



Road to Emergent Spacetime

From Quantum Mechanics to Quantum Gravity

Master's thesis for Physics and Astronomy Torbjörn Nilsson

Department of Physics CHALMERS UNIVERSITY OF TECHNOLOGY Gothenburg, Sweden 2019

MASTERS THESIS 2019

Road to Emergent Spacetime

From Quantum Mechanics to Quantum Gravity

Torbjörn Nilsson

Department of Physics Division of Mathematical Physics CHALMERS UNIVERSITY OF TECHNOLOGY Gothenburg, Sweden 2019 Road to Emergent Spacetime Authors: Torbjörn Nilsson Contact: torlachnil@gmail.com

© Torbjörn Nilsson, 2019.

TIFX61- Master's Thesis at Fundamental Physics (60 hp)

Supervisor: Martin Cederwall, Division of Mathematical Physics **Examiner:** Martin Cederwall, Division of Mathematical Physics

Department of Fundamental Physics Chalmers University of Technology SE-412 96 Gothenburg Telephone +46 31 772 1000

Cover: Factors in Hilbert space giving rise to an emergent spacetime. Deep field image credit: NASA, ESA, R. Ellis (Caltech), and the HUDF 2012 Team, edited by Torbjörn Nilsson.

Typeset in LATEX Printed by Chalmers Reproservice Gothenburg, Sweden 2019

Abstract

A long-standing issue in theoretical physics has been the difficulty of uniting Quantum Field Theory (QFT) and general relativity into a single theory of everything. There is ample evidence that reality is fundamentally quantum mechanical, and as such there should exist a quantum theory of gravity. Emergent spacetime is a novel approach to quantum gravity, wherein the usual method of starting with a classical theory and applying some kind of quantization recipe is reversed. Instead one begins with an abstract quantum theory and 'geometrizes' it using recently discovered relationships between entanglement and geometry.

This thesis is divided into three parts. The first part provides a comprehensive review of QFT, quantum information theory, string theory and how the AdS/CFT correspondence points towards an equivalence between spacetime connectivity and quantum entanglement. In the second part modern technical developments regarding the explicit emergence of spacetime from the entanglement structure of quantum states is reviewed. The main discovery that is reviewed is the explicit, fully controlled, emergence of second-order perturbative gravity with extra standard model fields from entanglement dynamics in conformal field theory. This result is then extended to include quantum corrections to the emergent gravitational theory by relating entanglement in the quantum theory to wormholes in the gravitational theory via the ER=EPR conjecture. The review is finished with an extension of the emergent spacetime program to discretized spacetimes, using the example of the AdS/MERA correspondence, which relates discretized anti-de Sitter to the MERA tensor network. In the third part some first steps towards the recovery of perturbative third-order gravitational dynamics from entanglement are carried out. The results show no inconsistencies, and the next step towards novel results is the characterization of the OPE between conformal primary scalars and the stress energy tensor in terms of symplectic forms in an auxiliary AdS space.

Acknowledgements

First and foremost I would like to thank my supervisor Martin Cederwall. This project has been an amazing opportunity to explore physics on my own terms, and your insight has always helped me not get lost on the way. I also want to thank Magdalena Eriksson for answering my questions, the countless discussions and the extensive feedback on my thesis. In addition, I want to thank my fellow thesis students: Daniel Erkensten, Simon Ekhammar and Sebastian Bergström, for fika, discussion and help finding errors in my references. I should also extend a thank you to the physics Discord server for feedback, encouragement, and answering several of my questions. Finally, I would like to extend a final thank you to Bengt Nilsson, Donald Marolf and (especially) Kostas Skenderis for taking the time to answer questions that I sent their way.

Contents

1	Intr	oducti	ion	1			
Ι	Pre	elimina	aries	3			
2	Quantum Information Theory						
	2.1	Quant	um Theory	5			
		2.1.1	Stochastic Quantum States	6			
		2.1.2	Multipartite Quantum States	9			
		2.1.3	Purification	10			
		2.1.4	Stochastic Quantum Operators	12			
	2.2	Inform	nation Theory	16			
		2.2.1	Shannon Entropy	17			
		2.2.2	Joint Entropy	18			
		2.2.3	Mutual Information	19			
		2.2.4	Relative Entropy	20			
	2.3	Quant	um Information Theory	21			
		2.3.1	Classical information from quantum systems	21			
		2.3.2	Quantum Entropy	22			
		2.3.3	Joint von Neumann Entropy	25			
		2.3.4	Conditional von Neumann Entropy	26			
		2.3.5	Quantum Mutual Information	26			
		2.3.6	Quantum Relative Entropy	27			
		2.3.7	A Quantum Second Law of Thermodynamics	28			
3	Elements of Quantum Field Theory 31						
	3.1	Path I	Integral QFT	31			
		3.1.1	Path Integral Formulation of Nonrelativistic Quantum Mechanics	31			
		3.1.2	From Quantum Mechanics to QFT	36			
		3.1.3	Applying the Path Integral Formalism	39			
		3.1.4	Formal Representation of States	45			
	3.2	Therm	nal Field Theory	48			
		3.2.1	Thermodynamical Ensembles	49			
		3.2.2	Thermal Field theory	51			
	3.3	The R	enormalization Group	55			
	3.4	Confo	rmal Field Theory	59			
		3.4.1	Conformal Algebra	59			
		3.4.2	Conformal Algebra for $d > 2$	61			
		3.4.3	Conformal Field Theory in General Dimension	63			
		3.4.4	Advanced Topics in 2d CFT	69			

4	\mathbf{Ent}	anglement Entropy of Quantum Fields 79				
	4.1	Entanglement Entropy in Axiomatic QFT				
		4.1.1 The Reeh-Schlieder Theorem				
		4.1.2 Reeh-Schlieder Corollary				
		4.1.3 Von Neumann Algebras in QFT				
		4.1.4 Tomita-Takesaki theory				
		4.1.5 Finite-Dimensional Case of Tomita-Takesaki Theory				
	4.2	Examples in QFT				
		4.2.1 Path Integral Approach to Rindler Space				
		4.2.2 Bisogano-Wichmann Approach to the Rindler Wedges				
		4.2.3 Semiclassical Black Hole Entropy				
		4.2.4 Ball-Shaped Regions in CFT				
	4.3	Entanglement Entropy in CFT				
	1.0	4.3.1 UV-divergences in the Entropy				
		4.3.2 BG-flow of Central Charge in 2d 120				
		4.3.3 Beplica Method 123				
5	Elei	nents of String Theory 131				
	5.1	Bosonic String Theory				
		5.1.1 String Perturbation Theory				
		5.1.2 String Theory Vertex Operators				
		5.1.3 Emergence of Gravity as a Massless Excitation				
	5.2	Superstring Theory				
		5.2.1 Superstring Low Energy Effective Action				
6	Introducing the AdS/CFT Duality 15					
	6.1	Introduction to Anti de Sitter and its Conformal Boundary				
	6.2	AdS/CFT Duality Piece by Piece				
		6.2.1 Scalar Field/CFT Correspondence				
		6.2.2 Yang-Mills/CFT Correspondence				
		6.2.3 Gravity/CFT Correspondence				
		6.2.4 Type IIB Supergravity				
	6.3	Why the Duality is Only Conjecture				
7	Hal	ographia Entangloment 101				
'	7 1	AdS/CFT at Finite Temperature 101				
	1.1	7.1.1 Spacetime Thermodynamics				
		7.1.1 Spacetime Thermodynamics				
		7.1.2 Add Diack Diane/ Infermal STM Duanty				
	7 9	The Dun Takayanagi (PT) Formula				
	1.2	1100 Kyu-1akayanagi (K1) Formula				
		7.2.2 Equivalence Detrice Denline Construction and Minimum 198				
		(.2.2 Equivalence Between Replica Construction and Minimal Area 200				
		(.2.5 Extremal Bulk Surface Ending on Spheres				
		(.2.4 Holographic Proof of Strong Subadditivity				
	7.3	Hubeny-Kangamani-Takayanagi (HRT) Formula				
	_ :	7.3.1 Heuristic Gravitational Derivation of HRT Formula				
	7.4	Entanglement Builds Geometry				

Π	Recent Developments	211
8	Universal Gravity from Entanglement Dynamics 8.1 Entanglement First Law 8.2 Einstein Equations From from CFT Entanglement First Law 8.2.1 Linearized Einstein Field Equations from $\delta S = \delta E$ 8.3 Quantum Fisher Information 8.4 Nonlinear Gravity from Fisher Information 8.4.1 Gravitational Side of the Derivation 8.4.2 CFT Side of the Derivation 8.4.3 Combining the Results	215 . 215 . 217 . 218 . 223 . 226 . 228 . 228 . 236 . 252
9	 Entanglement is Not Enough 9.1 Eternal Black Hole Equals Wormhole	255 . 256 . 263 . 267 . 273 . 273 . 277 . 284 . 284 . 291
10	 D Beyond Holography, de Sitter 10.1 It From BC Bit	295 . 296 . 298 . 299 . 303 . 309
Π	I Original Work	315
11	Gravity Emerging to Third Order 11.1 Gravitational Side 11.2 CFT Side	317 . 317 . 320
12	2 Summary & Conclusions	325
A	ppendices	i
A	Miscellaneous Proofs and Calculations A.1 Vanishing of equation (8.131) A.2 Derivation of Null Singularities of Wightman Propagators A.3 Naive Third Order Perturbation of Relative Entropy	iii . iii . iii . iv
Bi	ibliography	xvii

Chapter 1

Introduction

A long-standing issue in theoretical physics has been the difficulty of uniting quantum field theory (QFT) and general relativity (GR) into a single theory of everything. The problem is that GR becomes ill-defined after the quantization procedure that is central to QFT, meaning we have no quantum theory of gravity that may be related to the framework underlying the standard model.

Emergent spacetime is a novel approach to quantum gravity, wherein the usual method of starting with a classical theory and applying some kind of quantization recipe is reversed. Instead one begins with an abstract quantum theory and 'geometrizes' it using recently discovered relationships between entanglement and geometry. The main idea is that in contrast to the standard model forces, gravity may not be a theory of fundamental interactions. Instead gravity may be an emergent property of a priori nongeometric quantum systems.

The connection between entanglement and geometry has its roots in the famous result of Bekenstein and Hawking that black holes have entropy proportional to their surface area [1]. This entropy may be interpreted as being due to the exterior spacetime being entangled with the interior of the black hole. A direct consequence of the black hole entropy discovery was the realization that black holes had to maximize the entropy per volume to prevent the gravitational collapse of hot clouds violating the first law of thermodynamics. Based on the area-law entropy bound of gravitational systems it was conjectured that their physics may be encoded in a theory living on the boundary of that system, the logic being that entropy should be extensive in the fundamental degrees of freedom [2]. This holographic principle was given a precise interpretation in string theory in terms of the Anti de Sitter/Conformal Field Theory (AdS/CFT) duality, relating a string theoretical description of quantum gravity in spacetime to conformally symmetric QFT living on the boundary of that spacetime [3]. Although the AdS/CFT duality was discovered in a string theoretical context it has since been realized that it might hold more generally. These hints of a relation between quantum physics and gravity were given a much more profound meaning in 2006 with the Ryu-Takayanagi conjecture [4, 5]. Specifically the Ryu-Takayanagi prescription relates the entanglement entropy (a quantitative measure of entanglement) between complementary regions on the boundary, as computed in the boundary QFT, to the area of particular surfaces in bulk.

In 2009 Raamsdonk further realized the significance of this relation in combination with an AdS/CFT result called the *geodesic approximation* which, together with an upper bound on correlation functions from a quantity called mutual information, tells us that as entanglement between a region A and its complement B decreases the geometric distance in the bulk between points in A and B increases. A picture emerges in which spacetime begins to split as we try to disentangle parts of the boundary theory, finally resulting in a complete disconnection as entanglement is brought to zero, as imaged in figure 1.1. The remarkable conclusion is that entanglement in the quantum theory is necessary for there to be a connected spacetime at all.



Figure 1.1: Boundary of AdS represented as a sphere divided into two hemispheres. As we try to decrease the entanglement between these regions, the spacetime starts to pinch off and eventually splits completely.

The seeming connection between quantum entanglement and spacetime connectivity has inspired a new program of *quantum first* or *emergent* gravity, in which gravity is taken to emerge from the entanglement structure of some quantum theory. There are a number of approaches to such an emergent formulation of gravity, all of which take inspiration from the Ryu-Takayanagi formula [6, 7, 8, 9, 10]. The collection of all of these is referred to in this text as the field of 'emergent spacetime'. The aim of the present text is to lay out the path to these developments and then contribute with something new.

Structure and Purpose

The field of emergent spacetime is relatively new, and I find that there is a lack of approachable literature. My goal is to build a comprehensive text, starting at the end of introductory courses in gravity and QFT, building all the way up to modern research literature on emergent gravity. The text is divided into three parts, Preliminaries, Recent Developments and Original Work. All chapters depend on all preceding chapters with the exception of chapters 4 and 5 which may be read in any order.

In Part I: *Preliminaries* the relevant elements of quantum field theory, quantum information theory, string theory and the AdS/CFT duality are introduced. After this we propose and later derive the Ryu-Takayanagi relation. Finally it is shown more explicitly that if one tries to disentangle the local degrees of freedom of the boundary spacetime, the bulk- and boundary spacetimes split apart.

In Part II: *Recent Developments* the current state of the field of "emergent gravity" is reviewed. The three main topics covered are: the emergence of gravitational equations of motion from relative entropy constraints, the insufficiency of pure entanglement for a complete reconstruction of spacetime and tensor network models of emergent gravity. Tensor networks and proposals beyond entanglement involve more conjecture but also connect more closely to our own seemingly asymptotically de Sitter universe.

In Part III: Original Work some work towards the derivation of third order gravitational dynamics from boundary entanglement dynamics is done.

The aim of this text is first and foremost to be pedagogical. In calculations I make an effort to not omit any nontrivial steps, and if I do omit anything the idea is to state it clearly. In an effort to avoid what comes across as elitist practices in the literature, I make a conscious effort to not refer to things as easy. I also try to avoid the common malpractice of citing results without translating conventions. A reader that has read Part I: *Preliminaries* should be ready to understand contemporary literature relating to holographic entanglement entropy without having to consult external sources. After reading Part II: *Current Developments* the reader should be familiar with the details of some of the most recent developments in the area of emergent spacetime.

Part I Preliminaries

Chapter 2

Quantum Information Theory

Information theory is the quantitative study of the information content of random variables $x \in X$ sampled from some probability distribution p(x), where X is some set of values (an 'alphabet') for x. In the case where you wish to communicate information in terms of binary numbers, the expected information S required to unambiguously represent a microstate (x) of a system in some macrostate (p(x)) is

$$S = -\sum_{x} p(x) \log_2 p(x) , \qquad (2.1)$$

where the number 2 in the logarithm is due to the choice of measuring information in units of bits. To understand why this is the expected information content, let us consider a string *aaabbababb*... of messages a and b that appear with probabilities p, 1 - p. For a large number of messages N, a appears almost exactly p times. The number of possible messages is then

$$\frac{N!}{(pN)!((1-p)N)!} \approx \frac{N^N}{(pN)^{pN}((1-p)N)^{(1-p)N}} = \frac{1}{(p)^{pN}((1-p)N)^{(1-p)}} = 2^{NS}$$
(2.2)

where we have used Stirling's formula and $\sum_{x} x \log_2 x = -p \log_2(p) - (1-p) \log_2(1-p)$. The number of messages is then equal to the number of possible values that can be represented by NS bits and dividing by N we get the average length of a message given in bits.

The distribution p(x) is a "macrostate" since it reflects everything we know for certain about the system, while the individual values x are the individual microstates that we cannot predict exactly. In thermodynamics, the probability distribution function p(x) is usually determined entirely in terms of a few parameters, such as temperature and volume, and we call the values of these variables the macrostate.

The extension of classical information theory to the quantum case requires the extension of statistical physics to quantum mechanics. This is the subject matter of section 2.1. In section 2.2 we cover the core tools of classical information theory before generalizing these tools to the quantum case in section 2.3. We conclude this chapter by using the monotonicity of quantum relative entropy to derive a form of quantum second law of thermodynamics.

Throughout this chapter we try to note some specific examples of quantum informational connections to gravitational physics via the AdS/CFT correspondence.

2.1 Quantum Theory

Quantum mechanics treats the time evolution of the Schrödinger wavefunction or Heisenberg operators, and perfect knowledge of the state and the form of all operators is usually assumed. In this chapter we develop in a fairly minimal way a theory of *noisy quantum mechanics*, where we assume imperfect knowledge of the state and the form of the operators with which we act on the state. Much like in statistical mechanics, this is done by taking a statistical approach to quantum mechanics, i.e. by asking "What if my wavefunctions were randomly sampled from a distribution?". More explicitly, we develop the machinery of density operators, entanglement, purification and Choi/Kraus operators.

2.1.1 Stochastic Quantum States

In quantum mechanics, the only objects we compute are expectation values on the form $\langle \psi | A | \psi \rangle$, where ψ is some normalized wavefunction. If the state we want to measure comes from a probability distribution $\sigma(x)$ the expectation value becomes

$$\langle A \rangle = \sum_{x} \sigma^{2}(x) \langle \psi_{x} | A | \psi_{x} \rangle \quad .$$
(2.3)

It is in general most convenient to define $p = \sigma^2$, and the macrostate as the density matrix ρ defined as

$$\rho = \sum_{x} p(x) |\psi\rangle_{x} \langle\psi|_{x} , \qquad (2.4)$$

which is to be understood as the quantum generalization of a probability density function. The sum over x of all p(x) must be equal to one, since the probability of being in a state at all should be unity. We will show this normalization explicitly in a following paragraph. This reduces to a classical statistical description if all of the states ψ_x are orthogonal to each other so that projective measurements can perfectly distinguish them.

To obtain the expectation value of an operator A with respect to the density operator we compute

$$p(x) \langle \psi |_{x} A | \psi \rangle^{x} = \operatorname{Tr}[Ap(x) | \psi \rangle_{x} \langle \psi |^{x}]$$

= $\operatorname{Tr}[A\rho] ,$ (2.5)

and the probability of a particular eigensolution with eigenvalue a_j is obtained by finding the expectation value of a projection operator Π_j that projects onto the eigenspace of A with eigenvalue a_j . The projection operators must satisfy a completeness relation:

$$\sum_{j} \Pi_{j} = \mathbf{I} , \qquad (2.6)$$

where \mathbf{I} is the identity operator. This last example corresponds to a projective quantum measurement. Since we can obtain the probabilities for each of the results of a quantum measurement directly from the density operator, we say that it fulfills the role of the quantum state in the statistical description.

It is worth noting that while thinking about projective measurements is very useful for the operational interpretation of certain quantities and theorems in quantum information theory, they are in a sense unphysical since they break unitarity. Essentially, an experimenter performing a projective measurement should themselves be a solution of the equation of motion for the universe, so nothing non-unitary should ever be able to happen. As a quick check that projective measurements are not unitary, realize that projections are not invertible and therefore do not have unit determinant.

It is a popular view in quantum cosmology that that the evolution of the universe is always unitary, and the apparent unitarity breaking property of measurements are a result of the state of the universe becoming entangled with the measurement outcome [11, 8]. This view inevitably leads to the conclusion that our classical experience, in which the projective view applies to experiment, is but one of many terms in an entangled superposed state of the universe wavefunction. The additional terms (or *branches*) in the superposition, corresponding to different measurement outcomes, may be interpreted as additional classical universes. The theory that describes the behaviour of the wavefunction in such a paradigm as well as how the branches of the wavefunction become independent of each other is called decoherence theory. At the current time decoherence theory does not provide a fully rigorous alternative to the projective interpretation of experimental outcomes, but at least it is compatible with unitary quantum gravity.

Returning to density operators, there are a number of important properties that we should show. The density operator has unit trace:

$$Tr[\rho] = Tr[p(x) |\psi\rangle_x \langle \psi|^x]$$

= $p(x)Tr[|\psi\rangle_x \langle \psi|^x]$
= $p(x) \langle \psi|^x |\psi\rangle_x$ (2.7)
= $p(x)\delta^x_x$
= 1,

where we have used that $|\psi\rangle$ is normalized.

The density operator is Hermitian:

$$\rho^{\dagger} = (p(x) |\psi\rangle_x \langle \psi|^x)^{\dagger}$$

= $p(x)(\langle \psi|_x)^{\dagger} (|\psi\rangle^x)^{\dagger}$
= $p(x) |\psi\rangle_x \langle \psi|^x$, (2.8)

since p(x) is real.

The density operator is positive semi-definite, meaning that for any wavefunction φ , we have $\langle \varphi | \rho | \varphi \rangle \ge 0$, which can be shown as follows:

$$\langle \varphi | \rho | \varphi \rangle = \langle \varphi | p(x) | \psi \rangle_x \langle \psi |^x | \varphi \rangle$$

= $p(x) \langle \varphi | | \psi \rangle_x \langle \psi^x | | \varphi \rangle$
= $\sum_x p(x) | \langle \varphi | | \psi \rangle_x |^2 .$ (2.9)

It is clear that the last line here is positive definite since both the probabilities p(x) and the absolute squares must be either positive or zero.

Another important result is that the time evolution of the density in terms of the unitary time evolution operator $U(t,t_0)$ is given by

$$\rho(t) = U(t, t_0)\rho(t_0)U^{\dagger}(t, t_0),$$

meaning the density operator evolves in time like a Heisenberg operator. This unitary time evolution operator is given as usual in terms of the Hamiltonian by

$$U = e^{-iHt} \; .$$

This applies for any other operator that is defined in the Schrödinger picture, under the same operator the density operator transforms as a Heisenberg operator. The density operator does not uniquely specify the ensemble of states that gave rise to it. For example, if we have a two-level system with eigenstates $|0\rangle$, $|1\rangle$ we can have two distinct ensembles of states that span the Hilbert space:

$$\begin{split} p_{\varphi}(1) &= 1/2 \ , |\varphi\rangle_1 = |0\rangle \\ p_{\varphi}(2) &= 1/2 \ , |\varphi\rangle_2 = |1\rangle \\ p_{\psi}(1) &= 1/2 \ , |\psi\rangle_1 = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \\ p_{\psi}(2) &= 1/2 \ , |\psi\rangle_2 = \frac{|0\rangle - |1\rangle}{\sqrt{2}} \end{split}$$

These define completely different in-going wavefunctions, but the density operator is the same:

$$p(x)_{\psi} |\psi\rangle_{x} \langle\psi|^{x} = \frac{1}{2} \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \frac{\langle0| + \langle1|}{\sqrt{2}} + \frac{|0\rangle - |1\rangle}{\sqrt{2}} \frac{\langle0| - \langle1|}{\sqrt{2}} \right)$$

$$= \frac{1}{4} \left(|0\rangle\langle0| + |1\rangle\langle1| + |1\rangle\langle0| + |0\rangle\langle1| + |0\rangle\langle0| + |1\rangle\langle1| - |1\rangle\langle0| - |0\rangle\langle1| \right)$$

$$= \frac{1}{4} \left(2 |0\rangle\langle0| + 2 |1\rangle\langle1| \right)$$

$$= \frac{1}{2} |0\rangle\langle0| + \frac{1}{2} |1\rangle\langle1|$$

$$= p(x)_{\varphi} |\varphi\rangle_{x} \langle\varphi|^{x} .$$
(2.10)

This result has interesting implications for quantum information theory, we will see in section 2.3.2.

By the spectral theorem any density operator can be decomposed in terms of eigenvectors $|\phi_x\rangle_x$, $x \in \{0,1...,d-1\}$ where d is the dimension of the Hilbert space. Such a decomposition is called a *Schmidt decomposition*. This means that we are always able to define a "canonical" ensemble corresponding to the density operator. This decomposition is still not unique if the spectrum of the Hamiltonian is degenerate, since in this case the eigenvectors of the system are not uniquely defined. In quantum data compression, which we will not cover in this thesis, any ensemble that is obtained from the spectral theorem is also "maximally efficient" for information storage.

The purity $P(\rho)$ of a density operator is given by

$$P(\rho) \equiv \operatorname{Tr}[\rho^{\dagger}\rho] = \operatorname{Tr}[\rho^{2}],$$

and it is a measure of the noisiness of a quantum state. The purity of a pure (completely definite) state is exactly equal to one, while a mixed state has a purity that is strictly smaller than 1. This is straightforward to check by just diagonalizing the two factors of ρ with the same basis:

$$P(\rho) = \operatorname{Tr} \left[p(x)p(y) |\psi\rangle_x \, \delta^x{}_y \, \langle \psi|^y \right]$$

= $\sum_x p(x)^2$
 ≤ 1 , (2.11)

where we have used that the eigenvalues of ρ are the probabilities p(x). A simple argument for the inequality is that the sum of the p(x) is equal to 1, and squaring them can only make the sum smaller since $0 \le p(x) \le 1$, $\forall x$.

The purity is minimized by the maximally mixed state, often denoted Π , which is given by $\rho = \text{Diag}(1/N)$ where N is the dimension of the Hilbert space in which the state lives. This is

quite reminiscent of the entropy, where the highest entropy state for a system with finite degrees of freedom is given by the uniform distribution, and we will find that purity and entropy are related in section 2.3.2.

2.1.2 Multipartite Quantum States

Possibly the most important aspect of any quantum theory is the case of two or more quantum systems that are joined together. Multipartite systems are the source of most strange quantum effects, all of which have to do with entanglement.

Let us first consider the absolutely simplest case of two independent ensembles. Let us label them $\{p(x), |\psi\rangle_x\}, |\psi\rangle \in \mathcal{H}_A, \{p(y), |\phi\rangle_y\}, |\phi\rangle \in \mathcal{H}_B$, where the \mathcal{H} denote two different Hilbert spaces. The composite density operator ρ_{AB} is then given by the tensor product:

$$\rho_{AB} = \left(p(x) |\psi\rangle_x \langle \phi|^x \right) \otimes \left(p(y) |\phi\rangle_y \langle \psi|^y \right)
= p(x)p(y) \left(|\psi\rangle_x \langle \psi|^x \otimes |\phi\rangle_y \langle \phi|^y \right)
= \rho_A \otimes \rho_B$$
(2.12)

where we in the last line have just indicated that the composite density operator admits a very simple description in terms of the density operators of the subsystems (ρ_A, ρ_B) in the independent case. A state that can be decomposed in this manner is called a *separable state*. It can also be shown that any convex linear combination of separable states is also a separable state. Thus, for $\sum_x p(x) = 1$ we have that a state ρ in the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2 \ldots \otimes \mathcal{H}_N$ is separable if it can be written as a tensor product state:

$$\rho = \sum_{x=1}^{N} p(x)\rho_1^x \otimes \rho_2^x \dots \otimes \rho_N^x .$$
(2.13)

If this decomposition is not possible a system is said to be *entangled*. Note that the amount of entanglement and even whether there is entanglement at all depends on the choice of subspaces $\mathcal{H}_1 \ldots \mathcal{H}_N$, so entanglement is always defined relative to some choice of subsystem decomposition. Canonically the decomposition is made into a set of Hilbert spaces that are associated to different spacelike separated regions in spacetime, typically a set of different labs making measurements of the entangled state at the same time.

Local Density Operator

When given the global state defined on $\mathcal{H}_A \otimes \mathcal{H}_B$ it is often interesting to know what the density operator looks like for an observer that only has access to part of the system, for example \mathcal{H}_A .

The method to obtain a local density operator is called a *partial trace*, which as it sounds is a trace operation performed only on part of the Hilbert space. The partial trace of the operator X_{AB} over subsystem B is defined as follows:

$$\operatorname{Tr}_{B}[X_{AB}] \equiv \sum_{l} \left(\operatorname{id}_{A} \otimes \langle l|_{B} \right) X_{AB} \left(\operatorname{id}_{A} \otimes |l\rangle_{B} \right)$$
(2.14)

where $|l\rangle_B$ is any orthonormal basis for \mathcal{H}_B . This is indeed just a different way to write the trace operation since the trace is invariant under orthogonal transformations, meaning that

$$\operatorname{Tr}[X_B] = \sum_{l} \operatorname{Tr}[|l\rangle\langle l| X_B |l\rangle\langle l|] = \sum_{l} \langle l| |l\rangle\langle l| X_B |l\rangle = \sum_{l} \langle l| X_B |l\rangle .$$

The partial trace operator is interesting because it reduces a state on a larger Hilbert space to a smaller space. In addition to this it preserves local measurement outcomes, which is what makes it a good method for obtaining a local density operator. This follows quite simply, recall that the expectation of an operator Λ is $\text{Tr}[\Lambda\rho]$, so let us consider a local measurement on subsystem A, given by the operator $\Lambda_A \otimes \text{id}_B$. We wish to show that

$$\mathrm{Tr}[(\Lambda_A \otimes \mathrm{id}_B)\rho_{AB}] = \mathrm{Tr}_A \Big[\Lambda_A \mathrm{Tr}_B[\rho_{AB}]\Big]$$

A way to show this equality is to introduce the two orthonormal bases $|j\rangle_A$, $|k\rangle_B$. We can then write the RHS as follows:

$$\operatorname{Tr}_{A}\left[\Lambda_{A}\operatorname{Tr}_{B}[\rho_{AB}]\right] = \sum_{j} \langle j | \Lambda_{A}\left(\sum_{k} \left(\operatorname{id}_{A} \otimes \langle k |_{B}\right)\rho_{AB}\left(\operatorname{id}_{A} \otimes |k\rangle_{B}\right)\right) | j \rangle$$
$$= \sum_{j} \sum_{k} \left(\left(\operatorname{id}_{A} \langle j | \Lambda_{A} \otimes \langle k |_{B}\right)\rho_{AB}\left(\operatorname{id}_{A} | j \rangle \otimes |k\rangle_{B}\right)\right)$$
$$= \sum_{j} \langle j | \Lambda_{A}\rho_{A} | j \rangle$$
(2.15)

Where we in the last step have used the definition of partial trace. An interesting note to make is that the local density matrices ρ_A and ρ_B do not uniquely specify ρ_{AB} , since ρ_{AB} has more degrees of freedom.

2.1.3 Purification

In the following we denote the space of density operators on \mathcal{H} by $\mathcal{D}(\mathcal{D})$. Purification of a quantum state $\rho_A \in \mathcal{D}(\mathcal{H}_A)$ is an operation in which you entangle ρ_A with an environment E such that the state $\rho_{AE} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_E)$ is pure. It is always possible to purify a state by entangling it with an environment. This allows for a universe that is in an overall pure state having impure subsystems. More specifically, the density operators associated with local regions in a pure universe look like a thermal ensemble ($\rho_A = e^{-\beta H}$). In fact, we will learn in chapter 4 that thermality is an unavoidable consequence in any spacetime that contains an event horizon. An event horizon is a surface that prevents an observer from ever seeing the part of the universe that is behind it, so an observer who sees an event horizon is the spacetime analog of the observer with access to only subsystem A in the previous section.

Purification and the process of choosing a purification scheme are important because they may serve to decouple entanglement and statistical entropy, as is covered in for example [12]. Essentially if the overall state is pure, then the full entropy of a subsystem may be attributed to entanglement. If not, then the entropy of a given subsystem is a difficult mix of thermal and entanglement entropy. The distinction between entanglement and thermal entropies is important for example because only entanglement entropy is thought to be associated with geometry in the Ryu-Takayanagi conjecture. The purification of a system is not unique, and we will give two possible purifications now. The first is a "simplest possible" approach, the other is called canonical purification.

Simple Purification

Let us suppose that we have a local density operator

$$\rho_A = p(x) \left| \psi_A \right\rangle^x \left\langle \psi_A \right|_x$$

where $|x\rangle$ is an orthonormal basis of \mathcal{H}_A . A purification of this state is then given by

$$\rho_{AE} \equiv \sum_{x} \sum_{y} \sqrt{p(x)p(y)} \Big(|\psi_A\rangle^x \otimes |\psi_E\rangle^x \Big) \Big(\langle\psi_A|_y \otimes \langle\psi_E|_y \Big) ,$$

where $|\psi_E\rangle^x \in \mathcal{H}_E$ are orthonormal. The only restriction on \mathcal{H}_E here is that it must be of equal or higher dimension than \mathcal{H}_A . To show that this state is indeed a purification we need to check that it reduces to ρ_A under the partial trace over E, as well as confirming that it has purity $P(\rho_{AE}) = 1$.

Performing a partial trace over E gives

$$\operatorname{Tr}_{E}[\rho_{AE}] = \sum_{x} \sum_{y} \sum_{l} \sqrt{p(x)p(y)} \left(|\psi_{A}\rangle_{x} \langle \psi_{A}|^{y} \right) \otimes \left(\langle l| |\psi_{E}\rangle_{x} \langle \psi_{E}|^{y} |l\rangle \right) = \sum_{x} p(x) |\psi_{A}\rangle_{x} \langle \psi_{A}|^{x} ,$$

$$(2.16)$$

where we have used that we can pick the $|l\rangle$ such that $\langle l||\psi\rangle_x = \delta_x^l$.

We also wish to show that this state is pure by computing the purity:

$$P(\rho_{AE}) = \operatorname{Tr}\left[\left(\sum_{x,y} \sqrt{p(x)p(y)}(|\psi_{A}\rangle^{x} \otimes |\psi_{E}\rangle^{x})(\langle\psi_{A}|_{y} \otimes \langle\psi_{E}|_{y})\right)\right] \\ \times \left(\sum_{z,v} \sqrt{p(z)p(v)}(|\psi_{A}\rangle^{z} \otimes |\psi_{E}\rangle^{z})(\langle\psi_{A}|_{v} \otimes \langle\psi_{E}|_{v})\right)\right] \\ = \sum_{x,y,z,v} \sqrt{p(x)p(y)p(z)p(v)} \\ \operatorname{Tr}\left[(\langle\psi_{A}|_{v} \otimes \langle\psi_{E}|_{v})(|\psi_{A}\rangle^{x} \otimes |\psi_{E}\rangle^{x})(\langle\psi_{A}|_{y} \otimes \langle\psi_{E}|_{y})(|\psi_{A}\rangle^{z} \otimes |\psi_{E}\rangle^{z})\right]$$

$$= \sum_{x} p(x)\left(\sum_{y,z} \sqrt{p(y)p(z)}(\langle\psi_{A}|_{y} \otimes \langle\psi_{E}|_{y})(|\psi_{A}\rangle^{z} \otimes |\psi_{E}\rangle^{z})\right) \\ = \sum_{y} p(y) \\ = 1.$$

$$(2.17)$$

Another, more common way to show that the state suggested is pure is to just observe that ρ_{AB} can be written as $|\psi\rangle \langle \psi|$ where

$$|\psi\rangle \equiv \sum_{x} \sqrt{p(x)} |\psi_A\rangle^x \otimes |\psi_E\rangle^x$$

Since this $|\psi\rangle$ is a definite state ρ_{AB} is pure by definition.

Canonical Purification

We let ρ_A be a general density operator, and we let $\sqrt{\rho_A}$ be its positive semi definite square root, such that $\sqrt{\rho_A}\sqrt{\rho_A} = \rho_A$. The canonical purification is defined in terms of the following wavefunction

$$|\Psi_{AE}\rangle = (\sqrt{\rho_A} \otimes \mathrm{id}_E) |\Gamma\rangle_{AE} , \qquad (2.18)$$

where $|\Gamma\rangle_{AE}$ is the unnormalized maximally entangled wavefunction, given in terms of orthogonal basis kets by

$$|\Gamma\rangle_{AE} = \sum_{i} |i\rangle_A \otimes |i_E\rangle \quad . \tag{2.19}$$

The corresponding density operator ρ_{AE} is given by

$$\rho_{AE} = |\Psi_{AE}\rangle \langle \Psi_{AE}| = \sum_{i} \sum_{j} (\sqrt{\rho_A} \otimes \mathrm{id}_E) (|i_A\rangle \otimes |i_E\rangle) (\langle i_A| \otimes \langle i_E|) (\sqrt{\rho_A} \otimes \mathrm{id}_E) , \qquad (2.20)$$

where we have used that ρ is hermitian. Let us confirm that this is actually a purification of ρ_A . We first check the partial trace over E

$$\operatorname{Tr}_{E}[\rho_{AE}] = \operatorname{Tr}_{E}\left[\sum_{i}\sum_{j}(\sqrt{\rho_{A}}\otimes \operatorname{id}_{E})(|i_{A}\rangle\otimes |i_{E}\rangle)(\langle j_{A}|\otimes \langle j_{E}|)(\sqrt{\rho_{A}}\otimes \operatorname{id}_{E})\right]$$
$$= \sum_{i}\sum_{j}\operatorname{Tr}_{E}\left[(\sqrt{\rho_{A}}|i_{A}\rangle\langle j_{A}|\sqrt{\rho_{A}})\otimes(|i_{E}\rangle\langle j_{E}|)\right]$$
$$= \sum_{i}(\sqrt{\rho_{A}}|i_{A}\rangle\langle i_{A}|\sqrt{\rho_{A}})$$
$$= \rho_{A},$$

where in the final step we used the completeness relation. The state ρ_{AE} is pure on account of just being the outer product of two wavefunctions.

2.1.4 Stochastic Quantum Operators

In addition to having imperfect information about the quantum state, we also have imperfect information about all of the operators that may act on the state. In the operator case it is not clear whether we can just sum over probabilities times different density operators and still end up with sensible operators. Before we think about how to combine a distribution of operators in a consistent manner, we need to understand what restrictions there are on a general quantum operator. To do this we will follow the so called *axiomatic approach* to quantum evolution. All operators, including projective measurements and time evolution are special cases of the formalism we will develop here. This section follows quite closely section 4.4 in [13].

The axiomatic approach starts from three axioms that should hold for any quantum process, and are used to find a set of mathematical constraints that are satisfied by any quantum process. The resulting constraints are given by the *Choi-Kraus theorem*, which we will after motivating and explaining the constraints. Operators that fulfill the Choi-Kraus criteria are colloquially referred to as *Quantum Channels*, and we will denote these by \mathcal{N} . We will define a quantum channel as a linear, completely positive and trace preserving map that acts on a Hilbert space $\mathcal{H}_{\mathcal{A}}$ and takes it to a Hilbert space $\mathcal{H}_{\mathcal{B}}$. The two Hilbert spaces need not be distinct.

The Choi-Kraus Representation Theorem

These axioms are motivated by the reasoning that the state is represented by the density operator, and that quantum states should evolve into other quantum states. This means that a Quantum Channel should be a map between density operators. The input density operator can be anything that fulfills the criteria for being a density operator, including pure states, mixed states or even one share of an entangled state.

To go from the axioms to the Choi-Kraus theorem we need to introduce some notation. We denote the space of all density operators on the Hilbert space \mathcal{H} as $\mathcal{D}(\mathcal{H})$. In addition to this, we denote the space of linear operators that act on the space \mathcal{H} by $\mathcal{L}(\mathcal{H})$ and we denote the space of linear operators that act on the space \mathcal{H}_A , \mathcal{H}_B as $\mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$.

Our first axiom is that the quantum channel \mathcal{N} has the following action: $\mathcal{N}(\rho_A) \in \mathcal{D}(\mathcal{H}_{\mathcal{B}})$ if $\rho_A \in \mathcal{D}(\mathcal{H}_{\mathcal{A}})$. To enforce this requirement, we demand that \mathcal{N} is convex linear when acting on $\mathcal{H}_{\mathcal{A}}$, meaning that

$$\mathcal{N}\left(\lambda\rho_A + (1-\lambda)\sigma_A\right) = \lambda\mathcal{N}\left(\rho_A\right) + (1-\lambda)\mathcal{N}\left(\sigma_A\right) , \qquad (2.21)$$

where $\rho_A, \sigma_A \in \mathcal{D}(\mathcal{H}_A)$. We can give this formal requirement a more intuitive interpretation. Imagine that perform N measurements. We sample the distribution $\mathcal{N}(\rho_A) \lambda N$ times and the distribution $\mathcal{N}(\sigma_A) (1 - \lambda)N$ times. If we do not know λ beforehand, we measure the distribution

$$\kappa_B = \mathcal{N} \left(\lambda \rho_A + (1 - \lambda) \sigma_A \right) \;.$$

If we are then told which measurements come from what distribution, we learn that the measured result actually corresponds to

$$\kappa_B = \lambda \mathcal{N}(\rho_A) + (1 - \lambda) \mathcal{N}(\sigma_A)$$

Knowing what distributions we were sampling at what time should not alter the outcome of an experiment, so the two values of κ_B must be the same. In principle this is just a form of linearity relation that we have to normalize to keep the unit trace property of the in-going density matrix intact. In fact, we go ahead to require that the quantum channel \mathcal{N} is linear in all arguments, not just density operators. By combining this requirement with the convex linearity on density operators we find that \mathcal{N} is a *positive map*, meaning that $\mathcal{N}(X_A)$ is positive semi-definite for all positive semi-definite X_A .

If we were dealing with a classical system positivity of the operator \mathcal{N} would be a sufficient requirement on the quantum evolution, but we want the operator \mathcal{N} to be positive definite when acting on only one share of an entangled state as well. This requires a stronger condition called *complete positivity*.

A map $\mathcal{N} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ is completely positive if the map $\mathrm{id}_R \otimes \mathcal{N}$ is positive for a reference system R of arbitrary size. That is, if the subsystem A is entangled with an arbitrary system R, the composite operator acting on $H_R \otimes H_A$ is positive for all R.

Let us make this slightly more explicit, so it is clear what is going on. Let X_{RA} denote an arbitrary state in the Hilbert space $\mathcal{H}_R \otimes \mathcal{H}_A$, ad let us act on it with the map $\mathrm{id}_R \otimes \mathcal{N}$. We can expand the state X_{RA} in the basis for \mathcal{H}_R according to

$$\sum_{i,j} \ket{i} ig\langle j
vert \otimes X_A^{i,j}$$
 .

and then then write out the action of our operator

(

$$id_{R} \otimes \mathcal{N})(X_{RA}) = (id_{R} \otimes \mathcal{N}) \left(\sum_{i,j} |i\rangle \langle j| \otimes X_{A}^{i,j} \right)$$

$$\langle \text{linearity} \rangle = \sum_{i,j} (id_{R} \otimes \mathcal{N}) \left(|i\rangle \langle j| \otimes X_{A}^{i,j} \right)$$

$$= \sum_{i,j} (id_{R}) \left(|i\rangle \langle j| \right) \otimes \mathcal{N} \left(X_{A}^{i,j} \right)$$

$$= \sum_{i,j} |i\rangle \langle j| \otimes \mathcal{N} \left(X_{A}^{i,j} \right)$$
(2.22)

In terms of this calculation, the complete positivity criteria asks if positive semi-definiteness of X_{RA} implies the positive semi-definiteness of $\sum_{i,j} |i\rangle \langle j| \otimes \mathcal{N}(X_A^{i,j})$. Writing the criterion on this more explicit form is sometimes useful.

Finally, as a tie-in to the fact that \mathcal{N} takes density operators to other density operators, we require that \mathcal{N} is *trace-preserving*. Essentially, this means that \mathcal{N} cannot act on an operator that is not a density and turn it into a density, assuming that the unit trace condition was the only thing separating the input operator and a density. Mathematically this means that

$$\operatorname{Tr}[X_A] = \operatorname{Tr}[\mathcal{N}(X_A)], \ \forall X_A.$$

It is worth noting that trace preservation together with linearity implies the convex linearity and density to density conditions.

With this, we have asserted three axioms that completely specify a valid quantum channel: linearity, complete positivity and trace preservation. We are now ready to state the Choi-Kraus theorem as follows:

Theorem 2.1.1 (Choi-Kraus Theorem) A map $\mathcal{N} : \mathcal{L}(\mathcal{H}_{\mathcal{A}}) \to \mathcal{L}(\mathcal{H}_{B})$ is linear, completely positive and trace preserving if and only if it has a Choi-Kraus decomposition as follows:

$$\mathcal{N}(X_A) = V^l X_A V_l^{\dagger} \tag{2.23}$$

where $X_A \in \mathcal{L}(\mathcal{H}_A), V \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$ for all $l \in (0, 1, ..., d - 1)$,

$$V^l V_l^{\dagger} = i d_A \tag{2.24}$$

and $d \leq \dim(\mathcal{H}_{\mathcal{A}})\dim(\mathcal{H}_{\mathcal{B}}).$

The statement of the Choi-Kraus theorem is very important because it characterizes a very general operator in a restrictive manner. In the context of quantum gravity, it finds its application for example in extracting the degrees of freedom of a quantum causal theory [14]. The actual proof of the Choi-Kraus theorem does not provide any particularly important insight in the context of this thesis, and is readily found in [13].

An important observation due to the Choi-Kraus theorem is that the naive way of generating a stochastic quantum operator by summing according to $A_{tot} = \sum_{x} p(x)A_x$, where all A_x are unitary operators, *almost works*. If we set

$$V_x = \sqrt{p(x)}A_x$$

we obtain

$$V^x V_x^{\dagger} = \sum_x p(x) A_x A_x^{\dagger} = \sum_x p(x) \operatorname{id} = \operatorname{id} ,$$

where we used unitarity and that the probabilities sum to 1. Therefore, by putting a square root in front of the probabilities we can formulate a simplest possible ensemble of quantum operators.

Properties of Quantum Channels

In this section we list some basic properties of a general quantum channel, which are easily found by writing them on the Choi-Kraus form.

Consider two quantum channels, $\mathcal{N}_{A\to B}$, $\mathcal{M}_{B\to C}$ where the subscripts denote the Hilbert spaces between which the channels map. The serial concatenation of these quantum channels is also a quantum channel:

$$\mathcal{M}_{B\to C}(\mathcal{N}_{A\to B}(X_A)) = \sum_k \sum_l V_{\mathcal{M},k} V_{\mathcal{N},l} X_A V_{\mathcal{N},l}^{\dagger} V_{\mathcal{M},k}^{\dagger} = \sum_k \sum_l \left(V_{\mathcal{M},k} V_{\mathcal{N},l} \right) X_A \left(V_{\mathcal{M},k} V_{\mathcal{N},l} \right)^{\dagger}$$

where the Choi-Kraus operators are given by summing over one of the indices, and observing that the object to the left of X_A is the V_l in the Choi-Kraus theorem above.

We may also consider a parallel action where we have the quantum channels $\mathcal{N}_{A\to B} \otimes \mathcal{M}_{C\to D}$ that acts on the bipartite state X_{AB} . The action of these channels may be written either as

$$\left(\mathcal{N}_{A\to B}\otimes \mathrm{id}_{D}\right)\left(\left(\mathrm{id}_{A}\otimes\mathcal{M}_{C\to D}\right)\left(X_{AB}\right)\right)$$

or $\left(\mathrm{id}_{B}\otimes\mathcal{M}_{C\to D}\right)\left(\left(\mathcal{N}_{A\to B}\otimes\mathrm{id}_{C}\right)\left(X_{AB}\right)\right)$.

The point here being that it does not matter for the final result in which order you perform a local measurement on the A, B subspaces.

Finally, we wish to find and understand the adjoint (generalization of hermitian conjugate) of the linear operator \mathcal{N} . The adjoint of a matrix A is the unique linear operator A^{\dagger} that satisfies the condition

$$\langle u|Av\rangle = \langle A^{\dagger}u|v\rangle$$

for any vectors u, v. It is also possible to define an inner product for operators, the so called Hilbert-Schmidt inner product, defined as

$$\langle A, B \rangle = \operatorname{Tr}[A^{\dagger}B]$$
.

The definition of the adjoint of the map $\mathcal{N} : \mathcal{L}_A \to \mathcal{L}_B$ can then be defined as the unique map \mathcal{N}^{\dagger} that fulfills

$$\langle A, \mathcal{N}(B) \rangle = \langle \mathcal{N}^{\dagger}(A), B \rangle$$
 (2.25)

for all $X \in \mathcal{L}_A, Y \in \mathcal{L}_B$.

Another important definition is a *unital map*, which is a map that preserves the identity operator, meaning that

$$\mathcal{N}(\mathrm{id}_A) = \mathrm{id}_B$$
.

Now let us investigate the properties of a quantum channel in terms of the adjoint and unital properties. We know that a quantum channel has a Choi-Kraus decomposition, so the left hand side of the adjoint condition evaluates as

$$\langle Y, \mathcal{N}(X) \rangle = \langle Y, V^l X V_l^{\dagger} \rangle$$

$$= \operatorname{Tr}[Y^{\dagger} V^l X V_l^{\dagger}]$$

$$= \operatorname{Tr}[V_l^{\dagger} Y^{\dagger} V^l X]$$

$$= \operatorname{Tr}[(V_l^{\dagger} Y^{\dagger} V^l)^{\dagger} X]$$

$$= \langle (V_l^{\dagger} Y V^l), X \rangle$$

$$(2.26)$$

and we can read off the definition

$$\mathcal{N}^{\dagger}(Y) = \left(\mathcal{N}(Y)\right)^{\dagger}$$
.

The only tricks used in this calculation are essentially the cyclicity of the trace and the definition of the Hilbert-Schmidt inner product in the first and last steps. It follows from the restriction on the V's from the Choi Kraus theorem,

$$V_l^{\dagger} V_l = \mathrm{id}_A, \ V_l^{\dagger} V^l = \mathrm{id}_B \ ,$$

that the quantum channel \mathcal{N} is also a unital map. In conclusion, \mathcal{N} is a self-adjoint unital map.

2.2 Information Theory

The field of classical information theory, or *Shannon theory*, was founded in 1948 by Claude Shannon. Shannon's great insight was the use of probability theory to define quantitative measures of something that was at the time rather vague - the "information content" of a random variable. Let illustrate Shannon's definition of information by way of an example.

Suppose we have a random variable X that can take as values the symbols $x \in \{a, b, c, d\}$, and we wish to communicate these to someone else by encoding them as binary numbers. The most obvious way in which we could encode the four possible values of the random variable X is as follows:

$$\begin{array}{l} a \rightarrow 00 \ , \\ b \rightarrow 01 \ , \\ c \rightarrow 10 \ , \\ d \rightarrow 11 \ . \end{array}$$

In this scheme, each possible value of the random variable X is captured by two binary numbers, or two *bits*. In addition, given a sequence of bits it is never ambiguous what it means, since we know every pair of bits in the sequence uniquely corresponds to a specific value. Let us now specify an actual distribution for X, assigning the probabilities p as follows:

$$p(a) = 1/2 ,$$

 $p(b) = 1/8 ,$
 $p(c) = 1/4 ,$
 $p(d) = 1/8 .$

The insight of Shannon was that because each of the possible values of a random variable do not have the same probability, there is a sense in which their appearance does not carry as much information about the underlying distribution. Shannon defined the *surprise* or *information content* (i(x)) of the symbol x as follows:

$$I(x) = -\log_2\left(p(x)\right)$$

where the log base 2 is a result of the fact that we have taken as our unit of information the binary number. This also means that I(x) has the unit of bits. Looking back at the probabilities of each of the symbols we see that according to this definition we have

$$I(a) = 1$$
, $I(b) = 3$, $I(c) = 2$, $I(d) = 3$. (2.27)

So, why is this information measure useful? The results in equation (2.27) tell us that the symbol *a* carries one bit of information, *c* carries two bits, and so forth. The optimal way to uniquely encode the output of the random variable X is to encode the symbols with the same number of bits as they "contain" according to Shannon's information measure. When we say optimal we mean optimal in the sense that the average number of bits per symbol communicated will be as small as possible. The following scheme gives such an optimal encoding

$$\begin{array}{l} a \rightarrow 0 \ , \\ b \rightarrow 110 \ , \\ c \rightarrow 10 \ , \\ d \rightarrow 111 \ . \end{array}$$

The difficulty of finding this particular encoding lies entirely in making sure that a message consisting of a long string of bits is never ambiguous. The expected length per symbol of this coding scheme is the sum of the products of the number of bits required to encode a symbol times the probability of it appearing:

$$\frac{1}{2} \cdot 1 + \frac{1}{8} \cdot 3 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 = \frac{7}{4} ,$$

which we see is less than the expected length of 2 bits for the "most obvious" way in which we encoded the information at the start of this section. What we have calculated here is actually the *expected information content* of the random variable X, and can be written as

$$\sum_{x} p(x)I(x) = -\sum_{x} p(x)\log_2(p(x)) .$$
(2.28)

This quantity is important enough to have its own name: the *entropy* or *Shannon entropy* of the random variable X. Interestingly, the entropy as defined in statistical mechanics seem very closely related to a notion of "information" that has a very intuitive operational interpretation.

2.2.1 Shannon Entropy

There are a number of properties of the entropy

$$S(X) = -\sum_{x} p(x) \log(p(x))$$

that we defined in the previous section. We leave the base 2 implicit here, because the entropy makes sense for any base in which we wish to measure information. In the quantum case is is natural to work with *dits*, which are elementary components that can take *d* different values corresponding to *d*-level quantum systems. In thermodynamics there is some sense in which we measure information in terms of e^{k_B} , 'Boltzmann units', instead of base 2.

Some properties of entropy have already been mentioned, such that the entropy is maximized by the uniform distribution (for finite degrees of freedom). In addition to this the entropy is bounded below by zero. We will now prove two more properties of the entropy that are invoked when proving important theorems. In quantum information theory, reference is often made to the classical proofs, so it is very valuable to spend some effort on this.

Additivity

If we have two ensembles with independent random variables X, Y described by probability distributions p(x), p(y), the Shannon entropy is given by

$$S(X,Y) = -\sum_{x,y} p(x)p(y)\log(p(x)p(y))$$

= $-\sum_{x,y} p(x)p(y) (\log p(x) + \log p(y))$
= $-\sum_{x,y} p(x)p(y)\log p(x) - \sum_{x,y} p(x)p(y)\log p(y))$
= $S(X) + S(Y)$, (2.29)

where we have used $\sum_{x} p(x) = 1$ and likewise for y.

Non-negativity

The entropy S(X) is always positive for any random variable X with probability density p(x). This follows simply, because $\log(p(x)) \le 0$ due to $0 \le p(x) \le 1$. The entropy is then a sum of positive terms.

Concavity

The entropy S(X) is concave in the probability density p(x). This means that given three random variables with associated probability densities according to

$$X_1, p_1(x)$$

$$X_2, p_2(x)$$

$$X_t, p_t(x) = qp_1(x) + (1-q)p_2(x), q \in [0,1]$$

we have that

$$S(X_t) \ge qS(X_1) + (1-q)S(X_2)$$

Let us think about this in terms of the individual terms in the sum that makes up H:

$$-p_t(x)\log(p_t(x)) \ge -qp_1(x)\log(p_1(x)) - (1-q)p_2(x)\log(p_2(x))$$

Without loss of generality, assume that $p_2(x) \ge p_1(x)$. Then we have that $p_1(x) \le p_t(x) \le p_2(x)$. The right hand side defines a function that has a constant derivative with respect to q, and we have equality when q is 0 or 1. If the left hand side has a negative definite second derivative with respect to the parameter q then the inequality must hold for all values of $q \in [0,1]$. This can be understood as follows: the LHS is always bending downwards, while the RHS is a straight line. If the LHS crosses the straight line twice it must be above the RHS at all points between the crossings. Let us show that the second derivative of the LHS is negative (dropping the explicit x-dependencies):

$$\frac{\partial^2}{\partial q^2} \left[-\left(qp_1 + (1-q)p_2\right) \log\left(qp_1 + (1-q)p_2\right) \right] \\
= \frac{\partial}{\partial q} \left[-\left(p_1 - p_2\right) \log\left(qp_1 + (1-q)p_2\right) - \left(qp_1 + (1-q)p_2\right) \frac{p_1 - p_2}{qp_1 + (1-q)p_2} \right] \\
= \frac{\partial}{\partial q} \left[-\left(p_1 - p_2\right) \log\left(qp_1 + (1-q)p_2\right) - (p_1 - p_2) \right] \\
= \frac{-(p_1 - p_2)^2}{qp_1 + (1-q)p_2} \\
\leq 0 ,$$
(2.30)

where the final inequality follows because $q \in [0,1]$, $p_2 \ge p_1 \ge 0$ and from the positivity of the square. With this, we have shown that the concavity of S(X) holds for each x, and therefore for the entire quantity.

2.2.2 Joint Entropy

We know that the total entropy of two independent random variables is additive, but what if one of the random variables is conditioned on the other? The joint entropy S(X,Y) is defined as

$$S(X,Y) = -\sum_{x,y} p_{x,y} \log(p_{x,y}) , \qquad (2.31)$$

where the arbitrary two variable joint distribution may be decomposed as $p_{x,y} = p_x p_{y|x} = p_y p_{x|y}$. The conditional probability $p_{y|x}$ is to be read as "the probability of getting y given x".

In addition to the conditional probability we can define a conditional entropy

$$S(Y|X) = -\sum_{x,y} p_{y|x} \log p_{y|x} .$$

The conditional entropy is positive definite and is the expected information content of the random variable Y given that you already know X. The conditional entropy is always lesser than or equal to the Shannon entropy since already having more information can only reduce the information obtained when you measure Y, meaning

$$S(Y) \ge S(Y|X)$$
.

In terms of the marginal entropies S(X), S(Y) we can show that

$$S(X,Y) = S(X) + S(Y|X)$$
, (2.32)

This is done by just computing explicitly

$$S(X,Y) = -\sum_{x,y} p_x p_{y|x} \log(p_x p_{y|x})$$

$$= -\sum_{x,y} p_x p_{y|x} (\log p_x + \log p_{y|x})$$

$$= -\sum_x p_x \log p_x - \sum_{x,y} p_x p_{y|x} \log p_{y|x}$$

$$= S(X) + S(Y|X)$$

(2.33)

This can be generalized to a chain rule for the joint entropy

$$S(X_1, X_2 \dots X_n) = S(X_1) + S(X_2 | X_1) + \dots + S(X_n | X_{n-1} \dots X_1) .$$

The proof of this claim comes down to the decomposition of the joint probability that can be done as a product of probabilities that become conditioned on all previous probabilities

$$p(x_1, x_2...x_n) = p_{x_1}(p_{x_2|x_1})(p_{x_3|x_2x_1})...(p_{x_n|x_{n-1}...x_1})$$

Finally, we state that the entropy is *subadditive*, meaning that

$$S(X_1, X_2 \dots X_n) \leq \sum_{i=1}^n S(X_i)$$
.

Perhaps surprisingly, subadditivity (or rather, its quantum counterpart) is one of the most important properties of entropy. This is because the subadditivity of entropy in a quantum theory on the boundary implies nontrivial restrictions on gravitational quantities in the bulk on the other side of an AdS/CFT duality [8, 15].

2.2.3 Mutual Information

The mutual information is a function of two random variables X, Y, and measures how much information they have in common. The mutual information is defined according to

$$I(X;Y) \equiv S(X) - S(X|Y) ,$$

which is readily interpreted as the information content of X minus the information content of Xgiven that you know Y. In this sense it is a measure of how much measuring one of the two variables reduces your uncertainty about the other. The mutual information is symmetric, meaning that

$$I(X;Y) = I(Y;X) = S(Y) - S(Y|X)$$
.

In terms of the joint and marginal probability density functions $p_x, p_y, p_{x,y}$ we can express the mutual information as

$$I(X;Y) = \sum_{x,y} p_{x,y} \log\left(\frac{p_{x,y}}{p_x p_y}\right) ,$$

which can be shown by using that $p_{x|y} = \frac{p_{x,y}}{p_y}$ and some algebra. In quantum information theory, the mutual information of factors in a pure state is a measure of entanglement. The Ryu-Takayanagi conjecture states that the entanglement measured by the quantum mutual information is related to a surface area in a dual gravitational theory [4], letting us relate the entanglement structure of a quantum many-body system to an emergent geometry [8].

Conditional Mutual Information

We can also define a conditional mutual information, according to

$$I(X;Y|Z) = S(Y|Z) - S(Y|X,Z) = S(X|Z) - S(X|Y,Y) = S(X|Z) + S(Y|Z) - S(X,Y|Z)$$
(2.34)

where both X and Y may be conditioned on Z. This definition is symmetric in X, Y, just like the mutual information. In the context of conditional mutual information we have another theorem whose quantum version is extremely important, namely strong subadditivity, which is stated as follows:

$$I(X;Y|Z) \ge 0. (2.35)$$

This is a direct result of the concavity of the entropy as well as the positivity of mutual information. In the quantum case we will find the proof significantly more involved.

2.2.4**Relative Entropy**

Relative entropy is a quantity that quantifies how distinguishable two probability distributions are. The relative entropy is defined as

$$D(p||q) = \sum_{x} p(x) \log\left(\frac{p(x)}{q(x)}\right)$$

where the sum only goes over the values of x for which both distributions are nonzero and p,qare probability distribution functions. The relative entropy is not strictly a distance measure, since it is not symmetric in its arguments. The relative entropy has the operational interpretation that if you use the wrong probability distribution q to encode an information source that actually emits information with distribution p, you will require on average S(X) + D(p||q) bits to encode each microstate. Further, relative entropy is non-negative and the fact that conditioning does not increase entropy, non-negativity of mutual information and strong subadditivity all follow from the non-negativity of relative entropy.

Proof of Non-negativity of Relative Entropy

Intuitively the positivity of relative entropy can be made quite clear, for values of x where q(x) > p(x), the logarithm is negative, and for p(x) > q(x) the logarithm is positive, but the latter case is weighted more favorably since they have larger p(x) than q(x). Mathematically this reasoning is not sufficient.

Let us consider $q(x) = p(x) + t\delta(x)$ where $\sum_x \delta(x) = 0, t \ge 0, -p(x) \le t\delta(x) \le 1 - p(x)$, where the limits are chosen such that q(x) is always between 0 and 1. Then,

$$D(p||q)\Big|_{t=0} = \sum_{x} p(x) \log(1) = 0 .$$
(2.36)

If the derivative with respect to t is positive for all t then non-negativity of relative entropy follows. We just carry out the derivative

$$\frac{\mathrm{d}}{\mathrm{d}t} \left[\sum_{x} p(x) \log\left(\frac{p(x)}{p(x) + t\delta(x)}\right) \right] = \sum_{x} p(x) \frac{\mathrm{d}}{\mathrm{d}t} \left[\log(p(x)) - \log(p(x) + t\delta(x)) \right]$$
$$= -\sum_{x} \frac{\delta(x)}{1 + t\delta(x)p(x)}$$
$$\geq -\sum_{x} \delta(x) = 0 , \qquad (2.37)$$

which is what we wanted to show. In the last step we used that $p(x) \ge 0$ and then set p(x) = 0. That the signs work out to a simple inequality in the last step is not particularly obvious by inspection, since $\delta(x)$ has indeterminate sign. When $\delta(x)$ is negative, the overall expression is positive, but setting p(x) to zero increases the size of the denominator. When $\delta(x)$ is positive, the overall sign is negative, but setting p(x) to zero decreases the size of the denominator. Thus, setting p(x) has the same effect in both cases.

2.3 Quantum Information Theory

Quantum information theory is the generalization of the classical information theory of probability distributions to density operators, their quantum counterpart. Initially it will almost seem like we are just repeating classical information theory, but we will find some very quantum behaviour, such as negative conditional entropies.

We will begin with a quick look at the relationship between classical information and quantum systems, and will then move on to the purely quantum theory. In this section we will assume that the Hilbert spaces involved are at most countably infinite dimensional, so that matrix (or matrix-like) representations of the density operator ρ make sense, and we assume that the Hilbert space is in general factorizable. In chapter 4, treating entropy in quantum field theory we will encounter a wider class of Hilbert spaces, and generalize the quantum relative entropy to be introduced in section 2.3.6. Since the other information theoretical quantities may be derived from the relative entropy, this amounts to a generalization of all quantities to the field theory case.

2.3.1 Classical information from quantum systems

We can always encode classical information in quantum systems. Imagine for example that we based on a classical probability distribution p(x) generate states with the density matrix ρ_x . From these quantum states we can obtain classical information by acting with some measurement operator that can be decomposed into the projective set of operators Λ_y . The conditional probability of outcome y given x is then

$$p_{y|x} = \operatorname{Tr}[\Lambda_y \rho_x]$$
.

If we try to define the entropy of some quantum state ρ in a naive manner, we need a projective measurement such as $\Lambda_x = |x\rangle\langle x|$ to map it to a classical probability distribution. It is then clear that

$$S(X) = \sum_{x} p_x \log p_x$$

=
$$\sum_{x} \operatorname{Tr}[\Lambda_x \rho] \log (\operatorname{Tr}(\Lambda_x \rho))$$
 (2.38)

It turns out that the minimum Shannon entropy with respect to all measurements of the density operator ρ is equal to something called the *Von-Neumann entropy* or *quantum entropy*, which will be the focus of the next section. In the quantum case, the unit of information for a *d*-dimensional quantum system is one *qudit*. A qudit is simply a *d*-dimensional state in Hilbert space. The *qubit* is a two-dimensional *qudit*, in analogy with the *bit* being an object that can be in the two states '0' and '1'. In the literature qubit is often sloppily used and may technically refer to a qudit. We will later show that one qubit may communicate more than one classical bit of information, due to the superposition property of quantum states that the classical bit does not possess.

2.3.2 Quantum Entropy

The quantum entropy, or *von Neumann entropy* as it is commonly called, actually predates classical entropy. One could imagine that it would have a more complex definition than in the classical case due to the uncertainty principle, but since the density operator already captures all properties of a quantum state we only need something that is a direct function of the density operator. This quantity is found to be

$$S(A)_{\rho} = -\mathrm{Tr}[\rho \log \rho] ,$$

where the logarithm acts on the eigenvalues of $\rho \in \mathcal{D}(\mathcal{H}_A)$. We denote the von Neumann entropy by $S(\ldots)_{\rho}$, but it is not unusual to drop the density operator subscript. In these cases it should be clear from context whether a classical or quantum entropy is being referenced. The advantage of not having " ρ_A " be the argument is that we might have a density operator over several subsystems ρ_{ABC} , and we would like the notation for joint, conditional and relative entropies to remain similar to that of classical information theory.

The interpretation of the von Neumann entropy is almost exactly the same as in the classical case. That is, a measurement of the state

$$\rho_A = \sum_x p_x \left| x \right\rangle\!\!\left\langle x \right|$$

has an expected information content of $S(A)_{\rho}$ qubits. However the quantum and classical quantities do not behave the same way in terms of the probability distribution p_x . Let us consider the ensemble $\{p_x, \rho_x\}$ with the following distribution:

$$\begin{aligned} p_1 &= \frac{1}{4} \quad \rho_1 = |0\rangle\!\langle 0| \ , \\ p_2 &= \frac{1}{4} \quad \rho_2 = |1\rangle\!\langle 1| \ , \\ p_3 &= \frac{1}{4} \quad \rho_3 = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)(\langle 0| + \langle 1|) \ , \\ p_4 &= \frac{1}{4} \quad \rho_4 = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)(\langle 0| - \langle 1|) \ . \end{aligned}$$

To communicate this ensemble classically we require on average $-\log \frac{1}{4} = 2$ bits of information, since we have four outcomes with a uniform distribution. The quantum entropy of $\rho = \sum_{x} p_x \rho_x$ is then given by

$$S(A)_{\rho} = -\operatorname{Tr}\left[\frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)\log\left(\frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)\right)\right]$$

= $\operatorname{Tr}\left[\frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)\right]$
= 1, (2.39)

so we see that the above ensemble has an expected information content of only one qubit. This is maybe not very surprising because we in some sense bloated the quantum ensemble with extra values ρ_3 , ρ_4 that do not modify the total density operator compared to an ensemble with probabilities $\frac{1}{2}$ and state ρ_1 , ρ_2 . This is closely related to the discussion about the non-uniqueness relation between ensembles and density operators around equation (2.10).

The von Neumann entropy is non-negative and takes its minimum value of 0 for a pure state. The fact that a pure state has zero entropy may seem contrary to our earlier claim that the von Neumann entropy takes into consideration the uncertainty principle, since entropy is supposed to be a measure of uncertainty. The resolution is that for every pure state $\rho = |\psi\rangle\langle\psi|$ a projective measurement that has $|\psi\rangle\langle\psi|$ as one of its projectors will always return the same result - i.e. if we ask the optimal question we get zero information because the optimal question (choice of measurement) is also the answer to the question we are trying to ask eg. "what state is this wavefunction in?". A state with quantum entropy is then a state which admits no "perfect question", and will have entropy regardless of what measurements you try to make.

The maximum von Neumann entropy is given by the maximally mixed state, and it is equal to $\log d$ where d is the dimension of the Hilbert space. The von Neumann entropy is additive in the sense that for a system with density operator $\rho = \rho_A \otimes \rho_B$ we have that

$$S(AB)_{\rho} = S(A)_{\rho_A} + S(B)_{\rho_B} . \tag{2.40}$$

This is a straightforward consequence of the fact that the logarithm of a product is the sum of the logarithms. The quantum entropy is concave, meaning that for two density operators $\rho_1, \rho_2 \in \mathcal{D}(\mathcal{H}_A)$ it satisfies

$$qS(A)_{\rho_1} + (1-q)S(A)_{\rho_2} \le S(A)_{q\rho_1 + (1-q)\rho_2} , \qquad (2.41)$$

where $q \in [0,1]$. This follows by carrying out the same exact analysis as in equation (2.30) and keeping in mind that the logarithms are functions of matrices, defined by their power series. The first derivative is carried out explicitly in equation (2.46). Just like in the case of classical entropies we have that the marginal entropies satisfy the subadditivity condition

$$S(AB)_{\rho} \le S(A)_{\rho} + S(B)_{\rho} . \tag{2.42}$$

Proving the subadditivity condition is made much simpler by the development of relative entropy, done in section 2.3.6. In the next section we will properly define both the joint $(S(AB)_{\rho})$ and marginal quantum entropies introduced in (2.42).

An interesting corollary

An interesting property is that given a density operator $\rho_A \in \mathcal{H}_A$, dropping the off-diagonal elements will always increase the entropy. This property will be used to show the subadditivity of the von-Neumann entropy in section 2.3.6. If ρ_D contains just the diagonal elements of ρ_A we have

$$S(A)_{\rho_D} \ge S(A)_{\rho_A} . \tag{2.43}$$

This follows from concavity by defining $\rho(q) = (1 - q)\rho_D + q\rho_A$. The first derivative of $S(A)_{\rho(q)}$ is zero

$$\frac{\mathrm{d}}{\mathrm{d}q}S(A)_{\rho} = -\mathrm{Tr}\left[\dot{\rho}\log\rho + \rho\frac{\dot{\rho}}{\rho}\right] = -\mathrm{Tr}\left[\dot{\rho}\log\rho\right] , \qquad (2.44)$$

where $\dot{\rho} \equiv \frac{d\rho}{dq}$, and the final term is zero due to $\text{Tr}[\rho] \equiv 1 \Rightarrow \text{Tr}\left[\frac{d\rho}{dq} = 0\right]$. The first term vanishes at q = 0 because $\dot{\rho}(0) = \rho_A - \rho_D$ has only zeroes on the diagonal, and $\log \rho(0)$ is completely diagonal. This means that $\dot{\rho}(0) \log \rho(0)$ only has zeroes on the diagonal, causing the trace to vanish. Since the function is concave and the first derivative is zero at q = 0 we have that

$$S(A)_{\rho(0)} \ge S(A)_{\rho(1)} \Rightarrow S(A)_{\rho_D} \ge S(A)_{\rho_A}$$
 (2.45)

The expression $\stackrel{"\dot{\rho}}{\rho}$ " in the above calculation is an abuse of notation since these are matrices and the definition of log ρ is actually a series expansion. Due to the cyclicity of the trace we can pull the $\dot{\rho}$ term out of the series expansion and furthermore treat it as commuting with the density. Let us explicitly show that we are not wrong by defining $\rho' + 1 = \rho$ and doing some Taylor expansions:

$$\begin{aligned} \operatorname{Tr}[\rho \frac{\mathrm{d}}{\mathrm{d}q} \log \rho] &= \operatorname{Tr}\left[(\rho'+1) \frac{\mathrm{d}}{\mathrm{d}q} \log(\rho'+1) \right] \\ &= \operatorname{Tr}\left[(\rho'+1) \frac{\mathrm{d}}{\mathrm{d}q} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \rho'^n \right] \\ &= \operatorname{Tr}\left[(\rho'+1) \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \left(\dot{\rho} \rho'^{n-1} + \rho \dot{\rho} \rho^{n-2} + \ldots \right) \right] \\ \langle \text{cyclicity of trace, } [\rho',1] &= [\rho',\rho'] = 0 \rangle = \operatorname{Tr}\left[\dot{\rho} (\rho'+1) \sum_{n=1}^{\infty} (-1)^{n-1} \rho'^{n-1} \right] \\ &= \operatorname{Tr}\left[\dot{\rho} (\rho'+1) \sum_{n=0}^{\infty} (-1)^n \rho'^n \right] \\ &= \operatorname{Tr}\left[\dot{\rho} (\rho'+1) \frac{1}{1+\rho'} \right] \\ &= \operatorname{Tr}[\dot{\rho}] = 0 \;. \end{aligned}$$

The Taylor expansions are well defined since we know that the eigenvalues of ρ are between 0 and 1.

2.3.3 Joint von Neumann Entropy

The joint von Neumann entropy has a straightforward definition, for some multipartite state $\rho \in \mathcal{D}(\mathcal{H}_{A_1} \otimes \mathcal{H}_{A_2} \dots \mathcal{H}_{A_N})$ it is simply

$$S(A_1 A_2 \dots A_N)_{\rho} = -\operatorname{Tr}_{A_1 A_2 \dots A_N} \left[\rho \log \rho \right] , \qquad (2.47)$$

where we have made explicitly clear that we trace over all subsystems. In general, if we write a trace without subscripts, we mean a trace over the entire Hilbert space.

The marginal entropies are obtained by tracing out the subsystems you do not want. Given a state $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ we define

$$\rho_{AB} = \text{Tr}_C[\rho]$$
.

and we find that the marginal entropy is

$$S(AB)_{\rho} = \operatorname{Tr}_{AB} \left[\operatorname{Tr}_{C}[\rho] \log \left(\operatorname{Tr}_{C}[\rho] \right) \right] = \operatorname{Tr}_{AB}[\rho_{AB} \log \rho_{AB}] .$$

$$(2.48)$$

It is clear that $S(A)_{\rho}$ means different things if the total density operator is defined on $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ or $\rho \in \mathcal{D}(\mathcal{H}_A)$, since in the first case we need to trace out a subsystem. This is an unavoidable consequence of notation, and good to keep in mind.

We should also know how the marginal entropies relate to the concept of purification. We know that the maximally mixed state with entropy $\log d$, where d is the dimension of the Hilbert space can be made into a pure state by entangling it with an environment as in section 2.1.3. A strange property of quantum marginal entropies is that they can be larger than the joint entropy, which is strictly forbidden in the classical case! Pure multipartite states enjoy a much more precise relation than this. Given a multipartite state ρ_{ABCDE} , defined on subsystems A, B, C, D and E, we have that for any bipartite cut such as AB|CDE

$$S(AB)_{\rho_{ABCDE}} = S(CDE)_{\rho_{ABCDE}} , \qquad (2.49)$$

where $\rho_{AB} = \text{Tr}_{CDE}[\rho_{ABCDE}]$. That is, the complementary marginal entropies for any bipartite pure state are equal, and the total entropy is always zero. Let us show this for general bipartite system, in which the wavefunction admits the Schmidt decomposition

$$|\Psi\rangle_{AB} = \sum_{i} \sqrt{\lambda_i} |i\rangle_A \otimes |i\rangle_B \quad , \tag{2.50}$$

where $|i\rangle_A$, $|i\rangle_B$ are sets of orthogonal vectors on A and B respectively, $i < \text{Dim}(\mathcal{H}_A) \cdot \text{Dim}(\mathcal{H}_B)$ and $\sum_i \lambda_i = 1$. The marginal densities are then

$$\rho_A = \sum_i \lambda_i |i\rangle_A \langle i|_A , \quad \rho_B = \sum_i \lambda_i |i\rangle_B \langle i|_B , \qquad (2.51)$$

which is straightforwardly obtained by using the orthonormality of states as in equation (2.16). Since the two marginal densities must by definition admit the same spectral decomposition, they also have the same entropy. Since nothing prevents us from renaming subsystem A to AB and subsystem B to CDE, the theorem applies to the marginal entropies of any subsystem and its complement for an unlimited number of parties.

It is common to define the entropy of either subsystem of a pure state as the *entropy of en*tanglement of that subsystem - to indicate that all of the entropy in the subsystem is truly due to entanglement with an external system. This result, that a pure state may have subsystems that do have entropy, is of *utmost importance* in quantum gravity and warrants a little digression. In general, the universe is taken to be a pure state. This means that any finite region of space that has entropy must be entangled with the universe in such a way that it is purified. In the Ryu-Takayanagi conjecture, this entanglement of a quantum subsystem with its environment is identified with the area of a boundary cutting off a finite region in space from its surroundings.

In quantum field theory, the entropy of black holes occurs because the spacetime outside the black hole is entangled with the interior, and since the black hole interior is behind an event horizon observers in the external spacetime do not have access to the "black hole interior" subsystem. The Unruh effect, in which an accelerating observer sees a thermal bath even in a perfect QFT vacuum, is due to the fact that a constantly accelerating observer is effectively cut off from a part of spacetime by an event horizon, the entropy is then due to the vacuum on the other side of the event horizon being entangled with the region accessible to the accelerating observer. These two examples are covered in detail in 4.2.

2.3.4 Conditional von Neumann Entropy

There is no direct analogue of conditional entropy in the quantum theory, although we could reasonably construct something along the lines of section 2.3.1 in terms of measurements. Such constructions unfortunately either depend on what set of projective measurements you select or require a computationally very difficult optimization over all possible measurements limiting their usefulness.

The most useful definition of conditional von Neumann entropy is as follows: let $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, the conditional entropy S(A|B) is then the difference of the joint quantum entropy S(AB) and the marginal entropy S(B) according to

$$S(A|B) = S(AB)_{\rho} - S(B)_{\rho} .$$
(2.52)

This definition is inspired by the relation between joint entropy and conditional entropy in equation (2.32), and due to this it enjoys several of the properties of classical conditional entropy, with the notable exception of positivity. This conditional entropy has the property that conditioning does not increase entropy, since

$$S(A) \ge S(A|B) , \qquad (2.53)$$

which follows from the fact that $S(AB)_{\rho} \leq S(A)_{\rho} + S(B)_{\rho}$, which can be inserted in equation (2.52).

The conditional entropy can be negative, as is made clear by considering the fact that a pure state has zero entropy (S_{AB}) while it may have nonzero marginal entropies $(S_A = S_B \neq 0)$. An interpretation of this is that for an entangled state one can be more certain about the state of the entire system than about its parts. We saw this exact property in the previous section, where we can be completely certain about the overall state of the system, but its parts have entropy.

2.3.5 Quantum Mutual Information

The quantum mutual information is defined analogously to the classical mutual information. Given a state $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ we have that the mutual information $I(A; B)_{\rho}$ is given by

$$I(A;B)_{\rho} \equiv S(A)_{\rho} + S(B)_{\rho} - S(AB)_{\rho}$$

$$(2.54)$$

The quantum mutual information can also be rewritten in analogy with the classical case in terms of conditional entropies

$$I(A; B)_{\rho} = S(A)_{\rho} - S(A|B)_{\rho}$$

= $S(B)_{\rho} - S(B|A)_{\rho}$. (2.55)

The quantum mutual information is non-negative:

$$I(A;B) \ge 0$$
 . (2.56)

The proof of this claim, which is equivalent with the subbadditivity condition in equation (2.42), is simplified by the introduction of the relative entropy.

2.3.6 Quantum Relative Entropy

The quantum relative entropy is defined as follows

$$D(\rho||\sigma) = \operatorname{Tr}[\rho(\log \rho - \log \sigma)], \qquad (2.57)$$

where ρ and σ are both in the same space $\mathcal{D}(\mathcal{H})$. The quantum relative entropy has the same operational interpretation as the classical relative entropy, namely that if you use the wrong density operator σ to encode an information source that actually emits information with density operator ρ , you will require on average $S(A)_{\rho} + D(\rho||\sigma)$ qubits to encode each microstate. This claim is known as the quantum Stein's lemma.

Next, let us show that the quantum relative entropy is positive. The positivity of the quantum relative entropy can be used to imply the subadditivity of quantum entropy and the positivity of the mutual information. These inequalities have been used in the context of AdS/CFT duality to derive positive energy conditions on the gravitonal dual to any boundary CFT [16].

Our assertion is that

$$D(\rho||\sigma) = \operatorname{Tr}[\rho(\log \rho - \log \sigma)] \ge 0 .$$
(2.58)

To show this, let us diagonalize σ . In general, this means that we will not have diagonalized ρ . Let us define $A\sigma_D A^{-1} = \sigma$ and $A\rho' A^{-1} = \rho$, where σ_D is completely diagonal. Using the cyclicity of the trace we then have

$$\operatorname{Tr}[\rho(\log \rho - \log \sigma)] = \operatorname{Tr}[\rho'(\log \rho' - \log \sigma_D)]$$
(2.59)

Next, we know that when we perform the trace, the completely diagonal matrix $\log \sigma$ will only interact with the diagonal elements of ρ , so we may replace it with its diagonal. Thus defining ρ'_D as only the diagonal elements of ρ' , we have

$$\operatorname{Tr}[\rho'(\log \rho' - \log \sigma_D)] = \operatorname{Tr}[\rho' \log \rho' - \rho'_D \log \sigma_D)].$$
(2.60)

Let us now add and subtract $\text{Tr}[\rho'_D \log \rho'_D]$. This gives us that

$$\operatorname{Tr}[\rho' \log \rho' - \rho'_D \log \rho'_D + \rho'_D \log \rho'_D - \rho'_D \log \sigma_D)] = D(\rho'_D ||\sigma_D) + S(A)_{\rho_D} - S(A)_{\rho} .$$
(2.61)

We showed in equation (2.43) that the sum of the last two terms is positive. The relative entropy between two completely diagonal matrices is equal to a sum over the diagonal according to

$$D(\rho'_D||\sigma_D) = \sum_i r_i \log \frac{r_i}{s_i} , \qquad (2.62)$$

where r_i, s_i are the eigenvalues of ρ_D, σ_D respectively. Since $\sum_i s_i = \sum_i r_i = 1$, this quantity is just the classical relative entropy which we already showed is positive definite in section 2.2.4. Since $D(\rho||\sigma)$ is equal to the sum of two positive terms, we have shown equation (2.58).
Now, let us show the positivity of mutual information. The trick here is to find ρ and σ such that the relative entropy becomes exactly the mutual information, then positivity of the relative entropy directly implies the positivity of mutual information. We begin by letting $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a general density operator for a joint system. We then define $\sigma_{AB} = \rho_A \otimes \rho_B$, where ρ_A, ρ_B are the local density operators obtained by tracing out subsystems in ρ_{AB} . Then, using that

$$\log \sigma_{AB} = \log(\rho_A \otimes \rho_B)$$

= $\log \left((\rho_A \otimes \mathbf{1}_B) (\mathbf{1}_A \otimes \rho_B) \right)$
= $\log(\rho_A \otimes \mathbf{1}_B) + \log(\mathbf{1}_A \otimes \rho_B)$
= $\log(\rho_A) \otimes \mathbf{1}_B + \mathbf{1}_A \otimes \log(\rho_B)$, (2.63)

the relative entropy becomes

$$D(\rho_{AB}||\sigma_{AB}) = \operatorname{Tr} \left[\rho_{AB}(\log \rho_{AB} - \log \sigma_{AB})\right]$$

= Tr $\left[\rho_{AB}(\log \rho_{AB} - \log(\rho_A) \otimes \mathbf{1}_B - \mathbf{1}_A \otimes \log(\rho_B))\right]$
= $S(A)_{\rho_{AB}} + S(B)_{\rho_{AB}} - S(AB)_{\rho_{AB}}$
= $I(A; B)_{\rho_{AB}}$, (2.64)

where the last step is just the definition of the quantum mutual information.

Finally, we will show the strong subadditivity property of the von Neumann entropy. To do this we need to quote the monotonicity of quantum relative entropy. Showing the monotonicity of quantum relative entropy is very involved, and is done in for example [17] and [18]. The monotonicity of relative entropy under any quantum channel is stated as follows: given $\rho_{AB}, \sigma_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ we have

$$D(\rho_{AB}||\sigma_{AB}) \ge D(\mathcal{N}(\rho_{AB})||\mathcal{N}(\sigma_{AB})) .$$
(2.65)

The trace-out operation is a quantum channel since the partial trace takes density operators to other density operators and is linear in its arguments. Thus, the relative entropy decreases monotonically under the trace-out of subsystems.

Now, let us consider a tripartite system with density operator ρ_{ABC} . Let us also define $\sigma_{ABC} = \rho_A \otimes \rho_{BC}$, where density operators missing subscripts indicate local operators where the rest of the system has been traced out. The operators obtained by tracing out subsystem C are ρ_{AB} and $\sigma_A B = \rho_A \otimes \rho_B$. The monotonicity of quantum relative entropy then tells us that

$$D(\rho_{ABC}||\sigma_{ABC}) \ge D(\rho_{AB}||\sigma_{AB})$$

$$D(\rho_{ABC}||\rho_A \otimes \rho_{BC}) \ge D(\rho_{AB}||\rho_A \otimes \rho_B)$$

$$-S(ABC)_{\rho} + S(A)_{\rho} + S(BC)_{\rho} \ge -S(AB)_{\rho} + S(A)_{\rho} + S(B)_{\rho}$$

$$S(BC)_{\rho} + S(AB)_{\rho} \ge S(ABC)_{\rho} + S(B)_{\rho} , \qquad (2.66)$$

where the last line is the strong subadditivity condition.

2.3.7 A Quantum Second Law of Thermodynamics

It is possible to use the monotonicity of relative entropy under any quantum channel in an intuitive way to obtain something that looks like a quantum version of the Second Law of thermodynamics. This example is really due to Witten [19], but it is incredibly interesting and relevant to the gauge/gravity duality so it is worth repeating. Let us suppose we have a thermal system of temperature $T = \frac{1}{\beta}$ described by a density operator σ according to

$$\sigma = \frac{1}{Z} e^{-\beta H} . \tag{2.67}$$

Here, the eigenstates of the associated Hilbert space are the eigenfunctions of the Hamiltonian H. The partition function Z is a normalization such that $\text{Tr}[\sigma] = 1$. The relative entropy between an arbitrary density operator ρ and σ is then given by

$$D(\rho||\sigma) = \operatorname{Tr}[\rho(\log \rho - \log \sigma)]$$

= $-S(A)_{\rho} + \operatorname{Tr}[\rho(\beta H + \log Z)]$
= $\beta (E(\rho) - TS(A)_{\rho}) + T\sigma \log Z$, (2.68)

where we have defined $E(\rho) = \langle H \rangle_{\rho} = \text{Tr}[H\rho]$. The function $E(\rho)$ is just the expectation value of the total energy (Hamiltonian) with respect to the state ρ . The quantity in the large brackets is the Helmholtz free energy, given by

$$F(\rho) = E(\rho) - TS(A)_{\rho}$$
 (2.69)

For σ , the free energy is given by

$$F(\sigma) = \operatorname{Tr}\left[\sigma H - T\sigma \log \sigma\right]$$

= $\operatorname{Tr}\left[\sigma H - T\beta \sigma H - \log Z\right].$ (2.70)

Substituting these expressions into equation (2.68) we obtain

$$D(\rho||\sigma) = \beta(F(\rho) - F(\sigma)) . \qquad (2.71)$$

Now, imagine acting with a quantum channel that preserves the thermal state σ on both arguments in the relative entropy. Thus, we have a quantum channel \mathcal{N} acting according to $\mathcal{N}(\rho) = \rho', \mathcal{N}(\sigma) = \sigma$. The monotonicity of quantum relative entropy tells us that

$$D(\rho||\sigma) \ge D(\rho'||\sigma) . \tag{2.72}$$

Since σ is unchanged, inserting into equation (2.71) this implies

$$F(\rho) \ge F(\rho') . \tag{2.73}$$

This states that under any evolution that preserves the thermal state, the free energy decreases.

For a system with a constant total energy (a closed system), where the time evolution operator that preserves the thermal state also preserves the total energy, this precisely implies the second law of thermodynamics. This result is a simple illustration of how the second law of thermodynamics could be seen as a consequence of the deterministic time-evolution of a quantum system.

Chapter 3

Elements of Quantum Field Theory

The goals of this chapter are several. First the path-integral formulation of quantum field theory (QFT) is introduced, since it is commonly not included in an introductory course in quantum field theory despite being an incredibly powerful tool in modern physics. Then, the path integral machinery is used to understand several important properties of quantum field theory, such as Ward identities, the renormalization group and quantum field thermodynamics.

Finally, the main ideas of conformal field theory (CFT) are developed for the purpose of preparing the reader for future chapters. CFT is one half of the AdS/CFT duality which is the main topic of this text. CFT is thought to describe the very high- and low energy limits of the standard model. It is also important in our understanding of string theory. The presentation of CFT in this chapter is somewhat incomplete, but future chapters will successively give a more complete treatement of how CFT can be applied.

3.1 Path Integral QFT

The path integral approach to quantum field theory is a powerful tool that maintains manifest Lorentz invariance upon quantization, due to not requiring a Hamiltonian reformulation (singling out a preferred time direction) to introduce the canonical commutation relations. The downside is that unitarity is not manifest as it is in the canonical approach. In some sense this makes the canonical and path integral formulations of quantum field theory complementary in that some things that obscure in one are obvious in the other. Additionally, in contrast to Hamiltonian formalism the path integral formulation does not require the existence of a global timelike Killing vector in its definition making it much better suited for QFT on curved spacetimes.

This section borrows heavily in structure from the parts I consider best about the presentation of Peskin & Schroeder [20] and Anthony Zee [21]. Some later parts on some formal developments regarding path integration regions are based on lecture notes by Hartman [22].

3.1.1 Path Integral Formulation of Nonrelativistic Quantum Mechanics

The most illuminating presentation of the path integral is best started with a thought experiment regarding a generalization of the double slit experiment. This will give us a physical intuition about a 'sum over paths' formulation of quantum mechanics. We can then confirm our intuition via formal manipulation of the usual finite-time transition amplitude.

We will consider the facts that



Figure 3.1: To the left, the propagation paths that go into the computation of the result of the two-slit experiment. To the right the set of paths that should be summed over to predict the result of a "two consecutive three-slit experiment" are indicated.

- The transition amplitude in the absence of obstacles¹ between an initial state *i* and a final state *f* in the time *t* is given by $\langle f | e^{-iHt} | i \rangle$, where *H* is the Hamiltonian
- The transition amplitude in the case of an impenetrable screen with two pointlike holes is given by the sum of the transition amplitudes taking the initial state through each of the holes to the final state

which will tell us that the object $\langle f | e^{-iHt} | i \rangle$ in the absence of obstacles *must* admit an interpretation as a sum over all possible paths the initial state could possibly take to the final state in the finite time t. We then show formally that this representation of the transition amplitude exists, this representation is the so-called *path integral representation* of quantum mechanics.

In the double slit experiment, a particle is emitted at time t = 0 from some source at a point S. The particle is then to pass through a screen with two holes, say at points A_1 and A_2 , to then hit a detector at time t = T located at a position O as in figure 3.1. The amplitude of detection is then given through the superposition principle of quantum mechanics as the sum of the amplitudes of the paths $S \to A_1 \to O$ and $S \to A_2 \to O$.

A more general experiment would then be an N-slit experiment. Denoting the amplitude for the path through slit i as $S \to A_i \to O$ as $\mathcal{A}(S \to A_i \to O)$, we can write down the amplitude \mathcal{A}_{tot} for detection at O

$$\mathcal{A}_{\text{tot}} = \sum_{i} \mathcal{A}(S \to A_i \to O) \ . \tag{3.1}$$

Now we ask what happens if we place a second screen between S and the first screen, with holes labelled by the index j. Once again we sum over all possible paths as in the right of figure 3.1, giving us a double sum over the holes of the two screens:

$$\mathcal{A}_{\text{tot}} = \sum_{j} \sum_{i} \mathcal{A}(S \to A_j \to B_i \to O) .$$
(3.2)

So far, nothing seems strange. We are just applying the superposition principle twice. Here, we may make interesting observation. Since we are free to make an arbitrary number of holes, we may choose to make a continuously infinite number of holes in the screen labelled by j. Then there would no longer be a screen present to block the path of our particle.

¹In this context, 'obstacles' are really some difficult-to-write-down infinite square wells that enter into the potential term of the Hamiltonian. The double slit is such an obstacle, but it is modelled by a superposition principle instead of adding it to the potential.



Figure 3.2: In a limit with an infinite number of slits in an infinite stack of screens, our discussion tells us that we should sum over a continuous set of paths. These paths are not necessarily smooth, but to take the thought experiment further it is best to continue in a more formal manner.

If there is supposed to be a smooth limit as we take the number of holes in screen j to infinity it appears that despite the fact that there is no screen, the particle should for consistency take every straight path it possibly can through the surface the screen used to occupy. This result comes off as unintuitive, equation (3.2) does not *seem* mathematically equivalent to equation (3.1) in the limit of $j \in \mathbb{R}^2$. If the superposition principle holds, then every imaginable surface in space must be a candidate for a screen in which we have drilled an infinite number of holes. Thus consistency demands that we sum the contributions of literally every esoteric path imaginable to compute the total amplitude of propagation from point S to point O, even through empty space. It is at this point less confusing to go to formalism than it is to take the thought experiment further, the important part is the conclusion that there should be some 'sum over paths' formulation of quantum mechanics.

Reinterpreting the Transition Amplitude

The takeaway from the previous thought experiment is simply that the usual transition amplitude in quantum mechanics must admit a representation as a sum over paths to be consistent with the superposition principle solution to computing the result of the double slit experiment. Let us show that this is indeed the case.

The transition amplitude $\mathcal{A}_{q_I \to q_F}$ from an initial position eigenstate $|q_I\rangle$ to a final position eigenstate $|q_F\rangle$ is given by

$$\mathcal{A}_{q_I \to q_F} = \langle q_F | e^{-iHt} | q_I \rangle \quad . \tag{3.3}$$

There is nothing preventing us from splitting the time evolution operator according to

$$e^{-iHt} = \left(e^{-iH\delta t}\right)^N, \quad \delta t = \frac{t}{N}.$$
 (3.4)

Next, we insert a completeness relation between each factor of $e^{-iH\delta t}$. Remember that the completeness relation for position eigenstates is given up to normalization by $\int dq |q\rangle \langle q| = 1$ (in this section, whenever integration limits are not given, they go from minus to plus infinity). We then

obtain the expression

$$\mathcal{A}_{q_{I} \to q_{F}} = \langle q_{F} | e^{-iH\delta t} \left(\int \mathrm{d}q_{N-1} |q_{N-1}\rangle \langle q_{N-1} | \right) e^{-iH\delta t} \left(\int \mathrm{d}q_{N-2} |q_{N-2}\rangle \langle q_{N-2} | \right) \\ \dots \left(\int \mathrm{d}q_{1} |q_{1}\rangle \langle q_{1} | \right) e^{-iH\delta t} |q_{I}\rangle \\ = \left(\prod_{j=1}^{N-1} \int \mathrm{d}q_{j} \right) \langle q_{F} | e^{-iH\delta t} |q_{N-1}\rangle \langle q_{N-1} | e^{-iH\delta t} |q_{N-2}\rangle \langle q_{N-2} | \dots |q_{1}\rangle \langle q_{1} | e^{-iH\delta t} |q_{I}\rangle ,$$

$$(3.5)$$

where the Π -term just reminds us to integrate over all of the q_j . Let us now pick out and look at an individual factor $\langle q_{j+1} | e^{-iH\delta t} | q \rangle$. To continue we let δt become small, such that that $e^{-iH\delta t} \approx 1 - iH\delta t$ is a good approximation. That is, we let

$$\langle q_{j+1} | e^{-iH\delta t} | q \rangle \rightarrow_{\delta t \to 0} \langle q_{j+1} | (1 - iH\delta t) | q \rangle$$
 (3.6)

The purpose of equation (3.6) is that we want to disentangle the terms in H that depend on \hat{p} and \hat{q} , since the two operators do not commute. If we are dealing with a full expansion, this is in general very difficult. Although the Hamiltonian as an operator is not bounded from above, equation (3.6) should be valid as long as we grow the limits of the generalized integral $\lim_{R\to\infty} \int_{-R}^{R} dq(\ldots)$ slower than we take the limit $\delta t \to 0$. Given that the expression converges at all (which it should nonperturbatively since H is positive definite and thus e^{-iHt} is bounded), it should not matter how we take the limit.

Since the $|q_i\rangle$ are eigenstates of \hat{q} we have

$$\langle q_{j+1} | \hat{q}^n | q \rangle = \left(\frac{q_{j+1} + q_j}{2}\right)^n \tag{3.7}$$

for even n, obtained by writing $\hat{q} = \left(\frac{\hat{q}+\hat{q}}{2}\right)$ and acting to the left and the right with half the qs each. For terms on the form $\langle q_{j+1} | f(p) | q \rangle$ we need to introduce a complete set of states in the momentum basis $(\int dp | p \rangle \langle p | \sim 1)$, we then get to replace the operator \hat{p} by the eigenvalue p, and we obtain

$$\langle q_{j+1} | f(p) | q \rangle = \int dp f(p) \langle q_{j+1} | p \rangle \langle p | q \rangle$$

$$= \int dp f(p) e^{ip(q_{j+1} - q_j)} ,$$

$$(3.8)$$

using that $\langle p|q\rangle = e^{ipx}$.

Whenever we have terms f(q,p) in the Hamiltonian that are proportional to products of p's and q's we take them to be *Weyl ordered*, fulfilling $\langle q_{j+1} | f(\hat{q},\hat{p}) | q \rangle = h(q_{j+1} + q_j) \langle q_{j+1} | g(\hat{p}) | q \rangle$. An example is the following:

$$\langle q_{j+1} | \left(\hat{q}^2 \hat{p}^2 + 2\hat{q}\hat{p}^2 \hat{q} + \hat{p}^2 \hat{q}^2 \right) | q \rangle = (q_{j+1} + q_j)^2 \langle q_{j+1} | \hat{p}^2 | q \rangle , \qquad (3.9)$$

obtained by letting the \hat{q} 's act in whatever direction does not have them acting past a \hat{p} . Any analytic function of \hat{q} and \hat{p} can be Weyl-ordered, though it will generally introduce extra normal ordering constants that should appear in the potential.

The essence of this discussion is that we consider only Hamiltonians where we are able to express $\langle q_{j+1} | (1 - iH\delta t) | q \rangle$ as functions of the relevant eigenvalues in a simple manner, without having to consider operator ordering in any real sense. As long as we are working at first order, this picture

is rigorously valid, since operator ordering operations will only appear as a constant in V(q). We now know that for a general Hamiltonian $H(\hat{q}, \hat{p})$ we can write

$$\langle q_{j+1} | (1 - i\delta t H(\hat{q}, \hat{p})) | q_j \rangle = \int dp \left[1 - iH \left(\frac{q_{j+1} + q_j}{2}, p \right) \delta t \right] e^{ip(q_{j+1} - q_j)}$$

$$= \int dp \, e^{-iH \left(\frac{q_{j+1} + q_j}{2}, p \right)} e^{ip(q_{j+1} - q_j)} ,$$
(3.10)

where we have reused that $e^{-iH\delta t} = 1 - iH\delta t$ is a good approximation. Inserting back into the full transition amplitude we have, naming $q_F \to q_N$ and $q_I \to q_0$

$$\begin{aligned} \mathcal{A}_{q_{0} \to q_{N}} &= \left(\prod_{i=1}^{N-1} \int \mathrm{d}q_{j}\right) \langle q_{N} | e^{-iH\delta t} | q_{N-1} \rangle \langle q_{N-1} | e^{-iH\delta t} | q_{N-2} \rangle \langle q_{N-2} | \dots | q_{1} \rangle \langle q_{1} | e^{-iH\delta t} | q_{0} \rangle \\ &= \left(\prod_{j=1}^{N-1} \int \mathrm{d}q_{j} \mathrm{d}p_{j}\right) e^{-iH\left(\frac{q_{N}+q_{N-1}}{2}, p_{j}\right)} e^{ip(q_{N}-q_{N-1})\delta t} \dots e^{-iH\left(\frac{q_{1}+q_{0}}{2}, p_{j}\right)\delta t} e^{ip(q_{1}-q_{0})} \\ &= \left(\prod_{j=1}^{N-1} \int \mathrm{d}q_{j} \mathrm{d}p_{j}\right) e^{\sum_{j} ip_{j}(q_{j+1}-q_{j}) - iH\left(\frac{q_{j+1}+q_{j}}{2}, p_{j}\right)\delta t} \\ &= \left(\prod_{j=1}^{N-1} \int \mathrm{d}q_{j} \mathrm{d}p_{j}\right) e^{i\sum_{j} \delta t \left[p_{j}\frac{(q_{j+1}-q_{j})}{\delta t} - H\left(\frac{q_{j+1}+q_{j}}{2}, p_{j}\right)\right]}. \end{aligned}$$
(3.11)

Taking the continuum limit, $q_{j+1} - q_j \to dq$, $q_{j+1} \to q_j$ and $\delta t \to dt$, so we see that

$$\lim_{\delta t \to 0} \mathcal{A}_{q_0 \to q_N} = \left(\int \mathcal{D}q \mathcal{D}p \right) e^{i \int_0^T \mathrm{d}t [p\dot{q} - H(q,p)]} .$$
(3.12)

where we have defined $\dot{q} = \frac{\mathrm{d}q}{\mathrm{d}t}$ and the integration measure $\mathcal{D}q\mathcal{D}p$ is defined by

$$\int \mathcal{D}q\mathcal{D}p = \left(\lim_{N \to \infty} \prod_{j=1}^{N-1} \int \mathrm{d}q_j \mathrm{d}p_j\right) .$$
(3.13)

In words, for every point in time between the initial time t = 0 and the final time t = T, the position q and the momentum p may take any value, and we sum all of the contributions by integrating over them. We can also understand this as the claim that we sum over all functions q(t), p(t) such that $q(0) = q_I, q(T) = q_F$. In this sense, the object in equation (3.13) is a functional integral over the space of functions q(t), p(t).

In general we may be interested in the transition amplitude between initial and final states that are not position eigenstates. Calling the initial and final state Ψ_i, Ψ_f respectively we write

$$\langle \Psi_f | e^{-iHT} | \Psi_i \rangle = \int dq_i \int dq_f \langle \Psi_f | | q_f \rangle \langle q_f | e^{-iHT} | q_i \rangle \langle q_i | | \Psi_i \rangle$$

$$= \int dq_i \int dq_f \Psi_f(q_f) \langle q_f | e^{-iHT} | q_i \rangle \Psi(q_i)$$

$$= \left(\int \mathcal{D}q \mathcal{D}p \right) \Psi_f^*(q_f) e^{i \int_0^T dt [p\dot{q} - H(q,p)]} \Psi(q_i) ,$$

$$(3.14)$$

this time with no boundary condition on the endpoints q(0) and q(T), meaning we integrate over them too. Equivalently, we may insert the final and initial functions as boundary conditions for ϕ , according to

$$\left\langle \Psi_{f}\right|e^{-iHT}\left|\Psi_{i}\right\rangle = \left(\int_{\Psi_{i}}^{\Psi_{f}}\mathcal{D}q\int\mathcal{D}p\right)\mathrm{d}e^{-i\int_{0}^{T}\left[p\dot{q}-H(q,p)\right]},\qquad(3.15)$$

where we have defined the boundary condition wavefunctions to contain the boundary position and field data.

In general path integrals are difficult to compute, since without some very neat mathematical tricks you will have to go to the definition of the measures $\mathcal{D}q$ and $\mathcal{D}p$ and compute the integrals explicitly, then take the limit as $N \to \infty$. Luckily, we will see that there are a number of slick mathematical tricks available in the following sections.

3.1.2 From Quantum Mechanics to QFT

In the previous section, we found the path integral representation of the propagator corresponding to one particle in quantum mechanics. We would now like to extend this to a system of multiple particles, and then take a continuum limit to obtain a field theory. Let us consider the object

$$Z \equiv \langle 0| e^{-iHt} |0\rangle = \left(\int \mathcal{D}q\mathcal{D}p\right) e^{i\int_0^T \mathrm{d}t[p\dot{q} - H(q,p)]}$$
(3.16)

where the initial and final vacuum states have been absorbed as boundary conditions on the path integral.

To generalize to multiple particles in a way that does not quickly become mathematically intractable, we imagine that our quantum system is a two-dimensional lattice of particles in the x-yplane. The coordinate of the *a*:th particle q^a represents a displacement in the *z*-direction, and the potential V contains only terms that mix adjacent points. Zooming out a little, this configuration looks like an infinite drumskin in the x - y plane, and displacements in the *z*-direction look like excitations of the flat ground state of the drum. The tension between atoms of the drum sits as nearest-neighbour interactions in the potential. Our groundstate-groundstate correlation function now looks something like

$$Z = \left(\int \mathcal{D}q^a \mathcal{D}p^a\right) e^{\sum_a i \int_0^T \mathrm{d}t \left[p^a \dot{q^a} - H(q^a, p^a)\right]} , \qquad (3.17)$$

where we are now integrating over all possible q^a and p^a . To understand why we get a sum in the exponential over the index a, and not outside of the entire thing, consider that a completeness relation for all of the Hilbert spaces of each point a must be the tensor product of the individual completeness relations, i.e.

$$1 = \bigotimes_{a} \int \mathrm{d}q^{a} \left| q^{a} \right\rangle \left\langle q^{a} \right| \,, \tag{3.18}$$

and similarly for the momenta. The product of the terms then enters into the exponential as a sum.

At this point, we are almost doing field theory. Imagine we space our lattice so densely it looks continuous. The index a can now instead be labelled by the continuous vector $\vec{x} = \begin{pmatrix} x \\ y \end{pmatrix}$. Then, the mass of the a:th particle becomes the mass area density: $m^a \to \sigma(\vec{x})$. The positions $q^a(t)$ become the scalar field $\phi(t,\vec{x})$, the momenta $p^a(t)$ become the momentum density $p(t,\vec{x})$, and the sum over a becomes the integral $\int d^2x$. In this continuum limit we then find that

$$Z = \left(\int \mathcal{D}\phi(t,\vec{x})\mathcal{D}p(t,\vec{x})\right) e^{i\int_0^T \mathrm{d}t \int \mathrm{d}^2x \left[p(t,\vec{x})\dot{\phi}(t,\vec{x}) - H(\phi(t,\vec{x}),p(t,\vec{x}))\right]}, \qquad (3.19)$$

where we have put the dependency explicitly in the functional integration measure just to indicate that something has truly changed. An important thing to note is that the path integral is no longer a functional integral over all paths through space (all possible choices of q(t)), but rather a functional integral over all possible configurations of the field $\phi(t, \vec{x}), t \in [0,T]$. When we call this a "path integral" in the future, we really mean an integral over *configuration space*, despite 'possible sequences of field configurations' not coinciding with our usual notion of a 'path'.

For an arbitrary field ϕ in an arbitrary spatial dimension D, we find that the general correlator is given by

$$\langle \Psi_f | e^{-iHT} | \Psi_i \rangle = \int_{\Psi_i}^{\Psi_f} \mathcal{D}\phi \mathcal{D}p e^{i \int_0^T \mathrm{d}t \int \mathrm{d}^D x \left[p \dot{\phi} - H(\phi, p) \right]} , \qquad (3.20)$$

where \vec{x} is a vector with D elements, and we have suppressed the (t, \vec{x}) dependence of the fields in the exponent to clean up the expression. For this functional integral, the field configuration is restricted to $\Psi_f^*(T, \vec{x})$ and $\Psi_i(0, \vec{x})$ at the end points.

The main takeaway here is that we have shown that the quantum correlation function is equivalent to a *classical* integral over all paths, and that the construction generalizes readily to fields. Now, we are free to specify the Hamiltonian to describe some Lorentz invariant dynamics as we usually do in QFT.

Arguing for a More Fundamental Path-Integral Formulation

We have in the previous sections argued that the quantum mechanical propagator $\langle f | e^{iHt} | i \rangle$ can be represented as a classical integral over all configurations that the quantum system may take in the time t between the initial and final states. We then argued that we could generalize this result from a theory of particles to a theory of fields. Let us now consider the classical Hamiltonian in 1+3 dimensions for a relativistic scalar field ϕ ,

$$H(\phi,\pi) = \int d^3x \left(\frac{1}{2}\pi^2 + \frac{1}{2}(\nabla\phi)^2 + V(\phi)\right), \qquad (3.21)$$

where π is the momentum density. Note that the momentum density π can only appear squared due to Lorentz invariance. Then the correlation function between final and initial states Ψ_f and Ψ_i is given according to equation (3.20) by

$$\langle \Psi_f | e^{-i\hat{H}T} | \Psi_i \rangle = \int \mathcal{D}\phi \mathcal{D}\pi \, e^{i\int_0^T \mathrm{d}t \int \mathrm{d}^3x \left[\pi \dot{\phi} - \frac{1}{2}\pi^2 - \frac{1}{2}(\nabla \phi)^2 - V(\phi)\right]} \,, \tag{3.22}$$

where we do not (and do not have to) determine the quantized Hamiltonian \hat{H} . The fact that we do not have to determine the quantum time evolution operator \hat{H} is the very beauty of the path integral representation. In the canonical approach, we have to introduce commutation relations and promote the fields and momenta in the Hamiltonian to quantum operators, potentially doing violence to Lorentz invariance and other classic symmetries. By expressing the propagator in terms of the classical path integral, we have circumvented this problem. That being said, quantum anomalies may instead arise in the integration measure $\mathcal{D}\phi$ if ϕ enjoys a nonabelian gauge symmetry.

For the case of the scalar field, we find that we can actually perform the momentum integrals over $\pi(t, \vec{x})$, because the object in the exponential is really "just" a Gaussian integral. Let us show this explicitly

$$\langle \Psi_{f} | e^{-i\hat{H}T} | \Psi_{i} \rangle = \int \mathcal{D}\phi \mathcal{D}\pi \, e^{i\int_{0}^{T} \mathrm{d}t \int \mathrm{d}^{3}x \left[\pi \dot{\phi} - \frac{1}{2}\pi^{2} - \frac{1}{2}(\nabla\phi)^{2} - V(\phi)\right] }$$

$$\langle \text{complete the square} \rangle = \int \mathcal{D}\phi \mathcal{D}\pi \, e^{i\int_{0}^{T} \mathrm{d}t \int \mathrm{d}^{3}x \left[-\frac{1}{2}(\pi + \dot{\phi})^{2} + \frac{1}{2}\dot{\phi}^{2} - \frac{1}{2}(\nabla\phi)^{2} - V(\phi)\right] }$$

$$\langle \text{definition of } \mathcal{D}\pi \rangle = \int \mathcal{D}\phi \left(\lim_{N \to \infty} \prod_{n} \int \mathrm{d}\pi_{n}\right) e^{i\int_{0}^{T} \mathrm{d}t \int \mathrm{d}^{3}x \left[-\frac{1}{2}(\pi_{n} + \dot{\phi})^{2} + \frac{1}{2}\dot{\phi}^{2} - \frac{1}{2}(\nabla\phi)^{2} - V(\phi)\right] }$$

$$\langle \text{Gaussian integral} \rangle = \int \mathcal{D}\phi \left(\lim_{N \to \infty} \sqrt{2\pi^{N}}\right) e^{i\int_{0}^{T} \mathrm{d}t \int \mathrm{d}^{3}x \left[\frac{1}{2}\dot{\phi}^{2} - \frac{1}{2}(\nabla\phi)^{2} - V(\phi)\right] } .$$

$$(3.23)$$

Rigorously the last step requires that we demand that our Lorentz-signature integral is an analytic continuation from a Euclidean integral. By Wick-rotating to imaginary time according to $t \rightarrow -i\tau$ we get a real exponent to ensure convergence. We then analytically continue the result of the Gaussian integration back to real time and take this to be the correct answer. This is nearly always implicitly assumed in QFT since the interaction picture requires the introduction of a complex time shift.

One might worry additionally that the obtained expression diverges, but really it is just an artifact of us not having introduced the appropriate factor of 2π into the $1^{"}="\int dp |p\rangle \langle p|$ completion relation, so we can just drop this term. The reason for omitting this is that an overall factor is physically uninteresting, so we can keep things simple and just drop them. To do perturbation theory we will have to divide by a factor that contains the same set of normalization errors anyways, so this will not be an issue. With this, we have that the correlator is

$$\langle \Psi_f | e^{-i\hat{H}T} | \Psi_i \rangle = \int \mathcal{D}\phi e^{i\int_0^T dt \int d^3x \left[\frac{1}{2}\dot{\phi}^2 - \frac{1}{2}(\nabla\phi)^2 - V(\phi)\right]}$$

$$= \int \mathcal{D}\phi e^{i\int_0^T dt \int d^3x \left[-\frac{1}{2}(\partial_\mu\phi)^2 - V(\phi)\right]}$$

$$= \int \mathcal{D}\phi e^{i\int_0^T dt \int d^3x \mathcal{L}} ,$$

$$(3.24)$$

where we have recognized that we are left with an expression in terms of the Lagrangian density for the relativistic scalar field $\mathcal{L} = -\frac{1}{2}(\partial_{\mu}\phi)^2 - V(\phi)$. A nice thing about this expression is that the *quantum* Hamiltonian mechanics of our system is described by an object that depends solely on the *classical* action $S[\phi] = \int_0^T d^4 x \mathcal{L}$. This formulation of the quantum mechanical propagator therefore manifestly enjoys all of the symmetry properties of the Lagrangian, from Lorentz invariance to any eventual gauge symmetries. This is contrary to the case of canonical quantization, where we have to perform gauge fixing and introduce Ward identities to implement Lorentz invariance. While the Ward identities are not ad-hoc in Hamiltonian formalism, they arise less elegantly than from the path integral formalism.

The Hamiltonian dynamics that we use to describe canonical quantum theory are arbitrary; in a relativistic setting the Hamiltonian dynamics depend on the arbitrary choice of a timelike timeevolution vector, promoting time to something completely separate from the space coordinates. Because of this, a more fundamental way of looking at the Hamiltonian dynamics of a quantum theory might be to claim that the Hamiltonian dynamics are in fact *defined* by the right-hand side of equation (3.24). For any Lagrangian, the path integral will correspond to some operator \hat{H} , and we can differentiate with respect to T to find the corresponding Schrödinger equation. In this paradigm, the classical Lagrangian \mathcal{L} is the most fundamental specification of the quantum theory. An important property of this is that Hamiltonian dynamics can now be defined on spacetimes that do not admit a global timelike Killing vector. In general we expect the object

$$\langle \Psi_f | e^{-i\hat{H}T} | \Psi_i \rangle = \int \mathcal{D}\phi \, e^{i\int_0^T \mathrm{d}^4 x \mathcal{L}} = \int \mathcal{D}\phi e^{-iS[\phi]} \tag{3.25}$$

to converge because really extreme field configurations tend to interfere destructively while the configurations for which $S[\phi]$ varies slowly in ϕ will interfere constructively. Rigorously, to ensure the convergence of the path integral one wants perform a wick rotation to Euclidean time, substituting $t \to i\tau$. One performs the functional integral in Euclidean time, sees that it converges and analytically continues the answer in terms of real τ back to Lorentzian time, setting $\tau = -it$. Analyticity is generally assumed unless there is a very good reason not to, so we will be substituting back and forth without regard.

Classical Limit

It would be strange if in the classical limit $\hbar \to 0$ we did not recover classical mechanics. In the path integral formalism this turns out to happen in a remarkably simple way. Reinstating \hbar in the correlator we have

$$\langle \Psi_f | e^{-\frac{i}{\hbar}\hat{H}T} | \Psi_i \rangle = \int \mathcal{D}\phi \, e^{\frac{i}{\hbar} \int_0^T \mathrm{d}^4 x S[\phi]} \,. \tag{3.26}$$

In the limit $\hbar \to 0$, the integral is given by the stationary phase approximation, in terms of the stationary points of $S[\phi]$. Since $S[\phi]$ is a functional, the stationary points are given by the variation of the action $\delta_{\phi}S[\phi] = 0$, which is precisely the definition of the classical equations of motion. In this sense, the action principle of Hamilton follows from the path integral formulation being fundamental plus the assertion that the classical limit is given by $\hbar \to 0$.

3.1.3 Applying the Path Integral Formalism

In this section the path integral machinery is applied to write down general two-point correlation functions. Then higher order correlators are briefly discussed before introducing the generating functional method to compute example correlation functions. After this the perturbation expansion that describes interacting theories is described and Noether's theorem is generalized to the Ward identities. Finally some formal ways to write down vacuum states and amplitudes using Wickrotated path integrals are displayed.

Correlation Functions

The goal of this section is to recover a path integral expression for the Freeman-Dyson correlation functions. Let us for simplicity work with a Hermitian scalar field $\phi(x)$. We then wish to consider the object

$$\left\langle \phi_f \right| T[\phi(x_1)\phi(x_2)e^{-i\hat{H}2T}] \left| \phi_i \right\rangle = \int \mathcal{D}\phi \,\phi(x_1)\phi(x_2)e^{i\int_{-T}^{T} \mathrm{d}^4x\mathcal{L}} , \qquad (3.27)$$

where the boundary conditions on the path integral are $\phi(-T,\vec{x}) = \phi_i$, $\phi(T,\vec{x}) = \phi_f$ for some initial and final states ϕ_i, ϕ_f , and $\phi_S(x_1), \phi_S(x_2)$ are two time-independent field fields defined at times x_1^0, x_2^0 respectively. $T[\ldots]$ denotes time ordering on the operator side. The goal is to relate this object to the Freeman-Dyson two-point correlator $\langle \Omega | T(\hat{\phi}(x_1)\hat{\phi}(x_2)) | \Omega \rangle^2$ (where T is the time ordering operator and $\hat{\phi}$ is the Heisenberg operator that creates a quantum of the scalar field).

 $^{^{2}}$ If this notation is unfamiliar, chapter 4 in Peskin & Schroeder is recommended [20].

We can then decompose the path integral in the following manner.

$$\int \mathcal{D}\phi = \int \mathcal{D}\phi_1(\vec{x}) \int \mathcal{D}\phi_2(\vec{x}) \int_{\substack{\phi(x_1^0, \vec{x}) = \phi_1(\vec{x}) \\ \phi(x_2^0, \vec{x}) = \phi_2(\vec{x})}} \mathcal{D}\phi(x)$$
(3.28)

Here, we have imposed an additional constraint on the main path integral, so that in addition to the constraints at $\pm T$, we have a restriction at $\phi(x_1^0)$ and $\phi(x_2^0)$. The two path integrals over $\phi_1(\vec{x})$ and $\phi_2(\vec{x})$ are path integrals only over spatial configurations, as indicated by the three-vector argument. Since we integrate over all possible field configurations of ϕ_1 and ϕ_2 we truly have not changed the path integral at all.

Because of the imposition of extra boundary conditions, the correlator with the new integration measure will automatically be time-ordered. To see this consider the example when $x_1^0 < x_2^0$, we have that

$$\langle \phi_f | \phi(x_1)\phi(x_2)e^{-i\hat{H}2T} | \phi_i \rangle = \int \mathcal{D}\phi_1(\vec{x}) \int \mathcal{D}\phi_2(\vec{x}) \phi_1(\vec{x})\phi_2(\vec{x}) \langle \phi_f | e^{-iH(T-x_2^0)} | \phi_2 \rangle$$

$$\times \langle \phi_2 | e^{-iH(x_2^0-x_1^0)} | \phi_1 \rangle \langle \phi_1 | e^{-iH(x_1^0+T)} | \phi_i \rangle .$$
(3.29)

We can now turn the field $\phi_1(\vec{x})$ into an operator by observing that $\phi_1 |\phi_1\rangle = \hat{\phi}_1 |\phi_1\rangle$ ("inverting" the eigenvalue relation). The completion relation $\int \mathcal{D}\phi_1 |\phi_1\rangle\langle\phi_1|$ that we used to decompose our path integral can then be evaluated to eliminate the intermediate states and equation (3.29) turns into

$$\langle \phi_f | \phi(x_1)\phi(x_2)e^{-i\hat{H}2T} | \phi_i \rangle = \langle \phi_f | e^{-iH(T-x_2^0)}\hat{\phi}_2 e^{-iH(x_2^0-x_1^0)}\hat{\phi}_1 e^{-iH(x_1^0+T)} | \phi_i \rangle$$
(3.30)

The exponential containing x^0 's then combines with the Schrödinger operators to make Heisenberg field operators, since $[H,H] = 0 e^{H(A+B)} = e^{HA}e^{HB}$, and we are left with

$$\langle \phi_f | \phi(x_1)\phi(x_2)e^{-i\hat{H}2T} | \phi_i \rangle = \langle \phi_f | e^{-iHT}T \left[\phi_H(x_2)\phi_H(x_2) \right] e^{-iHT} | \phi_i \rangle \quad (3.31)$$

where $T[\ldots]$ is the time ordering operator. This is similar to the two-point correlator in the Freeman-Dyson approach, and really all that is missing is performing the same steps as they did. We project out the vacuum from the initial and final states by taking $T \to \infty(1 - i\epsilon)$, assuming that the initial and final states have any overlap with the vacuum. Let us briefly show how this comes about; we write ϕ_i in terms of the energy eigenstates $|n\rangle$ of the Hamiltonian as

$$|\phi\rangle = \sum_{n} |n\rangle \langle n|\phi_i\rangle \quad . \tag{3.32}$$

Then we have, keeping only the leading term

$$e^{-iHT} |\phi_i\rangle = \sum_n e^{-iE_nT} |n\rangle \langle n|\phi_i\rangle \xrightarrow{]} T \to \infty(1-i\epsilon)e^{-\infty(i+\epsilon)E_\Omega} |\Omega\rangle \langle \Omega|\phi_i\rangle \quad . \tag{3.33}$$

This term contains some difficult to compute factors that depend on the details of the overlap between the initial state and the vacuum. We deal with this by dividing off equation (3.31) without the extra Heisenberg operators. Then we find that in terms of path integrals we may express the two-point correlator in the Freeman-Dyson approach as

$$\frac{\langle \Omega | T[\phi(x_1)\phi(x_2)] | \Omega \rangle}{\langle \Omega | \Omega \rangle} = \lim_{T \to (1-i\epsilon)\infty} \frac{\int \mathcal{D}\phi \,\phi(x_1)\phi(x_2) e^{-\int_{-T}^{T} \mathrm{d}t \int \mathrm{d}x\mathcal{L}}}{\int \mathcal{D}\phi \, e^{-\int_{-T}^{T} \mathrm{d}t \int \mathrm{d}x\mathcal{L}}} \,. \tag{3.34}$$

For the higher order correlators we just insert more fields $\phi(x_i)$.

At this point it is possible to go on and derive the Feynman rules in without introducing any new concepts by going to the definition of the integration measures and taking limits. We will not do this, and will instead the introduce the generating functional method.

Generating Functional Method for Correlation Functions

The generating functional method of computing correlation functions on the form of equation (3.34) is quite formal, but a lot less work than going to definitions, explicitly evaluating integrals and taking limits. To introduce the generating functional method we need a little bit of formalism. The functional derivative δ is defined by the basis arises.

The functional derivative $\frac{\delta}{\delta J(x)}$ is defined by the basic axioms

$$\frac{\delta}{\delta J(x)}J(y) = \delta(x-y) , \quad \frac{\delta}{\delta J(x)} \int d^4 y J(y)\phi(y) = \phi(x) .$$
(3.35)

This definition can, viewing a function as a continuously infinite dimensional vector, be seen as the obvious generalization of the rules for discrete vectors

$$\frac{\partial}{\partial x_i} x_j = \delta_{ij} , \quad \frac{\partial}{\partial_i} \sum_j x_j k_j = k_i .$$
(3.36)

The functional derivative of more complicated functions is just performed using the usual productand chain rules of ordinary differentiation. We expect this to function in much the same way as ordinary differentiation for the simple reason that we see it as a limit $n \to \infty$ of ordinary derivatives on a vector of length n. We do have one additional trick, when the function in question depends on a derivative of J we may use partial integration. Two examples are

chain rule

partial integration

$$\frac{\delta}{\delta J(x)} e^{\int d^4 y J(y)\phi(y)} = \phi(x) e^{\int d^4 y J(y)\phi(y)} , \quad \frac{\delta}{\delta J(x)} \int d^4 y \,\partial_\mu J(y)\phi(y) = -\partial_\mu \phi(x) . \tag{3.37}$$

That is really all we need to start developing the generating functional formalism. In a scalar field theory, say with a free Klein-Gordon Lagrangian, we define the generating functional Z[J] in the following way

$$Z[J] = \int \mathcal{D}\phi e^{-i\int d^4x \left[\mathcal{L} + J(x)\phi(x)\right]} .$$
(3.38)

Then, the correlation functions can be computed by taking derivatives of the generating functional. Let us consider the two-point correlator for the free Klein-Gordon field:

$$\frac{\langle \Omega | T[\phi(x_1)\phi(x_2)] | \Omega \rangle}{\langle \Omega | \Omega \rangle} = \frac{1}{Z_0} \left(\frac{-i\delta}{\delta J(x_1)} \right) \left(\frac{-i\delta}{\delta J(x_2)} \right) Z[J] \Big|_{J=0} , \qquad (3.39)$$

where we have defined $Z_0 = Z[0]$. Each derivative with respect to J just brings out a factor ϕ in front of Z[J], so we have just recovered equation (3.34). This does not give us anything new, it just shows that we have not defined something fundamentally different from before. The next realization to make is that Z[J] can in fact be rewritten in a sneaky way. Let us consider specifically the case of the massive Klein-Gordon Lagrangian. This method works for any theory whose Green's function we can compute without introducing new formalism. We manipulate the exponent of Z[J]

$$\int d^4x \left[\mathcal{L} + J(x)\phi(x) \right] = \int d^4x \left[-\frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{1}{2} (m^2 - i\epsilon) \phi^2 + J(x)\phi(x) \right]$$

$$= \int d^4x \left[-\frac{1}{2} \phi (\partial_\mu \partial^\mu + m^2 - i\epsilon) \phi + J(x)\phi(x) \right],$$
(3.40)

where we have just performed a partial integration and dropped the boundary terms. The $i\epsilon$ term is a convergence factor, and comes from the fact that our integral is integrated in the $T \rightarrow \infty(1-i\epsilon)$ direction, which implicitly shifts the 0-components of the momenta in the Fourier space representation of $(\partial^2 - m^2)$, effectively appearing as a small shift of the mass-squared. This should be familiar as a necessary part of the derivation of the free Klein Gordon propagator.

We are now ready for the really sneaky step; we perform a shift of the field ϕ by adding another term

$$\phi' = \phi - (\partial_{\mu}\partial^{\mu} + m^2 - i\epsilon)^{-1}J(x) . \qquad (3.41)$$

Then we obtain (suppressing the x-dependence of J)

$$\int d^{4}x \left[\mathcal{L} + J\phi\right] = \int d^{4}x \left[-\frac{1}{2}(\phi' + (\partial_{\mu}\partial^{\mu} + m^{2} - i\epsilon)^{-1}J)(\partial_{\mu}\partial^{\mu} + m^{2} - i\epsilon)(\phi' + (\partial_{\mu}\partial^{\mu} + m^{2} - i\epsilon)^{-1}J) + J\phi\right]$$

$$= \int d^{4}x \left[-\frac{1}{2}\phi'(\partial_{\mu}\partial^{\mu} + m^{2} - i\epsilon)\phi' - \frac{1}{2}(\phi J + J\phi) - \frac{1}{2}J(\partial_{\mu}\partial^{\mu} + m^{2} - i\epsilon)^{-1}J + \mathcal{H}\phi\right]$$

$$= \int d^{4}x \left[-\frac{1}{2}\phi'(\partial_{\mu}\partial^{\mu} + m^{2} - i\epsilon)\phi' - \frac{1}{2}J(\partial_{\mu}\partial^{\mu} + m^{2} - i\epsilon)^{-1}J\right].$$
(3.42)

Since our change of functions was only a shift by a constant function, the Jacobian of the change of variables is unity, and we can just exchange $\mathcal{D}\phi \to \mathcal{D}\phi'$ in the path integral. We find the expression

$$Z[J] = \int \mathcal{D}\phi' e^{\int \mathrm{d}^4x \left[-\frac{1}{2}\phi'(\partial_\mu\partial^\mu + m^2 - i\epsilon)\phi' - \frac{1}{2}J(\partial_\mu\partial^\mu + m^2 - i\epsilon)^{-1}J\right]} \,. \tag{3.43}$$

But the first term in the exponent is just Z_0 . Even better, the second term does not depend on ϕ' Thus we can decompose according to

$$Z[J] = Z_0 e^{-\frac{i}{2} \int d^4 x J (\partial_\mu \partial^\mu + m^2 - i\epsilon)^{-1} J} .$$
(3.44)

Observing that the differential operator $(-\partial_{\mu}\partial^{\mu} - m^2 + i\epsilon)$ has a Green's function representation we may write

$$Z[J] = Z_0 \exp\left[-\frac{1}{2} \int d^4 x \, d^4 y J(x) (\partial_\mu \partial^\mu + m^2 - i\epsilon)^{-1} \delta(x - y) J(y)\right]$$

= $Z_0 \exp\left[-\frac{1}{2} \int d^4 x \, d^4 y \frac{d^4 p}{(2\pi)^4} J(x) \frac{ie^{-ip \cdot (x - y)}}{p^2 + m^2 - i\epsilon} J(y)\right]$
= $Z_0 \exp\left[-\frac{1}{2} \int d^4 x \, d^4 y J(x) D_F(x - y) J(y)\right],$ (3.45)

where we have defined the Feynman propagator D_F . This last form of Z[J] is the one we wish to use for practical reasons, and this calculation is intended as a crash reminder of how to get from differential operator to Feynman propagator. The first step is just using $J(x) = \int d^4y \delta(x-y)J(y)$, then, going to Fourier space $\partial \to p$ and $\delta \to 1$. Performing the *p*-integral then gives us the propagator. If this is all wholly unfamiliar, chapter two of [20] is recommended.

We are now ready to find some correlation functions. The two-point function is

$$\langle \Omega | T[\phi(x_1)\phi(x_2)] | \Omega \rangle = -\frac{\delta}{\delta J(x_1)} \frac{\delta}{\delta J(x_2)} \exp\left[-\frac{1}{2} \int d^4 x \, d^4 y J(x) D_F(x-y) J(y)\right]_{J=0}$$

$$= \frac{\delta}{\delta J(x_1)} \left(\frac{1}{2} \int d^4 x J(x) D_F(x-x_2) J(x_2) + \frac{1}{2} \int d^4 y J(x_2) D_F(x_2-y)\right) \frac{Z[J]}{Z[0]} \Big|_{J=0}$$

$$= \frac{1}{2} D_F(x_1-x_2) + \frac{1}{2} D_F(x_2-x_1)$$

$$= D_F(x_1-x_2) .$$
(3.46)

The first derivative acting on the exponential brings down two identical terms (since $D_F(x-y) = D_F(y-x)$), and the only terms that survive setting J = 0 are the ones that get hit by a second derivative with respect to J. This leads to the survival of only the terms outside of the exponential in the third step.

Let us work out the four-point correlator in a similarly explicit way, introducing only some cleaner notation. Let us set $\phi_1 \equiv \phi(x_1)$, $J_x \equiv J(x)$, $D_{x1} \equiv D(x - x_1)$ and so on. We then take the integration over repeated subscripts to be implied (an "Einstein integration convention" if you will). We can then start doing algebra

$$\langle \Omega | T[\phi_1 \phi_2 \phi_3 \phi_4] | \Omega \rangle = \frac{\delta}{\delta J_1} \frac{\delta}{\delta J_2} \frac{\delta}{\delta J_3} [\underbrace{-J_x D_{x4}}_{=-\frac{1}{2}(J_x D_{x4} + J_y D_{4y})}] e^{-\frac{1}{2}J_x D_{xy} J_y} \Big|_{J=0}$$

$$= \frac{\delta}{\delta J_1} \frac{\delta}{\delta J_2} [-D_{34} + J_x D_{x3} D_{y4}] e^{-\frac{1}{2}J_x D_{xy} J_y} \Big|_{J=0}$$

$$= \frac{\delta}{\delta J_1} [D_{23} J_y D_{y4} + J_x D_{x3} D_{24} + D_{34} J_x D_{x2} - J_y D_{y3} J_z D_{z2} J_x D_{x4}] e^{-\frac{1}{2}J_x D_{xy} J_y} \Big|_{J=0}$$

$$= D_{23} D_{14} + D_{13} D_{24} + D_{34} D_{12} .$$

$$(3.47)$$

The final step follows by the same reasoning as last time - upon setting J = 0, only terms that were linear in J outside the exponential survive.

We have discovered the beginnings of the general pattern in the canonical approach - the correlator is given by the sum of all possible contractions (the Wick contraction of two fields equals the propagator) between the involved fields. The general construction above can be used in the case of interacting theories as well; the general definition of the correlator in terms of path integrals is unchanged. In the case of an interacting theory the factor Z_0 is nontrivial, corresponding to the sum of all vacuum bubbles in the Freeman-Dyson approach.

Interacting theory

While we see that we can get a nice, tractable expression for the correlators of a free theory, it is unclear if we are as lucky when it comes to arbitrary interaction terms in the Lagrangian. Let us add a general interaction term \mathcal{L}_{int} to the Lagrangian and consider the generating functional:

$$Z[J] = \int \mathcal{D}\phi e^{-i\int \mathrm{d}^d x (\mathcal{L}_{free} + \mathcal{L}_{int}) + J\phi} . \qquad (3.48)$$

The integral is no longer Gaussian and we can no longer shift the exponent of equation (3.48) to find the exact correlation functions. Instead we need to use *perturbation theory*, we Taylor expand \mathcal{L}_{int} around the point where all coupling constants are zero. That is, we write

$$Z[J] = \left[e^{\int d^d x \mathcal{L}_{int}\left(\frac{-i\delta}{\delta(J(x))}\right)} \right] \mathcal{D}\phi e^{-i\int d^d x (\mathcal{L}_{free} + J\phi)}$$

$$= \left[e^{\int d^d x \mathcal{L}_{int}\left(\frac{-i\delta}{\delta(J(x))}\right)} \right] Z_0[J] .$$
(3.49)

Given that we already know $Z_0[J]$ from the previous section (where we can do the shift again since its a free theory) and we may take \mathcal{L}_{int} as given, there is nothing preventing us from evaluating the expansion.

An interesting note is that the perturbative expansion in equation (3.49) generally has zero radius of convergence. An expansion is said to have a radius of convergence R if

$$\sum_{k}^{\infty} \lambda^{k}(...) = \begin{cases} \text{finite if } |\lambda| < R\\ \text{divergent if } |\lambda| > R \end{cases}$$
(3.50)

Consider for example ϕ^4 -theory, with an interaction potential $V = \lambda \phi^4$. If λ is positive, the potential is bounded from below and small values of ϕ are preferred. If λ is slightly negative the potential is unbounded from below and ϕ will want to shoot off to infinity. Such a potential is clearly unphysical since it would be possible to extract an infinite amount of energy from the ever growing fields. Therefore the expansion diverges for arbitrarily small negative λ , and it must have a vanishing radius of convergence. Luckily for the many people trying to describe reality using perturbation theory, the expansion in equation (3.49) is still numerically accurate to order $\frac{1}{g}$, where g is the dimensionless coupling constant, as is eloquently presented in [23].

Symmetries and Ward Identities

Field transformations $\phi \to \phi + \delta \phi$ that leave the Lagrangian \mathcal{L} invariant are called symmetries. Noether's theorem tells us that for each infinitesimal symmetry of the Lagrangian there is an associated conserved current. In the path integral formulation of QFT we see that the symmetries of the Lagrangian also imply *Ward identities*, identities that the quantum operators of a theory must satisfy to preserve the symmetry. In the following we assume that symmetry transformations leave the integration measure $\mathcal{D}\phi$ invariant. This is not true in general, and leads to the introduction of anomalies in non-abelian gauge theories such as quantum chromodynamics and CFT.

Let us start from the generating functional Z[J] and perform the shift $\phi \to \phi' = \phi + \delta \phi$, letting $\delta \phi$ be arbitrary. The generating functional is assumed to be invariant under an infinitesimal shift, and we find by applying the functional derivative with respect to ϕ that

$$\delta_{\phi} Z[J] = i \int \mathcal{D}\phi e^{-i(S[\phi] + \int d^d x J\phi)} \int d^d x \left(\frac{\delta S}{\delta \phi} + J\right) \delta \phi = 0 , \qquad (3.51)$$

where we have assumed that the integration measure $\mathcal{D}\phi = \mathcal{D}\phi'$ is invariant.

This equation is not very useful by itself, since all physical quantities are obtained by applying functional derivatives with respect to J and then taking $J \rightarrow 0$. For a general correlator, after applying the functional derivatives and setting J = 0 equation (3.51) implies the Schwinger-Dyson equations

$$0 = \left\langle \frac{\delta S}{\delta \phi(x)} \phi(x_1) \phi(x_2) \dots \phi(x_n) \right\rangle - i \sum_{j=1}^n \langle \phi(x_1) \dots \phi(x_{j-1}) \delta(x - x_j) \phi(x_{j+1}) \dots \phi(x_n) \rangle .$$
(3.52)

In particular, we see that the Schwinger-Dyson equations imply the equations of motion $(\delta_{\phi}S[\phi] = 0)$ up to contact terms $(\delta(x - x_j))$, arising from when the integration variable field $\phi(x)$ overlaps with the field operators $\phi(x_j)$ in the correlator. In Euclidean signature the factor of *i* in front of the sum disappears.

Let us now assume that we perform a symmetry transformation. Then the variation of the Lagrangian reads

$$\delta \mathcal{L} = \frac{\partial \mathcal{L}}{\partial \phi} \delta \phi + \frac{\partial \mathcal{L}}{\partial (\partial_{\mu} \phi)} \partial_{\mu} \delta \phi = 0$$
(3.53)
$$\left\langle \text{add and subtract } \partial_{\mu} \left(\frac{\partial \mathcal{L}}{\partial (\partial_{\mu} \phi)} \right) \right\rangle = \partial_{\mu} \left(\frac{\partial \mathcal{L}}{\partial (\partial_{\mu} \phi)} \delta \phi \right) + \frac{\delta S}{\delta \phi} \delta \phi .$$

Defining the Noether current

$$\mathcal{J}^{\mu} = -\left(\frac{\partial \mathcal{L}}{\partial(\partial_{\mu}\phi)}\delta\phi\right) , \qquad (3.54)$$

we find that

$$\partial_{\mu}\mathcal{J}^{\mu} = \frac{\delta S}{\delta\phi}\delta\phi \ . \tag{3.55}$$

Inserting into the Schwinger-Dyson equation we find that

$$0 = \partial_{\mu} \langle \mathcal{J}^{\mu} \phi(x_1) \phi(x_2) \dots \phi(x_n) \rangle - i \sum_{j=1}^{n} \langle \phi(x_1) \dots \delta \phi(x_j) \delta(x - x_j) \dots \phi(x_n) \rangle$$
(3.56)

Then the classical statement $\partial_{\mu} \mathcal{J}^{\mu} = 0$ is, like the equations of motion, true up to contact terms. In this case, the form of the contact terms depends on the symmetry through $\delta\phi$. If $\delta\phi$ does not involve time derivatives we can generate it on the operator side with the *Noether charge Q* according to

$$[\hat{Q},\hat{\phi}] = i\delta\hat{\phi} , \qquad (3.57)$$

where

$$Q = \int \mathrm{d}^{d-1} \vec{x} \,\mathcal{J}^t \,. \tag{3.58}$$

In CFTs we will find that the contact terms can be (relatively) easily computed via the residue theorem due to the large amount of extra symmetry.

Operator Product Expansion (OPE)

The product of two local field operators $\mathcal{O}_i(x)$ and $\mathcal{O}_j(y)$ can in general be singular if they are defined in the same point x = y. For example, Poincaré invariance for the free scalar field tells us that

$$\langle 0 | \phi(x)\phi(y) | 0 \rangle \sim \frac{1}{|x-y|^2} ,$$
 (3.59)

which is clearly divergent as $x \to y$. In some cases, it is useful to expand the product of two operators in a Laurent series, keeping only the most singular terms. Explicitly, we assert that we can write down the *Operator Product Expansion* (OPE) on the form

$$\mathcal{O}_i(x)\mathcal{O}_i(y) = \sum_k F_{ij}{}^k(x-y)\mathcal{O}_k(y) , \qquad (3.60)$$

as x approaches y. That we can expect to expand the product of two local operators in terms of a single local operator is nontrivial because the operators are distribution-valued, and should be seen either as an axiom or as a result of axiomatic quantum field theory.

It is not readily obvious what information we gain from doing this, we are now only dealing with one operator but we also have a possibly infinite set of unknown coefficients in $F_{ij}{}^k$. The reason why we gain something by doing this is that the Ward identities and the renormalization group equations³ of a theory may impose severe restrictions on $F_{ij}{}^k$. For example, in two dimensional CFT (section 3.4.4) the full correlation functions are determined only by the residues of the contact terms as a result of the Ward identities.

3.1.4 Formal Representation of States

In the previous sections, we have only treated the path integral as a method of computing correlators. In this section we discuss how to formally construct states, starting with a path integral representation of the vacuum ket. This section borrows heavily from the great presentation of

 $^{^{3}}$ To be covered in section 3.3.

Hartman [22]. Thinking about path integrals in this way also turns out to be very useful for field theory on curved spacetimes, as well as quantum field thermodynamics.

We may consider the object

$$|\Psi\rangle = e^{-i\hat{H}T} |\phi_i\rangle = \int_{\phi_i} \mathcal{D}\phi e^{-i\int_0^T \mathrm{d}^4 x \mathcal{L}}$$
(3.61)

with the boundary condition $\phi = \phi_i$ at t = 0, and an undefined upper boundary condition at the variable time T a fully valid definition of the Schrödinger picture wavefunction. The object $|\Psi\rangle$ is a functional that takes a state $\langle \phi_f |$ represented as the insertion of an upper boundary condition $\phi = \phi_f$ at time T as an argument and translates it to a complex number according to

$$\langle \phi_f | \Psi \rangle = \int_{\phi_i}^{\phi_f} \mathcal{D}\phi e^{-i\int_0^T \mathrm{d}^4 x \,\mathcal{L}} \,, \qquad (3.62)$$

which is consistent with the operator form of the correlation function. The object

$$e^{-i\hat{H}T} = \int \mathcal{D}\phi e^{-i\int_0^T \mathrm{d}^4 x \,\mathcal{L}} \,, \tag{3.63}$$

with two unspecified boundary conditions at times 0 and T is then the path integral representation of the time evolution operator. Inserting operators like we did in the correlation functions of section 3.1.3 it becomes a path integral representation of a time-ordered correlator between the operators evaluated using the states provided as boundary conditions.

Formally we can take the time evolution to go in the complex direction, i.e. we set $T = -i\tilde{T}$ with \tilde{T} real and we substitute $t \to -i\tau$ to find

$$|\Psi\rangle = e^{-\hat{H}\tilde{T}} |\phi_i\rangle = \int_{\phi_i} \mathcal{D}\phi e^{\int_0^{-\bar{T}} \mathrm{d}^3 x \mathrm{d}\tau \mathcal{L}} .$$
(3.64)

But the Euclidean time evolution operator will dampen all but the lowest energy state, so we find that we can compute the vacuum state by considering

$$|\Omega\rangle = \lim_{\tilde{T} \to \infty} \int_{\phi_i} \mathcal{D}\phi e^{-\int_0^{\tilde{T}} \mathrm{d}^3 x \mathrm{d}\tau \mathcal{L}} , \qquad (3.65)$$

where $|\phi_i\rangle$ is some arbitrary initial state with some overlap with the vacuum state, and we do not specify a boundary condition at infinity. This $|\Omega\rangle$ is not normalized, and might even have infinite norm. In general, as has been the case earlier in this text, any correlation function of operators has to be divided by $\langle \Omega | \Omega \rangle$ to ensure a normalized result. More pictorially, we could represent the vacuum by

$$|\Omega\rangle = \left[\begin{array}{c} & & & \\ & & & & \\ & & & \\ & & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & &$$

here the dashed line tells us to insert a boundary condition $\phi_i(0, \mathbf{x})$, where \mathbf{x} denotes all spatial coordinates, and the region of integration is bounded with by lines. If we wish to compute states that are obtained in the Heisenberg picture as

$$|\Psi\rangle = A \left|\Omega\right\rangle \ , \tag{3.67}$$

we can insert the operator as a function at some constant time into the integrand of the path integral:

$$|\Psi\rangle = \lim_{\tilde{T} \to \infty} \int_{\phi_i} \mathcal{D}\phi A[x_1, \tau_1] e^{-\int_0^{\tilde{T}} d^3x d\tau \mathcal{L}} = \left[\begin{array}{c} & & & & \\ & & & & \\ & & & \\ & & & & \\ & & & \\ & & &$$

granted that (x_1,τ_1) is in the integration region of the action. This is of course entirely analogous to how we introduced field operators into the path integral when computing correlation functions. In this sense we have already proved that this representation of the operator A in the path integral formalism is consistent with the operator formalism.

Just like we create the vacuum ket with a path integral to $-\infty$ in Euclidean time, we can generate a vacuum bra via an integral to plus infinity in Euclidean time. The vacuum to vacuum correlator $\langle \Omega | \Omega \rangle$ is two path integrals to positive and negative infinity respectively in Euclidean time, and we should glue them together. To see that we should glue, consider inserting a completion relation;

$$\langle \Omega | \Omega \rangle = \sum_{i} \langle \Omega | \phi_i \rangle \langle \phi_i | \Omega \rangle \quad . \tag{3.69}$$

The first term is a path integral over the upper half plane, the second is a path integral over the lower half plane. Summing over all boundary conditions ϕ_i ensures continuity across the boundary, gluing the two halves together. Thus, pictorially we have

$$\langle \Omega | \Omega \rangle = -\infty$$

$$\tau \uparrow_{\downarrow} \times \\ \bullet \qquad (3.70)$$

Expectation values of local operators $\mathcal{O}_1, \mathcal{O}_2$ in the vacuum state $|\Omega\rangle$ can be represented in the same manner, just inserting operators in the right place:

$$\langle \Omega | \mathcal{O}_1 \mathcal{O}_2 | \Omega \rangle = \int \mathcal{D} \phi \mathcal{O}_1(x_1) \mathcal{O}_2(x_2) e^{-\int_{-T}^{T} \mathrm{d}\tau \int \mathrm{d}^3 x \mathcal{L}} = - \mathbf{e} \left[\begin{array}{c} & & \\ \mathcal{O}_{\mathbf{i}}(\mathbf{x}_{\mathbf{i}}) & & \\ &$$

Importantly, note that the operator ordering is important on the left-hand side, but not on the right. To understand expectation values of time-ordered operators that are inserted at some Lorentzian time t, we need to complicate things a bit. Essentially, we need to 'fold' the path integral at $\tau = 0$,

integrate in the Lorentzian time direction until we pick up the Lorentzian operators and integrate back to the Euclidean time axis. Pictorially, assuming $t_1 > t_2$ we have



(3.72)

This construction is analogous to the one carried out in equation (3.29) when we first found the path integral representation of correlation functions. As indicated, the path integral starts from $t = -i\infty$ and is evolved to t = 0 to construct the vacuum ket. Then a factor e^{-iHt_2} evolves the vacuum ket to time $t = t_2$, we insert the operator \mathcal{O}_2 and continue the evolution with the operator $e^{-iH(t_2-t_1)}$ to time t_2 . The path integral is then evolved back down to t = 0 by e^{iHt_2} and we complete the path integral by going to $t = i\infty$ to construct the vacuum bra.

Rarely do we need to perform such a folded path integral as in equation (3.72). We have two simpler options in its place;

- We can compute the Euclidean path integral with the insertion points as arbitrary parameters. In this case the correlator becomes as in equation (3.71), except it is a function of the insertion points of the Euclidean operators. We can then analytically continue the correlation function to Lorentzian time to find the time-ordered correlators. An interesting characteristic of this approach is that when operators enter each other's light cones, analyticity might break down.
- We can use an $i\epsilon$ -prescription to compute the Lorentzian path integral. The usual $i\epsilon$ prescription that we used in section 3.1.3 is just a deformation of the integration contour in equation (3.72) and computes the exact same quantity⁴.

3.2 Thermal Field Theory

In the section about the path integral formulation of quantum field theory it was claimed that there is a clear connection between thermodynamics and the path integral. In this section the basics of of *thermal field theory*, which is the generalization of thermodynamics to quantum fields shall be treated. More specifically, the real and imaginary time formalisms as well as thermal Green's functions will be introduced.

⁴Although the operators look to be inserted somewhere not on the Lorentzian time axis in the case of $t = (1+i\epsilon)\tau$, τ real, remember that at the end we take the limit $\epsilon \to 0$ in the sense that we can insert the operators arbitrarily close to where they are supposed to be defined.

The key observation in thermal field theory is that the density operator of the canonical ensemble⁵, $\rho = e^{-\beta \hat{H}}$ can be interpreted as Hamiltonian time evolution by $t = -i\beta$. Moreover, the expectation of operators \mathcal{O} in a state described by a density operator is given by $\text{Tr}[\rho \mathcal{O}]$. The trace tells you to glue the strips t = 0 and $t = -i\beta$ together, putting the path integral on a circle of circumference β .

3.2.1 Thermodynamical Ensembles

Remember from statistical mechanics that in a thermal ensemble, that is an ensemble that represents a state in thermal equilibrium. The probability distribution (p(x)) over microstates x that maximizes the entropy $S = -\sum_{x} p(x) \ln p(x)$ under the constraint that all known thermodynamical quantities are held constant is defined as thermal. Typical examples of such properties are the temperature T, the volume V and the total energy E.

In the quantum case where the state of the system is defined by a density operator

$$\rho = \sum_{n} p(n) \left| n \right\rangle \left\langle n \right| \tag{3.73}$$

maximizing the entropy entails picking the p(n) so that $S = -\text{Tr}[\rho \ln \rho]$ is maximized under the constraint $\sum_n p(n) = 1$. For a finite dimensional Hilbert space with no extra restrictions this is just given by ρ equal to the maximally mixed state, but for infinite dimensional Hilbert spaces we need some other physical constraint to end up with a finite entropy. In the following a brief reminder of the various thermal ensembles that come from different physical constraints is given.

Microcanonical Ensemble

The microcanonical ensemble is defined by the restriction that we know the energy E of the microstate of the system. The distribution that maximizes the entropy is still given by a uniform distribution, but this time only over the states n for which $\hat{H} |n\rangle = E_n |n\rangle$. Denoting the number of different states that have energy E by $\Omega(E)$ the entropy is given by $S(E) = \ln \Omega(E)$. In the case of QFT the operator \hat{H} has a continuous spectrum and we can safely define the inverse temperature

$$\frac{1}{T} = \frac{\mathrm{d}S}{\mathrm{d}E} \tag{3.74}$$

In section 4.2.3 we use that the energy of a black hole is directly related to its mass to express it as a microcanonical ensemble and then compute the entropy using equation (3.74).

Canonical Ensemble

Often, it is more convenient to work with the canonical ensemble than the microcanonical. We do not usually know the exact energy of the system under study. We then consider a system of fixed particle number N_a (where the index *a* denotes particle species), volume *V* connected to a reservoir of temperature *T*. The probability of finding the system in a state $|n\rangle$ with energy E_n is proportional to $e^{-\beta \hat{E_n}}$ where $\beta \equiv 1/T$ and *T* has units of energy since we set $k_B = 1$. The partition function is the normalization factor for the density operator and is given by

$$Z_{\rm can} = \sum_{n} e^{-i\beta E_n} = \operatorname{Tr}[e^{-\beta \hat{H}}]$$
(3.75)

⁵We set Boltzmann's constant k_B to unity in $\rho = e^{-\frac{1}{k_B}\beta \hat{H}}$. This convention will be kept for the rest of the text.

where we can perform the trace by picking the orthonormal basis in definition equation (2.14) to be the basis of energy eigenstates. The equilibrium state of the system is then described by the density operator

$$\rho_{\rm can} = \frac{e^{-\beta H}}{Z_{\rm can}} \,. \tag{3.76}$$

The expectation value for an operator \mathcal{O} is then given by

$$\langle \mathcal{O} \rangle = \text{Tr}[\mathcal{O}\rho_{\text{can}}]$$
 (3.77)

and the expectation value of the energy $\langle \hat{H} \rangle \equiv E$ for the ensemble is

$$E = \text{Tr}[\hat{H}\rho_{\text{can}}] = -\frac{\partial}{\partial\beta} \ln Z_{\text{can}} . \qquad (3.78)$$

The entropy S is defined by

$$\operatorname{Tr}[-\rho_{\operatorname{can}}\ln\rho_{\operatorname{can}}] = \operatorname{Tr}[\rho_{\operatorname{can}}(\beta\hat{H} + \ln Z_{\operatorname{can}})] = \beta E + \ln Z_{\operatorname{can}} , \qquad (3.79)$$

since $\text{Tr}[\rho_{\text{can}}] = 1$. Defining the free energy F = E - ST we find explicitly that

$$F = -T \ln Z_{\rm can} \ . \tag{3.80}$$

The free energy F is a thermodynamic potential of the canonical ensemble. F is a function of the parameters with which we defined the ensemble, meaning it is a function of the volume V, the particle number N_a and the temperature T, i.e.

$$F = F(T, V, N_a) . aga{3.81}$$

The free energy $F(T,V,N_a)$ is a potential in the sense that it is a scalar quantity, and the other interesting thermodynamic quantities can be obtained by considering its derivatives. Specifically, we have

$$S = -\left(\frac{\partial F}{\partial T}\right)_{V,N_a} , \quad p = -\left(\frac{\partial F}{\partial V}\right)_{T,N_a} , \quad \mu_a = \left(\frac{\partial F}{\partial N_a}\right)_{V,T} , \quad (3.82)$$

where S is the entropy, p the pressure and μ_a the chemical potential of the particle of species a, and subscripted quantities are held fixed. In terms of exact differentials, we have

$$dF = -pdV - SdT + \mu_a dN_a . aga{3.83}$$

In quantum field theory, the canonical ensemble is not quite general enough, since particles can be created and destroyed even in a closed system. Because of this we must introduce the *grand conical ensemble* in which particle number is allowed to fluctuate.

Grand Canonical Ensemble

In the grand canonical ensemble, the temperature T, the volume V and chemical potential μ_a are taken to be known. The grand canonical ensemble has the density operator

$$\rho_{\text{grand}} = \frac{e^{-\beta(\hat{H} - \mu_a \hat{Q}_a)}}{Z_{\text{grand}}} , \qquad (3.84)$$

where \hat{Q}_a is the quantum operator that counts the number of particles of species a in a state, and Z_{grand} is the grand canonical partition function

$$Z_{\text{grand}} = \text{Tr}[e^{-\beta(\hat{H} - \mu_a \hat{Q}_a)}] .$$
(3.85)

The operators \hat{Q}_a are identified with the conserved Noether charges associated with the specific field theory in question. In particular, this ensures that spontaneous creation of particle-anti particle pairs does not contribute to the particle number, only particle exhchange with some external reservoir is counted. For simplicity we take the \hat{Q}^a to commute (thus not considering non-abelian gauge symmetries) and we refer to $N_a = \langle Q_a \rangle$ as the *particle number* of species a.

The grand canonical potential $\Omega(T, V, \mu_a)$ is defined by

$$\Omega(T, V, \mu_a) = \langle \hat{H} \rangle - TS - \mu_a \langle \hat{Q}_a \rangle = E - TS + \mu_a N_a .$$
(3.86)

The entropy $S = \text{Tr}[\rho_{\text{grand}} \ln \rho_{\text{grand}}]$ is given explicitly by

$$S(T,V,\mu_a) = \beta E - \mu_a N_a + \ln Z_{\text{Grand}} . \qquad (3.87)$$

This implies that the grand canonical potential becomes

$$\Omega(T, V, \mu_a) = -TZ_{\text{grand}} . \tag{3.88}$$

The free energy F = E - TS is given by

$$F(T,V,\mu_a) = \Omega(T,V,\mu_a) + \mu_a N_a = -TZ_{\text{grand}} + \mu_a N_a .$$
(3.89)

3.2.2 Thermal Field theory

The main formal development in this section is the observation that we have to redevelop the concept of time-ordering when the time is no longer restricted to the real axis, as well as introducing the path integral representation of $\text{Tr}[\rho]$. After defining time ordering for a curve in the complex *t*-plane we can introduce generating functionals to compute thermal correlation functions.

Formally $e^{-\beta \hat{H}}$ is equal to the Hamiltonian time evolution operator $e^{-i\hat{H}t}$ if we allow ourselves to set $t = -i\beta$. A path integral representation of the density operator is then

$$e^{-\beta\hat{H}} = \int \mathcal{D}\phi e^{-\int_0^\beta \mathrm{d}^D x\mathcal{L}} , \qquad (3.90)$$

with unspecified boundary conditions. The trace of the operator is defined via a complete orthonormal basis;

$$\operatorname{Tr}[e^{-\beta\hat{H}}] = \sum_{i} \langle \phi_{i} | e^{-\beta\hat{H}} | \phi_{i} \rangle = \sum_{i} \int_{\phi_{i}}^{\phi_{i}} \mathcal{D}\phi e^{-\int_{0}^{\beta} \mathrm{d}^{D}x\mathcal{L}} .$$
(3.91)

This actually has a remarkably simple interpretation in terms of the formalism we developed in section 3.1.4. Let's say we are completely uninterested in continuing our path integral beyond $\text{Im}(t) = -\beta$, so we define the interesting part of the complex plane to lie in the strip $0 \ge \text{Im}(t) > -\beta$. The completion relation over ϕ_i in the integration limit then tells us to glue the integration region together, leaving us with no boundary conditions, but with the new restriction that the ϕ must be periodic in the imaginary time direction, with period β meaning $\phi(0, \vec{x}) = \phi(i\beta, \vec{x})$. This nice property of the trace operator is the central observation that lets us do thermal field theory. This is illustrated graphically in figure 3.3



Figure 3.3: Path integral representation of $\text{Tr}[e^{-\beta \hat{H}}]$ via a path integral in imaginary time. Gluing together the strip $0 \ge \text{Im}(t) > -\beta$ using the completion relation $\sum_i |\phi_i\rangle \langle \phi_i| = 1$, we see that the trace is equal to a path integral on the circle. To ensure analyticity of the path integral, the ϕ in the path integral must be periodic, meaning $\phi(0, \vec{x}) = \phi(i\beta, \vec{x})$. This elegant property of the trace operator is the central observation that lets us do thermal field theory.

In the case where we consider fermionic fields $\psi(t, \vec{x}, \text{ we obtain an antiperiodicity condition};$ $<math>\psi(0, \vec{x}) = -\psi(i\beta, \vec{x})$. This comes about because fermions are represented by anticommuting field variables, and the upper integration limit of the path integral corresponds to *minus* the fermion bra.

The goal is now to analyze thermal Green's functions defined by

$$G^{\mathcal{C}}(x_1\dots x_n) = \left\langle T_{\mathcal{C}}(\hat{\phi}(x_1)\hat{\phi}(x_2)\dots\hat{\phi}(x_n)) \right\rangle = \operatorname{Tr}\left[T_{\mathcal{C}}(\hat{\phi}(x_1)\hat{\phi}(x_2)\dots\hat{\phi}(x_n)\rho_\beta) \right], \quad (3.92)$$

where $T_{\mathcal{C}}$ is the time ordering, and ρ_{β} the density operator of the thermal ensemble under study. When we allow the time to become imaginary, it is not generally clear what "time ordering" is supposed to mean. We can take the time ordering $T_{\mathcal{C}}$ to only defined on a curve \mathcal{C} . We restrict \mathcal{C} to curves that can be written as a single valued function γ of a single real parameter τ according to $t = \gamma(\tau)$. We can then define delta- and step functions on the curve according to

$$\theta_{\mathcal{C}}(t-t') = \theta(\tau-\tau') , \quad \delta_{\mathcal{C}}(t-t') = \delta(\tau-\tau') .$$
(3.93)

We may then define time-ordering on \mathcal{C} :

$$T_{\mathcal{C}}(\hat{\phi}(x)\hat{\phi}(x')) = \theta_{\mathcal{C}}(t-t')\hat{\phi}(x)\hat{\phi}(x') + \theta_{\mathcal{C}}(t'-t)\hat{\phi}(x')\hat{\phi}(x) .$$
(3.94)

This definition of time ordering ensures that the operator with smallest τ appear on the far right. Causality then runs in increasing τ . Now we have the necessary tools to construct the path integral formulation of the thermal Green's functions, expressing them in terms of a generating functional. Using that we have defined delta functions on C we can also define a functional derivative

$$\frac{\delta J(x')}{\delta J(x)} = \delta_{\mathcal{C}}(t - t')\delta(\vec{x} - \vec{x}') \tag{3.95}$$

in analogy with section 3.35. We can then define the generating functional Z[J] for the thermal Green's functions. In analogy with equation (3.34) the generating functional is given by

$$Z[J] = \int \mathcal{D}\phi e^{-i\int_{\mathcal{C}} \mathrm{d}^D x [\mathcal{L} + J(x)\phi(x)]} .$$
(3.96)

The thermal green's function is then given by

$$G^{\mathcal{C}}(x_1 \dots x_n) = \frac{1}{Z_0} \left(\frac{1}{i}\right)^n \frac{\delta^n Z[J]}{\delta J(x_1) \dots \delta J(x_n)} \bigg|_{J=0}, \qquad (3.97)$$

where $Z_0 = Z[J = 0]$. In the following sections we do not explicitly re-use the curve time-ordering T_c , but we keep in mind its definition to express time ordering in the curve correctly in terms of ordinary real time-ordering operators.

Let us consider what curves C are allowed. Requiring that all thermal Green's functions are analytic with respect to their time arguments x_i^0 implies in particular

$$-\beta \le \operatorname{Im}(x_i^0 - x_j^0) \le \beta , \quad \forall i, j .$$
(3.98)

In the following two sections we will discuss two different curves that respect the above condition. In the first, we choose $t = -i\tau$ with $\tau \in [0,\beta]$ and real. This *imaginary time formalism* is interesting in the sense that all correlators can be computed in a simple manner from the Feynman rules of the vacuum theory. In principle these solutions can be analytically continued to include Lorentzian time-separated operators as well, but this is only possible in rare cases where we know the exact propagators. To cover more general cases, we will then develop the *Schwinger-Keldysh* formalism, where mixed signature curves similar to the one in equation 3.72 are allowed.

Real Time Formalism

In the real time formalism, the thermal green's functions are taken to only be defined for $t = -i\tau$ with $\tau \in [0, \beta]$ real. We can see this as a Euclidean time that lives on a circle. This Euclidean time is often referred to as *periodic time* or a *thermal circle* in applications. Bosonic fields satisfy periodic boundary conditions, and fermionic field satisfy antiperiodic boundary conditions. In particular, we can Fourier expand bosonic fields ϕ and fermionic fields ψ according to

$$\phi = \sum_{n} a_{n} e^{i\omega_{n}\tau} , \qquad \omega_{n} = \frac{2\pi n}{\beta}$$

$$\psi = \sum_{n} b_{n} e^{i\omega_{n}\tau} , \qquad \omega_{n} = \frac{(2n+1)\pi}{\beta} ,$$
(3.99)

where the frequencies ω_n are the so-called *Matsubara frequencies*.

The Green's functions for the theory are straightforwardly obtained if we know the Feynman rules of the original theory. This is because we have already generalized the functional derivative method of computing correlators in such a way that it is structurally exactly the same. Specifically, the translation from the Feynman rules of the vacuum theory are given by table 3.1.

If we know the Euclidean Green's function exactly we can analytically continue back to Lorentzian time to find time-ordered correlators. When doing this, analyticity commonly breaks down once operator insertions become null separated, a characteristic that we will use in chapter 4 to identify thermal correlators. In general we might have only a numerical expression for the Euclidean Green's functions. In such a situation it is important to have techniques for computing Lorentzian time-ordered Green's functions directly. These techniques are provided by Schwinger-Keldysh formalism, in which we consider a path integration region constructed out of segments with different metric signature.

	Vacuum theory $(T = 0)$		Imaginary Time Formalism $(T \neq 0)$
Propagator	$G(k_1 \dots k_n)$	\rightarrow	$\left(\frac{1}{i}\right)^n G_E(k_1 \dots k_n)$
Momentum (loop) integrals	$\int rac{\mathrm{d}^d k}{(2\pi)^d} f(\omega,ec{k})$	\rightarrow	$-ieta\sum_n\intrac{\mathrm{d}^{d-1}k}{(2\pi)^{d-1}}f(\omega_n,ec k)$
Momentum conservation	$(2\pi)^d \delta^{(d)}(k)$	\rightarrow	$-ieta(2\pi)^{d-1}\delta_{n,0}\delta^{(d-1)}(\vec{k})$

Table 3.1: Translation of momentum space Feynman rules of the vacuum theory to the thermal theory at temperature t. In general, this involves a factor of $\frac{1}{i}$ for every functional derivative we perform on the generating functional due to the form of the Euclidean action, as well as a discretization of the energy component of the momentum.

Schwinger-Keldysh Formalism

The Schwinger-Keldysh formalism gives us the tools to directly analyze time-ordered Green's function in a thermal state. In particular, we can understand transport processes in which the system is brought out of equilibrium for a finite time.

In this section, we consider a curve C that runs along the real time axis from t_i to t_f , then moves into the lower imaginary half plane to $t = t_f - i\sigma$, then returns to $t = t_i - i\sigma$. Finally, the curve goes out to $t = t_i - i\beta$ which, due to the trace is identified with $t = t_i$. In this sense, we have periodic boundary conditions for bosons, and antiperiodic boundary conditions for fermions along the entire curve. The form of the curve C depends on a free parameter σ which takes values $\sigma \in [0, \beta]$.

The action $S[\phi]^6$ may be split into four parts, one for each straight piece of \mathcal{C} ,

$$S[\phi] = \int_{\mathcal{C}} \mathrm{d}t \int \mathrm{d}^{d-1} \mathcal{L}(\phi(t,\vec{x}))$$

$$= \int_{t_i}^{t_f} \mathrm{d}t \int \mathrm{d}^{d-1} \mathcal{L}(\phi(t,\vec{x})) - i \int_0^{\sigma} \mathrm{d}\tau \int \mathrm{d}^{d-1} \mathcal{L}(\phi(t_f - i\tau,\vec{x})))$$

$$- \int_{t_i}^{t_f} \mathrm{d}t \int \mathrm{d}^{d-1} \mathcal{L}(\phi(t - i\sigma,\vec{x})) - i \int_{\sigma}^{\beta} \mathrm{d}\tau \int \mathrm{d}^{d-1} \mathcal{L}(\phi(t_i - i\tau,\vec{x}))) , \qquad (3.100)$$

where we have not explicitly written out the $\partial \phi$ dependence of the Lagrangian. Now, we write

$$\phi_1(t,\vec{x}) \equiv \phi(t,\vec{x}) , \quad \phi_2(t,\vec{x}) \equiv \phi(t-i\sigma,\vec{x}) .$$
(3.101)

Then, we define the sources for these fields

$$J_1(t,\vec{x}) \equiv J(t,\vec{x}) , \quad J_2(t,\vec{x})J(t-i\sigma,\vec{x}) .$$
 (3.102)

In terms of these fields and sources, the Lorentzian generating functional reads

$$Z[J_1, J_2] = \int \mathcal{D}\phi \exp\left(iS[\phi] + i\int_{t_i}^{t_f} \int d^{d-1}x(\phi_1(t, \vec{x})J_1(t, \vec{x}) - \phi_2(t, \vec{x})J_2(t, \vec{x}))\right) , \qquad (3.103)$$

as seen by inspecting the two real-time integral terms in equation (3.100). Since the fields ϕ_1 and ϕ_2 are defined on different lines parallel to the real time axis, they are independent. Varying the generating functional independently with respect to the two sources we obtain the Schwinger-Keldysh propagator

$$iG_{ab}(x-y) = i \begin{pmatrix} G_{11}(x-y) & -G_{12}(x-y) \\ -G_{21}(x-y) & G_{22}(x-y) \end{pmatrix} = \frac{1}{i^2} \frac{\delta^2 \ln Z[J_1, J_2]}{\delta J_a(x) \delta J_b(y)} \Big|_{J=0} .$$
(3.104)

 $^{^{6}}$ For simplicity of notation we consider only a bosonic field. In what follows, replace commutators by anticommutators for fermions.

In the operator formalism, we time order along the curve C. This means that the ϕ_2 always come "later" than the ϕ_1 . Additionally, we have to reverse the Lorentzian time ordering for the ϕ_2 since time runs backwards. We then obtain

$$iG_{11}(x-y) = \langle T(\phi_1(x)\phi_1(y)) \rangle, \quad iG_{12} = \langle \phi_2(y)\phi_1(x) \rangle iG_{21}(x-y) = \langle T(\phi_2(y)\phi_1(x)) \rangle, \quad iG_{22} = \langle \overline{T}(\phi_2(x)\phi_2(y)) \rangle ,$$
(3.105)

where \overline{T} denotes reverse time ordering. We can relate the components of the Schwinger-Keldysh propagator to the *retarded* and *advanced* Green's functions G_A and G_B , defined by

$$G_R(x-y) = -i\theta(x^0 - y^0) \left\langle \left[\phi(x), \phi(y)\right] \right\rangle$$
(3.106)

$$G_A(x-y) = -i\theta(y^0 - x^0) \left< [\phi(y), \phi(x)] \right> .$$
(3.107)

To find this relation, let us Fourier transform to momentum space, letting

$$G(k) = \int d^d x \, e^{-ikx} G(x) \;.$$
 (3.108)

Remember that the advanced and retarded Green's functions are obtained from a momentum space representation by shifting poles on the real axis either in the positive or negative imaginary directions. Due to this, we know that in Fourier space $G_R(k) = G_A^*(k)$. Using this information, the matrix elements of the Schwinger-Keldysh propagator $G_{ab}(k)$ are given by [24]

$$G_{11}(k) = \operatorname{Re}[G_R(k)] + i \operatorname{coth}\left(\frac{\omega\beta}{2}\right) \operatorname{Im}[G_R(k)] , \qquad (3.109)$$

$$G_{12}(k) = i \frac{2ie^{-(\beta - \sigma)\omega}}{1 - e^{-\beta\omega}} \text{Im}[G_R(k)] , \qquad (3.110)$$

$$G_{21}(k) = i \frac{2ie^{-\sigma\omega}}{1 - e^{-\beta\omega}} \text{Im}[G_R(k)] , \qquad (3.111)$$

$$G_{22}(k) = -\operatorname{Re}[G_R(k)] + i \operatorname{coth}\left(\frac{\omega\beta}{2}\right) \operatorname{Im}[G_R(k)] . \qquad (3.112)$$

In the AdS/CFT correspondence, black holes are related to thermal states on the gravity side. Because of this Schwinger-Keldysh contours become important for the holographic study of black hole physics. We will make the connection between black holes and thermal QFT more clear in chapter 7.

3.3 The Renormalization Group

In this section the renormalization group is introduced. The renormalization group is an incredibly powerful tool in studying the high and low energy properties of quantum field theory.

As an example, in quantum field theory we find that in the Lagrangian of quantum electrodynamics

$$\mathcal{L}_{\text{QED}} = \overline{\psi}(i\partial \!\!\!/ + ieA - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} , \qquad (3.113)$$

the coupling constant e does not represent the electric charge as we know it from experiment. In general this is due to the fact that some Feynman diagrams of the theory are divergent, meaning that finite scattering amplitudes are impossible if this e is finite.

To solve the problem of divergent diagrams, we must introduce so-called physical fields ψ_0, A_0 , setting $\psi = Z_{\psi}^{1/2} \psi_0$ where Z_{ψ} is some potentially infinite constant. The physical fields and couplings

correspond to the ones measured in experiment and are finite. In general, the original fields are called the *bare* fields, whilst the new ones are called "physical". We then make a change of variables such that the Lagrangian is on the form

$$\mathcal{L}_{\text{QED}} = \overline{\psi}_{0}(i\partial \!\!\!/ + ie_{0}A_{0} - m_{0})\psi_{0} - \frac{1}{4}F_{\mu\nu0}F_{0}^{\mu\nu} + \overline{\psi}_{0}\delta_{\psi}(i\partial \!\!\!/ + iA_{0})\psi_{0} - \overline{\psi}_{0}\delta_{m}\psi_{0} - \frac{\delta_{F}}{4}F_{\mu\nu0}F_{0}^{\mu\nu} = \mathcal{L}_{\text{phys}} + \mathcal{L}_{\text{counterterms}} .$$
(3.114)

The constants δ_{ψ} , δ_A , δ_m and δ_F are in general infinite, and are chosen via the method of renormalization such that they cancel loop corrections at some energy scale of choice E. In this example we need four counterterms since we are renormalizing four quantities, the mass, the electric charge, the electron field ψ and the vector potential A. In general, to make good experimental predictions, we need to perform an experimental measurement at some controlled energy scale to choose a counterterm that gives the correct physical constants. In renormalizable theories only a finite number of distinct subdiagrams diverge, meaning that it is possible to fix all of the couplings with a finite number of experiments.

In this scheme, we only obtain the coupling constant e_0 at some particular energy scale E. As we move away from that scale $e_0(E)$ changes in value. As an example: to photons of increasing energy the electron looks like it has increasing charge. The more general understanding of the running of the physical coupling constants with energy scale was the work of Kenneth Wilson, with the introduction of the *renormalization group*.

The renormalization group bases itself on the idea that the running of the couplings should be viewed as a dynamical flow in the space of coupling constants as you integrate out high momentum degrees of freedom. Specifically we compute what happens to the physical coupling g(E) under a small change of the energy scale $E \to E + dE$, and then we find a differential equation

$$\beta(g) = \frac{\mathrm{d}g}{\mathrm{d}E} , \qquad (3.115)$$

for which we study the behaviour using tools of nonlinear dynamics.

For simplicity of notation, we will do the explicits using a scalar field theory, specifically ϕ^4 as a simplest possible interacting field theory. The Lagrangian is given by

$$\mathcal{L} = (\partial_{\mu}\phi)^2 + \frac{m^2}{2}\phi^2 + \frac{g}{4!}\phi^4 , \qquad (3.116)$$

where g is the coupling constant of the interaction. We then introduce an ultraviolet cutoff Λ as our regularization scheme for the infinities of the theory, meaning we only keep states with some finite momentum $|k| \leq \Lambda$. Naively it might seem most natural to define the cutoff in Minkowski space, but the existence of lightlike momenta with |k| = 0 and arbitrarily large components means such a scheme will not succeed in suppressing divergences. It is therefore best to Wick rotate to Euclidean time, impose the momentum cutoff on the Euclidean momenta, and analytically continue back to the Lorentzian case at the end.

Having chosen a regularization, the starting point for Wilson's renormalization group is the generating functional

$$Z[J] = \int \mathcal{D}\phi e^{-S_E[\phi] + \int d^d x J(x)\phi(x)} , \qquad (3.117)$$

where we have Wick rotated to Euclidean signature. We then impose the ultraviolet cutoff Λ on the Euclidean momentum of the field. After imposing the cutoff the path integral takes the form

$$Z[J] = \int \mathcal{D}\phi_{|k| < \Lambda} e^{-S_E^{\text{eff}}[\phi;\Lambda] + \int d^d x J(x)\phi(x)}$$
(3.118)

where the path integral measure is given by

$$\mathcal{D}\phi_{|k|<\Lambda} = \prod_{|k|<\Lambda} \mathrm{d}\phi(k) , \qquad (3.119)$$

and the Wilsonian effective action S_E^{eff} is determined by

$$e^{S_E^{\text{eff}}[\phi;\Lambda]} = \int \mathcal{D}\phi_{|k|>\Lambda} e^{-\mathcal{S}_E} . \qquad (3.120)$$

In the case of ϕ^4 -theory, the effective action is given by the effective Lagrangian on the form

$$\mathcal{L}_{E}^{\text{eff}} = \frac{Z(\Lambda)}{2} \partial_{\mu} \phi \partial^{\mu} \phi + \frac{m^{2}(\Lambda)}{2} \phi^{2} + \frac{g(\Lambda)}{4!} \phi^{4} + \mathcal{O}\left(\frac{1}{\Lambda^{2}}\right) , \qquad (3.121)$$

where $Z(\Lambda), m^2(\Lambda), g(\Lambda)$ are all finite functions of Λ and the fields ϕ are the physical (renormalized) fields. The term $\mathcal{O}\left(\frac{1}{\Lambda^2}\right)$ represents higher order terms that arise from loop corrections, compensating for the removal of high k Fourier modes.

Wilson's approach to the renormalization group then consisted of studying what happens when we lower the cut-off from Λ to $b\Lambda$ with $b \in [0, 1)$. That is, we integrate out the degrees of freedom between $b\Lambda < |k| < \Lambda$ to obtain a new effective action

$$Z[J] = \int \mathcal{D}\phi_{|k| < b\Lambda} e^{S_E^{\text{eff}}[\phi; b\Lambda] + \int d^d x J(x)\phi(x)} , \qquad (3.122)$$

where

$$S_E^{\text{eff}}[\phi; b\Lambda] = \int \mathcal{D}\phi_{\Lambda > |k| > b\Lambda} e^{-\mathcal{S}_E^{\text{eff}}[\phi, \Lambda]} .$$
(3.123)

In the new effective action at scale $b\Lambda$, modes with Euclidean momenta between $b\lambda < |k|\Lambda$ are no longer explicitly present. These degrees of freedom are instead encoded in the coefficients Z, mg as well as the higher order terms that we suppressed in equation (3.121). It is clear the that the procedure of integrating out higher momenta is associative, i.e. taking $b_2 < b_1$ we can either first integrate out the momenta Λ and $b_1\Lambda$, and then between $b_2\Lambda$ and $b_1\Lambda$ or we may simply go all the way at once, integrating out all the momenta between $b_2\Lambda$ and Λ . This property is why the renormalization group is called a "group". Note that the process of integrating out higher momentum degrees of freedom is irreversible, so really the renormalization group is only half a group.

Imposing a cutoff on the momentum roughly corresponds to taking spacetime to be lattice, and the process of integrating out higher momenta corresponds to increasing the lattice spacing. This sort of reduction of the degrees of freedom of the theory is called *coarse graining*, since we are modelling the continuum spacetime with an increasingly coarse lattice.

The physical idea of the procedure just introduced is to formulate a differential equation in terms of the parameter b for the evolution of the Λ -dependent couplings. Keeping the finite energy physics fixed we can then use the differential equation to understand how the Λ -dependent couplings behave at very high and very low energies. In particular, we talk about the "flow" induced by the reduction of the cutoff, as $g(\Lambda) \to g(b\Lambda)$. The flow of the coupling g is described by the Callan-Symanzik beta function

$$\beta(g(\Lambda)) = \frac{\mathrm{d}g(\Lambda)}{\mathrm{d}b} , \qquad (3.124)$$

which describes how the coupling g varies when we integrate out momentum modes. Imagining that we could instead vary Λ directly, it is clear that

$$\Lambda db = d\Lambda \implies db = \frac{d\Lambda}{\Lambda} = d(\ln \Lambda) \quad , \tag{3.125}$$

letting us write the Callan-Symanzik beta function according to

$$\beta(g(\Lambda)) = \frac{\mathrm{d}g(\Lambda)}{\mathrm{d}\ln\Lambda} \ . \tag{3.126}$$

Now, without doing anything even remotely explicit we can separate variables and integrate, to find that

$$\ln \Lambda - \ln \Lambda_0 = \int_{g_0}^{g(\Lambda)} \frac{\mathrm{d}g'}{\beta(g')} , \qquad (3.127)$$

where g_0 is the coupling constant at the renormalization scale Λ_0 that we may obtain from experiment and a subtraction scheme in the usual way. In principle we can (and should) extend this analysis to the dimensionful couplings of the theory such as the masses. Usually, we define

$$\gamma_m \equiv \frac{1}{m} \frac{\mathrm{d}m}{\mathrm{d}(\ln\Lambda)} , \quad \gamma \equiv \frac{1}{2Z} \frac{\mathrm{d}Z}{\mathrm{d}(\ln\Lambda)} , \qquad (3.128)$$

where γ_m and γ are the *anomalous dimensions* of the mass and field, respectively. The difference to the case of dimensionless couplings is that the combination $m^2\phi^2$ is dimensionless when integrated over four dimensions, so the violation of energy scale invariance is measured by how much the dimensions of ϕ and m deviate from L^{-1} . The β -functions measure instead how badly the dimensionless coupling g deviates from having dimension L^0 .

Generalities of the Renormalization Group Flow

Let us now speak very generally about the renormalization group flow of a single coupling constant g, and then shortly discuss the case of several couplings g_i . Our starting point for the discussion is the "master equation",

$$\ln \Lambda - \ln \Lambda_0 = \int_{g_0}^{g(\Lambda)} \frac{\mathrm{d}g'}{\beta(g')} \,. \tag{3.129}$$

The behaviour of the coupling constants in its full generality is very complicated, the presence of $\beta(g)$ as a nonlinear function of the g's, makes equation (3.129) a nonlinear integral equation.

In the case where $\beta(g) \propto g^{\alpha}$ with $\alpha > 1$, the RHS of equation (3.129) converges as we take $g(\Lambda) \to \infty$. This is called a *Landau pole*, and means that at some finite energy scale Λ_{\max} , the coupling constant necessarily diverges. This is the case for QED and ϕ^4 -theory⁷. Very interestingly, the existence of a Landau pole tells us that these theories cannot describe physics in a continuum spacetime. The standard model contains Landau poles, meaning it cannot describe physics at arbitrary Λ . This problem of UV incompleteness might be solved by nonperturbative contributions to the β -functions, a conjecture called asymptotic safety.

For other field theories in which β is positive and $\beta \propto g^{\alpha}$ with $\alpha \leq 1$, it is clear that Λ_{\max} can go to infinity, although the coupling g does not necessarily stay finite.

Points g_* in coupling constant space for which the beta function $\beta(g_*) = 0$ are referred to as fixed points. When $\beta(g_*) = 0$, and $\beta(g_* - \delta) > 0$ with δ positive all lower energy couplings will flow towards g_* as $\Lambda \to \infty$. In this case g_* is called a *UV fixed point*.

If $\beta(g)$ is negative for small g then g decreases with increasing Λ , and g = 0 is a UV fixed point. Such theories are called *asymptotically free*. Quantum chromodynamics was discovered to have this property to first order in 1973 by David Gross and Frank Wilczek [25], earning them a Nobel Prize.

As a final note, theories at RG fixed points become *scale invariant*, since the physical dimensionless couplings become scale independent. It has been shown under some assumptions that this

⁷At least to first order, which is at risk of being invalid since the couplings blow up.

necessarily implies full conformal invariance in two and four dimensions. For this reason, it is believed that the RG flow of for example the Standard Model is a flow between two CFTs, from the UV fixed point to the IR fixed point of the couplings.

3.4 Conformal Field Theory

The study of conformal field theory (CFT) is motivated on many fronts. It describes condensed matter systems at criticality⁸. In addition to this, it is believed that both the low-and high energy limits of the standard model might be CFTs [26]. Thus, it is clearly very interesting to study CFT as a possible UV completion of the standard model.

Beyond this older motivation for studying CFT, string theory is described two-dimensional CFT living on its worldsheet. In addition to this, it was discovered in 1997 that CFT is dual to quantum gravity via the AdS/CFT correspondence [3].

This section is largely based on section 3.2 in [24]. In this section the Dynkin labels (chiral spins) of the Lorentz group will be referred to without introduction as these are unimportant for the rest of the text. There are quick introductions to these ideas in [24].

3.4.1 Conformal Algebra

The conformal group is the set of transformations of space that preserve angles. It is an extension of the Poincaré group, since rotations and translations are both angle preserving. The additional structure imposed by the conformal group is *local scale-invariance*, i.e. invariance under transformations that scale the coordinates $x^{\mu} \rightarrow \Omega(x)x^{\mu}$. In a Lorentzian spacetime, the conformal transformations become the most general possible transformations that locally preserve causality. In particular, spacelike separated points remains spacelike separated and timelike separated points remain timelike separated.

Let us now consider a spacetime with a metric $g_{\mu\nu}(x)$ given by $ds^2 = g_{\mu\nu}(x)dx^{\mu}dx^{\nu}$. Then, conformal transformations can be seen as the transformations that leave the metric invariant up to a positive function, i.e.

$$g_{\mu\nu} \to \Omega(x)^{-2} g_{\mu\nu}(x) \equiv e^{2\sigma(x)} g_{\mu\nu}(x)$$
 (3.130)

Such a transformation changes the length of the infinitesimal line element by $ds'^2 = e^{2\sigma(x)}ds^2$, but angles are preserved since we are not mixing any eigenvectors of the metric. Note that picking $\Omega = 1$ implies the Poincaré group, so the conformal group is a direct extension.

Now, let us determine the conformal transformations for a flat spacetime, with the metric $\eta_{\mu\nu}$. For an infinitesimal transformation $x^{\mu} \to x^{\mu} + \eta^{\mu}(x)$ we find that

$$\eta_{\mu\nu} \to \frac{\partial x^{\prime\rho}}{\partial x^{\mu}} \frac{\partial x^{\prime\sigma}}{\partial x^{\nu}} \eta_{\rho\sigma} = \eta_{\mu\nu} + \partial_{\mu}\epsilon_{\nu} + \partial_{\nu}\epsilon_{\mu} + \mathcal{O}(\epsilon^2) . \qquad (3.131)$$

Using the definition of a conformal transformation, we find that if we want equality to first order between equation (3.130) and equation (3.131) we obtain

$$\partial_{\mu}\epsilon_{\nu} + \partial_{\nu}\epsilon_{\mu} = 2\sigma(x)\eta_{\mu\nu} . \qquad (3.132)$$

Contacting both sides with $\eta^{\mu\nu}$ and using $\eta_{\mu\nu}\eta^{\mu\nu} = d$ we find that

$$\partial_{\mu}\epsilon^{\mu} = d \cdot \sigma(x) , \qquad (3.133)$$

⁸When a condensed matter system undergoes a phase transition, the correlation length blows up and the system becomes scale invariant (reflected by the β -functions of the renormalization group going to zero). It is then well described by a CFT.

where d is the spacetime dimension. Combining these equations to cancel $\sigma(x)$ we find

$$\partial_{\mu}\epsilon_{\nu} + \partial_{\nu}\epsilon_{\mu} = \frac{2}{d}\eta_{\mu\nu}\partial\cdot\epsilon , \qquad (3.134)$$

which is called the conformal Killing equation, since its solutions span the set of Killing vectors of a conformally symmetric theory.

Conformal Algebra for d = 2

In the case when d = 2 the conformal Killing equation (3.134) is remarkably simple. To find this, we Wick rotate to Euclidean coordinates, letting $t \to -i\tau$, making the metric positive definite. The conformal Killing equation then becomes three component equations. Two of these turn out to be equivalent, and we have

$$2\partial_0 \epsilon_0 = (\partial_0 \epsilon_0 + \partial_1 \epsilon_1)$$

$$\Rightarrow \quad \partial_0 \epsilon_0 = \partial_1 \epsilon_1 \tag{3.135}$$

and
$$\partial_1 \epsilon_0 = -\partial_0 \epsilon_1$$

with the equation for the 11 component equivalent to the 00 component of the equation. This is identified as the Cauchy-Riemann equation of complex analysis if we identify ϵ_0 as the real part and ϵ_1 as the imaginary part of a complex valued function, and x_0, x_1 as the real and imaginary parts of a complex number. It is convenient to introduce the complex coordinates $z = x_0 + ix_1, \overline{z} = x_0 - ix_1$. Back in the Lorentzian spacetime, z corresponds to the forward lightcone coordinate t + x, and \overline{z} corresponds to the backward lightcone coordinate t - x.

With this choice of coordinates $\epsilon(z) = \epsilon_0 + i\epsilon_1$ is a function of z and holomorphic. We can also define the antiholomorphic function $\overline{\epsilon}(\overline{z}) = \epsilon_0 - i\epsilon_1$, which depends only on \overline{z} . All (anti)holomorphic functions admit a Laurent expansion according to

$$\epsilon(z) = -\sum_{n \in \mathbb{Z}} \epsilon_n z^{n+1} , \quad \overline{\epsilon}(\overline{z}) = -\sum_{n \in \mathbb{Z}} \overline{\epsilon}_n \overline{z}^{z+1} .$$
(3.136)

The infinitesimal transformation is now given by $z \to z + \epsilon(z)$ and $\overline{z} \to \overline{z} + \overline{\epsilon}(\overline{z})$. The generators of a conformal transformation where the coefficients $\epsilon_n, \overline{\epsilon}_n$ are only nonzero for a single value n = kare given by

$$l_k = -z^{k+1}\partial_z , \quad \bar{l} = -\bar{z}^{k+1}\partial_{\bar{z}} . \tag{3.137}$$

To see that these are indeed the generators, consider that to first order a translation is proportional to the first derivative ∂_z , we then have to multiply by z^{k+1} since $\partial_z \epsilon(z)$ at some particular value of z is just a number. Lastly, we need a minus sign because the ϵ are defined as minus the sum over coefficients.

The commutation relations of the conformal algebra are given by

$$[l_{n}, l_{m}] = (z^{n+1}\partial_{z}z^{m+1}\partial_{z}) - (z^{m+1}\partial_{z}z^{n+1}\partial_{z})$$

$$= (m+1)(z^{n+1}z^{m}\partial_{z}) - (n+1)(z^{m+1}z^{n}\partial_{z})$$

$$= (m-n)z^{m+n+1}\partial_{z}$$

$$= (m-n)l_{m+n}$$

$$[\bar{l}_{n}, \bar{l}_{m}] = (m-n)\bar{l}_{m+n}$$

$$[l_{m}, \bar{l}_{n}] = 0.$$

(3.138)

Notably, the generators $\{l_{-1}, l_0, l_1\}$ and their complex conjugates together generate the subalgebra $sl(2, \mathbb{R}) \oplus sl(2, \mathbb{R})$ corresponding to global conformal transformations. Analogs of these will reappear

	$\epsilon^{\mu}(x)$	$\sigma(x)$	Generator
Translation	a^{μ}	0	P^{μ}
Lorentz transformation	$\omega^{\mu}{}_{\nu}x^{\nu}, \omega_{\mu\nu} = -\omega_{\nu\mu}$	0	$J_{\mu u}$
Dilatation	λx^{μ}	λ	D
Special conformal transformation	$b^{\mu}x^2 - 2(b\cdot x)x^{\mu}$	$-2(b \cdot x)$	K_{μ}

Table 3.2: Set of conformal transformations in d > 2 dimensions

in the higher dimensional case. The rest of the conformal generators in two dimensions do not have a counterpart in higher dimensions, and it is the presence of this large amount of extra symmetry that makes 2d CFT especially tractable (among field theories). We will further develop two-dimensional CFT in section 3.4.4.

3.4.2 Conformal Algebra for d > 2

The conformal Killing equation (3.134) in d > 2 dimensions has the general solution

$$\epsilon^{\mu}(x) = a^{\mu} + \omega^{\mu}{}_{\nu}x^{\nu} + \lambda x^{\mu} + b^{\mu}x^2 - 2(b \cdot x)x^{\mu}$$
(3.139)

where the scalar product denotes the contraction $b \cdot x = x_{\mu}b^{\mu}$. All of the parameters in equation (3.139) have a finite number of components, so we find that in d > 2 dimensions the conformal algebra is finite dimensional, unlike the d = 2 case. The parameters each have a geometric interpretation, given in table 3.2.

The generators of the translations a^{μ} are given by the momenta P^{μ} , the generators of the Lorentz boosts $\omega^{\mu}{}_{\nu}$ will be denoted $J_{\mu\nu}$, the dilatation (scale by λ) generator is D, and the generator of the special conformal transformations is K^{μ} . The full commutation relations of the conformal algebra are then given by

$$[J_{\mu\nu}, J_{\rho\sigma}] = i \left(\eta_{\mu\rho} J_{\nu\sigma} + \eta_{\nu\sigma} J_{\mu\rho} - \eta_{\mu\sigma} J_{\nu\rho} - \eta_{\nu\rho} J_{\mu\sigma} \right)$$

$$[J_{\mu\nu}, P_{\rho}] = -i \left(\eta_{\mu\rho} P_{\nu} - \eta_{\nu\rho} P_{\mu} \right)$$

$$[J_{\mu\nu}, K_{\rho}] = i \left(\eta_{\mu\rho} K_{\nu} - \eta_{\nu\rho} K_{\mu} \right)$$

$$[D, P_{\mu}] = i P_{\mu}$$

$$[D, K_{\mu}] = -i K_{\mu}$$

$$[D, J_{\mu\nu}] = 0$$

$$[K_{\mu}, K_{\nu}] = [P_{\mu}, P_{\nu}] = 0$$

$$[K_{\mu}, P_{\nu}] = -2i (\eta_{\mu\nu} - J_{\mu\nu}) .$$

(3.140)

It turns out that the conformal generators can be grouped in such a way that the conformal algebra is isomorphic to so(d,2). The generators in so(d,2) are denoted $\overline{J}_{AB} = -\overline{J}_{BA}, A \in [0,1\ldots,d+1]$ and fulfill the same commutation relation as the $J_{\mu\nu}$ with η replaced by $\overline{\eta} = diag(-1,1,\ldots,1,-1)$. This is found by identifying $\overline{J}_{\mu\nu} \equiv J_{\mu\nu}$. For the remaining 2d + 1 generators, $\overline{J}_{\mu d}, \overline{J}_{\mu(d+1)}$ and $\overline{J}_{d(d+1)}$ we need to express them in terms of P_{μ}, K_{μ} and D. Incidentally, P_{μ}, K_{μ} and D happen to have a total of 2d + 1 components. The generator $\overline{J}_{d,d+1}$ must be invariant under Lorentz transforms in the first d indices ($\mu = 0, \ldots, d-1$)), i.e. it must commute with $J_{\mu\nu}$. One of our generators already fulfills this, so we can identify

$$\overline{J}_{d,d+1} = -D . (3.141)$$

Similarly, $\overline{J}_{\mu d}$ and $\overline{J}_{\mu(d+1)}$ have one free vector index that transforms as a tensor under *d*-dimensional Lorentz boosts, so $\overline{J}_{\mu d}$ and $\overline{J}_{\mu(d+1)}$ must be linear combinations of K_{μ} , P_{μ} . To derive the precise relation, we just make a linear combination ansatz on the form $\overline{J}_{\mu d} = AP_{\mu} + BK_{\mu}$ and check the commutation relations to determine the constants. We omit this (rather messy) exercise in algebra and state the result

$$\overline{J}_{\mu d} = \frac{1}{2}(K_{\mu} - P_{\mu}) , \quad \overline{J}_{\mu(d+1)} = \frac{1}{2}(K_{\mu} + P_{\mu}) .$$
 (3.142)

What we have shown is then that the conformal group in a flat spacetime with signature (d-1, 1) is described by the symmetry algebra so(d, 2). No step in the previous derivation depends on the explicit choice of signature for $\eta_{\mu\nu}$, so the conformal group in a general spacetime with p spacelike dimensions and q timelike dimensions is represented by SO(p+1, q+1).

Finite Transformations

We have now fully characterized the infinitesimal properties of the conformal group in d > 2 dimensions. Let us now turn our eyes to finite transformations. Of interest are the scale transformations, special conformal transformations and the inversion

$$x^{\mu} \to \lambda x^{\mu}$$
, (3.143)

$$x^{\mu} \to \frac{x^{\mu} + b^{\mu} x^2}{1 + 2b \cdot x + b^2 x^2}$$
, (3.144)

$$x^{\mu} \to \frac{x^{\mu}}{x^2} . \tag{3.145}$$

The special conformal transformation and inversions are not globally defined, since they have singularities at $1 + 2b \cdot x + b^2 x^2 = 0$ and $x^2 = 0$ respectively. In order to have globally defined conformal transformations we need to add "points at infinity" to our spacetime. In technical terms, this is called considering the conformal compactification of $\mathbb{R}^{d-1,d}$, meaning essentially that we take the infinities to actually be part of our set, such that we are considering a compact (infinite) manifold. This is a Lorentzian analog of the Riemann sphere of complex analysis.

The inversion is the only finite transformation that is not obtainable by integrating infinitesimal transforms, since it is not connected to the identity. Any overall transformation that involves two inversions is connected to the identity, as an example the special conformal transformation can be written as a combination of two inversions and a translation.

For any conformal transformation we may define

$$\mathcal{R}^{\mu}{}_{\rho}(x) = \Omega(x) \frac{\partial x'^{\mu}}{\partial x^{\rho}}.$$
(3.146)

It is straightforward to see that $\mathcal{R}^{\mu}{}_{\rho}(x)$ is a local Lorentz transform, since

$$\mathcal{R}^{\mu}{}_{\rho}(x)\mathcal{R}^{\nu}{}_{\sigma}(x)\eta_{\mu\nu} = \Omega^2 \frac{\partial x'^{\mu}}{\partial x^{\rho}} \frac{\partial x'^{\nu}}{\partial x^{\sigma}} \eta_{\mu\nu} = \eta_{\rho\sigma} , \qquad (3.147)$$

remembering the definition of $\Omega(x)$ in terms of the scaling of η under conformal transformations given by equation (3.130). This will be useful for the construction of the conformal correlation functions. For the inversion $x'^{\mu} = x^{\mu}/x^2$ we have

$$\Omega(x) = x^2 \tag{3.148}$$

and the local Lorentz transformation is given by

$$\mathcal{R}^{\mu\nu} = x^2 \left(\frac{\delta^{\mu\nu}}{x^2} - 2\frac{x^{\mu}x^{\nu}}{x^4}\right) = \eta^{\mu\nu} - 2\frac{x^{\mu}x^{\nu}}{x^2} \equiv I^{\mu\nu}(x) , \qquad (3.149)$$

where we have defined the *inversion matrix* $I^{\mu\nu}(x)$. For two-points x, y we have that

$$I^{\mu\nu}(x'-y') = \mathcal{R}^{\mu}{}_{\rho}(x)\mathcal{R}^{\nu}{}_{\sigma}(y)I^{\rho\sigma}(x-y)$$
(3.150)

since the \mathcal{R} are linear and preserve the metric $\eta^{\mu\nu}$. In particular this implies that the scalar $(x-y)^2$ must transform in the following fashion

$$(x' - y')^2 = \frac{(x - y)^2}{\Omega(x)\Omega(y)} .$$
(3.151)

In addition we can use the inversion to define a vector that transforms nicely under conformal transformation. The vector Z^{μ} is defined in the point z and is constructed from three points x,y,z according to

$$Z^{\mu} = \frac{x^{\mu} - z^{\mu}}{(x - z)^2} - \frac{y^{\mu} - z^{\mu}}{(y - z)^2} , \qquad (3.152)$$

and it squares to

$$Z^{2} = \frac{(x-y)^{2}}{(x-z)^{2}(y-z)^{2}} .$$
(3.153)

Under a conformal transformation the vector Z^{μ} transforms covariantly, i.e. it transforms by a scaling plus times a local boost:

$$Z^{\prime\mu} = \Omega(z) \mathcal{R}^{\mu}_{\nu} Z^{\nu} . \qquad (3.154)$$

Similar vectors defined at the points x, y may be obtained via cyclic permutation.

3.4.3 Conformal Field Theory in General Dimension

With the characterization of the conformal algebra out of the way, we can now do field theory. In order to construct the representation of the conformal group in a field theory of general dimension we use the method of induced representations. The recipe is as follows; first we determine the properties of the fields $\phi(x)$ at x = 0, then we can use the momentum operator P^{μ} to translate the argument to an arbitrary point in spacetime to find the general transformation rule. We consider fields ϕ of arbitrary spin.

Fully solving CFTs in d > 2 is in general not possible unless we introduce additional symmetries to further constrain the theory, such as supersymmetry. The best we can do is use this method of induced representations to get some restrictions on the form of correlation functions. The greatest weakness of this approach is that since we are not coming from a canonical quantization, we have no idea if the quantum operators we specify are compatible with a unitary quantum theory. Despite this we can still gain enough a great deal of information about the behaviour of general CFTs. In addition, as we will learn in the next chapter, entanglement entropy can be computed without reference to the precise field content of the theory.

Let us first remind ourselves of what happens when we promote ϕ to an operator. Before, in terms of some generator T_a , the field transformed as

$$\phi \to \phi' = e^{-iT_a x^a} \phi = \phi(x) - iT_a x^a \phi + \mathcal{O}(x^2) \equiv \phi + \delta_T \phi .$$
(3.155)
But an operator $\hat{\phi}$ transforms under a change of basis according to

$$\hat{\phi} \to \phi + \delta_T \phi = e^{-iT_a x^a} \hat{\phi} e^{iT_a x^a} . \tag{3.156}$$

Thus, to first order in the parameter x we find that

$$\delta_T \hat{\phi} = x^a (-iT_a \hat{\phi} + i\hat{\phi}T_a) = -ix^a [T_a, \phi] , \qquad (3.157)$$

but $-ix^a$ is just the argument generating the finite transformation, dividing it off, we can write for infinitesimal transformations

$$\delta_T \hat{\phi} = [T, \hat{\phi}] , \qquad (3.158)$$

meaning that the infinitesimal transformation of $\hat{\phi}$ is *proportional* to $[T,\phi]$. In proceeding, we drop the hats on operators, keeping in mind that we should treat all objects as Heisenberg operators.

For the *d*-dimensional Lorentz transformations we postulate that

$$[J_{\mu\nu}, \phi(0)] = -\mathcal{J}_{\mu\nu}\phi(0) \tag{3.159}$$

where $\mathcal{J}_{\mu\nu}$ is some finite-dimensional representation of the Lorentz group. In addition we postulate that

$$[D, \phi(0)] = i\Delta\phi(0) , \qquad (3.160)$$

implying that ϕ transforms under dilatations as

$$\phi(x) \to \phi'(x') = \lambda^{-\Delta} \phi(x) . \qquad (3.161)$$

We call Δ the scaling dimension of ϕ . Specifically, all fields that transform under an irreducible representation of the conformal group must be eigenstates of the dilatation operator D. In radial quantization, which will be introduced in section 3.4.4, this is motivated by the fact that the dilation operator generates radial (time) translation, meaning it is the Hamiltonian.

By way of gauge fixing, it is sufficient to consider only *conformal primary fields* ϕ that are defined to satisfy the commutation relation

$$[K_{\mu}, \phi(0)] = 0 . (3.162)$$

We can then show that P_{μ} increases the scaling dimension of ϕ by considering the commutator

$$[D, P_{\mu}\phi(0)] = P_{\mu}[D, \phi(0)] + [D, P_{\mu}]\phi(0)$$

= $iP_{\mu}\Delta\phi(0) + iP_{\mu}\phi(0)$
= $i(\Delta + 1)P_{\mu}\phi(0)$ (3.163)

In a similar fashion we find that K_{μ} decreases the scaling dimension of ϕ . Based on equation (3.162) we see that the conformal primaries are of lowest possible scaling dimension in a given conformal multiplet. All other fields are *descendants* of this multiplet, obtained by acting with P_{μ} on the conformal primary.

At this point we have the transformation properties of the field at x = 0. Let us now introduce the translation operator $\mathcal{T}(a) = e^{-iP_{\mu}a^{\mu}}$ which acts on operators according to

$$\mathcal{T}(a)\phi(0)\mathcal{T}^{-1} = \phi(a)$$
 . (3.164)

The infinitesimal variation of the field generated by for example P_{μ} at x is

$$\delta_{P}\phi = [P_{\mu}, \phi(x)]$$

$$= [P_{\mu}, e^{-iP_{\mu}x^{\mu}}\phi(0)e^{iP_{\mu}x^{\mu}}]$$

$$= P_{\mu}e^{-iP_{\sigma}x^{\sigma}}\phi(0)e^{iP_{\rho}x^{\rho}} - e^{-iP_{\sigma}x^{\sigma}}\phi(0)e^{iP_{\rho}x^{\rho}}P_{\mu}$$

$$= i\partial_{\mu}(e^{-iP_{\mu}x^{\mu}}\phi(0)e^{iP_{\mu}x^{\mu}}) .$$
(3.165)

It is similarly possible to deduce the remaining commutation relations for a conformal primary field $\phi(x)$,

$$[P_{\mu}, \phi(x)] = -i\partial_{\mu}\phi(x) \equiv \mathcal{P}_{\mu}\phi(x) ,$$

$$[D, \phi(x)] = -i\Delta(x)\phi(x) - ix^{\mu}\partial_{\mu}\phi(x) \equiv \mathcal{D}\phi(x)$$

$$[J_{\mu\nu}, \phi(x)] = -\mathcal{J}_{\mu\nu}\phi(x) + i(x_{\mu}\partial_{\nu} - x_{\nu}\partial_{\mu})\phi(x) \equiv \tilde{\mathcal{J}}_{\mu\nu} ,$$

$$[K_{\mu}, \phi(x)] = \left(i(-x^{2}\partial_{\mu} + 2x_{\mu}x^{\rho}\partial_{\rho} + 2x_{\mu}\Delta) - 2x^{\nu}\mathcal{J}_{\mu\nu}\right)\phi(x) \equiv \mathcal{K}_{\mu}\phi(x) .$$
(3.166)

The new operators $\mathcal{P}_{\mu}, \mathcal{D}, \mathcal{J}_{\mu\nu}, \mathcal{K}_{\mu}$ can once again be combined to form a representation of so(p + 1, q + 1) i.e. a representation of the conformal algebra in p + q dimensions.

The general transformation $x \to x' = x + \epsilon(x)$, with ϵ given as in equation (3.139) now takes the form

$$\delta_{\epsilon}\phi(x) = -\epsilon_{\mu}\partial^{\mu}\phi(x) = -\left(\epsilon \cdot \partial + \frac{\Delta}{d}\partial \cdot \epsilon - \frac{i}{2}\partial_{\mu}\epsilon_{\nu}\mathcal{J}^{\mu\nu}\right)\phi(x) . \qquad (3.167)$$

Unitarity Bound on CFTs

Due to something called the *unitarity bound*, there is a lower bound on the scaling dimension Δ of the field ϕ . Finding the unitarity bounds is in general not easy, so we will just state them here.

Let us consider the subalgebra $so(1,1) \oplus so(3,1) \in so(4,2)$ of the conformal algebra, corresponding to dilatations and Lorentz transformations⁹. Additionally consider the decomposition of $so(3,1) = su(2)_L \oplus su(2)_R$ where the subscripts denote chirality.

We may then label the representations in this subalgebra by (Δ, j_L, j_R) where Δ is the scaling dimension and j_l, j_r the spin of the corresponding su(2) representations in which a general field may transform. In terms of these quantities, the unitarity bounds in 3 + 1 dimensions are given by

$$\Delta \ge 1 + j_L \text{ for } j_R \ne 0$$

$$\Delta \ge 1 + j_R \text{ for } j_L \ne 0$$

$$\Delta \ge 2 + j_L + j_R \text{ for both } j_R, j_L \ne 0$$
(3.168)

Some specific unitarity bounds are given as follows:

- Scalar fields transform in $(\Delta, 0, 0)$ and thus must have $\Delta \ge 1$
- A chiral spinor transforms in either $(\Delta, 1/2, 0)$ or $(\Delta, 0, 1/2)$ and must have $\Delta > 3/2$.
- A vector transforms in $(\Delta, 1/2, 1/2)$ and we find $\Delta \geq 3$.
- A symmetric traceless tensor transforms in $(\Delta, 1, 1)$ and we have $\Delta \ge 4$.

⁹To see this, consider that so(1,1) has only one element, sitting in the $\overline{J}_{d(d+1)}$

In general dimension, the unitarity bounds for scalars, spin 1/2 and spin s respectively are¹⁰

$$\begin{split} &\Delta \geq \frac{d}{2} - 1 \ , \\ &\Delta \geq \frac{d-1}{2} \ , \\ &\Delta \geq d+s-2 \end{split}$$

Conserved Charges of the CFT

Conformal symmetry is a continuous symmetry, and therefore Noether's theorem applies. For the translations the conserved current is the stress-energy tensor $T_{\mu\nu}$, while for the Lorentz boosts we have $N_{\mu\nu\rho} = x_{\nu}T_{\mu\rho} - x_{\rho}T_{\mu\nu}$. The associated charges are the momenta and angular momenta

$$P_{\mu} = \int d^{d-1}x \, T^{0}{}_{\mu} \,, \quad M_{\mu\nu} = \int d^{d-1}x \, (x_{\mu}T^{0}{}_{\nu} - x_{\nu}T^{0}{}_{\mu}) \,. \tag{3.169}$$

The scale and special conformal transformations give rise to the conserved currents

$$J_{D\mu} = x^{\nu} T_{\mu\nu} , \quad J_{K\mu\nu} = x^2 T_{\mu\nu} - 2x_{\nu} x^{\rho} T_{\mu\rho} . \qquad (3.170)$$

The charges that generate the symmetries are then

$$D = \int d^{d-1}x \, x^{\rho} T^{0}{}_{\rho} ,$$

$$K_{\nu} = \int d^{d-1}x \left(x^{2} T^{0}{}_{\nu} - 2x_{\nu} x^{\rho} T^{0}{}_{\rho} \right) .$$
(3.171)

Since all of these charges are conserved, we have some serious restrictions on the stress-energy tensor $T_{\mu\nu}$. It should be familiar that the restriction on the stress tensor from the Poincaré charge conservation is just that it is symmetric. In addition to this, scale invariance requires that the stress-energy tensor must be *traceless*. To see this, just impose conservation on $J_{D\mu}$:

$$0 = \partial^{\mu} J_{D\mu} = \partial^{\mu} (x^{\nu} T_{\mu\nu}) = \underbrace{= (\partial^{\mu} x^{\nu})}_{\delta^{\mu\nu}} T_{\mu\nu} + x^{\nu} (\underbrace{= \partial^{\mu} T_{\mu\nu}}_{0}) = T^{\rho}{}_{\rho} .$$
(3.172)

The restriction due to conservation of K_{ν} is not as elegant, so we do not state it here.

Correlation Functions

Thanks to conformal symmetry, the correlation functions of a conformally symmetric theory may be found without resorting to perturbation theory. This is due to the fact that the resulting Ward identities heavily restrict the form of two-, three- and four-point correlators.

The invariance of the action under symmetry transformations leads to the Ward identity (3.52) for correlation functions. Specifically, we find the dilatation Ward identity

$$\sum_{i}^{n} \left(x_{i}^{\mu} \frac{\partial}{\partial x_{i}^{\mu}} + \Delta_{i} \right) \left\langle \phi_{1}(x_{1}) \dots \phi_{i}(x_{i}) \dots \phi_{n}(x_{n}) \right\rangle = 0 , \qquad (3.173)$$

with Δ_i the scaling dimension of the field ϕ_i . We will now consider scalar fields, and then move on to more general conformal primary operators.

¹⁰These are given at the end of section 3 in [27].

Let us consider the two-point correlator for conformal primary scalar fields, abbreviating $\phi_1(x_1) = \phi_1$. Then the dilatation ward identity implies that

$$\langle \phi_1 \phi_2 \rangle \xrightarrow[x \to \lambda x]{} \lambda^{\Delta_1 + \Delta^2} \langle \phi_1 \phi_2 \rangle .$$
 (3.174)

Poincaré invariance additionally tells us that the two-point correlator can only depend on $(x_1 - x_2)^2$, so the general two-point correlator is just

$$\langle \phi_1 \phi_2 \rangle = \frac{C_{\phi_1 \phi_2}}{(x_1 - x_2)^{\Delta_1 + \Delta_2}} ,$$
 (3.175)

where here we have abbreviated $((x_1 - x_2)^2)^{(\Delta_1 + \Delta_2)/2} = (x_1 - x_2)^{\Delta_1 + \Delta_2}$ and we will keep doing so. Additionally, we showed in equation (3.151) using the inversion symmetry that

$$(x' - y')^2 = \frac{(x - y)^2}{\Omega(x)\Omega(y)} .$$
(3.176)

For consistency we then see that equation (3.175) can only hold if $\Delta_1 = \Delta_2$. Moreover we expect the constant $C_{\phi_1\phi_2}$ to be symmetric under the exchange $1 \leftrightarrow 2$. We can then diagonalize $C_{\phi_1\phi_2}$ in the space of conformal primary scalar operators \mathcal{O} such that it is only nonzero for conjugated operators $\mathcal{O}, \mathcal{O}^{\dagger}$. Rescaling the operators we can set $C_{\mathcal{O}\mathcal{O}^{\dagger}} = 1$ and we find that for a scalar conformal primary operator \mathcal{O} of scaling dimension Δ that

$$\langle \mathcal{O}(x_1)\mathcal{O}^{\dagger}(x_2)\rangle = \frac{1}{(x_1 - x_2)^{2\Delta}}$$
 (3.177)

Similarly, the three point function for the conformal primary scalar operators \mathcal{O}_i with scaling dimensions Δ_i reads

$$\langle \mathcal{O}_1 \mathcal{O}_2 \mathcal{O}_3 \rangle = \frac{C_{\mathcal{O}_1 \mathcal{O}_2 \mathcal{O}_3}}{(x_1 - x_2)^{\Delta_1 + \Delta_2 - \Delta_3} (x_2 - x_3)^{\Delta_2 + \Delta_3 - \Delta_1} (x_3 - x_1)^{\Delta_3 + \Delta_1 - \Delta_2}}, \qquad (3.178)$$

where the coefficient $C_{\mathcal{O}_1\mathcal{O}_2\mathcal{O}_3}$ is nontrivially constrained by the Ward identities.

In the more complicated case of arbitrary conformal primary operators \mathcal{O}^{I} where I denotes indices on which a representation of the orthogonal group O(p,q) acts. Possible examples in d = 4are the vector current J^{μ} and the energy-momentum tensor $T^{\mu\nu}$ that transform in the (1/2,1/2)and (1,1) representations of O(3,1) respectively. It is very tedious to work out, on a case-by-case basis what the correlators look like for each nonscalar operator. Nonetheless we can once again use the somewhat high-brow method of induced representations to obtain some useful results.

Remember that we could write any O(p,q) transformation via the local operator $\mathcal{R}^{\mu}{}_{\nu} = \Omega(x) \frac{\partial x'^{\mu}}{\partial x'^{\nu}}$. Let $D(\mathcal{R})^{I}{}_{J}$ denote the appropriate representation of the local boost/rotation operator \mathcal{R} acting on \mathcal{O}^{J} . Then a general conformal primary operator transforms as

$$\mathcal{O}^I \to \Omega(x)^{\Delta} D(\mathcal{R}(x))^I{}_J \mathcal{O}^J$$
 (3.179)

It is now possible to construct conformally covariant two-point correlators. For a field \mathcal{O} transforming as in equation (3.179) we can once again diagonalize the two-point constant so that it is only nonzero when \mathcal{O} is paired with its conjugate $\overline{\mathcal{O}}$ that transforms in the conjugate representation $\overline{\mathcal{O}} \to \Omega^{\Delta} \overline{\mathcal{O}}_I (D(\mathcal{R})^{-1})^I{}_J$. For general $\mathcal{O}, \overline{\mathcal{O}}$ living in irreducible representations of O(p,q) we can write

$$\langle \mathcal{O}^{I}(x)\overline{\mathcal{O}}_{J}(y) = \frac{C_{\mathcal{O}}}{(x-y)^{2\Delta}} D(I(x-y))^{I}{}_{J} , \qquad (3.180)$$

where I(x-y) is the inversion defined in equation (3.150) and $C_{\mathcal{O}}$ is an overall constant that like in the previous case can be set to 1 by redefining the operators. The presence of I(x-y) is essentially because it has the appropriate transformation properties in the indices I, J while also depending only on the Poincaré invariant (x - y).

As examples, let us consider the conserved currents $J_{D\mu}$ and $T_{\mu\nu}$, which transform in the fundamental representation of the O(p,q) group. The conserved currents were defined in equations equation (3.169) and onward in terms of d-1 dimensional integrals. In order to cancel to cancel the integration measures and eventual extra factors of x, P_{μ} , $J_{D\mu}$ and $T_{\mu\nu}$ must have conformal weights d - 1, d - 1 and d respectively. In combination with this, equation (3.180) tells us that for any conserved vector current,

$$\langle J_{\mu}(x)J_{\nu}(y)\rangle = \frac{C_J}{(x-y)^{2(d-1)}}I_{\mu\nu}(x-y) ,$$
 (3.181)

and for the stress tensor

$$\langle T_{\mu\nu}(x)T_{\rho\sigma}(y)\rangle = \frac{C_T}{(x-y)^{2d}}\mathcal{I}^T_{\mu\nu,\rho\sigma} , \qquad (3.182)$$

where

$$\mathcal{I}_{\mu\nu,\rho\sigma}^{T} = I_{\mu\alpha}(x-y)I_{\nu\beta}(x-y)\left(\frac{1}{2}(\delta_{\rho}^{\alpha}\delta_{\sigma}^{\beta} + \delta_{\rho}^{\beta}\delta_{\sigma}^{\alpha}) - \frac{1}{d}\eta_{\rho\sigma}\eta^{\alpha\beta}\right) , \qquad (3.183)$$

in which the last term projects onto tensors symmetric and traceless in $\rho\sigma$ to enforce the symmetric tracelessness of $T_{\rho\sigma}$ in the LHS. Since the projector is also symmetric traceless in the indices $\alpha\beta$ we find that it also enforces symmetric tracelessness in $\mu\nu$.

By using the conformally covariant vector Z defined in equation (3.152) it is possible to construct the general three-point correlator for conformal primary fields. The most general form of this expression reads

$$\langle \mathcal{O}_1^I \mathcal{O}_2^J \mathcal{O}_3^K \rangle = \frac{D_1 (I(x-z))^{I_{I'}} D_2 (I(y-z))^{J_{J'}} t^{I'J'K}(Z)}{(x-z)^{2\Delta_1} (y-z)^{2\Delta_2}}$$
(3.184)

where once again D denotes the appropriate representation of O(p,q) acting on the operators. The function $t^{I'J'K}(Z)$ is an arbitrary function that has to satisfy

$$t^{I'J'K}(\lambda Z) = \lambda^{\Delta_3 - \Delta_1 - \Delta_2} t^{I'J'K}(Z) , \qquad (3.185)$$

as well as

$$D_1(\mathcal{R})^{I}{}_{I'}D_2(\mathcal{R})^{J}{}_{J'}D_1(\mathcal{R})^{K}{}_{K'}t^{I'J'K'}(Z) = t^{IJK}(\mathcal{R}Z) .$$
(3.186)

These conditions are sufficient to ensure that the three-point correlator satisfies the conformal Ward identity equation in (3.173). A particularly interesting case is that of three conformal primary scalar fields, for which

$$t(Z) = C_{\mathcal{O}_1 \mathcal{O}_2 \mathcal{O}_3} \left(\frac{(x-z)(y-z)}{(x-y)} \right)^{\Delta_1 + \Delta_2 - \Delta_3}$$
(3.187)

is the most general t that satisfies equation (3.185) and equation (3.186). Inserting equation (3.187) into the three-point-function defined in equation (3.184) it is straightforward to see that it simplifies to equation (3.178) since the scalars transform in the trivial representation of O(p,q).

The scalar t(Z) is significant because it generically represents the leading order term in the operator product expansion:

$$\mathcal{O}_1^I \mathcal{O}_2^J \sim \frac{1}{C_{\mathcal{O}_3}} t^{IJK} (x - y) \overline{\mathcal{O}}_{3K}$$
(3.188)

where $C_{\mathcal{O}_3}$ is the coefficient of the $\langle \mathcal{O}_3 \overline{\mathcal{O}}_3 \rangle$, two-point function. In addition, reproducing this form of the three-point function is an important check of the AdS/CFT correspondence as we will see in section 6.2.1.

3.4.4 Advanced Topics in 2d CFT

In this section we specialize once again to 2d CFTs. 2d CFTs have an infinite number of symmetries, and therefore enjoy very severe restrictions under the conformal Ward identities. This means that two dimensional CFT is especially good for explicit results. We will be using examples in 2d CFT in a number of cases in the future chapters on entanglement in field theory, string theory and the AdS/CFT duality.

We found before that the conformal algebra in two dimensions takes the form of two copies of the Weyl algebra, with generators l_m, \bar{l}_m and commutation relations given by

$$[l_n, l_m] = (m - n)l_{m+n}$$

$$[\bar{l}_n, \bar{l}_m] = (m - n)\bar{l}_{m+n}$$

$$[l_m, \bar{l}_n] = 0.$$
(3.189)

Upon quantization, the conformal algebra acquires a *central charge* and turns into the *Virasoro* algebra, with commutation relations given by

$$[L_n, L_m] = (m-n)l_{m+n} + \frac{c}{12}(m^3 - m)\delta_{m+n,0}$$

$$[\overline{L}_n, \overline{L}_m] = (m-n)\overline{l}_{m+n} + \frac{\tilde{c}}{12}(m^3 - m)\delta_{m+n,0}$$

$$[L_m, \overline{L}_n] = 0.$$
(3.190)

We will discover the appearance of the central charge by requiring that the symmetries of the quantum theory are generated by the quantum conserved charges later in this section.

Holomorphic Stress-Energy Tensor

Remember that the conformal generators in 2d were formulated in terms of the complex coordinates z = t + ix and $\overline{z} = t - ix$, and that they are related to forward and backward light-cone coordinates by a Wick rotation. Let us furthermore define holomorphic derivatives

$$\partial_z \equiv \partial = \frac{1}{2}(\partial_t - i\partial_x) , \quad \partial_{\overline{z}} \equiv \overline{\partial} = \frac{1}{2}(\partial_t + i\partial_x)$$
 (3.191)

that fulfill $\partial z = \overline{\partial} \overline{z} = 1$ and $\partial \overline{z} = \overline{\partial} z = 0$. The Euclidean metric becomes

$$g_{zz} = g_{\overline{z}\overline{z}} = 0 , \quad g_{z\overline{z}} = \frac{1}{2} , \qquad (3.192)$$

analogously to how the metric becomes off-diagonal in lightcone coordinates in Minkowski space. Note that lowering a holomorphic index z makes an antiholomorphic index \overline{z} .

The tracelessness of the stress energy tensor $T^{\mu}{}_{\mu} = 0$ becomes $T_{z\overline{z}} = 0$, since the stress-energy tensor is symmetric and the trace is defined by $g^{\alpha\beta}T_{\alpha\beta}$. The conservation equation for the stress-energy tensor

$$\partial_{\mu}T^{\mu\nu} = 0 \tag{3.193}$$

then takes the form

$$\partial T^{zz} = \overline{\partial} T^{\overline{zz}} = 0 \Rightarrow \overline{\partial} T_{zz} = \partial T_{\overline{zz}} = 0 .$$
(3.194)

In other words, T_{zz} can be any holomorphic function, and $T_{\overline{z}\overline{z}}$ can be any antiholomorphic function. It is sometimes practical to write these simply as $T_{zz} \equiv T(z)$ and $T_{\overline{z}\overline{z}} \equiv T(\overline{z})$.

Radial Quantization

A remarkably elegant way of quantizing CFTs exists in two dimensions. In fact, this radial quantization can be formulated in any *d*-dimensional CFT whose manifold can be conformally mapped to $\mathcal{R} \otimes S^{d-1}$ ($x \in [0,2\pi)$ and $t \in [-\infty,\infty]$), but we will specify to the 2d case. Here, the radial coordinate plays the role of time, with the infinite past at the center and the infinite future in the point at infinity.

By performing a conformal transformation, i.e. letting $z \to f(z)$ where f is a holomorphic function we can pick new coordinates, calling them by the same name as before, according to

$$z = e^{2(\tau + ix)} \text{ and } \overline{z} = e^{2(\tau - ix)}$$
(3.195)

so that the infinite past $\tau = -\infty$ is mapped to the origin of the plane, and the point at infinity becomes the infinite future. In addition, we will find it very practical that the coordinate x becomes only a phase. Note that $e^{2(\tau+ix)}$ is a holomorphic function of the previous coordinate $z = \tau + ix$, so we only change the metric by a real prefactor. This can be easily removed by a Weyl transformation.

Remember that the stress energy tensor is the conserved charge that arises from Noether's theorem due to translation invariance. A general coordinate transformation that leaves a CFT invariant is of the type

$$z \to z + \epsilon(z) , \quad \overline{z} \to \overline{\epsilon}(\overline{z}) ,$$
 (3.196)

as we found in section 3.4.1, where ϵ and $\overline{\epsilon}$ are general holomorphic/antiholomorphic functions, respectively.

Going to the quantum theory, we require that the symmetry transformations are generated by their corresponding conserved charges. The classical translation generator is ∂_{μ} and the corresponding conserved charge is

$$P_{\mu} = \int \mathrm{d}x \, T^{0}{}_{\mu} \, . \tag{3.197}$$

To go to the quantum version of our classical symmetry algebra we would like to take $\partial \to P_z$ and $\overline{\partial} \to P_{\overline{z}}$. Thanks to our choice of radial coordinates this becomes very elegant, the momentum operator becomes an integral around a circle of constant radius, where the radius is the Euclidean time, which we indicate in the integrand by saying the integral around the circle \oint is centered on z = 0 according to

$$P_z = \frac{1}{2\pi i} \oint_{z=0} \mathrm{d}z T(z) , \qquad (3.198)$$

and similarly for $P_{\overline{z}}$. This is nonzero because T(z) is a meromorphic function.

When we go to the quantum theory, the conformal l_n generators turn into the Virasoro generators L_n . Let us first consider the case $l_{-1} = \partial$. The corresponding Virasoro generator becomes

$$l_{-1} = -\partial \to L_{-1} = -\frac{1}{2\pi i} \oint dz \, T(z) \;.$$
 (3.199)

For general n, we observe that for any conformal transformation acting as a translation $z \to z + \epsilon(z)$, we get a conserved current $\epsilon(z)T(z)$. The L_n generate the transformations $z \to z + \epsilon_n z^{n+1}$. Setting $\epsilon_z = \epsilon_n z^{n+1}$ and dividing off the parameter ϵ_n we see that for consistency between the expressions the correct expression for the Virasoro generators in terms of the stress energy tensor is

$$L_n = \frac{1}{2\pi i} \oint_{z=0} \mathrm{d}z \, z^{n+1} T \,. \tag{3.200}$$

Note that we have yet to establish any difference between L_m and l_m . We will in the following sections make the difference transparent by considering the general form of the two-point correlator $\langle T(y)T(z)\rangle$. This expectation will then be plugged into the definition of the Virasoro generators to find that there is an extra term in their commutation relation compared to the classical case.

2D Conformal Ward Identity

The conformal Ward identity in general form reads

$$\partial_{\mu} \langle J^{\mu}(y) \mathcal{O}_{1}(x_{1}) \dots \mathcal{O}_{n}(x_{n}) \rangle = \sum_{i} \delta(y - x_{i}) \langle \mathcal{O}_{1}(x_{1}) \dots \delta \mathcal{O}_{i}(x_{i}) \dots \mathcal{O}_{n}(x_{n}) \rangle , \qquad (3.201)$$

however in two dimensions we can rewrite this using Stokes' theorem (in general dimensions, we may rewrite in a similar way by singling out 'radial' and 'time' coordinates). First, we remember that this identity should be under an integral, and then we realize that the only nonzero contributions to the integral comes from when the integration variable y coincides with the x_i in the RHS. Then, we only need to perform our integration over some possibly disconnected region V such that all $x_i \in V$. Stokes theorem then tells us that

$$\int_{V} \mathrm{d}^{2} y \,\partial_{\mu} \langle J^{\mu}(y) \mathcal{O}_{1}(x_{1}) \dots \mathcal{O}_{n}(x_{n}) \rangle = \oint_{\partial V} \mathrm{d} y \, n_{\mu} \langle J^{\mu}(y) \mathcal{O}_{1}(x_{1}) \dots \mathcal{O}_{n}(x_{n}) \rangle \tag{3.202}$$

where n_{μ} is a unit normal to the boundary of the region. Picking the boundary to be a spatial slice of the region, adding the point at infinity since we are in a conformal compactification of Minkowski, this surface is closed, we find that $n = (\tau, 0)$ and this is just an integral over the x-axis at some time t.

In radial coordinates equation (3.202) takes the form

$$\frac{i}{2\pi} \oint_{\partial V} dz \left\langle J_z(z,\overline{z})\mathcal{O}_1(\sigma_1)\dots\mathcal{O}_n(\sigma_n) \right\rangle - \frac{i}{2\pi} \oint_{\partial V} d\overline{z} \left\langle J_z(z,\overline{z})\mathcal{O}_1(\sigma_1)\dots\mathcal{O}_n(\sigma_n) \right\rangle
= \sum_i \left\langle \mathcal{O}_1(\sigma_1)\dots\delta(\sigma_i - (z_i,\overline{z}_i))\mathcal{O}_i(\sigma_i)\dots\mathcal{O}_n(\sigma_n) \right\rangle,$$
(3.203)

where the integral over x has become an integral over a circle since the coordinates are related by $z = e^{2(\tau+ix)}$, $\overline{z} = e^{2(\tau-ix)}$. Note that an integral in increasing x results in the z and \overline{z} contour integrals running in opposite directions, so upon integration they will contribute with the same sign.

We can further divide the integration region into disjoint patches around each σ_i , and get something like equation (3.203) with no sum on the on the RHS and V a small region around σ_i . For currents that are generated by the conformal transformations, J_z is holomorphic and $J_{\overline{z}}$ is antiholomorphic, which means that the contour integral only picks up the residue;

$$\frac{i}{2\pi} \int_{\delta V} \mathrm{d}z J_z(z) \mathcal{O}_1(w) = -\mathrm{Res}_w[J_z \mathcal{O}_1] .$$
(3.204)

As $z \to w$, any pair of CFT operators may be expressed as an operator product expansion. The result of this integral tells us that the OPE between the two operators is of the form

$$J_z(z)\mathcal{O}_1(w,\overline{w}) = \dots + \frac{\operatorname{Res}[J_z\mathcal{O}_1]}{z-w} + \dots , \qquad (3.205)$$

meaning that given the OPE we may compute the residue and given the residue we know a term in the OPE.

Let us now specify $J_z(z)$ to be the current associated with the infinitesimal conformal transformations $\delta z = \epsilon(z)$. The associated conserved charge is $\epsilon(z)T(z)$, and since this charge generates the translation of an operator, we must have that

$$\delta_{\epsilon} \langle \mathcal{O}_1(w, \overline{w}) \rangle = \langle \epsilon(z) T(z) \mathcal{O}_1(w, \overline{w}) \rangle = -\operatorname{Res}(\epsilon(z) T(z) \mathcal{O}_1(w, \overline{w})) .$$
(3.206)

Similarly, for a transformation $\delta \overline{z} = \overline{\epsilon}(\overline{z})$ we find

$$\delta_{\overline{\epsilon}} \mathcal{O}_1(\sigma_1, \overline{\sigma}_1) = -\operatorname{Res}(\epsilon(\overline{z}) T(\overline{z}) \mathcal{O}_1(\sigma_1, \overline{\sigma}_1)) .$$
(3.207)

Equations (3.206)-(3.207) tell us that for any operator \mathcal{O} , if we know the OPE between \mathcal{O} and the components $T(z), \overline{T}(\overline{z})$ of the stress tensor this completely determines the transformation properties of the operator \mathcal{O} . This is because we care only about how operators transform inside correlation functions and the correlations functions are only sensitive to the singular terms of the OPE. Conversely, if we know how an operator transforms under conformal transformations, i.e. $\delta_{\epsilon}\mathcal{O}$, we automatically know part of its OPE with T and \overline{T} .

Let us finish this section with a quick reminder on how to compute residues. The residue of an expression is found as follows

$$\operatorname{Res}\left[\frac{f(z)}{(z-w)^{n}}\right] = \frac{1}{(n-1)!} \frac{\partial^{n-1}}{\partial z^{n-1}} f(z)\Big|_{z=w} .$$
(3.208)

Let us now use these results to explore the operator product expansions between common CFT field operators.

Operator Product Expansions

All operators transform under translations ($\delta z = \epsilon = \text{constant}$) as

$$\mathcal{O}(z-\epsilon) = \mathcal{O}(z) - \epsilon \partial \mathcal{O}(z) \dots ,$$
 (3.209)

telling us that the OPE of any operator \mathcal{O} with T is on the form

$$T(z)\mathcal{O}(w,\overline{w}) = \dots \frac{\partial \mathcal{O}(w,\overline{w})}{z-w} + \dots$$
 (3.210)

In two dimensions, the field ϕ is called a conformal primary field of conformal dimension (h, \tilde{h}) if

$$\phi(z,\overline{z}) \to \left(\frac{\partial w}{\partial z}\right)^h \left(\frac{\partial \overline{w}}{\partial \overline{z}}\right)^{\tilde{h}} \phi(w,\overline{w}) \tag{3.211}$$

under conformal transformations $z \to w(z)$. The conformal dimensions h, \tilde{h} are the 2d analog of the single scaling dimension Δ of the arbitrary dimensional theory.

Infinitesimally, we see that a primary field transforms under $(w, \overline{w}) \to (w, \overline{w}) + \epsilon(w, \overline{w})$ as

$$\delta_{\epsilon}\phi(w,\overline{w}) = h(\partial\epsilon(w))\phi(w,\overline{w}) + \epsilon(w)\partial\phi(w,\overline{w}) . \qquad (3.212)$$

Equation equation (3.208) then tells us that the OPE of $T(z)\phi(w,\overline{w})$ as $z \to w$ must be

$$T(z)\phi(w,\overline{w}) = \frac{h}{(z-w)^2}\phi(w,\overline{w}) + \frac{1}{z-w}\partial\phi(w,\overline{w})$$
(3.213)

so that the residues of the OPE reproduce the transformation $\delta_{\epsilon}\langle \phi(w,\overline{w})\rangle$. It is straightforward to see that an analogous property holds for $\overline{z} \to \overline{w}(\overline{z})$ in terms of \tilde{h} . This completely determines the transformation properties of conformal *primary* operators.

The stress-energy tensor is *not* a primary operator, but it has conformal weight 2. To understand this, realize that the total energy at a given time has conformal dimension 1 since $[E] = L^{-1}$. In turn, the total energy is an integral of the stress tensor over a spatial slice. In addition, the stress tensor has spin 2 since it is a symmetric tensor. In particular, this means it has eigenvalue 2 under all of the conformal transformation generators (roughly, rotations¹¹ + scalings), which is equivalent to the statement that T has conformal weight (2,0) and \overline{T} has weight (0,2). Thus, the operator product expansion T(z)T(w) must take the form

$$T(z)T(w) = \ldots + \frac{2T(w)}{(z-w)^2} + \frac{\partial T(w)}{z-w} \ldots$$
 (3.214)

where the first term is due to h = 2 and the second is universal for all operators. Each term in the expansion has total dimension $\Delta = 4$, and all operators that can appear in the expansion must be on the form

$$\frac{\mathcal{O}_n}{(z-w)^n} , \qquad (3.215)$$

where the conformal dimension of \mathcal{O}_n is 4-n. Unitary CFT's have no operators of any kind with conformal dimension $\Delta < 0$, so the most singular terms that can possibly appear must be of order $(z-w)^{-4}$. In addition, we expect the expansion to be symmetric under $z \leftrightarrow w$ since we take these operator equations to hold inside time-ordered correlation functions, ruling out odd powers of (z-w). The term $\partial T/(z-w)$ manages to be okay since the differential operator also changes sign under $z \leftrightarrow w$. Finally, the only operators of conformal dimension 0 must be proportional to the identity operator. With this line of reasoning, we conclude that

$$T(z)T(w) = \frac{c/2}{(z-w)^4} + \frac{2}{(z-w)^2}T(w) + \frac{1}{z-w}\partial T(w) + \text{finite terms}.$$
 (3.216)

Analogously, for the antiholomorphic stress-energy tensor component we have

$$\overline{T}(\overline{z})\overline{T}(\overline{w}) = \frac{\tilde{c}/2}{(\overline{z}-\overline{w})^4} + \frac{2}{(\overline{z}-\overline{w})^2}\overline{T}(\overline{w}) + \frac{1}{\overline{z}-\overline{w}}\partial\overline{T}(\overline{w}) + \text{finite terms}.$$
(3.217)

The constants c and \tilde{c} are the *central charges* of the CFT. These are arbitrary, but in some cases they need to take a certain value to preserve symmetries of the theory at the quantum level, as we will see in string theory.

Another important thing to note; the stress-energy tensor tells us how something transforms under a change of coordinates, which means equation (3.217) tells us how T transforms. It can be shown by inspecting equation (3.217) for an extended time that under $z \to \tilde{z}$ T transforms as

$$\tilde{T}(\tilde{z}) = \left(\frac{\partial \tilde{z}}{\partial z}\right)^{-2} \left[T(z) - \frac{c}{12}S(\tilde{z}, z)\right]$$
(3.218)

where $S(\tilde{z}, z)$ is the Schwarzian, defined by

$$S(\tilde{z},z) = \left(\frac{\partial^3 \tilde{z}}{\partial z^3}\right) \left(\frac{\partial \tilde{z}}{\partial z}\right)^{-1} - \frac{3}{2} \left(\frac{\partial^2 \tilde{z}}{\partial z^2}\right)^2 \left(\frac{\partial \tilde{z}}{\partial z}\right)^{-2}.$$
(3.219)

We will find in string theory, where we are working with an explicit stress tensor in terms of fields that the central charges corresponding to a single scalar field is $c = \tilde{c} = 1$, and if we have D independent scalar fields we obtain the central charge $c = \tilde{c} = D$. This hints at a deep

¹¹The spin in a 2d CFT is given by $h - \tilde{h}$, relating to the holomorphic and antiholomorphic modes being 'left' and 'right' moving sectors on a closed space. Embedding the compact space in a higher dimensional space, this appears as left- and right-handed spinning modes.

connection; the central charge c somehow measures the number of degrees of freedom in the CFT. This connection is made explicit by the *Cardy formula*

$$S = 2\pi \sqrt{\frac{c}{6} \left(L_0 - \frac{c}{24}\right)} \tag{3.220}$$

for the entropy of a CFT living on a circle of radius $R = L_0/E$ where E is the total energy of the state. In section 4.2 we also prove the c theorem which states that c decreases monotonically under renormalization group flow, further supporting the intuition that it somehow measures degrees of freedom.

Virasoro Algebra

We can now derive the quantum version of the conformal algebra by requiring that it is generated by the quantum operators corresponding to the classical symmetries. We will find that it is not exactly the same as its classical counterpart, the central charge will make an appearance as a *central extension* of the conformal algebra. Let us consider the commutator

$$[L_m, L_n] = \left[\oint \frac{\mathrm{d}z}{2\pi i} \, z^{m+1} T(z), \oint \frac{\mathrm{d}w}{2\pi i} \, w^{n+1} T(z)\right] \,. \tag{3.221}$$

As with all operator equations, we expect this to live inside a time-ordered correlation function. In particular, this means that terms further to the right are taken to occur at earlier time, which in our radial coordinates corresponds to a smaller radius. This means that the term L_n, L_m tells us to take the z-contour to be outside w while the term $-L_m L_n$ tells us take to the z-contour to be inside the w-contour and integrate in the opposite direction.

To make sense of this computation is we fix w and perform the z-integration. Since we have a fixed w we can connect the two z- contours by going back and forth along a line segment in the radial direction.

$$[L_m, L_n] = \oint \frac{dw}{2\pi i} \oint_{\gamma_w} \frac{dz}{2\pi i} z^{m+1} w^{n+1} T(z) T(w) = \oint \frac{dw}{2\pi i} \operatorname{Res} \left[z^{m+1} w^{n+1} \left(\frac{c/2}{(z-w)^4} + \frac{2}{(z-w)^2} T(w) + \frac{1}{z-w} \partial T(w) + \operatorname{finite} \right) \right] ,$$
(3.222)

where we have used that the curve γ_w is given by



where the equality between contours holds because the integrand is holomorphic except in the point z = w. To pick out the residue we need the 0th, 1st and 3rd order terms in the Taylor expansion

of z^{m+1} about z = w;

$$z^{m+1} = w^{m+1} + (m+1)w^m(z-w) + \frac{1}{2}m(m+1)w^{m-1}(z-w)^2 + \frac{1}{6}m(m^2-1)w^{m-2}(z-w)^3 + \dots$$
(3.224)

We can then read off the residues, finding that

$$[L_m, L_n] = \oint \frac{\mathrm{d}w}{2\pi i} w^{n+1} \left[w^{m+1} \partial T(w) + 2(m+1) w^m T(w) + \frac{c}{12} (m^3 - m) w^{m-2} \right]$$

= $\oint \frac{\mathrm{d}w}{2\pi i} w^{n+1} \left[-(n+m+2) w^m T(w) + 2(m+1) w^m T(w) + \frac{c}{12} (m^3 - m) w^{m-2} \right]$
= $\oint \frac{\mathrm{d}w}{2\pi i} \left[(m-n) w^{m+n+1} T(w) + \frac{c}{12} (m^3 - m) w^{m+n-1} \right]$
= $(m-n) L_{m+n} + \frac{c}{12} (m^3 - m) \delta_{m+n,0} .$ (3.225)

The last step used the definition of the Virasoro generator in terms of T for the first term, and that the integral around a circle of w^k is only nonzero for k = -1, you could see this as the integral picking out the residue at the origin of w^{m+n-1} .

This is the famous Virasoro algebra, and we see the appearance of a "central term" due to the requirement that the symmetry be generated by te quantum conserved charges. The reason the terms proportional to c is called central is because it commutes with all other terms in the algebra.

Weyl Anomaly

On a flat background spacetime the central charge may be chosen completely freely, but this is not so for a curved spacetime. Remember that scale invariance is encoded in the tracelessness of the stress tensor. On a curved background, we will find that

$$\langle T^{\alpha}{}_{\alpha} \rangle = -\frac{c + \tilde{c}}{12}R \tag{3.226}$$

where R is the Ricci scalar. This anomaly goes under the names Weyl anomaly, trace anomaly and conformal anomaly.

In two dimensions any metric can be made proportional to the flat metric such that $g_{\alpha\beta} = e^{2\omega}\delta_{\alpha\beta}$. In these coordinates we may write the Ricci scalar as

$$R = -2e^{-2\omega}\partial^2\omega , \qquad (3.227)$$

Equation equation (3.226) then tells us that for a CFT with central charge has at least one gauge invariant observable in $\langle T^{\alpha}{}_{\alpha} \rangle$.

To prove this, we need some intermediate results, and we want to work in the radial, holomorphic coordinates. We saw that the vanishing of the trace took the form $T_{z\bar{z}}$ at the beginning of this section, so this is where we will now find our violation. Let us consider the stress-energy conservation equation

$$\partial T_{z\overline{z}} = -\overline{\partial}T_{zz} \ . \tag{3.228}$$

Using this result and the fact that T_{zz} is a function of \overline{z} , the OPE of the off-diagonal stress tensor components may be simply expressed:

$$\partial_z T_{z\overline{z}} \partial_w T_{w\overline{w}} = \partial_{\overline{z}} T_{zz} \partial_{\overline{w}} T_{ww} = \partial_{\overline{z}} \partial_{\overline{w}} \left[\frac{c/2}{(z-w)^4} \dots \right] .$$
(3.229)

Here the RHS looks independent of z, but since it is not holomorphic at the singular points it ceases to be \overline{z} -independent. We may then use the identities

$$\int d^2 \sigma \partial_{\overline{z}} \frac{1}{z-w} = -\oint d\overline{z} \frac{1}{z-w} = 2\pi , \qquad (3.230)$$

as well as

$$\int d^2 \sigma \delta(z - w, \overline{z} - \overline{w}) = 1 , \qquad (3.231)$$

to claim $\partial_{\overline{z}} \frac{1}{z-w} = 2\pi \delta(z-w,\overline{z}-\overline{w})$ under the integral sign. This lets us rewrite

$$\partial_{\overline{z}}\partial_{\overline{w}}\frac{1}{(z-w)^4} = \frac{1}{6}\left(\partial_z^2\partial_w\frac{1}{z-w}\right) = \frac{\pi}{3}\partial_z^2\partial_w\partial_{\overline{w}}\delta(z-w,\overline{z}-\overline{w}) . \tag{3.232}$$

Inserting into the correlation function in equation (3.229) and dropping $\partial_z \partial_w$ on both sides we find

$$T_{z\overline{z}}T_{w\overline{w}} = \frac{c\pi}{6}\partial_z\partial_{\overline{w}}\delta(z-w,\overline{z}-\overline{w}) . \qquad (3.233)$$

Armed with the OPE between the off-diagonal stress tensor components, corresponding to the trace components in Cartesian coordinates, we are ready to find the Weyl anomaly. Let us assume that $\langle T^{\alpha}{}_{\alpha} \rangle = 0$ on a flat background. We will then look for an expression for $\langle T^{\alpha}{}_{\alpha} \rangle$ close to flat space. We can then use the definition of the stress tensor in terms of the variation of the metric, combined with the definition of the expectation of the stress-energy tensor in terms of path integrals.

$$\delta_{g} \langle T^{\alpha}{}_{\alpha}(\sigma) \rangle = \delta \int \mathcal{D}\phi e^{-S} T^{\alpha}{}_{\alpha}(\sigma) = \frac{1}{4\pi} \int \mathcal{D}\phi e^{-S} \left[T^{\alpha}{}_{\alpha}(\sigma) \int d^{2}\sigma' \sqrt{g} \delta g^{\beta\gamma} T_{\beta\gamma}(\sigma') \right] .$$
(3.234)

We can now restrict to a Weyl transformation, where $\delta g_{\alpha\beta} = 2\omega\delta_{\alpha\beta}$ and $\delta g^{\alpha\beta} = -2\omega\delta^{\alpha\beta}$. Then

$$\delta_g \left\langle T^{\alpha}{}_{\alpha}(\sigma) \right\rangle = -\frac{1}{2\pi} \int \mathcal{D}\phi e^{-S} \left[T^{\alpha}{}_{\alpha}(\sigma) \int \mathrm{d}^2 \sigma' \sqrt{g} \omega T^{\gamma}{}_{\gamma}(\sigma') \right] \,. \tag{3.235}$$

The stress tensors are operators, and we see that the integral over σ' is determined by the OPE between the trace terms. We now simply need to switch to holomorphic coordinates, keeping track of factors of 2 to find

$$T^{\alpha}{}_{\alpha}(\sigma)T^{\gamma}{}_{\gamma}(\sigma') = 16T_{z\overline{z}}T_{w\overline{w}} = \frac{8c\pi}{3}\partial_z\partial_{\overline{w}}\delta(z-w,\overline{z}-\overline{w}) .$$
(3.236)

The δ lets us replace $\partial_{\overline{w}} \to -\partial_{\overline{z}}$. We can then recast in Cartesian coordinates by using

$$-8\partial_z\partial_{\overline{z}} = \partial^2\delta(\sigma - \sigma') . \qquad (3.237)$$

Plugging into equation (3.235) and partial integrating to put derivatives on ω we find

$$\delta \langle T^{\alpha}{}_{\alpha} \rangle = \frac{c}{6} \partial^2 \omega(\sigma) . \qquad (3.238)$$

But the variation of $\delta \langle T^{\alpha}{}_{\alpha} \rangle$ about flat space is the entire trace, so by matching to our earlier expression for R to first order in ω we find

$$\langle T^{\alpha}{}_{\alpha} \rangle = -\frac{c}{12}R , \qquad (3.239)$$

completing the derivation. An analogous derivation gives the \tilde{c} part of the anomaly.

In general dimension, the Weyl anomaly is not directly proportional to the central charge. General dimensional CFTs are then characterized instead by two numbers, the Weyl anomaly A and the value of c as obtained from the two-point function of the stress energy tensor.

State-Operator Correspondence

In this section we discuss the property of CFTs that lets us do string perturbation theory, namely the *state-operator* correspondence. The statement of the correspondence is that there is a *one-toone correspondence* between *local* operators and states.

To understand how this correspondence comes about, let us remember that formally, we can define a time dependent state via a path integral with an initial condition ϕ_f and an open final boundary condition (see equation (3.61))

$$|\phi(t)\rangle = \int_{\phi_i} \mathcal{D}\phi e^{-i\int_{t_0}^t L[\phi]} .$$
(3.240)

The wavefunction(al) $|\phi(t)\rangle$ is completely determined by the initial condition ϕ_i . In the radial quantization scheme, time becomes radius, and the initial condition has to be specified on the entire surface of constant radius. Then, the ket takes the form

$$|\phi(r(t))\rangle = \int_{\phi_i} \mathcal{D}\phi e^{-\int_{r_0(t_0)}^r \mathrm{d}r L[\phi]}$$
 (3.241)

We can now take the initial condition to the infinite past, collapsing the inner circle of radius r_0 to a point. Then, the wavefunction is completely specified by weighting the path integral at the single point z = 0. This is exactly what we mean by a local operator insertion, meaning that every inequivalent local operator at z = 0 defines a unique state on the entire spacetime. Therefore, a general time (radius) dependent state in the CFT is described by

$$|\phi(r)\rangle = \int \mathcal{D}\phi e^{-\int_0^r \mathrm{d}r L[\phi]} \mathcal{O}(z=0) \ . \tag{3.242}$$

Particularly notable is the fact that $\mathcal{O} = \mathbb{1}$ gives the CFT vacuum state. This is because taking $\tau \to -\infty$ in Euclidean coordinates is precisely the trick we use in ordinary QFT to project out all but the vacuum state.

Similarly we can define the bra by defining an initial condition at $r = \infty$, which is a point. The correlator between two states is then

$$\langle \phi_2 | \phi_1 \rangle = \int \mathcal{D}\phi \mathcal{O}_1(0) \mathcal{O}_2(\infty) e^{-\int_0^\infty \mathrm{d}r L[\phi]} , \qquad (3.243)$$

where a completion relation in the middle glues the integration regions together.

3.4. Conformal Field Theory

Chapter 4

Entanglement Entropy of Quantum Fields

The entanglement entropy of a region of space in QFT is a quantity of great interest in modern fundamental physics. This interest in some sense came to be with the discovery of black hole thermodynamics by Bekenstein and Hawking [1, 28]. Since then, the entropy of quantum field theories has found relevance for example in the context of AdS/CFT correspondence [4, 3] and area-law systems in condensed matter [29].

Entanglement entropy in QFT is fundamentally much more subtle than the version of entanglement entropy that was encountered in section 2.3. In field theory, the entire universe is generally described by a single wavefunction. The wavefunction is not a local function of the spacetime coordinates, but rather some integral over the spacetime manifold (or formally constructed as a path integral over the manifold, depending on formalism). It then becomes very difficult to assign wavefunctions to only compact regions of spacetime. More in line with the descriptions of section 2.3, it is not possible to write the Hilbert space as $\mathcal{H}_A \otimes \mathcal{H}_B$ and assign the two Hilbert space factors to two spatially separated labs in a self-consistent way.

This means that it is not obvious how to define the partial trace operation so that we can pass from the global wavefunction to the local density operator. It is not useful to trace over basis states of the Hilbert space, since the Hilbert space does not decompose into factors that may be assigned to local regions. To try and make sense of this, there are two ways to proceed

- Ignore the previous problem, and perform the partial trace in a local region by gluing the path integral representation of $\text{Tr}_A[|\Omega\rangle\langle\Omega|]$ only across the region A, without admitting that there is no such thing as a complete basis of spatially local states to make this operation defined.
- Take the high road and use axiomatic field theory to find a self-consistent formulation of local properties of quantum fields, then hope that entanglement entropy is well defined in terms of these properties.

This chapter will develop both approaches, using the latter to show that despite technically being ill defined the first approach can produce correct results. In section 4.1 the formalism and important results of axiomatic QFT are introduced. In section 4.2 the Unruh effect is derived using both path integral and axiomatic methods and then applied to discover Hawking radiation. In section 4.3 some of the machinery developed is used to understand the entanglement properties of CFTs. These results tie into chapters 6 and 7 where the focus is on the AdS/CFT correspondence.

4.1 Entanglement Entropy in Axiomatic QFT

Axiomatic quantum field theory (AQFT) is the attempt to mathematically rigorously formulate QFT. It fell out of favor for a long time because there was no consistent way to reproduce the interaction picture used in scattering theory. Recently it has seen a resurgence due to its usefulness for computing the von Neumann entropy of quantum fields. The initial formulation of AQFT that is covered here is due to Streater and Wightman [30], while a later formulation in terms of so-called C^* algebras of local observables is due to Haag and Kastler [31]. The latter will be discussed in section 4.1.3.

Axiomatic QFT takes seriously the notion that the right hand side of the canonical commutation relations,

$$[\phi(0,x), \dot{\phi}(0,y)] = \delta(x-y) , \qquad (4.1)$$

is a distribution, not a function. The commutation relation implies that we should find an operatorvalued field $\phi(x)$ such that its commutator with $\dot{\phi}$ is a Dirac delta, but this does not make sense unless ϕ itself is an operator-valued distribution. If $\phi(x)$ is a distribution, terms such as $\lambda \phi^3(x)$ that appear in the equations of motion of ϕ^4 -theory are either infinite or at the very least undefined. It was concluded that a more indirect definition of the dynamics of QFT was needed for this to make mathematical sense.

For the commutation relation to make sense, it turns out that it is sufficient¹ to consider that ϕ is a *tempered distribution*. For the field operators to be non singular when acting on Hilbert space, it is then enough to smear with an infinitely differentiable test function f that goes to zero faster than any power of euclidean distance at infinity according to

$$\phi[f] = \int \mathrm{d}x \,\phi(x) f(x) \,. \tag{4.2}$$

The smeared operator $\phi[f]$ is finite as long as it is considered as acting on a class of states called the *vacuum sector* of Hilbert space, the meaning of which will be made more precise shortly. The type of smearing function that enters into this definition is said to live in the Schwarz space S of functions.

The product $\phi_1 \phi_2$ of two tempered distributions is not a tempered distribution, meaning that the smeared operator

$$(\phi_1 \phi_2)[f_1] = \int \mathrm{d}x \,\phi_1(x) \phi_2(x) f_1(x) \tag{4.3}$$

might diverge in a complicated manner despite the smearing. The basic example of this is the Dirac delta, $\int dx \delta(x) \delta(x) f(x) = \infty \cdot f(0)$. Luckily, Schwarz's kernel theorem tells us that the product of two tempered distributions can be arbitrarily closely approximated in the following way

$$(\phi_1\phi_2)[f_n] = \sum_i c_i\phi_1[f_{i,1}]\phi_2[f_{i,2}] , \qquad (4.4)$$

for each $f_n \in S$, where the index $i \in \mathbb{N}$. The Schwarz kernel theorem implies the validity of the *Operator Product Expansion* that we mentioned in sections 3.1.3 and 3.4.4. With this context in mind, let us now state the axioms of algebraic QFT that define the necessary structure on Hilbert space for the emergence of a relativistic QFT. These axioms are used repeatedly in the following sections.

¹In this case, 'sufficient' means that if we restrict to the simple class of tempered distributions, the axiomatic theory can still reproduce nontrivial dynamics for relativistic quantum field theories.

Axiom 4.1.1 (Definition of State) The states of a quantum theory are normalized vectors in a separable Hilbert space \mathcal{H} , and two vectors that differ by a phase define the same state.

This is just the usual definition of a quantum state, in the same form it has been since introductory quantum mechanics.

Axiom 4.1.2 (Poincaré Representation) The Hilbert space \mathcal{H} of a QFT contains a unitary representation, $(a,\Lambda) \to U(a,\Lambda)$ of the restricted, orthocronous Poincaré group $\mathcal{P}_{+}^{\uparrow}$. In \mathcal{H} there exists a unique vector Ω , up to a phase, that is invariant under all $U(a,\Lambda)$, and for all other vectors $\Psi \in \mathcal{H}$ the spectrum of P^{0} , the self-adjoint generator of the one-parameter time-translation group, is positive.

The restricted orthochronous Poincaré group is the Lie group of spacetime translations plus Lorentz transforms: $x \to \Lambda x + a$ such that $\det(\Lambda) = 1$ and $\Lambda^{00} > 0$. P^0 generates time translations, and as the 0 component of the simultaneous four-momentum, it defines the Hamiltonian. The statement that the spectrum of P^0 is positive for all Ψ and 0 for the vacuum is then the statement that the ground state has energy zero and all excited states have positive energy. The condition of positive energy combined with unitary operators $U(0,\Lambda)$ implementing the Lorentz boosts leads to the spectral condition; the simultaneous spectrum of P^{μ} must lie in the closed forward lightcone (since the restricted orthochronous Lorentz group can only take the timelike vector (E,0,0,0) to another timelike vector). The simultaneous momentum P^{μ} is the integral over a spacelike Cauchy slice of the momentum density.

In equation (4.2) we defined the smeared field $\phi[f] = \int dx \, \phi(x) f(x)$. This operator is unbounded in the general case, and we should see it as living in a correlation function $\langle \psi_2 | \phi | \psi_1 \rangle$. To make sense of an unbounded operator we must define a domain D of vectors such that if $\psi_1, \psi_2 \in D$, then $\langle \psi_2 | \phi | \psi_1 \rangle$ is finite.

A state that we know is in the domain of the field operator $\phi[f]$ is the Fock vacuum $|\Omega\rangle$. In addition, we can multiply the state $\phi[f] |\Omega\rangle$ by another field operator $\phi[g]$ to obtain a new state in the Fock-space of QFT. In general, the Fock space of states that can be created from the vacuum by field operators is in the domain D of $\phi[f]$. Since composite operators, such as $\phi_1\phi_2$ can be expanded via the Schwarz kernel theorem, $(\phi_1\phi_2)[f] |\Omega\rangle$ is also in D as long as we define the domain D to be closed under limit sequences. Being closed under limit sequences means exactly that if we have a sequence $|s_n\rangle \in D$ such that $\lim_n |s_n\rangle$ converges to $|\Psi\rangle$, we consider $|\Psi\rangle$ to lie in D. Setting $s_n = \sum_{i=1}^n c_i \phi_1[f_{i,1}] \phi_2[f_{i,2}]$ includes any state described by the Schwarz kernel theorem in D.

Axiom 4.1.3 (Vacuum Sector) The vacuum of the QFT, Ω , lies in the domain of any $\phi[f]$, and for any $f_1 \dots f_n \in S$ we have $\phi[f_1] \dots \phi[f_n] \Omega \in D$. We take D to be defined by the span of such vectors as we vary n and the test functions. By definition, D is the vacuum sector Hilbert space, meaning all states that can be obtained by acting on the vacuum with field operators. In the later sections we will refer to D as \mathcal{H}_0 , since it is the subspace of the full Hilbert space (\mathcal{H}) that is connected to the vacuum ($|0\rangle$ or $|\Omega\rangle$) by field operators.

Moreover, the smeared fields transform under $U(a,\Lambda)$ according to

$$U(a,\Lambda)\phi(f)U^{-1}(a,\Lambda) = \phi(f_{a,\Lambda})$$
(4.5)

where the test function transforms as the inverse element of the group:

$$f_{a,\Lambda} = f\left(\Lambda^{-1}(x-a)\right)$$

This is relying on the known analysis of free field theory in terms of creating many particle states by acting on a Fock vacuum with creation/annihilation operators. In addition, we have defined the necessary transformation property of the smearing functions for the smeared field to transform sensibly under the Poincaré group.

Axiom 4.1.4 (Hermiticity) If ϕ a is field operator defined on D, then its Hermitian conjugate ϕ^{\dagger} is also a field operator defined on D, defined by

$$\langle \psi_2 | \phi^{\dagger} | \psi_1 \rangle = \overline{\langle \psi_1 | \phi | \psi_2 \rangle} . \tag{4.6}$$

Relating to the usual form of QFT, we expect creation and annihilation operators to have the same domain, we expect them both to exist, and we expect them to be related by a Hermitian conjugate.

Axiom 4.1.5 (Tempered Distribution) For any pair of vectors $\psi_1, \psi_2 \in D$, the map $S \to \mathbb{C}$ given by $f \to \langle \psi_2 | \phi[f] | \psi_1 \rangle$ is continuous.

This is the definition of ϕ being a tempered distribution as long as it acts as an operator on states in D, and ensures that the smeared fields are finite operators on D.

Axiom 4.1.6 (Causality) Let the support of f, supp(f(x)), be defined as the set of all points for which $f(x) \neq 0$. Suppose that we have two test functions $f, g \in S$ such that all points in supp(f(x)) are spacelike separated from all points in supp(g(x)); then $\phi[f]\phi[g] = \phi[g]\phi[f]$ for bosons and $\phi[f]\phi[g] = -\phi[g]\phi[f]$ for fermions.

This is the Lorentz invariant statement of the equal time canonical commutation relations in ordinary QFT.

4.1.1 The Reeh-Schlieder Theorem

The Reeh-Schlieder theorem is a statement that any field content of a QFT in all of spacetime can be formulated in terms of boundary conditions supported only on a small, compact subspace. This is in stark contrast with the initial expectation that to define the QFT at arbitrary time t in all of space, we need to specify boundary conditions on a noncompact spacelike Cauchy surface such that all of spacetime is in the causal past or causal future of the surface.

To make the statement more precise and to prove the theorem, we need to make a few definitions. We assume that we are in Minkowski spacetime of D dimensions with mostly positive signature and one time dimension.

The initial state of a QFT is given by some state $|\Psi\rangle \in \mathcal{H}_0$ where \mathcal{H}_0 is the vacuum sector of the full Hilbert space \mathcal{H} . The vacuum sector of the Hilbert space is the set of states that can be created by acting on the vacuum with local field operators. One might think that \mathcal{H}_0 should be the same as \mathcal{H} , thinking that creation operators acting on the vacuum should span the entire Hilbert space. As an example, there is no construction of a state that is in a superposition of states with different charges (say color charge and electric charge), which is of course very important for consistency, since such a state would lead to possible projective measurements in which charges are projected out of the existence, breaking charge conservation. Principles such as charge- and baryon number conservation that select a part of Hilbert space that is "physical" are called *superselection principles*.

Following [32], we assume for simplicity that the set of local operators in our QFT is given by a Hermitian scalar field $\phi(x^{\mu})$, the generalization is given by additional operators for the additional

fields, including their commutation relations. We then introduce smeared fields

$$\phi[f] = \int \mathrm{d}^D x \, f(x)\phi(x) \,, \qquad (4.7)$$

where $f \in S$. Any state $|\Psi\rangle$ in the vacuum sector can be approximated arbitrarily closely by a linear combination of states on the form

$$|\Psi_{\vec{f}}\rangle = \phi_{f_1}\phi_{f_2}\dots\phi_{f_n}|\Omega\rangle \tag{4.8}$$

where $\phi_{f_i} \equiv \phi[f_i]$, and we have denoted the vacuum state Ω . That is, we can write

$$|\Psi\rangle \left(\phi_{f_1}, \phi_{f_2}, \dots, \phi_{f_n}\right) = \sum_k c_k \left|\Psi_{\vec{f_k}}\right\rangle \left(\phi_{f_1}\phi_{f_2}\dots\phi_{f_n}\right)$$
(4.9)

where $|\Psi\rangle$ is a multilinear function of the $\phi[f]$, and the RHS is a so called *kernel* of the product of the $\phi's$. The validity of equation (4.9) follows from Schwarz's kernel theorem as stated in equation (4.4). The smearing construction is due to the fact that although the individual fields $\phi(x^{\mu})$ define well-behaved distributions, their compositions might not, by introducing smearing we make sure that we generate finite-norm states in Hilbert space. We can then approximate infinite norm states as limit sequences of states on the form of equation (4.9).

A Cauchy hypersurface Σ is a complete spacelike surface on which you can define sufficient boundary conditions for a QFT, the perhaps simplest example is the hypersurface defined by the constraint t = 0. Classically, the equations of motion then give the field configuration for all $t \neq 0$. Usually, sufficient boundary conditions for a theory are the values of $\phi(x^{\mu})$ as well as the time derivative $\dot{\phi}(x^{\mu}) \forall x^{\mu} \in \Sigma$. To express sufficient boundary conditions in terms of the smearing functions f_i , we require the existence of a time derivative, meaning that the f_i need to have have support in a small neighborhood \mathcal{U} of Σ . In the example of the surface t = 0 we can express this as $supp(f_i) \in |t| < \epsilon$ where ϵ is an arbitrarily small parameter and $supp(f_i)$ denotes the support of the function f_i .

Let us now formulate the Reeh-Sclieder theorem in terms of the machinery we just introduced

Theorem 4.1.1 (Reeh-Schlieder theorem) Let Σ be a Cauchy hypersurface, let \mathcal{H}_0 be the vacuum sector of Hilbert space, let $\mathcal{V} \in \Sigma$ be a small open subset of Σ and $\mathcal{U}_{\mathcal{V}}$ a corresponding small neighbourhood of \mathcal{V} . Then, for $f_i : supp(f_i) \in \mathcal{U}_{\mathcal{V}}$ the states

$$|\Psi\rangle [\phi_{f_1}, \phi_{f_2}, \dots, \phi_{f_n}] = \sum_k c_k |\Psi_{\vec{f_k}}\rangle [\phi_{f_1}\phi_{f_2}\dots\phi_{f_n}] , \qquad (4.10)$$

are dense in \mathcal{H}_0 , meaning that any state in \mathcal{H}_0 may be arbitrarily closely approximated using only smeared operators with local support.

The implications of this theorem are as follows. The boundary condition for a QFT is a state $|\Psi\rangle$, which can be expressed in terms of the smeared fields $\phi[f]$ acting on the vacuum $|\Omega\rangle$. The Reeh-Schlieder theorem then tells us that we may pick f with compact support, meaning $\phi[f]$ has only compact support. This means that to specify sufficient initial conditions for the full theory we need only specify the $\phi[f_i]$ on a finite subset of a Cauchy slice, and the vacuum state $|\Omega\rangle$ will take care of extrapolating this initial condition to all of space. This is a first hint at the interconnectedness (entanglement) of the QFT vacuum. It should be noted that the Reeh-Schlieder theorem says nothing about causality, since the operators considered are not necessarily unitary, meaning they need not correspond to the time evolution generated by the Hamiltonian (or a man-made perturbation thereof), instead it is a statement about the difficulty of defining localized fields in QFT.

Proof of the Reeh-Schlieder Theorem

Suppose theorem 4.1.1 is false, then there exists a state $|\chi\rangle \in \mathcal{H}_0$ such that

$$\langle \chi | \psi_{\vec{f}} \rangle = 0 , \ \forall \ | \psi_{\vec{f}} \rangle , \ \forall f : supp(f) \in \mathcal{U}_{\mathcal{V}} .$$
 (4.11)

The intuition behind the proof of the Reeh-Schlieder theorem is that we have a complex valued analytic function, the inner product, that is zero on a finite region in the real (hyper)plane. The multi-dimensional edge of the wedge theorem of complex calculus then tells us that this function must be zero on the entire real plane. We will use a slightly less mathematically sophisticated and more physical proof to show this explicitly. This proof uses axioms 4.1.2, 4.1.3, 4.1.5 that we defined in the previous section.

In essence, what we will do in this proof is use Hamiltonian time evolution to move points in the arguments of $|\psi\rangle_{\vec{f}}$ outside of $\mathcal{U}_{\mathcal{V}}$ to show that if $\langle \chi | \psi \rangle_{\vec{f}} = 0$ inside of $\mathcal{U}_{\mathcal{V}}$ they are zero on the lightcone of $\mathcal{U}_{\mathcal{V}}$. Then, realizing that we are free to move in any timelike direction, we can zigzag backwards and forwards in time to move an argument in a spacelike direction (along the Cauchy surface) and we will see that the inner product must be zero in all of spacetime.

We begin by clarifying how to translate the functional representation of the field operators. Let x_i denote a point in spacetime such that $x_i \in \mathcal{V}$, and let \mathcal{U}_{x_i} denote a small neighbourhood of x_i such that this neighbourhood is contained in $\mathcal{U}_{\mathcal{V}}$. We then denote the field operators as

$$\phi(x_i) = \int \mathrm{d}^D x f_i(x) \phi(x) \ , \quad supp(f_i) \in \mathcal{U}_{x_i} \ . \tag{4.12}$$

This way of writing amounts to writing the ϕ s as a functional of the smearing functions f_i , parametrized by a 'center of support coordinate' x_i . In this proof we will treat the inner product as a complex valued function of the x_i . Then, we can write the translated field operators according to

$$\phi(x_i + \delta) = \int \mathrm{d}^D x f_i(x - \delta) \phi(x) , \quad supp(f_i) \in \mathcal{U}_{x_i - \delta} , \qquad (4.13)$$

noting that if we have translated by a sufficiently large timelike vector δ , $\mathcal{U}_{x_i-\delta}$ is no longer contained inside $\mathcal{U}_{\mathcal{V}}$.

Let us first show that

$$\langle \chi | \psi_{\vec{f}} \rangle = \langle \chi | \phi(x_1) \dots \phi(x_{n-1}) \phi(x_n) | \Omega \rangle$$
(4.14)

still vanishes if x_n is moved outside of $\mathcal{U}_{\mathcal{V}}$. We denote the timelike vector (1,0,0,0) by t, and translate x_n by ut for some real u. This can be written in terms of the Hamiltonian H as

$$\langle \chi | \phi(x_1) \dots \phi(x_{n-1}) \phi(x_n + ut) | \Omega \rangle = \langle \chi | \phi(x_1) \dots \phi(x_{n-1}) e^{iHu} \phi(x_n) e^{-iHu} | \Omega \rangle$$

= $\langle \chi | \phi(x_1) \dots \phi(x_{n-1}) e^{iHu} \phi(x_n) | \Omega \rangle \equiv g(u) ,$ (4.15)

where we have used that the Hamiltonian annihilates the vacuum in Minkowski space and defined the RHS as a function g(u). The definition of the Hamiltonian requires the selection of a preferred timelike vector t that we use to parametrize time evolution, the point being that the choice of t is actually arbitrary, and this reasoning can be extended to an arbitrary timelike vector.

It is clear that g(u) is 0 by definition for $u \in [-\epsilon_{\min}, \epsilon_{\max}]$ (rigorously, as long as $\mathcal{U}_{x_n+ut} \in \mathcal{U}_{\mathcal{V}}$), for some ϵ '. We analytically extend g(u) to complex values of u. The Hamiltonian as an operator is bounded below by eigenvalue zero, telling us that g(u) is holomorphic in the upper half plane, and continuous as one approaches the real axis. We will now use the holomorphicity of g(u) as well as its value of zero on a finite segment of the real axis to show that it must be zero on the entire real axis.



Figure 4.1: Using that g(u) is holomorphic in the upper half plane and zero on a finite segment on the real axis, we can show that it is holomorphic below the real axis as well. This is done by letting the Cauchy curve integral representation run along the segment where g(u) is zero, and then excluding the segment from the curve. The Cauchy integral formula remains holomorphic as we move the point u through the segment and into the lower half plane. We conclude that g(u) as represented by the Cauchy integral formula must be holomorphic on the segment $\operatorname{Re}(z) \in [-\epsilon_{\min}, \epsilon_{\max}]$.

If we knew that g(u) was holomorphic on the real axis, then it would admit a convergent Taylor series around u = 0 which would have to be identically zero since g(u) is zero in a finite segment of the real axis. Now we are not armed with holomorphicity on the real axis, only continuity as we approach it, so we instead observe that by the residue theorem g(u) admits a Cauchy integral representation in the upper half plane

$$g(u) = \frac{1}{2\pi i} \oint_{\gamma} \mathrm{d}u' \frac{g(u)}{u' - u} , \qquad (4.16)$$

where γ is a closed curve parametrized by u' that goes counter-clockwise around u. For a fixed γ , this representation is manifestly holomorphic as long as u remains in the interior of γ . On the boundary holomorphicity fails since the integrand will hit the singularity as $u' \to u$. The next step is to include $[-\epsilon_{\min}, \epsilon_{\max}]$ (and the lower half-plane). For the intuition of the following argument, see figure 4.1.

Since we know that g(u) is continuous as we approach the real axis, we can move the curve as close to the real axis as we want, consequently the curve can be put on the real axis in this segment. Since g(u) is zero on the segment we can drop this segment from the integration curve, and the Cauchy integral remains holomorphic even as the argument u is moved onto the real axis. From this, it follows that g(u) is holomorphic on a finite segment of the real axis because it is represented by a holomorphic Cauchy integral. Since g(u) therefore admits a convergent Taylor expansion on the finite segment it must be zero on the entire real axis. Consequently, we can move x_n out of $\mathcal{U}_{\mathcal{V}}$ without changing g(u). Picking a different timelike vector and translating x_n backwards in time, repeating the previous argument (except in the lower half plane), we can generate spacelike translation while keeping g(u) = 0.

Now we need to repeat the reasoning for x_{n-1} and x_n simultaneously. Since $e^{iHu}e^{-iHu} = 1$, we are left only with the expression

$$g'(u) = \langle \chi | \phi(x_1) \dots \phi(x_{n-1} + ut) \phi(x_n + ut) | \Omega \rangle = \langle \chi | \phi(x_1) \dots e^{iHu} \phi(x_{n-1}) \phi(x_n) | \Omega \rangle .$$
(4.17)

The operator e^{iHu} is still holomorphic in the upper half plane, and g'(u) is still zero on a finite segment of the real axis in u, and so, repeating the same reasoning as before, we can move the last

two points simultaneously outside of $\mathcal{U}_{\mathcal{V}}$ while keeping g(u) = 0. Repeating this for all points lets us move the entire set of field operators outside $\mathcal{U}_{\mathcal{V}}$ in succession.

With this, we have shown that if

$$\langle \chi | \psi_{\vec{f}} \rangle = 0 , \ \forall \ | \psi_{\vec{f}} \rangle , \forall f : supp(f) \in \mathcal{U}_{\mathcal{V}} ,$$

$$(4.18)$$

then

$$\langle \chi | \psi_{\vec{f}} \rangle = 0 , \ \forall f \in \mathcal{S}$$
 (4.19)

without the restriction of supp(f) being in $\mathcal{U}_{\mathcal{V}}$. Thus, we have proven the Reeh-Schlieder theorem; there are no nontrivial vectors in the vacuum sector of Hilbert space that are orthogonal to those generated by field operators with local support.

4.1.2 Reeh-Schlieder Corollary

Theorem 4.1.1 has interesting consequences for the relationship between local observables supported on spacelike separated spacetime regions, illustrated in this section by an example of negative local energy density. The more general implication is subtler in nature, and results in the applicability of so-called Tomita-Takesaki theory to QFT as well as a reformulation of axiomatic field theory in terms of von Neumann algebras.

Let $\mathcal{V} \in \Sigma$ be a subset such that its complement \mathcal{V}' in Σ is not empty. Since the two sets are disjoint, they are also spacelike separated. Assume further that they are contained in small, spacelike separated subsets of spacetime $\mathcal{U}_{\mathcal{V}}, \mathcal{U}_{\mathcal{V}'}$. Note that since these regions are adjacent, they share a boundary. The spacelike thickenings must then be performed in a lightlike direction. This is compatible with the previous discussion, the consequence is that some boundary conditions near this boundary need to be defined using lightcone coordinates so that the thickenings remain spacelike separated.

Now let A be any operator supported in $\mathcal{U}_{\mathcal{V}}$. Since $\mathcal{U}_{\mathcal{V}}$ is spacelike separated from $\mathcal{U}_{\mathcal{V}'}$ we have that

$$[A, \phi(x^{\mu})] = 0 , \quad \forall x \in \mathcal{U}_{\mathcal{V}'} .$$

$$(4.20)$$

Similarly, an operator A' supported in $\mathcal{U}_{\mathcal{V}'}$ fulfils

$$[A', \phi(x^{\mu})] = 0 , \quad \forall x \in \mathcal{U}_{\mathcal{V}} .$$

$$(4.21)$$

The Reeh-Schlieder theorem 4.1.1 applies to both \mathcal{V} and \mathcal{V}' . Suppose now that the operator A annihilates the vacuum state:

$$A\left|\Omega\right\rangle = 0. \tag{4.22}$$

Since A commutes with any operators $\phi(x^{\mu}) \ \forall x \in \mathcal{U}_{\mathcal{V}'}$ this implies that

$$A\phi[f_1]\phi[f_2]\dots\phi[f_n]|\Omega\rangle = 0 \ \forall f: \ supp(f) \in \mathcal{U}_{\mathcal{V}'} \ , \tag{4.23}$$

but the Reeh-Schlieder theorem tells us that $\phi_1[f_1]\phi[f_2]\dots\phi[f_n]|\Omega\rangle$ is dense in \mathcal{H}_0 , so the operator A must be zero acting on any state in the vacuum sector. This means that apart from the trivial operator A = 0, no locally supported operator can annihilate the vacuum.

To continue, and acquaint the reader with some important terminology, we now interject with some mathematical definition. First, let us define $\mathcal{A}_{\mathcal{U}}$ to be the algebra of operators supported in \mathcal{U} . In the previous discussion we have considered the regions $\mathcal{U} = \mathcal{U}_{\mathcal{V}}$ and $\mathcal{U}' = \mathcal{U}_{\mathcal{V}'}$, and the operators $A \in \mathcal{A}_{\mathcal{U}}, \mathcal{A}_{\mathcal{U}'}$. In mathematical terminology, a vector Ψ in a Hilbert space is called *cyclic* for $\mathcal{A}_{\mathcal{U}}$ if the states $A |\Psi\rangle, A \in \mathcal{A}_{\mathcal{U}}$ are dense in \mathcal{H}_0 . In words, if $|\Psi\rangle$ is cyclic in $\mathcal{A}_{\mathcal{U}}$ you can span the entire Hilbert space by acting on $|\Psi\rangle$ with operators in $\mathcal{A}_{\mathcal{U}}$. In this terminology, the statement of the Reeh-Schlieder theorem is explicitly that the vacuum state is cyclic in the local algebra.

The vector Ψ is said to be separating for $\mathcal{A}_{\mathcal{U}}$ if $A |\psi\rangle = 0$ implies that A = 0 for all states in \mathcal{H}_0 . With this terminology, the vacuum is clearly separating for both $\mathcal{A}_{\mathcal{U}}$ and $\mathcal{A}_{\mathcal{U}'}$. A more intuitive restatement is that a separating vector cannot be annihilated by a nontrivial operator in $\mathcal{A}_{\mathcal{U}}$. The separating property of the QFT vacuum is known as the Reeh-Schlieder corollary.

More generally, the Reeh-Schlieder theorem implies that in each superselection sector of Hilbert space, any vector on which the translation group acts holomorphically is cyclic and separating for $\mathcal{A}_{\mathcal{U}}$ and for $\mathcal{A}_{\mathcal{U}'}$. The restriction of holomorphic action of the translation group is exactly the requirement that was fulfilled by the vacuum in our proof of the Reeh-Schlieder theorem. While we used the flatness of Minkowski to make claims about the operator \hat{H} , the Reeh-Schlieder theorem holds for some states in curved spacetimes, though the states in question are not guaranteed to be physically interesting [33].

Let us now use the fact that the vacuum is separating for the algebra $\mathcal{A}_{\mathcal{U}}$ to demonstrate the nonpositive-definiteness of the local energy density. The total energy H annihilates only the vacuum Ω , and can be defined as the integral of the 00-component of the stress tensor T^{00} over a Cauchy surface (such as t = 0). The energy density T^{00} can be expressed in terms of smeared operators $T_f = \int_{\mathcal{V}} d^{D-1} x f(x) T^{00}(x)$, where f is any smooth function such that $supp(f) \in \mathcal{U}_{\mathcal{V}}$. The functionals T_f are contained in $\mathcal{A}_{\mathcal{U}}$. Poincaré invariance and $H |\Omega\rangle$ together imply that $\langle \Omega | T^{00}(x) | \Omega \rangle = 0$. Since $\langle \Psi | T^{00}(x) | \Psi \rangle \neq 0$ for general states $\Psi \in \mathcal{H}_0$, the separating property of the vacuum in $\mathcal{A}_{\mathcal{U}}$ implies that $T^{00} | \Omega \rangle \neq 0$, unless all states in \mathcal{H}_0 have zero energy (which may be considered a trivial exception). Therefore, there must be some state $\chi \in \mathcal{H}_0$, such that $\langle \chi | T_f | \Psi \rangle \neq 0$. We can now define \mathcal{W} as the two-dimensional subspace of \mathcal{H}_0 generated by the basis $\{\chi, \Omega\}$. Then the form of T_f restricted to the subspace \mathcal{W} is

$$T_f = \begin{bmatrix} 0 & \langle \chi | T_f | \Psi \rangle \\ \langle \Psi | T_f | \chi \rangle & 0 \end{bmatrix} .$$
(4.24)

A matrix on the form equation (4.24) is not positive semi-definite, implying that there are states $\tilde{\chi} \in \mathcal{W} \in \mathcal{H}_0$ in which the energy density has a negative expectation value.

An Intuitive but Incomplete Connection to Density Operators

A somewhat intuitive connection between the Reeh-Schlieder theorem and the density operator can be made in the case of a finite-dimensional Hilbert space on the form $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$. We will find explicitly that *all* degrees of freedom of a reduced density operator on one of the two subspaces are necessarily entangled.

Now, let \mathcal{A}_1 and \mathcal{A}_2 denote the algebras of all operators acting only on \mathcal{H}_1 and \mathcal{H}_2 respectively. A generic state in \mathcal{H} can be written in terms of bases for the component spaces as

$$|\Psi\rangle = \sum_{j \in [1, dim(\mathcal{H}_1)], \ k \in [1, dim(\mathcal{H}_2)]} c_{j,k} |j\rangle_1 \otimes |k\rangle_2 \quad , \tag{4.25}$$

where the $|j\rangle_1$ are basis vectors that span \mathcal{H}_1 , and the $|k\rangle_2$ span \mathcal{H}_2 .

The vector Ψ can be cyclic in both \mathcal{A}_1 and \mathcal{A}_2 if and only if $dim(\mathcal{H}_1) = dim(\mathcal{H}_2)$. To discuss the analogy with the Reeh-Schlieder theorem we thus have to specify to the case of both Hilbert spaces being of the same dimension D. In QFT, the issue of different-dimensional Hilbert spaces does not come up, since elements in both \mathcal{A}_1 and \mathcal{A}_2 act on the same space. By a change of basis kets, $|\Psi\rangle$ can be written

$$\sum_{k=1}^{D} c_k \left| k \right\rangle_1 \otimes \left| k \right\rangle_2 \ . \tag{4.26}$$

in terms of bases $|k\rangle_1, |k\rangle_2$ for $\mathcal{H}_1, \mathcal{H}_{\in}$. If Ψ is to be separating in \mathcal{A}_1, c_k can have no nonzero components. If there exists an index n for which $c_n = 0$, then there exists a nonzero operator $O = |n\rangle_1 \langle n|_1 \otimes |n\rangle_2 \langle n|_2$ that is nontrivial but fulfills $O |\Psi\rangle = 0$.

For the reduced density operators ρ_1 and ρ_2 defined in $\mathcal{D}(\mathcal{H}_1)$ and $\mathcal{D}(\mathcal{H}_2)$ respectively (notation as in section 2.1), the condition that all c_k are nonzero implies that the density operators are invertible. Since the density operators are invertible, they have only nonzero eigenvalues, implying that they describe a state in which *all* degrees of freedom are mixed.

This type of "universally" entangled state is not particularly interesting in typical applications of quantum information theory, but if spacetime emerges from entanglement as we are setting out to show it warrants further discussion. Specifically if entanglement is geometry, this is telling us that all degrees of freedom contribute to geometry. This seems to point towards universality of entanglement given a cyclic separating state being related to nothing less than the universal coupling of matter fields to gravity via the stress-energy tensor [34].

Some care has to be taken in applying this to QFT, since we would be claiming that we can factorize the Hilbert space in such a way that $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ where \mathcal{H}_1 has support only in a finite region of spacetime, \mathcal{H}_2 in its complement. This factorization is not possible in the literal sense. If it were there would exist overall pure states in the vacuum sector of the QFT Hilbert space where the local states ψ , χ are *also* pure, for example $\psi \otimes \chi, \psi \in \mathcal{H}_1, \chi \in \mathcal{H}_2$. The Reeh-Schlieder theorem already tells us that such states do not exist, so we are not allowed to factorize Hilbert space in this way.

Let us finish this aside by proving the claim about the invertibility of the reduced density operators by performing a partial trace, considering the case when $dim(\mathcal{H}_1) = dim(\mathcal{H}_2) = D$:

$$\rho_{1} = \sum_{k \leq D} \sum_{l \leq D} c_{k} c_{l}^{*} |k\rangle_{1} \langle l|_{1} \langle k|_{2} |l\rangle_{2}$$

$$= \sum_{k \leq D} \sum_{l \leq D} c_{k} c_{l}^{*} \delta_{kl} |k\rangle_{1} \langle k|_{1}$$

$$= \sum_{k \leq D} |c_{k}|^{2} |k\rangle_{1} \langle k|_{1} .$$

$$(4.27)$$

Since the spectrum of ρ_1 contains only positive nonzero eigenvalues, it is clearly invertible.

4.1.3 Von Neumann Algebras in QFT

In the previous section we spoke about the algebra $\mathcal{A}_{\mathcal{U}}$ associated with 'all operators' supported in a region $\mathcal{U}_{\mathcal{V}}$. In this section we will argue that the correct characterization of this algebra of observables is as a *von Neumann* algebra, which is a type of algebra of bounded operators with friendly closure properties.

In the previous section we restricted only to *simple operators*, operators that can be expressed as polynomials in smeared local fields. Simple operators are of course very useful and make up the basic machinery of regularization and operator product expansions. The problem with such a restriction is made intuitive by following setup.

Imagine that we have a slight thickening of a finite subset of a Cauchy surface that we call $\mathcal{U}_{\mathcal{V}}$, and an associated algebra $\mathcal{A}_{\mathcal{U}}$. Then, we would expect the algebra of observables $\hat{\mathcal{A}}_{\mathcal{U}}$ in the

domain of dependence $\hat{\mathcal{U}}_{\mathcal{V}}$ of the Cauchy surface to fulfill $\mathcal{A}_{\mathcal{U}} = \hat{\mathcal{A}}_{\mathcal{U}}$. The logic is that all operators with support in $\hat{\mathcal{U}}_{\mathcal{V}}$ are completely determined by a time evolution operator acting on the operators in $\mathcal{U}_{\mathcal{V}}$. The time evolution of operators in general results in exceedingly complex functions of the original operators, meaning that $\hat{\mathcal{A}}_{\mathcal{U}}$ cannot be described using only simple operators. Thus for the equality $\mathcal{A}_{\mathcal{U}} = \hat{\mathcal{A}}_{\mathcal{U}}$ to make sense, we need $\mathcal{A}_{\mathcal{U}}$ to contain all operators that can be made from simple ones by any imaginable Hamiltonian that satisfies the Wightman axioms.

The set of operators can be constructed with simple operators has a simple characterization. If F is any bounded function of a complex variable then $F(\phi_f)$ is also a bounded operator. More generally, we can have a bounded complex function of multiple variables, $F(\phi_{f_1}, \phi_{f_2} \dots \phi_{f_n})$, where $f_1, f_2 \dots f_n$ are n smearing functions.

Considering only bounded operators is nice. Since they are defined on all of Hilbert space they can be multiplied to make only other bounded operators, naturally forming an algebra. In addition the function of an operator is in general defined in terms of a Taylor expansion, and if it is a bounded operator we know that it can be arbitrarily closely approximated by a sequence of simple operators. A little more formally, Let $A_n, n = 1, 2...$ be a sequence of elements in $\mathcal{A}_{\mathcal{U}}$ such that the limit

$$\lim_{n \to \infty} A_n \left| \Psi \right\rangle = A \left| \Psi \right\rangle \tag{4.28}$$

exists, then we define

$$\lim_{n \to \infty} A_n \in \mathcal{A}_{\mathcal{U}} . \tag{4.29}$$

This prescription is called closure under a strong limit sequence², and it ensures that all bounded functions of the field operators are included in the algebra (as limit sequences of simple operators).

Another important detail to note, now that we have argued that all bounded operators that can be made from the field operators can be obtained as limit sequences of simple operators, is that if $A \in \mathcal{A}_{\mathcal{U}}$, then $A^{\dagger} \in \mathcal{A}_{\mathcal{U}}$. This property of *closure under Hermitian conjugation* is due to the fact that if $\phi_f = \int d^D f \phi$ is a smeared field in a given region, then $\phi_f^{\dagger} = \int d^D f^* \phi$ is also a smeared field. Any operator constructed from the field operators is then also closed under Hermitian conjugation. An algebra acting on a Hilbert space that is closed under Hermitian conjugation is called a *algebra. Therefore any reasonable notion of what we mean by the algebra of observables $\mathcal{A}_{\mathcal{U}}$ must at the very least be a *-algebra.

A *-algebra that is closed under strong (or weak) limits is called a *von Neumann algebra*. This leads us to conclude that $\mathcal{A}_{\mathcal{U}}$ must be a von Neumann algebra.

If $\mathcal{A}_{\mathcal{U}}$ is a von Neumann algebra on \mathcal{H} , then we can define its *commutant* $\mathcal{A}_{\mathcal{U}}'$ as the algebra of all operators that commute with operators in $\mathcal{A}_{\mathcal{U}}$. The commutant $\mathcal{A}_{\mathcal{U}}'$ is also a von Neumann algebra, seen by $A'_n \in \mathcal{A}_{\mathcal{U}}'$ realizing that since

$$\lim_{n \to \infty} [A, A'_n] \Psi = [A, \lim_{n \to \infty} A'] |\Psi\rangle = [A, A'] |\Psi\rangle = 0 , \forall n$$
(4.30)

both sides of the strong limit prescription exist, letting us consistently define

$$A' = \lim_{n \to \infty} A'_n \in \mathcal{A}_{\mathcal{U}}' . \tag{4.31}$$

Since \mathcal{A}' is closed under strong limits it is a von Neumann algebra. Interestingly, $\mathcal{A}_{\mathcal{U}}'$ is a von Neumann algebra even if $\mathcal{A}_{\mathcal{U}}$ is not closed. Any elements that are in both $\mathcal{A}_{\mathcal{U}}'$ and $\mathcal{A}_{\mathcal{U}}$ are called *centers*. These are operators that commute with all operators in both algebras, analogous to the central charge of the Virasoro algebra.

²Originally, this reasoning was carried out using a weak limit, i.e. demanding that $\lim_{n\to\infty} \langle \Psi | A_n | \Psi \rangle = \langle \Psi | A | \Psi \rangle$. It is a nontrivial result due to to von Neumann that this leads to the same algebra $\mathcal{A}_{\mathcal{U}}$.

When $\mathcal{A}_{\mathcal{U}}$ is a von Neumann algebra, the commutant relation is reciprocal, meaning that $\mathcal{A}_{\mathcal{U}}'' = (\mathcal{A}_{\mathcal{U}}')' = \mathcal{A}_{\mathcal{U}}$. If $\mathcal{A}_{\mathcal{U}}$ is not closed under strong limits, then $\mathcal{A}_{\mathcal{U}}''$ is called the *closure* of $\mathcal{A}_{\mathcal{U}}$ since, as we saw before, the commutant operator always defines an algebra closed under strong limits.

Relating more concretely to QFT, it was proposed by Haag and Schroeder that if \mathcal{U} and \mathcal{U}' are causal complements, i.e. \mathcal{U}' is maximal under the condition of being spacelike separated from \mathcal{U} , then the corresponding algebras $\mathcal{A}_{\mathcal{U}}$ and $\mathcal{A}_{\mathcal{U}'}$ are commutants, meaning they are maximal under the condition of commuting with each other [31]. This condition is often called Haag duality, and can be written

$$\mathcal{A}_{\mathcal{U}'} = \mathcal{A}_{\mathcal{U}}' \ . \tag{4.32}$$

Haag duality is a postulate of the Haag-Kastler approach to axiomatizing QFT, which is similar to the Wightman axiomatization from section 4.1, formulated in terms of properties of the algebras of local observables. Furthermore the postulate says that if \mathcal{U} is a union of open sets \mathcal{U}_{α} , then $\mathcal{A}_{\mathcal{U}}$ is the smallest von Neumann algebra that contains all of the $\mathcal{A}_{\mathcal{U}_{\alpha}}$. The Haag-Kastler framework is usually referred to as *local quantum field theory* (LQFT), since it is formulated in terms of the algebras of local observables. This framework is more amenable to curved spacetimes since it does not require us to have a neat Hamiltonian understanding of the spacetime in question.

The Haag duality postulate holds for complementary so-called Rindler wedges in Minkowski, as we will discuss in section 4.2. In addition, it seems to hold in several nontrivial quantum field theories, among them CFT in 1+1 dimensions [35]. There are however cases on curved backgrounds and in higher spin theories where Haag duality fails to hold. It is also arguable if Haag duality *should* hold in a fully quantum description of geometry.

On a related note, given a quantum description of geometry, the notion that any operators should commute exactly like the von Neumann algebras and their duals is up for debate [9]. Simply put, if we act with a gigantic number of creation operators in a region $\mathcal{U}_{\mathcal{V}}$, resulting infinite number of massive particles results in an infinitely large black hole. This stack of operators obviously affects local operators in the complement of $\mathcal{U}_{\mathcal{V}}$, so the commutation of spacelike separated operators can not be exact in quantum gravity.

4.1.4 Tomita-Takesaki theory

Tomita-Takesaki theory tells us about the properties of an algebra \mathcal{A} that has a cyclic separating vector Ψ . What we have shown with the Reeh-Schlieder theorem is that Tomita-Takesaki theory may be applied to QFT.

The starting point of Tomita-Takesaki theory is that we have an algebra \mathcal{A} of operators $A \in \mathcal{A}$, and there is a cyclic separating vector Ψ for \mathcal{A} . We begin by defining the antilinear Tomita operator $S_{\Psi} : \mathcal{H} \to \mathcal{H}$, defined by its action

$$S_{\Psi}A |\Psi\rangle = A^{\dagger} |\Psi\rangle \quad . \tag{4.33}$$

This definition of S_{Ψ} is only sensible thanks to the separating property $(A |\Psi\rangle = 0 \Rightarrow A = 0)$, so we have no situations where S_{Ψ} exhibits singular behaviour such as if $A |\Psi\rangle = 0, A^{\dagger} |\Psi\rangle \neq 0$. The cyclic property ensures that $A |\Psi\rangle$ is dense in \mathcal{H} , and together with the non-singular behaviour of S, it must map the dense set spanned by $A |\Psi\rangle$ to a dense set in \mathcal{H}_0 .

An antilinear operator A fulfills $A[|\chi\rangle + |\psi\rangle] = A |\chi\rangle + A |\psi\rangle$ and $cA |\psi\rangle = Ac^* |\psi\rangle$, where c^* is the complex conjugate of c. The "anti" simply refers to the conjugation property of the operator. In the same sense, an antiunitary operator squares to the identity but complex conjugates scalars.

It is worth noting, that the Schwarz kernel theorem told us that in QFT we can obtain what is essentially singular operators as a limit $n \to \infty$ of some $A_n |\Psi\rangle$. If we want the Tomita operator to be defined for all states in the vacuum sector, we also have to define its action on these limits as follows. Let $A_n, n = 1, 2...$ be a sequence of elements in $\mathcal{A}_{\mathcal{U}}$ such that the limits

$$x = \lim_{n \to \infty} A_n |\Psi\rangle$$
 and $y = \lim_{n \to \infty} A_n^{\dagger} |\Psi\rangle$ (4.34)

both exist. We then define

$$S_{\Psi}x = y . \tag{4.35}$$

By extending the definition of S_{Ψ} in this way, we have defined its action on all states in the vacuum sector of Hilbert space. We also ensure that S_{Ψ} is a so-called *closed* operator, meaning it maps onto a closed subspace.

A couple of facts about S_{Ψ} are that

$$S_{\Psi}^2 = 1$$
, (4.36)

and if we let the operator A be $\mathbf{1}$, we see that

$$S_{\Psi} |\Psi\rangle = |\Psi\rangle$$
 . (4.37)

The Tomita operator has an adjoint, S_{Ψ}^{\dagger} . Since it is antilinear the definition of the adjoint operator is that for all χ, ξ

$$\langle \xi | S | \chi \rangle = \langle \chi | S^{\dagger} | \xi \rangle , \qquad (4.38)$$

where the RHS is the complex conjugate of the definition for linear operators.

The Tomita operator for the commutant of $\mathcal{A}_{\mathcal{U}}$, i.e. the algebra of observables $\mathcal{A}_{\mathcal{U}}'$ supported in the complement of \mathcal{U} , is given by

$$S'_{\Psi} = S^{\dagger}_{\Psi} . \tag{4.39}$$

The action of S'_{Ψ} is given by the action on $A' \in \mathcal{A}_{\mathcal{U}}'$ as

$$S'_{\Psi}A' |\Psi\rangle = A^{'\dagger} |\psi\rangle$$

To show that $S'_{\Psi} = S^{\dagger}_{\Psi}$ we simply turn to the definition of the adjoint operator. Let $\chi = A\Psi$ and $\xi = A'\Psi$. Reeh-Schlieder tells us that both of these states are dense in Hilbert space, so they are adequate 'test states' for the definition of adjointedness. Going to the definition equation (4.38) we have

$$\begin{aligned} \langle \chi | S'_{\Psi} | \xi \rangle &= \langle \Psi | A^{\dagger} S'_{\Psi} A' | \Psi \rangle \\ &= \langle \Psi | A^{\dagger} A'^{\dagger} | \Psi \rangle \\ &= \langle \Psi | A'^{\dagger} A^{\dagger} | \Psi \rangle \\ &= \langle \xi | S_{\Psi} A | \Psi \rangle \\ &= \langle \xi | S_{\Psi} | \chi \rangle , \end{aligned}$$

$$(4.40)$$

Where the first and last lines define $S'_{\Psi} = S^{\dagger}_{\Psi}$, at least for all $|\Psi\rangle$ on which S'_{Ψ} is defined.

Since the Tomita operator is invertible it has a unique polar decomposition:

$$S_{\Psi} = J_{\Psi} \Delta_{\Psi}^{1/2} . \tag{4.41}$$

A polar decomposition is analogous to the polar form of complex numbers, in this case J_{Ψ} is an antiunitary matrix (loosely, the 'phase') and $\Delta_{\Psi}^{1/2}$ is a Hermitian, positive definite matrix (loosely, the 'absolute value'). These are called the *modular conjugation* and *modular operator*, respectively.

The *modular operator* is defined as

$$\Delta_{\Psi} = S^{\dagger}S . \tag{4.42}$$

The modular operator is self-adjoint, and it is positive definite due to the invertibility of S. Since $S |\Psi\rangle = S^{\dagger} |\Psi\rangle = |\Psi\rangle$ we have that

$$\Delta_{\Psi} |\Psi\rangle = |\Psi\rangle \quad . \tag{4.43}$$

This implies that for any function f,

$$f(\Delta_{\Psi}) |\Psi\rangle = f(1) |\Psi\rangle \quad . \tag{4.44}$$

In addition, since $S_{\Psi}^2 = 1$, we have that

$$J_{\Psi} \Delta_{\Psi}^{1/2} J_{\Psi} \Delta_{\Psi}^{1/2} = 1 \implies J_{\Psi} \Delta_{\Psi}^{1/2} J_{\Psi} = \Delta_{\Psi}^{-1/2} .$$
 (4.45)

Rewriting slightly, we have

$$J_{\Psi}^{2}(J_{\Psi}^{-1}\Delta_{\Psi}^{1/2}J_{\Psi}) = 1 \cdot \Delta_{\Psi}^{-1/2}$$
(4.46)

But both $(J_{\Psi}^{-1}\Delta_{\Psi}^{1/2}J_{\Psi})$ and Δ_{Ψ}^{-1} must be positive definite, therefore $J_{\Psi}^2 = 1$ (it is unitary and its square must be real positive). We showed earlier that $S'_{\Psi} = S_{\Psi}^{\dagger}$ where the prime denotes the Tomita operator of the commutant algebra. Comparing to the polar decomposition $S'_{\Psi} = J'_{\Psi}\Delta_{\Psi}^{1/2}$ we see that

$$S'_{\Psi} = S^{\dagger}_{\Psi} = \Delta^{1/2}_{\Psi} J^{\dagger}_{\Psi} = \Delta^{1/2}_{\Psi} J_{\Psi} = J_{\Psi} \Delta^{-1/2}_{\Psi}$$
(4.47)

implies that

$$J'_{\Psi} = J_{\Psi} , \quad \Delta'_{\Psi} = \Delta_{\Psi}^{-1} .$$
 (4.48)

Finally, since

$$J_{\Psi} \Delta_{\Psi}^{1/2} J_{\Psi} J_{\Psi} \Delta_{\Psi}^{1/2} J_{\Psi} = J_{\Psi} \Delta_{\Psi} J_{\Psi} = \Delta_{\Psi}^{-1} , \qquad (4.49)$$

we have that

$$J_{\Psi}f(\Delta_{\Psi})J_{\Psi} = \overline{f}(\Delta_{\Psi}^{-1}) . \qquad (4.50)$$

To understand this, consider that the definition of a function of an operator is a Taylor expansion, and insert a factor J_{Ψ}^2 between each Δ_{Ψ} in the expansion. The conjugation then comes from moving the leftmost J_{Ψ} past any coefficients in the expansion and the antiunitary property.

A function that we will find particularly useful is $f(\Delta_{\Psi}) = \Delta_{\Psi}^{is}$ with s real, where we find that

$$J_{\Psi} \Delta_{\Psi}^{is} J_{\Psi} = \Delta_{\Psi}^{is} . \tag{4.51}$$

Relative Tomita- and Modular Operators

The relative Tomita operator $S_{\Psi,\Phi}: \mathcal{H} \to \mathcal{H}$ is an antilinear operator defined by

$$S_{\Psi|\Phi}A|\Psi\rangle = A^{\dagger}|\Phi\rangle \quad , \tag{4.52}$$

and it is well defined for the same reasons as S_{Ψ} . The state Φ is any arbitrary state in the *entire* Hilbert space, and need not be cyclic nor separating. When Φ is not cyclic separating, $S_{\Psi|\Phi}$ is not generally invertible. It is usual to define the vectors in Hilbert space to be unit vectors,

$$\langle \Psi | \Psi \rangle = \langle \Phi | \Phi \rangle = 1 . \tag{4.53}$$

The definition of the relative Tomita operator is extended to all operators by the use of limiting sequences, as in the case of the original Tomita operator.

In principle, the Tomita operator can be decomposed as

$$S_{\Psi|\Phi} = J_{\Psi|\Phi} \Delta_{\psi|\Phi}^{1/2} , \qquad (4.54)$$

however there are some subtleties. If Φ is not separating, $S_{\Psi|\Phi}$ has a nontrivial null space. For the relative modular conjugation $J_{\Psi|\Phi}$ to be uniquely defined like its nonrelative cousin we define it to have the same null space as $S_{\Psi|\Phi}$. If Φ is not cyclic, then $S_{\Psi|\Phi}$ is not dense in \mathcal{H} . In the non-cyclic case $J_{\Psi|\Phi}$ is an antiunitary map from the orthocomplement of the null space of $S_{\Psi|\Phi}$ to the (not dense) subspace spanned by $S_{\Psi|\Phi}$. In the case where Φ is cyclic and separating, $J_{\Psi|\Phi}$ is completely antiunitary, and enjoys all the properties of its nonrelative cousin.

The relative modular operator is given in analogy with equation (4.42) by

$$\Delta_{\Psi|\Phi} = S_{\Psi|\Phi}^{\dagger} S_{\Psi|\Phi} . \tag{4.55}$$

The relative modular operator is not necessarily invertible, but it is positive definite just like the modular operator. The lack of invertibility is due to the lack of the cyclic and separating properties in Φ , preventing $S_{\Psi|\Phi}$ from always mapping between dense sets in \mathcal{H} of same size. Also, it is clear by definition that

$$\Delta_{\Psi|\Psi} = \Delta_{\Psi} , \quad S_{\Psi|\Psi} = S_{\Psi} .$$

Another useful property of the relative modular operator is that it is invariant under $\Phi \rightarrow \Phi' = A'\Phi$, where $A' \in \mathcal{A}'_{\mathcal{U}}$ and $A'^{\dagger}A' = \mathbf{1}$, that is, A' is a unitary element in the complementary spacetime algebra, $\mathcal{A}_{\mathcal{U}}'$. Let us check this, letting $A \in \mathcal{A}_{\mathcal{U}}$:

$$\langle A\Psi | \Delta_{\Psi,A'\Phi} | A\Psi \rangle = \langle A\Psi | S_{\Psi,A'\Phi}^{\dagger} S_{\Psi,A'\Phi} | A\Psi \rangle$$

$$= \langle A^{\dagger}A'\Phi | A^{\dagger}A'\Phi \rangle$$

$$= \langle A^{\dagger}\Phi | A'^{\dagger}A' | A^{\dagger}\Phi \rangle$$

$$= \langle A^{\dagger}\Phi | A^{\dagger}\Phi \rangle ,$$

$$(4.56)$$

where we have used that the operators A, A' commute. The last line is precisely the action of $\langle A\Psi | \Delta_{\Psi,\Phi} | A\Psi \rangle$.

As a final note, when it is not obvious, for example when we are working with several spacetime regions, we will denote the relative operators as

$$\Delta_{\Psi|\Phi}, A \in \mathcal{A}_{\mathcal{U}} \to \Delta_{\Psi|\Phi:\mathcal{U}} . \tag{4.57}$$

Relative Entropy in QFT

Having developed the Tomita- and modular operators, we are now ready to define the relative entropy in QFT. Consider an open spacetime region \mathcal{U} such that it is spacelike separated from some other nonempty region in spacetime. There is some algebra $\mathcal{A}_{\mathcal{U}}$ of operators with support only in \mathcal{U} . Then, letting Ψ be a cyclic separating vector for $\mathcal{A}_{\mathcal{U}}$ and Φ be any vector in \mathcal{H} , the relative entropy between states Ψ, Φ in the region \mathcal{U} is given by

$$D_{\Psi||\Phi}(\mathcal{U}) = -\langle \Psi| \log(\Delta_{\Psi|\Phi;\mathcal{U}}) |\Psi\rangle \quad . \tag{4.58}$$

It is by no means obvious that this is the field theoretical generalization of the quantum relative entropy defined in equation (2.57), but we will show that it reduces to it in the case of the typical factorizable systems considered in quantum information theory. In analog to the divergence of the relative entropy in equation (2.57), when the support of Φ is not contained in the support of Ψ , this version of the quantity may diverge when Φ is not a separating vector in $\mathcal{A}_{\mathcal{U}}$.

The relative entropy is zero whenever $|\Phi\rangle = A' |\Psi\rangle$, where A' is a unitary element in the commuting algebra. We already showed that $\Delta_{\Psi|\Phi} = \Delta_{\Psi|A'\Phi}$, so it suffices to show that

$$D_{\Psi||\Psi}(\mathcal{U}) = -\langle \Psi|\log(\Delta_{\Psi|\Psi;\mathcal{U}})|\Psi\rangle = -\langle \Psi|\log(S_{\Psi}^{\dagger}S_{\Psi})|\Psi\rangle = -\langle \Psi|\log(\mathbf{1})|\Psi\rangle = 0 , \quad (4.59)$$

where we used that $S_{\Psi|\Psi} = S_{\Psi}$ and $S|\Psi\rangle = |\Psi\rangle$ to observe that the argument to the logarithm is an identity operator when acting on the states involved.

Let us now show the non-negativity of the relative entropy. The first step is using the inequality $-\log \lambda \ge 1 - \lambda$ for real, positive λ . Since the relative modular operator is positive semi-definite, it has eigenvalues $\lambda \ge 0$, so the inequality applies. We showed in equation (2.46) the necessary machinery to claim that $\log \Delta$ and Δ are diagonalized by the same basis, meaning that the inequality for real numbers translates to the operator inequality

$$-\log \Delta_{\Psi|\Phi} \ge 1 - \Delta_{\Psi|\Phi} . \tag{4.60}$$

Let us now insert the LHS of this inequality into the definition of the relative entropy

$$D_{\Psi||\Psi}(\mathcal{U}) \ge \langle \Phi | \Phi \rangle - \langle \Psi | (\Delta_{\Psi|\Psi:\mathcal{U}}) | \Psi \rangle$$

= 1 - \langle \Psi | S^{\dagger}_{\Psi|\Phi} S_{\Psi|\Phi} | \Psi \rangle
= 1 - \langle \Psi | \Phi \rangle
= 0, (4.61)

since we assume normalized states. Thus, we have shown the positivity of relative entropy.

Monotonicity of Relative Entropy

In the quantum informational case (section 2.3.6), the relative entropy was monotone under the tracing out of subsystems in the argument. We would like to show a similar property here, namely that the relative entropy defined as in equation (4.58) is monotone under the shrinking of the spacetime region on which it is defined. In other words, we wish to show that for a region $\tilde{\mathcal{U}} \in \mathcal{U}$,

$$D_{\Psi||\Phi}(\mathcal{U}) \ge D_{\Psi||\Phi}(\tilde{\mathcal{U}}) . \tag{4.62}$$

The original proof of the monotonicity of relative entropy in the form considered in this section was given by Araki [36], although we are following the presentation of Witten [32]. For ease of notation, let us keep Ψ , Φ constant for the rest of this section, and abbreviate

$$\Delta_{\Psi|\Phi:\mathcal{U}} \equiv \Delta_{\mathcal{U}} , \quad \Delta_{\Psi|\Phi:\tilde{\mathcal{U}}} \equiv \Delta_{\tilde{\mathcal{U}}} ,$$

The monotonicity of relative entropy is a direct result of the operator inequality

$$\Delta_{\tilde{\mathcal{U}}} \ge \Delta_{\mathcal{U}} . \tag{4.63}$$

The mathematically sound statement of any operator inequality such as this, for operators that do not diagonalize simultaneously, is

$$\langle \chi | \Delta_{\tilde{\mathcal{U}}} - \Delta_{\mathcal{U}} | \chi \rangle \ge 0 , \quad \forall \chi \in \mathcal{H} .$$
 (4.64)

The statement equation (4.64) holds as long as the relative modular operators are bounded, so to include the non-invertible cases we in principle need to extend this inequality with the same limit sequence prescription as we have used before. Since both $\Delta_{\tilde{\mathcal{U}}}$ are positive semi-definite, an equivalent inequality statement is that

$$\frac{1}{s + \Delta_{\mathcal{U}}} \ge \frac{1}{s + \Delta_{\tilde{\mathcal{U}}}} , \qquad (4.65)$$

for all real s > 0. As long as we do not take s = 0, this form of the inequality covers the unbounded cases of the operators and has the correct limit behavior because when Δ is unbounded positive we just get $\frac{1}{s+\Delta} = 0$.

To see that this implies the same inequality we can consider the operator $P(t) = t\Delta_{\tilde{\mathcal{U}}} + (1-t)\Delta_{\mathcal{U}}$. If equation (4.64) holds, then $\frac{d}{dt}P(t) \ge 0$. We can check this by explicitly computing the time derivative:

$$\frac{d}{dt}\left(\frac{1}{s+P(t)}\right) = -\frac{1}{s+P(t)}\dot{P}(t)\frac{1}{s+P(t)} .$$
(4.66)

The RHS is negative, because it is of the form -ABA with B a positive operator (implied by equation (4.64)), and A a self-adjoint operator since it is a linear combination of the self-adjoint modular operators. Thus, $\frac{1}{s+P(t)}$ is decreasing in t, meaning $\frac{1}{s+P(0)} \ge \frac{1}{s+P(1)}$, and we have found that

$$\left(\Delta_{\tilde{\mathcal{U}}} \ge \Delta_{\mathcal{U}}\right) \Leftrightarrow \left(\frac{1}{s + \Delta_{\mathcal{U}}} \ge \frac{1}{s + \Delta_{\tilde{\mathcal{U}}}}\right).$$
 (4.67)

The arrow from the LHS to the RHS follows by observing that the sign in the LHS of equation (4.66) is entirely determined by \dot{P} , so assuming either equation (4.64) or equation (4.65) implies the other.

To show that equation (4.65) implies equation (4.62) we now use a nice trick,

$$\log P = \int_0^\infty \mathrm{d}s \left(\frac{1}{1+s} - \frac{1}{s+P}\right)$$
$$= \left[\log\left(\frac{1+s}{s+P}\right)\right]_{s=0}^{s=\infty}$$
$$= -\log\frac{1}{P}.$$
 (4.68)

We showed before that $\frac{1}{s+P(t)}$ is decreasing in t, and using the convergence of the integral to exchange derivative with respect to t and integration we see that $\log P(t)$ must be *increasing* in t. This in turn tells us that equation (4.64) implies

$$\log \Delta_{\tilde{\mathcal{U}}} \ge \log \Delta_{\mathcal{U}} . \tag{4.69}$$

What remains in the proof of monotonicity is now proving the inequality that was used:

$$\langle \chi | \frac{1}{s + \Delta_{\tilde{\mathcal{U}}}} - \frac{1}{s + \Delta_{\mathcal{U}}} | \chi \rangle \ge 0 , \quad \forall \chi \in \mathcal{H} .$$
 (4.70)

The essence in this proof is to use the fact that $\Delta_{\mathcal{U}} = S_{\Psi|\Phi:\mathcal{U}}^{\dagger} S_{\Psi|\Phi:\mathcal{U}}$, where $S_{\mathcal{U}} \equiv S_{\Psi|\Phi:\mathcal{U}}$ is a map from one Hilbert space \mathcal{H} to a possibly different space \mathcal{H}' . We can now define $\hat{\mathcal{H}} = \mathcal{H} \oplus \mathcal{H}'$, and consider the so-called graph Γ of $S_{\mathcal{U}}$, which is the set of all vectors on the form $(\chi, S_{\mathcal{U}}\chi)$. The graph Γ is a subspace of $\hat{\mathcal{H}}$, and it is a *closed* subspace since $S_{\mathcal{U}}$ is a closed operator. The closed property just says that if the limits

$$x = \lim_{n \to \infty} x_n , \quad y = \lim_{n \to \infty} S_{\mathcal{U}} x_n \tag{4.71}$$

both exist, they are also part of the subspace Γ . Since the Tomita operator and its relative version both were extended to act on such limit points, their image also contain them.

For a closed subspace such as $\Gamma \in \hat{\mathcal{H}}$ it is possible to define an orthogonal projection operator $\Pi : \hat{\mathcal{H}} \to \Gamma$. An orthogonal projection operator is necessarily bounded since it has only eigenvalues

that are one or zero, and as such it is defined on all states. The closed property of the subspace is necessary because of the boundedness of the projector; if Γ were not closed there would be points in $\hat{\mathcal{H}}$ that exist as limit points that would be projected outside of Γ .

With Γ the graph of $S_{\mathcal{U}}$ we want to find a projector Π that, acts on a column vector (χ, ψ) with $\chi \in \mathcal{H}, \psi \in \mathcal{H}'$ in the following way:

$$\Pi\begin{pmatrix}\chi\\\psi\end{pmatrix} = \begin{pmatrix}\eta\\S_{\mathcal{U}}\eta\end{pmatrix} . \tag{4.72}$$

Furthermore, we wish for $\Pi = \Pi^{\dagger}$ to ensure orthogonality of the eigenvectors, and $\Pi^2 = \Pi$ so that it actually is a projection. It can be straightforwardly checked that (denoting $S_{\mathcal{U}} \equiv S$ for cleanliness)

$$\Pi = \begin{bmatrix} (1+S^{\dagger}S)^{-1} & (1+S^{\dagger}S)^{-1}S^{\dagger} \\ S(1+S^{\dagger}S)^{-1} & S(1+S^{\dagger}S)^{-1}S^{\dagger} \end{bmatrix}$$
(4.73)

fulfills the requirements.

Let us verify explicitly that Π acts in the appropriate way on (χ, ψ) :

$$\Pi\begin{pmatrix} \chi\\ \psi \end{pmatrix} = \begin{pmatrix} (1+S^{\dagger}S)^{-1}\chi + (1+S^{\dagger}S)^{-1}S^{\dagger}\psi\\ S(1+S^{\dagger}S)^{-1}\chi + S(1+S^{\dagger}S)^{-1}S^{\dagger}\psi \end{pmatrix}$$

$$= \begin{pmatrix} (1+S^{\dagger}S)^{-1}(\chi+S^{\dagger}\psi)\\ S(1+S^{\dagger}S)^{-1}(\chi+S^{\dagger}\psi) \end{pmatrix}.$$
(4.74)

We see that Π has the correct action by letting $\eta = (1 + S^{\dagger}S)^{-1}(\chi + S^{\dagger}\psi)$.

Having introduced the necessary machinery, we can now prove our operator inequality. We have two densely defined operators $S_{\mathcal{U}}, S_{\tilde{\mathcal{U}}}$ that map from \mathcal{H} to \mathcal{H}' with graphs $\Gamma_{\mathcal{U}}, \Gamma_{\tilde{\mathcal{U}}}$. Let us further denote the corresponding projectors $\Pi_{\mathcal{U}}$ and $\Pi_{\tilde{\mathcal{U}}}$. We know that since $\tilde{\mathcal{U}} \in \mathcal{U}, S_{\mathcal{U}} |\chi\rangle = S_{\tilde{\mathcal{U}}} |\chi\rangle$ for all $|\chi\rangle$ for which $S_{\tilde{\mathcal{U}}}$ is defined. Therefore, $\Gamma_{\tilde{\mathcal{U}}}$ must be a subset of $\Gamma_{\mathcal{U}}$. Since $\Gamma_{\tilde{\mathcal{U}}}$ is a subset of $\Gamma_{\mathcal{U}}$, we obtain the operator inequality $\Pi_{\mathcal{U}} \geq \Pi_{\tilde{\mathcal{U}}}$, since the eigenvalues of $\Pi_{\mathcal{U}}$ are equal to 1 for all nonzero eigenvalues of $\Pi_{\tilde{\mathcal{U}}}$ (and for these eigenvalues, they have the same eigenvectors!), however $\Pi_{\mathcal{U}}$ may have additional nonzero eigenvalues. Thus, we have that $\langle \Psi | \Pi_{\mathcal{U}} | \Psi \rangle \geq \langle \Psi | \Pi_{\tilde{\mathcal{U}}} | \Psi \rangle$ for all $\Psi = (\chi, \psi)$. Specializing to $\psi = 0$ and using the explicit projectors as in equation (4.73) we find that the projector inequality $\Pi_{\mathcal{U}} \geq \Pi_{\tilde{\mathcal{U}}}$ reduces to

$$\langle \chi | \frac{1}{1 + S_{\mathcal{U}}^{\dagger} S_{\mathcal{U}}} | \chi \rangle \ge \langle \chi | \frac{1}{1 + S_{\tilde{\mathcal{U}}}^{\dagger} S_{\tilde{\mathcal{U}}}} | \chi \rangle \quad .$$

$$(4.75)$$

Redefining $S \to S/\sqrt{s}$ and dividing off an overall factor of s on both sides of the inequality yields

$$\langle \chi | \frac{1}{s + S_{\mathcal{U}}^{\dagger} S_{\mathcal{U}}} | \chi \rangle \ge \langle \chi | \frac{1}{s + S_{\tilde{\mathcal{U}}}^{\dagger} S_{\tilde{\mathcal{U}}}} | \chi \rangle \quad , \tag{4.76}$$

which is exactly equation (4.65), which we set out to show. With this, we have shown the monotonicity of relative entropy under the reduction of the spacetime region.

4.1.5 Finite-Dimensional Case of Tomita-Takesaki Theory

While we have already observed that QFT is not a theory of finite-dimensional Hilbert spaces, it is very useful to show Tomita-Takesaki theory in finite dimensions. It both has actual applications,

and serves as an explicit grounding of the ideas of the previous section. Certain ideas of this section extend to the infinite dimensional case [32].

In the finite-dimensional case, the existence of a factorization of the full Hilbert space as $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ is not fiction. Let us then define \mathcal{A} as the algebra of linear operators on \mathcal{H}_1 and \mathcal{A}' as the algebra of linear operators on \mathcal{H}_2 . An operator $A \in \mathcal{A}$ acts on \mathcal{H} as $A \otimes \mathbf{1}$, and an operator $A' \in \mathcal{A}'$ acts as $\mathbf{1} \otimes \mathcal{A}'$. The algebra \mathcal{A}' is maximal under the condition that it commutes with \mathcal{A} , so the algebras are each other's commutants. If the algebras are to share a cyclic separating vector the subspaces \mathcal{H}_1 and \mathcal{H}_2 need to be of the same dimension, as we saw when studying equation (4.26). Denoting orthogonal basis elements in \mathcal{H}_1 by $|k\rangle$ and in \mathcal{H}_2 by $|k\rangle'$, a general cyclic separating vector has the form

$$\Psi = \sum_{k} c_k \left| k \right\rangle \otimes \left| k' \right\rangle = \sum_{k} c_k \left| k, k \right\rangle , \qquad (4.77)$$

with all c_k nonzero.

Let us find the modular operators in this context. The definition of $S_{\Psi} : \mathcal{H} \to \mathcal{H}$ is

$$S_{\Psi}(A \otimes 1) |\Psi\rangle = (A^{\dagger} \otimes 1) |\Psi\rangle \quad . \tag{4.78}$$

Let us pick a matrix A by picking a basis $|k\rangle$ and defining its action as follows

$$A|i\rangle = |j\rangle, \quad A|k\rangle = 0 \text{ if } k \neq i .$$
 (4.79)

The action of the adjoint of A is then given by

$$A^{\dagger} |j\rangle = |i\rangle, \quad A^{\dagger} |k\rangle = 0 \text{ if } k \neq j .$$
 (4.80)

Then, extending to the full Hilbert space we have

$$(A \otimes \mathbf{1}) \sum_{k} |k,k'\rangle = c_i |j,i\rangle \quad , \quad (A^{\dagger} \otimes \mathbf{1}) \sum_{k} c_k |k,k'\rangle = c_j |i,j\rangle \quad .$$

$$(4.81)$$

The definition of S_{Ψ} then implies that

$$S_{\Psi}(c_i | j, i \rangle) = c_j | i, j \rangle \quad , \tag{4.82}$$

remembering that S_{Ψ} is antilinear we see that

$$S_{\Psi}|j,i\rangle = \frac{c_j}{c_i^*}|i,j\rangle \quad . \tag{4.83}$$

This actually completely characterizes S_{Ψ} , since the states $|i,j\rangle$ give a complete basis of \mathcal{H} (we just repeat the analysis for each choice of i,j in the operator A). The adjoint operator S_{Ψ}^{\dagger} acts as

$$S_{\Psi}^{\dagger} \left| i, j \right\rangle = \frac{c_j}{c_i^*} \left| j, i \right\rangle \ . \tag{4.84}$$

where the antilinear property has complex conjugated the prefactor relative to the linear definition of an adjoint operator. The modular operator is now straightforwardly computed, and its action is

$$\Delta_{\Psi} |j,i\rangle = S^{\dagger} \frac{c_j}{c_i^*} |j,i\rangle = \frac{c_j^*}{c_i} S^{\dagger} |i,j\rangle = \frac{|c_i|^2}{|c_j|^2} |j,i\rangle \quad .$$
(4.85)

We also wish to know the modular conjugation operator J_{Ψ} . Since

$$\Delta^{1/2} |j,i\rangle = \sqrt{\frac{|c_i|^2}{|c_j|^2}} |j,i\rangle \tag{4.86}$$

we can obtain J_{Ψ} by dividing the forefactor in the RHS off of equation (4.83):

$$J_{\Psi} \left| j, i \right\rangle = \sqrt{\frac{c_j c_i}{c_j^* c_i^*}} , \qquad (4.87)$$

where choice of conjugation in the square root is determined by the phase of $\frac{c_j}{c_{*i}}$.

Let us now move on to the relative operators. For this we need to define a second vector

$$|\Phi\rangle = \sum_{\alpha} d_{\alpha} |\phi\rangle_{\alpha} \otimes |\phi\rangle_{\alpha}' , \qquad (4.88)$$

where the $|\phi\rangle$ and $|\phi\rangle'$ are bases for $\mathcal{H}_1, \mathcal{H}_2$ respectively, with coefficients d_{α} . Let us abbreviate like in the previous case so that $|\phi\rangle_{\alpha} \otimes |\phi\rangle'_{\alpha} \equiv |\alpha, \alpha\rangle$. The state Φ is not necessarily cyclic nor separating for \mathcal{A} and \mathcal{A}' since we do not require $d_{\alpha} \neq 0$.

In the following, $|i,j\rangle$ denotes the basis states for Ψ , and $|\alpha,\beta\rangle$ for Φ . Repeating the method for the nonrelative operator, we pick $A \in \mathcal{A}$ such that

$$A|i\rangle = |\alpha\rangle, \quad A|k\rangle = 0, \text{ if } |k\rangle \neq i$$

$$(4.89)$$

implying that the adjoint is defined by

$$A^{\dagger} |\alpha\rangle = |i\rangle, \quad A^{\dagger} |\beta\rangle = 0, \text{ if } |\beta\rangle \neq \alpha .$$
 (4.90)

Extending to the full Hilbert space we have

$$(A \otimes \mathbf{1}) \sum_{k} c_{k} |k,k\rangle = c_{i} |\alpha,i\rangle \quad (A^{\dagger} \otimes \mathbf{1}) \sum_{\beta} d_{\beta} |\beta,\beta\rangle = d_{\alpha} |i,\alpha\rangle \quad .$$
(4.91)

The definition of $S_{\Psi|\Phi}$ (equation (4.78)) then gives us that

$$S_{\Psi|\Phi}A|j,i\rangle = S_{\Psi|\Phi}|\alpha,i\rangle \frac{d_{\alpha}}{c_i^*}|i,\alpha\rangle \quad . \tag{4.92}$$

The adjoint is given (not unlike the case in equation (4.84)) by

$$S_{\Psi|\Phi}^{\dagger}|i,\alpha\rangle = \frac{d_{\alpha}}{c_i^*}|\alpha,i\rangle \quad , \tag{4.93}$$

and the relative modular operator by

$$\Delta_{\Psi|\Phi} |\alpha, i\rangle = \frac{|d_{\alpha}|^2}{|c_i^2|} |\alpha, i\rangle \quad . \tag{4.94}$$

Some of the expressions just obtained are expressible in terms of density matrices. Let us assume in the following that Φ , Ψ are normalized so that $\sum_k |c_k|^2 = \sum_{\alpha} |d_{\alpha}|^2 = 1$.

The density operator for Ψ is

$$\rho_{12} = \sum_{k} \sum_{l} c_k c_l^* |k,k\rangle \langle l,l| \quad , \tag{4.95}$$

and for Φ it is given by

$$\sigma_{12} = \sum_{\alpha} \sum_{\beta} d_{\alpha} d_{\beta}^* |\alpha, \alpha\rangle \langle \beta, \beta| \quad , \tag{4.96}$$

where the 12 subscript indicates that these are density operators in $\mathcal{D}(\mathcal{H}_1 \otimes \mathcal{H}_2)$, i.e. density operators on the full Hilbert space. Performing a partial trace, it is straightforward to show (see equation (4.27)) that the reduced density operators are

$$\rho_{1} = \sum_{k} |c_{k}|^{2} |k\rangle \langle k| , \quad \rho_{2} = \sum_{k} |c_{k}|^{2} |k\rangle' \langle k|'$$
(4.97)

and

$$\sigma_1 = \sum_{\alpha} |d_{\alpha}|^2 |\alpha\rangle \langle \alpha| \ , \quad \sigma_2 = \sum_{\alpha} |d_{\alpha}|^2 |\alpha\rangle' \langle \alpha|' \ . \tag{4.98}$$

Comparing to equation (4.85) and equation (4.94) we see that

$$\Delta_{\Psi} = \rho_1 \otimes \rho_2^{-1} \quad \Delta_{\Psi|\Phi} = \sigma_1 \otimes \rho_2^{-1} , \qquad (4.99)$$

which is well defined since we proved in equation (4.27) that the ρ must be invertible.

The relative entropy in QFT was given in equation (4.58), and in terms of density operators we see that for the present case it is

$$D_{\Psi||\Phi} = -\langle \Psi | \log \left(\sigma_1 \otimes \rho_2^{-1} \right) | \Psi \rangle$$

= $\operatorname{Tr}_{12} \left[\rho_{12} \log \left(\sigma_1 \otimes \rho_2^{-1} \right) \right]$
= $\operatorname{Tr}_{12} \left[\rho_{12} \left(-\log(\sigma_1 \otimes \mathbf{1}) + \log(\mathbf{1} \otimes \rho_2) \right) \right]$
= $\operatorname{Tr}_1 \left[-\rho_1 \log(\sigma_1) \right] + \operatorname{Tr}_2 \left[\rho_2 \log(\rho_2) \right) \right].$ (4.100)

Since ρ_1 and ρ_2 have the same dimensions and eigenvalues, $\text{Tr}_2[\rho_2 \log(\rho_2)] = \text{Tr}_1[\rho_1 \log(\rho_1)]$, and we have

$$D_{\Psi||\Phi} = \operatorname{Tr}[\rho_1(\log(\rho) - \log(\sigma)], \qquad (4.101)$$

which is the definition of relative entropy in quantum information theory that we covered in section 2.3.6.

The Modular Conjugation Operator in Finite Dimension

In the previous section, an interesting fact is that since the reduced ρ_1 and ρ_2 have the same spectrum, they are dual under the exchange $|i\rangle \leftrightarrow |i\rangle'$. The same of course also goes for σ_1 and σ_2 in terms of the bases $|\alpha\rangle$, $|\alpha\rangle'$. We are free to pick the basis $|i\rangle$ relative to $|i\rangle'$ such that the phases of the c_i are minus the phases of the c_j . In this case, the antiunitary operator J_{Ψ} becomes a flip of basis vectors;

$$J_{\Psi}|i,j\rangle = \sqrt{\frac{c_j c_i}{c_j^* c_i^*}}|j,i\rangle = |j,i\rangle \quad . \tag{4.102}$$

The existence of a natural antiunitary operator J_{Ψ} of this character suggests that we can think of the state $|j,i\rangle$ as a matrix, and J_{ψ} as the Hermitian conjugation operator. Since \mathcal{H}_1 and \mathcal{H}_2 are of the same dimensions, we can identify the spaces as each other's duals, thinking of \mathcal{H}_1 as the space of column vectors and \mathcal{H}_2 as the space of row vectors.

We can then interpret an element in \mathcal{H} as an $n \times n$ matrix that maps from \mathcal{H}_1 to itself, with n the dimension of \mathcal{H}_1 . The inner product $\langle x|y\rangle$ on \mathcal{H} (= $\mathcal{H}_1 \otimes \mathcal{H}_2$) can then be interpreted as

$$\operatorname{Tr}_{\mathcal{H}_1}[x^{\dagger}y] , \qquad (4.103)$$
where y, x are matrices. Since x is daggered, what is happening is that we are contracting the row spaces (\mathcal{H}_2) of x and y in the matrix product, and then the trace performs the column space (\mathcal{H}_1) part of the inner product.

The action of operators $A \in \mathcal{A}$ act on this representation of \mathcal{H} as

$$x \to Ax\mathbf{1}^T = Ax , \qquad (4.104)$$

and operators A' in the commutant A' act as

$$x \to \mathbf{1}xA'^T = xA'^T \tag{4.105}$$

where A'^T is the transpose of the matrix A'. The identity in equation (4.104) comes from the identity piece of $A \otimes \mathbf{1}$ acting on $\mathcal{H}_1 \otimes \mathcal{H}_2$; the transposed matrix to the right acts on the row space (\mathcal{H}_2) , and the matrix on the right acts on the column space (\mathcal{H}_1) . Having interpreted the states in this way as matrices, Ψ becomes

$$\Psi = \rho^{1/2} , \qquad (4.106)$$

where we have exchanged $|i\rangle \otimes |i\rangle' \rightarrow |i\rangle \langle i|'$ in the definition of Ψ equation (4.77), and used that for c_i real positive $c_i = \sqrt{|c_i|^2}$.

We found that the relative modular operator was given by $\Delta_{\Psi|\Phi} = \sigma_1 \otimes \rho_2^{-1}$. Its action on a state in this new representation is

$$\Delta_{\Psi|\Phi} x = \sigma_1 x \rho_2^{-1T} , \qquad (4.107)$$

however, since \mathcal{H}_1 is dual to \mathcal{H}_2 we have $\rho_2^T = \rho_1$, and thus

$$\Delta_{\Psi|\Phi} x = \sigma_1 x \rho_1^{-1} . (4.108)$$

For later use, this implies that

$$\langle \Psi | \Delta_{\Psi|\Phi}^{\alpha} | \Psi \rangle = \operatorname{Tr}_{1} \left[\rho^{1/2} \Delta_{\Psi|\Phi}^{\alpha} \rho^{1/2} \right]$$

$$= \operatorname{Tr}_{1} \left[\sigma_{1}^{\alpha} \rho_{1} \rho_{1}^{-\alpha} \right]$$

$$= \operatorname{Tr}_{1} \left[\sigma_{1}^{\alpha} \rho_{1}^{1-\alpha} \right]$$

$$(4.109)$$

where we have made repeated use of the cyclicity of the trace and the Hermiticity of the density matrix.

A similarly simple relation between Φ and its density operators is not possible, since the splitting of \mathcal{H} relied on the existence of a natural antiunitary operator J_{Ψ} exchanging the bases of the subspaces. If we are only interested in σ_1 , we can act on Φ with a unitary element $u \in \mathcal{A}'$ without changing σ_1 . Once we specify the states in \mathcal{H} as matrices acting on \mathcal{H}_1 , the state Φ will correspond to some such matrix, that admits a polar decomposition $\Phi = PU$ where P is positive and U is unitary. Letting $P = \sigma_1^{1/2}$ we can transform away the unitary part by acting on the row space (\mathcal{H}_2) since states are only defined up to a unitary, letting us write $\Phi = \sigma^{1/2}$.

Modular Automorphism Group in Finite Dimensional Hilbert Space

Let us now state the main theorems of Tomita-Takesaki theory. We will use the representations of the Tomita and modular operators found in the finite-dimensional case to show these theorems in a simple setting, however most of them hold true in the infinite-dimensional case(s) as well. We will then sketch how these properties are generalized to infinite-dimensional Hilbert spaces. The interesting properties of the Tomita- and Modular operators regard the *Modular automorphism group*, which is the group of unitary transforms on the form Δ_{Ψ}^{is} , $s \in \mathbb{R}$. We already know by virtue of the previous section that in the finite-dimensional case $\Delta_{\Psi}^{is} = \rho_1^{is} \otimes \rho_2^{is}$. Using this explicit formula we see that for $A \in \mathcal{A}$

$$\Delta_{\Psi}^{is}(A \otimes \mathbf{1}) \Delta_{\Psi}^{-si} = \rho_1^{is} A \rho_1^{-is} \otimes \mathbf{1} .$$
(4.110)

The important thing to note here is that the RHS of equation (4.110) is on the form $B \otimes \mathbf{1}$ and is therefore in \mathcal{A} . What we have observed is that "conjugation" by the modular group maps \mathcal{A} to itself, and the same holds for \mathcal{A}' . Summarizing, we can write this as

$$\Delta_{\Psi}^{is} \mathcal{A} \Delta_{\Psi}^{-si} = \mathcal{A} , \quad \Delta_{\Psi}^{is} \mathcal{A}' \Delta_{\Psi}^{-si} = \mathcal{A}' .$$
(4.111)

In contrast, the modular conjugation operator exchanges the two algebras \mathcal{A} and \mathcal{A}' . We saw this in the previous section when we picked a basis such that all of the c_i in $\Psi = \sum_i c_i |i\rangle \otimes |i\rangle'$ were real positive, then $J_{\Psi} |i,j\rangle = |j,i\rangle$. A matrix $M \in \mathcal{A} = M \otimes \mathbf{1}$ can be written as its spectral decomposition

$$M \otimes \mathbf{1} = \sum_{i} \sum_{j} \lambda_{i}(|i\rangle \otimes |j\rangle)(\langle i| \otimes \langle j|) , \qquad (4.112)$$

where $|i\rangle$, $|j\rangle$ are orthonormal bases for \mathcal{H}_1 , \mathcal{H}_2 respectively. Now, having picked J_{Ψ} such that it flips j,i we see explicitly that

$$J_{\Psi}(M \otimes \mathbf{1}) J_{\Psi} = J_{\Psi} \left(\sum_{i} \sum_{j} \lambda_{i} (|i\rangle \otimes |j\rangle) (\langle i| \otimes \langle j|) \right) J_{\Psi}$$

$$= \sum_{i} \sum_{j} \lambda_{i}^{*} J_{\Psi} (|i\rangle \otimes |j\rangle) (\langle j| \otimes \langle i|)$$

$$= \sum_{i} \sum_{j} \lambda_{i}^{*} (|j\rangle \otimes |i\rangle) (\langle j| \otimes \langle i|)$$

$$= \sum_{j} \sum_{i} \lambda_{j}^{*} (|i\rangle \otimes |j\rangle) (\langle i| \otimes \langle j|)$$

$$= \mathbf{1} \otimes M^{*},$$

(4.113)

where the complex conjugation of the matrix eigenvalues is due to the anti- property of J_{Ψ} . Since \mathcal{A} and \mathcal{A}' are closed under complex conjugation, we summarize this property as

$$J_{\Psi}\mathcal{A}J_{\Psi} = \mathcal{A}' , \quad J_{\Psi}\mathcal{A}'J_{\Psi} = \mathcal{A} .$$
(4.114)

We will see that in QFT, J_{Ψ} is related to the CPT operator.

Finally, the *relative modular group* is, perhaps unsurprisingly, defined by the group of unitary transformations on the form $\Delta_{\Psi|\Phi}^{is}$, $s \in \mathbb{R}$. We saw in the previous section that in the finite dimensional case it can be represented by $\Delta_{\Psi|\Phi}^{is} = \sigma_1^{is} \otimes \rho_2^{-is}$, and for $(A \otimes \mathbf{1}) \in \mathcal{A}$ we have

$$\Delta^{is}_{\Psi|\Phi} A \Delta^{-is}_{\Psi|\Phi} = \sigma^{is}_1 A \sigma^{-is} \otimes \mathbf{1} .$$
(4.115)

It is clear that the relative modular group also maps \mathcal{A} to itself, and likewise for \mathcal{A}' . The relative modular conjugation does however have an additional interesting property - since no ρ 's appear it is completely independent of the cyclic separating vector Ψ . Thus, for any other cyclic separating vector Ψ' we have

$$\Delta^{is}_{\Psi|\Phi} A \Delta^{-is}_{\Psi|\Phi} = \Delta^{is}_{\Psi'|\Phi} A \Delta^{-is}_{\Psi'|\Phi} .$$
(4.116)

The properties of equations (4.111), (4.114) and (4.116) are regarded as the main theorems of *Tomita-Takesaki theory*. In the general case of infinite dimensional von Neumann algebra with cyclic separating vectors they are difficult to prove, although they remain true.

Extension to the Infinite-Dimensional Case

In this section we will provide a brief overview of the generalization to von Neumann algebras that are obtainable as a limit sequence of matrix algebras on a product space of an infinite number of qubits. This picture is thought to hold rigorously in QFT. We will then discuss some of the essential differences between finite-dimensional and infinite-dimensional Tomita-Takesaki theory, mainly discussing the analyticity properties of the modular operator, Δ_{Ψ}^{is} , $s \in \mathbb{C}$.

In the case where the infinite dimensional Hilbert space may be obtained as a limit $n \to \infty$ of matrix algebras, it is relatively straightforward to prove the Tomita-Takesaki theorems. It is believed that the algebra associated to an open spacetime region \mathcal{U} in QFT is of this type [32]. The main idea is that one can think of the degrees of freedom in \mathcal{U} as an infinite collection of qubits. Picking *n* of these qubits one gets an algebra \mathcal{M}_n of $2^n \times 2^n$ matrices acting on the product space $\mathcal{H} = \bigotimes_{i \leq n} \mathcal{H}_i$ where the \mathcal{H}_i have dimension 2. Beginning with one qubit and adding more one gets an ascending chain of algebras $\mathcal{M}_1 \subset \mathcal{M}_2 \subset \ldots \mathcal{M}_n \subset \ldots \subset \mathcal{A}_{\mathcal{U}}$ with $\mathcal{A}_{\mathcal{U}}$ as its limit. At each finite step in the chain one can define the modular operators $\Delta_{\Psi}^{(n)}$, $\Delta_{\Psi|\Phi}^{(n)}$ and the modular conjugation $J_{\Psi}^{(n)}$ as approximations of their full counterparts, and show explicitly that they fulfill equations (4.111), (4.114) and (4.116). One then shows that the sequence of operators converge in an appropriate sense to the full operators, and since they fulfill equations (4.111), (4.114) and (4.116) at each step they will also do so in the limit $n \to \infty$.

Depending on the details of the construction, one can end up with Type I, Type II, and Type III von Neumann algebras $\mathcal{A}_{\mathcal{U}}$. Type I algebras are obtained by considering a finite Hilbert space, in this case we have an algebra of bounded operators. An algebra of bounded operators on an infinite dimensional Hilbert space is said to be of type I_{∞} . The type II and III algebras are constructed by considering a vector space V where the elements are complex matrices $a, b \in V$. This is made into a Hilbert space by defining the inner product

$$\langle a, b \rangle = \operatorname{Tr}[a^{\dagger}b] . \tag{4.117}$$

A 2x2 complex matrix can be represented as a tensor product $W \otimes W'$ of two two-component complex vectors W and W'. In this sense, a and b are entangled qubit pairs. We then define the algebra \mathcal{M} and its commutant \mathcal{M}' as the algebras acting the row- and vector spaces of Vrespectively. That is, \mathcal{M} acts on W and \mathcal{M}' acts on W'. The type II algebra is then roughly obtained by tensoring n copies of V to make the Hilbert space, and then tensoring n copies of $\mathcal{M} \otimes \mathcal{M}'$ to make the von Neumann algebra. One obvious consequence as $n \to \infty$ is that since we have n maximally entangled qubits, the entanglement entropy in either factor $(W \otimes)^n$ or $(W' \otimes)^n$ is infinite. Type III algebras are obtained by allowing for non-maximal entanglement in the qubit pair. The entanglement entropy of a qubit pair can be parametrized by $\lambda \in [0,1]$. Generalizations the basic type III algebras are obtained by letting the entanglement in the qubit pair depend on n, i.e. $\lambda \to \lambda_n$. If $\lambda_n \to 1$ sufficiently fast, we recover the type I_{∞} algebra from before. If λ_n converges to 0 slowly, we get a new type of algebra called type III₀. If λ_n does not converge we obtain a Type III₁ algebra.

It can be shown [32] that the algebra constructed from factors $\mathcal{M}_{(n)} \otimes \mathcal{M}'_{(n)}$ such that λ_n does not converge as $n \to \infty$ has the properties of the local algebras of QFT. Thus the local algebras of QFT are of type III_1 , this is mostly related to the fact that the modular operators for such an algebras have as their spectrum the the positive real axis, and we will see that the modular operator in QFT is related to the the simultaneous momentum operator.

A fundamental fact about these algebras is that they have no irreducible representation. This is like the fact that the Reeh-Schlieder theorem tells us that the local algebra of an arbitrarily small spacetime region is dense in Hilbert space - there is no minimal region below which local operators stop being dense in Hilbert space.

The infinite-dimensional case of Tomita-Takesaki theory differentiates itself when we consider operators on the form Δ_{Ψ}^{iz} , where z is not restricted to the real axis. For a matrix algebra there is no issue, $\Delta_{\Psi}^{iz} = e^{iz \log \Delta_{\Psi}}$ has only bounded eigenvalues, meaning $e^{iz \log \Delta_{\Psi}}$ is an entire holomorphic function in z. In QFT Δ_{Ψ} may be unbounded when acting on some states, and more care has to be taken in the analysis.

Restricting to the case of Δ_{Ψ} acting on a vector $A |\Psi\rangle$, $A \in \mathcal{A}$ we can in a simple way prove some holomorphicity properties of the modular group. First, we can show that $\Delta^{1/2}A |\Psi\rangle$ has finite norm

$$\begin{split} \langle \Delta^{1/2} A \Psi | \Delta^{1/2} A \Psi \rangle &= \langle A \Psi | \Delta_{\Psi} | A \Psi \rangle \\ &= \langle A \Psi | S^{\dagger} S | A \Psi \rangle \\ &= \langle A^{\dagger} \Psi | A^{\dagger} \Psi \rangle \quad , \end{split}$$
(4.118)

where the final state is of finite norm since the eigenvalues of A^{\dagger} are conjugate to those of A. We can then (reminding ourselves that Δ_{Ψ} is Hermitian) use that for $0 \leq r \leq 1$, the inequality $\lambda^r < \lambda + 1$ implies the operator inequality $\Delta_{\Psi}^r < \Delta_{\Psi} + 1$. Starting from the previous calculation, we see that

$$\langle A\Psi | A\Psi \rangle + \langle A^{\dagger}\Psi | A^{\dagger}\Psi \rangle > \langle \Delta_{\Psi}^{r/2} A\Psi | \Delta_{\Psi}^{r/2} A\Psi \rangle \quad . \tag{4.119}$$

Since the lefthand side of the inequality is finite, and Δ_{Ψ}^{is} , $s \in \mathbb{R}$ is unitary, we know that $\Delta_{\Psi}^{is}\Delta_{\Psi}^{r}A|\Psi\rangle$ is bounded for all $s \in R$ and $r \in [0, 1/2]$. Thus, $\Delta^{iz}A|\Psi\rangle$ is continuous in the strip $0 \geq \operatorname{Im}(z) \geq -1/2$ and holomorphic in its interior. We learned in equation (4.48) that replacing \mathcal{A} with \mathcal{A}' replaces Δ_{Ψ} with Δ_{Ψ}^{-1} , so we see that $\Delta_{\Psi}^{iz}A'|\Psi\rangle$, $A' \in \mathcal{A}'$ is continuous in the strip $0 \leq \operatorname{Im}(z) \leq 1/2$.

The Modular Hamiltonian and Time-Ordering from Analytic Continuation

In this section we look at the holomorphicity of the function $F(z) = \langle \Psi | B \Delta_{\Psi}^{iz} A | \Psi \rangle$ where initially, $z \in \mathbb{R}$. We find that the function F(z) can be related to a "modular Hamiltonian", and corresponds to a two-point correlation function, where different operator orderings are obtained via analytic continuation. This analytic behaviour of correlation functions is the same as for the thermal field theories we considered in section 3.2.2. We then sketch the extension of this construction to higher order correlators.

We consider the function

$$F(z) = \langle \Psi | B \Delta_{\Psi}^{iz} A | \Psi \rangle \quad . \tag{4.120}$$

If z = s - ir, we have

$$F(s-ir) = \langle \Delta^{1/2r} B^{\dagger} \Psi | \Delta_{\Psi}^{is} A | \Delta^{1/2r} \Psi \rangle \quad (4.121)$$

and we know that the states $|\Delta^{1/2r}B^{\dagger}\Psi\rangle$ and $|\Delta^{1/2r}\Psi\rangle$ are of finite norm for $r \in [0, 1]$ from the previous section. Therefore, the function F(z) is continuous in the strip $0 \ge \text{Im}(z) \ge -1$ and holomorphic in the interior. Let us determine the appearance of F(z) on the lower end of the strip,

defined by z = s - i,

$$F(s-i) = \langle \Delta^{1/2} B^{\dagger} \Psi | \Delta_{\Psi}^{is} A | \Delta^{1/2} \Psi \rangle$$

$$\left\langle J_{\Psi} = J_{\Psi}^{-1}, J_{\Psi} \Delta^{1/2} = S_{\Psi} \right\rangle = \langle J_{\Psi} S_{\Psi} B^{\dagger} \Psi | \Delta_{\Psi}^{is} | J_{\Psi} S_{\Psi} A \Psi \rangle$$

$$\left\langle \text{antiunitarity of } J_{\Psi} \right\rangle = \langle B^{\Psi} | J_{\Psi} \Delta_{\Psi}^{is} J_{\Psi} | A^{\dagger} \Psi \rangle^{*}$$

$$\left\langle J_{\Psi} \Delta_{\Psi}^{is} J_{\Psi} = \Delta_{\Psi}^{is} \text{equation } (4.51) \right\rangle = \langle B^{\Psi} | \Delta_{\Psi}^{is} | A^{\dagger} \Psi \rangle^{*}$$

$$= \langle A^{\dagger} \Psi | \Delta_{\Psi}^{-is} | B \Psi \rangle$$

$$= \langle \Psi | A \Delta^{-is} B | \Psi \rangle ,$$
(4.122)

where what we have obtained is F(s), but with opposite operator ordering.

Let us illustrate what this means using the finite dimensional case. Let $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ and let \mathcal{A} be the algebra of matrices acting on \mathcal{H}_1 . We consider again the density matrix $\rho = |\Psi\rangle \langle \Psi|$ (coming from the cyclic separating state Ψ), and the reduced density matrix $\rho_1 = \text{Tr}_2[\rho]$. Then, we can define the *modular Hamiltonian* H according to $\rho_1 = e^{-H}$. In the definition of F(z) we can replace $\Delta_{\Psi}^{iz} A |\Psi\rangle$ by $\Delta_{\Psi}^{iz} A \Delta_{\Psi}^{-iz}$ since $\Delta_{\Psi} |\Psi\rangle = |\Psi\rangle$. We then have $\Delta_{\Psi}^{iz} A \Delta_{\Psi}^{-iz} |\Psi\rangle = \rho_1^{iz} A \rho_1^{-iz} |\Psi\rangle =$ $e^{-izH} A e^{izH} |\Psi\rangle$ (as in equation (4.110)). For any operator \mathcal{O} that acts only on \mathcal{H}_1 , the expectation value of the operator in the state Ψ is given by $\langle \mathcal{O} \rangle_{\Psi} = \text{Tr}_1[\rho_1 \mathcal{O}] = \text{Tr}_1[e^{-H} \mathcal{O}]$. Finally, we can write down

$$F(z) = \text{Tr}_1[e^{-H}Be^{-izH}Ae^{izH}]$$
(4.123)

Using our work in equation (4.122) we can now easily write down the expressions for F(s) and F(s-i) as

$$F(s) = \operatorname{Tr}_1[e^{-H}Be^{-isH}Ae^{isH}], \quad F(s-i) = \operatorname{Tr}_1[e^{-H}Ae^{-izH}Be^{izH}]$$
(4.124)

In the physical interpretation, s represents time, $\hat{A}(s) = e^{-isH}Ae^{isH}$ is a Heisenberg operator at time -s, and the two functions F(s) and F(s-i) are two-point correlation functions in a thermal ensemble with Hamiltonian H, as in section 3.2.2.

Let us now consider a slightly more general case of an infinite-dimensional factorizable Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$. Given the definition $\rho = e^{-H}$, if the trace in equation (4.123) is to be finite, the eigenvalues of ρ approach zero for some eigenstates, meaning the modular Hamiltonian H is necessarily unbounded above. In addition, the unit trace property and non-negativity of ρ implies that H is positive definite. In this case both iz and 1 - iz must have a non-negative real part to ensure the convergence of the trace, leading to the restriction $0 \ge \text{Im}(z) \ge -1$ that we observed without assuming a factorization earlier. Extending this reasoning in the factorizable case, we can easily generalize the two-point correlator F(z) to a three-point correlator (and more). Consider the function

$$F(z_1, z_2) = \operatorname{Tr}_1 \left[e^{-H} A e^{-iz_1 H} B e^{i(z_1 - z_2) H} C e^{iz_2 H} \right].$$
(4.125)

It is holomorphic when $\text{Im}(z_1) \leq 0$, $\text{Im}(z_1 - z_2) \geq 0$ and $1 + \text{Im}(z_2) \geq 0$, as this ensures that all exponents have negative real part, where the cyclicity of the trace pairs up the -H and iz_2H exponents. These holomorphicity statements can be proven without using the assumption of factorization, and we will see an example of this in the next section. We will not see any explicit example of the case with more than two operators.



Figure 4.2: A spacetime divided into two Rindler wedges \mathcal{U}_R , defined by $x \ge |t|$, and \mathcal{U}_L , defined by $x \le -|t|$. Each Rindler wedge is the domain of dependence of the regions $x \ge 0$ and $x \le 0$. It is readily seen that an exchange $x \to -x$, $t \to -t$ exchanges the two regions. This is why in this decomposition of spacetime, choosing the modular conjugation as $J_{\Psi} = \text{CRT}$ is a reasonable choice, where the purpose of the charge conjugation is to implement anti-unitarity.

4.2 Examples in QFT

In this section we describe the application of Tomita-Takesaki theory to some simple cases in QFT, namely the Rindler decomposition of Minkowski and the case of an accelerating observer in a Minkowski background. The results of these examples allow for a particularly elegant derivation of the semiclassical entropy of a black hole in section 4.2.3. These results also prove even more versatile, and we use them as the basis for a general class of results regarding the entanglement entropy of ball-shaped regions in CFT. As we will see, a sequence of conformal transformations may deform a ball into the Rindler geometry. This fact lets us study the entanglement entropy of conformal balls in section 4.183.

The General Setup

We consider a Minkowski spacetime M_D of general dimension D, with one timelike direction and D-1 spacelike. We further single out one space coordinate (x) and split the metric such that

$$ds^2 = -dt^2 + dx^2 + d\vec{y}^2 , \qquad (4.126)$$

where \vec{y} is a vector of D-2 spacelike coordinates.

We let the Cauchy surface Σ be defined by t = 0. We let \mathcal{V}_R be the open half-space in Σ defined by x > 0, and \mathcal{V}_L be the open half space defined by x < 0. We will call the *domain of dependence* of \mathcal{V}_R , \mathcal{U}_R , the right wedge. The right wedge \mathcal{U}_R is defined by |t| < x. Similarly, the left wedge \mathcal{U}_L is the domain of dependence of \mathcal{V}_L , and it is defined by $x \leq -|t|$. These two regions are often referred to as Rindler spaces. Finally, we let \mathcal{A}_R and \mathcal{A}_L denote the algebras of observables supported on \mathcal{U}_R and \mathcal{U}_L respectively. The left and right wedges are depicted in figure 4.2.

Now, let $|\Omega\rangle$ be the vacuum state of a QFT on M_D . Our goal is to use that the vacuum is cyclic separating for \mathcal{A}_R and \mathcal{A}_L to determine the modular operators Δ_{Ψ} and J_{Ψ} for observations in the right wedge \mathcal{U}_R .

CPT (charge conjugation, parity, time reversal) is a universal symmetry of QFT (in 4 spacetime dimensions). The inevitability of CPT invariance is built in if you assume that QFT can be obtained as the analytic continuation of a theory in euclidean signature³. This is because QFT is constructed to be rotation(Lorentz) invariant, and a rotation that inverts four axes has positive determinant, meaning it can be obtained by a sequence of infinitesimal rotations⁴. Analytic continuation back to Minkowski then converts this overall parity transform to CPT. This clearly fails to hold in odd dimensions, since a rotation with determinant -1 is not in the connected component of the rotation group. A better symmetry that holds in general dimension is obtained by replacing parity by reflection in one axis, that is, an operator $R : R(F(\vec{x})) = F(-\vec{x})$. In the following sections we will make use of this charge, reflection, time reversal (CRT) symmetry of QFT in general dimension.

The CRT operator is antiunitary⁵ and it can be chosen so that it exchanges the regions \mathcal{U}_R and \mathcal{U}_L in the setup we just described, thus exchanging the algebras \mathcal{A}_R and \mathcal{A}_L . We will see in the following example that it becomes natural to identify J_{Ω} with the CRT operator.

4.2.1 Path Integral Approach to Rindler Space

The path integral approach to the determination of Δ_{Ω} and J_{Ω} in right Rindler space is not mathematically rigorous, but will nonetheless lead to an interesting and correct determination of the operators. We assume the reader is familiar with everything that was discussed in section 3.1.

In this setup, we continue to Euclidean signature, setting $t = -i\tau$. Remembering that any state is evolved to the vacuum as we let τ go to infinity, the vacuum state Ω can then be obtained as the path integral over the half-space $\tau \leq 0$ as a function of some boundary values at $\tau = 0$. More explicitly, we saw in equation (3.65) that

$$|\Omega\rangle = \lim_{\tilde{T} \to \infty} \int \mathcal{D}\phi e^{-\int_0^{\tilde{T}} \mathrm{d}^{D-1}x \mathrm{d}\tau \mathcal{L}} |\phi_i\rangle \quad .$$
(4.127)

In the present context, we can think of $|\phi_i\rangle$ as some arbitrary initial state that is some function of the field operators ϕ_L , ϕ_R supported on the left and right wedges respectively.

Let us now pretend that we can factorize the full Hilbert space according to $\mathcal{H} = \mathcal{H}_R \otimes \mathcal{H}_L$, where \mathcal{H}_R and \mathcal{H}_L are the degrees of freedom supported in x > 0, x < 0 respectively. Then we know that \mathcal{A}_R acts only on \mathcal{H}_R and the same for the left wedge. We now wish to start with the pure density operator $\rho = |\Omega\rangle \langle \Omega|$ and trace over \mathcal{H}_L . To do this, we write the vacuum as a functional of the field operators ϕ_L, ϕ_R :

$$|\Omega\rangle \equiv |\Omega\rangle \left[\phi_L, \phi_R\right] \,. \tag{4.128}$$

The field operators can be seen as analogues of the basis vectors in the finite-dimensional case. Then, the partial trace on the density operator $\rho = |\Omega\rangle (\phi'_L, \phi'_R) \langle \Omega | (\phi_L, \phi_R)$ is performed by identifying $\phi'_L = \phi_L$ and integrating over ϕ_L (formally, the completion relation that defines the trace does not exist on the operator side, but thinking purely in terms of path integrals the gluing still makes sense):

$$\rho_R = \left(\int_{\phi_R}^{\phi_R'} \mathcal{D}\phi_L \left| \Omega \right\rangle \left(\phi_L, \phi_R' \right) \left\langle \Omega \right| \left(\phi_L, \phi_R \right) \right) . \tag{4.129}$$

where the boundary conditions ϕ_R, ϕ'_R are undetermined fields with support only on \mathcal{U}_R . The bra in the integrand can be computed, as noted above as an integral over $\tau < 0$, and the ket (since

 $^{^{3}}$ The rigorous proof of the CPT theorem depends on the holomorphicity property we found when proving the Reeh-Schlieder theorem.

⁴More formally, the inversion of all four axes lies in the connected component of the rotation group

⁵Specifically, it is the charge conjugation operator that is antiunitary.



Figure 4.3: The integration regions $W_{2\pi}$ and W_{θ} . The density operator is obtained from the path integral over $W_{2\pi}$ with boundary conditions ϕ_R and ϕ'_R . The density operator can be thought of as the limit $\theta \to 2\pi$ of the spacetime W_{θ} , that admits the operator interpretation of a rotation of the initial ϕ_R in the $\tau - x$ -plane by an angle θ . Formally, we are picking a new "Rindler time" θ and identifying the Hamiltonian time evolution with a rotation. This rotation analytically continues to Lorentzian signature to generate Lorentz boosts.

it is a complex conjugate) can be computed as an integral over $\tau > 0$. We take ϕ_R and ϕ'_R to be arbitrary states with some vacuum overlap, while also viewing them as the starting points for computing the vacuum state in the sense of equation (4.127). The local states that have overlap with the vacuum are dense in the vacuum sector of Hilbert space thanks to the separating property of $|\Omega\rangle$ for local algebras.

To set $\phi'_L = \phi_L$ we glue together the time integrals along x < 0. This gluing gives us a path integral over the two dimensional euclidean space $W_{2\pi}$ obtained from the $\tau - x$ -euclidean space by making a cut along x > 0. In this path integral we need to specify the boundary states $\langle \phi'_R |$ and $|\phi_R \rangle$ just above and just below the x > 0 half axis. We display the spacetime with the cut and boundary conditions in figure 4.3.

Looking at ρ_R as the path integral defined in equation (4.129) with two unspecified boundary conditions, it has the operator interpretation of evolving the state $|\phi_R\rangle$ by rotating it in the $x - \tau$ plane 2π radians onto the state $\langle \phi'_R |$. Let us consider only part of this operator, the path integral in a Euclidean wedge W_{θ} with opening angle θ , as displayed to the right in 4.3. The wedge operator \mathcal{O}_{θ} is a Euclidean rotation of the $x - \tau$ system by angle θ , i.e. the wedge operator acts according to

$$\mathcal{O}_{\theta} |\phi_R\rangle [\phi(x,\tau)] = |\phi_R\rangle [\phi(x',\tau')] , \quad \begin{pmatrix} \tau' \\ x' \end{pmatrix} = R_{\theta} \begin{pmatrix} \tau \\ x \end{pmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{pmatrix} \tau \\ x \end{pmatrix} . \tag{4.130}$$

Working in Lorentzian time $t = -i\tau$, we see that the rotation operator R_{θ} can be written as a Lorentz boost:

$$R_{\theta} \begin{pmatrix} t \\ x \end{pmatrix} = \begin{bmatrix} \cos \theta & i \sin \theta \\ i \sin \theta & \cos \theta \end{bmatrix} \begin{pmatrix} t \\ x \end{pmatrix} = \begin{bmatrix} \cosh(-i\theta) & -\sinh(-i\theta) \\ -\sinh(-i\theta) & \cosh(-i\theta) \end{bmatrix} \begin{pmatrix} t \\ x \end{pmatrix} .$$
(4.131)

The RHS is a Lorentz boost by a negative imaginary boost parameter of the t - x-plane. The generator of a Lorentz boost can be written as an integral over the initial value surface t = 0:

$$K = \int_{t=0} \mathrm{d}x \mathrm{d}^{D-2} y x T_{00}(x) \ . \tag{4.132}$$

The 00-component of the stress-energy tensor is precisely the energy density, which is the generator of local time evolution. By weighting the generator of time translation by the distance from the origin x, we obtain precisely a rotation of the entire x-axis in the t - x-plane. With this definition, for positive boost angles we are rotating the right wedge backwards in time, and the left wedge forwards in time as in figure 4.3, where the direction of the rotation comes from e^{-iHt} defining a positive time translation. Formally we can decompose the boost generator K according to

$$K = K_R - K_L \tag{4.133}$$

where K_R and K_L are the partial Lorentz boost generators on their respective wedges:

$$K_{R} = \int_{t=0, x \ge 0} dx d^{D-2} y \, x T_{00}(x, \vec{y}) ,$$

$$K_{L} = -\int_{t=0, x \ge 0} dx d^{D-2} y \, x T_{00}(-x, \vec{y}) ,$$
(4.134)

where the minus sign in front of K_L comes from the substitution $x \to -x$. The minus signs from the interval, dx and x in the integrand come out to an overall sign.

The unitary operator that implements a Lorentz boost by the real boost parameter η is given by $e^{-i\eta K}$. Setting $\eta = -i\theta$ we find that the path integral on the wedge W_{θ} must define the operator $e^{-\theta K}$, but K_L has only support on the on the left-half space so only K_R survives. Thus, setting $\theta = 2\pi$ to obtain our full path integral we have learned that

$$\langle \phi_R' | \rho_R | \phi_R \rangle = \langle \phi_R' | e^{-2\pi K_R} | \phi_R \rangle \tag{4.135}$$

for the right wedge. Similarly by instead tracing out \mathcal{U}_R , we obtain for \mathcal{U}_L that

$$\langle \phi_L' | \rho_L | \phi_L \rangle = \langle \phi_L' | e^{-2\pi K_L} | \phi_L \rangle \quad . \tag{4.136}$$

More specifically, it is clear the the density operators for the left and right wedges are

$$\rho_R = e^{-2\pi K_R}, \quad \rho_L = e^{-2\pi K_L}.$$
(4.137)

Combining the results for the left- and right wedges for the density operator we can determine the modular operator Δ_{Ω} . We know that (since the operators K_L, K_R commute)

$$\Delta_{\Omega} = \rho_R \otimes \rho_L^{-1} = e^{-2\pi K_R} e^{2\pi K_L} = e^{-2\pi K} .$$
(4.138)

While not very interesting in the present case, we find that the modular Hamiltonian H_R is given by

$$H_R = 2\pi \int_{t=0, x \ge 0} \mathrm{d}x \mathrm{d}^{D-2} y x T_{00}(x, \vec{y}) \ . \tag{4.139}$$

The fact that the modular Hamiltonian, which a priori has nothing to do with the Hamiltonian that determines the time evolution of the state, is a simple function of the field theory Hamiltonian is a remarkable property of Rindler space. This property will let us understand the entanglement properties of ball-shaped regions in CFTs, where the modular Hamiltonian is obtained by tracking how $T_{00}(x)$ transforms under a pair of conformal transformations from the Rindler wedge to the ball. In fact, this is the central result that lets us study spacetime emerging from holographic entanglement entropy in Part II.

Now let us consider a state obtained by acting on the vacuum with an operator in the local algebra $A |\Omega\rangle$, $A \in \mathcal{A}_R$. Let us assume that well defined operators A can be defined by smearing a local operator $\phi(x,\tau=0)$ such that the smeared operator has support only in \mathcal{U}_R . In general, the only point where this might be an issue is at x = 0, where this prohibits any smearing in time, opening up the possibility of ultraviolet divergence at x = 0 (which does indeed exist, and is related to the fact that we are not allowed to factorize the Hilbert space in the way that we have done).

Under our assumption that the Hilbert space factorizes, the state $A |\Omega\rangle$ can be obtained via a path integral on the lower $x - \tau$ half-plane by inserting the operator A into the path integration so that it is weighted by $Ae^{-iS[\phi]}$, on the x > 0 boundary. We now consider the state

$$\Delta^{\alpha} A \left| \Omega \right\rangle = e^{-2\pi\alpha K} A \left| \Omega \right\rangle = e^{-2\pi\alpha K_R} e^{2\pi\alpha K_L} A \left| \Omega \right\rangle \ . \tag{4.140}$$

We know that the operator A is in the right wedge algebra, so it must commute with $e^{2\pi\alpha K_L}$.

The picture is that the state $|\Omega\rangle$ is rotated counterclockwise in the $x - \tau$ plane by an angle $2\pi\alpha$, while we are still inserting the operator in the same spot. Rotating the $x - \tau$ plane while keeping the insertion point of the operator A has the effect of rotating the integration domain as well as the $\tau = 0$ boundary. Once $\alpha = \frac{1}{2}$, the path integral is defined over the upper half plane, and the boundary that used to lie in the left wedge now coincides with the insertion point of A. Changing coordinates by rotating, we see that the vacuum state should look the same, and that A has become $A[\phi[-x,0]] = \tilde{A}$ through the action of Δ_{Ω}^{α} . The operator \tilde{A} has support on what was originally the left wedge, so it must lie in \mathcal{A}_L . Thus, for $A \in \mathcal{A}_R$ we have

$$\Delta_{\Omega}^{1/2} A(x) \left| \Omega \right\rangle = \tilde{A}(-x) \left| \Omega \right\rangle \tag{4.141}$$

for some $\tilde{A} \in \mathcal{A}_L$. A similar statement will of course hold for the converse case. In this sense, the operator $\Delta^{1/2}$ seems to exchange the algebras \mathcal{A}_R and \mathcal{A}_L .

We cannot continue increasing α beyond the value of $\frac{1}{2}$ since it puts A outside of the integration region. In terms of operator formalism, this is like trying to determine the action of an operator by first acting on an arbitrary state with the operator A at some time $t = -t_0$, then defining initial conditions at a later time t = 0, which obviously makes no sense.

What we have learned is that the operator $\Delta_{\Omega}^{\alpha} |A\rangle \Omega$ is well defined in the closed interval $0 \leq \alpha \leq \frac{1}{2}$. In addition, there is no problem acting with a unitary operator $\Delta_{\Omega}^{is} A |\Omega\rangle$, so really the claim is that Δ_{Ω}^{iz} is holomorphic in the strip $0 > \text{Im}(z) > -\frac{1}{2}$ and continuous on the boundary, precisely the conditions we found in equation (4.119).

For our final exercise, we want to determine the modular conjugation J_{Ω} . We know that S_{Ω} is supposed to act as

$$S_{\Omega}A|\Omega\rangle = J_{\Omega}\Delta_{\Omega}^{1/2}A|\Omega\rangle = A^{\dagger}|\Omega\rangle \quad . \tag{4.142}$$

Suppose for simplicity that the local algebra is generated by a scalar field. Then, to determine J_{Ω} it is sufficient to consider the operator ϕ and $\frac{\partial \phi}{\partial t} = \dot{\phi}$. Since both ϕ and its time derivative are Hermitian, by definition

$$S_{\Omega}\phi |\Omega\rangle = \phi^{\dagger} |\Omega\rangle , \quad S_{\Omega}\dot{\phi} |\Omega\rangle = \dot{\phi}^{\dagger} |\Omega\rangle .$$
 (4.143)

The action of $\Delta_\Omega^{1/2}$ on general operators was found above, as

$$\Delta_{\Omega}^{1/2}\phi(0,x,\vec{y}) |\Omega\rangle = \phi(0,-x,\vec{y}) |\Omega\rangle ,$$

$$\Delta_{\Omega}^{1/2}\dot{\phi}(0,x,\vec{y}) |\Omega\rangle = -\dot{\phi}(0,-x,\vec{y}) |\Omega\rangle ,$$
(4.144)

where the additional minus sign on the time derivative comes from the fact that the action of $\Delta^{1/2}$ rotates the time direction by π radians as well. This determines the action of J_{Ω} :

$$J_{\Omega}\phi(0, -x, \vec{y}) |\Omega\rangle = \phi^{\dagger}(0, x, \vec{y}) |\Omega\rangle$$

$$-J_{\Omega}\dot{\phi}(0, -x, \vec{y}) |\Omega\rangle = \dot{\phi}^{\dagger}(0, x, \vec{y}) |\Omega\rangle \quad .$$
(4.145)

Inserting a $J_{\Omega}^2 = 1$ and using that $J_{\Omega} |\Omega\rangle = |\Omega\rangle$ this implies, when comparing equation (4.143) and equation (4.145) that

$$J_{\Omega}\phi(0, -x, \vec{y})J_{\Omega} = \phi^{\dagger}(0, x, \vec{y}) , \quad -J_{\Omega}\dot{\phi}(0, -x, \vec{y})J_{\Omega} = \dot{\phi}^{\dagger}(0, x, \vec{y})$$
(4.146)

We see that J_{Ω} must be an antiunitary operator (since it is supposed to conjugate the field), that maps $x \to -x$, $t \to -t$ and $\vec{y} \to \vec{y}$. In other words,

$$J_{\Psi} = \operatorname{CRT} \,, \tag{4.147}$$

the charge conjugation⁶-reflection-time reversal operator.

In this simple example of the Rindler spacetime, have explicitly verified that the operators Δ_{Ω} and J_{Ψ} fulfill the Tomita-Takesaki theorems (equations (4.111) and (4.114)). More explicitly, since Δ_{Ω}^{iz} , $z \in \mathbb{R}$ generates Lorentz boosts by a real parameter, it maps the left and right wedges to themselves and therefore also preserves the algebras $\mathcal{A}_R, \mathcal{A}_L$, which is precisely equation (4.111). The CRT operator J_{Ω} exchanges the left- and right wedges, exchanging $\mathcal{A}_R, \mathcal{A}_L$ which is precisely equation (4.114). The fact that J_{Ω} precisely exchanges the two algebras $\mathcal{A}_R, \mathcal{A}_L$ tells us that they are commutants, meaning they are maximal under the condition of commuting. Explicitly showing these properties of J_{Ω} and Δ_{Ω} is how Haag duality equation (4.32) was originally proven for complementary Rindler wedges by Bisogano and Wichmann [37].

4.2.2 Bisogano-Wichmann Approach to the Rindler Wedges

It would be nice to repeat the success of the above result without treating the boundary x = 0imprecisely, and without claiming a false factorization of Hilbert space. Since $J_{\Omega} = \text{CRT}$ definitely exchanges the two wedges as in equation (4.146) we do not need to modify our reasoning on that point. This means we only need to justify that for $A \in \mathcal{A}_R$ we have

$$\Delta_{\Omega}^{1/2}\tilde{A}(x)\left|\Omega\right\rangle = A(-x)\left|\Omega\right\rangle \tag{4.148}$$

for some $\tilde{A} \in \mathcal{A}_L$. Then, must show that Δ_{Ω}^{is} maps \mathcal{A}_R to itself while also fulfilling equation (4.148). What we need to show is

- identifying $\Delta_{\Omega} = e^{-2\pi K}$, where K is the simultaneous Lorentz boost generator given by equation (4.132) makes the automorphism group $\Delta_{\Omega}^{is}, s \in \mathbb{R}$ map the entire local algebra of operators supported in the right wedge to itself,
- analytically continuing the modular automorphism group to complex arguments, $\Delta_{\Omega}^{iz}, z \in \mathbb{C}$ gives us a domain of holomorphicity 0 > Im(z) > -1/2 when acting on states of the form $A |\Omega\rangle$ for all $A \in \mathcal{A}_R$,

⁶Remember that the Hermitian conjugate of the field has opposite charge. More explicitly one could start with a particular gauge theory, find the conserved charge and confirm that it is odd under J_{Ω} .

showing these two properties ensures that J_{Ω} and Δ_{Ω} fulfill equation (4.114) and equation (4.111) respectively (at the same time). Let us define the operator $A \in \mathcal{A}_R$ according to

$$A = \phi(t_1, x_1.\vec{y_1})\phi(t_2, x_2.\vec{y_2})\dots\phi(t_n, x_n.\vec{y_n}) , \qquad (4.149)$$

represented in the path integral by insertion in the right wedge \mathcal{U}_R at points $p_i = (t_i, x_i, \vec{y}_i), i = 1, 2...n$. In addition to this we can take the points p_i to be spacelike separated from each other. Since this makes the field operators commute, we can then order them such that

$$x_j - x_i > |t_j - t_i|, \ j > i$$
 . (4.150)

It is enough to consider operators on this form because the Reeh-Schlieder theorem still holds under the spacelike separation restriction⁷.

The statement in equation (4.150) is Lorentz invariant (in the t-x-plane), since a Lorentz boost can only map spacelike vectors to other spacelike vectors. This ensures that Δ_{Ω}^{is} precisely maps \mathcal{A}_R to itself, since for real *s* the operator Δ_{Ω}^{is} is just a Lorentz boost by a real boost parameter. The action of Δ_{Ω}^{is} is defined by

$$\Delta_{\Omega}^{is} |\Psi\rangle [t, x] \to |\Psi\rangle [t', x'] , \qquad (4.151)$$

where

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = \begin{bmatrix} \cosh 2\pi s & -\sinh 2\pi s \\ -\sinh 2\pi s & \cosh 2\pi s \end{bmatrix} \begin{pmatrix} t \\ x \end{pmatrix} .$$
(4.152)

Then, remembering that $K |\Omega\rangle = 0$, we insert a factor $e^{-2i\pi Ks}e^{2i\pi Ks}$ between between the operator and vacuum, finding that the Heisenberg picture operator under this transformation is

$$\phi(\mathbf{x}'(s)) = e^{2i\pi Ks} \phi(\mathbf{x}) e^{-2i\pi Ks} , \qquad (4.153)$$

where we have defined $\mathbf{x}(s) = (t', x')$, $\mathbf{x} = (t, x)$ as in equation (4.152). Thus, we find that

$$e^{2i\pi Ks}\phi(\mathbf{x}_1,\vec{y_1})\phi(\mathbf{x}_2,\vec{y_2})\dots\phi(\mathbf{x}'_n,\vec{y_n})|\Omega\rangle = \phi(\mathbf{x}'_1(s),\vec{y_1})\phi(\mathbf{x}'_2(s),\vec{y_2})\dots\phi(\mathbf{x}'_n(s),\vec{y_n})|\Omega\rangle$$

$$\equiv F(s) , \qquad (4.154)$$

where we have defined the function F(s) for convenience. We would like to analytically continue to the complex domain, defining $F(z) \in \mathbb{C}$. First we wish to show that the domain of holomorphy for F(z) is defined the following way. Let $\mathbf{x}'_i = u_i + iv_i$, then F(z) is holomorphic if v_1 and $v_{i+1} - v_i$ are future timelike for all v_i . Using the spacetime translation generator P and defining $X_i = (\mathbf{x}'_1(z), \vec{y_1})$ (suppressing the explicit z dependence) we may write

$$F(z) = \phi(\mathbf{x}'_{1}(z), \vec{y_{1}})\phi(\mathbf{x}'_{2}(z), \vec{y_{2}}) \dots \phi(\mathbf{x}'_{n}(z), \vec{y_{n}}) |\Omega\rangle$$

= $e^{iP_{1}\cdot X_{1}}\phi(0)e^{iP_{2}\cdot (X_{2}-X_{1})}\phi(0)e^{iP_{2}\cdot (X_{3}-X_{2})} \dots e^{iP_{n}\cdot (X_{n}X_{n-1})}\phi(0)e^{-iP_{n}\cdot X_{n}} |\Omega\rangle$, (4.155)

where the \cdot denotes the Lorentz inner product $P_i X_i = -tH + \vec{x} \cdot \vec{p}$. The vacuum should be invariant under translation, so the last factor $e^{-iP_n \cdot X_n}$ does not matter. For a general vector $\mathbf{x}'_i = u_i + iv_i$ the generic exponents look like (suppressing the uninteresting \vec{y} direction since it is has no imaginary part):

$$iP \cdot (X_{i+1} - X_i) = -itH(u_{i+1}^t - u_i^t) + iP_x(u_{i+1}^t - u_i^t) + tH(v_{i+1}^t - v_i^t) - P_x(v_{i+1}^t - v_i^t) . \quad (4.156)$$

⁷The Reeh-Schlieder theorem tells us we can translate the field operators in spacelike directions, so we could use the procedure of the proof of the Reeh-Schlieder theorem to put any set of field operators A on the form equation (4.149) before successively moving them out of $\mathcal{U}_{\mathcal{V}}$.

But we know that e^R where $\operatorname{Re}(R < 0)$ is holomorphic, so really we want

$$\operatorname{Re}[iP \cdot (X_{i+1} - X_i)] = H(v_{i+1}^t - v_i^t) - P_x(v_{i+1}^t - v_i^t) \le 0 , \qquad (4.157)$$

which is precisely the condition that v_1 and $v_{i+1} - v_i$ are future timelike, based on what exponents appear in equation (4.155). To see this, remember that the simultaneous momentum of a physical field, P lies in the closed forward lightcone as defined by the Wightman axioms. We then realize that the product of two closed forward timelike vectors must be positive. The object in equation (4.157) is minus the Lorentz product of P and $v_{i+1} - v_i$, so the condition implies that since P is timelike, so is $v_{i+1} - v_i$.

Now, continuing on to F(z), we claim that if the points \mathbf{x}_i are chosen according to equation (4.150) then F(z) is holomorphic precisely when 0 > Im(z) > -1/2. Since the statement of equation (4.150) is Lorentz invariant, it suffices to check that the statement holds for Re(z) = 0. Even further, because of how we have chosen the points \mathbf{x}_i we only need to check the condition for i = 1, since all $\mathbf{x}_{i+1} - \mathbf{x}$ are in the wedge x > |t|. Inserting $i\alpha$ as the boost parameter we find

$$\mathbf{x}' = \begin{pmatrix} t'\\ x' \end{pmatrix} = \begin{bmatrix} \cosh 2i\pi\alpha & -\sinh 2i\pi\alpha\\ -\sinh 2i\pi\alpha & \cosh 2i\pi\alpha \end{bmatrix} \begin{pmatrix} t\\ x \end{pmatrix} = \begin{pmatrix} \cos 2\pi\alpha t - ix\sin 2\pi\alpha\\ \cos 2\pi\alpha x - it\cos 2\pi\alpha \end{pmatrix} .$$
(4.158)

Since x > |t| the complex part of \mathbf{x}' is future timelike as long as $-x \sin 2\pi\alpha > -t \sin 2\pi\alpha$, which is true for $\alpha \in (-1/2,0)$, since in this region $-\sin(2\pi\alpha)$ is positive. With this, we have proved that

- identifying $\Delta_{\Omega} = e^{-2\pi K}$ makes the automorphism group $\Delta_{\Omega}^{is}, s \in \mathbb{R}$ map the entire local algebra of operators supported in the right wedge to itself (this was obvious since Lorentz boosts map spacelike to spacelike, and cannot exchange the wedges),
- $\Delta_{\Omega}^{iz}, z \in \mathbb{C}$ gives us a domain of holomorphicity 0 > Im(z) > -1/2 when acting on states of the form $A |\Omega\rangle$ for all $A \in \mathcal{A}_R$,

which is what we set out to show.

Accelerating Observer

We have seen that if we split Minkowski into two complementary wedges, and consider an observer who only has access to the right wedge, the observer will find that their part of spacetime is described by a thermal density operator $\rho_R = e^{-2\pi K}$. The observer who has access precisely to a Rindler wedge of Minkowski spacetime is the constantly accelerating observer, and the expectation value of operators \mathcal{O} measured by this observer are therefore given by $\langle \mathcal{O} \rangle = \text{Tr}[\mathcal{O}e^{-2\pi K}]$. The world line of the constantly accelerating observer is parametrized by

$$\begin{pmatrix} t(\tau) \\ x(\tau) \end{pmatrix} = R \begin{pmatrix} \sinh(\frac{\tau}{R}) \\ \cosh(\frac{\tau}{R}) \end{pmatrix} , \qquad (4.159)$$

where τ is the proper time of the observer, and the proper acceleration is a = 1/R. As before, we suppress the \vec{y} direction since nothing interesting happens. For simplicity, we abbreviate $\begin{pmatrix} t(\tau) \\ x(\tau) \end{pmatrix} = \mathbf{x}(\tau)$.

We imagine that the observer probes the Minkowski vacuum Ω by measuring a local operator \mathcal{O} and its adjoint \mathcal{O}^{\dagger} along the worldline $\mathbf{x}(\tau)$. For simplicity, we consider only the two-point functions, $\langle \Omega | \mathcal{O}(\mathbf{x}(\tau_1)) \mathcal{O}^{\dagger}(\mathbf{x}(\tau_2)) | \Omega \rangle$ and $\langle \Omega | \mathcal{O}^{\dagger}(\mathbf{x}(\tau_2)) \mathcal{O}(\mathbf{x}(\tau_1)) | \Omega \rangle$. Due to Lorentz invariance, we



Figure 4.4: Rindler spacetime with sample trajectories of a constantly accelerating observer. The right hand trajectory is the physical trajectory, while the downwards-moving trajectory on the left will be obtained as an analytic continuation of the other. An observer on the physical trajectory has an event horizon on the region rightwards lightlike separated from the origin, since timelike geodesics can never pass this surface from the left.

know this correlator depends only on $\tau_1 - \tau_2$ so without loss of generality we can set $\tau_1 = \tau$ and $\tau_2 = 0$. We now write these as two functions

$$F(\tau) = \langle \Omega | \mathcal{O}(\mathbf{x}(\tau)) \mathcal{O}^{\dagger}(\mathbf{x}(0) | \Omega \rangle$$

$$G(\tau) = \langle \Omega | \mathcal{O}^{\dagger}(\mathbf{x}(0)) \mathcal{O}(\mathbf{x}(\tau)) | \Omega \rangle .$$
(4.160)

The insight of Unruh was that these two functions have the analytic properties of a thermal correlator. The basic properties of thermal correlators is, as we covered in section 3.2.2 are that analyticity breaks down if any two arguments (in this case τ_2 and τ_1) are separated by more than β . Slightly more mathematically, there is a function $H(\tau)$ that is holomorphic on a strip in the complex plane whose boundary values are $F(\tau)$ and $G(\tau)$ on the two boundaries of the strip that are canonically given by $\text{Im}(\tau) = 0$ and $\text{Im}(\tau) = i\beta$.

To have the same insight as Unruh, we first analytically continue the observer's trajectory. We set $\tau/R = s + i\theta$, with s, θ real, and find that

$$\mathbf{x}(\tau) = R \begin{pmatrix} \sinh(s\cos\theta) + i\cosh(s\sin\theta)\\ \cosh(s\cos\theta) + i\sinh(s\sin\theta) \end{pmatrix} .$$
(4.161)

Then, we have for the imaginary part that

$$\operatorname{Im}(\mathbf{x}(\tau)) = R \sin \theta \begin{pmatrix} \cosh(s) \\ \sinh(s) \end{pmatrix} . \tag{4.162}$$

As we arrived at in equation (4.157), $F(\tau)$ is holomorphic when $\text{Im}(\mathbf{x}(\tau))$ is future timelike and $G(\tau)$ when $\text{Im}(\mathbf{x}(\tau))$ is past timelike. The vector $(\cosh s, \sinh s)$ is always future timelike, so $F(\tau)$ is holomorphic in the strip $0 < \theta < \pi$ and $G(\tau)$ in the strip $\pi < \theta < 2\pi^8$, entirely due to the sign

⁸Equivalently, we could choose $-\pi < \theta < 0$ but such a choice is inconvenient later.

of $\sin \theta$. The function $F(\tau)$ is equal to the original correlator at $\theta = 0$, and at the boundary $\theta = \pi$ it is equal to

$$F(R(s+i\pi)) = \langle \Omega | \mathcal{O}(-\mathbf{x}(Rs)) \mathcal{O}^{\dagger}(\mathbf{x}(0)) | \Omega \rangle \quad .$$
(4.163)

For $G(\tau)$ the situation is similar; for $\theta = 2\pi$ it is clearly just the original correlator. Then, at $\theta = \pi$ it is given by

$$G(R(s+i\pi)) = \langle \Omega | \mathcal{O}^{\dagger}(\mathbf{x}(0)) \mathcal{O}(-\mathbf{x}(Rs)) | \Omega \rangle \quad .$$
(4.164)

The crucial observation is now that $\mathbf{x}(0)$ and $-\mathbf{x}(Rs)$) are always spacelike separated, since $-\mathbf{x}(Rs)$) lies in the left wedge of figure 4.4. Therefore, the operators commute and we find that

$$F(R(s+i\pi)) = G(R(s+i\pi)) .$$
(4.165)

We have thus found two functions: $F(\tau)$ that is holomorphic for $0 < \theta < \pi$ and $G(\tau)$ that is holomorphic for $\pi < \theta < 2\pi$. In addition, at $\theta = \pi$ they are equal. We can therefore define a single function on the combined strip $0 < \theta < 2\pi$ according to

$$H(\tau) = \begin{cases} F(\tau) \text{ if } \pi R \ge \operatorname{Im}(\tau) \ge 0\\ G(\tau) \text{ if } 2\pi R \ge \operatorname{Im}(\tau) \ge \pi R \end{cases}$$
(4.166)

The function $H(\tau)$ is holomorphic in the combined strip and continuous on the boundaries, granted that we can show that it must be holomorphic at $\text{Im}(\tau) = \pi R$. To show holomorphicity at $\text{Im}(\tau) = \pi R$ we realize that we in the strip $0 < \theta < \pi$ can define $H(\tau)$ as a Cauchy integral with a boundary that runs partly along $\theta = \pi$ (since it is holomorphic inside the strip). In the strip $\pi < \theta < 2\pi$ we can do the same, letting part of the contour run along $\theta = \pi$ in the opposite direction. We then pick the two curves so that the part along $\theta = \pi$ cancels, using that the sum of two holomorphic functions is another holomorphic function, we find that H must be holomorphic on the line $\theta = \pi$.

Conclusively, $H(\tau)$ is holomorphic on a strip of width $\beta = 2\pi R$ and its limits at the boundaries are given by the two operator orderings F(s), G(s) of the correlation function. The constantly accelerating observer has proper acceleration $a = \frac{1}{R}$, and finds that the vacuum has temperature

$$T = \frac{1}{2\pi R} = \frac{a}{2\pi} , \qquad (4.167)$$

which indeed is the celebrated expression for the Unruh temperature.

4.2.3 Semiclassical Black Hole Entropy

We found in equation (4.167) that proper acceleration is related to an observed thermal state in Minkowski spacetime. Finding this result was a long trip through the foundations of axiomatic QFT, but the result we have found is remarkably powerful thanks to the equivalence principle of general relativity. In fact, it is now very straightforward to obtain the entropy of a black hole, the famous result of Bekenstein and Hawking [1, 28].

Let us now turn our eyes to a QFT in a curved spacetime. Consider an observer that is fixed *very close* to the horizon of a black hole. Invoking the equivalence principle, there is a locally flat (Minkowski) spacetime that our observer sees if it is moving along a geodesic. The fact that our observer is fixed implies that it must be counteracting the acceleration of the black hole with its own proper acceleration, i.e. we have an accelerating observer in a locally Minkowski spacetime. We know that the temperature observed by the accelerating observer is

$$T = \frac{a}{2\pi} , \qquad (4.168)$$

where a is the acceleration. Furthermore, we know that the acceleration at the event horizon of a black hole is given by Newtons law as

$$a = \frac{G_{\rm N} M_{\rm BH}}{R_{\rm BH}^2}$$
 (4.169)

We also know that the Schwarzschild radius is given (in four dimensions) by

$$R_{\rm BH} = 2G_{\rm N}M_{\rm BH} \ . \tag{4.170}$$

Therefore, the temperature of the black hole is given by

$$T = \frac{1}{8\pi G_{\rm N} M_{\rm BH}} . \tag{4.171}$$

Notably, very massive black holes are cold. Since we know the total energy of the black hole, as well as its temperature, we can interpret the black hole as a microcanonical ensemble to calculate its entropy. Remember that for the microcanonical ensemble, we have from equation (3.74) that

$$\frac{1}{T} = \frac{\mathrm{d}S}{\mathrm{d}E} \ . \tag{4.172}$$

The total energy of the black hole is simply $M_{BH}(c^2)$ since it is completely static. Substituting $dE \rightarrow dM$ and inserting our expression for T, we find the integral relation

$$\int_{0}^{S(M=M_{\rm BH})} \mathrm{d}S = \int_{0}^{M_{\rm BH}} \mathrm{d}M8\pi G_{\rm N}M \ . \tag{4.173}$$

Where we have required that the entropy of the black hole goes to zero as the mass goes to zero, since there should just be a Minkowski vacuum in the zero mass limit. We find by performing the integration that

$$S_{\rm BH} = 4\pi G_{\rm N} M^2$$
 (4.174)

Expressing the area of the black hole horizon as $A = 4\pi (2G_{\rm N}M)^2$ we find

$$S_{\rm BH} = \frac{A}{4G_{\rm N}} \ . \tag{4.175}$$

This computation is a so-called *semiclassical* approximation of black hole physics. It is not classical since we are in principle considering particle anti-particle pairs that would usually be Feynman bubble diagrams, which are one loop. It is not a full quantum derivation since we do not consider quantum gravitational corrections, and we are not considering the back-action of the particle-antiparticle pair on the background spacetime.

An incredibly important property of the black hole entropy is that it is completely independent of the particle content of our effective field theory. It is for this reason that the semiclassical physics of black holes are thought to hold for all consistent extensions of the standard model. This means that reproducing the leading area term of the black hole entropy is the first check that any theory of quantum gravity must pass.

Both string theory and loop quantum gravity (LQG) can manage to find the area law for the entropy by counting microstates explicitly. Perturbative string theory comes with the caveat that it is difficult to find Schwarzschild-like black hole solutions in 3+1 large dimensions, but for example 5-dimensional calculations give a quantitative agreement with the semiclassical result. The case in LQG is that there is a free parameter called the Barbero-Immirzi parameter that is fixed only

by the black hole entropy computation, and there is a lack of other successful computations with which the consistency of the black hole value of the Barbero-Immirzi parameter can be checked.

Finally, the black hole entropy result is the first hint at an apparent paradox. Let us imagine that we have a quantum theory of gravity, wherein time evolution is described by a unitary operator. It is well known that the von Neumann entropy of a system is invariant under unitary transformations. We can start with a universe that is nearly in the vacuum state, containing only two very energetic particles that are sent towards each other. This is a completely definite state with zero entropy. When these particles collide they should form a black hole. The black hole has nonzero entropy meaning that something non-unitary has happened, violating a central postulate of quantum mechanics.

This might be okay, maybe the entropy and non-unitarity only exists because we are not describing the interior of the black hole. The result in equation (4.168) tells us that the black hole has temperature and should radiate its content back into the universe as thermal radiation. This means that as some point in the future we will end up with a universe containing no black holes and lots of thermal radiation. Thermal radiation has entropy, and now we can not use the excuse of not knowing what happens inside the black hole. This leaves us with a paradox, black hole entropy seems to violate the unitarity of quantum mechanics. Since entropy is a measure of information, this is called the *information paradox* and it is one of the important open questions of quantum gravity.

A notable ambiguity in the present discussion of black holes is that we have assumed an event horizon to apply the Unruh effect. If the black hole radiates until it disappears, this indicates that on the quantum level there was no event horizon to prevent matter in the interior from escaping. The way in which the analysis of black hole physics has proceeded has then been to try and assume that the Hawking radiation is secretly pure, by virtue of photons emitted early in the lifespan of the black hole being entangled with those emitted late. This assumption, which saves quantum mechanics, will force us to give up on either the smoothness⁹ of the black hole horizon or the locality of physics as we will learn in chapter 9.

4.2.4 Ball-Shaped Regions in CFT

The Unruh-Wichmann result is both incredibly general and incredibly useful. Let us now see how it can be applied to find the modular group for ball-shaped regions in a CFT. This section is based on an example in chapter 6.1 of [38] that in turn follows [39].

Let us denote the ball-shaped spatial region B_d , and the d + 1-dimensional Minkowski background M^{d+1} . We furthermore choose coordinates such that the ball is centered at the origin at time t = 0 and has radius R. The domain of dependence (or *causal diamond*) D_B of the ball is a double cone, with the two endpoints $p^{\pm} = (\pm R, 0, 0, 0)$. In CFTs conformal transformations are a symmetry, and such transformations may deform D_B . We will now find the modular group for the ball-shaped regions by using two conformal transformations to turn D_B into a Rindler wedge. We know that the modular Hamiltonian for a Rindler wedge is given by (equation (4.139))

$$H_R = 2\pi K_R = 2\pi \int_{t=0,x>0} \mathrm{d}x \mathrm{d}^{D-2} y \, x T_{00}(x, \vec{y}) \,, \qquad (4.176)$$

where the local density operator is $\rho_R = e^{H_R}$. We can then construct the local density operator ρ_B by tracking how T transforms as we invert the conformal transformations to go back from the Rindler wedge to D_B . The sequence of geometries is given in figure 4.5.

 $^{^{9}}$ In this case, by 'smooth' we mean that the black hole horizon is not distinguished by curvature, i.e. the geometry is nonsingular.



Figure 4.5: The sequence of geometries for the CFT ball modular Hamiltonian computation. The domain of dependence D_B of the ball B_d in polar coordinates is mapped by a conformal transformation to a hyperbolic space tensored with a timelike coordinate τ . The region filled in with red lines keeps track of the transformation of B_d and the region in grey follows the transformed D_B . This space is then mapped to the Rindler half-space by another conformal transformation, in which B_d is mapped to the slice $z \ge 0$, $X^0 = 0$. The naming of the coordinates in each step follows the conventions of equations (4.177), (4.179), (4.180) and (4.181).

Let us begin the sequence of coordinate transformations by looking at polar coordinates in which the ball takes a simple form

$$ds^{2} = -dt^{2} + dr^{2} + r^{2}d\Omega_{d-2}^{2}. \qquad (4.177)$$

We can make a coordinate transformation

$$t = R \frac{\sinh\left(\frac{\tau}{R}\right)}{\cosh u + \cosh\left(\frac{\tau}{R}\right)} , \quad r = R \frac{\sinh u}{\cosh u + \cosh\left(\frac{\tau}{R}\right)} , \quad (4.178)$$

putting the metric on the form

$$\mathrm{d}s^2 = \frac{1}{\left[\cosh u + \cosh\left(\frac{\tau}{R}\right)\right]^2} \left(-\mathrm{d}\tau + R^2 \left(\mathrm{d}u^2 + \sinh^2 u \mathrm{d}\Omega_{d-2}^2\right)\right) \ . \tag{4.179}$$

Up to an overall scale factor this metric describes a hyperbolic space ("Euclidean AdS") tensored with the time direction. Metrics related by an overall scale factor are conformally equivalent. By observing that as $\tau \to \pm \infty$, $t \to \pm R$, while as $u \to \infty$, $r \to R$, the causal wedge has been mapped to $\tau \in \mathbb{R}$ and $u \ge 0$, i.e our coordinates precisely cover D_B .

The metric equation (4.179) can be related to the Rindler spacetime by rewriting the hyperbolic part in half-plane coordinates (see equation (6.28)). The metric becomes

$$ds^{2} = \frac{1}{\left[\cosh u + \cosh\left(\frac{\tau}{R}\right)\right]^{2}} \left(-d\tau^{2} + \frac{dz^{2} + \sum_{i=2}^{d-1} dX^{i} dX^{i}}{z^{2}} \right)$$

$$\equiv \frac{1}{z^{2} \left[\cosh u + \cosh\left(\frac{\tau}{R}\right)\right]^{2}} \left(-z^{2} d\tau^{2} + dz^{2} + \sum_{i=2}^{d-1} dX^{i} dX^{i} \right) , \qquad (4.180)$$

where in the final term, the expression in parentheses is the metric of a Rindler observer, and the factor outside parentheses can be removed by a conformal scaling. The Rindler metric can be brought to Cartesian coordinates via

$$X^1 \pm X^0 = z e^{\pm \frac{i}{R}} . ag{4.181}$$

The entangling surface at r = R in the original coordinates has been mapped to z = 0, which is the entangling surface of the Rindler spacetime. We know from section 4.2.2 that the modular Hamiltonian for the Rindler wedge is

$$H_R = 2\pi K_R = 2\pi \int_{t=0,x>0} \mathrm{d}x \mathrm{d}^{D-2} y \, x T_{00}(x, \vec{y}) \,. \tag{4.182}$$

It is possible to show by reversing the coordinate changes and finding the Schwarzian of the coordinate change that the modular Hamiltonian on the ball takes the form

$$H_B = 2\pi \int_{B_d} \mathrm{d}^{d-1} x \frac{R^2 - r^2}{2R} T_{00}(x) , \qquad (4.183)$$

where $r = \sqrt{\sum x_i^2}$ is a radial coordinate.

4.3 Entanglement Entropy in CFT

In this section some modern results about the entanglement properties of quantum fields are presented. These derivations do not use the Tomita-Takesaki machinery, but rely on some of the general properties that we discovered in the previous sections, such as the fact that the entropy of a subregion is equal to the entropy of the minimal causal diamond containing that subregion. These examples specifically regard CFTs, and are directly relevant for the AdS/CFT duality, as we will see in chapters 6 and 7.

4.3.1 UV-divergences in the Entropy

In the previous sections we have skirted around the issue of a divergent UV contribution to the entanglement entropy very close to the boundary between complementary spacetime regions, as in the case of the Rindler wedges. In the explicit example of the accelerating observer in Minkowski, the observer will see an infinitely large surface with temperature $T = a/(2\pi)$ so the divergence of the entropy is obvious. In the example of a black hole, the integration constant in equation (4.173) ate the universal vacuum-vacuum divergence thus regularizing the black hole entropy. Such a regularized, *vacuum-subtracted* entropy is known in the literature as the *Casini entropy*.

The divergent part of the entanglement entropy comes solely from the vacuum due to the fact that if you zoom in far enough, any finite energy state will look like arbitrarily high-energy random vacuum fluctuations. This means that to understand the structure of the divergent contributions to the entanglement entropy we need only consider the vacuum state, $\rho = |\Omega\rangle\langle\Omega|$. Since the divergences come from very high energy physics, we expect them to be very localized. In particular, this means that the entanglement entropy should be the integral of a local property of the entangling surface between a region and its complement. Causally complementary regions in a d + 1-dimensional spacetime have an entangling surface that is d - 1-dimensional. A surface that has two (n) fewer dimensions than the full spacetime is formally referred to as a "codimension 2 (n)" surface.

Let us write down the entropy of a region A as the most general local, Lorentz invariant expression of the surface parameters

$$S_A^{\text{div}} = \int_{\partial A} \mathrm{d}^{d-1} x \sqrt{h} F(K_{ab}(x), h_{ab}(x)) , \qquad (4.184)$$

where h is the induced metric on ∂A , K is the spacelike extrinsic curvature $\nabla_a \hat{n}^a$ on ∂A constructed from h and F is an arbitrary analytic function thereof. We can expand F as a power series in K. The extrinsic curvature K has dimension of inverse length since its a directional derivative. The global vacuum is a pure state, so we know that the complement of A has the same entropy and the same boundary surface, but the intrinsic curvature K has the opposite sign since the normal points inwards. If the two regions have the same entropy it is clear that only even powers of K can be included in the expansion. At zeroth order in K the entropy integral should be proportional to L^{d-1} where L is a characteristic length of the surface, and each factor of K should contribute a factor $\sim 1/L$. Thus we have an expansion of the form

$$S_A^{\text{div}} \sim a_1 L^{d-1} + a_2 L^{d-3} + \dots ,$$
 (4.185)

where the coefficients a_i are theory dependent and depend on the induced metric h, while the powers of L are universal. The leading contribution to the entanglement entropy is proportional to the d-1-dimensional volume of the entangling surface, i.e. "area". This should be expected, the leading entropy contribution is extremely high energy modes coupling at short distance, and the number of such modes is proportional to the area. The divergences sit in the coefficients a_i as we shall soon see. In theories that are not scale invariant, the divergence sits purely in a_1 this coefficient is associated with the L^{d-1} scaling expected of a local quantity integrated over ∂A .

Let us sit back for just a moment and think about this result. We are claiming that in *any* quantum theory, the leading term in the entanglement entropy is proportional to the area. In addition, we recently showed in section 4.2.3 that the black hole entropy may be interpreted in terms of a Rindler observer in the vacuum that is entangled with degrees of freedom on the other side of an event horizon. The claim that the entropy of black holes as observed from the outside *truly* is due to entanglement with the interior is supported by the above derivation, since entanglement entropy is precisely the type of entropy with a leading area contribution. It would be strange for the black hole entropy to be an honest thermal entropy for the simple reason that such an entropy should be extensive. This realization that the leading black hole entropy is truly due to entanglement across a horizon is the intuition that led to the holographic principle, which itself led to the AdS/CFT correspondence and the *Ryu-Takayanagi proposal*, which we discuss in section 7.2.

Scale Invariant Theories

In the case of a scale invariant theory, the only dimensionful quantities in the problem are the inverse energy ϵ_{UV} and the characteristic length L. The entropy is supposed to be an observable, meaning it must be scale invariant. By dimensional analysis, since the a_i by definition do not get to depend on the length scale, we need $a_1 \sim \epsilon_{UV}^{1-d}$, $a_2 \sim \epsilon_{UV}^{3-d}$ and so forth. Defining new dimensionless coefficients b_i the expansion then reads in even dimension

$$S_A^{CFT} \sim b_{d-1} \left(\frac{L}{\epsilon_{UV}}\right)^{d-1} + \sim b_{d-3} \left(\frac{L}{\epsilon_{UV}}\right)^{d-3} + \ldots + b_1 \frac{L}{\epsilon_{UV}} + \tilde{S} + \mathcal{O}(\epsilon_{UV}/L)$$
(4.186)

and for odd d:

$$S_A^{CFT} \sim b_{d-1} \left(\frac{L}{\epsilon_{UV}}\right)^{d-1} + b_{d-3} \left(\frac{L}{\epsilon_{UV}}\right)^{d-3} + \dots + b_2 \left(\frac{L}{\epsilon_{UV}}\right)^2 + \tilde{S} \log \frac{L}{\epsilon_{UV}} + \text{const} + \mathcal{O}(\epsilon_{UV}) , \qquad (4.187)$$

where the log term comes from a $\sim \int d^{d-1}x x^{1-d}$ in the K power expansion. The b_i and \tilde{S} depend on the theory but not on the energy- or length scales ϵ_{UV} and L. In a theory that is not scale invariant, there are additional dimensionful couplings, so the scale invariance of S_A is broken. We do not prove this, but the terms proportional to \tilde{S} are independent on the choice of UV regulator, leading to \tilde{S} being referred to as the renormalized entanglement entropy.

In 1+1-dimensional CFT, the entanglement entropy of a partial strip of length L_0 is given by

$$S_{2dCFT} = \frac{c}{3} \ln \frac{L_0}{\epsilon_{UV}} ,$$
 (4.188)

which matches the expectation of a leading log divergence from equation (4.187). We show that the leading divergence indeed takes this form in section 4.3.3.

4.3.2 RG-flow of Central Charge in 2d

In section 3.4.4 we encountered the central charge of two-dimensional CFT. We hinted that it measures the number of degrees of freedom of the theory by claiming that in string theory, each scalar and spinor give a positive contribution to it. Renormalization involves integrating out high energy degrees of freedom, which if our intuition is true must mean that the central charge should decrease. The proof that this is the case in 2d CFT is called the *c*-theorem and it can be proven quite simply using the machinery of causal wedges and strong subadditivity of quantum entropy. The c-theorem was originally proved in the 80s by Zamolodchikov without reference to entanglement entropy. The version of the proof we are about to consider is much simpler and due to Casini and Huerta [40], showcasing the power of the quantum information machinery in QFT.

c-theorem

The c-theorem states that the central charge c decreases monotonically under a unitary, Lorentz invariant RG-flow. This can be show relatively simply by considering the following setup. Consider two spacelike lines A, B of equal proper length in the $\tau - \sigma$ plane as is in figure 4.6, where we have also labelled three linez X, Y and Z that are defined in terms of the endpoints of A, B. Here, τ is the timelike CFT coordinate and σ the spatial coordinate. The shaded region is the causal wedge belonging to either the line $A \cup B$ or $X \cup Y \cup Z$. Two regions with the same minimal causal wedge must have the same entropy since they are completely determined in terms of each other by unitary time evolution.

By inspecting figure 4.6 and comparing causal wedges, it is clear that

$$S_A = S_{X \cup Y} , \qquad S_B = S_{Y \cup Z}$$

$$S_{A \cup B} = S_{X \cup Y \cup Z} , \qquad S_{A \cap B} = S_Y ,$$
(4.189)

where $A \cap B$ denotes the intersection of the causal wedges of A and B respectively. Strong subadditivity (equation (2.66)) implies that

$$S_A + S_B \ge S_{A \cup B} + S_{A \cap B} , \qquad (4.190)$$

or equivalently

$$S_{XY} + S_{YZ} \ge S_Y + S_{XYZ} \tag{4.191}$$

where we have left the unions implicit. We can parametrize the length l of the respective regions by

$$l(A) = l(B) = \sqrt{rR}$$
, $l(Y) = R$, (4.192)



Figure 4.6: Two spacelike lines A and B of equal proper length, together with other the related lines X, Y and Z, expressed in terms of the time coordinate τ and the spatial coordinate σ . The shaded region is the causal wedge or domain of dependence of $A \cup B$. Equivalently, it is the domain of dependence of the line $X \cup Y \cup Z$. The endpoints of A and B lie on the lightcone of the point P and are thus related to Y via a Lorentz boost about P.

where R is the proper distance between the outer endpoints of the causal wedge and r is the proper length of Y. Then, since the entanglement entropy in a 2d CFT on any interval can only depend on the proper length of the region, strong subadditivity reads

$$2S(\sqrt{rR}) \ge S(R) + S(r) . \tag{4.193}$$

Letting $R = r + \epsilon$, expanding in powers of ϵ/r for the argument and in powers of the total perturbation for the entropies, equation (4.193) becomes

$$2S\left(r+r\frac{1}{2}\frac{\epsilon}{r}-r\frac{1}{8}\frac{\epsilon^2}{r^2}\right) \ge S(r+\epsilon)+S(r)$$

$$2\left[S(r)+S'(r)\left(\frac{\epsilon}{2}-\frac{1}{8r}\epsilon^2\right)+\frac{1}{8}S''(r)\epsilon^2+\mathcal{O}(\epsilon^3)\right] \ge 2S(r)+S'(r)\epsilon+\frac{1}{2}S''(r)\epsilon^2$$

$$(4.194)$$

Notably the zeroth and first order terms in ϵ are the same on both sides, and we have dividing by ϵ^2 and multiplying by r that

$$-\frac{1}{4r}S'(r) + \frac{1}{4}S'' \ge \frac{1}{2}S''(r)$$

$$0 \ge \frac{1}{4}\left(S'(r) + rS''(r)\right) .$$
(4.195)

We can define the C-function as

$$C(r) = rS'(r) , \quad C'(r) \le 0$$
 (4.196)

where the inequality is equivalent to strong subadditivity. Notably, the function C is monotonic as a function of interval size - increasing the size of the strips decreases the value of C.

So far, we have not invoked scale invariance. Let us now consider the case in which we have a scale invariant theory. At any RG fixed point, a generic QFT is scale invariant and in 2d this necessarily implies the full conformal invariance. We have an explicit expression for the entanglement entropy under the assumption of scale invariance in equation (4.188):

$$S_{cft}(r) = \frac{c}{3} \ln \frac{r}{\epsilon_{UV}} , \qquad (4.197)$$

where c is the central charge. Then, the C-function has an explicit expression

$$C_{cft} = \frac{c}{3}$$
 . (4.198)

If the QFT in question is not scale invariant, the RG flow describes a transition between an UV CFT and an infrared CFT. We know at the fixed points that the central C-function must be

$$C_{UV} = \frac{c_{UV}}{3} , \quad C_{IR} = \frac{c_{IR}}{3}.$$
 (4.199)

We interpret the physical distance r, that is the length of the strip Y, as the renormalization scale. Very high energy corresponds to very short distances, so by increasing r we are inducing an RG flow from the UV to the IR. But we know that $C'(r) \leq 0$ for all r, so

$$\int_{r_{UV}}^{r_{IR}} \mathrm{d}r \, C' \le 0 \, . \tag{4.200}$$

Then the fundamental theorem of calculus tells us that

$$c_{UV} \ge c_{IR} , \qquad (4.201)$$

proving the c-theorem, namely that the central charge decreases under RG flow.

A very interesting fact is that we never had to refer to the explicit field content of the theory. This behaviour is completely general given any Lorentz invariant QFT with IR and UV fixed points. This result can also be generalized to some higher dimensions as done in [40].

4.3.3 Replica Method

In section 4.2.1 we used a path integral construction and a claimed Hilbert space factorization to compute the local density operator in the Rindler half-space. The generalization of this construction to a finite region of arbitrary shape is horribly complicated. We will in this section outline the so-called *replica construction*, as well as the steps to deriving the entanglement entropy of a finite line segment in a 1 + 1-dimensional CFT. We base the following discussion on the review by Cardy and Calabrese [41], as well as the presentation by Rangamani and Takayanagi in [42]. The example here does not let us compute something we could not have computed with the modular Hamiltonian of section 4.2.4, but the path integral version generalizes to multiple disconnected regions in a tractable way. In addition the replica construction admits a holographic interpretation as a gravitational problem using the AdS/CFT correspondence as we will see in section 7.2.2.

Let us in the following consider a general QFT whose Hilbert space we claim we can factorize according to $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. We take the Hilbert spaces A and B to be belonging to causally complementary spacetime regions. The full quantum state is assumed to be a pure state.

There is a generalization of the von Neumann entropies called the Renyi entropies $S_A^{(n)}$, which we introduce in the following section. The computation of the Renyi entropies has us evaluate the quantity ρ_A^n , for integer n, and then analytically continue the result to $n \in \mathbb{R}$. This leads to a path integral over several copies of the $\tau - x$ -plane glued together in a certain way. In CFT we will see that we can express this as a path integral over a single plane with so-called *twist operators* \mathcal{T} inserted. The replica trick is general to d-dimensional balls in d + 1-dimensional spacetimes, but we will carry out the explicits in 1 + 1 dimensions.

Renyi Entropies

In the case of CFTs, it turns out to be useful to define the Renyi entropies as

$$S_A^{(n)} \equiv \frac{1}{1-n} \ln \operatorname{Tr}[\rho_A^n]$$
 (4.202)

In the case when ρ is pure we have $S_A^{(n)} = S_B^{(n)}$. We know that $\operatorname{Tr}_A[\rho_A^n]$ is finite for all positive integers n since the eigenvalues of ρ are between 0 and 1. We can then analytically continue the Renyi entropies to any n, taking its expression in terms of the eigenvalues of ρ to define it. Then, the Renyi entropies reduce to the usual entropy in the limit $n \to 1$. To see this we diagonalize ρ according to $\rho = \sum \lambda_j |j\rangle\langle j|$ and remember that ρ has unit trace. Then,

$$\operatorname{Tr}\rho^{n} = \sum \lambda_{j}^{n} = 1 - \sum_{j} (\lambda_{j} - \lambda_{j}^{n}) \equiv 1 - f(n) . \qquad (4.203)$$

We can then rewrite f(n) in a tricky manner

$$f(n) = \sum_{j} \lambda_{j} (1 - \lambda_{j}^{n-1})$$

= $\sum_{j} \lambda_{j} (1 - e^{(n-1)\ln\lambda_{j}})$
= $\sum_{j} \lambda_{j} (1 - n) \ln\lambda_{j} + \mathcal{O}((n-1)^{2})$
= $(1 - n) \operatorname{Tr} \rho_{A} \ln\rho_{A} + \mathcal{O}((n-1)^{2})$. (4.204)

Inserting back into the Renyi entropies we see

$$S_A^{(n)} = \frac{1}{1-n} \ln \left(1 - (1-n) \operatorname{Tr}[\rho_A \ln \rho_A] \right)$$

= $-\operatorname{Tr}[\rho_A \ln \rho_A] + \mathcal{O}((1-n)^2)$ (4.205)

and we see that we indeed recover the von Neumann entropy as $n \to 1$. Another way to write the limit is

$$S_A = -\frac{\partial}{\partial n} \operatorname{Tr} \rho_A^n \bigg|_{n=1}$$
(4.206)

where the single derivative tells us to keep the first coefficient in the n power expansion of $\text{Tr}\rho^n$.

The Replica Path Integral

In section 4.2.1 we wrote the general expression for the local density operator on the right Rindler wedge in terms of a path integral by gluing together the path integral along the left wedge. We then ended up with a path integral over the spacetime $W_{2\pi}$ with a cut along the half axis x > 0. The basic ingredient was that the vacuum could be written as in equation (3.66). If we want to find the expectation value of time-ordered operators in the state ρ_A the path integration must be carried out on a mixed-signature generalization of the Euclidean $\tau - x$ -plane using the Schwinger-Keldysh formalism we introduced in section 3.2.2, but we will here consider only the simplest case. As long as the global state is time translation symmetric, a pure euclidean time analysis such as the one we will now carry out is sufficient.

The vacuum $|\Omega\rangle$ is time tanslation invariant. We would like compute the quantity $\operatorname{Tr}[\rho_A^n]$ when A is a strip of finite length l, i.e. $x \in [0, l]$ and ρ_A is the associated local density operator. We know that ρ_A is given in terms of $\phi_A \in \mathcal{H}_A$ by,

$$\langle \phi_A^f | \rho_A | \phi_A^i \rangle = \left(\int_{\phi_A^i}^{\phi_A^f} \mathcal{D}\phi_L | \Omega \rangle \left(\phi_B, \phi_R' \right) \langle \Omega | \left(\phi_B, \phi_R \right) \right) , \qquad (4.207)$$

where the trace over B glues together the path integral outside of the region A. Graphically, we

can write the contour as

$$\langle \phi_{A}^{f} | \rho_{A} | \phi_{A}^{i} \rangle = \infty \left[\begin{array}{c} & & & \\ & \uparrow^{*} & & \\ & & & & \\ & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\$$

Let us think about how we can understand the quantity ρ^n in the path integral language. The trick is to insert completion relations between each factor of ρ so we have, denoting the initial and final states ϕ^0_A and ϕ^n , that

$$\left\langle \phi_{A}^{n} \right| \rho_{A}^{n} \left| \phi_{A}^{0} \right\rangle = \sum_{j=1}^{n-1} \sum_{i_{j}} \left\langle \phi_{A}^{f} \right| \rho_{A} \left| \phi_{A}^{i_{n-1}} \right\rangle \left\langle \phi_{A}^{i_{n-1}} \right| \rho_{A} \dots \left| \phi_{A}^{i_{1}} \right\rangle \left\langle \phi_{A}^{i_{1}} \right| \rho_{A} \left| \phi_{A}^{0} \right\rangle \right\rangle$$

$$(4.209)$$

Here, each factor of $\langle \phi_A^{i_{j+1}} | \rho_A | \phi_A^{i_j} \rangle$ is a path integral over the $\tau - x$ -plane as in equation (4.208). Therefore, ρ_A^n is clearly a path integral over n replicas of this plane. Furthermore, the completion relation between each of the factors of ρ_A instructs us to glue the planes together. Let us label the planes by j, where the zeroth plane belongs to the factor $\langle \phi_A^{i_1} | \rho_A | \phi_A^0 \rangle$. Then, the lefthand completion relation on each factor tells us to glue the top side of the $x \in [0,l]$ cut of the j:th plane to the j + 1:th plane. Taking the trace of ρ_A^n then instructs us to glue the remaining cuts on the n - 1:th and zeroth planes together. The integration surface is most understandable by inspecting a graphical representation:



where we have replaced the completion relations with an identification of boundary conditions, denoted by a line connecting the strips on which the boundary fields were previously defined. Note that the cyclicity of the trace means that the path integral construction is symmetric under a cyclic symmetry with n distinct elements. This symmetry is generally referred to as the *replica symmetry*.

If we want to use the Renyi entropies to find the von Neumann entropy we want to be able to analytically continue the above construction from $n \in \mathbb{N}$ to $n \in \mathbb{R}$, since the continuous limit $n \to 1$ makes no sense if $S_A^{(n)}$ is not defined for noninteger n. In general a function defined only for integer values does not admit a unique analytic continuation. There is a mathematical result called Carlson's theorem that states that functions that are defined on integers z have a unique analytic continuation to the complex plane if they grow slower than an exponential as $z \to \pm i\infty$.

This means that if you can compute the Renyi entropies as an analytic function f(n), and we can argue from physical principles that they are well behaved at imaginary infinity, we should be able to take the values for non integer n seriously. In general, failure of the replica construction may imply that there is some interesting physics at play since it invalidates the physical arguments for the nice behaviour at infinity [38]. We will not in the following encounter a case in which the analytic continuation behaves poorly.

Single Interval in CFT_2

We consider the vacuum of a 1 + 1-dimensional CFT. We let the Lorentzian time be t = 0, without loss of generality. We define A to be the strip $-a \le x \le a$, $\tau = 0$. The replica path integral has us take n copies of the $\tau - x$ -plane with slits along A and glue them as in equation (4.210). We want to find the entanglement entropy of this finite line, which is equal to the von Neumann entropy since the global state is pure. The following calculation skips some computational steps, the purpose is to make the main ideas clear.

We can now use the fact that we are working with a CFT; the glued together planes is a complicated geometry but it can be mapped to something much simpler. It is in fact possible to map this to the complex plane, together with the identification under the *n*-fold replica symmetry as in figure 4.7. The singular behaviour of the orbifold theory at the fixed points at the origin and the point at infinity reflect the fact that the geometry of the n-sheeted surface is very sharp at -a, a. This is captured by so-called *twist operators* \mathcal{T} , inserted at the boundary points -a, a.

By going to holomorphic coordinates z = x + it and $\overline{z} = x - it$ we can continue the analysis. The explicit coordinate change to the orbifold plane is given for the k:th plane by $w \to z(w)$

$$z(w) = \left(\frac{w+a}{w-a}\right)^{1/n+k/n} .$$
 (4.211)

The fields that live on single copies of the complex plane $\phi(z,\overline{z})_k$, $k \in [0,n-1]$ together with the gluing conditions fuse into a single field $\varphi(w,\overline{w})$ obeying so-called *twisted boundary conditions*. Figure 4.7 aims to make this relation clear. In this case, the twisted boundary condition states that $\varphi(x,\tau) = \mathcal{R}(\frac{2\pi}{n})\varphi(x,\tau)$ where $\mathcal{R}(\theta)$ implements a rotation in the $\tau - x$ -plane by θ radians.

The stress-energy tensor in any of the original planes has expectation value $\langle T_k(w) \rangle = 0$. Let us do the following analysis for the zeroth plane and invoke replica symmetry for the generalization to all planes. This transformation of the coordinates induces a change in the stress-energy tensor as in equation (3.218), which after some work can be seen to have the expectation value

$$\langle T(w) \rangle \to \langle T(z) \rangle = \frac{c}{24} \left(1 - \frac{1}{n^2} \right) \frac{(2a)^2}{(w-a)^2 (w+a)^2} ,$$
 (4.212)

where we used that $\langle T_k(w) \rangle = 0$ and the surviving term is the expectation value of the Schwarzian. The antiholomorphic stress tensor component transforms in the same manner. We will leave the computation of the Schwarzian until after the rest of this derivation.

The expectation value on the orbifold $\langle T(w) \rangle$ is proportional the three-point function $\langle T(w)\mathcal{T}_n(-a)\mathcal{T}_{-n}(a) \rangle$ in the original zeroth plane. The fields \mathcal{T}_n are called the *branch-point twist fields* and have conformal



Figure 4.7: To the left, integration regions for each of the replicas, where the lines indicate gluing boundary conditions. Using conformal symmetry we may label each plane by an index k and let $z(w) = \left(\frac{w+a}{w-a}\right)^{(1+k)/n}$ taking us from n w-planes to a single orbifold, defined on right. The geometry on the right is an "orbifold" because the origin and the point at infinity are fixed points of the replica symmetry.

weights

$$h_n = \overline{h}_n = \frac{c}{24} \left(1 - \frac{1}{n^2} \right) . \tag{4.213}$$

The two-point correlator is given by $\langle \mathcal{T}_n(-a)\mathcal{T}_{-n}(a)\rangle = \epsilon/(2a)^{2(h_n+\overline{h}_n)}$. The constant ϵ accounts for the fact that the constant in the two-point function is not fixed by the conformal Ward identities. It is possible to relate ϵ to the UV cutoff ϵ_{UV} introduced in section 4.3.1.

Thanks to replica symmetry, by inserting these operators into all of the *n* planes and computing the three-point correlation functions as if there were just *n* independent CFTs on the plane we can obtain the expectation value of T(z) on the orbifold geometry. We know from chapter 3.4.4 that the expectation of $\langle T\mathcal{O} \rangle$ in any state determines the conformal dimension of \mathcal{O} . Looking at $\text{Tr}[T\rho]$ term by term in the completion relation that defines the trace, this tells us that ρ^n living on the orbifold transforms as if it was the *n*:th power of the two-point correlator $\langle \mathcal{T}_n(-a)\mathcal{T}_{-n}(a)\rangle$ on a single sheet. Therefore,

$$\operatorname{Tr}[\rho^{n}] = c_{n} \langle \mathcal{T}_{n}(-a)\mathcal{T}_{-n}(a) \rangle^{n} = c_{n} \left| \frac{\epsilon}{2a} \right|^{\frac{c}{6}(n-1/n)} , \qquad (4.214)$$

where 2a is the length of the interval A and c_n is an undetermined constant (that depends on n). The function we have obtained looks like it is defined outside of the integers, so we will assume that the necessary conditions hold and furthermore assume that $n \in \mathbb{R}$. We can now recover the von Neumann entropy

$$S_{A} = -\frac{\partial}{\partial n} \operatorname{Tr}_{A} \rho_{A} \bigg|_{n=1} = -\frac{\partial}{\partial n} \left(c_{n} \left| \frac{\epsilon}{2a} \right|^{\frac{c}{6}(n-1/n)} \right) \bigg|_{n=1}$$
$$= -\left[c_{n} \ln\left(\frac{2a}{\epsilon}\right) \frac{c}{6} \left(-1 - \frac{1}{n^{2}} \right) (2a^{-\frac{c}{6}(n-\frac{1}{n})}) + c_{n}' \left| \frac{\epsilon}{2a} \right|^{\frac{c}{6}(n-1/n)} \right] \bigg|_{n=1}$$
$$= c_{1} \frac{c}{3} \ln\left(\frac{2a}{\epsilon}\right) - c_{1}' .$$
(4.215)

This obeys the general form predicted by equation 4.187. It is possible to check that the condition $\text{Tr}[\rho] = 1$ sets $c_1 = 1$. Denoting the length of the interval by L = 2a we obtain the form of the

entropy of the line element as given in equation (4.188). The constant term $-c'_1$ is in general dependent on the operator content of the relevant CFT, while the first term is universal. Keeping only the universal term we have obtained

$$S_A = \frac{c}{3} \ln\left(\frac{L}{\epsilon}\right) \tag{4.216}$$

for the finite line element in a 2d CFT.

It is possible to extend the above reasoning to a CFT on a cylinder of circumference R. This requires a conformal transformation from the plane with complex coordinate w to the cylinder with complex coordinate w', taking the form

$$w = \tan \frac{\pi w'}{R} , \qquad (4.217)$$

The Renyi entropy now consists of *n* copies of a cylinder with the gluing conditions. The idea is that the stress tensor has zero expectation value on the cylinder, but by considering how it transforms under the sequence of transformations equation (4.217) followed by equation (4.211) we can understand once again how ρ^n should transform in the orbifold picture by looking at the expectation of *T*. The additional contribution becomes a factor $\left[\frac{L}{\pi}\cos\left(\frac{\pi L}{R}\right)\right]^{-\frac{c}{6}(n-1/n)}$ to multiply onto equation (4.214).

In the case of a spatial circle, we have

$$S_A = \frac{c}{3} \ln\left(\frac{R}{\pi\epsilon} \sin\frac{\pi L}{R}\right) \,. \tag{4.218}$$

Note that for $L \ll R$ this reduces to the flat case. In the case where we have a thermal state of temperature $\beta = R$ on a non-compact spacetime, the entropy is given by

$$S_A = \frac{c}{3} \ln \left(\frac{\beta}{\pi \epsilon} \sinh \frac{\pi L}{\beta} \right) \,. \tag{4.219}$$

The compactification on the thermal circle is related to the cylinder by the change of coordinates

$$w'' = e^{\frac{2\pi}{\beta}w'} . (4.220)$$

The results on the spatial and thermal circles are simply related by the substitution $R \to i\beta$, which is related to the fact that they correspond to a compactification either of the real or imaginary directions in the $x - \tau$ plane.

Computing the Schwarzian

Let us compute the Schwarzian of the replica \rightarrow orbifold change of coordinates. It is defined by

$$S[z,w] \equiv \frac{\partial_w^3 z}{\partial_w z} - \frac{3}{2} \left(\frac{\partial_w^2 z}{\partial_w z} \right)^2 \,. \tag{4.221}$$

Here we consider only the zeroth plane, noting that replica symmetry generalizes the result to all n planes. That means that we have

$$z(w) = \left(\frac{w+a}{w-a}\right)^{1/n} . \tag{4.222}$$

The first three derivatives of z(w) are given by

$$\partial_w z = -\frac{1}{n} \frac{2a}{(w-a)^2} \left(\frac{w+a}{w-a}\right)^{\frac{1}{n}-1} , \qquad (4.223)$$

$$\partial_w^2 z = \frac{1}{n} \left(\frac{1}{n} - 1\right) \frac{4a^2}{(w-a)^4} \left(\frac{w+a}{w-a}\right)^{\frac{1}{n}-2} + \frac{2}{n} \frac{1}{n} \frac{2a}{(w-a)^3} \left(\frac{w+a}{w-a}\right)^{\frac{1}{n}-1} , \qquad (4.224)$$

$$\partial_w^3 z = -\frac{1}{n} \left(\frac{1}{n} - 1\right) \left(\frac{1}{n} - 2\right) \frac{8a^3}{(w-a)^4} \left(\frac{w+a}{w-a}\right)^{\frac{1}{n}-3}$$
(4.225)

$$-\frac{6}{n}\left(\frac{1}{n}-1\right)\frac{4a^{2}}{(w-a)^{5}}\left(\frac{w+a}{w-a}\right)^{\frac{1}{n}-2} \\ -\frac{6}{n}\frac{2a}{(w-a)^{4}}\left(\frac{w+a}{w-a}\right)^{\frac{1}{n}-1}.$$

Now we just need to combine the parts into the terms that feature in the Schwarzian. We have that

$$\frac{\partial_w^3 z}{\partial_w z} = \left(\frac{1}{n} - 1\right) \left(\frac{1}{n} - 2\right) \frac{4a^2}{(w-a)^2(w+a)^2} , + \frac{6}{n} \left(\frac{1}{n} - 1\right) \frac{2a}{(w-a)^2(w-u)} + \frac{6}{(w-a)^2}$$
(4.226)

and

$$\frac{\partial_w^2 z}{\partial_w z} = -\left(\frac{1}{n} - 1\right) \frac{2a}{(w-a)(w+a)} - \frac{2}{w-a} . \tag{4.227}$$

We need the square of the latter, given by

$$\left(\frac{\partial_w^2 z}{\partial_w z}\right)^2 = \left(\frac{1}{n} - 1\right)^2 \frac{4a^2}{(w-a)^2(w+a)^2} + \left(\frac{1}{n} - 1\right) \frac{8a}{(w-a)^2(w+a)} + \frac{4}{(w-a)^2}$$
(4.228)

Now, we are ready to compute the Schwarzian, which becomes

$$S[z,w] = \left(\frac{1}{n} - 1\right) \left(\frac{1}{n} - 2\right) \frac{4a^2}{(w-a)^2(w+a)^2} + \frac{6}{n} \left(\frac{1}{n} - 1\right) \frac{2a}{(w-a)^2(w-u)} + \frac{6}{(w-a)^2} - \left(\frac{1}{n} - 1\right)^2 \frac{6a^2}{(w-a)^2(w+a)^2} - \left(\frac{1}{n} - 1\right) \frac{12a}{(w-a)^2(w+a)} - \frac{6}{(w-a)^2}.$$
(4.229)

We can rearrange and see that

$$S[z,w] = \left[\left(\frac{1}{n} - 1\right) \left(\frac{1}{n} - 2\right) - \frac{3}{2} \left(\frac{1}{n} - 1\right)^2 \right] \frac{4a^2}{(w+a)^2(w-a)^2}$$

$$= \frac{1}{2} \left(1 - \frac{1}{n^2}\right) \frac{4a^2}{(w+a)^2(w-a)^2} ,$$
 (4.230)

which indeed gives the claimed expression in equation (4.212) since in the present case the stress tensor is proportional $\frac{c}{12}S[z,w]$.

Chapter 5

Elements of String Theory

In this chapter the basics elements of string theory are presented. This includes an introduction to string perturbation theory and the α' -expansion. The low energy effective action for bosonic strings in a gravitational background is derived as an illustrative example using the nonlinear sigma approach. The low energy effective action is then formulated in the presence of more complicated background fields. The extension to superstring theory is made, and the low energy effective action of superstring theory is given.

String theory is important for the motivation of fully quantum variants of the AdS/CFT duality to be introduced in chapter 6. Specifically, unitary CFTs have been discovered through the conformal bootstrap program to require the existence of an infinite number of conformal primary operators of increasing conformal weight. In the AdS/CFT duality this corresponds to a theory in the bulk containing an infinite number of increasingly massive fields. A quantum theory of gravity that realizes this constraint is string theory, as will be apparent in this chapter.

5.1 Bosonic String Theory

In this section we sketch the construction of bosonic string theory, since the bosonic sector of superstring theory is very similar to bosonic string theory. The purpose is to remind the reader of the main points of the theory while establishing notation, as well as making explicit some advanced topics, such as string perturbation theory, vertex operators and the α' -expansion. The reason for introducing these topics in the bosonic theory is that they are similar in principle to the corresponding superstring constructions, but algebraically much cleaner. This will let us develop an intuitive understanding for the underlying framework in a tractable example, before just citing the results in the supersymmetric case.

The basic idea of string theory is to consider the idea that rather than point particles, the fundamental objects in the world are extended strings. Just as a point particle traces out a (timelike) worldline in spacetime as it moves, a string will sweep out a 1 + 1-dimensional worldsheet.

For a point particle in a general spacetime, the equations of motion minimize the proper length of the worldline, described up to a constant by the action

$$S = \int \sqrt{-g} \mathrm{d}\lambda \,\,, \tag{5.1}$$

where λ parametrizes the geodesic. For the relativistic string, the action is given by the the proper area of the worldsheet Σ according to

$$S_{NG} = -\frac{1}{2\pi\alpha'} \int_{\Sigma} \mathrm{d}^2 \sigma \sqrt{-g} , \qquad (5.2)$$

where $g_{\mu\nu}$ is the induced metric on the surface Σ . The square root $\sqrt{-g}$ denotes the square root of minus the determinant of $g_{\mu\nu}$. The proper area functional for the string worldsheet is also called the *Nambu-Goto* action.

The string world sheet Σ can be parametrized by the proper time τ and the spacelike coordinate σ , usually referred to as *world sheet coordinates*. It is often practical to write these as $\sigma^{\alpha} = (\sigma^0, \sigma^1) = (\tau, \sigma)$. The coordinate σ lies in $[0, \sigma_0]$, where the best choice of σ_0 depends on what type of string you are considering. The embedding of the string into the target spacetime is given by the functions $X^M(\tau, \sigma)$, where the indices M are indices of the target spacetime. The more explicit form of the Nambu-Goto action is now given in a target spacetime with metric g_{MN} by

$$S_{NG} = -\frac{1}{2\pi\alpha'} \int d\tau d\sigma \sqrt{-\text{Det}\left(\frac{\partial X^M}{\partial\sigma^{\alpha}}\frac{\partial X^M}{\partial\sigma^{\beta}}g_{MN}\right)},$$
(5.3)

where the determinant is over the induced metric $\tilde{g}_{\alpha\beta}$. The induced metric is obtained in the usual fashion by

$$\tilde{g}_{\alpha\beta} = \frac{\partial X^M}{\partial \sigma^{\alpha}} \frac{\partial X^M}{\partial \sigma^{\beta}} g_{MN} , \qquad (5.4)$$

i.e. we have just performed a coordinate change onto the surface described by $X(\tau, \sigma)$. Formally this operation is the *pullback* of the target spacetime metric onto the world sheet coordinates.

The constant α' is related to the string length scale l_s by $\alpha' = l_s^2$, and the tension of the string is given by $T_{F1} = \frac{1}{2\pi\alpha'}$. Due to the square root in equation (5.3) it is in general difficult to handle, especially in a path integral quantization scheme. A canonical quantization of the Nambu-Goto action is readily found in Zwiebach [43].

To successfully quantize string theory we introduce an auxiliary wordsheet metric $h_{\alpha\beta}(\sigma)$ and write down the *Polyakov action*

$$S_P = -\frac{1}{4\pi\alpha'} \int d^2\sigma \sqrt{-h} h^{\alpha\beta} \partial_\alpha X^M \partial_\beta X^N g_{MN} , \qquad (5.5)$$

where $h^{\alpha\beta}$ is the inverse of $h_{\alpha\beta}$. The Polyakov action is dynamically equivalent to the Nambu-Goto action in the sense that is has the same Hamiltonian and the same equations of motion. The auxiliary metric $h^{\alpha\beta}$ is symmetric and 2×2 so naively it has three independent components. The h's only appear in the combination $\sqrt{-h}h^{\alpha\beta}$ which is invariant under Weyl rescalings

$$h_{\alpha\beta} \to h_{\alpha\beta}{}' = \Omega^2(\tau, \sigma) h_{\alpha\beta} , \quad h^{\alpha\beta} \to h^{\alpha\beta\prime} = \Omega^{-2}(\tau, \sigma) h^{\alpha\beta\prime} ,$$
 (5.6)

which in two dimensions this leads to the restriction $\det(\sqrt{-h}h^{\alpha\beta}) = -1$. To recover the Nambu-Goto action we can vary the action with respect to $h_{\alpha\beta}$ to find

$$0 = \frac{\delta S_P}{\delta h^{\alpha\beta}} = -\frac{\sqrt{-h}}{4\pi\alpha'} \left(\partial_\alpha X^M \partial_\beta X^N g_{MN} - \frac{1}{2} h_{\alpha\beta} h^{\rho\sigma} \partial_\rho X^M \partial_\sigma X^N G_{MN} \right) .$$
(5.7)

Equation equation (5.7) can be used to eliminate the auxiliary metric $h_{\alpha\beta}$ from the action and reobtain the Nambu-Goto acion. In addition, it gives rise to constraints on the fields X^M of the theory known as the *Virasoro* constraints. When performing a manifestly covariant quantization of the Nambu-Goto action these have to be put in by hand, but they fall naturally out of the Polyakov action. In general we will in the following work with the Polyakov action. An additional point is that we can define the world-sheet stress energy tensor

$$T_{\alpha\beta} = -\frac{4\pi\alpha'}{\sqrt{h}}\frac{\delta S_P}{\delta h^{\alpha\beta}} = \partial_{\alpha}X^M\partial_{\beta}X^Ng_{MN} - \frac{1}{2}h_{\alpha\beta}h^{\rho\sigma}\partial_{\rho}X^M\partial_{\sigma}X^NG_{MN} , \qquad (5.8)$$

which is implied by equation (5.7) to vanish.

Let us now remind ourselves of all the symmetries of the action S_P , now restricting to the case of a *D*-dimensional Minkowski target spacetime, setting $G_{MN} = \eta_{MN}$:

• Poincaré invariance in target spacetime. The action S_P is manifestly Poincaré invariant since it is constructed only from Lorentz invariant combinations of derivatives of the coordinates. The action of Poincaré transformations on the fields in S_P is given by

$$X^M \to \Lambda^M{}_N X^N + a^M , \quad h_{\alpha\beta} \to h_{\alpha\beta} .$$
 (5.9)

• Reparametrization Invariance. The parametrization of the worldsheet is arbitrary, and the invariance of the Nambu-Goto (and therefore also Polyakov) action under reparametrizations is a central property. Under a reparametrization $\sigma^{\alpha} \rightarrow f^{\alpha}(\sigma)$ the fields transform according to

$$X^M \to X^M$$
, $h_{\alpha\beta}(\tau,\sigma) \to \frac{\partial f^{\sigma}}{\partial \sigma^{\alpha}} \frac{\partial f^{\sigma}}{\partial \sigma^{\beta}} h_{\rho\sigma}(\tau',\sigma')$ (5.10)

In the case of the Polyakov action, the cancellation is readily seen since $h^{\alpha\beta}$ transforms in negative powers of ∂f while $\sqrt{-h}$ transforms by positive powers.

• Weyl transformations. The Polyakov enjoys one additional symmetry, under Weyl rescalings, as we defined in equation (5.6).

This set of local symmetries allows us to choose a gauge in which the worldsheet metric $h^{\alpha\beta}$ is proportional to $\eta_{\alpha\beta} = \text{Diag}(-1,1)$. This is called the *conformal gauge*, where

$$h_{\alpha\beta} = e^{2\omega(\tau,\sigma)} \eta_{\alpha\beta} .$$
 (5.11)

The Polyakov action in conformal gauge is

$$S_P = \frac{1}{4\pi\alpha'} \int d^2\sigma \left(\partial_t X^M \partial_t X^N - \partial_\sigma X^M \partial_\sigma X^N\right)$$
(5.12)

and the equations of motion reduce to a simple wave equation

$$(\partial_t^2 - \partial_\sigma^2) X^M = 0 . (5.13)$$

In addition to the equations of motion we need to supply reasonable boundary conditions as well as the Virasoro constraints in equation (5.7). Introducing lightcone coordinates

$$\sigma^{+} = \frac{\tau + \sigma}{\sqrt{2}} , \quad \sigma^{-} = \frac{\tau - \sigma}{\sqrt{2}} ,$$

the action takes the form

$$S_P = \frac{1}{4\pi\alpha'} \int d^2 \sigma \partial_+ X^M \partial_- X^N$$
(5.14)

and the equations of motion as well as the Virasoro constraints take the simple form

$$\partial_{+}\partial_{-}X^{M} = 0 ,$$

$$\partial_{+}X^{M}\partial_{+}X_{M} = 0 ,$$

$$\partial_{-}X^{M}\partial_{-}X_{M} = 0 .$$
(5.15)

Finally, we need to impose the boundary condition

$$\partial_{\sigma} X^{M} \delta X_{M} \Big|_{\sigma=0}^{\sigma_{0}} \tag{5.16}$$

to ensure that range of the parameter $\sigma \in [0, \sigma_0]$ is always the same.

Additionally, we can go to holomorphic and antiholomorphic coordinates $z = e^{2(\tau+ix)}$ and $\overline{z}e^{2(\tau-ix)}$ on the string worldsheet, at which point the Polyakov action takes the form

$$S_P = \frac{1}{4\pi\alpha'} \int d^2 2\partial X^M \overline{\partial} X_M . \qquad (5.17)$$

Imposing Boundary Conditions, Mode Expansions

The equations of motion are solved by Fourier mode expansions of X^M in terms of left- and right moving waves, depending only on σ^+ , σ^- respectively. We write

$$X^{M} = X_{L}^{M}(\sigma^{+}) + X_{R}^{M}(\sigma^{-})$$
(5.18)

and Fourier expand the components as

$$X_{L}^{M}(\sigma^{+}) = \frac{\tilde{x}_{0}^{M}}{2} + \frac{\alpha'}{2}\tilde{p}^{M}\sigma^{+} + i\sqrt{\frac{\alpha'}{2}}\sum_{n\neq 0}\frac{\tilde{\alpha}_{n}^{M}}{2}e^{-in\sigma^{+}} ,$$

$$X_{R}^{M}(\sigma^{-}) = \frac{x_{0}^{M}}{2} + \frac{\alpha'}{2}p^{M}\sigma^{-} + i\sqrt{\frac{\alpha'}{2}}\sum_{n\neq 0}\frac{\alpha_{n}^{M}}{2}e^{-in\sigma^{-}} .$$
(5.19)

The centre of mass position of the string is given by the sum of the x_0^M coordinates, which come from the zero mode of the Fourier expansion. The constants \tilde{p}^M and p^M are the center of mass momenta of the the left- and right moving modes, and appear as non-periodic solutions to the equation of motion. Reality of X^M further imposes that $\alpha_n^{M\dagger} = \alpha_{-n}^M$.

In the case of closed strings, we set $\sigma_0 = 2\pi$. Then, the periodic boundary condition

$$X^{M}(\tau,\sigma) = X^{M}(\tau,\sigma+2\pi)$$
(5.20)

ensures that we automatically satisfy equation (5.16). Moreover the left-and right moving modes are already periodic by construction, except for the aperiodic terms proportional to σ^{\pm} which cancel if we set $p^M = \tilde{p}^M$. Moreover, there is no reason not to set $x_0 = \tilde{x}_0$

For open strings it is most convenient to set $\sigma_0 = \pi$. To satisfy equation (5.16) we can pick either Dirichlet (D) or Neumann (N) boundary conditions on each of the endpoints $\sigma = 0$ and $\sigma = \pi$. Explicitly, letting $\overline{\sigma}$ denote either endpoint, the possible conditions are:

Neumann
$$\partial_{\sigma} X^{M}(\tau, \overline{\sigma}) = 0$$
,
Dirichlet $\delta X^{M}(\tau, \overline{\sigma}) = 0$. (5.21)

The Neumann condition describes a string with a freely moving endpoint that is always parametrized by the same range of σ , and it can be shown that the endpoint must always move at the speed of light. The Dirichlet boundary condition describes a string attached at the endpoint to something. In principle, we can apply these boundary conditions independently to each coordinate in X^M , describing a string whose ends can move freely inside some subspace of the full space.

A subspace on which the endpoint of a string is constrained is called a *D*-brane. D-branes necessarily decay in bosonic string theory but it turns out they are charged under the fields of

superstring theory, making them conserved. We require the D-brane to be a physical object because Dirichlet boundary conditions to not conserve the string center of mass momentum, and the lost momentum has to be carried by something. The action of D-branes can be formulated as the pullback of the global metric onto the worldvolume swept out by the D-brane as it moves in time. The action constructed in this is a part of the so-called *Dirac-Born-Infeld* action that describes the dynamics of D-branes.

Since we can impose boundary conditions separately for both ends of the string, we have three inequivalent choices of boundary conditions, DD, ND and NN. The three possible boundary conditions couple the left- and right moving waves so that we end up with only one expansion. The mode expansions are as follows:

NN:
$$X^M(\tau,\sigma) = x_0^M + 2\alpha' p^M \tau + i\sqrt{2\alpha'} \sum_{n \neq 0} \frac{\alpha_n^M}{n} e^{-in\tau} \cos n\sigma$$
 (5.22)

DD:
$$X^{M}(\tau,\sigma) = x_{i}^{M} + \frac{1}{\pi}(x_{f}^{M} - x_{i}^{M})\sigma + i\sqrt{2\alpha'}\sum_{n\neq 0}\frac{\alpha_{n}^{M}}{n}e^{-in\tau}\sin n\sigma$$
(5.23)

ND:
$$X^M(\tau,\sigma) = x_f^M + i\sqrt{2\alpha'} \sum_n \frac{\alpha_n^M}{n} e^{-in\tau} \cos\left((n+\frac{1}{2})\sigma\right) ,$$
 (5.24)

where x_i denotes the fixed starting point and x_f the fixed ending point defined by the Dirichlet boundary conditions.

On these mode expansions we still have to impose the Virasoro constraints. These are are $T_{++} = \partial_+ X^M \partial_+ X_M = 0$ and similarly for T_{--} . We write these down on a compact form by first introducing the Virasoro modes

$$\tilde{L}_m = \frac{1}{2} \sum_n \tilde{\alpha}_m^M \tilde{\alpha}_{m,M} , \quad L_m = \frac{1}{2} \sum_n \alpha_m^M \alpha_{m,M} , \qquad (5.25)$$

where $\alpha_0^M = p^M$ and similarly for $\tilde{\alpha}_0^M$ In terms of these we can write the worldsheet stress-energy tensor as

$$T_{++} = \alpha' \sum_{m} \tilde{L}_m e^{-im\sigma^+} , \quad T_{--} = \alpha' \sum_{m} L_m e^{-im\sigma^-} .$$
 (5.26)

Then, the Virasoro constraints take the form

$$\tilde{L}_m = L_m = 0 \quad \forall m \; . \tag{5.27}$$

In the case of the open string the α^M and $\tilde{\alpha}^M$ are proportional, so we only have to implement $L_m = 0$.

Quantization of Open Strings

Classical strings have a particle spectrum that is completely nonsensical with respect to observation since the effective mass of the string may take any value. A theory of reality must thus be a quantized string theory, which we will now formulate. The starting point is as usual imposing the canonical commutation relations

$$[X^{M}(\tau,\sigma),\Pi^{M}(\tau,\sigma')] = i\eta^{MN}\delta(\sigma-\sigma') , \qquad (5.28)$$
where the canonical momentum $\Pi(\tau, \sigma)$ is defined by

$$\Pi^{M}(\tau,\sigma) = \frac{\partial \mathcal{L}}{\partial(\partial_{t}X^{M})} = \frac{\partial_{t}X^{M}(\tau,\sigma)}{2\pi\alpha'} .$$
(5.29)

We then go to the mode expansions, meaning we have to work through each set of boundary conditions separately. For example, for open strings with NN boundary conditions we find that in terms of the modes the only nonzero commutation relations are

$$[X_0^M, p^N] = i\eta^{MN} \quad [\alpha_m^M, \alpha_n^N] = m\eta^{MN}\delta_{m, -n} .$$
(5.30)

We can then define creation and annihilation operators by

$$a_m^M = \frac{1}{\sqrt{m}} \alpha_m^M \ , \alpha_m^{M\dagger} = \frac{1}{\sqrt{m}} \alpha_{-m}^M \ , \tag{5.31}$$

yielding the relations

$$[a_m^M, a_n^{N\dagger}] = \eta^{MN} \delta_{m,n} , \quad [a_m^M, a_n^N] = [a_m^{M\dagger}, a_n^{N\dagger}] = 0$$
(5.32)

for $m, n \ge 0$. For each M, m the creation operator a_m^M describes the Hilbert space of a harmonic oscillator, except for the case of M = 0 where we find that $[a_m^0, a_m^0^{\dagger}] = -1$, describing states of negative norm. In order to obtain a sensible quantum theory we we need these negative norm states to be ruled out by the Virasoro constraints, and this indeed turns out to be the case.

From now on, let us work in light-cone coordinates, $X^{\pm} = X^0 \pm X^{D-1}$. We denote the directions transverse to the lightcone X^I , $I \in [1, D-2]$. In these coordinates it is straightforward to solve the Virasoro constraints. We can pick

$$X^{+} = x_{0}^{+} + 2\alpha' p^{+} \tau , \qquad (5.33)$$

and the Virasoro constraints will completely determine X^- up to a constant x_0^- in terms of X^+ and the X^I .

This means that the dynamical degrees of freedom are all of the transverse modes $x_0^I, p^I, a_m^I, a_m^{I\dagger}$, as well as p^+ and x_0^- . We can now define the ground states of the quantum string, $|0, k\rangle$ by the action of the transverse oscillators

$$p^{M}|0,k\rangle = k^{M}|0,k\rangle$$
, $a^{I}|0,k\rangle = 0 \ \forall I \in [1,D-2]$. (5.34)

As is apparent the k in $|0,k\rangle$ is hiding a vector index.

Note that since the momenta are not described by harmonic oscillators, there is a continuum of momentum eigenstates that can serve as ground states for the oscillators. A general excited state is obtained by acting on a ground state with creation operators

$$|N,k\rangle = \left[\prod_{I=1}^{D-2} \prod_{n=1}^{\infty} \frac{(a_n^{I\dagger})^{N_{I,n}}}{\sqrt{N_{I,n}}}\right]|0,k\rangle \quad , \tag{5.35}$$

where $N_{I,n}$ denotes the number of I direction excitations of oscillator mode n are in the state $|N,k\rangle$. The N in $|N,k\rangle$ should, like the k not be understood as a single number but as a collection of all the $N_{I,n}$. Despite this, it is convenient to also take the number N to be defined

$$N = n \sum_{I,n} N_{I,n} . (5.36)$$

where the extra factor n is due to the oscillator mode with index n carries n units of energy.

The state $|N,k\rangle$ satisfies

$$a_n^{I\dagger} a_n^{I} | N, k \rangle = N_{I,n} | N, k \rangle \quad , \tag{5.37}$$

making $a_n^{I\dagger}a_n^{I}$ the number operator of the *n*:th mode of the *I*-direction oscillator.

We now wish to impose the Virasoro constraints, and begin by expressing the Virasoro modes in terms of the creation and annihilation operators. Combining equations (5.25) and (5.31) we see that

$$L_{0} = \alpha' p^{M} p_{M} + \sum_{I=1}^{D-2} \sum_{n=1}^{\infty} n \frac{1}{2} (a_{n}^{I\dagger} a_{n}^{I} - a_{n}^{I} a_{n}^{I\dagger})$$

$$= \alpha' p^{M} p_{M} + N + \frac{1}{2} \sum_{I} \sum_{n} n [a_{n}^{I\dagger}, a_{n}^{I}]$$

$$= \alpha' p^{M} p_{M} + N + \frac{D-2}{2} \sum_{n} n .$$

(5.38)

It is customary to name the divergent sum and its forefactor a. Since it is only due to a normalordering ambiguity it could be removed or set to a finite number by redefining the classical expressions so that they are *Weyl ordered* before quantization as in the discussion around equation (3.9). In principle the choice of a is arbitrary but only a specific choice will preserve Lorentz invariance.

We then continue to follow custom and redefine $L_0 \equiv \alpha' p^M p_M + N$ so that the Virasoro constraints read

$$(L_0 + a) |\psi\rangle = 0$$
, $L_m |\psi\rangle = 0$. (5.39)

where

$$a = \frac{D-2}{2} \sum_{n} n'' = " - \frac{D-2}{24}$$
(5.40)

where the final result can be obtained via Riemann-zeta regularization, or more rigorously by constructing the quantum Lorentz generators and demanding the right commutation relations¹. The latter of these is more a derivation of the right choice of the arbitrary parameter a, but the result happens to coincide with the Riemann-zeta result for reasons unknown to man.

The L_0 Virasoro constraint now tells us how to find the mass of the excited string. The mass operator $M^2 = -p^M p_M$ must be given by

$$M^{2} = -p^{M} p_{M} = \frac{1}{\alpha'} \left(N + \frac{2 - D}{24} \right) .$$
 (5.41)

We are now ready to discuss the mass spectrum of the open string. The first observation to make is that the vacuum state $|0,k\rangle$ has

$$\alpha' M^2 = -\frac{D-2}{24} , \qquad (5.42)$$

which is negative for D > 2. Thus, bosonic string theory predicts the existence of tachyons, an indicator that our supposed vacuum is unstable.

Let us continue in spite of this complication and see what we find at the first excited mass level of bosonic open string theory. What we find is a state of the form

$$a_1^I |k,0\rangle$$
, with $\alpha' M^2 = -\frac{D-26}{24}$. (5.43)

This is a field with one vector index and momentum k. In particular, it is a vector boson with D-2 degrees of freedom. A vector field in D dimensions with D-2 degrees of freedom must be

 $^{^{1}}$ cf. section 12.5 in [43], in principle this is similar to how we found the central charge in section 3.4.4.

massless, because massive relativistic fields have D-1 degrees of freedom. More formally, since a^{I} has D-2 independent components and transforms as a *D*-dimensional vector under Lorentz boosts in the target spacetime, it must live in a massless ("short") representation of the Lorentz group. This tells us that unless string theory lives in D = 26 Lorentz invariance is broken. We can now write down some of the field content of open NN strings:

Ν	$ \psi angle$	$\alpha' M^2$	Degrees of freedom
0	0,k angle	-1	1
1	$ a_1^{I\dagger} \ket{0,k}$	0	D-2
2	$a_1^{I\dagger}a_1^{J\dagger}\ket{0,k}$	1	$\frac{(D-2)(D-1)}{2}$
2	$ ~a_{1}^{I\dagger}~ 0,k angle$	1	D-2

Table 5.1: The first few excited states of bosonic open strings with NN boundary conditions. Here D = 26 but it is more transparent to not insert numbers in the degree of freedom count.

Quantization of Closed Strings

The quantization of closed strings is performed in the same way as for open strings. The main difference is that the left- and right handed modes are decoupled, so we have two sets of oscillators a_n^I and \tilde{a}_n^I . This means that the Virasoro modes are now uncoupled as well, and we get two copies of the Virasoro constraints. In addition the mode expansions are just those in equation (5.19) so definitions of the creation/annihilation operators differ by a factor of $\sqrt{2}$ from the open string case.

The Virasoro constraints for the closed string read

$$\hat{L}_{n} |\psi\rangle = L_{n} |\psi\rangle = 0 , \quad n \neq 0,$$

$$(\tilde{L}_{0} - \tilde{a}) |\psi\rangle = (L_{0} - a) |\psi\rangle = 0 ,$$
(5.44)

with analogous definitions to the open string case. The normal-ordering constants are now given by

$$a = \tilde{a} = -\frac{D-2}{12} \ . \tag{5.45}$$

A generic closed string state is now given by $|\tilde{N}, N, k\rangle$, defined by

$$|\tilde{N}, N, k\rangle = \left[\prod_{I=1}^{D-2} \prod_{n=1}^{\infty} \frac{(\tilde{a}_{n}^{I\dagger})^{\tilde{N}_{I,n}} (a_{n}^{I\dagger})^{N_{I,n}}}{\sqrt{N_{I,n}\tilde{N}_{I,n}}}\right] |0, 0, k\rangle \quad ,$$
(5.46)

where $|0,0,k\rangle$ is the vacuum annihilated by all a_n^I , \tilde{a}_n^I and momentum eigenvalue k^M . The mass of the string is given by solving the L_0 and $\tilde{L_0}$ restrictions for p^2 in terms of the other terms and adding the solutions. The mass squared operator is then

$$M^{2} = \frac{2}{\alpha'} \left(\sum_{I=1}^{D-2} \sum_{n=1}^{\infty} n(N_{I,n} + \tilde{N}_{I,n}) + \frac{2-D}{12} \right) .$$
 (5.47)

Subtracting the L_0 - and $\tilde{L_0}$ -Virasoro conditions we find the *level matching condition*

$$0 = (L_0 - \tilde{L}_0 + \tilde{a} - a) |\psi\rangle$$

$$\left\langle \tilde{a} = a , \tilde{p}^M = p^M \right\rangle = \sum_{I=1}^{D-2} \sum_{n=1}^{\infty} n(N_{I,N} - \tilde{N}_{I,N})) |\psi\rangle$$

$$= (N - \tilde{N}) |\psi\rangle$$
(5.48)

where we have defined N and \tilde{N} analogously to the open string case:

$$N \equiv \sum_{I=1}^{D-2} \sum_{n=1}^{\infty} n N_{I,n} , \quad \tilde{N} \equiv \sum_{I=1}^{D-2} \sum_{n=1}^{\infty} n \tilde{N}_{I,n} .$$
 (5.49)

The lowest energy states of the open string are the vacuum

$$|0,0,k\rangle$$
, with $\alpha' M^2 = \frac{2-D}{6}$ (5.50)

and the first excited state

$$a_1^{I\dagger} \tilde{a}_1^{J\dagger} |0,0,k\rangle$$
, with $M^2 = \frac{26 - D}{6\alpha'}$. (5.51)

For D = 26 this describes a transverse rank two tensor. Since the tensor is transverse, Lorentz invariance demands that it is massless. We can decompose this tensor into an antisymmetric tensor B^{MN} , a symmetric traceless tensor g^{MN} and a scalar ϕ . The symmetric tensor g may be identified with a metric perturbation of spacetime, B is the Kalb-Ramond field and ϕ is called the *dilaton*. The choice of this decomposition comes from the requirement that each component should transform as an irreducible representation of the Lorentz group. We will show in section 5.1.3 how exactly the dynamical spacetime metric $G_{\mu\nu}$ and $g_{\mu\nu}$ are related.

5.1.1 String Perturbation Theory

What we have discussed in the previous section is a theory of free non-interacting strings and it would be interesting to extend to the interacting case. Our intuition from QFT tells us that the way to go about this is to add various nonlinear terms to the Polyakov action, but there is no way to do this without breaking our gauge symmetries. We will in this discussion specialize all the details to closed string theory since our main interest is quantum gravity.

Instead, it turns out that the Polyakov action of free strings contains all of the information we need to know about interacting strings. One way to see that this is believable is to consider what a string Feynman diagram would look like;



This diagram describes two open strings joining for a while and then splitting again. This is the analogue of a ϕ^4 scattering process in ordinary scalar QFT. In Feynman diagrams information about interactions is encoded at vertices but the string diagram has no natural place to insert these.

The question of what to compute arises, as it seems we can not amputate the legs and only look at vertices. One thing we could try is to compute the probability of a particular configuration of strings in some particular state propagating into some other configuration in finite time. This would be a very path integral approach where we would stipulate initial curves for the initial two strings, and summing over all intermediate states that end on some other two fixed strings. How to do this is unknown, and and we will argue now that such a quantity would not be an observable anyways. The important observation to make is that all observables of a theory must be *gauge invariant*, and due to conformal invariance the string does not know if its worldsheet has traveled far or not². Therefore a finite-time scattering diagram does not make sense to compute. We can, however compute the string S-matrix by pulling all of the external string states out to infinity.



The point at infinity is left invariant by all infinitesimal (gauge) symmetries of the theory, so a gauge invariant amplitude can be computed. Each of the external string states will map to a point, and assigned a separate free string state with some spacetime momentum p.

Due to the *state operator correspondence* of two-dimensional CFT's as was covered in section 3.4.4 we know that each of the external string states are equivalent to some operator insertion at the endpoints. Inserting the operators and then using conformal symmetry to map the endpoints back to a finite distance we end up with a worldsheet with the topology of a sphere, with an operator insertion for each external string. The operators that correspond to the external string states are called *vertex operators*. To compute scattering amplitudes we then need to sum over all conformally inequivalent sets of operator insertions, a thing we will not discuss in detail.

Let us now consider the path integral of the Polyakov action after Wick rotation to Euclidean space:

$$Z_p = \sum_{\text{all topologies of } \Sigma} \int_{\Sigma} \mathcal{D}X^M \mathcal{D}h_{\alpha\beta} e^{-S_p[X,h]} , \qquad (5.52)$$

where the sum over topologies is because we allow strings to join and split, meaning we have to allow any number of intermediate interactions.

For closed strings, the sum over all topologies is a sum over all two-dimensional oriented surfaces without boundary. For open strings it is a sum over two-dimensional surfaces with a boundary. In both cases the surface is completely characterized by the number of holes in it, referred to as the *genus g*. The sum over topologies then becomes a sum over g, and the only question that remains is how to weight the contributions from different topologies in the string perturbation expansion.

It turns out that we can augment the Polyakov action with an extra term that classically does nothing at all, we define

$$S_{\text{string}} = S_p - \chi , \quad \chi = \lambda \frac{1}{4\pi} \int_{\Sigma} d^2 \sigma \sqrt{h} R$$
 (5.53)

where R is the Ricci scalar of the world sheet. It is possible to check that the new term is reparametrization and Weyl invariant.

²This would be analogous to off-shell correlation functions in QFT such as $\langle \phi(x_1)\phi(x_2)\rangle \sim f(|x-y|)$ for a CFT, which represents a Feynman diagram where all of the external legs may have arbitrary momenta. Since |x-y| is not conformally invariant, local off-shell correlators are not observables in a CFT. This is one of the early clues towards the AdS/CFT duality, any theory of quantum gravity will have symmetries (diffeomorphisms) that render local observables nonsense, so *all* off-shell degrees of freedom must live on the boundary of spacetime which is left invariant under the gravitational symmetries. This is also why we do not know how to handle quantum gravity in de Sitter space, there is no way to send the endpoints to infinity in de Sitter.

Someone who has taken a course in general relativity might recognize χ as the Einstein-Hilbert term describing the dynamics of spacetime. In two dimensions, gravity has no propagating degrees of freedom³ so it has no dynamics, rendering the additional term trivial at the classical level. When we have a nontrivial topology however, χ is the *Euler number*⁴ and it is related to the genus of σ_g by $\chi = 2 - 2g$.

The path integral now has the form

$$\sum_{g} \int_{\Sigma} \mathcal{D}X^{M} \mathcal{D}h \, e^{-S_{\text{string}}[X,h]} = \sum_{g} e^{-2\lambda(1-g)} \int_{\Sigma} \mathcal{D}X^{M} \mathcal{D}h \, e^{-S_{p}[X,h]} \,.$$
(5.54)

We see that $e^{-\lambda}$ takes the role of the string coupling constant. Note that although we have made no approximation in writing this down, the expansion is still an asymptotic expansion, meaning

As long as $e^{\lambda(2g)} \ll 1$ we have a good perturbative expansion. If this condition does not hold, we really have no way of analyzing string theory, since we do not have a nonperturbative formulation.

It is customary to define the string coupling constant as

$$g_s \equiv g_{\text{closed}} = g_{\text{open}}^2 = e^\lambda , \qquad (5.55)$$

where the relation between the open and closed string couplings are due to the fact that the Euler number of a plane with boundaries goes as $\chi = 2 - g$, since when we add a hole to the open string worldsheet we add a boundary and no handles. We are in general not interested in the specifics of high-energy string scattering and will now set up the necessary tools to treat the low-energy approximation of strings where only the massless modes are excited.

5.1.2 String Theory Vertex Operators

We said in the previous section that there are operators that create external string states via insertion at a point. In this section we derive the form of these *vertex operators*, which represent the absorption or emission of a physical string mode $|N,k\rangle$ on some point on the worldsheet. The notation in this section may differ from other sources since we are working in lightcone gauge.

In the case of an open string, the vertex operator must act on the boundary of the worldsheet, while for a closed string it may act somewhere on the interior. The quantum operator on Hilbert space \hat{V}_{ϕ} has no explicit coordinate dependence, so as usual it must be constructed via an integral over all possible insertion points of some $V_{\phi}(\sigma)$ that has coordinate dependence.

For closed and open strings respectively, the vertex operator is then on the form

$$\hat{V}_{\text{closed}} = g_s \int_{\Sigma} d^2 z \, V_{\phi}(z,\overline{z}) \,, \quad \hat{V}_{\text{open}} \sqrt{g_s} \oint_{\delta \Sigma} ds \, V_{\phi}(s) \,, \qquad (5.56)$$

where ϕ labels the specific state being emitted or absorbed at the vertex and Σ denotes the string worldsheet. We will in this section focus on the closed string due to our interest in gravity.

The vertex operator must be invariant under the Weyl+diffeomorphism symmetry of the string worldsheet. The integration measure d^2z has conformal dimension (-1, -1), so $V_{\phi}(z, \overline{z})$ must have conformal weight (1,1). In addition we know from section 3.4.4 that the vacuum is obtained by letting $\hat{V} = \mathbb{1}$. Then, the excited string states must be produced by some appropriate representation of the Fock-space operators of section 5.1.

³We saw this when we were able to pick the conformal gauge.

⁴The Euler numbers of the closed- and open string surfaces are given by the Gauss-Bonnet theorem as 2 - 2h - b where h is the number of handles and b the number of boundaries.

The bosonic string vacuum with momentum k is generated by

$$|0,0,k\rangle = e^{ik \cdot X} |0,0,0\rangle$$
, (5.57)

so the vertex operator should contain a factor : $e^{ik \cdot X}$:. This exponential has to be normal-ordered, since the vertex operator should be defined in a normal-ordered string correlation function. In the following we will leave the normal-ordering implicit. The operator $e^{ik \cdot X}$ has conformal dimension $\alpha'k$ for open strings and $(\alpha'k^2/4, \alpha'k^2/4)$ for closed strings⁵.

The tachyon is the ground state for the bosonic string theory, and therefore has the vertex operator

$$\hat{V}_{\text{tachyon}} \sim \int \mathrm{d}^2 \sigma \, e^{ikX} \;.$$
 (5.58)

Weyl invariance requires the integral to have weight (0,0), but the operator e^{ikX} has weight $(\alpha'k^2/4, \alpha'k^2/4)$. Since the integration measure has weight (-1, -1) this is only consistent if

$$\alpha' k^2 / 4 = 1 \Rightarrow \alpha' M_{\text{tachyon}}^2 = -4 , \qquad (5.59)$$

using that $k^2 = -M^2$. This is exactly the mass of the closed string tachyon in D = 26 dimensions as we saw in equation (5.50), so our definitions so far seem self-consistent.

For general excited states we know that we will obtain on-shell string states by acting on the ground state with creation operators, according to

$$|\tilde{N}, N, k\rangle = \left[\prod_{I=1}^{D-2} \prod_{n=1}^{\infty} \frac{(\tilde{a}_{n}^{I\dagger})^{\tilde{N}_{I,n}} (a_{n}^{I\dagger})^{N_{I,n}}}{\sqrt{N_{I,n}\tilde{N}_{I,n}}}\right] |0, 0, k\rangle , \qquad (5.60)$$

where the level matching condition requires $\tilde{N} = N$. Since the creation operators are the Fourier modes of X, we have the identity

$$\alpha_n^{I\dagger} = \frac{1}{\pi} \oint z^{-n} \partial X^I \mathrm{d}z \;, \tag{5.61}$$

which suggests that we can simply replace $\alpha_n^{M\dagger}$ with the residue

$$\alpha_n^{I\dagger} \to \frac{2i}{(m-1)!} \partial^m X^I , \quad n > 0 .$$
(5.62)

According to this proposal a closed string vertex operator is then given up to normalization by

$$\hat{V}_{\phi} = \int \mathrm{d}^2 \sigma \,: \prod_i \overline{\partial}^{m_i} X^{I_i}(\overline{z}) \prod_i \partial^{n_j} X^{J_j}(z) e^{ikX(z,\overline{z})} : \xi_{IJ} \,\,, \tag{5.63}$$

where the $\overline{\partial}$ on the first X is due to the fact that it comes from the antiholomorphic set of oscillators \tilde{a} in equation (5.60). The contraction with the constant two component tensor ξ_{IJ} is necessary since the operator \hat{V}_{ϕ} is supposed to be an invariant. The excited states should fulfill the level matching condition, and the ground state has to fulfill the Virasoro constraints to ensure that the vertex operator has conformal dimension (0,0).

⁵We prove this claim for closed strings in the next subsection.

Let us now explicitly write out the vertex operators of the massless excited states of the bosonic closed string, that is, the Kalb-Ramond, metric and dilaton fields. We have

$$\hat{V}_h = \int d^2 \sigma : \overline{\partial} X^I(\overline{z}) \partial X^J(z) e^{ikX(z,\overline{z})} : h_{IJ} , \qquad (5.64)$$

$$\hat{V}_B = \int d^2 \sigma : \overline{\partial} X^I(\overline{z}) \partial X^J(z) e^{ikX(z,\overline{z})} : B_{IJ} , \qquad (5.65)$$

$$\hat{V}_{\phi} = \int \mathrm{d}^2 \sigma : \overline{\partial} X^I(\overline{z}) \partial X_I(z) e^{ikX(z,\overline{z})} : \phi , \qquad (5.66)$$

where h_{IJ} is a symmetric traceless constant tensor, and B_{IJ} is an antisymmetric constant tensor. These are written in the conformal gauge, as well as in light-cone coordinates. Since the massless states have $\alpha' k^2 = 0$, the term : e^{ikX} : has conformal dimension (0,0). The restrictions on the $p^$ component of the spacetime momentum ensures that the normal-ordered operator is a conformal primary with only transverse degrees of freedom. Note that the trace in the dilaton vertex operator is only over transverse coordinates. In section 5.4.1 in [44] there is an analysis of vertex operators in covariant quantization.

Proof that e^{ikX} has conformal weight $(\alpha'k^2/4, \alpha'k^2/4)$

In the preceding section, we claimed that that e^{ikX} has conformal weight $(\alpha'k^2/4, \alpha'k^2/4)$. We will now show this in complex coordinates z, \overline{z} .

We need to build a fair bit of machinery to show this, which is why it gets its own section. The stress energy in the quantum theory is

$$T(z) = -\frac{1}{\alpha'} : \partial X \partial X :$$
(5.67)

where the : denote normal-ordering. Usually, we would define normal-ordering by putting all annihilation operators to the right, but if we wish to not make any reference to creation and annihilation operators, an equivalent definition is

$$\partial X \partial X := -\lim_{z \to w} (\partial X(z) \partial X(w) - \langle \partial X(z) \partial X(w) \rangle$$
(5.68)

guaranteeing that the VEV is $\langle : \partial X \partial X : \rangle = 0$. We can then show that ∂X is a primary field of weight 1. We consider the OPE of $T(z)\partial X(w)$

$$T(z)\partial X(w) = -\frac{1}{\alpha'} : \partial_z X \partial_z X : \partial_w X .$$
(5.69)

The lefthand side lives inside a correlation function so it is time-ordered, while the righthand side is just a product of normal-ordered objects. For the definition in equation (5.69) to make sense we need to make a time-ordered expression of the LHS. The way to make a time-ordered expression out of a normal-ordered one is given by Wick's theorem, we should sum over all possible pairwise Wickcontractions of the operators in question. This means that we replace the pair by the appropriate propagator

$$\partial \overline{X(z)} \partial \overline{X}(w) = -\frac{\alpha'}{2} \frac{1}{(z-w)^2} , \qquad (5.70)$$

coming from the fact that $X(z)X(y) = \ln(z-y)$ in two dimensions⁶, and hitting twice with derivatives. Then, we have

$$T(z)\partial X(w) = -\frac{2}{\alpha'}\partial X(z)\left(\frac{\alpha'}{2}\frac{1}{(z-w)^2} + \text{finite term}\right)$$
(5.71)

⁶The Xs have the equation of motion $\partial^2 X = \delta(x)$, which is solved in Fourier space by the Green's function $G_F \sim \int d^2 k \frac{e^{ikx}}{k^2}$.

Here, the finite term is the totally normal-ordered part : $T(z)\partial X(w)$: that has VEV 0. In general, the Ward identities ensure that all correlators with T depend only on the residues of the OPE, meaning also that these also fully determine the transformation properties.

Now, we can Taylor expand $\partial X(z)$ about z = w and we obtain

$$T(z)\partial X(w) = \frac{\partial X}{(z-w)^2} + \frac{\partial^2 X(w)}{(z-w)} + \text{finite terms}$$
(5.72)

which is the OPE expected of a primary operator of weight 1. Let us now consider the OPE of $\partial X(z)$ with $:e^{ikX}:$

$$\partial X(z) : e^{ikX(w)} := \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \partial X : X(w)^n :$$

= $\sum_{n=1}^{\infty} \frac{(ik)^n}{(n-1)!} : X(w)^{n-1} : \left(-\frac{\alpha'}{2}\frac{1}{z-w}\right) + \dots$ (5.73)
= $-\frac{\alpha' k}{2} : \frac{e^{ikX(w)}}{z-w} : + \dots$

where we have used $\partial X(z)X(w) = -\left(\frac{\alpha'}{2}\frac{1}{z-w}\right)$, with *n* possible contractions, and ... denotes fully normal-ordered terms.

Using the result in equation (5.73) we find that

$$T(z): e^{ikX} := -\frac{1}{\alpha'}: \partial X(z)\partial X(z):: e^{ikX}:$$

$$= -\frac{1}{\alpha'} \left(-\frac{\alpha'^2 k^2}{4} \cdot \frac{e^{ikX}}{z-w} - i\alpha' k \frac{\partial X(z)e^{ikX}}{z-w} + \text{finite} \right)$$

$$= \frac{\alpha' k^2}{4} \cdot \frac{e^{ikX}}{z-w} + ik \frac{\partial X(z)e^{ikX}}{z-w} + \text{finite terms}.$$
(5.74)

Here, the finite terms are due to the fully normal-ordered object : $T(z)e^{ikx}$:, the first terms is obtained by two contractions of ∂X with : e^{ipx} : and the second term comes from a single contraction, where we have to keep the normal-ordered remnant because it contributes to the residue. Taylor expanding ∂X around w and taking ∂_z as a derivative with respect to w we can write this as

$$T(z): e^{ikX} := \frac{\alpha'k^2}{4} \frac{e^{ikX}}{z-w} + \frac{\partial_w : e^{ikX}}{z-w} + \text{finite terms}$$
(5.75)

and we see that this indeed is the OPE for a conformal primary operator of weight $(\alpha' k^2/4, \alpha' k^2/4)$. This follows because the analysis for \overline{z} is exactly the same.

5.1.3 Emergence of Gravity as a Massless Excitation

Let us consider what the Polyakov action would look like on a curved spacetime. This just means that we should allow for a non-flat G_{MN} in the Polyakov action:

$$S_P = -\frac{1}{4\pi\alpha'} \int d^2\sigma \sqrt{-h} h^{\alpha\beta} \partial_\alpha X^M \partial_\beta X^N G_{MN} . \qquad (5.76)$$

There are two problems with just writing this expression down. First off, we relied on the flat spacetime metric to quantize our string theory. In addition to this, the bosonic string theory already contains a graviton field, and it would be very strange indeed to have *two* graviton fields. Let us now demonstrate (working in Euclidean signature) that these two fields are the same by expanding G_{MN} around flat space

$$G_{MN} = \eta_{MN} + H_{MN}(X) . (5.77)$$

The string partition function is then

$$Z = \sum_{\text{topo}} \int \mathcal{D}X \mathcal{D}h \, e^{-S_{\text{String}}-V} = \sum_{\text{topo}} \int \mathcal{D}X \mathcal{D}h \, e^{-S_{\text{String}}} \left(1 - V + \frac{1}{2}V^2 + \dots\right)$$
(5.78)

where

$$V = \frac{1}{4\pi\alpha'} \int d^2\sigma \sqrt{h} h^{\alpha\beta} \partial_{\alpha} X^M \partial_{\beta} X^N H_{MN}(X) , \qquad (5.79)$$

which is just the vertex operator associated to the graviton state of the bosonic string, as we showed in equation (5.64). The diffeomorphism symmetry in a theory of gravity⁷ allows us to pick the traceless transverse gauge for the metric perturbation⁸ for H, setting $H_{MN} = H_{IJ}$ where IJ are the transverse indices. We then construct light cone coordinates out of the other two directions, and go to complex Euclidean coordinates to obtain

$$V = \frac{1}{2\pi\alpha'} \int d^2 z \,\partial X^I \overline{\partial} X^J H_{IJ}(X) \,. \tag{5.80}$$

Now, identifying $H_{IJ} = e^{ikx}h_{IJ}$ where h_{IJ} is the constant symmetric traceless tensor in equation (5.64) we see that $V \sim \hat{V}_h$, i.e. it is (up to a constant) the graviton vertex operator.

With this, we know that inserting a single factor of \hat{V}_h into the path integral generates a single graviton state. Inserting e^{-V} corresponds to a coherent state of gravitons, reproducing equation (5.78). Thus the traceless symmetric massless vertex operator of string theory *really* is the generator of gravity.

In a similar way, it can be shown that the dilaton and Kalb-Ramond vertex operators when exponentiated add to the Polyakov action

$$e^{\hat{V}_B}e^{\hat{V}_\phi}: S_P \to S_P + S_{B,\phi} = S_P + \int d^2\sigma \ \sqrt{-h} \left(\epsilon^{\alpha\beta}\partial_\alpha X^M \partial_\beta X^N B_{MN} + \alpha' R_h \phi(X)\right)$$
(5.81)

where we have, without proving that this is the right answer, undone the choice of conformal and lightcone gauge, and the fields B_{MN} and ϕ are defined to contain the factor e^{ikx} . Notably, we see right away that the dilaton term resembles the χ term which we used to augment the Polyakov action. It turns out that in fact, the string coupling is given by

$$g_s = e^{\langle \phi \rangle} , \qquad (5.82)$$

where $\langle \phi \rangle$ is the VEV of the dilaton field.

We see that in general by inserting exponentiated vertex operators in the string path integral, we are able to able to turn on *background fields* in the Polyakov action. The \hat{V}_h turns on gravity, \hat{V}_B turns on the Kalb-Ramond Field and \hat{V}_{ϕ} turns on the dilaton field. The fact that these look like classical additions to the Polyakov action is the very definition of the exponents e^{-V} creating coherent states.

 $^{^{7}\}mathrm{By}$ writing the action with a general curved metric, the Polyakov action is manifestly general covariant in target spacetime coordinates.

⁸This is usually covered in a lecture about gravitational waves in a GR course c.f. chapter 6 in [45].

Einstein's Equations from the α' -Expansion

In the previous section we found that the traceless symmetric part of the massless states of closed bosonic string theory generate gravity. We still have no reason to believe that this stringy quantum gravity actually obeys Einstein's equations in the classical limit. Our tool for finding this result is the so-called α' -expansion, which let us formalize when we can use perturbation theory to analyze the worldsheet CFT of string theory. Note that the α' and g_s expansions are two distinct expansions, due to this it is said that string theory has two expansions. One expansion in the genus of the worldsheet (g_s) , and one in the shape of the worldsheet (α') .

The basic recipe is to renormalize the worldsheet CFT in terms of the embedding coordinates X with the background fields with G_{MN} acting as the coupling constants of the theory. Demanding that worldsheet conformal symmetry be conserved on the quantum level is equivalent to demanding the vanishing of the β -functions arising from renormalization. We will then find that the vanishing of the β -functions in a CFT implies that the target spacetime metric G_{MN} must obey the vacuum Einstein equations. This demonstrates that quantum string theory contains classical Einstein gravity in an appropriate limit.

Let us consider the Polyakov action in conformal gauge, on a curved spacetime

$$S = \frac{1}{4\pi\alpha'} \int d^2\sigma \, G_{MN}(X) \partial_\alpha X^M \partial^\alpha X^N \,. \tag{5.83}$$

We can expand the action about a classical solution, for example a string just sitting at a point X_0^{μ} , letting

$$X^M(\sigma) = X_0^M + \sqrt{\alpha'} Y^M(\sigma) , \qquad (5.84)$$

where Y^M represents a dynamical fluctuation which we assume to be small. To have [Y] = 0 so that statements such as $Y \ll 1$ make sense we have added a factor α' to compensate for the fact that [X] = -1. We can now expand the Lagrangian, finding

$$G_{MN}(X)\partial_{\alpha}X^{M}\partial^{\alpha}X^{N} = \alpha' \left[G_{MN}(X_{0}) + (\nabla_{L}G_{MN}(X_{0}))\sqrt{\alpha'}Y^{L} + (\nabla_{K}\nabla_{L}G_{MN}(X_{0}))\frac{\alpha'}{2}Y^{L}Y^{K} + \dots \right] \partial_{\alpha}Y^{M}\partial^{\alpha}Y^{N}$$
(5.85)

where the ∇_K are general covariant derivatives. Each of coefficients in the Taylor expansion of G_{MN} about the classical solution X_0 determines the coupling constants for an interaction term for the fluctuating Y-fields. The theory has an infinite number of terms, and therefore an infinite number of coupling constants, all of which depend on the function G_{MN} .

We would like to know when the theory is weakly coupled, so that we can do perturbation theory. This requires the whole infinite set of coupling constants to be small, a seemingly daunting thing to check, but we can perform a crude analysis. Let us suppose that the target space has a characteristic radius of curvature, so that

$$\frac{\partial G}{\partial X} \sim \frac{1}{r_c} \ . \tag{5.86}$$

Since [G] = -2 and [X] = -1 we have $[r_c] = -1$. The effective dimensionless coupling in equation (5.85) is given by

$$g_{eff} \sim \frac{\sqrt{\alpha'}}{r_c} ,$$
 (5.87)

since each derivative on G gives a factor $1/r_c$ and each order in Y gives a factor $\sqrt{\alpha'}$.

This implies that we can use perturbation theory to study the CFT in equation (5.85) as long as the target spacetime metric varies only on scales much greater than $\sqrt{\alpha'}$, i.e. when the radius of curvature is significantly larger than the string length. In particular, this means that on spacetimes containing a black hole singularity, the world-sheet CFT is strongly coupled and perturbation theory breaks down, but the theory is perturbative on an AdS background with large radius of curvature.

String Renormalization

The perturbative theory in the dimensionless coupling $\frac{\sqrt{\alpha'}}{r_c}$ that we defined in equation (5.85) is manifestly conformally invariant in the worldsheet coordinates, at least at the classical level. Let us now renormalize the theory, and see what conformal invariance requires of the β -functions. This is in analogy with the renormalization group approach we introduced in section 3.3.

In the current theory, we have an infinite number of couplings, but they all depend on the same function $G_{\mu\nu}(X)$. It is therefore reasonable to speak of a β functional $\beta(G)$ that schematically should be of the form

$$\beta_{\mu\nu}(G) \sim \frac{\mathrm{d}G_{\mu\nu}(X;\Lambda)}{\mathrm{d}\Lambda} ,$$
 (5.88)

where Λ represents some energy cutoff scale. Conformal invariance is only quantum mechanically preserved if

$$\beta_{\mu\nu}(G) = 0$$
 . (5.89)

Let us now compute the β functional for the α' model at one loop. We will impose a dimensional regularization scheme to isolate the one-loop divergence and subtract off a counterterm to cancel it. The β functional vanishes if the counterterm vanishes.

To perform this analysis, we import some knowledge about general relativity. Diffeomorphism invariance lets us pick *Riemann normal coordinates*⁹ such that the expansion in $X^M = X_0^M + \sqrt{\alpha'}Y^M$ gives

$$G_{MN}(X) = \delta_{MN} - \frac{\alpha'}{3} \mathcal{R}_{MLNK}(X_0) Y^L Y^K + \mathcal{O}(Y^3) , \qquad (5.90)$$

where \mathcal{R}_{MLNK} is the spacetime Riemann curvature tensor. Then, to order Y^4 the α' action (equation (5.85)) becomes

$$S = \frac{1}{4\pi} \int d^2 \sigma \,\partial_\alpha Y^M \partial^\alpha Y_M - \frac{\alpha'}{3} \mathcal{R}_{MLNK} Y^L Y^K \partial_\alpha Y^M \partial^\alpha Y^N \,. \tag{5.91}$$

This is a field theory with a typical kinetic term plus a quartic interaction. The Feynman rule in Fourier space is

 $\sim \alpha' \mathcal{R}_{MLNK} k_{\alpha}^{M} k^{\alpha N}$ (5.92)

where k_{α}^{M} is the world-sheet momentum in the α directions for the scalar field Y^{M} at index M. The propagator in momentum space is

$$\langle Y^M(\sigma)Y^N(\sigma')\rangle = 2\pi\delta^{MN} \int \frac{\mathrm{d}^2k}{(2\pi)^2} \frac{e^{ik(\sigma-\sigma')}}{k^2} \,. \tag{5.93}$$

⁹This is essentially just the equivalence principle. We can always pick coordinates such that the Christoffel symbols vanish near X_0 , killing off any first-order variations in the metric as a function of position.

The divergence in the theory comes from the one-loop snail diagram

$$= \int \frac{\mathrm{d}^2 k}{(2\pi)} \frac{\mathcal{R}_{MLNK} k^M_\alpha k^{\alpha N}}{k^2} , \qquad (5.94)$$

where σ, σ' are the starting-and ending positions of the loop. We see clearly that this diagram is horribly divergent. To make sense of the divergence let us first renormalize the loop propagator, then we will fix the divergence due to the vertex by redefining the coupling G_{MN} .

To regularize the divergence we introduce dimensional regularization, working in $d = 2 + \epsilon$ worldsheet dimensions, where ϵ is small. Then the propagator becomes

$$\langle Y^{M}(\sigma)Y^{N}(\sigma')\rangle = 2\pi\delta^{MN} \int \frac{\mathrm{d}^{2+\epsilon}k}{(2\pi)^{2+\epsilon}} \frac{e^{ik(\sigma-\sigma')}}{k^{2}} \to \frac{\delta^{MN}}{\epsilon} \text{ as } \sigma \to \sigma' .$$
 (5.95)

To cancel the divergence in the propagator as well as the loop, we make the replacement

$$\mathcal{R}_{MLNK}Y^LY^K\partial_{\alpha}Y^M\partial^{\alpha}Y^N \to \mathcal{R}_{MLNK}Y^LY^K\partial_{\alpha}Y^M\partial^{\alpha}Y^N - \frac{1}{\epsilon}\mathcal{R}_{MN}\partial_{\alpha}Y^M\partial^{\alpha}Y^N , \qquad (5.96)$$

where $\mathcal{R}_{MN} = \mathcal{R}_{MLNK} \delta^{LK} + \mathcal{O}(Y^5)$ is the Ricci tensor. It is possible to show that this can be absorbed by the wavefunction renormalization $Y^M \to Y^M + \frac{\alpha'}{6\epsilon} \mathcal{R}^M{}_N Y^N$ together with

$$G_{MN} \to G_{MN} + \frac{\alpha'}{\epsilon} \mathcal{R}_{MN}$$
 (5.97)

If the quantum theory on the worldsheet should retain its conformal invariance, the coupling constant G must be independent of the scale parameter ϵ , which is only possible if

$$\beta_{MN}(G) = \alpha' \mathcal{R}_{MN} = 0 . \tag{5.98}$$

In other words, the requirement for the preservation of conformal invariance of the string worldsheet theory is that the target spacetime obeys the vacuum Einstein equations with no cosmological constant.

Turning on all the Background Fields

In the previous section we understood how the string couples to a background metric generated by a coherent graviton state. We also know that the massless sector of the bosonic string contains two more background fields, described by the action

$$S_{\text{eff}} = S_G + \int d^2 \sigma \,\sqrt{-h} \left(\epsilon^{\alpha\beta} \partial_\alpha X^M \partial_\beta X^N B_{MN} + \alpha' R_h \phi(X) \right) \,, \tag{5.99}$$

where S_G is the gravitational action for which we just did perturbation theory. The field B_{MN} is the string analogue of the Maxwell field A_M in that it couples to the string worldsheet in the same way as the Maxwell field couples to the particle worldline. Just like in the case of Maxwell theory, the *B* field has a corresponding field strength

$$\partial_{[K}B_{MN]} = H_{KMN}$$
 or, using differential forms $H = dB$. (5.100)

The three form H fulfills the same role as the torsion of general relativity, meaning that string gravity is not generally torsion free.

The one-loop β function(al)s for this theory are given by

$$\beta_{MN}^G(G) = \alpha' \left(\mathcal{R}_{MN} + 2\nabla_M \nabla_N \phi - \frac{1}{4} H_{MLR} H_N^{LR} \right)$$

$$\beta_{MN}^B(G) = \alpha' \left(-\frac{1}{2} \nabla^L H_{LMN} + \nabla^L \phi H_{LMN} \right)$$

$$\beta^{\phi}(G) = \alpha' \left(\frac{D - 26}{6\alpha'} - \frac{1}{2} \nabla^2 \phi + \nabla_M \phi \nabla^M \phi - \frac{1}{24} H_{MNL} H^{MNL} \right) .$$
(5.101)

Right away we see that if we turn off the Kalb-Ramond field we must have D = 26 and g_{MN} must satisfy the vacuum Einstein field equations.

Low Energy Effective Action

We can take an alternative viewpoint to this discussion about β -functions; we postulate that "for consistency, our string theory must remain conformally invariant even in the quantum case"¹⁰. Then, by definition the background fields must fulfill the requirement that all of the β -functions in equation (5.101) vanish. We *define* the vanishing of the β -functions to be the equations of motion for the background fields in a consistent quantum string theory. The *low-energy effective action* is then the action that has the vanishing of the β -functions as its equations of motion. The action is called low-energy because:

- no higher energy massive string states are in the background,
- we assume that $\sqrt{\alpha'}/r_c$ is small,
- we work at one loop in perturbation theory to determine the β -functions.

The low-energy effective action of bosonic string theory is given by

$$S = \frac{1}{2\tilde{\kappa}^2} \int \mathrm{d}^D X \sqrt{-g} e^{-2\phi} \left(\mathcal{R} + 4\nabla_M \phi \nabla^M \phi - \frac{1}{12} H_{MNR} H^{MNR} - \frac{2(D-26)}{3\alpha'} + \mathcal{O}(\alpha') \right) .$$
(5.102)

Note that the kinetic term for the dilaton ϕ has the wrong sign, and in addition, the term $\sqrt{-g}e^{-2\phi}\mathcal{R}$ is not *quite* the usual Einstein-Hilbert term. By a field redefinition we can obtain the *Einstein-frame* effective action (in contrast to equation (5.102), which is the *string frame* effective action). The Einstein-frame action is

$$S = \frac{1}{2\kappa^2} \int \mathrm{d}^D X \sqrt{\tilde{G}} \left(\tilde{\mathcal{R}} - \frac{4}{D-2} \nabla_M \tilde{\phi} \nabla^M \tilde{\phi} - \frac{1}{12} e^{-\frac{8}{D-2}\tilde{\phi}} H^2 - \frac{2(D-26)}{3\alpha'} e^{\frac{4}{D-2}\tilde{\phi}} \mathcal{O}(\alpha') \right)$$
(5.103)

and it is obtained by the following substitutions

$$\phi = \phi - \phi_0$$

$$\tilde{G}_{MN}(X) = e^{-4\tilde{\phi}/(D-2)} G_{MN} ,$$
(5.104)

where $\phi_0(=\langle \phi \rangle)$ is the constant mode of ϕ , usually taken to be the the asymptotic value at infinity. Here, $\kappa = \kappa_0 e^{\phi_0} = \kappa_0 g_s$. The biggest trouble in checking the substitution is computing the new Ricci scalar \tilde{R} from \tilde{G}_{MN}

¹⁰This is sometimes referred to as the *Weyl anomaly free* method of finding the low energy effective action since we are demanding that there is no Weyl anomaly breaking the conformal symmetry after renormalization.

5.2 Superstring Theory

We saw in the previous section that bosonic string theory is a quantum theory that contains gravity and seems in general not to be divergent. Despite these attractive features two big problems remain; the theory contains tachyons and it contains no fermions. The real world has no tachyons and lots of fermions, so something seems to be amiss. In the following we will detail the Ramond-Neveu-Schwarz (RNS) formalism of superstrings.

Superstring theory is the minimal supersymmetric extension of the bosonic string theory of the previous section. In conformal gauge the super-Polyakov action reads

$$\mathcal{S} = -\frac{1}{4\pi\alpha'} \int \mathrm{d}^2\sigma \,\eta^{\alpha\beta} \left(\partial_\alpha X^M \partial_\beta X^N + i\overline{\Psi}^M \gamma_\alpha \partial_\beta \Psi^N \right) g_{MN}(X) \,. \tag{5.105}$$

Just as in the bosonic case, we consider at first a target spacetime that is flat, i.e. $g_{MN} = \eta_{MN}$. The Ψ^M are spinors on the worldsheet, and by picking a nice basis of the two-dimensional gamma matrices, we can pick them to be Majorana spinors, $\Psi^M = (\psi^M_-, \psi^M_+)^T$ with two real components. A possible representation of the worldsheet gamma matrices is

$$\gamma^{0} = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \ \gamma^{1} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}.$$
(5.106)

The fermionic part of the super Polyakov action can then be written as

$$S_F = \frac{i}{4\pi\alpha'} \int d^2\sigma \left(\psi^M_- \partial_+ \psi_{-,M} + \psi^M_+ \partial_- \psi_{+,M}\right) \,. \tag{5.107}$$

The equations of motion then take the same form as in the bosonic sector

$$\partial_{-}\psi_{+}^{M} = \partial_{+}\psi_{-}^{M} = 0 , \qquad (5.108)$$

which describes left- and right moving waves. In two dimensions these equations are known as the Weyl conditions, so the spinors are often referred to as Majorana-Weyl spinors.

The action is invariant under the supersymmetry transformations¹¹

$$\delta_{\epsilon} X^{M} = \bar{\epsilon} \Psi^{M} , \quad \delta_{\epsilon} \Psi^{M} = \gamma^{\alpha} \partial_{\alpha} X^{M} , \qquad (5.109)$$

where the parameter ϵ is a constant infinitesimal Majorana spinor and $\overline{\epsilon} = i\epsilon^{\dagger}\gamma^{0}$ denotes Dirac conjugation. In addition to the equations of motion, the variation of the action gives rise to boundary terms since the variation of the derivative terms requires an integration by parts

$$\delta S_F = \frac{i}{4\pi\alpha'} \int d\tau \left(\psi^M_- \delta \psi_{-M} - \psi^M_+ \delta \psi_{+M} \right) \Big|_{\sigma=0}^{\sigma=\pi} .$$
(5.110)

The analysis of the bosonic part of the string action is exactly the same as in the bosonic case, so let us now look at the fermionic parts of open and closed superstrings.

¹¹Note that this supersymmetry is only a supersymmetry on the worldsheet. That the worldsheet supersymmetry implies spacetime supersymmetry is nontrivial, and performing this analysis using the manifestly spacetime supersymmetric *Green-Schwarz* action requires the introduction of a lot of machinery which we will not reuse.

Open Superstrings

For open strings the boundary conditions $\sigma = 0$ and $\sigma = \pi$ need to vanish separately since they do not define the same point. The action does not care about the overall sign of the full spinor Ψ^M , so we can without loss of generality set

$$\psi^M_+(\tau,0) = \psi^M_-(\tau,0) . \tag{5.111}$$

The choice of the relative sign at the other end of the string then give rise to the *Ramond* or *Neveu-Schwarz* sectors

R:
$$\psi_{+}^{M}(\tau,\pi) = \psi_{-}^{M}(\tau,\pi)$$

NS: $\psi_{+}^{M}(\tau,\pi) = -\psi_{-}^{M}(\tau,\pi)$
(5.112)

The corresponding mode expansions are given by

$$\mathbf{R:} \ \psi^{M}_{\pm}(\tau,\sigma) = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} d_{n}^{M} e^{-in\sigma_{\pm}}$$
$$\mathbf{NS:} \ \psi^{M}_{\pm}(\tau,\sigma) = \frac{1}{\sqrt{2}} \sum_{r \in \mathbb{Z} + \frac{1}{2}} b_{r}^{M} e^{-ir\sigma_{\pm}}$$
(5.113)

where the Fourier modes d_n, b_r are Grassman-valued for consistency with the Grassman-valued full spinor. To quantize the Neveu-Schwarz sector promote the Fourier modes to operators with the anticommutation relations

$$\{d_m^M, d_n^N\} = \eta^{MN} \delta_{m, -n} , \quad \{b_r^M, b_s^N\} = \eta^{MN} \delta_{r, -s} .$$
(5.114)

We define positively moded operators to be annihilation operators, and the negatively moded operators are creation operators. Then, the vacuum $|0\rangle$ is defined to be annihilated by all positively moded operators. There is a complication regarding the Ramond vacuum, since the operator d_0^M changes the state without changing its energy, and cannot be defined to annihilate it. The Ramond vacuum is therefore degenerate. Since $\{d_0^M, d_0^N\} = \eta^{MN}$ is up to a normalization the Clifford algebra, and the vacuum remains a vacuum under the action of this algebra, the vacuum state $|0,k\rangle_R$ of the Ramond sector must be a spacetime spinor with $2^{D/2}$ components. In some dimensions these can be divided left- and right handed components, and we can define a chirality operator Γ^{11} analogously to $\gamma 5$ in four dimensions.

Going to lightcone gauge, we find that $\Psi^+ = 0$ and that Ψ^- is determined in terms of the transverse coordinates to impose the Virasoro constraints. Then the mass-squared operators (before normal-ordering, and including the bosonic creation operators a_n^I) are given by

$$\mathbf{R:} \ \alpha' M^2 = \frac{1}{2} \left(\sum_{n \in \mathbb{Z}} n a_{-n}^I a_n^I + \sum_{m \in \mathbb{Z}} m d_{-m}^I d_m^I \right)$$

$$\mathbf{NS:} \ \alpha' M^2 = \frac{1}{2} \left(\sum_{n \in \mathbb{Z}} n a_{-n}^I a_n^I + \sum_{r \in \mathbb{Z} + \frac{1}{2}} r b_{-r}^I b_r^I \right),$$

$$(5.115)$$

where the sum over the transverse index I is left implicit. To normal-order, we obtain for the bosonic part as in equation (5.38)

$$\frac{1}{2} \sum_{n \in \mathbb{Z}} n a_{-n}^{I} a_{n}^{I} = \sum_{n=1}^{\infty} n a_{-n}^{I} a_{n}^{I} - \frac{D-2}{24} , \qquad (5.116)$$

for the fermionic sectors

$$\mathbf{R:} \ \frac{1}{2} \sum_{m \in \mathbb{Z}} m d^{I}_{-m} d^{I}_{m} = \sum_{m=1}^{\infty} m d^{I}_{-m} d^{I}_{m} + \frac{D-2}{24}$$
$$\mathbf{NS:} \ \frac{1}{2} \sum_{r \in \mathbb{Z} + \frac{1}{2}} r b^{i}_{-r} b^{i}_{r} = \sum_{r=1/2}^{\infty} r b^{I}_{-r} b^{i}_{r} - \frac{D-2}{48} \ .$$
(5.117)

where all of the explicit terms $\sim (D-2)$ come from the analytic continuation of divergent sums. Notably, the normal-ordering constant for the Ramond fermions exactly cancels the bosonic ordering constant, meaning $|0,k\rangle_R$ is a massless spacetime spinor. This means that the Ramond vacuum splits into two sets of chiral spinors that do not talk to each other.

For the Neveu-Schwarz sector we have the normal-ordering constant

$$a_{NS} = -\frac{D-2}{48} - \frac{D-2}{24} = -\frac{D-2}{16} , \qquad (5.118)$$

so the ground state of the NS sector is tachyonic. The lowest energy excitation in the NS sector is a transverse vector, is obtained by $b_{-1/2} |0,k\rangle_{NS}$ and has mass squared

$$\alpha' M^2 = \frac{1}{2} - \frac{D-2}{16} . (5.119)$$

We expect a transverse vector excitation to be massless so we find that the critical dimension for superstring theory is given by D = 10. Just like in the bosonic case, a more rigorous derivation of the critical dimension is found by requiring that the Lorentz group is correctly generated by the appropriate conserved charges of the quantum theory.

We have yet to see the promised absence of the tachyon. This comes about due to the GSO projection, a necessary truncation of the spectrum to preserve supersymmetry. 'Deriving' the GSO projection requires continuing without the truncation, studying the superstring partition function and demanding that (super)conformal symmetry is intact. To perform the GSO projection, we introduce the G-parity operator G. In the NS-sector it is given by

$$G_{\rm NS} = (-1)^{F+1}$$
, with $F = \sum_{r=1/2}^{\infty} b_{-r}^I b_r^I$, (5.120)

where F counts the number of world-sheet fermions. As is apparent, the G-parity operator specifies whether there are an odd or even number of worldsheet fermions, with G = +1 for an odd number of fermions. In the Ramond sector we have

$$G_{\rm R} = \Gamma_{11} (-1)^E$$
, with $E = \sum_{n=1}^{\infty} d_{-n}^I d_n^I$, (5.121)

where Γ_{11} is the chirality operator as defined as the 10d analog of γ_5 in 4d QFT. A spinor that fulfills $\Gamma_{11}\Psi = \pm \Psi$ has definite chirality, and is called a Weyl spinor.

To perform the GSO projection we keep only states of positive G-parity. This means that we keep only states with an odd number of *b*-oscillator excitations in the NS sector, and for the Ramond sector we project onto an odd number of *d*-oscillator excitations if the ground state has negative chirality, and an even number of *d*-oscillator excitations if the ground state has positive chirality. Notably, this eliminates the tachyon from the ground state of the NS sector and the NS ground state becomes the vector boson $b_{-1/2}^{I} |0,k\rangle_{NS}$. Additionally, this means that all states in the Ramond sector are spacetime fermions, while all states in the NS sector are spacetime bosons.

Note that the massless bosonic state in the NS sector has 8 independent components. We said earlier that the Ramond vacuum was a Majorana spacetime spinor with $2^{D/2} = 32$ real components. Since $8 \neq 32$ this goes against our earlier claim that RNS string theory has spacetime supersymmetry. This problem is solved by two things, firstly only the transverse degrees of freedom are physical, so the physical Ramond ground state is actually a spinor in eight Euclidean dimensions, with 16 real components. In addition to this, by requiring that the vacuum has definite chirality we get down to 8 real components. Thus, we have a matching of the spacetime fermionic and bosonic degrees of freedom for the massless states of the theory¹².

Closed Superstrings

The analysis of closed superstrings is now fairly straightforward. In fact, the difference is mainly that the left-and right moving modes are decoupled and we can pick Ramond or Neveu-Schwarz boundary conditions separately for left- and right movers. We then get four different sectors for the closed string: NS-NS, NS-R, R-NS and R-R. In addition, we get an additional degree of freedom in choosing the relative chirality between the ground states of the two Ramond sectors, leading to type IIA or IIB string theory.

The R-R and NS-NS sectors give rise to spacetime bosons, while the R-NS and NS-R sectors give rise to spacetime fermions. This follows since the tensor product of an even number of fermions is bosonic, while the tensor product of a boson and a fermion is a fermion.

The mass-squared operator in closed string theory is given by

$$\frac{1}{2}\alpha' M^2 = \alpha' M_L^2 + \alpha' M_R^2$$
(5.122)

where M_L^2 and M_R^2 are the mass squared operators of the open string theories that make up the left-and right moving sectors.

Let us denote a Ramond vacuum of positive chirality $|+\rangle_R$, then we can write down the massless spectrum of type IIA string theory as

$$\begin{array}{lll} (\mathbf{R-, R+):} & |-\rangle_R & \otimes |+\rangle_R & \otimes |p^+, p^I\rangle \\ (\mathbf{NS, R+):} & b_{-1/2}^I |0\rangle_{\mathrm{NS}} & \otimes |+\rangle_R & \otimes |p^+, p^I\rangle \\ (\mathbf{R-, NS):} & |-\rangle_R & \otimes b_{-1/2}^I |0\rangle_{\mathrm{NS}} & \otimes |p^+, p^I\rangle \\ (\mathbf{NS, NS):} & b_{-1/2}^I |0\rangle_{\mathrm{NS}} & \otimes b_{-1/2}^J |0\rangle_{\mathrm{NS}} & \otimes |p^+, p^I\rangle. \end{array}$$

$$(5.123)$$

The NS-NS sector ground state transforms as the tensor product of two eight component vectors, just like the massless state in bosonic string theory. In the same way we can decompose the NS-NS ground state according to

(NS, NS):
$$b_{-1/2}^{I} |0\rangle_{\rm NS} \otimes b_{-1/2}^{J} |0\rangle_{\rm NS} \equiv \phi \oplus g^{IJ} \oplus B^{IJ}$$
 (5.124)

where the dilaton ϕ is the trace part of the tensor, g^{IJ} is traceless symmetric and B^{IJ} is antisymmetric. Turning off all other background fields, these behave exactly as in the bosonic case under the α' -expansion except now we are in D = 10.

 $^{^{12}}$ That spacetime supersymmetry holds for higher excited modes of the string is highly nontrivial, relying on an obscure identity found by Carl Gustav Jacob Jacobi in 1829. For a more detailed discussion we refer the reader to section 14.6 in [43].

From the R-R sector we have the tensor product of two spacetime spinors, with a total of $8^2 = 64$ degrees of freedom. The components of spacetime spinors commute, so we want to write the R-R sector as a direct sum of antisymmetric tensors. In addition, since the two Ramond sectors have opposite chiralities, we want the tensors to transform in the $(\frac{1}{2}, \frac{1}{2})$ of the Lorentz group. We then find that

$$(\mathbf{R}-,\mathbf{R}+): \quad |-\rangle_R \otimes R |+\rangle \equiv A^I \oplus A^{IJK}$$

$$(5.125)$$

where the three-form A^{IJK} is antisymmetric in its indices and has 56 independent components. Like the Kalb-Ramond field the three form field has a natural field strength and gauge symmetry. The field A^{I} corresponds to a photon.

The (NS-R) and (R,NS) fields have one vector and one spinor index and we can decompose the tensor products as

(NS-R):
$$b_{-1/2}^{I} |0\rangle_{\rm NS} \otimes |+\rangle_{-} \equiv \lambda^{1} \oplus \psi_{M}^{1}$$

(R-NS): $|+\rangle_{R} \otimes b_{-1/2}^{I} |0\rangle_{\rm NS} \equiv \lambda^{2} \oplus \psi_{M}^{2}$, (5.126)

where the λ are 8-component spinors called dilatinos and the ψ_M are 56 component vector-spinors called gravitinos. The NS-R and R-NS sectors are very similar, although they have opposite chiralities.

For type IIB string theory we carry out the same reasoning, letting both Ramond sectors have the same chirality. We find that the field content is

$$(\mathbf{R}+, \mathbf{R}+): \quad A_{(0)} \oplus A_{(2)} \oplus A_{(4)}^+$$

$$(\mathbf{NS}, \mathbf{R}+): \quad \lambda^1 \oplus \psi_M^1$$

$$(\mathbf{R}+, \mathbf{NS}): \quad \lambda^2 \oplus \psi_M^2$$

$$(\mathbf{NS}, \mathbf{NS}): \quad \phi \oplus q^{IJ} \oplus B^{IJ}.$$

$$(5.127)$$

The hidden difference in the mixed sectors is that the λ 's and ψ 's have the the same chirality in IIB theory. The R-R fields are the *axion* $A_{(0)}$, a second 2-form Kalb-Ramond field $A_{(2)}$ and a self dual 4-form field $A_{(4)}^+$.

5.2.1 Superstring Low Energy Effective Action

The low-energy effective action of the type IIA and type IIB are obtained in analogy to the low energy effective action of the bosonic string theory, the main difference being that we have to introduce superspace formalism, recreate the vertex operators, and redo the α' -expansion with a *super* conformal theory on the worldsheet. A main new ingredient is the necessity of superconformally invariant vertex operators containing a sum of different fields that transform into each other under the supersymmetry transformations.

The full derivation is well outside the scope of this text, and we will just state the effective actions, noting that they are determined by the vanishing of the α' -expanded β -functions. They may also be derived using the more modern pure spinor formalism, which is once again outside the scope of this text. These actions are taken from Ammon & Erdmenger [24].

Type IIA Supergravity

In terms of the field content of the massless type IIA superstring the bosonic part of the low energy effective action is given by

$$S_{\text{IIA}} = \frac{1}{2\kappa^2} \left[\int \mathrm{d}^{10} x \sqrt{-g} \left(e^{-2\phi} (\mathcal{R} + 4\nabla_M \phi \nabla^M - \frac{1}{2} |H|^2) - \frac{1}{2} |F_{(2)}|^2 - \frac{1}{2} |\tilde{F}_{(4)}|^2 \right) - \frac{1}{2} \int B \wedge F_{(4)} \wedge F_{(4)} \right],$$
(5.128)

where H = dB is a three form, $F_{(2)} = dA_{(1)}$ is a two form and $F_4 = dA_{(3)}$ is a four form. These three quantities are all simply the field strengths corresponding to their respective fields in the the IIA massless spectrum. Lastly, we have defined $\tilde{F}_{(4)} = dA_{(3)} - A_{(1)} \wedge H$.

It can be shown that this action is the same as the action of 11-dimensional supergravity compactified on a circle of radius $R_{11} = g_s^{2/3} l_p$ where l_p is the Planck length. Eleven dimensional supergravity is the unique theory in 11d that has local supersymmetry and no massless field of spin higher than 2. The bosonic part of the 11d supergravity action is given by

$$S_{11} = \frac{1}{2\kappa_{11}^2} \left[\int \mathrm{d}^{11}x \sqrt{-g} \left(\mathcal{R} - \frac{1}{2} |F_4|^2 \right) - \frac{1}{6} \int A_{(3)} \wedge F_{(4)} \wedge F_{(4)} \right] \,. \tag{5.129}$$

Type IIB Supergravity

In terms of the field content of the massless type IIB superstring the bosonic part of the low energy effective action is given by

$$S_{\text{IIA}} = \frac{1}{2\tilde{\kappa}^2} \left[\int d^{10}x \sqrt{-g} \left(e^{-2\phi} (\mathcal{R} + 4\nabla_M \phi \nabla^M - \frac{1}{2} |H|^2) - \frac{1}{2} |F_{(1)}|^2 - \frac{1}{2} |\tilde{F}_{(3)}|^2 - \frac{1}{4} |\tilde{F}_{(5)}|^2 \right) - \frac{1}{2} \int A_{(4)} \wedge H \wedge F_{(3)} \right],$$
(5.130)

where $\tilde{F}_{(3)} = dA_{(2)} - A_{(0)}H$, $\tilde{F}_{(5)} = F_{(5)} - \frac{1}{2}A_{(2)} \wedge H + \frac{1}{2}B \wedge F_{(3)}$.

5.2. Superstring Theory

Chapter 6

Introducing the AdS/CFT Duality

In this chapter the famous Anti de Sitter (AdS)/Conformal Field Theory (CFT) duality is introduced. An unorthodox approach is taken, in which the string theory argument for the duality is presented last, after the weaker form of the duality due to Witten has been thoroughly explored. The intent is that this approach is more constructive, we first consider in a fully controlled setting parts of the duality before ever considering a full realization.

First an introduction to the geometry of AdS is given, with special focus on how a noncompact space may be said to have a "boundary". We show that rotations and translation in the interior AdS space define conformal transformation of its boundary. A correspondence is then shown between bulk and boundary physics for one type of field at a time. Specifically, scalar fields, *p*-forms and metric deformations of AdS are considered. These partial results only relate some specific observables, and do not consider a full theory with all of its observables on either side of the duality. To understand the full duality we consider the original argument of Juan Maldacena relating supergravity on the spacetime $AdS_5 \times S^5$ to so-called $\mathcal{N} = 4$ Super Yang-Mills theory in 3+1 dimensions as outlined in [3]. Maldacena's form of the duality relates a full interacting quantum gravitational theory with additional fields to a QFT in flat space. Quantum gravity is in general poorly understood, and has no universally successful framework. On the other hand QFT on flat backgrounds is the most experimentally successful theory to date, and as such the proposed duality garnered great interest.

The chapter is concluded with a remark on what bulk theories can reasonably be dual to a boundary CFT, noting that explicit realizations of AdS/CFT have either strings in the bulk or they are carried out in the case of AdS_3/CFT_2 . In the latter case the CFT has a large amount of extra symmetry (Virasoro vs. regular conformal group), so the requirements on the bulk theory are not as steep.

6.1 Introduction to Anti de Sitter and its Conformal Boundary

In this section we try to make clear the connection between anti de Sitter (AdS) space and conformal field theory. We then explain the connection to the holographic principle by realizing that the boundary of AdS is the *conformal compactification* of *d*-dimensional Minkowski space M_d . The boundary of AdS is only defined up to a conformal transformation, as we will discover.

Thanks to this relation, instead of AdS/CFT duality being a one-to-one map between a CFT in M_d and quantum gravity in AdS_{d+1} with the two theories living in completely unrelated spacetimes, we have additional reason to the believe the CFT really lives on the boundary of the AdS spacetime.

In addition, since the boundary of AdS is only defined up to a conformal transformation, if

a theory that is relativistically invariant on AdS is supposed to be described by a theory on the boundary of AdS, that theory has to be invariant under conformal transformations. If an operator or correlator in the bulk of AdS transforms under some element of SO(d,2), then the corresponding object in the boundary CFT must transform in the same way.

The Boundary of AdS

The simplest way of understanding the symmetries of AdS_{d+1} is obtained by embedding it into d + 2-dimensional Minkowski as the surface satisfying

$$\tilde{\eta}_{MN}X^MX^N = X_1^2 + X_2^2 + \dots X_d^2 - X_0^2 - X_{d+1}^2 = -L^2$$
(6.1)

where $X_0, X_1 \dots X_{d+1}$ are the coordinates of M_{d+2} with metric $\tilde{\eta}$ and signature (d,2). The constant L is referred to as the AdS radius. The LHS of equation (6.1) is just the definition of $X_M X^M$ in the d+2 dimensional Minkowski spacetime, so the AdS surface is obviously invariant under SO(d,2). This is exactly the same symmetry group as a CFT_d in a d dimensional Minkowski spacetime with signature (d-1,1).

To find the conformal boundary of AdS_{d+1} we take the limit when all coordinates X_M become large. That is, we let $L^2 \to 0$, in which case AdS_{d+1} approaches the light-cone in M_{d+2} described by

$$X_1^2 + X_2^2 + \dots X_d^2 - X_0^2 - X_{d+1}^2 = 0.$$
(6.2)

We then define the "boundary" of AdS as the set of all lines on the lightcone originating from the origin of $\mathbb{R}^{d,2}$. More formally we can write this as the set of all points on the lightcone, together with an identification of points that are related by a real scale factor.

$$\partial \mathrm{AdS}_{d+1} = \left\{ X : X \in \mathbb{R}^{d,2}, X \neq 0, \tilde{\eta}_{MN} X^M X^N = 0 \right\}$$

$$(6.3)$$

together with the identification

$$X \sim \lambda X \tag{6.4}$$

where λ is some real number. By adding the scale identification, we are saying that all points on the lightcone in M_{d+2} are identified with the points at infinity for which $L^2 = 0$ makes sense, meaning they are identified with points that actually are at the boundary of AdS. This way of saying it might seem roundabout, but it lets us understand the topology of the AdS boundary in a simple way. We can represent any element in δAdS_{d+1} by X that fulfill

$$\sum_{i=1}^{d} X_i^2 = 1 , (6.5)$$

where we have chosen the scaling λ such that $X_0^2 + X_{d+1}^2 = 1$ in the RHS. In fact, the equation

$$X_0^2 + X_{d+1}^2 = 1 (6.6)$$

defines a circle S¹, and equation (6.5) defines the d-1 dimensional sphere S^{d-1}. The points X and -X are not identified in S¹ and S^{d-1}, but in our definition of δAdS_{d+1} we said that $X \sim -X$. Therefore the conformal boundary of AdS is has the topology $(S^1 \otimes S^{d-1})/\mathbb{Z}_2$. Note that this version of AdS is periodic in the timelike coordinate, and we would like to be able to consider AdS spaces that are not time-periodic. To do this, we pass to the *universal covering* of AdS, defined by taking the time direction to not be periodic. The only subtlety that is raised is that the universal covering can not be embedded into flat M_{d+2} , and we end up decompactifying the time direction. This does not spoil the following analysis showing that the boundary of AdS is a conformal compactification of Minkowski, as the different choice of λ lets us turn the boundary time aperiodic.

To find the claimed result that the boundary of AdS_{d+1} is the conformal compactification of M_d let us introduce the coordinates

$$u = X_{d+1} + X_d$$
, $v = X_{d+1} - X_d$ (6.7)

and rewrite equation (6.2) according to

$$uv = \eta_{\mu\nu} X^{\mu} X^{\mu} , \qquad (6.8)$$

where $\mu \in [0, d-1]$, $\eta_{\mu\nu}$ is the metric of M_d with signature (d-1, 1). We still have the identification under scalings $X \sim \lambda X$. This means that whenever $v \neq 0$ we can rescale so that v = 1, i.e $\lambda = 1/v$. For given X^{μ} , u is completely determined so that the set of points that fulfill equation (6.8) are simply any $X^{\mu} \in \mathbb{R}^{d-1,1}$. There is one caveat however, and that is the case where v = 0. To be able to scale by $\lambda = 1/v$ we need to add points at infinity for the X^{μ} , and by inspecting

$$0 = \eta_{\mu\nu} X^{\mu} X^{\mu} , \qquad (6.9)$$

we see that the set of points we need to add are defined by a lightcone. This is exactly the conformal compactification of Minkowski we found necessary to include conformal inversions on the lightcone in chapter 3.4.3.

It is actually possible to see that SO(d, 2) acts on the conformal boundary of AdS as a natural representation of the conformal group. Let us first consider the one parameter group of Lorentz boosts between X_{d+1} and X_d . It acts as

$$\begin{pmatrix} X'_{d+1} \\ X'_d \end{pmatrix} = \begin{bmatrix} \cosh \theta & -\sinh \theta \\ -\sinh \theta & \cosh \theta \end{bmatrix} \begin{pmatrix} X_{d+1} \\ X_d \end{pmatrix}$$
(6.10)

meaning that u and v transform as

$$\begin{pmatrix} u'\\v' \end{pmatrix} = \begin{bmatrix} \cosh\theta - \sinh\theta & 0\\ 0 & \cosh\theta + \sinh\theta \end{bmatrix} \begin{pmatrix} u\\v \end{pmatrix} .$$
(6.11)

We are scaling so that v = 1 to fix the arbitrary λ , so really this transformation implements

$$u' = \frac{\cosh \theta - \sinh \theta}{\cosh \theta + \sinh \theta} u = e^{-2\theta} u , \qquad (6.12)$$

but this means that that we are scaling the line element on the RHS of equation (6.8), so the boost between u and v with boost parameter θ implements a scaling of the metric by $e^{-2\theta}$ on the conformal boundary of AdS. We have found that that the Lorentz generator $J_{d,d+1}$ generates dilatations, which exactly the embedding of dilatations into SO(d,2) we had in section 3.4.2. All rotations within the boundary Minkowski space itself trivially just generate the Lorentz group in d dimensions.

It is now an exercise in algebra to check that general SO(d,2) transformations between X^{μ} and X^{d}, X^{d+1} correspond to some combination of special conformal transformations and spacetime translations in the boundary Minkowski spacetime.

Notice that since the Minkowski space at the boundary is not invariant under the symmetry transformations of the AdS embedding, the boundary of AdS is only defined up to a conformal transformation. Thus there is no specific metric induced on the boundary of AdS, only a *conformal structure*. A conformal structure refers to an equivalence class of metrics that are related by conformal transformation. In this example of pure AdS the boundary is in the *conformally flat* equivalence class of metrics, i.e. metric on the form $e^{2w}\eta^{\mu\nu}$.

Causal Connection Between AdS Boundary and Bulk

Let us now consider the important properties of AdS space itself. The most important property of AdS is that it is a solution to Einstein's equation, and for some time it was even thought to describe the asymptotic behaviour of our observable universe. The second most interesting property of AdS is that despite it being a noncompact manifold, a light signal can reach the boundary in finite time (and return again!). In this sense, AdS has the causal structure of a solid cylinder of finite radius and physics in the interior are causally connected to the physics on the boundary. This means that physics in AdS must take boundary conditions at infinity into account, in stark contrast to the usual field theoretical practice of tossing boundary terms in the equations of motion.

A choice of coordinates that solves the constraint in equation (6.1) is

$$X_{0} = \sqrt{L^{2} + r^{2}} \cos(t/L)$$

$$X_{d+1} = \sqrt{L^{2} + r^{2}} \sin(t/L)$$

$$X_{i}X^{i} = r^{2}$$
(6.13)

leading to the AdS metric

$$ds^{2} = -\left(1 + \frac{r^{2}}{L^{2}}\right)dt^{2} + \frac{dr^{2}}{1 + \frac{r^{2}}{L^{2}}} + r^{2}d\Omega_{d-1}^{2} , \qquad (6.14)$$

where $r \in [0, \infty), t \in [-\infty, \infty]$ and Ω_{d-1} is the metric on the unit sphere S^{d-1} . Note that we naturally have a periodicity in time so that $t \sim t + \frac{2\pi}{L}$, so in choosing $t \in \mathbb{R}$ we have passed to the universal covering space.

The metric of AdS_{d+1} is a solution of Einstein's equations with a negative cosmological constant

$$R_{\mu\nu} - \frac{R}{2}g_{\mu\nu} + \Lambda g_{\mu\nu} = 0 , \qquad (6.15)$$

where the cosmological constant Λ is given by

$$\Lambda = -\frac{d(d-1)}{2L^2} \ . \tag{6.16}$$

Knowing the metric it is straightforward to compute the time it takes for a photon to travel to the AdS boundary. For a photon, $ds^2 = 0$, so assuming $d\Omega_{d-1} = 0$ we can easily solve for coordinate time in terms of r, separate variables and integrate for the elapsed time. It is clear by inspection of the metric that $ds^2 = 0$ is solved by

$$\mathrm{d}t = \frac{L^2}{L^2 + r^2} \mathrm{d}r \; .$$

From here we just integrate both sides to find that

$$t = \int_0^\infty \frac{L^2}{L^2 + r^2} dr$$
$$= [L \arctan(r)]_0^\infty$$
$$= \frac{\pi}{2}L .$$

Then, the coordinate time for a photon to go to infinity and back from the origin again is given by πL . The main point here is that the bulk of the AdS spacetime can communicate with the boundary in finite time.



Figure 6.1: AdS imaged as a cylinder, with the coordinate ρ in the metric equation (6.18) written out. In grey is the part of AdS covered by the Poincaré patch. In red the geodesic of a photon travelling between two antipodal boundary points is imaged, with a travel time of $\Delta t = \pi$. Before the decompactification of the time direction, the Poincaré patch covers exactly half of AdS.

The causal structure we found for AdS can be made clearer by change of coordinates. In the following we set L = 1. We define the new coordinate $\rho \in [0, \pi/2)$ via

$$r = \tan \rho \tag{6.17}$$

and we find that

$$ds^{2} = \frac{1}{\cos^{2}\rho} \left(-dt^{2} + d\rho^{2} + \sin^{2}(\rho) d\Omega_{d-1}^{2} \right) .$$
(6.18)

The geodesics of photons do not see an overall factor on the metric (also called a "Weyl factor"), so as far as photons are concerned the metric of AdS is the same as that of a solid cylinder with metric

$$ds^{2} = -dt^{2} + d\rho^{2} + \sin^{2}(\rho)d\Omega_{d-1}^{2} .$$
(6.19)

At the boundary $(r = \infty \text{ or } \rho = \pi/2)$ we see that AdS has the topology of a sphere S^{d-1} tensored with the time direction, that is, the topology is given by $\mathbb{R} \times S^{d-1}$. This representation is illustrated in figure 6.1.

Poincaré Coordinate Patch

We may pass into another coordinate system in which many calculations simplify. These are called Poincaré coordinates, and they come with the downside that they do not cover all of AdS. The *Poincaré patch* that is covered by the Poincaré coordinates is highlighted in figure 6.1.

The Poincaré coordinates correspond to a different choice of coordinates that solves equation

(6.1):

$$X_{0} = \frac{L^{2}}{2r} \left(1 + \frac{r^{2}}{L^{4}} [\vec{x}^{2} - t^{2} + L^{2}] \right)$$

$$X_{d+1} = \frac{rt}{L}$$

$$X^{i} = \frac{rx^{i}}{L}, \forall i \in [1, \dots, d-1]$$

$$X^{d} = \frac{L^{2}}{2r} \left(1 + \frac{r^{2}}{L^{4}} [\vec{x}^{2} - t^{2} + L^{2}] \right)$$
(6.20)

where $r \ge 0$. The metric takes the form

$$ds^{2} = \frac{L^{2}}{r^{2}}dr^{2} + \frac{r^{2}}{L^{2}}\left(-dt^{2} + d\vec{x}^{2}\right) .$$
(6.21)

Here, the conformal boundary is located at $r = \infty$. A way to more explicitly see the form of the conformal boundary is to pass to boundary-appropriate coordinates. Defining $z = \frac{L^2}{r}$ the metric takes the form

$$ds^{2} = \frac{L^{2}}{z^{2}} \left(-dt^{2} + dz^{2} + d\vec{x}^{2} \right) .$$
 (6.22)

We see that up to the factor L^2/z^2 that can be Weyl transformed away, this is just Minkowski. The conformal boundary is at z = 0 and a particular structure is made very clear, the boundary of AdS is characterized by a double pole at the boundary. We will expand on this claim the the following subsection about Euclidean AdS.

Euclidean AdS

In the previous section we worked in a Lorentzian signature, but the identification of the boundary of AdS and Minkowski also holds in Euclidean signature. In general it may be convenient to do field theory in Euclidean language, as it is here that we can straightforwardly prove uniqueness theorems. Standard practice is to assume that Lorentzian field theory is related to Euclidean field theory by analytic continuation.

To obtain the Euclidean version of AdS we Wick-rotate the time coordinate $t \to i\tau$. It is then possible to show that Euclidean AdS_{d+1} can be identified with the unit ball B_{d+1} in \mathbb{R}^{d+1} with metric

$$\mathrm{d}s^2 = \frac{4\Sigma_{i=0}^d \mathrm{d}y_i^2}{(1-|y|^2)^2} , \qquad (6.23)$$

with coordinates $y_0 \ldots y_d$ and the unit ball is defined by $\sum |y|^2 < 1$. We can compactify the unit ball to obtain the closed unit ball $\overline{B_{d+1}}$, whose boundary is S_d . The fact that S_d is the boundary of $\overline{B_{d+1}}$ is the Euclidean version of the statement that the conformal compactification of M_d is the boundary of AdS_{d+1} . This can be seen by considering that the conformal compactification of R^d is obtained by adding a single point at infinity, tying R^d together into the sphere S^d .

The metric on $\overline{B_{d+1}}$ does not extend to the boundary, since it is singular for |y| = 1. To obtain a metric that is defined on the boundary S_d we can pick a function f on $\overline{B_{d+1}}$ that has a first order zero at |y| = 1, for example $f = 1 - |y|^2$, and rescale ds^2 :

$$\mathrm{d}\tilde{s}^2 = f^2 \mathrm{d}s^2 \ . \tag{6.24}$$

The function f is only defined up to a general function that approaches a constant value at the boundary. This class of functions can be fully captured by e^w where w is any real function on $\overline{B_{d+1}}$ plus an eventual minus sign (that cancels out). In other words, we have the equivalence relation

$$f \sim \pm f e^w \tag{6.25}$$

with w any real function on $\overline{B_{d+1}}$, inducing a conformal transformation on the boundary such that

$$\mathrm{d}\tilde{s}^2 \to e^{2w} \mathrm{d}\tilde{s}^2 \;, \tag{6.26}$$

for the metric of the S^d . Just like we found in the Lorentzian case, Euclidean AdS has a boundary that is only defined up to a conformal transformation, and it is conformally equivalent to \mathbb{R}^d plus a point at infinity.

A useful choice of coordinates for Euclidean AdS is obtained by the substitution $r = \tanh(|y|/2)$, and introducing angular coordinates, putting the metric on the form

$$\mathrm{d}s^2 = \mathrm{d}y^2 + \sinh^2 y \mathrm{d}\Omega^2 \;, \tag{6.27}$$

where $d\Omega$ is the metric on the unit sphere and $0 \le y \le \infty$. In these coordinates, the boundary is located at $y = \infty$.

Finally, another useful representation of Euclidean AdS is as an upper half plane $x^0 > 0$ with coordinates $x^0 \dots x^d$ and metric

$$ds^{2} = \frac{1}{x_{0}^{2}} \left(\sum_{i=0}^{d} (dx^{i})^{2} \right) .$$
 (6.28)

In this representation the boundary consists of a space that looks like \mathbb{R}^d at $x^0 \to 0$ plus a point at infinity given by $x^0 = \infty$. The boundary when $x^0 = \infty$ defines a point since the metric vanishes in this limit. Just as previously, this defines a conformally flat boundary structure.

Asymptotically AdS Spacetimes

If we are interested in gravity we clearly cannot only consider AdS space, since any nontrivial matter distribution will cause the metric to deviate from equation (6.14). We are therefore interested in the class of *asymptotically AdS spacetimes*, meaning any spacetime that have boundary topology $\mathbb{R} \times S^{d-1}$ and whose metric approaches equation (6.14) at large radii.

The most important asymptotically AdS spacetime is probably the Schwarzschild-AdS black hole, with metric given by

$$ds^{2} = -f(r)dt^{2} + \frac{dr^{2}}{f(r)} + r^{2}d\Omega_{d-1} , \qquad (6.29)$$

where the form function f(r) is

$$f(r) = 1 + r^2 - \frac{16\pi GM}{(d-1)\Omega_{d-1}} \frac{1}{r^{d-2}} .$$
(6.30)

At large radii, the term proportional to GM in f(r) goes to zero, so the metric obviously approaches AdS at infinity.

A more general way to consider asymptotically AdS spacetimes is via the Fefferman-Graham representation of the metric. To define this it is practical to use the boundary-adapted coordinates equation (6.22) on the form

$$ds^{2} = \frac{L^{2}}{z^{2}} \left(dz^{2} + h_{\mu\nu}(z, x^{\mu}) dx^{\mu} dx^{\nu} \right)$$
(6.31)

where the index runs over all but the coordinate z. On this form, all asymptotically AdS metrics are described by a $h_{\mu\nu}(z, x^{\mu})$ such that the limit $\lim_{z\to 0} h_{\mu\nu}(z, x^{\mu}) = h_{\mu\nu}^{0}$ is finite. Multiplying the metric by z^2 , $h_{\mu\nu}^{0}$ defines the *boundary metric*. We see that picking $h_{\mu\nu}^{0} = \eta_{\mu\nu}$ returns the boundary coordinates of pure AdS. We will later see examples where the boundary metric is not in the equivalence class of conformally flat metrics.

The coordinate z may be interpreted as the inverse energy scale of a CFT living on the boundary space. To see this, remember that the boundary of AdS is only defined up to a conformal transformation. If we let $x^{\mu} \to \alpha x^{\mu}$, the compensating transformation of z is $z \to \alpha z$ to preserve the metric. In a field theory, the energy E goes as E/α under a rigid scale transformation. The limit $z \to 0$ thus corresponds to the high energy limit of the boundary theory, with bigger z probing lower and lower energies of the CFT. In this sense, the radial coordinate $r = L^2/z$ of AdS can be seen as the renormalization scale of the boundary CFT becoming a genuine physical dimension.

6.2 AdS/CFT Duality Piece by Piece

In this section we show the correspondence between operators in a CFT on the boundary of an AdS spacetime and fields in the bulk by considering some basic examples. We will treat in some detail all of the pieces that go into the full AdS/CFT duality between stringy quantum gravity in the bulk and supersymmetric Yang-Mills theory on the boundary.

The essence of the AdS/CFT duality lies in a fairly simple observation. If we solve the classical equations of motion for some theory in the interior with field values at infinity as boundary conditions we may find that there is a unique field configuration on the interior that can satisfy this boundary condition. More succinctly, in Euclidean AdS the boundary fields completely determine the bulk fields.

We then extend this classical correspondence to the quantum level by postulating that a conformal field theory on the boundary and whatever theory in the bulk have the same *generating functional* in the path integral sense. We test this conjecture by computing the boundary CFT correlation functions using the bulk theory generating functional and finding that they agree with the expected correlators found in chapter 3.4.3.

We will formulate the relationship between a CFT on the boundary and the following theories in the bulk:

- scalar field theory,
- Yang-Mills gauge theory,
- gravity.

The reason for not including spinors is that we will be mainly working with the so-called *classical* supergravity approximation, setting all spinor fields to zero. A treatment of the correspondence between spinors in AdS and quasi-primary conformal spinors in the boundary CFT can be found in [46].

The following scalar field discussion is based on the presentation by Witten in [47] with some elements included from Freedman [48]. The discussion of Yang-Mills fields is also based on Witten's paper. The discussion about gravity follows a paper by M. Henningson and K. Skendris [49]. Finally, the string theoretical realization of the duality is based on the presentation by Zwiebach [43].

6.2.1 Scalar Field/CFT Correspondence

The first, absolutely simplest case to consider is the massless scalar field living in AdS_{d+1} . It is in the section that we will be most explicit about checking the duality, although the general machinery applies to *p*-form and gravitational fields as well. In this section we take AdS to be a static gravitational background, meaning the scalar field does not couple via the stress-energy tensor to gravity. In this way we only have to deal with the dynamics of the scalar field.

By a massless scalar field we mean a scalar field obeying the equation of motion

$$\nabla_M \nabla^M \Phi = 0 , \qquad (6.32)$$

where ∇_M is the general covariant derivative. We will find that the field in the interior is completely determined by the field configuration at the boundary. We will then use this finding to conjecture a correspondence between the dynamics of boundary fields and fields in the interior.

A basic fact about AdS is that given the field $\phi(\Omega), \Omega \in S^d$ on the boundary, there is a unique solution of the Laplace equation in the bulk $\Phi(y) y \in \overline{B_{d+1}}$ that satisfies the boundary condition $\Phi(y) = \phi(\Omega) \quad \forall y \in S^d$. The uniqueness depends on the fact that there is no nonzero squareintegrable solution of the Laplace equations. A square integrable solution goes to zero at infinity and can therefore be added to an existing solution without spoiling the boundary condition. If there existed a solution where $\phi(\Omega) = 0$ then we would have

$$0 = -\int_{B_{d+1}} \mathrm{d}^{d+1} y \sqrt{g} \Phi \nabla_M \nabla^M \Phi$$

= $-\int_{B_{d+1}} \mathrm{d}^{d+1} y \Phi \partial_M (\sqrt{g} \partial^M \Phi)$
= $\int_{B_{d+1}} \mathrm{d}^{d+1} y \sqrt{g} |d\Phi|^2$ (6.33)

where the first equality is due to the equation of motion setting the integrand to zero, the second equality uses the definition of the general covariant derivative, and in the third equality we have used $\phi(\Omega) = 0$ to toss the boundary terms. Equation equation (6.33) implies that $d\Phi(y) = 0$ everywhere since the integral of $(d\Phi)^2$ is zero, setting $\Phi = 0$ everywhere if it is zero on the boundary.

Working with the equation (6.27) representation of AdS, the Laplace equation takes the form

$$\left(-\frac{1}{(\sinh y)^d}\frac{\mathrm{d}}{\mathrm{d}y}(\sinh y)^d\frac{\mathrm{d}}{\mathrm{d}y} + \frac{L^2}{\sinh^2 y}\right)\Phi = 0 , \qquad (6.34)$$

where the operator L^2 is the angular part of the Laplacian. If we expand ϕ in terms of spherical harmonics f_{α} according to

$$\Phi = \sum_{\alpha} \Phi_{\alpha}(y) f_{\alpha}(\Omega) \tag{6.35}$$

the equations of motion look for large y like

$$\frac{\mathrm{d}}{\mathrm{d}r}e^{dy}\frac{\mathrm{d}}{\mathrm{d}y}\Phi_{\alpha} = 0 , \qquad (6.36)$$

with solutions $\Phi_{\alpha} \sim 1$ and $\Phi_{\alpha} \sim e^{-dy}$. Only one linear combination of these is smooth at y = 0, and this solution has a finite value at infinity. If this was not the case there would exist square integrable solutions relating the non-unique solutions with the same boundary value, but we already proved that square integrable solutions do not exist.

Therefore, for every partial wave f_{α} there is a unique solution of the Laplace equation in radial direction for each constant value at infinity. If we expand $\phi(\Omega)$ in the boundary in terms of spherical harmonics as

$$\phi(\Omega) = \sum_{\alpha} c_{\alpha} f_{\alpha} , \qquad (6.37)$$

we see that any function $\phi(\Omega)$ on S_d will define a unique solution of the Laplace equation in the interior by extending $c_{\alpha} \to \Phi_{\alpha}(r)$ and requiring $\lim_{r\to\infty} \Phi_{\alpha}(r) = c_{\alpha}$.

The conclusion we can draw from this example is that the scalar field in the interior of AdS_{d+1} is completely determined by its values at infinity due to the nonexistence of square integrable solutions to the Laplace equation. In addition, this determination holds for all times, so in some sense the dynamics of the boundary field completely determine what the dynamics of the field on the interior must be. It is not unreasonable to think that if we quantize the two field theories, the one on the boundary and the one in the interior, they might be described by the same Hilbert space and quantum dynamics.

More explicitly, since we have a correspondence for all times we conjecture that

some scalar QFT in AdS_{d+1} is dynamically equivalent to some scalar CFT on ∂AdS_{d+1} .

Note that we are not claiming that these theories should somehow supplement each other, but rather that the two theories are interchangeable.

Ansatz for the Explicit AdS/CFT Duality

Let ϕ denote values of the bulk field Φ to the boundary of AdS_{d+1} . We assume that the duality between the the bulk and boundary theories should couple the boundary field ϕ to some conformal field \mathcal{O} via a vertex operator on the form $\int_{\mathrm{S}^d} \phi \mathcal{O}$. The boundary field ϕ is conformally invariant, so for the the vertex operator to have conformal weight zero, the operator \mathcal{O} must have weight d.

We would like to compute the correlation functions $\langle \mathcal{O}(x_1)\mathcal{O}(x_2)\ldots\mathcal{O}(x_n)\rangle$ for $x_i \in S^d$. In addition, we hope that the boundary CFT is so well-behaved that we can have a well defined generating functional according to

$$Z_{CFT}[\phi] = \int \mathcal{DO} \, e^{-S_{CFT} + \int_{\partial B_{d+1}} \phi \mathcal{O}} \,, \tag{6.38}$$

where the path integration contour is chosen to project out the vacuum at the boundaries and S_{CFT} is the Euclidean action of the CFT in question. For the bulk theory, we have the partition function

$$Z_B[\phi] = \int_{g_B=h} \mathcal{D}g \int_{\Phi=\phi} \mathcal{D}\Phi \, e^{-S_B} \,, \qquad (6.39)$$

where the boundary field configurations enter as boundary conditions on the path integral. By the boundary equality " $g_B = h$ " we mean that we integrate over the set of metrics in the bulk g_B that induce the conformal structure h at the boundary. We take the Euclidean bulk action S_B to be the action for supergravity or string theory in the interior of (Euclidean) AdS.

The precise ansatz for the relationship between a boundary field theory on the boundary and a field theory in the bulk is then

$$Z_B = Z_{CFT} {.} {(6.40)}$$

The bulk partition function Z_B is in general not fully known, and is often handled via a *classical* approximation wherein we assume the path integral is dominated by the classical solution, letting

$$Z_B[\phi] \to \tilde{Z}_B[\phi] = e^{-S_B[\Phi]} \tag{6.41}$$

where the Euclidean bulk action is integrated over the field configurations that solve the classical equations of motion with the desired boundary conditions. In cases where the classical bulk action is not unique, we should sum over all classical solutions, since none of the classical solutions are suppressed by Planck's constant as a quantum effect. In general the bulk theory is a theory of quantum gravity, and the condition for small quantum corrections is $\frac{\sqrt{\alpha'}}{r_c} \ll 1$ with r_c the characteristic radius of AdS and $\sqrt{\alpha'}$ the string length. This was covered in section 5.1.3 in the discussion around equation (5.84) when performing the α' -expansion around a classical solution of stringy gravity. In the case where we imagine a non stringy bulk theory, there is no stringent notion of what "small" quantum corrections mean.

Checking the Proposed Duality for Two-point Functions

Let us now carry out some sample calculations. The idea is to compute the partition functions for the interior and boundary theories and find that they match. We shall begin with an AdS theory in classical approximation containing a massless scalar with action

$$S[\Phi] = \frac{1}{2} \int_{B_{d+1}} \mathrm{d}^{d+1} y \sqrt{g} |\mathrm{d}\Phi|^2 \ . \tag{6.42}$$

We denote by $S[\phi]$ the *on shell action* obtained by inserting the solution Φ with boundary value ϕ as the argument of the action.

We assume that the boundary value ϕ sources some state \mathcal{O} in the boundary CFT. We would like to compute a two-point correlation by evaluating

$$\left\langle \mathcal{O}(x)\mathcal{O}(x')\right\rangle = \frac{1}{Z[0]} \frac{\delta}{\delta\phi(X_1)} \frac{\delta}{\delta\phi(X_2)} Z_B[\phi]\Big|_{\phi=0} , \qquad (6.43)$$

and show that it is the expected value that we'd obtain from $Z_{CFT}[\phi]$.

To solve the Laplace equation in the bulk for some boundary field ϕ we should first look for a Green's function solution K that defines a delta function at some point P on the boundary. A convenient way to do this is to pick the upper-half plane representation of Euclidean AdS with metric 6.28

$$ds^{2} = \frac{1}{x_{0}^{2}} \left(\sum_{i=0}^{d} (dx_{i})^{2} \right) , \qquad (6.44)$$

taking the point P to be the point at $x^0 = \infty$. The boundary condition (δ at $x^0 = \infty$) as well as the metric is translation invariant in the x^i , so the solution K will posses this symmetry and depend only on x_0 . We can then write down the Laplace equation

$$\frac{\mathrm{d}}{\mathrm{d}x_0} x_0^{-d+1} \frac{\mathrm{d}}{\mathrm{d}x_0} K(x_0) = 0 \ . \tag{6.45}$$

We want a solution that vanishes at $x_0 = 0$ and divergent at $x_0 = \infty$. The solution that has this property is

$$K = cx_0^d av{6.46}$$

where c is some constant. This solution diverges at infinity and is zero on the $x_0 = 0$ boundary. To show that this really defines a delta function, let us make an SO(1,d+1) transformation to make P a finite point. We let

$$x_i \to \frac{x_i}{x_0^2 + \sum_{j=1}^d x_j^2} , \quad i \in [0, d] ,$$
 (6.47)

mapping P to the origin, and transforming K to

$$K(x) = c \frac{x_0^d}{\left(x_0^2 + \sum_{j=1}^d x_j^2\right)^d} .$$
(6.48)

The boundary is still defined by $x_0 = 0$ and $x_0 = \infty$, so we want the delta function to be a delta function in the x_j coordinates. By dimensional analysis, $\int dx_1 \dots dx_j K(x)$ is independent of x_0 . Also, as $x_0 \to 0$ K(x) vanishes unless all of the $x_j = 0$. In addition K is positive. Therefore, as $x_0 \to 0$, K becomes a delta function at $x_j = 0$ which, if c is chosen correctly, has unit coefficient.

Defining $\mathbf{x} = (x_1, x_2 \dots x_d)$ we can now write down a solution Φ with an arbitrary boundary function ϕ as a convolution of the Green's function solution with the boundary function in the usual way

$$\Phi(x_0, \mathbf{x}) = c \int d\mathbf{x}' K(x_0, \mathbf{x} - \mathbf{x}') \phi(\mathbf{x}') = c \int d\mathbf{x}' \frac{x_0^d}{(x_0^2 + |\mathbf{x} - \mathbf{x}'|^2)^d} \phi(\mathbf{x}') .$$
(6.49)

As $x_0 \to 0$ it is straightforward to compute the partial derivative

$$\frac{\partial \Phi}{\partial x_0} \sim dc x_0^{d-1} \int \mathrm{d}\mathbf{x}' \frac{\phi(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^{2d}} + \mathcal{O}(x_0^{d+1}) \ . \tag{6.50}$$

We now have a classical solution of the equations of motion for any boundary field. We can express the action as an integral over this solution, letting us write $S[\Phi]$ as a boundary integral:

$$\frac{1}{2}\int \mathrm{d}^{d+1}x\,\sqrt{g}|\mathrm{d}\Phi|^2 = \frac{1}{2}\int \mathrm{d}^d\mathbf{x}\,\sqrt{h}\Phi(\hat{n}\cdot\nabla)\Phi - \frac{1}{2}\int_{B_{d+1}}\mathrm{d}^{d+1}x\,\sqrt{g}\Phi\nabla^2\Phi \tag{6.51}$$

where h is the induced metric on the boundary, the last term is zero when we impose the equations of motion and the boundary integral is performed over the boundary $x_0 = 0$ with normal \hat{n} . In principle, we explicitly have to take the limit $x_0 \to 0$ to get the final result. On the boundary one has $\sqrt{h} = x_0^{-d}$, $(\hat{n} \cdot \nabla) = x_0 \frac{\partial}{\partial x_0}$. Then, using $\Phi(0, \mathbf{x}) = \phi(\mathbf{x})$ and inserting equation (6.50) we can write the on-shell action

$$S[\phi] = \frac{cd}{2} \int d\mathbf{x} d\mathbf{x}' \frac{\phi(\mathbf{x})\phi(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^{2d}} , \qquad (6.52)$$

and thus the on-shell partition function is,

$$Z_B[\phi] = e^{-\frac{cd}{2} \int d\mathbf{x} d\mathbf{x}' \frac{\phi(\mathbf{x})\phi(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^{2d}}} .$$
(6.53)

The two-point correlator is simple, since only terms with no remaining factors of ϕ survive we find

$$\langle \mathcal{O}(x)\mathcal{O}(x')\rangle = \frac{cd}{|\mathbf{x} - \mathbf{x}'|^{2d}},$$
(6.54)

which is indeed the expected behaviour of the correlation function between two conformal primary fields of conformal dimension d according to section 3.4.3. We have explicitly verified that the

AdS/CFT correspondence correctly relates two-point correlators for massless scalars Φ in the d+1 dimensional bulk and conformal primary fields \mathcal{O} of scaling dimension d on the boundary.

Despite this success there is a problem glaring at us; by definition our partition function can not produce a nonzero three-point correlator. Nontrivial three-point functions exist as long as nontrivial two-point functions do in a CFT, so this tells us that to recover the boundary CFT we need to write down a more general scalar field theory in AdS that respects the SO(d,2) symmetry. Before dealing with this, let us first generalize to the case of a massive scalar field.

Massive Scalar Field

Let us now consider the case of a massive scalar field in the interior of AdS_{d+1} and what sort of conformal fields it corresponds to on the boundary. We will find that the massive scalar field with mass m^2 in the bulk corresponds to a conformal primary operator \mathcal{O} in the boundary with mass scaling dimension $\Delta = \frac{1}{2}(d + \sqrt{d^2 + m^2})$.

We begin by considering the classical bulk action

$$S[\Phi] = \frac{1}{2} \int d^{d+1}y \sqrt{g} \left(|d\Phi|^2 + m^2 \Phi^2 \right) .$$
 (6.55)

Representing Euclidean AdS as a ball according to equation (6.27) we obtain the equations of motion

$$\left(-\frac{1}{(\sinh y)^d}\frac{\mathrm{d}}{\mathrm{d}y}(\sinh y)^d\frac{\mathrm{d}}{\mathrm{d}y} + \frac{L^2}{\sinh^2 y} + m^2\right)\Phi = 0 , \qquad (6.56)$$

similarly to the massless case (equation (6.34)). For large y the L^2 terms remains irrelevant as it is suppressed by a factor e^{-2y} . For large y, $\sinh(y) \sim e^y/2$ and the equation motion reduces to

$$\left(e^{-dy}\frac{\mathrm{d}}{\mathrm{d}y}e^{dy}\frac{\mathrm{d}}{\mathrm{d}y} - m^2\right)\Phi = 0.$$
(6.57)

The two linearly independent solutions to the differential equation for large y are $e^{\lambda y}$ where λ is a solution to the characteristic equation

$$\lambda(\lambda + d) = m^2 . ag{6.58}$$

To prevent the problem of physical tachyons and vacuum instability, we consider only m^2 such that the equation has real solutions¹. We denote by λ_+, λ_- the larger and smaller solutions respectively. Note that this restriction on m^2 imposes $\lambda_+ \ge -\frac{1}{2}$ and $\lambda_- \le -\frac{d}{2}$. By uniqueness², there is only one linear combination of the two solutions that extends smoothly over the interior of AdS. At infinity, this solution is dominated by $e^{\lambda_+ y}$. In contrast to the massless case this solution does not have a constant term at infinity, so we cannot solve this via the spherical harmonics as we did back then.

¹This is the analogue of requiring $m^2 > 0$ in flat space. The main idea is that in AdS the scalar field obtains a kinetic energy contribution from the boundary so that "true" AdS masslessness does not occur until $m^2 = -d^2/4$. In section two of [50] a loophole to this requirement is found, extending analysis of scalars in AdS to $-d^2/4 - 1 < m^2 < -d^2/4$. The main difference in this regime is that there are two, instead of one solution in the interior that satisfies the boundary condition, corresponding to picking either Δ_+ or Δ_- as the exponent for the boundary behaviour. This is in principle because we cannot guarantee that one solution dominates over the other when the solutions oscillate.

²Uniqueness for bulk solutions of the massive scalar field in AdS was shown in [51] by Breitenlohner and Freedman. Uniqueness is obtained by explicitly solving the differential equation with hypergeometric functions in the cylinder (equation (6.18)) representation of AdS and finding that there is a unique solution that is regular at the origin.

The closest we can get to such a solution is to pick an arbitrary positive function f with a simple zero on the boundary. An example function would be $f = e^{-y}$. We can then hope that there is a solution to the equations of motion that looks like

$$\Phi \sim f^{-\lambda_+} \phi_m \tag{6.59}$$

with an arbitrary boundary function ϕ_m (that only depends on the angular coordinates).

It is important to understand what kind of object the field ϕ_m is. How ϕ_m is defined as a function depends on the choice of the arbitrary function f with a simple zero on the boundary, this is the exact same situation as in the case when we wanted to extend the metric on B_{d+1} to the conformal boundary. Transforming $f \to e^w f$ changes the metric according to $d\tilde{s}^2 \to e^{2w} d\tilde{s}^2$. If Φ is to behave as an honest function on the boundary, it must be independent of the choice of f, and we see that under $f^{-\lambda_+} \to e^{-\lambda_+ w} f^{-\lambda_+}$ the boundary field must transform as $\phi_m \to e^{\lambda_+ w}$. The conclusion we draw is that this boundary field ϕ_m must be a conformal primary field of weight $-\lambda_+$.

We are now ready for two-point functions. If there is a boundary CFT on $\partial \text{AdS}_{d+1}$ with a coupling $\int \phi_m \mathcal{O}$ we have concluded that the operator must have conformal dimension $d + \lambda_+$ to cancel the dimension of ϕ_m and the integration measure. We can verify that the the scalar field/CFT duality holds by computing the two-point correlator using the classical approximation of the bulk partition function.

First, let us find the explicit form of a function Φ that obeys the massive wave equation and behaves as $f^{-\lambda_+}\phi_m$ at infinity. Once again we go to the half-space representation of AdS because it is comparatively simple to single out a lone point at the boundary. The metric is

$$ds^{2} = \frac{1}{x_{0}^{2}} \left(\sum_{i=0}^{d} (dx_{i})^{2} \right) , \qquad (6.60)$$

where the boundary is given by the hypersurface $x_0 = 0$ as well as a single point $x_0 = \infty$.

Repeating the massless argument, we want a solution K that is zero at all points of the boundary except for $x_0 = \infty$, defining a boundary delta function. Then, since the point at infinity is invariant under translations we find that K must depend only on x_0 and the massive wave equation becomes

$$\left(x_0^{d+1}\frac{\mathrm{d}}{\mathrm{d}x_0}x_0^{-d+1}\frac{\mathrm{d}}{\mathrm{d}x_0} - m^2\right)K(x_0) = 0.$$
(6.61)

As before, the ansatz for a solution that vanishes for $x_0 = 0$ and is nonzero at $x_0 = \infty$ is $K(x_0) = cx_0^N$. Solving for N we find $K = x_0^{d+\lambda_+}$. We then perform the inversion

$$x_i \to \frac{x_i}{x_0^2 + |\mathbf{x}|^2} ,$$
 (6.62)

mapping $x_0 = \infty$ to the origin, and transforming K to

$$K(x) = \frac{x_0^{d+\lambda_+}}{\left(x_0^2 + |\mathbf{x}|^2\right)^{d+\lambda_+}} .$$
(6.63)

By the same arguments as in the massless case this is proportional to a delta function as $x_0 \to 0$. We can then write down a general solution that gives rise to some boundary field ϕ_m at infinity by convoluting with the Green's function

$$\Phi(x_0, \mathbf{x}) = c' \int d\mathbf{x}' \frac{x_0^{d+\lambda_+}}{(x_0^2 + |\mathbf{x} - \mathbf{x}'|^2)^{d+\lambda_+}} \phi_m(\mathbf{x}') .$$
(6.64)

As $x_0 \to 0$ we see that

$$\left. \frac{\partial \Phi}{\partial x_0} \right|_{x_0=0} \sim c'(d+\lambda_+) x_0^{d+\lambda_+-1} \int \mathrm{d}\mathbf{x}' \frac{\phi_m(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|^{2(d+\lambda_+)}} \,. \tag{6.65}$$

The action $S[\Phi]$ can then be rewritten as a surface term by using the equations of motion to set the remaining bulk terms to zero after partial integration, yielding the on-shell action

$$S[\phi_m] = c' \frac{d + \lambda_+}{2} \int d\mathbf{x} d\mathbf{x}' \frac{\phi_m(\mathbf{x})\phi_m(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^{2(d + \lambda_+)}} .$$
(6.66)

The two-point function is proportional to $|\mathbf{x} - \mathbf{x}'|^{-2(d+\lambda_+)}$, as expected for a conformal field \mathcal{O} of weight $d + \lambda_+$.

We have thus found a correspondence between massive scalar fields of mass m in AdS_{d+1} and conformal primary fields \mathcal{O} of conformal weight

$$\Delta = d + \lambda_{+} = \frac{1}{2} (d + \sqrt{d^2 + 4m^2}) . \qquad (6.67)$$

The lower bound on Δ is set by $m^2 = -d^2/4$, giving us $\Delta \ge d/2$. Therefore, in a CFT with a scalar field dual there are no conformal primary fields with weight less than d/2. We will later find that the addition of *p*-form fields in the bulk may give rise to conformal fields of minimum weight d/2 - p. In addition, [50] extends the analysis of scalar fields down to the unitarity bound of the boundary CFT at $\Delta = \frac{d-2}{2}$.

Two Point-Correlators Are Given By Geodesic Lengths

There is an alternate way of finding the two-point correlators of massive scalar field. The insertion of conformal operators \mathcal{O} on the boundary corresponds to δ -function boundary conditions for the bulk scalar ϕ . For the purpose of the two-point function we should therefore be able to replace the full scalar field action by the relativistic point particle action³. The point particle action in general relativity is just the length of a geodesic times the energy density of the point particle along the worldline,

$$S_m[X] = m \int_0^T \mathrm{d}s \sqrt{g_{\mu\nu}} \frac{\partial X^{\mu}}{\partial s} \frac{\partial X^{\nu}}{\partial s} , \qquad (6.68)$$

where s is the affine parameter. Note that the argument of the square root is nonnegative since we are in Euclidean signature. In Lorentzian signature, the metric term is $\sqrt{-g}$ and this method may fail for spacelike separated operators \mathcal{O} . For any spacetime that can be obtained as an analytic continuation of a Euclidean spacetime, the geodesic approximation holds. This will be true for almost every spacetime considered in this text.

In AdS the length of any geodesic that is not exactly lightlike diverges at radial infinity, meaning we need to put the boundary at a finite radius r_b and take a limit $r_b \to \infty$ to regulate the geodesic length. Let us now consider two conformal operators inserted at points (t_1, l) , $(t_2, 0)$ on the boundary. The corresponding partition function on the gravity side is a path integral over all paths connecting the boundary points (t_1, l, r_b) , $(t_2, 0, r_b)$. By going to the semiclassical approximation, it is a sum over all paths that solve the geodesic equation while ending on the boundary points. Then we have

$$\langle \mathcal{O}(t_1, l)\mathcal{O}(t_2, 0)\rangle = \int \mathcal{D}X e^{-mL[X]} \approx e^{-mL[x]} ,$$
 (6.69)

³Technically, we should prove that the point particle action and relativistic scalar fields both invert the Klein-Gordon differential operator with a δ boundary condition.
where L[x] denotes the length of the bulk geodesic x that solves the classical equations of motion. The semiclassical approximation corresponds to large m since all L[X] that are not minimal are strongly damped in the path integral. Restoring the AdS radius L, and using that $\Delta/L \approx m$ for large m, we can write this as

$$\langle \mathcal{O}(t_1, l)\mathcal{O}(t_2, 0)\rangle \sim e^{-\frac{\Delta L[x]}{L}}$$
 (6.70)

This result provided an important clue towards the Ryu-Takayanagi conjecture to be covered in section 7.2, relating the entanglement entropy between subregions on the boundary to d-1-dimensional "areas" in the bulk. In chapter 4.3.3 we found that the entanglement entropy between complementary regions in a CFT₂ theory was given by the expectation value of so-called twist operators \mathcal{T}_n . This expression, using our newfound geodesic approximation is given by

$$c_n \langle \mathcal{T}_n(-a) \mathcal{T}_{-n}(a) \rangle^n = e^{-\frac{2n\Delta L[x]}{L}} , \qquad (6.71)$$

where the factor 2 comes from the fact that the twist operators have both holomorphic and antiholomorphic weights $\Delta_n = \frac{c}{12}(1-1/n^2)$. Then, the von Neumann entropy is

$$S_{A} = -\frac{\partial}{\partial n} e^{-\frac{2n\Delta L[x;-a,a]}{L}} \bigg|_{n=1}$$

$$= 2\frac{\partial(n\Delta_{n})}{\partial_{n}} \frac{L[x;-a,a]}{R} e^{-\frac{2n\Delta L[x;-a,a]}{L}} \bigg|_{n=1}$$

$$= \left[\frac{c}{12}(1-n^{-2}) + \frac{c}{6}n^{-1}\right] \bigg|_{n=1}$$

$$= \frac{c}{6} \frac{L[x;-a,a]}{R} , \qquad (6.72)$$

where we have added the boundary conditions as extra arguments to the geodesic length L. To completely translate this result to gravitational language we need to cite a future result from section 6.2.3, namely that gravity in asymptotically AdS₃ space is dual to a CFT with central charge $c = 3R/2G_N^{(3)}$ where $G_N^{(3)}$ is the 3d gravitational constant. Then the entropy becomes

$$S_A = \frac{L[x; -a, a]}{4G_N^{(3)}} . (6.73)$$

Since in two spatial dimensions a length is the analog of an area, equation (6.73) is similar to what we would expect from the entropy of a black hole in AdS_3 with circumference L. We will investigate this connection more deeply in chapter 7. We have yet to introduce the tools necessary to compute the length of the geodesic L and check that the obtained result actually matches the CFT result. This will also be checked in chapter 7.

Interacting Scalar Fields and *n*-Point Functions

We have now formulated and confirmed scalar field/CFT duality for two-point correlation functions. To find higher order correlation functions we need to introduce interaction terms. The "classical approximation" is then obtained by working only at tree level in a so-called *Witten diagram*, a Feynman diagram with the external legs anchored on the boundary of AdS. We will now demonstrate how this works and present the rules for computing n-point correlation functions, following the presentation of Freedman and van Proeyen in [48].

The starting point is now the more general action

$$S[\Phi] = \int \mathrm{d}^{d+1} y \sqrt{g} \left(\frac{1}{2} |\mathrm{d}\Phi|^2 + \frac{1}{2} m^2 \Phi^2 + \frac{1}{3} b \Phi^3 + \frac{1}{4} c \Phi^4 \dots \right) , \qquad (6.74)$$

where ... denotes terms of increasing order in Φ . The equations of motion are then

$$(-\nabla_M \nabla^M + m^2)\Phi + b\Phi^2 + c\Phi^3 + \dots = 0.$$
 (6.75)

We are once again interested in constructing the solution for an arbitrary boundary field by going the to half-plane coordinates and finding a K that gives a delta on the boundary. Since K depends only on x_0 we we have the equations of motion

$$\left(x_0^{d+1}\frac{\mathrm{d}}{\mathrm{d}x_0}x_0^{-d+1}\frac{\mathrm{d}}{\mathrm{d}x_0} - m^2\right)K(x_0) + bK^2(x_0) + cK^3(x_0) + \ldots = 0.$$
(6.76)

The ansatz $K = x^{d+\lambda_+}$ is no longer a valid solution to the differential equation. To deal with this we will consider only a perturbative solution around the solution to the non-interacting theory. We let $K_{\Delta} = x^{d+\lambda_+}$ be the zeroth order Green's function which we call the *bulk to boundary propagator*. We then write the scalar field as

$$\Phi_{1}(x) = c' \int d^{d} \mathbf{x}' K_{\Delta}(x_{0}, \mathbf{x} - \mathbf{x}') \phi(\mathbf{x}')$$

$$\Phi = \Phi_{1} + b \int d^{d+1} x' \sqrt{g} G(x - x') \Phi_{1}^{2}(x') + \dots$$
(6.77)

The perturbative solution is a perturbation around a solution that is hilariously divergent as $z_0 \rightarrow 0$. This is because the zeroth order term defines a δ distribution on the boundary. The square of the zeroth-order term is then the square of a distribution on the boundary, which we absolutely cannot expect to converge when integrated over the boundary⁴. To deal with this we will have to renormalize the theory by introducing a cutoff in the bulk; we let the boundary be defined by $z_0 = \epsilon$ and renormalize in terms of this cutoff parameter, this is called *holographic renormalization theory*, and we will touch upon this later.

At this point, the interaction terms require that we specify the bulk to bulk propagators G(x-x') that satisfy

$$\sqrt{g}(-\nabla'_{M}\nabla^{M'} + m^{2})G(x - x') = \delta(x - x')$$
(6.78)

where ∇'_M denotes the general covariant derivative with respect to x'. Defining $\Delta = d + \lambda_+$ it can be shown that this is solved in terms of the hypergeometric function F(a,b;c;z)

$$G_{\Delta}(u) = \tilde{C}_{\Delta}(2u^{-1})^{\Delta}F\left(\Delta, \Delta - d + \frac{1}{2}; 2\Delta - d + 1; -2u^{-1}\right)$$

$$\tilde{C}_{\Delta} = \frac{\Gamma(\Delta)\Gamma(\Delta - \frac{1}{2}d + \frac{1}{2})}{(4\pi)^{(d+1)/2}\Gamma(2\Delta - d + 1)L^{d-1}} \quad u = \frac{(x - x')^2}{2x_0x'_0}$$
(6.79)

where the hypergeometric function F(a,b;c;d) is given by

$$F(a,b;c;z) = \sum_{n} \frac{(a)_{n}(b_{n})}{(c)_{n}} \frac{z^{n}}{n!}$$

$$(q)_{n} = \begin{cases} 1, & n = 0\\ q(q+1)\dots(q+n-1), & n > 0 \end{cases}$$
(6.80)



Figure 6.2: The perturbative expansion in the scalar field Φ_1 that solves the massive wave equation in AdS can be interpreted in terms of Witten diagrams. The interior of the disk represents AdS space, the boundary represents the space where the boundary CFT lives.

Inserting the perturbative solution $\Phi(x)$ into the classical action in equation (6.74) we find an expansion of the on shell action $S[\phi]$ in powers of ϕ . The terms in this expansion can be interpreted as Witten diagrams, like those in figure 6.2. Analogously to the Feynman rules for Feynman diagrams, we then have computation rules as follows:

- The external points are fixed.
- For each vertex in the bulk with coordinate x, integrate over all possible interaction points, $\int d^{d+1}x \sqrt{g(x)}$.
- For each n-point vertex, add a coupling factor (such as b for the ϕ^3 vertex) corresponding to the relevant interaction terms of the Lagrangian. In addition to this, multiply by the same combinatoric weights as in Feynman diagrams.
- Each bulk to boundary line carries a factor K_{Δ} , and each bulk to bulk line carries a factor of G_{Δ} .
- Sum over all diagrams with the same external states.
- In the classical approximation, only include diagrams with no internal loops.

As is the case for the usual QFT, when we generalize this formalism to interacting *gauge* theories some diagrams may be anomalous, meaning the diagrams do not have the same symmetries as the original action. In this case anomaly cancellation may put restrictions on the field content of the boundary CFT.

As a tractable example, let us compute the three-point function. This diagram contains no bulk-bulk propagators, and going to the usual half-plane representation of AdS the diagram is given by

$$A(x,y,z) = \int \frac{\mathrm{d}^{d+1}w}{w_0^{d+1}} \left(\frac{w_0}{|w-x|^2}\right)^{\Delta} \left(\frac{w_0}{|w-y|^2}\right)^{\Delta} \left(\frac{w_0}{|w-z|^2}\right)^{\Delta} .$$
(6.81)

⁴This is fundamentally the same thing that causes the perturbative expansion in ordinary QFT to diverge.

where $|w - x|^2 = w_0^2 + |\vec{w} - \vec{x}|^2$ since x, y, z all live on the boundary where the zeroth component is zero.

We can check that this amplitude transforms in the correct way under an inversion, sending all coordinates $q_M = \frac{q'_M}{q'^2}$. To figure out how A(x,y,z) transforms, we showed in section 3.4.2 using local Lorentz transforms and an "inversion matrix", that under inversion

$$(x-y)^{2} = \frac{(x'-y')^{2}}{|x'|^{2}|y'|^{2}} .$$
(6.82)

The transformation of A is then fairly easy to deduce. The integration measure carries d + 1 factors of w in denominator and numerator so it is clearly invariant. The factor $1/|w|^2$ from the denominators in the remaining terms cancels to the transformation of w_0 and only the $1/|x|^2$, $1/|y|^2$, $1/|z|^2$ factors survive in the denominators. Then, since we do not integrate over the x, y, z we can move them outside the integral and find

$$A(x,y,z) = |x'|^{2\Delta} |y'|^{2\Delta} |z'|^{2\Delta} A(x',y',z') .$$
(6.83)

Note that it is very straightforward to generalize this to a coupling $b\phi_1\phi_2\phi_3$ between three distinct scalar fields of different masses that are dual to conformal field operators of weights $\Delta_1, \Delta_2, \Delta_3$ since the derivation used no "cross-talk" between the propagators. This means that we can easily describe a dual CFT containing conformal primary fields of varying conformal weight Δ by considering a theory in the bulk wherein scalar fields of different mass couple to each other. With this generalization we get the scale factor $|x'|^{2\Delta_1}|y'|^{2\Delta_2}|z'|^{2\Delta_3}$.

To see that this is the transformation property of a three-point function for conformal primary scalars, remember from section 3.4.3, equation (3.178) that

$$\langle \mathcal{O}_1 \mathcal{O}_2 \mathcal{O}_3 \rangle = \frac{C_{\mathcal{O}_1 \mathcal{O}_2 \mathcal{O}_3}}{(x-y)^{\Delta_1 + \Delta_2 - \Delta_3} (y-z)^{\Delta_2 + \Delta_3 - \Delta_1} (z-x)^{\Delta_3 + \Delta_1 - \Delta_2}}, \qquad (6.84)$$

with $C_{\mathcal{O}_1\mathcal{O}_2\mathcal{O}_3}$ a constant that is constrained by Ward identities. It is straightforward to use the inversion matrix argument here as well, the |x| factor will have exponent

$$\Delta_1 + \Delta_2 - \Delta_3 + \Delta_3 + \Delta_1 - \Delta_2 = 2\Delta_2 , \qquad (6.85)$$

and similarly for |y| and |z|. Setting $\Delta_1 = \Delta_2 = \Delta_3 = \Delta$ we see that this reduces to the exact same transformation rule as in equation (6.83). Using the local Lorentz transform formalism of section 3.4.3 it is also straightforward to also show that A(x,y,z) is completely invariant under global SO(d)transformations as well as translations in the *d*-dimensional boundary space. This guarantees that after performing the integral in A(x,y,z) we will obtain an expression that is proportional to the unique spacetime form of the three-scalar correlator in equation (6.84).

It would be interesting to perform the integral explicitly, to directly determine the constant $C_{\mathcal{O}_1\mathcal{O}_2\mathcal{O}_3}$ in the dual CFT. The integral is in general difficult to handle, but we can use translation symmetry together with inversion on the boundary to simplify. First, we move the point $\vec{z} \to 0$, so that $A(x,y,z) \to A(x-z,y-z,0) \equiv A(u,v,0)$. This simplifies the third term in A so that it is just

$$\left(\frac{w_0}{|w-z|^2}\right)^{\Delta_3} \to \left(\frac{w_0}{|w|^2}\right)^{\Delta_3} = w_0^{'\Delta_3} ,$$
 (6.86)

where we have taken the liberty of letting the scalar fields have different mass since it adds no complication. Performing the inversion of all coordinates, letting $q = \frac{q'}{q^2}$, we find that

$$A(u',v',0) = |u'|^{2\Delta_1} |v'|^{2\Delta_2} \int \frac{\mathrm{d}^{d+1}w'}{w_0^{'d+1}} \left(\frac{w_0'}{|w'-x'|^2}\right)^{\Delta_1} \left(\frac{w_0'}{|w'-y'|^2}\right)^{\Delta_2} w_0^{'\Delta_3} . \tag{6.87}$$

This integral can be computed using Feynman parameter methods to find that

$$A(u',v',0) = |u'|^{2\Delta_1} |v'|^{2\Delta_2} \frac{a(\Delta_1 \Delta_2 \Delta_3)}{|u'-v'|^{\Delta_1 + \Delta_2 - \Delta_3}} , \qquad (6.88)$$

where

$$a(\Delta_1, \Delta_2, \Delta_3) = \frac{\pi^{d/2}}{2} \frac{\Gamma\left(\frac{1}{2}(\Delta_1 + \Delta_2 + \Delta_3 - d)\right)}{\Gamma(\Delta_1)\Gamma(\Delta_2)\Gamma(\Delta_3)} \times \Gamma\left[\frac{1}{2}(\Delta_1 + \Delta_2 - \Delta_3)\right] \Gamma\left[\frac{1}{2}(\Delta_2 + \Delta_3 - \Delta_1)\right] \Gamma\left[\frac{1}{2}(\Delta_3 + \Delta_1 - \Delta_2)\right] .$$

By reverting to the old coordinates we can see that equation (6.88) is indeed the CFT three-point function with $a(\Delta_1, \Delta_2, \Delta_3) = C_{\mathcal{O}_1 \mathcal{O}_2 \mathcal{O}_3}$.

$$A(u,v,0) = \frac{1}{|u|^{2\Delta_1}|v|^{2\Delta_2}} |u|^{\Delta_1 + \Delta_2 - \Delta_3} |v|^{\Delta_1 + \Delta_2 - \Delta_3} \frac{a(\Delta_1, \Delta_2, \Delta_3)}{|u - v|^{\Delta_1 + \Delta_2 - \Delta_3}}$$

$$= \frac{a(\Delta_1, \Delta_2, \Delta_3)}{|u - v|^{\Delta_1 + \Delta_2 - \Delta_3} |v|^{\Delta_2 - \Delta_1 + \Delta_3} |u|^{\Delta_1 - \Delta_2 + \Delta_3}}$$

$$\langle u = x - z, v = y - z \rangle = \frac{a(\Delta_1, \Delta_2, \Delta_3)}{|x - z|^{\Delta_1 - \Delta_2 + \Delta_3} |x - y|^{\Delta_1 + \Delta_2 - \Delta_3} |y - z|^{\Delta_2 - \Delta_1 + \Delta_3}}.$$

(6.89)

Thus, we have found more evidence for the scalar field/CFT correspondence by computing the bulk three-point functions and comparing it to the CFT expectation. In principle, we could now go on to try and show that *all n*-point functions match the CFT expectation, but that is beyond the scope of this text.

For higher *n* correlators, we need to introduce the cutoff $z_0 = \epsilon$ to regularize the correlation functions. It is then nontrivial to show that the divergent parts of the Witten diagrams can be cancelled by local counterterms, resulting in holographic renormalization. This is sketched in more detail in a following section on gravity. Also, without introducing gravity, we will run into problems of consistency because it turns out that the boundary stress-energy tensor is related to perturbations of the bulk metric. All CFTs have a stress tensor so a duality that holds to all orders for all operators is likely impossible to realize without gravity.

6.2.2 Yang-Mills/CFT Correspondence

In a theory containing Yang-Mills fields A with some field strength two-form F = DA, with $D = d + [A \land \cdot]$ the gauge covariant derivative, the free massless equation of motion is usually given by the minimal Yang-Mills equation

$$D * F = 0$$
, (6.90)

or other, more general equations of motion obtained by adding higher order gauge invariant terms to the Lagrangian. A standard example is the addition of Chern-Simons terms to the action, proportional to a power of $\text{Tr}[(F \wedge)^n]$. For gauge fields we hope that there is an analog of the result for scalar fields stating that there is a unique extension (up to gauge transformations) to the bulk for each boundary field A_0 . We will find that this can only hold for A_0 sufficiently close to zero. The equations of motion for nonabelian gauge theories reduce to Maxwell to first order in A_0 , which we can use to show that there is a unique interior extension of A_0 to first order for all gauge fields by explicitly solving the Maxwell equations in the interior of AdS. This explicit proof will be carried out in the next section. To higher order in A, following Witten [47] we can find a topological reason for why uniqueness should fail for general A_0 . Let us carry out this proof while assuming d even, we will then comment on the odd case. Let us consider a gauge theory on $S^2 \times S^d$ with nonzero $\frac{d}{2} + 1$:th Chern class, meaning

$$\int_{\mathbf{S}^2 \times \mathbf{S}^d} \operatorname{Tr}\left[(F \wedge)^{\frac{d}{2}+1} \right] = \int_{\mathbf{S}^2 \times \mathbf{S}^d} \operatorname{Tr}\left[F \wedge F \wedge \ldots \wedge F \right] \neq 0 .$$
(6.91)

In general we cannot expect the *n*:th Chern class to be zero for a nontrivial boundary field A_0 . If every gauge field on S^d could be extended uniquely up to gauge transformations to the interior B^{d+1} , then by making the extension for each point on S^2 one could uniquely extend the theory on $S^2 \times S^d$ to $S^2 \times B^{d+1}$. Using the graded Leibniz rule for the exterior derivative d and the fact that dA is a two-form we know that

$$\mathrm{dTr}[F^n] = n\mathrm{Tr}[F^{n-1} \wedge \mathrm{d}F] = n\mathrm{Tr}[F^{n-1} \wedge (\mathrm{d}^2A + \mathrm{d}A \wedge A - A \wedge \mathrm{d}A)] = n\mathrm{Tr}[F^{n-1} \wedge 0] = 0 ,$$

and we have by partial integration that

$$0 = \int_{\mathbf{S}^2 \times \mathbf{B}^{d+1}} \mathrm{d}\mathrm{Tr}\left[F \wedge \ldots \wedge F\right] = \int_{\mathbf{S}^2 \times \mathbf{S}^d} \mathrm{Tr}\left[F \wedge \ldots \wedge F\right],\tag{6.92}$$

contradicting the supposition that we have a gauge theory with nonvanishing $\frac{d}{2} + 1$:th Chern class. To extend this argument to odd dimensions, replace $S^2 \to S^1$ and consider a nonvanishing $\frac{d+1}{2}$: th Chern class.

The point that lead to this contradiction was the assertion that *every* choice of boundary field uniquely extends to the interior so we have only proven that some care has to be taken. The fact that we can find a unique solution explicitly to first order implies that the problem arises when A_0 is too large, although we have found no natural reference scale with respect to which "large" should be defined.

Uniqueness and Boundary-Boundary Correlation for U(1) Field

Suppose we have a free U(1) gauge theory. We want to find a Green's function that gives us a delta on the boundary, which as usual is most easily carried out in the half-plane representation of Euclidean AdS_{d+1} . Thanks to translation invariance in the boundary coordinates we may consider as ansatz the one-form $A = f(x_0) dx^i$ for some fixed i > 0. Then, $F = dA = f'(x_0) dx^0 \wedge dx^i$. Then, we can compute *F:

$$*F = \frac{1}{x_0^{d-3}} f'(x_0) (-1)^i \mathrm{d}x^1 \wedge \mathrm{d}x^2 \dots \hat{\mathrm{d}x^i} \dots \wedge \mathrm{d}x^d$$
(6.93)

where by dx^i we mean that the differential dx^i should be excluded from the d-1-fold wedge product. The factor $x_0^{-(d-3)}$ comes from the factor of $\sqrt{g} = x_0^{-(d-1)}$ in the Hodge dual as well as the two factors of the metric coming from the contraction of the Levi-Civita symbol with the two lowered indices of dA. The equation of motion d * F = 0 becomes

$$\frac{\partial}{\partial x^0} \left(\frac{1}{x_0^{d-3}} f'(x_0) \right) = 0 , \qquad (6.94)$$

which is solved by $f(x_0) \propto x_0^{d-2}$ (which also happens to diverge at $x_0 = \infty$). Picking the constant for later convenience we can then use as our ansatz

$$A = \frac{d-1}{d-2} x_0^{d-2} \mathrm{d}x^i \ . \tag{6.95}$$

We then perform an inversion to put put our divergence at the origin, setting

$$A = \frac{d-1}{d-2} \left(\frac{x_0}{(x_0^2 + |\mathbf{x}|^2)^d} \right)^{d-2} d\left(\frac{x^i}{x_0^2 + |\mathbf{x}|^2} \right) .$$
(6.96)

We can then use a gauge transformation $A_{\mu} \to A + d\alpha(x)$ where α is any 0-form to rewrite this. With the goal of cancelling the peculiar differential we pick

$$\alpha(x) = \frac{-1}{d-2} \frac{x_0^{d-2} x_i}{(x_0^2 + |\mathbf{x}|^2)^{d-1}} .$$
(6.97)

We then note that by some clever rewriting and using the product rule, we can cancel the original A with

$$\begin{aligned} x_0^{d-2} x_i^{-(d-2)} d\left(\frac{-1}{d-2} \frac{x_i^{d-1}}{(x_0^2 + |\mathbf{x}|^2)^{d-1}}\right) &= -\frac{x_0^{d-2} x_i^{-(d-2)}}{d-2} (d-1) \frac{x_i^{d-2}}{(x_0^2 + |\mathbf{x}|^2)^{d-2}} d\left(\frac{x_i}{x_0^2 + |\mathbf{x}|^2}\right) \\ &= -\frac{d-1}{d-2} \frac{x_0^{d-2}}{(x_0^2 + |\mathbf{x}|^2)^{d-2}} d\left(\frac{x_i}{x_0^2 + |\mathbf{x}|^2}\right) . \end{aligned}$$
(6.98)

After this, we are left only with the terms where the exterior derivative acts on x_0 and the decomposition $d(x_i^{-(d-2)}) x_i^{d-1}$ in the numerator, yielding

$$A = \frac{x_0^{d-2}}{(x_0^2 + |\mathbf{x}|^2)^{d-1}} \mathrm{d}x^i - \frac{x_0^{d-3} x_i}{(x_0^2 + |\mathbf{x}|^2)^{d-1}} \mathrm{d}x^0 .$$
(6.99)

This solution is a Green's function for the boundary field at $x_0 = 0$, so for a general boundary field $A_0 = \sum a_i dx^i$ where $i \in [1,d]$ we write

$$A(x_0, \mathbf{x}) = \int d\mathbf{x}' \frac{x_0^{d-2} a_i(\mathbf{x}')}{(x_0^2 + |\mathbf{x} - \mathbf{x}'|^2)^{d-1}} dx^i - x_0^{d-3} dx^0 \int d\mathbf{x}' \frac{(x^i - x^{i'}) a_i(\mathbf{x}')}{(x_0^2 + |\mathbf{x} - \mathbf{x}'|^2)^{d-1}} .$$
 (6.100)

Then, F = dA is expressed by

$$F = (d-1)x_0^{d-2} dx^0 \wedge \int d\mathbf{x}' \frac{a_i(\mathbf{x}') dx^i}{(x_0^2 + |\mathbf{x} - \mathbf{x}'|^2)^{d-1}} - 2(d-1)x_0^{d-1} dx^0 \wedge \int d\mathbf{x}' \frac{a_i(\mathbf{x}') dx^i}{(x_0^2 + |\mathbf{x} - \mathbf{x}'|^2)^d} - 2(d-1)x_0^{d-3} dx^0 \wedge \int d\mathbf{x}' \frac{(x^i - x^{i'})a_i(\mathbf{x}')(x_k - x'_k) dx^k}{(x_0^2 + |\mathbf{x} - \mathbf{x}'|^2)^d} + \dots$$
(6.101)

where ... denotes terms that do not contain dx^0 . To correctly add the terms it is important to keep track of in which order the differentials appear after acting with the exterior derivative. We now want to write down the on-shell action. The Maxwell action in the absence of charges is

$$S = \frac{1}{2} \int_{B_{d+1}} F \wedge *F = \frac{1}{2} \int_{S^d} A \wedge *F , \qquad (6.102)$$

where in principle we should work with a limit prescription, taking the limit as $x_0 \to 0$. Inserting the above expression for F and using the usual trick of setting $A = A_0$ on the boundary it is possible to work out that the on-shell action is

$$S[A_0] = \int d\mathbf{x} d\mathbf{x}' \, a_i(\mathbf{x}) a_j(\mathbf{x}) \left(\frac{\delta^{ij}}{|\mathbf{x} - \mathbf{y}|^{2d-2}} - \frac{2(x - x')^i (x - x')^j}{|\mathbf{x} - \mathbf{x}'|^{2d}} \right) \,. \tag{6.103}$$

Which is exactly the result we stated for a CFT conserved vector current in equation (3.181), realized by inserting the definition of the inversion matrix. It is straightforward to see that the first term should have a δ^{ij} since after acting with the Hodge on $(dx^0 dx^i)$ the only basis vector that does not act as in $dx^k \wedge F^* = 0$ while also being a boundary coordinate is precisely dx^i .

With this we have seen that in addition to there being a unique extension to the interior in the case of an Abelian gauge field A, we also reproduce a two-point correlation function that matches the CFT expectation for a (conserved) boundary vector field.

p-Form Gauge Fields

Physically interesting theories in AdS such as supergravity also contain higher form fields, such as the Kalb-Ramond two-form $B_{\mu\nu}dx^{\mu}\wedge dx^{\nu}$ and the A-fields introduced in section 5.2. They transform not only as conformal scalars, as they may have nonzero eigenvalues under transformations other than dilatations. In general, *p*-form fields A_p couple to a (d-p)-form conformal field \mathcal{O} on the boundary according to

$$\int_{\partial \mathrm{AdS}_{d+1}} A \wedge \mathcal{O} \ . \tag{6.104}$$

We can see a general p-form field as as an object that behaves as a scalar in addition to containing p differentials. Each of the differentials dx^{μ} carry weight -1 under scaling. In the massless case, this just means that we require \mathcal{O} to be of weight d - p. Note that \mathcal{O} is not expected to be a conformal primary since the boundary field A_{p0} is expected to transform under boundary Lorentz transformations. For p = 1 this means that \mathcal{O} transforms like a boundary vector, with conformal weight d - 1, which is exactly the result we explicitly obtained in equation (6.103).

If instead we consider the case of massive *p*-form field that behaves as $f^{-\lambda_+}A_{p0}$ near the boundary, we expect that \mathcal{O} should have weight $\Delta = d - p + \lambda_+$, where λ_+ is the larger eigenvalue that solves equation (6.58).

6.2.3 Gravity/CFT Correspondence

The goal of this discussion is to understand a theory that contains both quantum fields and gravity in the bulk of the AdS spacetime. Let us therefore consider under what spacetime deformations the argument of the previous sections hold. The proof for the uniqueness of solutions Φ given a boundary field ϕ used the following:

- uniqueness of the solution holds as long as the manifold has a boundary on which to define boundary conditions,
- the existence of nontrivial solutions relied on spherical symmetry for large r.

Based on this, any metric that looks like AdS_{d+1} for large r should have unique, nontrivial solutions to the Laplace equation for the scalar field Φ given some boundary field ϕ .

The dynamics of gravity in this case are described by Einstein's equations with a negative cosmological constant

$$R_{\mu\nu} - \frac{R}{2}g_{\mu\nu} - \frac{d(d+1)}{2L^2}g_{\mu\nu} = 0 , \qquad (6.105)$$

where L is the characteristic radius of curvature for the AdS_{d+1} space and $-\frac{d(d-1)}{2L^2} = \Lambda$. Any metric X_{d+1} on $\overline{B_{d+1}}$ with a double pole on the boundary will induce a conformal structure on S^d . Furthermore due to a theorem of Graham and Lee, any conformal structure induced on the

boundary that is sufficiently close to the usual one⁵ arises by the procedure in equation (6.24) from a unique⁶ metric that obeys the Einstein equations and has a double pole at the boundary.

Therefore, to include gravity we simply allow the boundary CFT to exist on a spacetime that is not conformally flat. We know that a conformal field theory on a nonflat background will have a Weyl anomaly, and this will indeed appear as a result of the necessity of holographic renormalization.

Holographic Renormalization of the Gravitational On-Shell Action

Let us now study the relationship between the on-shell partition function for gravity in the interior of AdS and the boundary CFT in some detail. We will follow the discussion of Henningson and Skendris [49]. This derivation takes a fair bit of input from difficult mathematical results, and a fully explicit derivation is outside the scope of this text.

Since AdS has a boundary we need to include a boundary term if we want to properly write down the Einstein Hilbert action. This boundary term contains the so-called second fundamental form or extrinsic curvature, $K = \nabla_{\mu} n^{\mu}$, where n^{μ} is a unit normal to the boundary. The action then reads

$$S = \frac{1}{16\pi G} \left[\int_{B_{d+1}} \sqrt{g} \left(-\frac{1}{2}R + \Lambda \right) - 2 \int_{\mathbf{S}^d} \sqrt{\gamma} 2K \right] , \qquad (6.106)$$

where γ is the induced metric on the boundary. Since we are including gravity, the metric on the ball may be a nonstatic solution to the Einstein equation inducing a boundary metric that is not conformally flat. By multiplying the field equations by $g^{\mu\nu}$ we find that

$$R = 2(d+1)\frac{\Lambda}{d-1} \ . \tag{6.107}$$

This result leads us to conclude that the on-shell action for gravity is badly divergent. The integral over the ball B_{d+1} is proportional to the volume of AdS, which is infinite. There is also the question of how to interpret the boundary term, since we know that AdS does not actually induce a metric on the boundary.

To try and get further, we regularize the action by picking a function f on B_{d+1} with a simple zero on the boundary and multiplying the metric. This breaks conformal invariance, but it also introduces a proper metric on the boundary via equation (6.24). The divergence of the volume integral is regulated by integrating only over the region $B(\epsilon)$ defined by $f > \epsilon$. The boundary term may then be well defined since we just integrate over the boundary of the finite region $B(\epsilon)$ finding no infinities. The expectation is that divergences will appear as we take the limit $\epsilon \to 0$.

To pick a good defining function f, the expectation is that we need to know very well how the metric behaves near the boundary of AdS. For any asymptotically (Euclidean) AdS spacetime, we can pick near boundary Fefferman-Graham coordinates as in equation (6.31) with z > 0 and metric

$$ds^{2} = \frac{L^{2}}{z^{2}} \left[dz^{2} + h_{ij}(x,z) dx^{i} dx^{j} \right]$$
(6.108)

where h_{ij} is the transverse metric, fulfilling $\lim_{z\to 0} h_{ij} = \gamma_{ij}$. To make the divergences on the boundary local, Fefferman and Graham found that it was convenient to pick the coordinate system

⁵The "usual" conformal structure is a boundary metric that is related to the flat metric by a conformal transformation. "Sufficiently close" then means that the induced boundary metric is conformally equivalent to a sufficiently mild deformation of flat space. Quantifying exactly what "sufficiently small" means is beyond the scope of this thesis and we refer to [52]. A similar theorem for scalar fields and *p*-forms when the boundary is almost conformally flat is called the Fefferman-Graham theorem.

⁶Here, unique means unique up to diffeomorphisms.

with $\rho = z^2$ on B_{d+1} , with the metric

$$ds^{2} = \frac{L^{2}}{4}\rho^{-2}d\rho^{2} + \rho^{-1}h_{ij}(\rho, x)dx^{i}dx^{j} , \qquad (6.109)$$

These coordinates are special in the sense that they give a natural definition of the defining function f that makes the divergent part of the action depend locally on the boundary metric γ_{ij} . When we have a local γ_{ij} dependence the addition of counterterms becomes well-defined and renormalization can succeed.

In the following, let us denote $h' = \frac{\partial h}{\partial \rho}$. We furthermore treat the metric as a matrix h instead of a tensor h_{ij} to save space. With these conventions and the choice of coordinates in equation (6.109) the Einstein field equations become

$$\rho \left(2h'' - 2h'h^{-1}h' + \operatorname{Tr}[h^{-1}h']h' \right) + L^2 Ric(h) - (d-2)h' - \operatorname{Tr}[h^{-1}h']h = 0$$

$$\nabla_i \operatorname{Tr}[\gamma_{(0)}^{-1}h'] - \nabla^j h'_{ji} = 0$$

$$\operatorname{Tr}[h^{-1}h''] - \frac{1}{2} \operatorname{Tr}[h^{-1}h'h^{-1}h'] = 0$$
(6.110)

where ∇_i is the covariant derivative constructed from h, Ric(h) denotes the Ricci tensor corresponding to the metric h, and the indices in the second equality denote the free matrix indices of the expression.

The method of Fefferman and Graham was to attempt to solve these equations via a formal power series in ρ . They showed that by picking the defining function according to $f = \rho$ the boundary metric γ uniquely defines $h(\rho, x)$ on the interior of AdS order by order in ρ . For each order in ρ , one applies a derivative with respect to ρ and takes the limit $\rho \to 0$. An interesting note is that a similar ansatz is necessary for near boundary analysis of gravity with extra fields, and similar problems exist when certain integer powers of ρ have nonzero coefficients. The detailed solution of the equations and determination of the coefficients can be found in [53].

For odd d one found that the metric may be written as

$$h(\rho, x) = \gamma_{(0)} + \rho \gamma_{(2)} + \rho^2 \gamma_{(4)} \dots$$
(6.111)

where $\gamma_{(k)}(x)$ is given by some covariant combination of the boundary metric γ , its Riemann tensor and covariant derivatives constructed from γ . Here and henceforth, a subscript in parentheses denotes the number of derivatives with respect to x_i . Since $[\rho] = (\text{length})^2$ we need two derivatives per order of ρ for dimensional reasons in the expansion.

For d even the iterative procedure breaks down at order $\rho^{d/2}$ where a logarithmic divergence appears. To cancel the divergence a logarithmic term is added and the expansion takes the form

$$h(\rho, x) = \gamma_0 + \rho \gamma_{(2)} \dots + \rho^{d/2} \gamma_{(d)} + \tilde{\gamma}_{(d)} \rho^{d/2} \ln \rho + \mathcal{O}(\rho^{d/2+1}) .$$
 (6.112)

Here, the $\gamma_{(k)}$ are covariant up to and including k = d/2, while beyond that covariance will fail. All terms except the traceless transverse part of $\gamma_{(d)}$ are local functions of the leading term $\gamma_{(0)}$. The trace of $\tilde{\gamma}_{(d)}$ $\text{Tr}[\gamma^{(0)}\tilde{\gamma}_{(d)}] = 0$ vanishes identically and $\tilde{\gamma}_{(d)}$ is covariantly conserved, meaning $\nabla^i \tilde{\gamma}_{(d)ij} = 0$. We will see that the breakdown at d/2 is not too bad, because all terms beyond order d/2 in the action are finite.

Next, we introduce the regularization in terms of ϵ , letting $B(\epsilon)$ be defined by $\rho > \epsilon$ and S_{ϵ} be the surface at $\rho = \epsilon$. In the near boundary coordinates, using that the full Ricci scalar is given by

equation (6.107) it can be checked that the action takes the form

$$S[h] = \frac{1}{16\pi G} \left[\frac{d}{L} \int_{B(\epsilon)} \mathrm{d}\rho \mathrm{d}^d x \, \rho^{-d/2 - 1} \sqrt{h} + \int_{S_\epsilon} \mathrm{d}^d x \rho^{-d/2} \left(\underbrace{-\frac{2d}{L} \sqrt{h} + \frac{4}{L} \rho \partial_\epsilon \sqrt{h}}_{K} \right) \right] \,, \qquad (6.113)$$

where the lower limit for the ρ integration is $\rho = \epsilon$. Formally, it is possible to perform the radial integration in the first term. We can then define the Lagrangian by $S = (16\pi G)^{-1} \int_{S_{\epsilon}} d^d x \mathcal{L}(\epsilon, x)$ and write it as an expansion in ρ . We then have, for odd and even d respectively

$$\mathcal{L}_{\text{odd}} = \sqrt{h_{(0)}} \left(\epsilon^{-d/2} a_{(0)} + \epsilon^{-d/2+1} a_{(2)} + \dots + \epsilon^{-1/2} a_{(d-1)} \right) + \text{finite terms}$$
(6.114)

$$\mathcal{L}_{\text{even}} = \sqrt{h_{(0)}} \left(\epsilon^{d/2} b_{(0)} + \epsilon^{-d/2+1} b_{(2)} + \dots + \epsilon^{-1} b_{(d-2)} - \log \epsilon b_{(d)} \right) + \text{finite terms} , \qquad (6.115)$$

where the coefficients $a_{(k)}, b_{(k)}$ involve covariant expressions containing k derivatives on h. Note that the term $\rho^{d/2}\gamma_{(d)}$ in equation (6.112) that was dynamically undetermined for even d is a finite constant in ρ and so does not enter into the renormalization procedure. This means that all of the infinities are local, covariant expressions of γ , there is a finite number of them, and their form is determined by the classical equations of motion. In principle, it is possible to compute all of the counterterms, subtract them and obtain a renormalized action $S_R[h]$.

Holographic Weyl Anomaly

By inducing an actual metric h on the boundary instead of a conformal equivalence class, we cannot expect the renormalized theory to respect conformal invariance. Let \mathcal{L}_R denote the corresponding renormalized boundary Lagrangian, where the integration over ρ in the bulk part of the action has been carried out.

We want to know what happens when we act with a infinitesimal conformal transformation $\delta h_{(0)} \rightarrow 2\sigma(x)h_{(0)}$ on the boundary metric. The variation will take the general form

$$\delta_{\sigma} S_R = -\int_{\mathbf{S}^d} \mathrm{d}^d x \sqrt{h_{(0)}} \mathcal{A} \delta \sigma \tag{6.116}$$

where \mathcal{A} is called the *anomaly*, since it encodes how badly conformal invariance is violated. For odd d, \mathcal{A} vanishes while in the even case

$$\mathcal{A} = \frac{1}{16\pi G} (-2b_{(d)}) \ . \tag{6.117}$$

To see this, consider the case when σ is a constant parameter, representing a uniform scale transformation of the boundary space. Since a scale transformation of the boundary is induced by a redefinition of the defining function⁷ $\delta \rho = 2\sigma \rho$ the regularized action must be invariant under the joint transformation $\delta \epsilon = \epsilon 2\sigma$, $\delta h_{(0)} = h_{(0)} 2\sigma$.

The terms that are proportional to a negative power of ϵ are separately invariant, so the variation of the ln ϵ and the finite terms must cancel. Then, since $\sqrt{h_0}b_{(d)}$ is invariant and $\delta \log \epsilon = 2\sigma$ the finite part of the Lagrangian must transform as $-2\sigma b_{(d)}$. The finite part of the Lagrangian is all that is left after renormalization, so we find that the anomaly must take the form of equation (6.116).

⁷The coordinate ρ is related to the near horizon coordinate z by $\rho = z^2$. In turn the near boundary coordinates z are the inversion of the Euclidean AdS radial coordinates. Therefore, ρ has conformal dimension 2. More rigorously, see [53] section 4 where they implement the boundary conformal transformation as a special diffeomorphism that preserves the form of the metric equation (6.109).

An AdS_3/CFT_2 Example

Let us compute explicitly the Weyl anomaly in the case where the bulk theory is AdS_3 . Since the Weyl anomaly on the CFT₂ side takes the form given by equation $(3.239)^8$

$$\langle T^{\alpha}{}_{\alpha} \rangle = -\frac{c}{24\pi} R , \qquad (6.118)$$

we will be able to read off the central charge of a CFT that is dual to an asymptotically AdS_3 spacetime. In AdS_3 with the conventions of this section the perturbative expansion for the metric is

$$h(\rho, x) = \gamma_0 + \rho \gamma_{(2)} + \tilde{\gamma}_{(2)} \rho \ln \rho + \mathcal{O}(\rho^2) , \qquad (6.119)$$

taking γ_0 as known and as a solution to the Einstein field equations.

For the Einstein field equations we want the derivatives and inverse of h

$$h' = \gamma_{(2)} + \tilde{\gamma}_{(2)} (1 + \ln \rho) ,$$

$$h'' = \frac{\tilde{\gamma}_{(2)}}{\rho} ,$$

$$h^{-1} = \gamma_{(0)}^{-1} + \mathcal{O}(\rho) .$$

(6.120)

The iterative procedure is to evaluate the Einstein equations equation (6.110) in the limit $\rho \to 0$, then take a ρ derivative of the equations and repeat the process for the next order. This is done d/2 - 1 times, so in the AdS₃/*CFT*₂ case, everything is determined by the zeroth order equations in the limit $\rho \to 0$. Using that $\lim_{\rho\to 0} \rho \ln(\rho) = 0$ we find

$$\begin{split} \rho 2h''|_{\rho=0} &= 2\tilde{\gamma}_{(2)} ,\\ -2\rho h' h^{-1} h'|_{\rho=0} &= 0 ,\\ \rho \mathrm{Tr}[h^{-1} h'] h'|_{\rho=0} &= 0 ,\\ L^2 Ric(h)|_{\rho=0} &= L^2 Ric(\gamma_{(0)}) ,\\ (d-2)h'|_{d=2,\rho=0} &= 0 ,\\ -\mathrm{Tr}[h^{-1} h'] h|_{\rho=0} &= -\lim_{\rho \to 0} \mathrm{Tr}[\gamma_0^{-1}(\gamma_{(2)} + \tilde{\gamma}_{(2)}(1+\ln\rho))]\gamma_{(0)} . \end{split}$$
(6.121)

In the last line, we see clearly why we must define $\tilde{\gamma}_{(2)}$ to be traceless. Inserting this definition we have

$$-\operatorname{Tr}[h^{-1}h']h|_{\rho=0} = -\operatorname{Tr}[\gamma_0^{-1}\gamma_{(2)}]\gamma_{(0)} .$$
(6.122)

The vanishing of the covariant divergence of $\tilde{\gamma}$ is similarly meant to exclude the logarithmic divergence for the second row of equation (6.110). Using this the Einstein equations become

$$2\tilde{\gamma}_{(2)} + L^{2}Ric(\gamma_{(0)}) = \operatorname{Tr}[\gamma_{0}^{-1}\gamma_{(2)}]\gamma_{(0)} ,$$

$$\nabla_{i}\operatorname{Tr}[\gamma_{0}^{-1}\gamma_{(2)}] - \nabla^{j}\gamma_{ji}^{(2)} = 0 ,$$

$$\operatorname{Tr}[\gamma_{(0)}^{-1}\tilde{\gamma}_{(2)}] = 0 = \frac{1}{2}\operatorname{Tr}[\gamma_{0}^{-1}\gamma_{(2)}\gamma_{0}^{-1}\gamma_{(2)}] .$$
(6.123)

⁸In the string theory chapter we defined the stress tensor as $\frac{4\pi\alpha'}{\sqrt{h}}\frac{\delta S}{\delta G}$, while in gravitational contexts it is conventional to define it by $\frac{2}{\sqrt{h}}\frac{\delta S}{\delta G}$, so with the current conventions the Weyl anomaly picks up an extra factor $\frac{1}{2\pi}$. We implicitly used the latter definition in writing down equation (6.116).

Taking the trace of the first equation (by contracting the present matrices by $\gamma_{(0)}^{-1}$) we find

$$L^2 R_{(0)} = 2 \operatorname{Tr}[\gamma_0^{-1} \gamma_{(2)}] , \qquad (6.124)$$

where $R_{(0)}$ is the zeroth order Ricci scalar. In two dimensions the Riemann tensor has only one independent component and it follows that the Ricci tensor is proportional to the metric. This means that $\gamma_{(0)}$ is necessarily Einstein and the Ricci tensor is given by $\frac{R_{(0)}}{2}\gamma_{(0)}$, which inserted together with equation (6.124) into the first line of equation (6.123) tells us that

$$2\tilde{\gamma}_{(2)} = \underbrace{\left(\operatorname{Tr}[\gamma_0^{-1}\gamma_{(2)}] - \frac{L^2 R_{(0)}}{2}\right)}_{=0} \gamma_{(0)} = 0 , \qquad (6.125)$$

eliminating $\tilde{\gamma}_{(2)}$ from all other equations. The most general solution to the second row of equation (6.123) that respects equation (6.124) is

$$\gamma_{(2)} = \frac{1}{2} \left(R_{(0)} \gamma_{(0)} + T \right) , \qquad (6.126)$$

where T is some symmetric tensor (essentially an integration constant) satisfying

$$\nabla^j T_{ji} = 0$$
, $\operatorname{Tr}[\gamma_0^{-1}T] = 0$. (6.127)

The tensor T has three independent components and equation (6.127) constitutes three equations, two of them differential. This means that we need to supply additional boundary conditions to uniquely determine T. The boundary conditions may break covariance or locality of the tensor Tif picked irresponsibly. To compute the Weyl anomaly we do not need the full metric tensor, so we can just carry on by evaluating the on-shell action equation (6.113) of the gravitational theory using our derived solution

$$h = \gamma_{(0)} + \rho \gamma_{(2)} . \tag{6.128}$$

To perform the computation, we first Taylor expand:

$$\sqrt{h} \approx \sqrt{\gamma_{(0)}} + \frac{\rho}{2} \sqrt{\gamma_0} \operatorname{Tr}[\gamma_{(0)}^{-1} g_{(2)}] + \mathcal{O}(\rho^2) ,$$
 (6.129)

where we have applied the usual identity for the metric variation. Inserting into the on-shell action and performing the radial integration we find

$$16\pi GS[h] = \frac{2}{L} \int_{S} d^{2}x \int_{\epsilon}^{\infty} d\rho \frac{1}{\rho^{2}} \left(\sqrt{\gamma_{(0)}} + \frac{\rho}{2} \operatorname{Tr}[\gamma_{(0)}^{-1}g_{(2)}] \right) + \int_{S_{\epsilon}} d^{2}x \frac{1}{\epsilon} \left(-\frac{2d}{L} \left(\sqrt{\gamma_{(0)}} + \frac{\epsilon}{2} \operatorname{Tr}[\gamma_{(0)}^{-1}g_{(2)}] \right) + \frac{4}{L} \epsilon \partial_{\rho} \left(\sqrt{\gamma_{(0)}} + \frac{\epsilon}{2} \operatorname{Tr}[\gamma_{(0)}^{-1}g_{(2)}] \right) \right) + \text{finite} = \int_{S} d^{2}x \frac{2}{L} \frac{1}{\epsilon} \sqrt{\gamma_{(0)}} - \frac{\ln(\epsilon)}{2} \operatorname{Tr}[\gamma_{(0)}^{-1}g_{(2)}] - \frac{4}{L\epsilon} \sqrt{\gamma_{(0)}} + \text{finite} .$$
(6.130)

From here, we can read off the coefficient $b_{(2)}$ as defined in equation (6.115):

$$b_{(2)} = \frac{1}{L} \operatorname{Tr}[\gamma_{(0)}^{-1} g_{(2)}] = \frac{L}{2} R_{(0)} .$$
(6.131)

The anomaly is then given by

$$\mathcal{A} = \frac{1}{16\pi G} (-2b_{(d)}) = -\frac{LR_{(0)}}{16\pi G} .$$
(6.132)

Comparing to the form of the Weyl anomaly in equation (6.118), given by

$$\langle T^{\alpha}{}_{\alpha} \rangle = -\frac{c}{24\pi} R , \qquad (6.133)$$

we see that if

$$c = \frac{3L}{2G} , \qquad (6.134)$$

the two values agree. We have thus found an explicit relation between the central charge of the boundary CFT and the AdS radius L of the bulk spacetime.

CFT Dual of the Metric

Like we have previously, to finalize the relationship between the bulk and boundary theories we want γ_{ij} to source an operator in the boundary CFT.

The metric transforms as a conformal operator of weight two as well as spin 2. It should source a boundary operator through the coupling $\int d^d x \sqrt{\gamma} \gamma_{ij} \mathcal{O}^{ij}$. The factor $\sqrt{\gamma} \gamma_{ij}$ has conformal weight zero and spin 2, for the same reason as the Polyakov action is Weyl invariant. Therefore, the operator \mathcal{O} should be a spin 2 object with conformal weight d. These are exactly the properties of the stress energy tensor.

Thus, we can in a simple manner argue that the boundary stress tensor is related to the boundary metric in AdS/CFT duality. At this point it is possible to compare with the Holographic Weyl anomaly and conclude that the exact relation is given by [53]

$$\langle T_{ij} \rangle = \frac{dL^{d-1}}{16\pi G_N} g_{(d)ij} + X_{ij}[g_{(n)}]$$
(6.135)

where $X_{ij}[g_{(n)}]$ is a functional of all $g_{(n)}$ with n < d. X_{ij} vanishes in odd boundary dimensions, because it encodes the conformal anomaly due to the bulk gravity. This is covered in more detail in section 3 of [53]. When there are no divergent terms as $\rho \to 0$, equation (6.135) can be put on the form

$$\langle T_{ij} \rangle = \frac{dL^{d-1}}{16\pi G_N} h_{ij}(0,x) .$$
 (6.136)

6.2.4 Type IIB Supergravity

We found in the previous section under what conditions it is reasonable for scalar field theories, Yang-Mills fields and gravity in the bulk to be dual to conformal field theories on the boundary of AdS. Given that all of the ingredients we analyzed, as well as fermions in AdS all have CFT duals on the boundary, we can anticipate that any sufficiently nice combination of these fields in the bulk should have a CFT dual. All of the explicit calculations in the previous sections relied on the validity of a semiclassical approximation in the bulk.

It was the discovery due to Maldacena [3] that there was in fact a correspondence between a particular CFT in M_d and IIB supergravity in $AdS_d \otimes S^5$, *including* quantum corrections to the latter. This means that while string theory is not necessary for the AdS/CFT duality, it is presently our strongest tool for going beyond a semiclassical bulk approximation. Another important property of the duality is that string theory is finite to at least to genus 5 in the string perturbation expansion [54]. Finiteness is also enjoyed by the CFT we are about to discuss.

An equivalence between a QFT that was finite to all orders and quantum gravity was big news, since at the time there was no renormalizable theory of quantum gravity in terms of the experimentally successful framework of QFT. Said differently, naively quantum gravity does not work in QFT due to non-renormalizablity of the Einstein-Hilbert action, but via AdS/CFT there suddenly exists a renormalizable (or possibly even finite) holographic description of quantum gravity in terms of QFT. Another very promising property of the duality is that it relates a weakly coupled string theory to strongly coupled QFT and vice-versa, so we can use the perturbative regime of either theory to study the poorly understood nonperturbative behaviour of the other.

Let us now get an overview of how the duality could be anticipated, and how it was tested. We begin by clarifying the context:

- One of our best candidates for quantum gravity is superstring theory. In the low energy limit, that is at tree level and to first order in the α' -expansion, superstring theory is described by type IIA or type IIB supergravity as in section 5.2.1.
- Super Yang-Mills theory with gauge group SU(N) and four generations of fermionic superpartners (i.e. " $\mathcal{N} = 4$ ") is a QFT that is finite to all orders. Notably $\mathcal{N} = 4$ SYM is also conformally invariant because it can be shown that the β -functions vanish identically to all orders in perturbation theory [55].

The first statement of the AdS/CFT duality regarded the equivalence between $\mathcal{N} = 4$ SYM on the boundary of AdS₅ and type IIB superstring theory on AdS₅ × S⁵. Here, both the AdS₅ and S⁵ spaces have the same radius L. The SYM theory is specified by two dimensionless parameters, the gauge group label N and the Yang-Mills coupling $g_{\rm YM}$. The string theory side is specified by the the length ratio $L/\sqrt{\alpha'}$ and the string coupling g_s .

Aside from both theories having the same number of parameters, they also share symmetries. The conformal group in d dimensions is SO(d,2), moreover the $\mathcal{N} = 4$ SYM group has a socalled *R*-symmetry group given by SU(4). SU(4) is isomorphic to SO(6) (a so-called exceptional isomorphism). AdS₅ and S⁵ can both be embedded in higher dimensional flat spaces to see that they have isometry groups SO(4,2) and SO(6). Since the compactified type IIB supergravity is invariant under global transformations that preserve the background manifold we see that the bosonic part of SYM and a relativistic theory on AdS₅ × S⁵ share the same symmetries.

That the two theories share the same supersymmetries is a nontrivial check and requires the introduction of Green-Schwarz formalism, the manifestly spacetime supersymmetric formulation of string theory.

Explicitly, the statement of the conjectured duality is that

 $\mathcal{N} = 4$ Super Yang Mills is dynamically equivalent to Type IIB Superstring theory on $\mathrm{AdS}_5 \times \mathrm{S}^5$ with radii L and N units of F_5 flux on S^5 , with the parameters $N, g_{YM}, \sqrt{\alpha'}, g_s$ related by $g_{YM}^2 = 2\pi g_s$ and $2g_{YM}^2 N = L^4/\alpha'^2$.

Here dynamically equivalent means that the theories are equivalent on the quantum level, i.e. they have the same partition function⁹. Often, it is more convenient to speak of the *t'Hooft coupling* $\lambda = g_{YM}^2 N$ instead of the Yang-mills coupling g_{YM} .

Explicitly proving the duality for general values of all the parameters has not been done. Instead, it is possible to take limits and consider weaker forms of the duality. The duality comes in the following forms

⁹Sometimes it is easier to show a duality by working with Hamiltonian formalism in both theories and finding that the theories have the same Hilbert space and dynamics, as in the case of T-duality in string theory c.f. chapter 17 in [43].

	$\mathcal{N} = 4$ SYM side	String Theory side
Strongest form	N, λ arbitrary	Full quantum string theory
Strong form	$N \to \infty \ , \lambda$ arbitrary	Classical string theory, $g_s \to 0, \alpha'/L^2 \neq 0$
Weak form	$N \to \infty \ , \lambda \ \text{large}$	Classical supergravity, $g_s \to 0, \alpha'/L^2 \to 0$

The strong form of the duality is well motivated by the fact that in the limit $N \to \infty$, λ arbitrary we obtain a so-called planar limit on the CFT side, in which the Feynman diagrams take on the appearance of string scattering diagrams.

Showing the Weak Form of the Duality and Motivating the Strongest Form

While it is indeed easier to understand the stringy side in the weak form of the AdS/CFT duality, we cannot apply the direct method of computing correlation functions using the supergravity partition function and then comparing to the SYM expectation. Due to this, the following argument will be completely different from the very direct approach used in the simple cases of the previous sections.

The weak form of the AdS/CFT duality lets us tune the coupling $g_s N$ to any finite value. It was observed that string theory may simplify considerably in both of the limits $g_s N \ll 1$, $g_s N \gg 1$. Let us only give a rough sketch of the original argument, since in the future we will be specializing to theories that only have gravity in the bulk. We will allow for both open and closed strings and work in a flat background spacetime. We will find that the open string sector will describe SYM for $g_s N \ll 1$ and be hidden begin and event horizon in the latter. The closed string sector will describe flat 10d supergravity in the first limit. In the second there will be two decoupled closed superstring sectors, one describing $AdS_5 \otimes S^5$ supergravity in the near-horizon region of the stack of D-branes, and one describing flat 10d supergravity far from the black hole. Since string theory is expected to be well defined for all values of the coupling, and the difference between the two limits is the exchange of $\mathcal{N} = 4$ SYM and supergravity on $AdS_5 \times S^5$ it is concluded they are equivalent. This analysis is carried out in significantly more detail in section 5.2 of [24].

We will first state some facts about D-branes that we did not discuss in chapter 5.

- The mass of a single D-brane in string theory is proportional to g_s .
- The mass of N coincident D-branes is proportional to the tunable coupling $g_s N$.
- D3 branes are charged under the 4-form field A_4 , which is present in the IIB superstring theory, meaning they exist and are stable.
- Open strings ending on a stack of N D3-branes describe massless U(N) Yang-Mills theory living on the 3 + 1 dimensional worldvolume of the D3 branes in the low energy limit.
- The gravitational constant in superstring theory goes as $G^{(10)}\sim g^2(\alpha')^4$

Remember that the low energy limit is obtained when $\alpha' \ll L^2$ where L is the AdS/S radii, as we saw when performing the α' -expansion in 5.1.3. In this limit we are allowed to do supergravity, although the limits $g_s N \ll 1$, $g_s N \gg 1$ will corresponds to very different configurations in the supergravity picture. An interesting note is that the limit $N \to \infty$ corresponds to a limit of big central charge c_{eff} on the CFT side, since the central charge measures degrees of freedom as we showed for 2d with the c-theorem in section 4.3.2.

We can begin by considering a configuration of N coincident D3-branes at zero coupling $g_s = 0$. We have free open superstrings ending on the coincident D-branes, describing a supersymmetric U(N) gauge theory with only massless fields. We also have free closed strings, but there are no nontrivial background field configurations since $g_s = 0$. The closed and open strings do not interact, so we have two completely decoupled theories. Let us note that in U(N) Yang-Mills theory, the U(1) part decouples and the dynamics are described by SU(N) Yang-Mills theory.

Let us now allow g_s to be nonzero, such that $g_s N \ll 1$. Since the D-branes are effectively massless for very small g_s , we may treat the D-branes as free meaning their worldvolume is flat. The coupling between open and closed strings is likewise proportional to g_s so they do not interact. Since $\alpha' \to 0$ there are only massless modes, and we are still dealing with free theories that are decoupled from each other living effectively on a flat background. Thus, for $g_s N \ll 1$ we have two decoupled theories: Type IIB closed strings in flat 10-dimensional space and SYM on the 3 + 1-dimensional worldvolume of the coincident D3 branes.

We now let $g_s N \gg 1$, the most notable effect of this is to make gravitational effects important. We are still in a low energy regime $(\alpha' \to 0)$, so we are still dealing only with massless modes. In addition to the gravitational effect, the D3 branes carry charge under the $A_{(4)}$ field that appears in the low energy effective action of IIB string theory. What we obtain is some very heavy, charged configuration i.e. something similar to a charged black hole.

Let us denote the coordinates that are parallel to the D3 brane worldvolume as x^{μ} , $\mu \in [0,3]$, and the transverse coordinates by x^i , $i \in [4,9]$. The solution of IIB supergravity that preserves the supersymmetry (supersymmetry relates geometrical and field degrees of freedom) in the presence of N D3-branes is given by the ansatz

$$ds^{2} = H(r)^{-1/2} \eta_{\mu\nu} dx^{\mu} dx^{\nu} + H(r)^{1/2} \delta_{ij} dx^{i} dx^{j} ,$$

$$e^{2\phi} = g_{s}^{2}$$

$$A_{(4)} = \left(1 - H(r)^{-1}\right) dx^{0} \wedge dx^{1} \wedge dx^{2} \wedge dx^{3} + \dots ,$$
(6.137)

where $r \equiv x_i x^i$ and the ... denote terms that ensure self-duality of $dA_{(4)}$. In the present situation, we are only interested in the geometry described by the metric ds^2 . The IIB equations of motion fix

$$H(r) = 1 + \left(\frac{L}{r}\right)^4$$
, (6.138)

where L is undetermined. From string theory, it is known that the flux $dA_{(4)}$, which carries one differential in the *i* directions must be quantized and proportional to the number of D-branes. It can be shown [24] that this restriction sets

$$L^4 = 4\pi g_s N \alpha'^2 . (6.139)$$

The background spacetime separates into two different regions for $r \gg L$ and $r \ll L$ respectively. For $r \ll L$ we see that $H \approx 1$ is a good approximation, describing a flat 10d spacetime. The case $r \ll L$ corresponds to the *near horizon metric*, where the metrics read

$$ds^{2} = \frac{r^{2}}{L^{2}} \eta_{\mu\nu} dx^{\mu} dx^{\nu} + \frac{L^{2}}{r^{2}} \delta_{ij} dx^{i} dx^{j}$$

$$ds^{2} = \frac{L^{2}}{z^{2}} \left(\eta_{\mu\nu} dx^{\mu} dx^{\nu} + dz^{2} \right) + L^{2} ds^{2}_{S^{5}}$$
(6.140)

where we have split the 6d space into the sphere S^5 and a radial coordinate $z = L^2/r$. The term in parentheses describes AdS₅. The D-branes and open strings have disappeared behind an event horizon. In the low energy limit, we have two separate background on which closed strings propagate that do not talk to each other. Therefore, in the limit $g_s \gg 1$ our spacetime is described by IIB supergravity on AdS₅×S⁵ as well as supergravity on flat 10d spacetime, and the two theories are decoupled from each other.

We have mulled over a significant number of details, but the picture is clear. We have two limits of a single theory, namely string theory. In the limit $g_s \ll 1$ we have something that describes

 $\mathcal{N} = 4$ SYM in 3+1 dimensions + IIB supergravity in flat 10d.

While still working with the exact same theory, having only tuned the parameter $g_s N \gg 1$ we obtained something that was described by

IIB supergravity on $AdS_5 \times S^5 + IIB$ supergravity in flat 10d.

The proposal is that since both limits are just the same theory with a different value of the coupling, we can equate them and subtract the flat supergravity from both sides, meaning we are left with

 $\left(\mathcal{N}=4 \text{ SYM in } 3+1 \text{ dimensions}\right) = \left(\text{IIB supergravity on } \operatorname{AdS}_5 \times \operatorname{S}^5\right)$.

Currently there exists a number of more explicit tests for the duality, and we will see some of these, relating to entanglement entropy calculations.

6.3 Why the Duality is Only Conjecture

In the previous sections we have seen a number of remarkable results. We have seen that scalar field theory in the bulk may be used to compute correct correlation functions for conformal primaries on the boundary. We have seen similarly that some vector fields, *p*-form gauge fields and even fermions [46] also have boundary CFT duals. In addition to this, we have even seen that the correspondence even works for gravity!

It almost seems like all of these results paint a picture of AdS/CFT duality as obviously true, so it is strange that we need to consider the roundabout string-theoretical argument of the previous section. If nice Standard Model like theories containing gravity, scalars and fermions already have all their quantum correlation functions computed by a boundary CFT, did we not already solve quantum gravity? The answer lies in that we have yet to write down the Lagrangian of a CFT that has only the listed conformal primary operators of the previous sections in their spectrum. Such details of d > 2 CFTs are in general poorly understood, although recent progress has been made via the conformal bootstrap¹⁰ equation [56].

In hindsight, failure of the duality is connected to the fact that quantum theories should be *unitary*. In addition to the unitarity bounds CFTs actually need to satisfy additional conditions to be unitary. We would probably have encountered this in section 6.2.1 had we tried to check the unitarity of the S-matrix or higher order correlation functions. Specifically, a unitary CFT in d > 2 dimensions needs to have an infinite number of primaries of increasing conformal weight, as obtained in section 10.3 of [56].

On the gravitational side of the duality, this corresponds to adding an infinite tower of increasingly massive fields. There are two well known types of theory that do this. String theory naturally has an infinite tower of massive string excitations. The second known possibility is gravity that contains extra compact dimensions. Gravity that has extra compact dimensions automatically contains an infinite number of massive particles, because standing wave components of the metric around the compact space (S^5) look like massive particles to the noncompact spacetime (AdS₅). These massive modes are called Kaluza-Klein towers, because the idea of studying gravity with

 $^{^{10}}$ The "bootstrap" refers to the fact that the bootstrap equation in principle lets you compute all *n*-point correlation functions of a CFT without ever writing down a Lagrangian, requiring only a set of conformal primary operators to be specified. The formula obtained however is an infinite sum that has no analytic expression (except in very special cases), so the sums have to be computed numerically.

extra compact dimensions was first introduced by Oskar Klein and Theodor Kaluza. In addition to this string theory itself contains an infinite number of massive excitations as we saw in chapter 5.

It is then (once again in hindsight) possible to understand why such a roundabout realization was necessary. Stringy supergravity in $AdS_5 \otimes S^5$ was the first explicit realization of a higherdimensional AdS/CFT duality between two full theories with defined Lagrangians. An interesting peculiarity of this is that to make the duality hold you need to keep all of the Kaluza-Klein modes wrapping the S⁵. This is because these are the fields that posses the SO(6) part of the symmetry shared by the two theories on each side of the duality. It seems to be a general feature that CFTs that can be well approximated by bulk supergravity must have a large number of degrees of freedom 'N large' and admit a 1/N expansion. This expansion looks inherently stringy, as is noted in section 1.7.5 of [24]. Thus while nothing we have seen obviously tells us that the AdS/CFT duality requires string theory, explicit realizations seem to conspire towards it.

In the remainder of this text, we will generally not specify even what spectrum of operators we consider in the boundary CFT. This is because we cannot solve CFT's in general dimension, even perturbatively (since CFT's are scale agnostic, "small" perturbations are not defined). Fortunately we *are* able to compute the entanglement entropy, which will be our probe into the general bulk physics dual to a weakly specified boundary CFT. Even better, as we discovered in chapter 4 the leading entanglement entropies are field content independent, so if entanglement is enough there really is no need for further specification. Next we will explore the holographic entanglement entropy, a modern development that may even imply that the emergence of bulk gravity may be a universal property of any boundary CFT [34].

Chapter 7

Holographic Entanglement

We saw in the previous chapter that a bulk/boundary correspondence that respects the basic axioms of CFT on the boundary appears for a fairly general class of bulk AdS fields. We also saw the big picture of how the first explicit realization in terms of physically interesting theories on both sides of the correspondence was motivated. AdS/CFT is a vast field, and we will now pick a slim specialization - the relationship between entanglement on the CFT side and geometry on the gravity side of the correspondence.

In this section we begin by looking at the relationship between AdS spacetimes containing black holes and boundary CFTs. Then, we motivate the Ryu-Takayanagi proposal and its relativistic cousin via considerations of black hole physics and the general leading area law divergence of the entanglement entropy found in section 4.3.1. Finally, we finish with a first look at how geometry may appear from entanglement information.

The logic roughly goes as follows. The bulk and boundary theories have the same partition functions implying they have the same entropy. On the gravity side entropies are generally related to the areas of event horizons. The Ryu-Takayanagi conjecture takes inspiration from this relation and conjectures that bulk areas are instead the holographic duals of boundary entanglement entropies. This can be proved by way of a bulk extension of the Replica trick, as can a covariantized version of the Ryu-Takayanagi relation called the Hubeny-Rangamani-Takayangi (HRT) relation.

7.1 AdS/CFT at Finite Temperature

In this section we consider $\mathcal{N} = 4$ SYM on the boundary and semiclassical¹ supergravity in the bulk $AdS_5 \times S^55$ spacetime. We will look at the bulk solutions that are dual to a thermal state on the boundary. This means that the SYM theory in question lives in three spatial dimensions times a circle, where the Euclidean time direction has been curled up into a circle as in section 3.2.2. We are required to take a strong coupling limit in which SYM is poorly understood, but we will find interesting results despite this obstacle.

This section should be seen as a primer to the relationship between entropy in the bulk and entropy on the boundary, as well as some of the subtleties of the relation. We will find that while it seems possible to relate entanglement in the boundary theory to the surface areas of black holes in the bulk, the black hole solution is some times not unique in addition to not being thermodynamically favored.

The key idea is that the entropy is for a field theory at temperature T is formally given in terms

¹Entropy in spacetimes is a quantum effect, as we learned in section 4.2.3. Fortunately the semiclassical entropy is field content independent, so we need only consider the geometry of the bulk theory.

of the partition function

$$S = \frac{\partial}{\partial T} \left(T \ln Z \right) \,\,, \tag{7.1}$$

and since both theories have the same partition function as stated by AdS/CFT, they must have the same entropy at the same temperature. We will now proceed in three steps

- We introduce spacetime thermodynamics as obtained by analytic to Euclidean time and demanding periodicity.
- We consider an AdS black brane solution, compute its entropy and compare to knowledge about the entropy of thermal SYM
- We consider thermal CFTs living on $S^{5}1 \times S^{5}d$ where $S^{5}1$ is the periodic Euclidean time. We find that such a configuration has two distinct bulk solutions.

7.1.1 Spacetime Thermodynamics

Chapter 3.2.2 taught us that in field theory, we can think of a thermal state as one that is periodic in imaginary time. It turns out that this can be extended to spacetimes, where we can create "thermal spacetimes" by analytically continuing to Euclidean time and demanding periodicity. Maximally symmetric spacetimes may be given an arbitrary temperature, but for example the Schwarzschild black hole admits only a single temperature. This is because unless we pick a particular periodicity the Euclidean metric gets a conical singularity at the event horizon which is unphysical given the smoothness of the original spacetime. The following argument is carried out using a black hole in flat space but it holds in de Sitter and AdS as well, since a conical singularity can never be sourced by a finite curvature term.

Let us look explicitly at the Schwarzschild black hole, with metric

$$ds^{2} = -\left(1 - \frac{r_{h}}{r}\right)dt^{2} + \left(1 - \frac{r_{h}}{r}\right)^{-1}dr^{2} + r^{2}d\Omega^{2}.$$
 (7.2)

We can go to near-horizon coordinates by letting $r = r_h(1 + \epsilon)$ and keeping only terms that are first order in ϵ . We then have the metric

$$\mathrm{d}s^2 = -\epsilon \mathrm{d}t^2 + \frac{r_h^2}{\epsilon} \mathrm{d}\epsilon^2 + r_h^2 \mathrm{d}\Omega^2 \;. \tag{7.3}$$

We can then make the change of coordinates $\epsilon = \rho^2/4r_h^2$, putting the metric on the form

$$ds^{2} = -\frac{\rho^{2}}{4r_{h}}dt^{2} + d\rho^{2} + d\Omega^{2} . \qquad (7.4)$$

This is known as the Rindler metric, describing an observer with constant acceleration $a = 1/2r_h$ in a flat background. Let us now analytically continue to $i\tau = t$, obtaining the Euclidean Rindler metric

$$ds^{2} = \frac{\rho^{2}}{4r_{h}}d\tau^{2} + d\rho^{2} + d\Omega^{2} .$$
(7.5)

The $\tau - \rho$ part of this metric looks like polar coordinates for the plane $ds^2 = d\rho^2 + \rho^2 d\theta$, with $\theta = \tau/2r_h$. This only describes a flat plane if θ has exactly the periodicity $\theta \sim \theta + 2\pi$, otherwise we are looking at a cone. Cones have curvature at the tip, but we know that Schwarzschild solves

the vacuum Einstein equations at the horizon meaning there should be no curvature. Therefore, for consistency we must have $\tau \sim \tau + 4\pi r_h$. The temperature of the black hole is therefore

$$T = \frac{1}{4\pi r_h} . \tag{7.6}$$

Also, matching $a = 1/2r_h$ we see that $T = \frac{a}{2\pi}$ matching the result of Bisogano and Wichmann for the Unruh effect that we saw in section 4.2.2.

7.1.2 AdS Black Brane/Thermal SYM Duality

We now have a basic understanding of spacetime thermodynamics, and so we know what kind of spacetimes could reasonably be dual to thermal SYM. Let us now consider the black-brane solution to supergravity on $AdS_5 \times S^55$, that we claim is dual to thermal SYM on the boundary. This subsection is the last time we will take external input from supergravity/SYM. The following is based on section 11.2 in Ammon & Erdmenger [24].

This black brane solution is similar to the one we imported in section 6.2.4, except that we do not require the solution to retain the full supersymmetry. Indeed, SYM on a thermal circle has supersymmetry explicitly broken because the periodic boundary conditions for fermionic and bosonic fields differ by a sign. This solution corresponds to so-called non-extremal black branes with metric given by

$$ds^{2} = H(r)^{-1/2} \left(-f(r)dt^{2} + d\vec{x} \right) + H(r)^{1/2} \left(\frac{dr^{2}}{f(r)} + r^{2}d\Omega_{5}^{2} \right)$$
(7.7)

where f(r) is the blackening factor

$$f(r) = 1 - \left(\frac{r_h}{r}\right)^4 \tag{7.8}$$

and

$$H(r) = 1 + \frac{L^4}{r^4} . (7.9)$$

If we do not wish to involve string theory, we may see this as just a planar black hole in 4+1 dimensions. In the following analysis, we then drop the S⁵5 part of the metric as well as the vol(S⁵5) contribution to the entropy.

Here there is an event horizon at $r = r_h$. We can perform the same analysis as for the Schwarzschild black hole in the previous section to find the temperature. We take a near-horizon limit by letting $r/L \ll 1$ and make a change of coordinates to $z = L^2/r_h$ to isolate the S⁵5 part of the metric. We then define $z_h = L^2/r_h$ and find that the metric of equation (7.7) takes the form

$$ds^{2} = \frac{L^{2}}{z^{2}} \left(-\left(1 - \frac{z^{4}}{z_{h}^{4}}\right) dt^{2} + \left(1 - \frac{z^{4}}{z_{h}^{4}}\right)^{-1} dz^{2} + d\vec{x}^{2} \right) + L^{2} d\Omega_{5}^{2} .$$
(7.10)

This is vaguely similar to an AdS₅ black hole tensored with the sphere S⁵5, with the main difference being that the horizon is not compact. To analyze the temperature of this spacetime the spherical part is unimportant, so let us ignore it for now. Going to Euclidean time $t = i\tau$, the remaining part is

$$ds^{2} = \frac{L^{2}}{z^{2}} \left(\left(1 - \frac{z^{4}}{z_{h}^{4}} \right) d\tau^{2} + \left(1 - \frac{z^{4}}{z_{h}^{4}} \right)^{-1} dz^{2} + d\vec{x}^{2} \right) .$$
(7.11)

We can now pick near-horizon coordinates, anticipating that we will find something similar to polar coordinates by expanding the metric to lowest order in ρ defined by

$$z = z_h \left(1 - \frac{\rho^2}{L^2} \right) . \tag{7.12}$$

The Euclidean metric becomes

$$ds^{2} = \frac{4\rho^{2}}{z_{h}^{2}}d\tau^{2} + d\rho^{2} + \frac{L^{2}}{z_{h}^{2}}d\vec{x} . \qquad (7.13)$$

Once again by identifying $\theta = 2\tau/z_h$, we see that the $\rho - \tau$ coordinates resemble the plane in polar coordinates $ds^2 = d\rho^2 + \rho^2 d\theta^2$. Requiring the absence of a conical singularity we see that we need $\tau \sim \tau + \pi z_h$. Therefore the temperature of the non-extremal black-brane solution is

$$T = \frac{1}{\pi z_h} . \tag{7.14}$$

We find that the AdS black brane should be dual to thermal SYM at temperature $T = 1/\pi z_h$.

We can compute the entropy of this theory. In the boundary field theory, the physical interpretation is that of a regular entropy since the boundary theory is in a thermal state. In the bulk theory, the interpretation is that we are computing the entanglement entropy between the interior of the black brane and the exterior. We will use the semiclassical value for the black hole entropy, given by

$$S_{\rm BH} = \frac{A}{4G} \tag{7.15}$$

where G is the five-dimensional Newton constant, which can be expressed² as $G = \pi L^3/2N^2$. The area of the horizon is given by

$$A = d^3 \vec{x} \sqrt{g_{3d}|_{z=z_h}} \operatorname{vol}(S^5 5) , \qquad (7.16)$$

where we are integrating over the noncompact \vec{x} coordinates in the Euclidean metric equation (7.13). The horizon is tensored with S⁵5, so the area in AdS₅ must be multiplied by vol(S⁵5). The \vec{x} part of the metric is diagonal, so the determinant is simply L^6/z_h^6 . Then we have that

$$A = \pi^{6} L^{8} T^{3} \operatorname{vol}(\mathbb{R}^{3}) , \qquad (7.17)$$

i.e. the horizon area is proportional to the volume of three dimensional flat space. Since the entropy on the field theory side is an honest entropy of a thermal state this is good news as such entropy is supposed to be extensive. Inserting the value for G we see that

$$S_{\rm BH} = \frac{\pi^2}{2} N^2 T^3 \text{vol}(\mathbb{R}^3) .$$
 (7.18)

By AdS/CFT duality this is equal to the entropy of the SYM plasma at temperature T, so we should identify $S_{\text{SYM}} = S_{\text{BH}}$. The field theory side of this calculation is not understood, while the gravity side calculation is quite simple. This is a case of using the AdS/CFT duality to perform an impossible computation on the field theory side via a simple computation on the gravity side.

²This comes from dimensional reduction from 10d, i.e. $G = G^{(10)}/\text{vol}(S^55)$ where $G^{(10)} \sim g^2(\alpha')^4$ and the AdS/CFT duality identifications have been used to solve for L and N.

7.1.3 Gravity Dual of Thermal CFT on a Sphere

Let us now consider a CFT living on $\mathbb{R} \times S^5 d$. If the CFT is in a thermal state, we compactify the time direction, and we find something that lives on $S^5 1 \times S^5 d$. There are now two dimensionful quantities, $\beta = 1/T$ and $\beta' = 1/l$ where l is the radius of the boundary sphere. There are two metrics that may have $S^5 1 \times S^5 d$ as their boundary, namely

• d+2-dimensional thermal AdS with metric

$$ds^{2} = \left(1 + \frac{r^{2}}{L^{2}}\right) d\tau^{2} + \left(1 + \frac{r^{2}}{L^{2}}\right)^{-1} dr^{2} + r^{2} d\Omega_{d}^{2} , \qquad (7.19)$$

where we impose periodicity in the Euclidean time direction $\tau \sim \tau + \beta$.

• The d+2 dimensional AdS-Schwarzschild solution with Euclidean metric

$$ds^{2} = f(r)d\tau^{2} + f^{-1}(r)dr^{2} + r^{2}d\Omega_{d}^{2} , \qquad (7.20)$$

where

$$f(r) \equiv 1 - \frac{\mu}{r^{d-1}} + \frac{r^2}{L^2} , \qquad (7.21)$$

where μ is related to the black hole mass. The black hole event horizon is at radius r_h which is given by the larger root of f(r) = 0.

Thermal AdS may have any temperature and still maintain regularity. For the black hole we may find the temperature by the usual regularity analysis. Near the horizon, expanding in powers of $r - r_h$ we find

$$ds^{2} = f'(r_{h})(r - r_{h})d\tau^{2} + \frac{1}{f'(r_{h})(r - r_{h})}dr^{2} + r_{h}^{2}d\Omega_{d}^{2} , \qquad (7.22)$$

we define $\rho^2 = f'(r_h) \frac{\rho^2}{4}$ to get

$$ds^{2} = (f'(r_{h}))^{2} \frac{\rho^{2}}{4} d\tau^{2} + d\rho^{2}$$
(7.23)

for the $\tau - \rho$ part. This looks like polar coordinates with $\tau = 2\theta/|f'(r_h)|$ and regularity demands $\tau \sim \tau + 4\pi/|f'(r_h)|$. Therefore, the black hole has temperature

$$T = \frac{|f'(r_h)|}{4\pi} = \frac{(d+1)r_h^2 + (d-1)L^2}{4\phi L^2 r_h} .$$
(7.24)

The space of possible black holes is parametrized by $r_h > 0$. Black holes in AdS have the peculiar property that at small radii, they have the same type of negative specific heat as black holes in flat space, but at radii larger than $r_h = L$ they have positive specific heat. This means that AdS black holes in d+2-dimensions have a minimum temperature, which can be deduced from equation (7.24) to be given by

$$T_{\min} = \frac{1}{2\pi L} \sqrt{(d+1)(d-1)}$$
(7.25)

at $r_h = L$. For temperatures below T_{\min} there is no black hole dual of the boundary spacetime. For $T \geq T_{\min}$ there are two possible solutions, and we must discriminate between them by computing the free energy of the system. The system with less free energy is then thermodynamically favored³. The free energy is given by

$$F^{(i)} = -T \ln Z = T S^{(i)}_{\text{on-shell}} ,$$
 (7.26)

³In principle, the semiclassical approximation tells us to add up all classical solutions in the partition function, but the one with larger free energy will dominate the expression $\exp\left[S_{\text{on-shell}}^{(1)} + S_{\text{on-shell}}^{(2)}\right]$

where i = 1, 2 labels thermal AdS and the black hole respectively. Here we have used that on the gravity side we are in a classical approximation. It is possible to show by inserting the solutions into the Einstein-Hilbert action with the Gibbs-Hawking boundary term that

$$\Delta F = F^{(2)} - F^{(1)} = \frac{r_h^{d-1}}{2\kappa_{d+2}^2} \operatorname{vol}(S^5 d) \left(1 - \frac{r_h^2}{L^2}\right) .$$
(7.27)

We see that for $r_h < L$, thermal AdS is thermodynamically preferred, meaning that small black holes in AdS are expected to decay into thermal AdS. For $r_h > L$ we have the opposite situation, meaning large AdS black holes are stable, because they are in equilibrium with their radiation bouncing off the AdS boundary. Note that black holes take over as thermodynamically favored just as they gain positive specific heat. This means that the gravity side dual of the boundary always has positive specific heat, which an important consistency check given that a thermal CFT always has positive specific heat.

At $r_h = L$ there is a phase transition called the *Hawking-Page transition*, at which point the temperature is

$$T = \frac{d}{2\pi L} . \tag{7.28}$$

This predicts a phase transition in the dual field theory, from having very few degrees of freedom at low energy to having a very large number of degrees of freedom above some transition energy. This essentially means that for a boundary CFT to have a gravity dual, it must be "gapped" meaning that most of the field content has finite mass starting at the transition energy.

As a final note, thermal AdS has zero entropy in the semiclassical approximation. Despite this thermal AdS is thermodynamically favored over small AdS black holes due to their bizarre negative specific heat. This means that in these examples, the leading gravity side entropy can still be interpreted in terms of entanglement across a horizon.

7.2 The Ryu-Takayanagi (RT) Formula

In the previous section we saw that the entropies of bulk and boundary theories are related. In the cases where the boundary CFT was dual to a black hole, the gravitational side of the entropy calculation admits an interpretation as entropy due to entanglement across the black hole horizon. In fact, the area of the entangling surface and the entropy are directly proportional.

In section 4.3.1 we saw that the leading term for the entanglement entropy of any finite region in spacetime is proportional to the area, lending more credence to the claim that the black hole entropy is directly related to entanglement.

The Ryu-Takayangi proposal in a sense reverses the bulk relation by conjecturing that areas in the bulk are actually the duals of entanglement entropy on the CFT side. More succinctly, area and entanglement entropy are not proportional, they are *equivalent* [4].

Consider a theory in AdS_{d+1} that is the holographic dual of *d*-dimensional CFT_d . Restrict now to a finite d-1-dimensional region A in a constant time slice with boundary ∂A . Let B be the complement of A, meaning B covers the rest of the constant time slice and shares boundary with A, i.e. $\partial A = \partial B$. It is well known from chapters 2.3 and 4 that the entanglement entropy S has a few basic properties, namely

- complementarity (for bipartitions of pure states), S(A) = S(B),
- subadditivity, $S(A) + S(B) (S(A \cup B) \ge 0,$

• strong subadditivity, $S(A \cup B) + S(B \cup C) \ge S(A \cup B \cup C) + S(B)$ for any regions A, B and C.

Intuition from black hole physics tells us that the bulk dual of the entanglement entropy in the boundary theory should be some kind of d-1-dimensional 'area' in the bulk. This area should be completely specified by a well chosen boundary quantity that is the same for the regions A and B, since this ensures that the dual of entanglement entropy is the same for the regions A and B. There is exactly one boundary quantity that can determine d-1 dimensional surface in the bulk spacetime that is common to A and B, namely the boundary $\partial A = \partial B$. Since the bulk area is supposed to be determined by this boundary, we better demand that the bulk area is the area of a surface that shares boundary with A and B. In addition to this, there should be a unique way of choosing this area.

We know that the von Neumann entropy in quantum information theory is equivalent to the optimal choice of measurements that minimizes the entropy S_A in the sense that if we used a particular set of projective measurements to turn the quantum state into a classical probability distribution, the choice of projective measurement basis that minimizes the classical entropy gives exactly the von Neumann entropy. Therefore, the natural proposal is that the bulk surface that we will henceforth call γ_A should be the surface in the bulk spacetime that has the minimum possible area under the constraint that $\partial \gamma_A = \partial A$. This is the information we need to really understand the actual Ryu-Takayanagi proposal.

Ryu and Takayanagi proposed that the entanglement entropy S_A of the region A is dual to the area of a particular d-1-dimensional surface γ_A in the bulk spacetime. From black hole physics, we know that the entropy per unit area is given by $1/4G_N$ and as such we postulate the specific relation

$$S_A = \frac{\text{Area of } \gamma_A}{4G_N^{d+1}} . \tag{7.29}$$

Here γ_A is the d-1-dimensional static, minimal surface in AdS_{d+1} that has the same boundary as A, i.e. $\partial \gamma_A = \partial A$. By static we mean that like the region A, γ_A is a surface that extends only in the spatial directions. In general γ_A has a divergent area, related to the fact that only lightlike lines see the distance to the AdS boundary as finite. This is in line with the expectation of UVdivergences of the entropy on the field theory side. To uniquely specify the surface in a spacetime with topological defects (such as a black hole) we must require that γ_A is homologous to A meaning it can be continuously deformed into A. Notably this means that if there is a black hole at the center of the dual spacetime, $\gamma_A \neq \gamma_B$, corresponding to the fact that the boundary CFT is not in a pure state.

Strong (and regular) subadditivity are not manifest in this proposal, but it can be proven by assuming the null energy condition as we will see in section 7.2.4. This has a particularly interesting implication on the gravitational side of the AdS/CFT duality. In general relativity, the so-called null energy conditions postulate a positive local energy density, ruling out for example traversable wormholes. While these are "physical postulates" on the gravity side, they become necessary properties of geometry when the CFT side of the duality is taken as fundamental.

We proceed by an example AdS_3 gravitational computation of the entanglement entropy, showing a match with boundary CFT expectation in 1+1 dimensions. We then outline a more general argument for the correctness of the proposal originally due to Lewkowycz and Maldacena. In addition to this we derive the bulk extremal surface corresponding to boundary spherical regions and give a particularly simple holographic proof of strong subadditivity. For more examples where the Ryu-Takayanagi prescription matches known CFT expectations, the original reference is recommended [4].

7.2.1 AdS_3/CFT_2 , Entanglement Entropy of Single Interval

We consider a 1+1d CFT on the boundary of AdS_3 with curvature radius R. It should be noted that while the explicits of AdS_3 gravity are poorly understood, we expect on very general grounds⁴ that it should be dual to a boundary CFT with central charge

$$c = \frac{3R}{2G_N^{(3)}} , \qquad (7.30)$$

where G_N is the 3d Newton's constant.

In section 4.3.3 we showed that the entanglement entropy of a finite strip A of length l in a 1+1d CFT on a circle with circumference L is given by

$$S_A = \frac{c}{3} \ln \left(\frac{L}{\pi \epsilon} \sin \frac{\pi l}{L} \right) \,. \tag{7.31}$$

We want to rederive this result by computing the dual bulk property, which in this case is $L_{\gamma}/4G_N^{(3)}$ where L_{γ} is the length of a bulk geodesic ending on the endpoints of A. Note that since we are to compute the length of a geodesic that connects two points on the boundary, we are going to have to introduce a cutoff radius. This necessarily puts the boundary CFT on a circle with some circumference L, leading to the relevance of the above CFT result. We will work with the global AdS coordinates given by

$$ds^{2} = R^{2} \Big(-\cosh^{2}(\rho) dt^{2} + d\rho^{2} + \sinh^{2}(\rho) d\theta^{2} \Big) .$$
(7.32)

For the calculation to make sense we need to introduce a cutoff at a finite radius $\rho = \rho_0$ and consider the bulk to be the bounded region $\rho < \rho_0$. A cutoff at large radius ρ_0 is equivalent to a short-distance cutoff ϵ on the CFT side, with the relation [58][47]

$$e^{\rho_0} \sim \frac{L}{\epsilon}$$
 (7.33)

Note that the exact $\mathcal{O}(1)$ coefficient is ambiguous, because it depends on the specific field content of the boundary CFT which is in general not known. The region [0,l] on the cylinder may be identified with $\theta = [0, 2\pi \frac{l}{L}]$. We are then looking for the length of the geodesic γ_A that connects the points $(\rho, \theta) = (\rho_0, 0)$ and $(\rho, \theta) = (\rho_0, \pi \frac{l}{L})$ with t kept constant. The relevant geometry is illustrated in figure 7.1, where the spatial boundary of AdS is represented as a solid circle.

In AdS_{d+1} , the simplest way to compute the form of a geodesic is to go to the embedding space M_{d+2} . The geodesics are then defined by the intersection of the embedded AdS_{d+1} space and a two-dimensional hyperplane such that the normal of the embedded AdS space as well as the endpoints of the geodesic are contained in the hyperplane [4, 59]. In the flat embedding space M_4 the geodesic turns out to be (parametrized by $R\lambda \in [0, \ln(\sqrt{\frac{\alpha+1}{\alpha-1}})]$)

$$\vec{X} = \frac{R}{\sqrt{\alpha^2 - 1}} \sinh\left(\frac{\lambda}{R}\right) \cdot \vec{x} + R \left[\cosh\left(\frac{\lambda}{R}\lambda/R\right) - \frac{\alpha}{\sqrt{\alpha^2 - 1}} \sinh\left(\frac{\lambda}{R}\right)\right] , \quad (7.34)$$

 $^{^{4}}$ Brown and Henneaux originally derived this result based on general requirements on the global charges of the AdS theory in 1986 [57]. Note that they give the resulting Virasoro algebra in terms of Poisson brackets, so reading off the correct form is nontrivial. An alternative derivation of the result using holographic renormalization is given by Henningson and Skenderis in [49] and section 6.2.3



Figure 7.1: Schematic illustration of the curve γ_A in relation to A. In a) we look only at a spatial slice, introducing the coordinates ρ and θ as they appear in equation (7.32). In b) we show the embedding of the time slice into the full spacetime. The curve γ_A can be interpreted as the bulk event horizon enforcing that an observer only has access to subsystem A. The shaded region is then the region of spacetime accessible to the spacetime observer dual to a CFT observer with only access to A.

where $\alpha = 1 + 2 \sinh^2 \rho_0 \sin^2(\pi l/L)$ and

$$\vec{x} = (\cosh \rho_0 \cos t, \cosh \rho_0 \sin t, \sinh \rho_0, 0)$$

$$\vec{y} = (\cosh \rho_0 \cos t, \cosh \rho_0 \sin t, \sinh \rho_0 \cos\left(\frac{2\pi l}{L}\right), \sinh \rho_0 \sin\left(\frac{2\pi l}{L}\right)) , \qquad (7.35)$$

where t is the fixed time of the time slice on which A lives. Note that \vec{x} and \vec{y} point in the two directions of the endpoints. As $R\lambda$ goes from 0 to $\ln\left(\sqrt{\frac{\alpha+1}{\alpha-1}}\right)$ we see that \vec{X} goes from pointing entirely in the \vec{x} direction to pointing entirely in the \vec{y} direction. We can then integrate to find L_{γ}

$$L_{\gamma} = \int d\lambda \sqrt{\frac{\partial X^{\mu}}{\partial \lambda} \frac{\partial X^{\nu}}{\partial \lambda} \eta_{\mu\nu}}$$
(7.36)

where $\eta = \text{Diag}(-1, -1, 1, 1)$ is the flat metric of the embedding space. It is possible to show that the length of the geodesic is given by

$$\cosh\left(\frac{L_{\gamma}}{R}\right) = 1 + 2\sinh^2(\rho_0)\sin^2\left(\frac{\pi l}{L}\right) . \tag{7.37}$$

For large ρ_0 we can replace the hyperbolic sines by exponentials on both sides, since L_{γ} diverges with ρ_0 . Then,

$$L_{\gamma} = R \ln\left(e^{2\rho_0} \sin^2 \frac{\pi l}{L}\right) = 2R \ln\left(e^{\rho_0} \sin \frac{\pi l}{L}\right) \,. \tag{7.38}$$

Inserting into the Ryu-Takayanagi formula in equation (7.29) we find

$$S_A = \frac{R}{2G_N^{(3)}} \ln\left(e^{\rho_0} \sin\frac{\pi l}{L}\right) \,, \tag{7.39}$$

and now using the duality relations $c=3R/2G_N^{(3)}$ and $e^{\rho_0}\sim \frac{L}{\epsilon}$ we obtain

$$S_A \sim \frac{c}{3} \ln\left(\frac{L}{\epsilon} \sin\frac{\pi l}{L}\right)$$
 (7.40)

This precisely coincides with the known CFT result on the circle in equation (7.31). Remember that for the CFT result on the circle we threw away a constant term that depended on the explicit field content, and as such there is no reason to worry about the approximate equality in this expression. Thus we see that our natural relationship between entanglement entropy on the boundary and minimal surfaces in the bulk produces the correct result for a 2d CFT.

7.2.2 Equivalence Between Replica Construction and Minimal Area

Another interesting argument for the validity of the Ryu-Takayanagi proposal (equation (7.29)) was proposed seven years later by Lewkowycz and Maldacena in 2013 [60]. Their result required only that the spacetime be time-reflection symmetric about the spatial slice of the previous section, weakening the requirement of a static state. In addition to this proof, the Ryu-Takayangi proposal was proven in greater generality by Lewkowycz and Parrikar in 2018 using the Iyer Wald formalism we will introduce in chapter 8 [61].

In the 2013 proof [60] looked at the boundary geometry corresponding to the *n*-sheeted replica space of section 4.3.3. According to AdS/CFT the path integral over this space should be equal to the partition function for bulk gravity with the *n*-sheeted boundary geometry as its boundary. Taking the classical limit on the gravity side, the gravitational partition function just becomes $e^{-S_{\text{grav}}}$ where S_{grav} is integrated over a classical solution satisfying the boundary conditions. This means that in principle we have a way of computing the Renyi entropies gravitationally without ever making the Ryu-Takayanagi conjecture. They argue that the classical gravitational problem reduces to the computation of a minimal area as the limit $\lim_{n\to 1} \partial_n S_A^{(n)}$ is taken. Let us coarsely sketch how this comes about.

The *n*:th Renyi entropy on the boundary is a path integral over all of Euclidean time over n replicas of the boundary surface. This can be conformally mapped to a single manifold that has simple topology such as the orbifold plane in section 4.3.3.

For a general multi-sheeted surface we can define an angular coordinate θ that takes us to the next plane as we pass through an angle 2π , as in figure 7.2. In the case of n = 1 this imposes the periodicity of all fields in θ :

$$\psi_B(\theta) = \psi_B(\theta + 2\pi) , \qquad (7.41)$$

since in the case of a single sheet the trace over ρ_A tells us that the surface A is the same no matter which direction it is approached from.

As we consider geometries corresponding to the n > 1:th Renyi entropies, the gluing tells us that $\theta \to \theta + 2\pi$ should take us to the next sheet. After n sheets we should return to the first, so the periodicity of the coordinate becomes $\theta \sim \theta + 2\pi n$. The replica construction also carries a discrete symmetry, the replica symmetry Z_n , so the periodicity of the ψ_B remains the same. In principle this replica symmetry may be broken in a particular solution. There is a more in depth discussion of this in the reference [60], but the essence is that if the bulk AdS spacetime is allowed to contain bad tachyons that are not given positive kinetic energy by their boundary terms, the resulting instability breaks replica symmetry. Thus replica symmetry is roughly broken by the same things that break the bulk-boundary uniqueness relations in AdS/CFT. We continue, assuming that replica symmetry holds.

The next observation was that the classical solutions in the bulk for n > 1 contain a d - 1dimensional surface that is fixed under the action of the replica symmetry. Let us call this surface



Figure 7.2: Illustration of meaning of angular coordinate θ . *a*) is the case when the boundary is 1 + 1d, and *b*) illustrates the case of a 2 + 1d boundary. These examples are dual to AdS₃ and AdS₄ respectively. Approaching the region *A* from the future takes us to the next replica and approaching *A* from the past takes us to the previous replica. This means that $\theta \to \theta + 2\pi$ realizes the replica symmetry and that the coordinate θ has period $2\pi n$ for the n : th Renyi entropy. The boundary ∂A is fixed under the replica symmetry, and the fixed bulk surface must be either disjoint or anchored to ∂A .

 γ_A . The surface γ_A corresponds to points in the bulk where the circle parametrized by θ shrinks to zero size⁵. Note that the boundary ∂A as it appears in a single copy of the spacetime is naturally preserved under the replica symmetry. Due to this, the fixed surface is expected to fulfill $\partial \gamma_A = \partial A$ as $n \to 1$.

We can write the metric near γ_A as

$$ds^{2} = n^{2}dr^{2} + r^{2}d\theta^{2} + \dots$$
(7.42)

where r and θ are the coordinates that are transverse to γ_A and ... denotes its internal coordinates. Note that the factor n^2 is to prevent a conical singularity appearing in the bulk spacetime, the real angle around the fixed surface in a single replica is $\alpha = \theta/n$.

The next step is analytic continuation. To do the continuation, we let n be noninteger, and consider the case $n = 1 + \epsilon$ with ϵ small. The trick used by the original authors was to let $\theta \sim 2\pi$ for the computation of the action. The action is then evaluated by integrating θ from 0 to 2π and multiplying by $1 + \epsilon$, the idea being that you are integrating the same function over a whole period and it is only the size of the measure $d\theta$ that changes. This should introduce a singularity near r = 0 since we are forcing the metric (and any other potential bulk fields) to have period 2π in a spacetime that is actually $(2+2\epsilon)\pi$ -periodic. On general grounds, this singularity is integrable and vanishes smoothly as $\epsilon \to 0$. This means that the analytically continued solution can be expanded in terms of ϵ near n = 1 and the limit

$$S_A = -\frac{\partial}{\partial \epsilon} \text{Tr} \rho_A^{1+\epsilon} \bigg|_{\epsilon=0}$$
(7.43)

can be evaluated. Here we have rewritten equation (4.206) letting $n = 1 + \epsilon$ and used that the local density operator is $\text{Tr}[\rho_A] = e^{S_{\text{grav}}}$, where S_{grav} the on-shell action.

⁵Note that if the bulk spacetime has nontrivial topology, such as thermal AdS, the circle may not be contractible in the bulk. In this case, the lack of fixed points under Z_n leads to zero entropy.

A generic Einstein-Hilbert action in the bulk will give rise to some set of equations of motion. Demanding that the equations of motion be satisfied to leading order in ϵ will describe an extremal surface satisfying $\partial \gamma_A = \partial A$ and the action integral will compute the proper area of said surface. To first order in ϵ , $\rho_A^{1+\epsilon}$ will be proportional to S_{grav} , meaning S_A will be proportional to the proper area of a minimal surface. This shows that the gravitational dual of the Renyi entropy (under suitable assumptions) reduces to the computation of the area of a minimal surface. This is exactly the Ryu-Takayanagi proposal.

Let us see how this comes about. We first write down the metric, picking so-called Gaussian coordinates about the fixed surface of the replica symmetry. Heuristically demanding smoothness as we go around the θ circle the metric takes the near γ_A form

$$ds^{2} = n^{2}dr^{2} + r^{2}d\theta^{2} + b_{i}d\theta dy^{i} + g_{ij}dy^{i}dy^{j}$$

$$g_{ij} = h_{ij} + r^{n}\cos(\theta)K_{ij}^{1} + r^{n}\sin(\theta)K_{ij}^{2} + \mathcal{O}(r^{2})$$

$$R = r + \mathcal{O}(r^{3}) , \quad b_{i} = \mathcal{O}(r^{2}) ,$$
(7.44)

where r is the normal to γ_A , the y_i are parallel to γ_A and K_{ij}^{α} , $\alpha \in \{1,2\}$ are the two extrinsic curvature tensors. The intrinsic metric h_{ij} depends only on the y_i while the extrinsic curvatures capture the θ and r-dependence of the metric on the fixed surface. We take the linear combination of the extrinsic curvatures which is timelike to be zero. This corresponds to a static or time-reflection symmetric spacetime at the constant time slice of A. Then the equations of motion that follow will simply serve to set the spatial part to zero as well, defining an extremal surface.

Next we should compute the Einstein field equations for this metric. The Ricci curvature receives divergent contributions proportional to

$$R_{\mu\nu} \sim \epsilon \frac{K_{ij}^{\alpha} g^{ij}}{r} . \tag{7.45}$$

The requirement of smoothness at r = 0 to first order in ϵ leads to the vanishing of the traces of the extrinsic curvatures

$$g^{ij}K^1_{ij} = g^{ij}K^2_{ij} = 0 av{7.46}$$

which is the equation of motion of a extremal surface. The next step is to compute the onshell action. The on-shell action does not a priori look like a computation of the proper area of the minimal surface whose equations of motion are equation (7.46). It turns out that by using diffeomorphism invariance to pick a particular gauge it reduces to an area computation [42].

7.2.3 Extremal Bulk Surface Ending on Spheres

Let us now consider a region on the boundary that is a d-1-dimensional ball of radius R. We will find that the extremal surface with the same boundary as the ball is a d-1-dimensional halfsphere with radius R extending into the bulk. In this section we settle for just finding the extremal surface, although the entanglement entropy can technically be computed on both the CFT and gravity sides of the duality. In chapter 8 we will use this result to prove on general grounds that spacetimes whose Ryu-Takayanagi areas correctly reproduce the entanglement entropies of CFT balls necessarily satisfy the Einstein field equations to first and second order in perturbations about pure AdS. This is another argument for the validity of the Ryu-Takayanagi formula for any CFT dual to a theory of gravity.

The result we are about to derive comes from notes by Veronika E. Hubeny [62]. Hubeny leaves out only elementary (if quite gruesome) steps, so her text is a good reference for some extremal surface practice. Let us begin by representing our AdS spacetime in Poincaré coordinates:

$$ds^{2} = \frac{1}{z^{2}} \left[-dt^{2} + dz^{2} + d\vec{x}^{2} \right]$$
(7.47)

where we have set the AdS radius to L = 1. Note that Poincaré coordinates do not cover all of AdS, so we need to assume that the boundary ball may be contained in a single Poincaré patch. The boundary region is a ball of radius R centered on $\vec{x} = 0$, so the the extremal surface problem is more easily solved picking more appropriate boundary-spherical coordinates such that the metric becomes

$$ds^{2} = \frac{1}{z^{2}} \left[-dt^{2} + dr^{2} + r^{2} d\Omega_{d-2}^{2} + dz^{2} \right] .$$
(7.48)

The extremal surface \tilde{A} minimizes the area functional $A[\tilde{A}]$, given by

$$A(G, X_{ext}) = \int_{\tilde{A}} \mathrm{d}^d \sigma \sqrt{-g} , \qquad (7.49)$$

where g is the induced metric on \tilde{A} and G is the bulk metric. The induced metric g is related to the bulk metric G by a pullback

$$g_{ab} = G_{\mu\nu} \frac{\partial X^{\mu}}{\partial \sigma^a} \frac{\partial X^{\nu}}{\partial \sigma^b} .$$
(7.50)

Let us parametrize the bulk surface by picking σ^1 equal to r and σ^i with $i \in \{2...d\}$ to be the angles of the (d-2)-sphere. We then solve for z in terms of r by varying the area functional with respect to the bulk coordinates. By picking the coordinates in this way, the induced metric also becomes diagonal with the form

$$g_{ab} = \frac{1}{z^2} \begin{bmatrix} \left(\frac{\partial z}{\partial r}\right)^2 + \left(\frac{\partial r}{\partial r}\right)^2 & \mathbf{0} \\ \mathbf{0} & r^2 G_{\Omega} \end{bmatrix} , \qquad (7.51)$$

where G_{Ω} is the metric of the d-2 dimensional unit sphere, with determinant 1. It is now simple to evaluate the determinant according to

$$A = \int_0^R \mathrm{d}r \,\mathrm{d}^{n-2}\Omega \sqrt{-\det\left(\frac{\mathrm{d}X^{\mu}}{\mathrm{d}\sigma^a}\frac{\mathrm{d}X^{\nu}}{\mathrm{d}\sigma^b}G_{\mu\nu}\right)} = V_{d-2}\int_0^R \mathrm{d}r\frac{\sqrt{1+\dot{z}^2(r)}}{z(r)^{d-1}}r^{d-2} , \qquad (7.52)$$

where $V_{d-2} = \frac{2\pi^{(d-1)/2}}{\Gamma((d-1)/2)}$ is the volume of the unit n-2 sphere, obtained by integrating the angleindependent integrand over the spherical measure. At this point, we find the Euler Lagrange equations by varying with respect to z, dropping the explicit r dependence:

$$\begin{split} \delta_{z}A &= V_{d-2} \int_{0}^{R} \mathrm{d}r \left(\frac{r^{d-2}\dot{z}}{z^{d-1}\sqrt{1+\dot{z}^{2}}} \delta\dot{z} - (d-1) \frac{r^{d-2}\sqrt{1+\dot{z}^{2}}}{z^{d}} \delta z \right) \\ &= -V_{d-2} \int \mathrm{d}r \left((d-2) \frac{r^{d-3}\dot{z}}{z^{d-1}\sqrt{1+\dot{z}^{2}}} + \frac{r^{d-2}\ddot{z}}{z^{d-1}\sqrt{1+\dot{z}^{2}}} - (d-1) \frac{r^{d-2}\dot{z}^{2}}{z^{d}\sqrt{1+\dot{z}^{2}}} \right) \\ &\quad - \frac{r^{d-2}\ddot{z}\dot{z}\dot{z}^{2}}{z^{d-1}(1+\dot{z}^{2})^{3/2}} + (d-1) \frac{r^{d-2}\sqrt{1+\dot{z}^{2}}}{z^{d}} \right) \delta z + V_{d-2} \left[\frac{r^{d-2}}{z^{d-1}\sqrt{1+\dot{z}^{2}}} \delta z \right]_{0}^{R} \\ &= V_{d-2} \int \mathrm{d}r \frac{r^{d-2}}{z^{d-1}\sqrt{1+\dot{z}^{2}}} \left((d-2) \frac{\dot{z}}{r} + \ddot{z} - (d-1) \frac{\dot{z}^{2}}{z} \\ &\quad - \frac{\ddot{z}\dot{z}^{2}}{1+\dot{z}^{2}} + (d-1) \frac{1+\dot{z}^{2}}{z} \right) + V_{d-2} \left[\frac{r^{d-2}}{z^{d-1}\sqrt{1+\dot{z}^{2}}} \delta z \right]_{0}^{R} \\ &= V_{d-2} \int \mathrm{d}r \frac{r^{d-2}}{z^{d-1}(1+\dot{z}^{2})^{3/2}} \left(\ddot{z} + (1+\dot{z}^{2}) \left[\frac{d-1}{z} + \frac{d-2}{r} \dot{z} \right] \right) + V_{d-2} \left[\frac{r^{d-2}}{z^{d-1}\sqrt{1+\dot{z}^{2}}} \delta z \right]_{0}^{R}$$

$$(7.53)$$

To find the equations of motion, we just need the object in parentheses to vanish as well as for the boundary term at the end to vanish for arbitrary δz . We know that at z = 0, the surface should be anchored on the circle, so z(0) = R. The equation of motion turns out to be solved by the simple function

$$z(r) = \sqrt{R^2 - r^2} \tag{7.54}$$

for general d. To verify, observe that

$$\dot{z} = -\frac{r}{\sqrt{R^2 - r^2}} ,$$

$$\ddot{z} = -\frac{1}{\sqrt{R^2 - r^2}} - \frac{r^2}{(R^2 - r^2)^{3/2}} = -\frac{R^2}{(R^2 - r^2)^{3/2}} ,$$

$$\sqrt{1 + \dot{z}^2} = \sqrt{1 + \frac{r^2}{R^2 - r^2}} = \frac{R}{\sqrt{R^2 - r^2}} .$$
 (7.55)

Inserting the last line into the boundary term we see that it vanishes at r = 0. The boundary term diverges at r = R. This is related to the fact that the area functional also diverges close to the boundary of AdS. Let us also verify that the proposed expression also solves the equations of motion:

$$\begin{aligned} \ddot{z} + (1 + \dot{z}^2) \left[\frac{d-1}{z} + \frac{d-2}{r} \dot{z} \right] &= -\frac{R^2}{(R^2 - r^2)^{3/2}} + \frac{R^2}{R^2 - r^2} \left(\frac{d-1}{\sqrt{R^2 - r^2}} - \frac{d-2}{r} \frac{r}{\sqrt{R^2 - r^2}} \right) \\ &= -\frac{R^2}{(R^2 - r^2)^{3/2}} + \frac{R^2}{(R^2 - r^2)^{3/2}} \\ &= 0 \;. \end{aligned}$$

$$(7.56)$$

Thus, the extremal surface ending on a d-2-dimensional sphere on the boundary is half a d-1-dimensional sphere, defined by

$$z^2 + \vec{x}^2 = R$$
, $z \ge 0$, (7.57)

in the original Poincaré coordinates. With this the area functional is possible to compute although it is divergent. We impose a lower cutoff $z = \epsilon$ and find

$$A = \frac{2\pi^{(d-1)}/2}{\Gamma\left(\frac{d-1}{2}\right)} \int_{\epsilon}^{R} \frac{R}{(R^2 - r^2)^{d/2}} r^{d-2} .$$
(7.58)

This correctly computes the entanglement entropy of ball shaped regions via the RT formula if we set

$$L = \left(\frac{\Gamma(\frac{d}{2})}{\pi^{\frac{d}{2}}} 8\pi G_N a^*\right)^{\frac{1}{d-1}} , \qquad (7.59)$$

where L is the AdS radius and a^* is the trace anomaly of the boundary CFT [63].

7.2.4 Holographic Proof of Strong Subadditivity

The gravitational side of strong subadditivity admits a simple proof. That this can be shown is important, since if extremal areas and entropies are dual, they must have the same properties. We can deduce in a simple manner from the following picture that the Ryu-Takayanagi surfaces for regions A, B and C satisfy strong subadditivity:



The horizontal line represents the boundary of AdS. The area of γ_B is minimal under the condition of ending at ∂B , and is therefore less than the area of the the part of $\gamma_{AB} \cup \gamma_{BC}$ in solid grey. γ_{ABC} is a minimal area under the condition of ending on $\partial(ABC)$ and therefore has less area than the part of $\gamma_{AB} \cup \gamma_{BC}$ in solid black. The area of the union is the sum of the areas. Since entropy is proportional to area, this implies strong subadditivity. In the case where the regions A, B and C are disjoint the definition of the minimal surfaces is generally determined from the field theory side, so strong subadditivity is manifest [4].

Equivalently, by splitting $\gamma_{AB} \cup \gamma_{BC}$ in a different way we obtain a different formulation of strong subadditivity



Allowing for a negative energy density may turn γ_A or γ_{ABC} into maximal instead of minimal surfaces, ruining the validity of the proof. Additionally, this proof relies on the spacetime being either time independent or A, B and C being defined on the same time slice. This is not always

the case, and we will consider the consequences of this by looking at the Hubeny-Rangamani-Takayanagi proposal. Proving strong subadditivity in the time dependent case is due to Aron Wall [64].

7.3 Hubeny-Rangamani-Takayanagi (HRT) Formula

The Ryu-Takayanagi proposal in equation (7.29) is not Lorentz covariant and it is only defined for surfaces in constant time slices. A covariant generalization carries with it some difficulties.

A first difficulty is the fact that a minimal area is no longer an intuitive candidate, introducing timelike deformations of the surface where there previously was none automatically reduces the area since the timelike direction carries negative signature. To resolve this difficulty, the first suggestion by Hubeny, Rangamani and Takayanagi [5] was that we should look for any *extremal* surfaces \mathcal{M}_A with boundary ∂A and then pick the one with minimal spacelike area. The covariant holographic entanglement entropy is then given by

$$S_A = \frac{\mathcal{M}_A}{4G_N^{d+1}} \tag{7.60}$$

where d is the dimension of the boundary spacetime.

7.3.1 Heuristic Gravitational Derivation of HRT Formula

Just like in the static case, some time after the original proposal it was found that it was possible to argue its validity by understanding the gravitational dual of time-dependent replica construction. The original derivation was carried out in 2016 by Dong, Lewkowycz and Rangamani [65].

The essential difference to the static case is that on the CFT side we cannot use the trick of integrating over n copies of a Euclidean spacetime. Once we have genuine time-dependence we have to glue together n Schwinger-Keldysh⁶ contours across A. In addition, now that we have dropped time independence A does not have to be defined in a constant time slice, although we take it to be spacelike. The Schwinger-Keldysh contours contain an extra backwards- and forwards segment in the Lorentzian time direction at some Euclidean time $\tau = T$. This evolution introduces a kink in the spacetime at the Euclidean time T along some Cauchy slice Σ_T such that $A \in \Sigma_T$ in the boundary spacetime.

It is customary to do away with the complex time integration and instead consider any state $|\Psi\rangle (t = \infty)$ as a boundary condition. Then, the path integral is performed over two copies of the causal past J_T^- of Σ_T , glued across $A \in \Sigma_T$. In principle we should add the an $i\epsilon$ prescription to avoid eventual singularities on the real time axis. This involves sending one copy of J_T^- to $t = -\infty(1 + i\epsilon)$ corresponding to preparing the vacuum ket, and the other copy to $t = -\infty(1 + i\epsilon)$ to prepare the vacuum bra. The relevant Schwinger-Keldysh countour, as well as a single copy of the *n*-fold contour can be found in figure 7.3

To extend this to a bulk spacetime, the trick lies in extending the Cauchy surface on the boundary into the bulk, yielding the slice $\tilde{\Sigma}_T$. The bulk extension of a boundary Cauchy slice into the bulk is not unique⁷, it only requires that all points on the bulk Cauchy surface $\tilde{\Sigma}_T$ are spacelike

 $^{^{6}}$ We discussed the Scwhinger-Keldysh contour in section 3.2.2. In this example we are not splitting the contour into a series of straight segments since we, like in section 7.2.2, only need to understand the behaviour of a special fixed surface under the replica symmetry.

⁷In the time independent case, there is a unique way of going from one bulk Cauchy surface to the next. In such a case the curve that extremizes \mathcal{M}_A is obtained by picking the $\tilde{\Sigma}_T$ that has maximal volume (which will be the constant time slice), and finding the $\mathcal{M}_A \in \tilde{\Sigma}_T$ that minimizes the proper area of \mathcal{M}_A . When A is a spacelike surface, this reduces exactly to the RT formula.



Figure 7.3: Schwinger-Keldysh contour for $\text{Tr}[\rho]$ (a)) and a single copy of the integration for the computation of ρ^n (b)). In red are the sides of A in the imaginary time direction that take you to the next or previous replicas depending on whether you pass from the positive or negative τ -direction respectively. Each sheet expanding backwards in Lorentzian time is a copy of the causal past J_T^- of Σ_T . The *n*-fold replica construction involves a path integral over 2n copies of J_T^- as well as the gluing across A.

separated as well as $\partial \tilde{\Sigma}_T = \Sigma_T$. This means that $\tilde{\Sigma}_T$ may be any spacelike surface in the causal diamond with Σ_T as its boundary. The bulk Scwhinger-Keldysh integration region then looks as in figure 7.4.

Now that we have constructed the spacetime corresponding to the Schwinger-Keldysh contour, the rest of the derivation follows analogously to section 7.2.2. There is a replica symmetry where an angular coordinate θ takes us from one replica to the next. There may be a surface in the bulk that is fixed under this replica symmetry. The positive energy density requirement is replaced with a covariant analog, the null convergence condition given by $R_{ab}k^ak^b \geq 0 \forall k^a : k_ak^a = 0$. The rest of the derivation boils down to arguing that the near fixed surface metric takes the general form

$$ds^{2} = n^{2}dr^{2} + r^{2}d\theta^{2} + b_{i}d\theta dy^{i} + g_{ij}dy^{i}dy^{j} ,$$

$$g_{ij} = h_{ij} + r^{n}\cosh(\theta)K_{ij}^{1} + r^{n}\sinh(\theta)K_{ij}^{2} + \mathcal{O}(r^{2}) ,$$

$$R = r + \mathcal{O}(r^{3}) , \quad b_{i} = \mathcal{O}(r^{2}) ,$$
(7.61)

which looks like a Lorentzian analog of equation (7.44). Working to first order in ϵ when $n = 1 + \epsilon$ once again gives divergent contributions to the Ricci tensor

$$R_{\mu\nu} \sim \epsilon \frac{K_{ij}^{\alpha} g^{ij}}{r} . \tag{7.62}$$

The requirement of smoothness at r = 0 to first order in ϵ leads to the vanishing of the traces of the extrinsic curvatures

$$g^{ij}K^1_{ij} = g^{ij}K^2_{ij} = 0. (7.63)$$

The main difference to the previous case is that time independence does not automatically set the timelike combination of the extrinsic curvatures to zero and we get a spacetime extremal codimension two surface instead of simply a spatially extremal codimension two surface living in a constant time slice. The effective action again reduces to the computation of a proper area via a convenient choice of gauge.


Figure 7.4: Bulk extension of boundary Schwinger-Keldysh contour (figure 7.3). The two legs are the two copies of the bulk causal past \tilde{J}_T^- of $\tilde{\Sigma}_T$ and the boundary space is represented by a circle. The separation of the legs represents their separation in the imaginary time direction. The boundary folding surface Σ_T is the boundary of $\tilde{\Sigma}_T$ and $\tilde{\Sigma}_T$ is a spacelike surface. The region Aacross which there is gluing is indicated in red. The volume in this image self-intersects.

The HRT proposal was given a more rigorous treatment in 2016 by Aron Wall [64]. Wall also introduces a new formulation of the HRT proposal in terms of so-called maximin surfaces, defined by minimizing the area of \mathcal{M}_A for some arbitrary choice of $\tilde{\Sigma}_T$ while requiring $\mathcal{M}_A \in \tilde{\Sigma}_T$ and then maximizing the area with respect to the choice of $\tilde{\Sigma}_T$. The difference to the original proposal in the introduction to this section seems minor, but is mathematically essential.

Wall then goes on to prove the existence HRT surfaces for a general class of spacetimes, including ones containing black holes. In addition to this, the max/min surface formulation is necessary to extend the proof of strong subadditivity to the time dependent case.

7.4 Entanglement Builds Geometry

An interesting implication of the Ryu-Takayanagi formula is that areas are the bulk duals of entanglement in the boundary CFT. This means that given a CFT we should be able to reconstruct the bulk geometry. For every inequivalent way of splitting the CFT into two spatial regions, we can compute the area of the bulk effective event horizon by finding the entanglement entropy on the CFT side. We then want to find some bulk manifold M that can have these values of the extremal areas. This problem should be highly overconstrained since the entanglement entropies are a function on the space of subsets of the boundary spacetime while generally, the bulk geometry is described by a handful of functions.

Due to this, a boundary field theory that can conceivably have a geometrical dual it must have a very specific entanglement structure. Given a CFT with this specific type of entanglement structure we can reconstruct the bulk geometry. An exception to this is spacetimes that contain regions that are not penetrated by any extremal surface, such as the interior of black holes⁸. Thus without adding some sort of bulk quantum correction to the Ryu-Takayanagi prescription we cannot reconstruct black hole interiors. Such a region is in the literature called an *entanglement shadow*.

The derivation of Lewkowycz and Maldacena in section 7.2.2 already taught us that the Ryu-Takayanagi formula corresponds to a classical theory in the bulk, so the inaccessibility of black

 $^{^{8}}$ It is shown in [62] that no extremal surfaces may penetrate event horizons, at least in static spacetimes. Due to this, RT surfaces fail to probe black hole interiors.



Figure 7.5: To the left, we have divided a field theory living on a sphere into two hemispheres A and B. As we tune the inverse temperature β towards larger values, the symmetry of the problem fixes the shape of γ_A while tuning the area. As we increase β , the RT formula equation (7.29) tells us that the interfacing area between the hemispheres begins to pinch off. In addition, the monotonous relationship between entanglement and distance leads us to expect that reducing entanglement drives the hemispheres apart.

hole interiors is in some sense expected, as we expect quantum gravity to be important in this regime. Despite this the Ryu-Takayanagi formula can teach us a lot about the relationship between classical geometry and entanglement. We will see this by considering what happens to the bulk spacetime when you turn off the entanglement between two complementary regions A and B on the boundary. The following argument is due to Mark van Raamsdonk [15], and the construction is carried out explicitly in [66].

Consider a CFT living on a spatial ball. We may arbitrarily split the ball into two hemispheres, A and B as in figure 7.5. The density operators on A and B will describe two copies of the same thermal state, an example of a *thermofield double*. The thermofield double is constructed by entangling energy eigenstates $|r_i\rangle \in \mathcal{H}_A$ and $|l_i\rangle \in \mathcal{H}_B$ according to

$$|TFD\rangle = \frac{1}{\sqrt{Z(\beta)}} \sum_{i} e^{-\frac{1}{2}\beta E_i} |r_i, l_i\rangle \quad .$$
(7.64)

Tracing out either copy leaves a normally weighted thermal state

$$\rho_A = \frac{1}{Z(\beta)} \sum_i e^{-\beta E_i} |r_i\rangle \langle r_i| \quad .$$
(7.65)

Very importantly, the parameter β tunes the entanglement of the left and right hemispheres; $\beta = 0$ corresponds to the maximally mixed state, and $\beta = \infty$ puts both hemispheres in an individual pure state. Note that only a very specific temperature $T = 1/\beta$ corresponds to the CFT vacuum, since in the vacuum the temperature corresponding to a a boundary horizon (∂A) is fixed by the Unruh effect.

Let us now consider what happens when we tune the inverse temperature β . Note that the boundary geometry stays fixed under the following procedure, despite how it appears in the pictures, where we illustrate how the bulk geometry is deformed. The symmetry of the problem as illustrated in figure 7.5 tells us that the RT surface has to be a flat disk, regardless of what temperature the local states have. As we increase β , the area of γ_A must decrease, but reflection symmetry keeps it in a fixed shape. Therefore, the spacetime must start to *pinch off* as we increase β . Finally, as we bring $\beta \to \infty$ the two spacetimes should disconnect completely as in figure 7.6, since they are connected across a surface with vanishing area.



Figure 7.6: The limit as we completely disentangle the two local states in A and B is two separate spacetimes. This may be taken to imply the necessity of quantum entanglement for classical geometry to exist at all.

In addition, we can think about the mutual information of the two hemispheres. The mutual information is defined by

$$I(A,B) = S_A + S_B - \underbrace{S_{A \cup B}}_{0}, \qquad (7.66)$$

where the last term is zero since the full state is pure. The mutual information is positive definite as we showed in chapter 2.3, and represents an upper bound on the maximum correlation between operators in regions A and B. But we also showed in equation (6.70) that in terms of the geodesic distance $L[X, x_A, x_B]$, two-point correlators are described by a *geodesic approximation* given by

$$\langle \mathcal{O}_A(x_A)\mathcal{O}_B(x_B)\rangle = e^{-\frac{\Delta L[X;x_A,x_B]}{L_{AdS}}}.$$
(7.67)

As we bring $\beta \to \infty$, $S_A = S_B \to 0$, so the mutual information becomes upper bounded by a vanishing number. But the mutual information represents an upper bound on $\exp\left(-\frac{\Delta L[X;x_A,x_B]}{L_{AdS}}\right)$, so we conclude that the length of the geodesic connecting two points on the hemispheres A and B increases as $\beta \to \infty$. This means that in addition to the two halves of the spacetime pinching off, they are also driven apart as we try to disentangle them. Note that as the mutual information approaches zero, it is not clear whether there is a geometrical interpretation of the entangling surface and the geodesic may stop making sense.

The conclusion of this thought experiment is that without entanglement, spacetime seems to literally disconnect! Entanglement is the crucial ingredient rquired for bulk geometry to emerge from a boundary field theory. This relates back the the Reeh-Schlieder theorem of section 4.1.1, where we realized that in a continuum QFT there can be no pure local states. This means that to have pure 'local' states, spactime must be discontinuous at the boundary of the local region. This means that our result is definitely consistent with conventional wisdom coming from axiomatic field theory.

It is not entirely clear to what extent this conclusion holds, since as we commented some spacetimes contain an entanglement shadow that is not described by the Ryu-Takayanagi prescription. In addition given an honest CFT on the boundary the Hilbert space does not literally factorize into \mathcal{H}_R and \mathcal{H}_L , so the construction should be subjected to a more high-brow algebra of observables analysis. Despite this, at this point we should be readily convinced that entanglement and geometry are closely related. In Part II we will set out to explore the explicit mechanisms by which entanglement builds geometry.

Part II

Recent Developments

Introduction to Part II

In this part a number of recent results and conjectures about how entanglement information builds up spacetime will be treated. This part is a review of the different ideas that have sprung up, and are to some degree of abstraction inspired by the Ryu-Takayanagi conjecture. We will henceforth refer to the set of these ideas as the field of *emergent spacetime*.

Each of the three chapters in Part II aim to follow a handful of authors, trying to point out the main publications, results and lines of reasoning in a particular branch of "emergent spacetime". These authors are picked as representatives of their branches by me the author, but they are by no means the only contributors to their fields.

- Chapter 8 follows Mark van Raamsdonk through a number of papers. In this chapter we will discover how Einstein's general relativity emerges up to second order in perturbations from the completely general *entanglement first law* in the boundary CFT. This chapter involves no conjectures and very controlled configurations, so the results stand on *very* firm ground.
- Chapter 9 follows Juan Maldacena and Leonard Susskind through two papers [67, 7] to set the stage for the ER = EPR, complexity=action and complexity=volume conjectures. Inspired by this discovery we realize the importance of the microscopic structure of the CFT boundary states for the interior geometry. Finally a particular form of 'black hole microstate cosmology' is considered in which an FLRW braneworld cosmology living on the other side of a black hole is expressed holographically. The simplest version of such a configuration is found to be unstable.
- Chapter 10 investigates some wilder conjecture, giving the basics of the relation between tensor networks and cosmological spacetimes. We begin by relating the tensor network view-point to the approach of chapter 8 by showing that the precise interpretation of the boundary field theory may not be important for the emergence of classical geometry in the bulk using controlled machinery. After this, we give an introduction to tensor networks and the specific network MERA. We demonstrate the AdS/MERA and dS/MERA correspondences and check some consistency conditions coming from known gravitational physics, mostly following the PhD thesis of Charles Cao [68].

An alternate classification is that chapters 8, 9 and 10 are respectively the conservative, moderate and extreme versions of 'emergent gravity'.

Chapter 8

Universal Gravity from Entanglement Dynamics

In this chapter we try to understand in detail the precise relationship geometry and entanglement. This begins with the understanding that the geometry described by Einstein's field equations is dynamical. If entanglement equals geometry this means that we need a way to describe the dynamics of entanglement. We do this in section 8.1, where we introduce the first law of entanglement dynamics. We will see that small variations in entanglement are related to the modular energy of the CFT, similarly to how geometrical deformations are related to the stress energy in GR. In section 8.2 we show that *if* there exists a geometry that correctly computes the entanglement entropies for every ball shaped region in any given CFT, this geometry must be uniquely described by Einstein's field equations to first order in perturbations about AdS. In the following sections we extend the result to second order in perturbations, closely following a string of papers with Raamsdonk as a common author [69, 63, 15].

The most striking feature about the results of this chapter is that no conjectures are invoked and the results are very universal. With very minimal assumptions we find that *any* CFT admits states that have a dual gravitational description. In the second order calculation of section 8.4 the essential component is the construction of a very specific class of CFT states that closely resemble the coherent states of quantum mechanics. In this sense it seems that *the existence of a subspace of* gravitational states is a universal property of conformal field theories, at least if you only demand gravity to second order in perturbations about AdS.

8.1 Entanglement First Law

The entanglement first law relates the first order variation of the entanglement entropy to a variation of the expectation value of the Hamiltonian in a CFT. This relationship was first derived by Casini et al. [69] for the special case of spherical regions, and later as a special case of a relationship between modular Hamiltonians and entropy variations by Raamsdonk et al. [70].

Let us consider a CFT in d+1-dimensional Minkowski space. We choose a spatial region A and define the local density operator as ρ_A . Let us further define the complement of A, \overline{A} . Assuming that the full CFT is in the pure state $|\Psi\rangle$, the local density operator in A is given by

$$\rho_A = \operatorname{Tr}_{\overline{A}} \left[|\Psi\rangle\!\langle\Psi| \right] \,. \tag{8.1}$$

The modular Hamiltonian H_A is defined by

$$\rho_A = e^{-H_A} \ . \tag{8.2}$$

In the general case, the modular Hamiltonian is not related to the usual one in a tractable manner. For well chosen subregions and theories such as spatial balls in a CFT the modular Hamiltonian may be related to the physical one, as we saw in section 4.2.4.

We now consider the behaviour of the von Neumann entropy under an arbitrary variation of the state $|\Psi\rangle$. It is given by

$$\delta S_A = \delta \left(-\text{Tr} \left[\rho_A \ln \rho_A \right] \right)$$

= $-\text{Tr} \left[\delta \rho_A \ln \rho_A + \rho_A \frac{\delta \rho_A}{\rho_A} \right]$
= $\text{Tr} \left[\delta \rho_A H_A \right]$
= $\delta \langle H_A \rangle$, (8.3)

where in the second step we used that normalization of $\rho = \rho_A + \delta \rho$ together with linearity of the trace implies that $\text{Tr}[\delta \rho] = 0$. Therefore, for any spatial region we have that under an arbitrary perturbation of an arbitrary state,

$$\delta S_A = \delta \left\langle H_A \right\rangle \ . \tag{8.4}$$

The Ryu-Takayanagi conjecture gives us a gravitational interpretation of the lefthand side in terms of extremal surfaces. The righthand side is related to the asymptotic metric as in the Fefferman-Graham analysis of section 6.2.3. Another interesting property is that the equation equates geometry on the left with an energy on the right, reminiscent of the Einstein field equations.

In section 8.2 we show, specifying $|\Psi\rangle$ to be the vacuum state, that equality of both sides of this expression to first order imply the linearized Einstein field equations around pure AdS.

Hamiltonian-Entropy Relation for Spatial Balls in a CFT

For a CFT living in d + 1 dimensional Minkowski space we may relate the modular Hamiltonian to the stress energy tensor of the theory. We showed this in section 4.2.4 by using conformal transformations to relate the ball and the Rindler wedge. We then used the known form of the Rindler Hamiltonian as well as the transformation of the stress tensor by a Schwarzian under the inversion of these conformal transformations to obtain equation (4.183)

$$H_A = 2\pi \int_A \mathrm{d}^d x \frac{R^2 - r^2}{2R} T_{00}(x) , \qquad (8.5)$$

where R is the radius of the ball, and r is a radial coordinate with origin at the center of the ball and T_{00} is the expectation value of the 00-component of the stress energy tensor. The only quantity in this expression that depends on the state $|\Psi\rangle$ is the stress energy tensor, so we find that

$$\delta \langle H_A \rangle = 2\pi \int_A \mathrm{d}^d x \frac{R^2 - r^2}{2R} \delta T_{00}(x) \equiv \delta E_A^{\mathrm{hyp}} , \qquad (8.6)$$

where we have defined the hyperbolic energy E_A^{hyp} of the region A.

8.2 Einstein Equations From from CFT Entanglement First Law

In the previous section we found the relation

$$\delta S_A = \delta E_A^{\text{hyp}} . \tag{8.7}$$

The lefthand side may be interpreted in terms of extremal surfaces via the Ryu-Takayanagi formula. The righthand side may be interpreted in terms of the "boundary metric" as defined in the Fefferman-Graham analysis of section 6.2.3. We will now proceed to find the bulk duals of the leftand righthand sides of equation (8.7). We will discover that equation (8.7) holds for all possible choices of ball A in all boundary Lorentz frames *if and only if* the bulk Einstein field equations are satisfied to first order in perturbations about pure AdS. In the following we will consider a d+1 dimensional CFT dual to AdS_{d+2} .

Gravitational Computation of δS

The Ryu-Takayanagi relation tells us that S_A is given by the area of a codimension two surface \tilde{A} in the bulk that fulfills $\partial \tilde{A} = \partial A$. Explicitly, S_A is related to area of \tilde{A} by

$$S_A = \frac{\operatorname{Area}[A]}{4G_N} \,. \tag{8.8}$$

The surface \tilde{A} extremizes the area functional

Area
$$(G, X_{ext}) = \int_{\tilde{A}} \mathrm{d}^d \sigma \sqrt{-g}$$
 (8.9)

where g is the induced metric on \tilde{A} and X_{ext} describes the embedding of the extremal surface in the bulk spacetime. The induced metric g is related to the bulk metric G by a pullback

$$g_{ab} = G_{\mu\nu} \frac{\partial X^{\mu}}{\partial \sigma^a} \frac{\partial X^{\nu}}{\partial \sigma^b} .$$
(8.10)

We want to consider perturbations about pure bulk AdS, and we will use Poincaré coordinates in which the metric takes the form

$$ds^{2} = G^{0}_{\mu\nu}dx^{\mu}dx^{\nu} = \frac{1}{z^{2}}(-dt^{2} + dz^{2} + d\vec{x}^{2}) . \qquad (8.11)$$

Here, the AdS radius L has been set to 1. The boundary ∂A is defined in these coordinates by

$$\vec{x}^2 = R^2 , \quad z = 0 .$$
 (8.12)

We showed in section 7.2.3 that the bulk extremal surface \tilde{A} is the half-sphere defined by

$$z^2 + \vec{x^2} = R^2$$
, $z \ge 0$. (8.13)

Let us now consider a small variation of the area functional A, letting $G_{\mu\nu} = G^0_{\mu\nu} + \delta G_{\mu\nu}$ and $X = X_{ext} + \delta X$. This changes the extremal surface according to

$$A[G^0, X_{ext}] \to A[G^0 + \delta G, X_{ext} + \delta X] .$$
(8.14)

The original surface was extremal, so the variation of X doesn't contribute to first order:

$$A[G_0, X_{ext} + \delta X] = A[G_0, X_{ext}] + \mathcal{O}\left((\delta X)^2\right).$$
(8.15)

Therefore we may consider a variation only with respect to the metric of the area functional. We find

$$\delta_G A = \int \mathrm{d}^d \sigma \frac{1}{2} \sqrt{-g_0} g_0^{ab} \delta g_{ab} \tag{8.16}$$

where δg_{ab} means $\frac{\partial X^{\mu}}{\partial \sigma^a} \frac{\partial X^{\nu}}{\partial \sigma^b} \delta G_{\mu\nu}$. We have thus shown that the first order perturbation of the area functional A is just the perturbation of the bulk metric integrated over the extremal surface. Parametrizing the extremal half-sphere by $\sigma^i = x^i$, explicitly performing the pullback and inserting $S = A/4G_N$ one finds that

$$\delta S = \frac{R}{8G_N} \int d^d x (\delta_{ij} - \frac{1}{R^2} x_i x_j) H^{ij} .$$
 (8.17)

Gravitational Computation of δE_A^{hyp}

We would now wish to give a gravitational interpretation to the RHS of equation (8.7). Let us consider the Fefferman-Graham near boundary representation of AdS with L = 1

$$ds^{2} = \frac{1}{z^{2}} (dz^{2} + dx_{\mu} dx^{\mu} + h_{\mu\nu}(x, z) dx^{\mu} dx^{\nu}) . \qquad (8.18)$$

This metric only differs from pure AdS in Poincaré coordinates by the perturbation $h_{\mu\nu}(x,z)$. In section 6.2.3 we defined the boundary metric $\gamma_{\mu\nu} = h_{\mu\nu}(x,0)$, and argued that it was related to the boundary stress tensor by equation (6.136):

$$\langle T_{\mu\nu} \rangle = \frac{d}{16\pi G_N} \gamma_{ab} \ . \tag{8.19}$$

Then, inserting into equation (8.6) we find that

$$\delta E_A^{\text{hyp}} = \frac{d}{16G_N} \int_A d^d x \frac{R^2 - r^2}{R} \delta \gamma_{00}(x) .$$
 (8.20)

This is an integral of the boundary metric $\gamma = h(0,x)$ over the boundary region A.

8.2.1 Linearized Einstein Field Equations from $\delta S = \delta E$

We can now use the results we just derived to prove that $\delta S = \delta E$ is equivalent to the perturbed metric $G = G^0 + \delta G$ obeying the linarized Einstein field equations (EFE). We will specialize to 2+1 boundary dimensions as the simplest nontrivial example, but we will use a method that generalizes in a straightforward way to arbitrary dimension. In section 4 of [70] a proof is carried out without assuming analyticity of the metric, but it contains several steps that are specific to 2+1 boundary dimensions¹.

In the following we will multiply both sides of $\delta S = \delta E$ by $8G_N R$ for tidiness. Since we are in three boundary dimensions, d = 3. We denote the boundary region by A and the bulk half-sphere with the same boundary as A by S_A . Equations (8.17) and (8.20) then tell us that for a disk of

¹Another alternative proof is given by [69], but they only show that $\delta S = \delta E$ is satisfied by metrics satisfying the linearized Einstein equations, i.e. only one direction of the equivalence.

any radius R, centered at any boundary spatial point (x_0, y_0) the bulk surface integral

$$8G_N R\delta S \equiv \delta \hat{S} = \int_{S_A} dx dy \left[h_{xx} \left(\sqrt{R^2 - x^2 - y^2}, t, x + x_0, y + y_0 \right) (R^2 - x^2) + h_{yy} \left(\sqrt{R^2 - x^2 - y^2}, t, x + x_0, y + y_0 \right) (R^2 - y^2) - 2h_{xy} \left(\sqrt{R^2 - x^2 - y^2}, t, x + x_0, y + y_0 \right) xy \right]$$

$$(8.21)$$

is equal to the boundary surface integral

$$8G_N R\delta E^{\text{hyp}} \equiv \delta \hat{E} = \frac{3}{2} \int_A \mathrm{d}x \mathrm{d}y \left(R^2 - x^2 - y^2\right) h_{tt}(0, x + x_0, y + y_0) \,. \tag{8.22}$$

The linearized Einstein equations about an AdS background are given by

$$h_{\mu}{}^{\mu} = 0 , \quad \partial_{\mu}h^{\mu\nu} = 0 , \quad \frac{1}{z^4}\partial_z \left(z^4\partial_z h_{\mu\nu}\right) + \partial^2 h_{\mu\nu} = 0 .$$
 (8.23)

These are the $zz, z\mu$ and $\mu\nu$ components of the bulk einstein field equations

$$R_{MN} - \frac{1}{2}Rg_{MN} - 3g_{MN} = 0 \tag{8.24}$$

for the metric in 8.18. We have used that in our units, the cosmological constant is -3 in four spacetime dimensions.

$\mathbf{EFE} \Rightarrow \delta \mathbf{S} = \delta \mathbf{E}$

The key ingredient in this particular proof is to assume that teh entropy S_A and hyperbolic energy E_A^{hyp} can be Taylor expanded as a function of the radius R of the boundary ball. This gives us an infinite set of relations, order by order in R relating δS_A and δE_A^{hyp} . This implements the balls of "all radii" condition mentioned before. The goal is to show that given that h solves the linearized EFE (equation (8.23)), $\delta S = \delta E$ is satisfied.

It suffices to expand in powers of z, since z is defined by $z = \sqrt{R^2 - x^2 - y^2}$. We define

$$h_{\mu\nu}(z,x,y) = \sum_{n=0} z^n h_{\mu\nu}^{(n)}(x,y).$$
(8.25)

Note that in this notations $h_{\mu\nu}(0,x,y) = \gamma_{\mu\nu} = h_{\mu\nu}^{(0)}$. It is possible to show that the integrals in equations (8.21) and (8.22) can be performed explicitly term by term in an expansion. To do this we expand h(z,x,y) over A in powers of x, y as well. The general result used is

$$\int_{A} \mathrm{d}x \mathrm{d}y (R - x^2 - y^2)^{n/2} x^{2m_x} y^{2m_y} = R^{n+2m_x+2m_y+2} I_{n,m_x,m_y}$$
(8.26)

where

$$I_{n,m_x,m_y} = \frac{\Gamma(m_x + \frac{1}{2})\Gamma(m_y + \frac{1}{2})\Gamma(\frac{n}{2} + \frac{1}{2})}{\Gamma(\frac{n}{2} + m_x + m_y + 2)} .$$
(8.27)

Then, by inserting into equation (8.22) we find that

$$\delta \hat{E} = \frac{3}{2} \sum_{m_x, m_y=0}^{\infty} R^{2+2m_x+2m_y} I_{2,m_x,m_y}(\partial_x)^{2m_x} (\partial_y)^{2m_y} h_{tt}^{(0)}(t, x_0, y_0)$$
(8.28)

and for equation (8.21)

$$\delta \hat{S} = \sum_{n,m_x,m_y=0} R^{n+2m_x+2m_y+2} \left[\frac{1}{(2m_x)!(2m_y)!} \left(I_{n,m_x,m_y} - I_{n,m_x+1,m_y} \right) (\partial_x)^{2m_x} (\partial_y)^{2m_y} h_{xx}^{(n)}(t,x_0,y_0) \right. \\ \left. + \frac{1}{(2m_x)!(2m_y)!} \left(I_{n,m_x,m_y} - I_{n,m_x,m_y+1} \right) (\partial_x)^{2m_x} (\partial_y)^{2m_y} h_{xx}^{(n)}(t,x_0,y_0) \right. \\ \left. - 2R^2 \frac{1}{(2m_x)!(2m_y)!} I_{n,m_x+1,m_y+1} (\partial_x)^{2m_x+1} (\partial_y)^{2m_y+1} h_{xy}^{(n)}(t,x_0,y_0) \right]$$

$$(8.29)$$

Note that the $m_x + 1$ type factors in the *I*'s are there to match the powers of *R* between terms. The next step is to use the linearized Einstein equations to eliminate $h_{xy}^{(n)}$ from equation (8.29). Inserting the expansion of *h* the linearized EFE (equation (8.23)) become

$$h_{tt}^{(n)} = h_{xx}^{(n)} + h_{yy}^{(n)}$$
(8.30)

$$\partial_t h_{tt}^{(n)} = \partial_x h_{tx}^{(n)} + \partial_y h_{ty}^{(n)} \tag{8.31}$$

$$\partial_t h_{xt}^{(n)} = \partial_x h_{xx}^{(n)} + \partial_y h_{xy}^{(n)} \tag{8.32}$$

$$\partial_t h_{yt}^{(n)} = \partial_x h_{yx}^{(n)} + \partial_y h_{yy}^{(n)} \tag{8.33}$$

$$h_{\mu\nu}^{(n)} = \frac{1}{n(n+3)} (\partial_t^2 - \partial_x^2 - \partial_y^2) h_{\mu\nu}^{(n-2)} , \quad n \ge 2$$
(8.34)

$$h_{\mu\nu}^{(1)} = 0 . (8.35)$$

equation (8.30) is just the traceless condition for each order in z with the components written out. The same can be said about equations (8.31)-(8.33), they are just the order by order divergence-free condition of the linearized EFE. The last two equations follow straightforwardly by inserting the expansion into the last of the linearized EFE in equation (8.23).

To solve for $h_{xy}^{(n)}$ in terms of the other components, let us start by applying two time derivatives to both sides of equation (8.30). By then inserting equations (8.31)-(8.35) in sequence we find

$$\begin{split} \partial_t^2 h_{tt}^{(n)} &= \partial_t^2 \left(h_{xx}^{(n)} + h_{yy}^{(n)} \right) \ , \\ \partial_t \left(\partial_x h_{tx}^{(n)} + \partial_y h_{ty}^{(n)} \right) &= \partial_t^2 \left(h_{xx}^{(n)} + h_{yy}^{(n)} \right) \ , \\ \partial_x \left(\partial_x h_{xx}^{(n)} + \partial_y h_{xy}^{(n)} \right) &+ \partial_y \left(\partial_x h_{yx}^{(n)} + \partial_y h_{yy}^{(n)} \right) &= \partial_t^2 \left(h_{xx}^{(n)} + h_{yy}^{(n)} \right) \ , \\ \partial_x^2 h_{xx}^n + \partial_y^2 h_{yy}^{(n)} + 2 \partial_x \partial_y h_{xy}^{(n)} &= \partial_t^2 \left(h_{xx}^{(n)} + h_{yy}^{(n)} \right) \ , \\ \partial_x^2 h_{xx}^n + \partial_y^2 h_{yy}^{(n)} + 2 \partial_x \partial_y h_{xy}^{(n)} &= (\partial_x^2 + \partial_y^2) (h_{xx}^{(n)} + h_{yy}^{(n)}) + (n+2)(n+5) \left(h_{xx}^{n+2} + h_{xx}^{n+2} \right) \ , \\ &= 2 \partial_x \partial_y h_{xy}^{(n)} &= (\partial_x^2 h_{yy}^{(n)} + \partial_y^2 h_{xx}^{(n)} + (n+2)(n+5) \left(h_{xx}^{n+2} + h_{xx}^{n+2} \right) \ . \end{split}$$

Since derivatives commute it is straightforward to use this last equation to eliminate $h_{xy}^{(n)}$ from equation (8.29). We find that the variation of the gravitational entropy is

$$\delta \hat{S} = \sum_{n,m_x,m_y=0}^{\infty} R^{n+2m_x+2m_y} \left(\frac{1}{(2m_x)!(2m_y)!} \partial_x^{2m_x} \partial_y^{2m_x} C_{n,m_x,m_y}^{xx} h_{xx}^{(0)}(t,x_0,y_0) + \frac{1}{(2m_x)!(2m_y)!} \partial_x^{2m_x} \partial_y^{2m_x} C_{n,m_x,m_y}^{yy} h_{yy}^{(0)}(t,x_0,y_0) \right)$$
(8.36)

where the coefficients C for $n \ge 2$ are

$$\begin{split} C_{n,m_x,m_y}^{xx} &= I_{n,m_x,m_y} - I_{n,m_x+1,m_y} - \frac{2m_y}{2m_x+1} I_{n,m_x+1,m_y} - \frac{n(n+3)}{(2m_x+1)(2m_y+1)} I_{n-2,m_x+1,m_y+2} \\ &= 0 \ , \\ C_{n,m_x,m_y}^{yy} &= I_{n,m_x,m_y} - I_{n,m_x,m_y+1} - \frac{2m_x}{2m_y+1} I_{n,m_x,m_y+1} - \frac{n(n+3)}{(2m_x+1)(2m_y+1)} I_{n-2,m_x+1,m_y+2} \\ &= 0 \ . \end{split}$$

whereas for n = 1 and n = 0 they are

$$C_{1,m_x,m_y}^{xx} = I_{1,m_x,m_y} - I_{1,m_x+1,m_y} - \frac{2m_y}{2m_x+1}I_{1,m_x+1,m_y} = \frac{4}{3}I_{3,m_x,m_y} ,$$

$$C_{1,m_x,m_y}^{yy} = I_{1,m_x,m_y} - I_{1,m_x,m_y+1} - \frac{2m_x}{2m_y+1}I_{1,m_x,m_y+1} = \frac{4}{3}I_{3,m_x,m_y} ,$$

$$C_{0,m_x,m_y}^{xx} = I_{0,m_x,m_y} - I_{0,m_x+1,m_y} - \frac{2m_y}{2m_x+1}I_{0,m_x+1,m_y} = \frac{3}{2}I_{2,m_x,m_y} ,$$

$$C_{0,m_x,m_y}^{yy} = I_{0,m_x,m_y} - I_{0,m_x,m_y+1} - \frac{2m_x}{2m_y+1}I_{n,m_x,m_y+1} = \frac{3}{2}I_{2,m_x,m_y} ,$$

after simplification by using the definition of I (and the Γ 's inside). Using that $h_{\mu\nu}^{(1)} = 0$ as well as the results for the C's, only the n = 0 term of the sum in equation (8.29) survives and we have

$$\delta \hat{S} = \sum_{m_x, m_y=0}^{\infty} R^{2m_x + 2m_y + 2} \frac{1}{(2m_x)!(2m_y)!} \frac{3}{2} I_{2, m_x, m_y} \partial_x^{2m_x} \partial_y^{2m_x} \left(h_{xx}^{(0)} + h_{yy}^{(0)}\right)$$
$$= \sum_{m_x, m_y=0}^{\infty} R^{2m_x + 2m_y + 2} \frac{1}{(2m_x)!(2m_y)!} \frac{3}{2} I_{2, m_x, m_y} \partial_x^{2m_x} \partial_y^{2m_x} \left(h_{tt}^{(0)}\right)$$
$$= \delta \hat{E} .$$
(8.37)

where we have once again used equation (8.33). We may conclude that assuming that h satisfies the linear Einstein equations implies that $\delta S = \delta E^{\text{hyp}}$ is satisfied.

$\delta S = \delta E \Rightarrow \mathbf{EFE}$

We just showed that $\delta S = \delta E$ is satisfied by solutions to the Einstein field equations. We will now show that this applies *only* if *h* satisfies the linearized EFE. That is, we will prove that the unique theory compatible with the Ryu-Takayanagi formula for the entanglement entropy of all boundary spatial balls of all radii in all Lorentz frames to first order is linearized Einstein gravity. If we take the Ryu-Takayanagi formula as a fundamental relationship between entanglement and geometry, then this provides us a tool that lets us view Einstein gravity as the unique geometric theory that emerges from the CFT entanglement structure.

The trick here is to use a proof by contradiction. Let $h_{\mu\nu}^{\text{EFE}}$ denote the bulk metric perturbation that satisfies the linearized EFE's in the bulk as well as $h_{\mu\nu}^{\text{EFE}}(0,x) = (16\pi G_N/3)T_{\mu\nu}$. Let $h_{\mu\nu}$ denote a supposed different metric perturbation that satisfies the same boundary condition and $\delta S = \delta E^{\text{hyp}}$, but not the EFE. It follows that $\Delta = h - h^{\text{EFE}}$ satisfies

$$\Delta_{\mu\nu}(z=0,t,x,y) = 0.$$
(8.38)

The goal is top show that this restriction, as well as the restrictions coming from $\delta S = \delta E^{\text{hyp}}$ for all radii and Lorentz frames sets $\Delta = 0$ in all of the bulk. Explicitly, we have

$$\delta \hat{S} = 0 = \int_{A} dx dy \left[\Delta_{xx} \left(\sqrt{R^{2} - x^{2} - y^{2}}, t, x + x_{0}, y + y_{0} \right) (R^{2} - x^{2}) + \Delta_{yy} \left(\sqrt{R^{2} - x^{2} - y^{2}}, t, x + x_{0}, y + y_{0} \right) (R^{2} - y^{2}) - 2\Delta_{xy} \left(\sqrt{R^{2} - x^{2} - y^{2}}, t, x + x_{0}, y + y_{0} \right) xy \right].$$

$$(8.39)$$

We use the same integral relation as before (equation (8.26)) and expand Δ in powers of z to solve it. This means that we assume that Δ is equal to its multivariate Taylor expansion, i.e. that Δ is analytic. We solve order by order in R to implement the all radii condition. By defining

$$\Delta_{\mu\nu} = \sum_{n=0}^{\infty} z^n \Delta_{\mu\nu}^{(n)}(0, x, y) , \qquad (8.40)$$

and inserting into equation (8.39), we find that

$$\delta \hat{S} = \sum_{n,m_x,m_y=0} R^{n+2m_x+2m_y+2} \left[\frac{1}{(2m_x)!(2m_y)!} \Big(I_{n,m_x,m_y} - I_{n,m_x+1,m_y} \Big) (\partial_x)^{2m_x} (\partial_y)^{2m_y} \Delta_{xx}^{(n)}(t,x_0,y_0) \right. \\ \left. + \frac{1}{(2m_x)!(2m_y)!} \Big(I_{n,m_x,m_y} - I_{n,m_x,m_y+1} \Big) (\partial_x)^{2m_x} (\partial_y)^{2m_y} \Delta_{xx}^{(n)}(t,x_0,y_0) \right. \\ \left. - 2R^2 \frac{1}{(2m_x)!(2m_y)!} I_{n,m_x+1,m_y+1} (\partial_x)^{2m_x+1} (\partial_y)^{2m_y+1} \Delta_{xy}^{(n)}(t,x_0,y_0) \right] .$$

$$(8.41)$$

Using this, vanishing of terms at order \mathbb{R}^{N+2} implies

$$\Delta_{xx}^{(N)}(t,x_0,y_0) + \Delta_{yy}^{(N)}(t,x_0,y_0) = \sum_{m_x,m_y \neq 0} C_{xx}^{N,m_x,m_y} \partial_x^{2m_x} \partial_y^{2m_y} \Delta_{xx}^{(N-2m_x-2m_y)}(t,x_0,y_0) C_{yy}^{N,m_x,m_y} \partial_x^{2m_x} \partial_y^{2m_y} \Delta_{yy}^{(N-2m_x-2m_y)}(t,x_0,y_0) C_{xy}^{N,m_x,m_y} \partial_x^{2m_x-1} \partial_y^{2m_y-1} \Delta_{xy}^{(N-2m_x-2m_y)}(t,x_0,y_0) .$$

$$(8.42)$$

The C coefficients can be read off from equation (8.41), for the first few orders they are given by (suppressing the (t,x_0,y_0) argument of the $\Delta^{(n)}$)

$$\begin{aligned} \Delta_{xx}^{(0)} + \Delta_{yy}^{(0)} &= 0 , \\ \Delta_{xx}^{(1)} + \Delta_{yy}^{(1)} &= 0 , \\ \Delta_{xx}^{(2)} + \Delta_{yy}^{(2)} &= -\frac{1}{4} (\partial_y^2 \Delta_{xx}^{(0)} + \Delta_x^2 \Delta_{yy}^{(0)}) - \frac{3}{20} (\partial_x^2 \Delta_{xx}^{(0)} + \Delta_y^2 \Delta_{yy}^{(0)}) + \frac{1}{5} (\partial_x \partial_y \Delta_{xy}^{(0)}) , \\ \Delta_{xx}^{(3)} + \Delta_{yy}^{(3)} &= -\frac{1}{6} (\partial_y^2 \Delta_{xx}^{(1)} + \Delta_x^2 \Delta_{yy}^{(1)}) - \frac{1}{6} (\partial_x^2 \Delta_{xx}^{(1)} + \Delta_y^2 \Delta_{yy}^{(1)}) + \frac{1}{9} (\partial_x \partial_y \Delta_{xy}^{(1)}) . \end{aligned}$$

$$\end{aligned}$$

$$\tag{8.43}$$

This set of equations determines the linear combination $\Delta_{xx}^{(n)} + \Delta_{yy}^{(n)}$ to all orders in terms of lower order terms. Apart from the constraint that $\Delta_{\mu\nu}(0,t,x,y) = 0$, implying $\Delta_{\mu\nu}^{(0)} = 0$, the remaining components of $\Delta_{\mu\nu}^{(0)}$ are completely unconstrained.

To get further we need to demand that equation (8.39) holds in in an arbitrary Lorentz frame on the boundary. Explicitly we may consider a general Lorentz boost in three dimensions, given by

$$\Lambda = \begin{bmatrix} \gamma & \gamma \beta_x & \gamma \beta_y \\ \gamma \beta_x & 1 + \beta_x^2 \frac{\gamma^2}{\gamma + 1} & \beta_x \beta_y \frac{\gamma^2}{1 + \gamma} \\ \gamma \beta_y & \beta_x \beta_y \frac{\gamma^2}{1 + \gamma} & 1 + \beta_y^2 \frac{\gamma^2}{\gamma + 1} \end{bmatrix}$$
(8.44)

Then, since $\Delta_{\mu\nu}$ is supposed to be a tensor, we have

$$\Delta_{xx} + \Delta_{yy} = \Lambda_x^{\ \mu} \Lambda_x^{\ \nu} \Delta_{\mu\nu} + \Lambda_y^{\ \mu} \Lambda_y^{\ \nu} \Delta_{\mu\nu} , \quad \Delta_{xy} = \Lambda_x^{\ \mu} \Lambda_y^{\ \nu} \Delta_{\mu\nu} .$$
(8.45)

Up to a constant this tells us that

$$\Delta_{xx} + \Delta_{yy} \sim \Delta_{ii} + 2\beta_i \Delta_{it} + \beta^2 (\Delta_{tt} - \frac{1}{2}\Delta_{ii}) + (\beta_i \beta_j - \frac{1}{2}\delta_{ij}\beta^2)\Delta_{ij}$$
(8.46)

The first equation in (8.43) is still automatically satisfied since the vanishing of a tensor is a Lorentz invariant statement. For the second equation, we find that

$$\Delta_{ii}^{(1)} + 2\beta_i \Delta_{it}^{(1)} + \beta^2 (\Delta_{tt}^{(1)} - \frac{1}{2} \Delta_{ii}^{(1)}) + (\beta_i \beta_j - \frac{1}{2} \delta_{ij}^{(1)} \beta^2) \Delta_{ij}^{(1)} = 0 .$$
(8.47)

For fixed x_0, y_0 this is a polynomial equation for β_i . Equation equation (8.47) must vanish for an arbitrary boost, so the polynomial must be zero for all β , rendering it identically zero. At order β^0 this implies

Z

$$\Delta_{ii}^{(1)} = 0 . (8.48)$$

To order β we find

$$\Delta_{it}^{(1)} = 0 . (8.49)$$

To order β^2 we have two equations,

$$\Delta_{tt}^{(1)} = \frac{1}{2} \Delta_{ii}^{(1)} = 0$$

$$\Delta_{ij}^{(1)} = \delta_{ij} \frac{1}{2} \Delta_{kk}^{(1)} = 0 .$$
(8.50)

Thus, $\Delta_{\mu\nu}^{(1)} = 0$. In equation (8.43), all terms beyond the third have a nontrivial righthand side. This righthand side luckily only depends on lower order $\Delta^{(n)}$, the two first of which we set to zero. Thus, we only have to prove that the lefthand side has to vanish at all orders in n given that the righthand side is zero. This follows trivially by setting $1 \rightarrow n$ in the preceding analysis. By induction we then have that $\Delta_{\mu\nu}^{(n)} = 0$ for all n, proving that the first law of entanglement together with the Ryu-Takayanagi formula uniquely imply the linear EFE in the bulk as long as we consider only analytic metrics.

8.3 Quantum Fisher Information

In the previous section we introduced the first law of entanglement for infinitesimal variations around a reference state. It would be useful to find higher order constraints on the relationship between the Hamiltonian and the variation of the entropy. This is provided by the positivity of relative entropy, whose second order variation is called the quantum Fisher information. This quantity will let us extend the result of the previous section to second order in perturbation theory, as well as letting us include bulk matter fields.

The first law of entanglement has a natural generalization to finite perturbations Δ in the form of the inequality

$$\Delta \langle H_A \rangle - \Delta S_A \ge 0 . \tag{8.51}$$

To show this, remember the definition of the quantum relative entropy as

$$S(\rho||\rho_0) = \operatorname{Tr}\left[\rho \ln \rho - \rho \ln \rho_0\right] . \tag{8.52}$$

The relative entropy is non-negative, as we showed in full generality in section 4.1.4. Let $H_0 = -\ln \rho_0$ be the modular Hamiltonian of the reference state. Then,

$$S(\rho||\rho_0) = \operatorname{Tr} \left[\rho \ln \rho + \rho_0 \ln \rho_0 - \rho_0 \ln \rho_0 - \rho \ln \rho_0\right]$$

= Tr [-(\rho - \rho_0) \ln \rho_0 + \rho \ln \rho - \rho \ln \rho_0]
= \Delta \langle H_0 \rangle - \Delta S

together with positivity of relative entropy automatically implies equation (8.51). In addition, monotonicity of relative entropy implies that for $A \in B$,

$$S(\rho_A || \rho_{0,A}) \le S(\rho_B || \rho_{0,B}) . \tag{8.54}$$

The modular Hamiltonian is related to the boundary CFT stress tensor by equation (4.183), so for spatial balls on the boundary we have

$$2\pi \int_{A} \mathrm{d}^{d} x \frac{R^{2} - r^{2}}{2R} T_{00}(x) - \Delta S_{A} \ge 0$$
(8.55)

where the lefthand side must increase monotonically with R due to monotonicity

$$\frac{\mathrm{d}}{\mathrm{d}R} \left(2\pi \int_{A} \mathrm{d}^{d} x \frac{R^{2} - r^{2}}{2R} T_{00}(x) - \Delta S_{A} \right) \ge 0 .$$
(8.56)

Positivity of the relative entropy implies that $\rho = \rho_0$ is a minimum of $S(\rho||\rho_0)$. This means that first-order variations away from ρ_0 must vanish. Therefore, we have that

$$S(\rho_0 + \delta \rho || \rho_0) = \delta \langle H_0 \rangle - \delta S_A = 0 , \qquad (8.57)$$

which is just us recovering the first law of entanglement. Let us in the following denote $\frac{d}{d\epsilon}(\ldots) \equiv (\ldots)$ and take $\rho = \rho_0 + \epsilon \delta \rho + \epsilon^2 \delta^2 \rho \ldots$ We may now expand to second order and find

$$\delta^{2}S(\rho_{0} + \delta\rho ||\rho_{0}) = \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}\epsilon} \operatorname{Tr} \left[\dot{\rho} \ln \rho + \rho(\ln \rho) - \dot{\rho} \ln \rho_{0}\right] \Big|_{\epsilon=0}$$

$$= \frac{1}{2} \operatorname{Tr} \left[\ddot{\rho} \ln \rho + 2\dot{\rho}(\ln \rho) + \rho(\ln \rho) - \ddot{\rho} \ln \rho_{0}\right] \Big|_{\epsilon=0}$$

$$= \frac{1}{2} \operatorname{Tr} \left[\dot{\rho}(\ln \rho)\right]$$

$$\equiv \langle \delta\rho, \delta\rho \rangle$$
(8.58)

where we have defined $\langle \delta, \delta \rangle$, the quantum Fisher information. The penultimate step uses the fact that $\operatorname{Tr}[\rho(\ln \rho)] = \operatorname{Tr}[\delta\rho] = 0 \ \forall \epsilon$, meaning $\partial_{\epsilon}^{n} \operatorname{Tr}[\rho(\ln \rho)] = 0$ for all n. Specifically the identity for

n = 1 was used. The vanishing of the trace of the variation follows as usual because the state perturbation must conserve the unit trace of the density operator.

The quantum Fisher information can be promoted to an inner product on the tangent of the manifold defined by the set of states ρ_0 according to

$$\langle \delta \rho_1, \delta \rho_2 \rangle \equiv \frac{1}{2} \left(\langle \delta \rho_1 + \delta \rho_2, \delta \rho_1 + \delta \rho_2 \rangle - \langle \delta \rho_1, \delta \rho_1 \rangle - \langle \delta \rho_2, \delta \rho_2 \rangle \right) . \tag{8.59}$$

Since the relative entropy is positive definite the Fisher information metric is nondegenerate and non-negative, meaning it may be treated as a Riemannian metric.

We may obtain an explicit form of the Fisher information that will be of use in the next section. Let us denote $\rho(\epsilon) = \rho_0 + \epsilon \delta \rho + \epsilon^2 \delta^2 \rho \dots$ We may then use the identities²

$$-\ln(X) = \int_0^\infty \frac{\mathrm{d}s}{s} \left(e^{-sX} - e^{-s} \right) , \qquad (8.60)$$

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}e^{A+\epsilon B} = \int_0^1 \mathrm{d}x \, e^{Ax} B e^{(1-x)A} \tag{8.61}$$

to rewrite the third line of equation(8.58) according to

$$\frac{1}{2} \frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} S(\rho(\epsilon)||\rho_0)\Big|_{\epsilon=0} = \frac{1}{2} \int_0^1 \mathrm{d}x \int_0^\infty \mathrm{d}s \mathrm{Tr} \left[\delta\rho e^{-sx\rho_0} \delta\rho e^{-(1-x)s\rho_0}\right]$$
(8.62)

Here we have first applied equation (8.60) to the logarithmic term in equation (8.58), keeping only the zeroth order term in ϵ . The application of equation (8.61) to obtain equation (8.62) is then straightforward. By picking a basis in which ρ_0 is diagonalized we can express the trace in terms of the eigenvalues of ρ_0 , ρ_a as

$$(\delta\rho)^{a}{}_{b}e^{-x\rho_{b}s}(\delta\rho)^{b}{}_{a}e^{-(1-x)\rho_{a}s} = e^{x(\rho_{a}-\rho_{b})s-\rho_{a}s}(\delta^{2}\rho)^{b}{}_{a}(\delta\rho)^{a}{}_{b}$$
(8.63)

where sums over a and b are implied. For $\rho_a = \rho_b$ the integration over x is trivial, while in the case $\rho_a \neq \rho_b$ it is elementary. We find that

$$\frac{1}{2} \frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} S(\rho(\epsilon)||\rho_0)\Big|_{\epsilon=0} = \frac{1}{2} \int_0^\infty \mathrm{d}s \left[\sum_{\rho_a=\rho_b} e^{-\rho_a s} (\delta\rho)^b{}_a (\delta\rho)^a{}_b + \sum_{\rho_a\neq\rho_b} e^{-\rho_a s} \frac{1}{(\rho_a-\rho_b)s} \left(e^{(\rho_a-\rho_b)s} - 1 \right) (\delta\rho)^b{}_a (\delta\rho)^a{}_b \right] \\
= \frac{1}{2} \int_0^\infty \mathrm{d}s \left[\sum_{\rho_a=\rho_b} e^{-\rho_a s} (\delta\rho)^b{}_a (\delta\rho)^a{}_b + \sum_{\rho_a\neq\rho_b} \frac{e^{-\rho_b s} - e^{-\rho_a s}}{(\rho_a-\rho_b)s} (\delta\rho)^b{}_a (\delta\rho)^a{}_b \right] \tag{8.64}$$

The first term is simple to integrate, while for the second we reuse equation (8.60):

$$\int_{0}^{\infty} \mathrm{d}s \sum_{\rho_{a} \neq \rho_{b}} \frac{e^{-\rho_{b}s} - e^{-s} - e^{-\rho_{a}s} + e^{-s}}{(\rho_{a} - \rho_{b})s} (\delta\rho)^{b}{}_{a} (\delta\rho)^{a}{}_{b} = \sum_{\rho_{a} \neq \rho_{b}} \frac{\ln \rho_{a} - \ln \rho_{b}}{(\rho_{a} - \rho_{b})} (\delta\rho)^{b}{}_{a} (\delta\rho)^{a}{}_{b} .$$
(8.65)

²The first identity is shown by working in a basis that diagonalizes X. One then notes that both sides are zero for X = 1. By differentiating both sides with respect to X one sees that they have the same derivative at every point, implying equality of the full functions. The second identity is shown by expanding both sides in powers of ϵ and performing the integration explicitly.

And we see finally that

$$\frac{1}{2} \frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} S(\rho(\epsilon) ||\rho_0) \Big|_{\epsilon=0} = \frac{1}{2} \sum \frac{1}{\rho_a} (\delta\rho)^b{}_a (\delta\rho)^a{}_b + \frac{1}{2} \sum_{\rho_a \neq \rho_b} \frac{\ln \rho_a - \ln \rho_b}{(\rho_a - \rho_b)} (\delta\rho)^b{}_a (\delta\rho)^a{}_b .$$
(8.66)

There is another representation of the Fischer information available that ends up becoming proportional time-ordered correlations functions in modular time. This representation is the key result enabling the second order CFT analysis in the next section. It is possible to show that

$$\frac{1}{4} \int_{-\infty}^{\infty} \frac{e^{\frac{isx}{2\pi}}}{1 + \cosh(s)} = \frac{x}{2(e^{x/2} - e^{-x/2})} , \qquad (8.67)$$

by performing a contour integral around the upper half plane. Inserting $x = \ln \left(\frac{\rho_b}{\rho_a}\right)$ we see that

$$\frac{x}{2(e^{x/2} - e^{-x/2})} = \sqrt{\rho_a \rho_b} \frac{\ln \rho_b - \ln \rho_a}{2(\rho_b - \rho_a)}$$
(8.68)

so that the integral representation using equation (8.67) is given by

$$\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2}S(\rho(\epsilon)||\rho_0)\Big|_{\epsilon=0} = \int_{-\infty}^{\infty} \mathrm{d}s\frac{1}{1+\cosh(s)}\frac{1}{\sqrt{\rho_a\rho_b}}\left(\frac{\rho_b}{\rho_a}\right)^{\frac{\imath s}{2\pi}}(\delta\rho)^{b}{}_{a}(\delta\rho)^{a}{}_{b} \tag{8.69}$$

Restoring matrix notation (inserting ρ_a between contracted *a*'s et cetera), we find

$$\frac{1}{2} \frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} S(\rho(\epsilon)||\rho_0)\Big|_{\epsilon=0} = \frac{1}{4} \int_{-\infty}^{\infty} \mathrm{d}s \, \frac{1}{1+\cosh(s)} \mathrm{Tr}\left[\delta\rho\rho_0^{-\frac{1}{2}-\frac{is}{2\pi}} \delta\rho\rho_0^{-\frac{1}{2}+\frac{is}{2\pi}}\right] \,. \tag{8.70}$$

Making the change of variables $s \to s \pm i\pi(1-\varepsilon)$ we obtain the alternate representations

$$\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2}S(\rho(\epsilon)||\rho_0)\Big|_{\epsilon=0} = -\frac{1}{2}\int_{-\infty}^{\infty}\mathrm{d}s\,\frac{1}{4\sinh^2(\frac{s\pm i\epsilon}{2})}\mathrm{Tr}\left[\rho_0^{-1}\delta\rho\rho_0^{\pm\frac{is}{2\pi}}\delta\rho\rho_0^{\pm\frac{is}{2\pi}}\right]\,,\tag{8.71}$$

where the cyclicity of the trace has been used to summarize both substitutions into one expression. The relevant hyperbolic identities used are $\cosh(s + i\pi) = -\cosh(s)$ and $\cosh(s) - 1 = 2\sinh^2(s/2)$. This form of the Fisher information will be used in the next section to make a connection with a particular bulk energy density sourced by bulk matter fields, a result first arrived at in [70].

8.4 Nonlinear Gravity from Fisher Information

In [71] it is established that the quantum Fisher information associated to a boundary region A is equivalent to the gravitational "canonical energy", which is associated to a bulk region anchored to A. We will introduce the canonical energy properly in section 8.4.1. In [63] it is shown that the equality of canonical energy and Fisher information to second order in perturbations about the CFT vacuum implies that the EFE must be satisfied to second order in perturbations about pure AdS. In this section the goal is to follow and recreate this result, originally obtained in [63].

To find the promised result, a particular class of excited CFT states on the boundary are considered, constructed via a path integral. In addition the two parameters of boundary CFT, a^* and \tilde{C} must be taken into account. For the explicit calculation the specific relation

$$\tilde{C} = \frac{\pi^d (d-1)}{\Gamma(d-2)} C = a^*$$
(8.72)



Figure 8.1: Near-boundary region of AdS in Fefferman-Graham coordinates. A is the ball shaped region under consideration on the boundary, with the associated RT surface \tilde{A} indicated in red. In light grey is the domain of dependence of A, $\mathcal{D}(A)$. In dark grey is the spacelike region bounded by A and \tilde{A} , Σ_A .

is necessary, where C is the central charge as determined by the normalization of the stress tensor two-point function and a^* is the universal $\mathcal{O}(1)$ term in the CFT entanglement entropy. a^* is the trace (or Weyl) anomaly present in CFTs on a background that is not conformally flat.

On the gravitational side, it is first shown that there exists a metric

$$G = G_{\text{AdS}} + \epsilon G^{(1)} + \epsilon^2 G^{(2)} \dots$$
(8.73)

that gives the right Ryu-Takayanagi area for the bulk extremal surfaces corresponding to all boundary balls A. Here, a lot of difficulty is introduced by the fact that we need to consider both variations in the shape of the extremal surface and the metric perturbation. This is solved by picking a "Hollands-Wald" gauge that cancels the deformation of the surface itself, imposing extra gauge conditions on the metric between the extremal surface \tilde{A} and the boundary region A. The geometry under consideration for all of this section is imaged in figure 8.1

The main gravitational result will be that the gravitational interpretations of $\langle H \rangle_{\sigma} = E_A$ and S_A , E^{grav} and S^{grav} fulfill the identity

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} \left(\delta E^{\mathrm{grav}} - \delta S^{\mathrm{grav}}\right) = \int_{\Sigma_A} \omega(g, \frac{\mathrm{d}g}{\mathrm{d}\epsilon}, \mathcal{L}_{\xi_A}g) + \int_{\Sigma_A} \mathcal{G}$$
(8.74)

where the objects in the integrand are to be defined later and Σ_A is the spacelike volume enclosed by S_A and A. In contrast to the first-order case E^{grav} is not just related to the asymptotic Fefferman-Graham perturbation, but contains contributions from bulk matter. S^{grav} is defined by the HRT formula using the second-order perturbed metric in equation (8.73).

Correspondingly on the CFT side we have the identity

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}(E_A - S_A) = \frac{\mathrm{d}}{\mathrm{d}\epsilon}(S(\sigma + \epsilon\rho||\sigma))$$
(8.75)

at first order in ϵ , as is clear from replacing $\Delta \rightarrow \epsilon$ in equation (8.58). We require the equivalence of the left hand sides of equations (8.74) and (8.75), and so the goal becomes to show that

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}(S(\sigma + \epsilon\rho||\sigma)) = \int_{\Sigma_A} \omega(g, \frac{\mathrm{d}g}{\mathrm{d}\epsilon}, \mathcal{L}_{\xi_A}g) + \int_{\Sigma_A} \mathcal{G}$$
(8.76)

implies the Einstein equations to second order. Applying a second d/d ϵ to the equation and letting $\epsilon \rightarrow 0$ makes the LHS the quantum Fisher information. Explicitly, the second order master equation becomes

$$\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} (S(\sigma + \epsilon\rho||\sigma))\Big|_{\epsilon=0} = \int_{\Sigma_A} \omega(g^{(0)}, \delta g^{(1)}, \mathcal{L}_{\xi_A} \delta g^{(1)}) - \int_{\Sigma_A} 2\xi_A^q E_{ab}^{(2)} \epsilon^b \tag{8.77}$$

where $E^{(2)}$ denotes the second order Einstein tensor. The first term on the RHS defines the *canonical energy* of the gravitational theory, defined as the quantity that is conserved under the modular time translation ξ_A generated by the modular Hamiltonian of the region Σ_A .

Relation to Previous First Order Calculation

Setting $\epsilon = 0$ in equation (8.76) extracts the first order in ϵ of $S(\sigma + \epsilon \rho || \sigma)$, but we know that the first derivative of the relative entropy vanishes. It turns out the righthand side becomes

$$\int_{\Sigma_A} \xi_A^a E_{ab}^{(1)} \boldsymbol{\epsilon}^b = 0 , \qquad (8.78)$$

where ϵ^a is a volume form, ξ^a_A is the timelike Killing vector field generated by the bulk modular Hamiltonian and $E^{(1)}_{ab}$ is the first order Einstein tensor. In this formalism, it is possible to perform an analog of the proof of section 8.2 by demanding that equation (8.78) holds for all balls in all frames showing that $E^{(1)}_{\mu\nu}$ must vanish. The meaning of these gravitational quantities will be made clear in the following sections.

8.4.1 Gravitational Side of the Derivation

In this subsection we define the Hollands-Wald gauge. We then introduce the covariant phase space formulation of general relativity that lets us write down the relevant gravitational identities. We also formulate the gravitational identity away from the Hollands-Wald gauge and include bulk matter field contributions. This section assumes some familiarity with general relativity beyond the level of the rest of this text.

Choice of Hollands-Wald Gauge

To make clear what the Hollands-Wald gauge is, we will work in Fefferman Graham coordinates with

$$ds^{2} = \frac{1}{z^{2}} \left[dz^{2} + h_{\mu\nu}(z, x^{\mu}) dx^{\mu} dx^{\nu} \right] , \qquad (8.79)$$

and pure AdS given by $h_{\mu\nu} = \eta_{\mu\nu}$

In a Poincaré patch of AdS, the bulk extremal surface \tilde{A} that computes the entanglement entropy of the boundary ball A is a half-sphere defined by

$$z^2 + (x - x_0)^2 = R^2 . aga{8.80}$$

There is a region Σ_A that is has $\tilde{A} \cup A$ as it's boundary. The domain of dependence of Σ_A represents a Rindler wedge of AdS. There exists a Killing vector ξ_A that vanishes at \tilde{A} given by

$$\xi_A = -\frac{2\pi}{R}(t-t_0) \left[zn_z + (x^i - x_0^i)n_i \right] + \frac{\pi}{R} [R^2 - z^2 - (t-t_0)^2 - (\vec{x} - \vec{x_0})^2] n_t , \qquad (8.81)$$

where n_{μ} is the unit covector in the μ direction. This corresponds to the vector pointing tangentially to the worldline of a Rindler observer with constant acceleration, and may be seen as the vector field generated by the bulk modular Hamiltonian of the AdS Rindler wedge defined by the domain of dependence of Σ_A .

Near the boundary of AdS, ξ_A approaches

$$\xi_A = -\frac{2\pi}{R}(t - t_0) \left[(x^i - x_0^i)n_i \right] + \frac{\pi}{R} [R^2 - (t - t_0)^2 - (\vec{x} - \vec{x_0})^2] n_t .$$
(8.82)

defining an asymptotic symmetry of AdS. This has to be a symmetry of the boundary CFT, so equation (8.82) will impose a relation between the boundary CFT and the bulk gravitational theory³.

Let us now consider instead the perturbed spacetime

$$G(\epsilon) = G_{\text{AdS}} + \epsilon G^{(1)} + \epsilon^2 G^{(2)} \dots$$
(8.83)

Let us denote by $\tilde{A}(\epsilon)$ the extremal surface associated to the boundary ball in this geometry. In section 7.2.2 the fact that \tilde{A} has fixed coordinates to first order in ϵ is implied by \tilde{A} being extremal in the unperturbed spacetime. It turns out that we can pick a gauge for the metric G such that the same coordinate embedding X for the extremal surface remains extremal to all orders in ϵ . The existence of such a gauge near extremal surfaces was proven by Hollands and Wald in the context of perturbations about a black hole background [72]. This allows us to put \tilde{A} at the same coordinates to second order for all metrics $G(\epsilon)$. In addition, remaining freedom in the choice of coordinates makes sure that ξ_A with the same coordinate description as in the unperturbed AdS continues to satisfy the Killing equations on \tilde{A} .

The explicit gauge conditions become

- that the variation of the area functional vanishes to first order in both X and G. This is equivalent to the vanishing of the trace of the extrinsic curvature to first order in ϵ ,
- The Killing equation $\nabla_{(\mu}(\xi_A)_{\nu)} = 0$ is satisfied on \tilde{A} ,

where both of the restrictions only need to hold on the surface \tilde{A} . The variation of the area functional is proportional to $\int \sqrt{-g} g^{\alpha\beta} \delta g_{\alpha\beta}$ where g is the induced metric on the extremal surface. Let us pick coordinates such that the extremal surface is parametrized by $\sigma^{\alpha} = X^{\alpha}$, $X^{\overline{\alpha}} = (\text{constant})$ where α and $\overline{\alpha}$ are tangential and orthogonal coordinates to the surface, respectively. The induced metric is

$$g_{\alpha\beta} = \frac{\partial X^{\mu}}{\partial \sigma^{\alpha}} \frac{\partial X^{\nu}}{\partial \sigma^{\beta}} G_{\mu\nu} = G_{\alpha\beta} , \qquad (8.84)$$

where we used our choice of coordinates. The variation of the area functional with respect to the bulk metric is given by:

$$\delta_G A[G, X] = \frac{1}{2} \int \sqrt{-g} g^{\alpha\beta} \delta g_{\alpha\beta} . \qquad (8.85)$$

Note that X does not depend on the metric perturbation because we are not re-solving for the extremal surface. The vanishing of the trace of the extrinsic curvature to first order in δG is equivalent to $\delta_X A[G + \delta G, X] = 0$ for all δ_X . We need only consider terms proportional to $\delta_X \delta_G[G,X]$ since it is given that the original surface was extremal. Finally, we may without loss of generality take the shape variations δX^{ρ} to be orthogonal to the extremal surface. We then have

$$\delta_X \delta_G A[G, X] = \frac{1}{2} \int \sqrt{-g} \left[\frac{1}{2} g^{\gamma \delta} \delta_X g_{\gamma \delta} g^{\alpha \beta} \delta g_{\alpha \beta} + \delta_X g^{\alpha \beta} \delta g_{\alpha \beta} + g^{\alpha \beta} \delta_X \delta g_{\alpha \beta} \right]$$
(8.86)

³This is essentially how the Brown-Henneaux paper finds the central charge of the boundary CFT in [57].

where unsubscripted δ are with respect to G. We will need to know three variations

$$\delta_X g_{\alpha\beta} = 2 \frac{\partial X^{\mu}}{\partial \sigma^{\alpha}} \frac{\partial (\delta X^{\nu})}{\partial \sigma^{\beta}} G_{\mu\nu} + \frac{\partial X^{\mu}}{\partial \sigma^{\alpha}} \frac{\partial X^{\nu}}{\partial \sigma^{\beta}} \frac{\partial G_{\mu\nu}}{\partial X^{\rho}} \delta X^{\rho}$$

$$= 2 \frac{\partial (\delta X^{\nu})}{\partial \sigma^{\beta}} G_{\alpha\nu} + \frac{\partial G_{\alpha\beta}}{\partial X^{\rho}} \delta X^{\rho} \equiv 2 \delta X^{\rho} K_{\rho;\alpha\beta} ,$$

$$\delta_X g^{\alpha\beta} = -g^{\alpha\gamma} g^{\beta\delta} \delta_X g_{\gamma\delta} ,$$

$$\delta_X \delta g_{\alpha\beta} = \left[2 \frac{\partial X^{\mu}}{\partial \sigma^{\alpha}} \frac{\partial (\delta X^{\nu})}{\partial \sigma^{\beta}} + \frac{\partial X^{\mu}}{\partial \sigma^{\alpha}} \frac{\partial X^{\nu}}{\partial \sigma^{\beta}} X^{\rho} \frac{\partial}{\partial X^{\rho}} \right] \delta G_{\mu\nu}$$

$$= 2 \frac{\partial (\delta X^{\nu})}{\partial \sigma^{\beta}} \delta G_{\alpha\nu} + \frac{\partial \delta G_{\alpha\beta}}{\partial X^{\rho}} \delta X^{\rho} ,$$
(8.87)

where in the first variation we have used that the shape variation defines the extrinsic curvature⁴. In the following, let us use the notation ∂_{ρ} for the ρ derivative. We then have

$$\delta_X \delta_G A[G, X] = \frac{1}{2} \int \sqrt{-g} \delta X^{\rho} G^{\gamma \delta} K_{\rho;\gamma \delta} g^{\alpha \beta} \delta G_{\alpha \beta} - \int \sqrt{-g} \delta X^{\rho} K^{\alpha \beta}_{\rho} \delta G_{\alpha \beta} + \frac{1}{2} \int \sqrt{-g} \delta X^{\rho} G^{\alpha \beta} \partial_{\rho} \delta G_{\alpha \beta} + \int \sqrt{-g} G^{\alpha \beta} G_{\rho \beta} \partial_{\alpha} \delta X^{\rho} .$$
(8.88)

The first term is zero because the original surface was extremal, meaning the trace of the extrinsic curvature vanishes. Knowing that the final expression is to be covariant, the only possible result is

$$\delta_X \delta_G A[G, X] = \int \sqrt{-g} \left(\frac{1}{2} \delta X^{\overline{\alpha}} \nabla_{\overline{\alpha}} \delta G^{\alpha}{}_{\alpha} + G^{\alpha\beta} \delta G_{\overline{\alpha}\alpha} \nabla_{\beta} \delta X^{\overline{\alpha}} \right) = \int \sqrt{-g} \left(\frac{1}{2} \nabla_{\overline{\alpha}} \delta G^{\alpha}{}_{\alpha} - G^{\alpha\beta} \nabla_{\beta} \delta G_{\overline{\alpha}\alpha} \right) \delta X^{\overline{\alpha}} ,$$
(8.89)

where in the partial integration we have required that the leading asymptotic of δG vanishes at the boundary, equivalent to demanding that the spacetime remain asymptotically AdS.

Following the convention of Raamsdonk et al. in [63] the vanishing of the expression in parentheses gives us the explicit differential gauge condition

$$\left(\nabla^{(0)}_{\alpha}\delta G^{\alpha}{}_{\overline{\alpha}} - \frac{1}{2}\nabla^{(0)}_{\overline{\alpha}}\delta G^{\alpha}{}_{\alpha}\right)_{\tilde{A}} = 0 \ . \tag{8.90}$$

The explicit first-order condition to preserve the Killing equation is [63]

$$\delta G^{\alpha}{}_{\overline{\alpha}} = 0 \ , \tag{8.91}$$

$$\delta G^{\overline{\alpha}}{}_{\beta} - \frac{1}{2} \delta^{\overline{\alpha}}{}_{\beta} \delta G^{\overline{\alpha}}{}_{\overline{\beta}} = 0 \tag{8.92}$$

Gravitational Side of the Derivation Using Covariant Phase Space

We will now derive the gravitational identity in equation (8.74), using the covariant phase space formalism first introduced by Lee and Wald in [73]. In principle this formalism may be applied to the linear order case as well, but the proof carried out in section 8.2 serves as an elementary introduction to the main idea of implying the field equations through the Ryu-Takayanagi relation.

The main point of this formalism is that it guarantees the existence of a differential form χ such that $\int_{\Sigma_A} d\chi = \int_A \chi - \int_{\tilde{A}} \chi = \delta_{\epsilon} (E^{\text{grav}} - S^{\text{grav}})$. This means that restrictions on the boundary

 $^{^{4}}$ This is of course exactly why extremality under shape deformations is equivalent to vanishing of the extrinsic curvature, clear by inserting the shape variation into equation (8.85).

integrals over A, \tilde{A} that determine $E^{\text{grav}}, S^{\text{grav}}$ can be transformed via Stokes' theorem into differential restrictions in all of Σ_A , similarly to the relation between integral and differential forms of the Maxwell equations. This is important because it translates intractable nonlocal restrictions to local differential restrictions.

Let us assume that our theory is determined by a generally covariant Lagrangian density, i.e. a Lagrangian describing gravity plus additional fields. We can write this Lagrangian as the integral of a differential form L

$$L = \mathcal{L}\boldsymbol{\epsilon} \;, \tag{8.93}$$

where $\boldsymbol{\epsilon}$ is the usual volume form

$$\boldsymbol{\epsilon} = \frac{1}{(d+1)!} \sqrt{-g} \epsilon_{a_1 \dots a_{d+1}} \mathrm{d} x^{a_1} \wedge \dots \wedge x^{a_{d+1}} .$$
(8.94)

It will also be useful to define the induced volume form on a hypersurface of codimension k, given by

$$\epsilon_{a_1...a_k} = \frac{1}{(d+1-k)!} \sqrt{-g} \epsilon_{a_1...a_k b_{k+1}...b_{d+1}} \mathrm{d}x^{b_{k+1}} \wedge \ldots \wedge x^{b_{d+1}}$$
(8.95)

where the *a*'s are normal indices and the *b*'s are tangential indices of the hypersurface. To compute the volume of a hypersurface Σ one need only contract with a set of unit vectors in the space orthogonal to Σ and integrate $V_{\Sigma} = \int \epsilon_{a_1...a_k} n^{a_1} \dots n^{a_k}$. Since the free indices of the induced volume form sum only over directions orthogonal to the hypersurface, it is zero except when fed a set of linearly independent unit normals.

Here, we take the Lagrangian to depend on the metric G as well as potential matter fields ϕ_{α} , i.e. $\mathcal{L} = \mathcal{L}(G, \phi_{\alpha})$ The variation of the Lagrangian is the equations of motion plus an eventual boundary term. More explicitly we may write

$$\delta(\mathcal{L}\boldsymbol{\epsilon}) = -E^{ab}\delta G_{ab}\boldsymbol{\epsilon} - E^{\alpha}_{\phi}\delta\phi_{\alpha}\boldsymbol{\epsilon} + \mathrm{d}\boldsymbol{\theta}(g,\delta g,\phi_{\alpha},\delta\phi_{\alpha}) \ . \tag{8.96}$$

Here, $E^{ab} = 0$ defines the equations of motion for the metric G, and $E^{\alpha}_{\phi} = 0$ define the equations of motion for the matter fields. The $d\theta$ term is the total derivative term that may appear when integrating by parts to remove the $\delta \partial \phi$ type terms while deriving the equations of motion.

The basic ingredient of the covariant phase space formulation of gravity is a "symplectic current two-form" W. In a general classical system, dynamics can be seen as a flow in phase space⁵ governed by Hamilton's equations $\dot{X} = \xi_H$ where ξ is a vector field determined by

$$\xi_H \cdot W = \mathrm{d}H \ . \tag{8.97}$$

Here by \cdot we mean a contraction, and on the righthand side H is a Hamiltonian function, i.e. a 0-form. Being more explicit, ξ_H is determined by

$$\xi^{\mu}_{H}W_{\mu\nu}\mathrm{d}x^{\nu} = \partial_{\nu}H\mathrm{d}x^{\nu} . \tag{8.98}$$

The symplectic form takes as arguments any two perturbations $(\delta G_1, \delta \phi_1)$ and $(\delta G_2, \delta \phi_2)$ and is expressed directly in terms of the form $\boldsymbol{\theta}$ as the integral over a Cauchy surface of

$$\omega(G;\delta G_1,\delta\phi_1,\delta G_2,\delta\phi_2) = \delta_1 \boldsymbol{\theta}(G,\phi,\delta G_2,\delta\phi_2) - \delta_2 \boldsymbol{\theta}(G,\phi,\delta G_1,\delta\phi_1) .$$
(8.99)

⁵Phase space is constructed by using the fields and their derivatives as an orthogonal basis. A given initial condition is a point in this space, and it's time evolution traces out a trajectory. This space can be geometrized because it admits a natural nondegenerate symplectic two-form to act like a metric.

We can consider the metric transformation $g \to g + \mathcal{L}_X g$ where $\mathcal{L}_X = \nabla_a X_b + \nabla_b X_a$ is the Lie derivative. This transformation induces a particular flow on phase space which can be seen as generated by a phase space Hamiltonian H_X . Specifically we define the phase space Hamiltonian by demanding that for any other deformation,

$$\delta H_X = W(\delta G, \delta_\phi, \mathcal{L}_X G) . \tag{8.100}$$

The differential form θ is also associated with Noether's theorem. Since we consider a gravitational theory we have diffeomorphism invariance. Therefore, the transformation $\delta x = X$ is a symmetry for any X. Then there exists a conserved current that can be found by the usual Noether procedure that in terms of the present notation takes the form

$$J_X = \boldsymbol{\theta}(\delta_X G) - X \cdot L , \qquad (8.101)$$

where the conservation equation takes the (on-shell) form

$$dJ_X = 0$$
 . (8.102)

Because this holds for all X, J_X is an exact form up to terms that vanish on-shell, and can therefore be written as

$$J_X = \mathrm{d}Q_X + X^a C_a \tag{8.103}$$

where Q is some differential form and C_a are quantities that vanish on-shell. By picking $X = \xi_A$ and picking the Cauchy surface that we integrate over to be Σ_A , the conserved current is related to the modular Hamiltonian of the AdS-Rindler wedge. Let us now apply this formalism to the present scenario with gravity and a scalar field.

The main result (equation (8.74)) follows from the fact that for any metric g and vector field X, ω applied to the first order perturbations $\delta g, \delta \phi_{\alpha}$ as well as the Lie derivatives

$$(\mathcal{L}_X G)_{ab} = \nabla_a X_b + \nabla_b X_a , \quad \mathcal{L}_X \phi = X^a \nabla_a \phi \tag{8.104}$$

is a total derivative up to a term \mathcal{G} that vanishes on shell. That is, we can write

$$\omega(G;\delta G,\delta \phi, \mathcal{L}_X G, \mathcal{L}_X \phi) = \mathrm{d}\chi(G, \phi, \delta G, \delta \phi, X) - \mathcal{G}(G, \phi, \delta G, \delta \phi, X) .$$
(8.105)

where χ is related to the previous discussion in that $\chi(G,\phi) = \delta Q_X(G) + X \cdot \theta(G,\delta G)$. This is true under very general considerations [72], but we will instead explicitly consider gravity with a cosmological constant coupled only via the stress tensor to matter, meaning the gravitational equations of motion will be

$$E^{ab} \equiv E^{ab}_{\text{grav}}(G) - \frac{1}{2}T^{ab}(g,\phi_{\alpha})$$
 (8.106)

If we assumed Einstein gravity with no stringy higher derivative corrections, $E_{\text{grav}}^{ab}(G)$ would be the Einstein tensor. Let us assume that we have the equations of motion of a free scalar field plus Einstein gravity

$$E_{\phi} = \left(\nabla_a \nabla^a - m^2\right)\phi,$$

$$E_{ab} = \frac{1}{16\pi} \left(R_{ab} - \frac{1}{2}G_{ab}R + G_{ab}\Lambda\right)\phi - \frac{1}{2}T_{ab}(\phi),$$

$$T_{ab}(\phi) = \left(\nabla_a \phi \nabla_b \phi - \frac{1}{2}g_{ab}(\nabla_c \phi \nabla^c + m^2 \phi^2)\right).$$

(8.107)

where all of the ∇ s here and in the following equations are built from the unperturbed metric. When deriving these equations of motion from the appropriate Lagrangian, one finds the boundary terms

$$\boldsymbol{\theta}_{\phi}(\phi,\delta\phi) = \boldsymbol{\epsilon}_{a}\phi\nabla^{a}\delta\phi ,$$

$$\boldsymbol{\theta}_{\text{grav}}(G,\delta G) = \frac{1}{16\pi}\boldsymbol{\epsilon}_{a}\left(G^{ac}G^{bd} - G^{ad}G^{bc}\right)\nabla_{d}\delta G_{bc} .$$
(8.108)

where ϵ is the induced volume form on the boundary. With these definitions we can evaluate

$$\omega_{\phi}(\phi;\delta_{1}\phi,\delta_{2}\phi) = \boldsymbol{\epsilon}_{a} \left(\delta_{1}\phi\nabla^{a}\delta_{2}\phi - \delta_{2}\phi\nabla^{a}\delta_{1}\phi\right) , \qquad (8.109)$$

where the $\delta_1 \delta_2$ terms cancel since variations commute. After considerably more algebra, we have

$$\omega_{\text{grav}}(G,\delta_1 G,\delta_2 G) = \frac{1}{16\pi} P^{abcdef} \left(\delta_2 G_{bc} \nabla_d \delta_1 G_{ef} - \delta_1 G_{bc} \nabla_d \delta_2 G_{ef} \right) , \qquad (8.110)$$

where we have defined

$$P^{abcdef} = G^{ae}G^{bf}G^{cd} - \frac{1}{2}\left(G^{ad}G^{be}G^{ef} + G^{ab}G^{cd}G^{ef} + G^{ae}G^{bc}G^{fd} - G^{ad}G^{bc}G^{ef}\right) .$$
(8.111)

We will now state the form of χ , noting that it is straightforward to check that the following χ and \mathcal{G} satisfy equation (8.105)

$$\chi(G,\delta G,X) = \frac{1}{16\pi} \epsilon_{ab} \left(\delta G^{ac} \nabla_c X^b - \frac{1}{2} \delta G_c^{\ c} \nabla^a X^b + X^c \nabla^b \delta G^a_{\ b} - X^b \nabla_c + X^b \nabla^a \delta G_c^{\ c} + 16\pi X^b \delta \phi \nabla^a \phi \right)$$
$$\mathcal{G}(G,\delta_\epsilon G) = X^c \epsilon_c E^{ab} \delta_\epsilon g_{ab} + X^c \epsilon_c E^\phi \delta_\epsilon \phi - 2X^a \delta_\epsilon \left(E_{ab} \epsilon^b \right)$$
(8.112)

To find our geometric identity (equation (8.74)) we apply equation (8.105) to the case where the metric is our family of perturbed metrics about AdS $G(\epsilon)$ so that $\delta G = \partial_{\epsilon} G|_{\epsilon=0}$ and $X = \xi_A$. We then integrate over the surface Σ_A bounded by A and \tilde{A} . It follows from Stokes' theorem that

$$\int_{A} \chi - \int_{\tilde{A}} \chi = \int_{\Sigma_{A}} \omega + \int_{\Sigma_{A}} \mathcal{G} , \qquad (8.113)$$

where the minus sign is bookkeeping the orientation of the surfaces involved. The first integral is related to the variation of the area functional. Using that the surface is extremal the coordinate variations vanish and we have

$$\delta_{\epsilon} \operatorname{Area} = \frac{\delta A rea}{\delta G} \delta_{\epsilon} G = \frac{1}{2} \int_{\tilde{A}} \epsilon_{ab} n^{[a} n^{b]} G^{cd} \, \delta_{\epsilon} G_{cd}$$
(8.114)

where we have used $\det(G) = e^{\operatorname{Tr}[\ln(G)]}$ and that the volume form contains the determinant of the metric. Since at \tilde{A} the Killing vector ξ_A vanishes, we see that

$$\chi = \frac{1}{16\pi} \epsilon_{ab} \left(\delta G^{ac} \nabla_c \xi^b - \frac{1}{2} \delta G_c{}^c G^{ad} \nabla_d \xi^b \right) .$$
(8.115)

Using that $\xi_A^a = 0$ we only need the partial derivative part of the covariant derivative, finding that

$$\nabla_a \xi^b = \partial_a \xi^b f = -2\pi \left(G_{ac} n^{(c} n^{b)} - \delta_a{}^b \right) , \qquad (8.116)$$

where the symmetrization has been normalized. Since the volume form is antisymmetric in a and b, the second term in equation (8.115) is killed by contraction. What survives is

$$\chi = -\frac{1}{8} \epsilon_{ab} \left(G_{cd} n^{(d} n^{b)} \right) \delta G^{ac} .$$
(8.117)

Next we combine antisymmetry in a, b and symmetry in the raised a, c indices to exchange a and d, obtaining

$$\chi = \frac{1}{8} \epsilon_{ab} G_{cd} n^{[a} n^{b]} \delta G^{cd} . \qquad (8.118)$$

Upon integration we find that

$$\int_{\tilde{A}} \chi = \frac{1}{8} \int_{\tilde{A}} \epsilon_{ab} n^{[a} n^{b]} G_{cd} \delta G^{cd} = \frac{1}{4} \delta_{\epsilon} \operatorname{Area}[\tilde{A}] \equiv \delta_{\epsilon} S^{\operatorname{grav}} , \qquad (8.119)$$

where the final definition is the assumption of the Ryu-Takayanagi relation.

Just like in the first order case, the gravitational interpretation of the energy variation should be an integral over A of the boundary metric up to some forefactor. More precisely, using the holographic stress tensor relation equation (6.136) we wrote down in the first order case equation (8.6) that

$$\delta E^{\text{grav}} = \langle H_A \rangle = \frac{dL^{d-3}}{16G_N} \int_A d^{d-1}x \frac{R^2 - r^2}{R} \delta G_{00}(z=0) .$$
(8.120)

This can be written in a more covariant form as

$$\frac{dL^{d-3}}{16\pi G_N} \int_A \varepsilon^a \xi^b(z=0) \delta G_{ab}(z=0) , \qquad (8.121)$$

where ε is an induced volume form with respect to the boundary space, constructed $\sqrt{G(z=0)}$. Note that $\varepsilon \neq \varepsilon$

This is also given by the integral of χ , but this time over A. Introducing the appropriate dimensionful constants we have [63]

$$\int_{A} \chi = \frac{dL^{d-3}}{16\pi} \delta_{\epsilon} \left(\int_{A} \zeta_{A}^{a} g_{ab}^{(2)} \boldsymbol{\varepsilon}^{b} \right) \equiv \delta E^{\text{grav}} , \qquad (8.122)$$

where ζ_A^a is the boundary value of ξ_A^a and $g_{ab}^{(d-2)}$ is the part of the asymptotic metric *h* at order z^2 , describing the finite part of the boundary metric.

We have thus shown the key gravitational identity, since equation (8.113) now implies

$$\delta_{\epsilon} \left(E^{\text{grav}} - S^{\text{grav}} \right) = \int_{\Sigma_A} \omega_{\text{grav}}(G, \delta_{\epsilon} G, \mathcal{L}_{\xi_A} G) + \omega_{\phi}(\phi, \delta_{\epsilon} \phi, \mathcal{L}_{\xi_A} \phi) + \int_{\Sigma_A} \mathcal{G} . \tag{8.123}$$

We also wish to obtain a perturbative identity to second order in ϵ . Imposing that the first order perturbations solve the field equations, this is given by

$$\delta_{\epsilon}^{(2)}(E^{\text{grav}} - S^{\text{grav}}) = \int_{\Sigma_A} \left[\omega_{\text{grav}}(\delta G, \mathcal{L}_{\xi_A} \delta G) + \omega_{\phi}(\delta \phi, \mathcal{L}_{\xi_A} \delta \phi) - 2\xi_A^a \left(\delta_{\epsilon}^{(2)} E_{ab} \right) \boldsymbol{\epsilon}^b \right] , \qquad (8.124)$$

assuming that the Hollands-Wald gauge conditions are satisfied and given that $\mathcal{L}_{\xi_A} G^{(0)} = 0$. Here $E_{ab}^{(2)}$ is the second order Einstein equations coupled to matter, given by

$$E_{ab}^{(2)} = \left(\frac{\delta^2 E_{ab}}{\delta G_{cd} \delta G_{ef}} \delta G_{cf} \delta G_{ef} + \frac{\delta E_{ab}}{\delta G_{cd}} \delta^2 G_{cd}\right) + \frac{\delta^2 (E_{ab})_{\phi}}{\delta^2 \phi} \delta \phi \delta \phi .$$
(8.125)

Gravitational Identity Away From Hollands-Wald Gauge

The second order result that we wrote down equation (8.74) should be equal to the Fisher information of the region A in the dual field theory. When comparing to the field theory computation of the Fisher information, it is convenient to write down the second order object away from the Hollands-Wald gauge. This is because the field theory calculation is best carried out in de Donder gauge.

The basic ingredient is that a general first order coordinate transformation $x^a \to x^a + V^a$ may transform a general first order metric perturbation h into another perturbation γ that satisfies the Hollands-Wald gauge conditions. The perturbations are related by

$$\gamma = h + \mathcal{L}_V G , \qquad (8.126)$$

where G is the full metric. Given h and that γ has to satisfy the Hollands-Wald gauge, the vector field V is fully constrained. Since diffeomorphisms are symmetries, the results about the transformation properties of ω still hold. Using linearity of $\omega_{\text{grav}}(\gamma, \mathcal{L}_{\xi_A}\gamma)$ in the listed arguments we have

$$\omega_{\text{grav}}(\gamma, \mathcal{L}_{\xi_A}\gamma) = \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A}h) + \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A}\mathcal{L}_V G) + \omega_{\text{grav}}(\mathcal{L}_V G, \mathcal{L}_{\xi_A}(h + \mathcal{L}_V G)) .$$
(8.127)

When one of the arguments of ω_{grav} is pure gauge (i.e. a Lie derivative of the full metric G), and the other satisfies the linearized Einstein equations we know that $\omega_{\text{grav}}(\delta G, \mathcal{L}_X G) = d\chi(\delta G, X)$ is a total derivative. Then we can reexpress equation (8.127) as

$$\omega_{\text{grav}}(\gamma, \mathcal{L}_{\xi_A}\gamma) = \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A}h) + d\chi(h, [\xi_A, V]) - d\chi(\mathcal{L}_{\xi_A}(h + \mathcal{L}_V G), V) , \qquad (8.128)$$

where we have used that ω is symmetric in it's arguments and defined

$$[\xi_A, V]^a = \xi^b_A \partial_b V^a - V^b \partial_b \xi^a_A , \qquad (8.129)$$

using that $\mathcal{L}_{\xi_A}G = 0$ to write $\mathcal{L}_{\xi_A}\mathcal{L}_VG = [\mathcal{L}_{\xi_A}, \mathcal{L}_V]G = \mathcal{L}_{[\xi_A,V]}G$. Using that the quantity $h + \mathcal{L}_VG$ satisfies the Hollands-Wald gauge condition, $\chi(\mathcal{L}_{\xi_A}(h + \mathcal{L}_VG), V)$ vanishes on \tilde{A} . To show this, let us abbreviate $\dot{\gamma} \equiv \mathcal{L}_{\xi_A}(h + \mathcal{L}_VG)$. The perturbation $h + \mathcal{L}_VG$ is in Hollands-Wald gauge, so $\dot{\gamma}$ vanishes on \tilde{A} . Then $\chi(\dot{\gamma}, V)$ simplifies to (see definition in equation (8.112))

$$\chi(\dot{\gamma}, V)_{\tilde{A}} = \boldsymbol{\epsilon}_{ab} \left(\nabla^{b} \dot{\gamma}^{a}{}_{c} V^{c} - \nabla_{c} \dot{\gamma}^{ac} V^{b} + \nabla^{a} \dot{\gamma}_{c}{}^{c} V^{b} \right)$$

$$= 2\boldsymbol{\epsilon}_{+-} \left(\nabla^{[-} \dot{\gamma}^{+]}{}_{c} V^{c} - \nabla_{c} \dot{\gamma}^{c[+} V^{-]} + \nabla^{[+|} \dot{\gamma}_{c}{}^{c} V^{|-]} \right) , \qquad (8.130)$$

where we have used that the unit normals to the RT surface can chosen to be lightcone directions. The goal is now to split this into indices parallel to the RT surface α and normal indices $\overline{\alpha}$. We can pick the part of V that is parallel to the RT surface to vanish on the surface, so that $V^{\mu}|_{\tilde{A}} = V^{\overline{\alpha}}$.

All of the covariant derivatives are with respect to the unperturbed metric, and are compatible with the unperturbed metric that raises and lowers indices. We can lower the antisymmetrized \pm indices and rearrange, keeping only the orthogonal components $V^{\overline{\alpha}}$

$$\chi(\dot{\gamma}, V)_{\tilde{A}} \sim \boldsymbol{\epsilon}_{+-} \left(\nabla_{[-} \dot{\gamma}_{+]\overline{\alpha}} V^{\overline{\alpha}} - \nabla_{c} \dot{\gamma}_{[+}^{c} V_{-]} + \nabla_{[+|} \dot{\gamma}_{c}^{c} V_{|-]} \right) = \boldsymbol{\epsilon}_{+-} \left(\nabla_{[-} \dot{\gamma}_{+]\overline{\alpha}} V^{\overline{\alpha}} - (\nabla_{\alpha} \dot{\gamma}^{\alpha}_{[+} + \nabla_{\overline{\alpha}} \dot{\gamma}^{\overline{\alpha}}_{[+}) V_{-]} + \nabla_{[+|} (\dot{\gamma}_{\alpha}{}^{\alpha} + \dot{\gamma}_{\overline{\alpha}}{}^{\overline{\alpha}}) V_{|-]} \right) = \boldsymbol{\epsilon}_{+-} \left(\nabla_{[-} \dot{\gamma}_{+]\overline{\alpha}} V^{\overline{\alpha}} - \nabla_{\overline{\alpha}} \dot{\gamma}^{\overline{\alpha}}_{[+} V_{-]} + \nabla_{[+|} \dot{\gamma}_{\overline{\alpha}}{}^{\overline{\alpha}} V_{|-]} \right) = 0 , \qquad (8.131)$$

where in the penultimate step we used that $\nabla_{\pm}\dot{\gamma}^{\alpha}{}_{\alpha}|_{\tilde{A}} = \nabla_{\alpha}\dot{\gamma}^{\alpha}{}_{\pm}|_{\tilde{A}} = 0$ (coming from the fact that $\dot{\gamma}$ vanishes on \tilde{A} by virtue of γ being in Hollands-Wald gauge) and the last step can be checked by summing over $\overline{\alpha} = \{+, -\}$, explicitly writing out the antisymmetrizations, using the symmetry of the metric, metric compatibility of the ∇ to simultaneously raise/lower equal indices and explicitly finding the cancellations⁶.

Picking V to vanish sufficiently fast as we approach z = 0, $\chi(\mathcal{L}_{\xi_A}(h + \mathcal{L}_V G), V)$ also vanishes on A and we find that

$$\int_{\Sigma_A} \mathrm{d}\chi(\mathcal{L}_{\xi_A}(h + \mathcal{L}_V G), V) = 0 .$$
(8.132)

Then, the integral over Σ_A of $\omega_{\text{grav}}(\gamma, \mathcal{L}_{\xi_A}, \gamma)$ can be written as

$$\int_{\Sigma_A} \omega_{\text{grav}}(\gamma, \mathcal{L}_{\xi_A}, \gamma) = \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A}h) + \int_{\tilde{A}} \chi(h, [\xi_A, V]) .$$
(8.133)

Inserting the definition $\gamma = h + \mathcal{L}_V g = h + \nabla_{(a} V_{b)}$ into the Hollands-Wald gauge one obtains by straightforward calculation an explicit condition on the vector field V in terms of h, given by:

$$\left(\nabla_{\alpha} \nabla^{\alpha} V_{\overline{\alpha}} + [\nabla_{\alpha}, \nabla_{\overline{\alpha}}] V^{\alpha} \right)_{\tilde{A}} = - \left(\nabla_{\alpha} h^{\alpha}{}_{\overline{\alpha}} - \frac{1}{2} \nabla_{\overline{\alpha}} h^{\alpha}{}_{\alpha} \right)_{\tilde{A}}$$

$$\mathcal{L}_{V} g^{(0)}|_{\tilde{A}} = -\mathcal{L}_{\xi_{A}} h|_{\tilde{A}} .$$

$$(8.134)$$

The first of these is an inhomogenous Laplace equation in V, and can be solved by typical Green's function methods, while the second equation determines the first derivative of V away from the RT surface.

With this, the second order gravitational result in a general gauge is

$$\delta^2(E_A^{\text{grav}} - S_A^{\text{grav}}) = \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \omega_\phi(\delta_\epsilon \phi, \mathcal{L}_{\xi_A} \phi) - 2\xi_A^a \left(\delta_\epsilon^{(2)} E_{ab}\right) \epsilon^b + \int_{\tilde{A}} \chi(h, [\xi_A, V]) \ . \ (8.135)$$

This is the form of the gravitational identity that is most readily related to the CFT result that we will now derive.

8.4.2 CFT Side of the Derivation

The CFT side of the derivation is made significantly more complicated than the first order case by the fact that the Fisher information depends directly on the state perturbation in the field theory. To accommodate this novel ingredient, this section proceeds in a number of distinct steps:

- 1. First we construct a class of CFT perturbations that is distinctly 'classical' in the sense that they represent coherent states.
- 2. Given this class of state perturbations, we are able to rewrite the Fisher information as an integral of a two-point correlation function.
- 3. By using explicit maps from CFT correlation functions to Witten diagrams similarly to section 6.2.1 we can recast the correlation functions in terms of Hollands-Wald like formalism in an auxiliary AdS spacetime. We will specifically specifically consider the two-point functions of scalar fields and the stress energy tensor.

 $^{^{6}}$ This computation has been relegated to appendix A.1.

4. After a long journey of pushing integration contours around and applying the Residue theorem, we arrive at a CFT result that may be written as

$$\delta_{\epsilon}^2 S(\rho_A || \rho_A^{(0)}) = \frac{\tilde{C}_T}{a^*} \left[\int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \int_{\tilde{A}} \chi(h, [\xi_A, V)] \right] + \int_{\Sigma_A} \omega_{\phi}(h, \mathcal{L}_{\xi_A} \phi)$$
(8.136)

where V is a vector field taking us from de Donder gauge to the Hollands-Wald gauge. The introduction of a de Donder gauge is required for us to be able to integrate stress-tensor two-point functions. As can be observed, this only differs from the gravitational result (equation (8.135)) by the Einstein tensor. This means that demanding equality between the CFT and gravitational results for boundary balls in all asymptotic Lorentz frames will imply the vanishing of the EFE to second order.

Sourcing Classical Bulk States

As we mentioned in section 7.4, there should not be a clear one-to-one correspondence between CFT states and classical geometries. Because of this, to source perturbations to the CFT vacuum that may conceivably have a classical geometric dual, it turns out that it is good to stick to 'classical-like' quantum states, i.e. coherent states.

The CFT vacuum can be constructed as an integral over the Euclidean half-space $\tau < 0$, where τ is the imaginary component of the Lorentzian time. We then consider a one parameter family of operators $\lambda_{\alpha}(\epsilon)\mathcal{O}_{\alpha}$ inserted into the path integral as such

$$|\psi_{\lambda}(\epsilon)\rangle = \int \mathcal{D}\phi e^{-\int_{-\infty}^{0} \mathrm{d}\tau \int \mathrm{d}^{d-1}x(\mathcal{L}_{CFT}[\phi] + \lambda_{\alpha}(x;\epsilon)\mathcal{O}_{\alpha}(x))} , \qquad (8.137)$$

where $\lambda_{\alpha}(x; \epsilon) = \epsilon \lambda_{\alpha}(x) + \mathcal{O}(\epsilon^2).$

This is similar to how background fields were turned on in string theory. Some care has to be taken so that we do not define perturbations with infinite energy, which is achieved by demanding that the λ_{α} vanish sufficiently fast as we approach $\tau = 0$. In [74] it is shown that under a coherent perturbation the entanglement entropy of spherical regions may be computed as a perturbative series in the coupling λ , provided that λ vanishes at $\tau = 0$.

The sources λ_{α} can be thought of as boundary conditions for a Euclidean asymptotically AdS spacetime. Specifically, the correlation function $\langle \psi_{\lambda}(\epsilon) | \psi_{\lambda}(\epsilon) \rangle$ may be interpreted as a path integral over two τ half-planes glued together across $\tau = 0$. The gravitational field equations can be solved in the bulk asymptotically AdS space that has the full path integration region as its boundary, with the sources λ_{α} determining the asymptotic values of bulk fields. The $\tau > 0$ portion has complex conjugated sources by virtue of coming from the bra. We can then pick out the $\tau = 0$ slice as the initial condition for a Lorentzian spacetime, setting $\tau = 0$ and analytically continuing according to $(\phi_{\alpha}, \partial_{\tau}\phi_{\alpha}) \rightarrow (\phi_{\alpha}, i\partial_{t}\phi_{\alpha})$.

To obtain the most general Lorentzian spacetimes it is necessary to let the sources be complex. In [75] an explicit understanding of the map between Euclidean sources and Lorentzian spacetimes has been found. In fact it seems that at the linear level any Lorentzian spacetime may be arbitrarily closely approximated by a good choice of Euclidean sources although highly localized states require large values of the source couplings, invalidating perturbation theory.

Relative Entropy from Two-Point Functions

The first nonzero contribution to the variation of the relative entropy under a perturbation of the state is given by the quantum Fisher information. It is given by

$$\delta_{\epsilon}^{(2)}S(\rho_A||\rho_A^{(0)}) = 2F(\delta\rho,\delta\rho) = -\int_{-\infty}^{\infty} \frac{\mathrm{d}s}{4\sinh^2\left(\frac{s\pm i\epsilon}{2}\right)} \mathrm{Tr}\left((\rho_A^{(0)})^{-1}\delta\rho(\rho_A^{(0)})^{\pm\frac{is}{2\pi}}\delta\rho(\rho_A^{(0)})^{\pm\frac{is}{2\pi}}\right), \tag{8.138}$$

where $F(\delta\rho, \delta\rho)$ is the quantum Fisher information. By finding an explicit CFT expression for $\delta\rho$, we will see that this quantity reduces to an integral of two-point correlation functions for conformal primary operators \mathcal{O}_{α} .

We want an expression for $\delta \rho$ in terms of ϵ . First we want to construct the local density operator ρ_A by gluing the Euclidean path integration region across A, and demanding that $\lambda_{\alpha}(\tau, x) \equiv \lambda_{\alpha}(-\tau, x)$ to implement that we are constructing the density operator out of an excited state and its complex conjugate.

$$\langle \phi^- | \rho_A | \phi^+ \rangle = \frac{1}{N_\lambda} \int_{\phi(A^-) = \phi^-}^{\phi(A^+) = \phi^+} \mathcal{D}\phi e^{-\int_{-\infty}^{\infty} \mathrm{d}\tau \int \mathrm{d}^{d-1} x (\mathcal{L}_{CFT}[\phi] + \lambda_\alpha(x;\epsilon) \mathcal{O}_\alpha(x))}$$
(8.139)

where A^+ and A^- denotes the region A approached from the euclidean future and past respectively, the states ϕ^-, ϕ^+ are arbitrary local states that serve as a notational auxiliaries, appearing as boundary conditions for the path integral.

To construct the perturbation of the density operator ρ we use that the local density operator can be expanded as a formal series in $\epsilon \lambda$ according to

$$\rho_A = \rho_A^{(0)} + \epsilon \int \mathrm{d}^d x \lambda_\alpha(x) \rho_A^{(0)} \mathcal{O}_\alpha(x) + \mathcal{O}(\epsilon^2) \ . \tag{8.140}$$

Here, note that we have taken x to also include the time direction and we have turned \mathcal{O}_{α} into a time-dependent operator. To define the time dependent operator we need to introduce the vacuum modular Hamiltonian $H_A = -\ln(\rho_A^{(0)})$. Euclidean evolution with the operator $H_A/2\pi$ generates the modular time translation vector field

$$\zeta_A = -\frac{1}{R}\tau x^i n_i + \frac{1}{2R}(R^2 + \tau^2 - x_i x^i)n_\tau , \qquad (8.141)$$

where the ns denote unit vectors and the index i runs only over spatial indices. Note that this differs from the case in equation (8.81) because we have yet to continue to Lorentzian signature, and we have divided by 2π . We then define the time dependent operator \mathcal{O} according to

$$\mathcal{O}_{\alpha}(x) \equiv \mathcal{O}_{\alpha}(\tau, \vec{x}) \equiv e^{\tau \frac{H_A}{2\pi}} \mathcal{O}_{\alpha}(0, \vec{x}) e^{-\tau \frac{H_A}{2\pi}} \frac{\Omega^{\Delta}(\tau, \vec{x})}{\Omega^{\Delta}(0, \vec{x})} .$$
(8.142)

Since the modular flow generates a conformal transformation, we must include a conformal factor to make sure our time dependent conformal primary continues to transform as a primary. $\Omega(\tau, \vec{x})$ is the conformal factor that relates a choice of boundary metric to the flat case according to $g_{\mu\nu} = \Omega^2 \eta_{\mu\nu}$.

This conformal factor does not do anything when we are working in a Minkowski representation of the boundary space, but in the present case it is practical to make a conformal transformation taking the boundary space from \mathbb{R}^d to $S^1 \times \mathbb{H}^{d-1}$. Before making the change of coordinates, let us write out the explicit form of the second order metric variation given $\delta \rho$ as read off from equation (8.140):

$$\delta_{\epsilon}^{(2)}S(\rho_{A}||\rho_{A}^{(0)}) = 2F\left(\int d^{d}x_{a}\lambda_{\alpha}(x_{a})\rho_{A}^{(0)}\mathcal{O}_{\alpha}(x_{a}), \int d^{d}x_{b}\lambda_{\alpha}(x_{b})\rho_{A}^{(0)}\mathcal{O}_{\alpha}(x_{b})\right)$$

$$= -\int d^{d}x_{a}\int d^{d}x_{b}\lambda_{\alpha}(x_{a})\lambda_{\alpha}(x_{b})\int \frac{ds}{4\sinh^{2}\left(\frac{s\pm i\epsilon}{2}\right)}\operatorname{Tr}\left(\mathcal{O}_{\alpha}(x_{a})\rho_{A}^{(0)}(\rho_{A}^{(0)})^{\pm\frac{is}{2\pi}}\mathcal{O}_{\alpha}(\tau_{b},\vec{x})(\rho_{A}^{(0)})^{\mp\frac{is}{2\pi}}\right)$$

$$= -\int d^{d}x_{a}\int d^{d}x_{b}\lambda_{\alpha}(x_{a})\lambda_{\alpha}(x_{b})\int \frac{ds}{4\sinh^{2}\left(\frac{s\pm i\epsilon}{2}\right)}\left(\frac{\Omega^{\Delta}(\tau_{b},\vec{x}_{b})}{\Omega^{\Delta}(\tau_{b}\pm is,\vec{x}_{b})}\right)^{\pm1}\operatorname{Tr}\left(e^{-H_{A}}\mathcal{O}_{\alpha}(x_{a})\mathcal{O}_{\alpha}(\tau_{b}\pm is,\vec{x})\right)$$

$$(8.143)$$

where we have used $e^{-H_A} = \rho_A^{(0)}$. Here it is useful to pick the sign of the \pm as positive when $\tau_a > \tau_b$ and as negative when $\tau_b > \tau_a$. We can then sum up the two τ integration regions with their respective choice of sign, yielding a time-ordered two-point function for the \mathcal{O}_{α} in the vacuum state $\rho_A^{(0)} = e^{-H_A}$. Specifically, when picking the negative sign we use symmetry in the arguments of $F(\delta\rho_1, \delta\rho_2)$ to exchange $x_a \leftrightarrow x_b$. Then we use cyclicity of the trace to move the $(\rho_A^{(0)})^{\frac{is}{2\pi}}$ factor on the right to the operator $\mathcal{O}(x_b)$ that now appears first in the trace. This cancels the sign in the time evolution, and we get equation (8.143) with the opposite ordering of the conformal primaries. A linear combination of correlation functions that switches ordering depending on whether $\tau_a > \tau_b$ is the definition of a time-ordered correlator. The obtained expression then becomes

$$\delta_{\epsilon}^{(2)}S(\rho_{A}||\rho_{A}^{(0)}) = -\int \mathrm{d}^{d}x_{a} \int \mathrm{d}^{d}x_{b}\lambda_{\alpha}(x_{a})\lambda_{\alpha}(x_{b}) \int \frac{\mathrm{d}s}{4\sinh^{2}\left(\frac{s+i\epsilon\mathrm{sgn}(\tau_{a}-\tau_{b})}{2}\right)} \frac{\Omega^{\Delta}(\tau_{b},\vec{x}_{b})}{\Omega^{\Delta}(\tau_{b}+is,\vec{x}_{b})} \left\langle \mathcal{T}\left[\mathcal{O}_{\alpha}(x_{a})\mathcal{O}_{\alpha}(\tau_{b}+is,\vec{x})\right] \right\rangle$$

$$(8.144)$$

where \mathcal{T} indicates time ordering in τ direction. This is the promised result that we can obtain the relative entropy from two-point functions (and a conformal factor plus some integrals).

Note that we need to time-order the Euclidean times because our correlation function formally lives on a mixed signature Schwinger-Keldysh contour as in equation (3.72). For correlation functions on this contour to make sense, we need a notion of path time ordering that respects the causality of the Lorentzian segment with respect to the order of the operators, resulting in the necessity of ordering the Euclidean times.

Going to Hyperbolic Coordinates

To put the equation (8.144) on a form that is amenable to explicit results we wish to change from the usual Poincaré coordinates of the bulk AdS space, inducing a conformal transformation of the boundary from Minkowski to hyperbolic coordinates.

 AdS_{d+1} may be embedded in the embedding space $\mathbb{R}^{1,d+1}$. Let us give the parametrization of AdS by Poincaré and hyperbolic coordinates.

Poincaré:
$$X^{A} = \left(\frac{1+z^{2}+x^{2}}{2z}, \frac{1-x^{2}-z^{2}}{2z}, \frac{\vec{x}}{z}\right)$$
,
Hyperbolic: $X^{A} = \left(rY^{I}, \sqrt{r^{2}-1}\cos(\tau), \sqrt{r^{2}-1}\sin(\tau), rY^{m}\right)$.
(8.145)

Note here that for Poincaré, z is the near coordinate measuring distance from the boundary and x is the d-dimensional vector of boundary coordinates. For the hyperbolic coordinates, the hyperbolic factor is described by

$$Y \equiv (Y^{I}, Y^{m}) = \left(\frac{1+u^{2}+\vec{x}^{2}}{2u}, \frac{1-\vec{x}^{2}-u^{2}}{2u}, \frac{\vec{x}}{u}\right) \in \mathbb{H}^{d-1} , \qquad (8.146)$$

fulfilling the restrictions that $Y^I > 0$ and $Y^2 \equiv -(Y^I)^2 + Y^m Y^m = -1$. Here, \vec{x} is a vector with d-2 spacelike components.

Points on the boundary of AdS $S^1 \times \mathbb{H}^{d-1}$ are given in the hyperbolic description by

$$P \equiv (P^I, P^{II}, P^{\mu}) = \left(Y^I, \cos\tau, \sin\tau, Y^m\right) \in S^1 \times \mathbb{H}^{d-1}$$
(8.147)

such that the points P are on the forward lightcone of the origin in the flat embedding space, i.e. $P \cdot P = 0, P^I > 0$. The hyperbolic description is related to the flat description

$$P = \left(\frac{R^2 + x^2}{2R}, \frac{R^2 - x^2}{2R}, x^{\mu}\right)$$
(8.148)

via a conformal transformation. This conformal transformation carries the conformal factor

$$\Omega(\tau, Y) = R^{-1}(Y^I + \cos \tau) .$$
(8.149)

The hyperbolic representation of AdS has the induced metric (defining $x^m = (t = -i\tau, t, u, \vec{x})$)

$$g^{(0)} = -(r^2 - 1)dt^2 + \frac{dr^2}{(r^2 - 1)} + \frac{r^2}{u^2}(du^2 + d\vec{x}^2) .$$
(8.150)

In these coordinates the metric is reminiscent of the Schwarzschild metric, so these coordinates are referred to as hyperbolic black hole coordinates. The patch $r^2 > 1$ covers exactly the domain of dependence of σ_A , i.e. a so-called Rindler wedge. In these coordinates the Killing field ξ_A is given by $2\pi n_t$. Note that τ was periodic, and the Euclidean Killing vector field generates translations around the S^1 . After this continuation to a black-hole like representation this corresponds to the fact that the black hole has a particular temperature.

In these coordinates, equation (8.144) becomes

$$\delta_{\epsilon}^{2}S(\rho_{A}||\rho_{A}^{(0)}) = -\int \mathrm{d}\boldsymbol{\mu} \int \frac{\mathrm{d}s}{4\sinh^{2}\left(\frac{s+i\epsilon\mathrm{sgn}(\tau_{a}-\tau_{b})}{2}\right)} \left\langle \mathcal{T}\left[\mathcal{O}_{\alpha}(\tau_{a},Y_{a})\mathcal{O}_{\alpha}(\tau_{b}+is,Y_{b})\right] \right\rangle , \qquad (8.151)$$

where we have defined

$$\int \mathrm{d}\boldsymbol{\mu} = \int_0^{2\pi} \mathrm{d}\tau_a \int_{\mathbb{H}^{d-1}} \mathrm{d}^{d-1} Y_a \int_0^{2\pi} \mathrm{d}\tau_b \int_{\mathbb{H}^{d-1}} \mathrm{d}^{d-1} Y_b \,\lambda_\alpha(\tau_a, Y_a) \lambda_\alpha(\tau_b, Y_b) \Omega^{\Delta - d}(\tau_a, Y_a) \Omega^{\Delta - d}(\tau_b, Y_b) \,.$$

$$(8.152)$$

The main ingredient now is to use that the Fisher information is determined in terms of timeordered two-point functions. We will proceed to specialize to a deformation with only a single scalar field, and then generalize the method to stress tensor correlators.

Symplectic Formalism for Scalar Deformations

We now specialize equation (8.151) to the case of a single scalar field. This will occur in two steps. First we show that we can formally compute the two-point function by computing a symplectic flux ω_{ϕ} defined in an auxiliary AdS space close to its boundary. The second step is to perform the ds integral and half the integrals in $d\mu$. Finally, we are able to reexpress the remaining integrals in the same form as the gravitational expression.

The main ingredient is that the CFT two-point function may be written in terms of a quantity in an auxiliary AdS-Rindler wedge that is the domain of dependence $\mathcal{D}(\Sigma_A)$ of the region previously called Σ_A , dual to the causal diamond $\mathcal{D}(A)$ on the boundary. The "existence" of this auxiliary AdS is purely formal. We will make use of only the relations between scalar two-point functions in the bulk and on the boundary which are rigorously true, as the ones in section 6.2.1. Notably this means that no conjectures regarding AdS/CFT are invoked, only explicitly proven operator maps are used.

The quantity in the auxiliary AdS spacetime that computes the two-point functions is an asymptotic symplectic flux $W_{\mathcal{D}(A)}$ that takes as arguments bulk-to-boundary propagators of the auxiliary AdS spacetime. To obtain correlators for the boundary theory we take a limit where we evaluate the symplectic flux close to the boundary. Mathematically, we define

$$- \langle \mathcal{O}(\tau, Y_a) \mathcal{O}(is, Y_b) \rangle = W_{\mathcal{D}_{\mathcal{A}}}(K_E, K_R)$$

$$\equiv \lim_{r_0 \to \infty} \int_{r_B = r_0} dt_B dY_B \omega_\phi \left(K_E(r_B, it_B, Y_B | \tau, Y_a), K_R(r_B, t_B | s, Y_b) \right) , \qquad (8.153)$$

where ω_{ϕ} is the symplectic two-form density⁷ for the bulk scalar ϕ dual to the operator \mathcal{O} on the boundary, K_E is a Euclidean bulk-to-boundary propagator and K_R is a retarded bulk-to-boundary propagator (compensating for the fact that it has to propagate forwards in imaginary euclidean time, i.e. lorentzian time). Here, we have used time translation invariance to define $\tau = \tau_a - \tau_b$. The integration region is a constant radial slice in the AdS spacetime at $r_B = r_0$ close to the AdS boundary. We use the hyperbolic black hole coordinates from equation (8.150) in which the induced metric on the integration region has the form

$$g^{(0)} = -(r_B^2 - 1)dt^2 + \frac{dr_B^2}{(r_B^2 - 1)} + r_B^2 dY_B^2 .$$
(8.154)

To correctly adapt the hyperbolic black hole coordinates to a particular boundary ball, we just pick $r_B = 1, t_B = 0$ to coincide with the Ryu-Takayanagi surface \tilde{A} as denoted in the leftmost diagram of figure 8.2. The region $r_B \ge 1$ then covers only the interior of the causal wedge imaged in figure 8.1.

Let us now properly define the propagators that we introduced in equation (8.153). We will just cite the form of the bulk propagators for asymptotically AdS spacetimes, referring to chapter 23.10 of [48] for a derivation. The Euclidean propagator in equation (8.153) K_E is sourced at euclidean time τ but it ends at real time t_B so the endpoint argument has to be analytically continued, yielding

$$K_E(r_B, it_B, Y_B | \tau, Y_a) = \frac{D_\Delta}{\left(-2r_B Y_B \cdot Y_a - \sqrt{r_B^2 - 1}\cos(\tau - it_B)\right)^\Delta},$$
(8.155)

where $D_{\Delta} = \pi^{-d/2} \Gamma(\Delta) / \Gamma(\Delta - d/2)$. The retarded propagator

$$K_R(r_B, t_B, Y_B|s, Y_b) = i\Theta(t_B - s) \Big(K_+(r_B, t_B, Y_B|s, Y_b) - K_-(r_B, t_B, Y_B|s, Y_b) \Big)$$
(8.156)

⁷This is a the same quantity as in the Hollands-Wald formalism of the gravitational calculation. A review of the field theoretical variant of this formalism can be found in [76].

where θ is the Heaviside step function and K_{\pm} are Wightman propagators defined by

$$K_{\pm}(r_B, t_B, Y_B | s, Y_b) = \lim_{\varepsilon \to 0^+} K_E(r_B, t_B, Y_B | is \mp \epsilon, Y_b)$$
$$= \lim_{\varepsilon \to 0^+} \frac{D_{\Delta}}{\left(-2r_B Y_B \cdot Y_a - \sqrt{r_B^2 - 1}\cosh(s - t_B \pm i\epsilon)\right)^{\Delta}}.$$
(8.157)

We can now verify that equation (8.153) is reasonable by using that the propagators represent solutions that approach delta functions as $r_0 \to \infty$.

Let us first consider the symplectic flux for general perturbations δ_1, δ_2 that satisfy the linearized equations of motion. Keeping the limit $r \to 0$ implicit, we have

$$W_{\mathcal{D}(A)}(\delta_{1}\phi, \delta_{2}\phi) = \int_{r_{B}=r_{0}} \omega_{\phi}(\delta_{1}\phi, \delta_{2}\phi)$$

$$= \int_{r_{B}=r_{0}} \epsilon^{M}(\delta_{1}\phi\partial_{M}\delta\phi_{2} - \delta_{2}\phi\partial_{M}\delta\phi_{1})$$

$$\cong \int_{r_{B}=r_{0}} \mathrm{d}t_{B}\mathrm{d}Y_{B} r_{B}^{d}(\delta_{1}\phi\partial_{r_{B}}\delta\phi_{2} - \delta_{2}\phi\partial_{r_{B}}\delta\phi_{1}) , \qquad (8.158)$$

where we used that the normal to the surface of constant radius is in the radial direction and the factor r_B^d comes from the square root of the determinant of the metric. Since the $\delta_i \phi$ solve the linearized equations of motion they asymptotically behave as

$$\delta_i \phi \sim a_i(t_B, Y_B) \, r_B^{-d+\Delta} + b_i(t_B, Y_B) \, r_B^{-\Delta} \quad (i \in \{1, 2\})$$
(8.159)

which inserted into equation (8.158) gives

$$W_{\mathcal{D}(A)}(\delta_1\phi, \delta_2\phi) = (2\Delta - d) \int dt_B dY_B(a_1b_2 - a_2b_1) .$$
(8.160)

From equation (8.153) we can use the properties of the propagators to read off the coefficients, noting that $\delta_1 = K_E$ and $\delta_2 = K_R$. Let us also denote by G_E, G_R as the Green's functions of the boundary theory. We then have that

$$\lim_{r_0 \to \infty} K_E \sim r^{-\Delta} \Rightarrow \begin{cases} a_1 = 0\\ b_1 = \frac{1}{2\Delta - d} G_E(it_B | \tau, Y_a) \end{cases}$$

$$\lim_{r_0 \to \infty} K_R \sim \delta^{d-1} (Y_B - Y_b) \delta(t_B - s) r^{-d+\Delta} + r^{-\Delta} \Rightarrow \begin{cases} a_2 = \delta^{d-1} (Y_B - Y_b) \delta(t_B - s) \\ b_2 = \frac{1}{2\Delta - d} G_R(it_B | \tau, Y_a) \end{cases}$$

$$(8.161)$$

where we defer the details of the asymptotic form of the propagators (especially K_R) to [63]. The expressions in terms of Γ s for the explicit correlators are well known in the AdS/CFT literature, and we saw some of this for three-point functions in section 6.2.1. The euclidean two-point function G_E in the boundary space is given by

$$G_E(\tau_1, Y_1 | \tau_2, Y_2) \equiv \langle \mathcal{O}(\tau_1, Y_1) \mathcal{O}(\tau_2, Y_2) \rangle = \frac{2\Delta - d}{\pi^{d/2}} \frac{\Gamma(\Delta)}{\Gamma(\Delta - d/2)} \frac{1}{\left(-2Y_1 \cdot Y_2 - 2\cos(\tau_1 - \tau_2)\right)^{\Delta}},$$
(8.162)

in the present case analytically continued to complex τ_2 (i.e a Lorentzian bulk point). Inserting the values for the coefficients into equation (8.160) one can then see that

$$-G_E(\tau, Y_a|is, Y_b) = W_{\mathcal{D}(A)}(K_E, K_R)$$
(8.163)



Figure 8.2: Images depicting a slice of the bulk causal wedge corresponding to the boundary region A. In a), the main boundaries are denoted, with A, \tilde{A} as points because they are codimension 2. In b), the range of t_B that is spacelike separated from a boundary point s is shown as a dashed line. In c) the boundary interval that is spacelike separated from a particular bulk point t_B . The dashed regions may be excluded from the first integral performed, since the Lorentzian propagators vanish for spacelike separated points. In addition, the retarded propagator is only nonzero when $t_B > s$.

proving that the asymptotic symplectic flux computes the two-point function.

With this, we may write equation (8.151) on the following form

$$\delta_{\epsilon}^2 S(\rho_A || \rho_A^{(0)})_{\phi} = \int \mathrm{d}\boldsymbol{\mu} \int_{-\infty}^{\infty} \frac{\mathrm{d}s}{4\sinh^2(\frac{s+i\epsilon \mathrm{sgn}(\tau_b - \tau_a)}{2})} W_{\mathcal{D}(A)}(K_E, K_R)$$
(8.164)

The reason that we wished to go this route of symplectic forms is two-fold. Firstly, it relates cleanly to the Hollands-Wald formalism of the previous section and secondly the integrated symplectic form is invariant under deformations of the integration surface. This second property is because we chose deformations that satisfy the equations of motion, making ω_{ϕ} a closed form on spacetime, i.e. $\omega_{\phi} = d\chi$ where χ is some one form.

Performing the Integrals

The previous result lets us deform the integration region and evaluate $W_{\mathcal{D}(A)}(K_E, K_R)$ at the horizon of the hyperbolic black hole, taking $r_B \to 1$. The computation of the desired integrals then becomes an exercise in Residue calculus with a fairly involved branch-cut and pole structure.

We write

$$\delta_{\epsilon}^2 S(\rho_A || \rho_A^{(0)})_{\phi} = \lim_{r_0 \to 1} \int \mathrm{d}\boldsymbol{\mu} \int_{-\infty}^{\infty} \frac{\mathrm{d}s}{4\sinh^2(\frac{s+i\epsilon \operatorname{sgn}(\tau_b - \tau_a)}{2})} \int_{r_B = r_0} \mathrm{d}t_B \mathrm{d}Y_B \omega_{\phi}(K_E, K_R)$$
(8.165)

The retarded propagator $K_R(r_B, t_B, Y_B|s, Y_b)$ is only nonzero when $t_B > s$ and by causality it is also only nonzero when the bulk and boundary points are timelike separated. This means that we may toss some of the integration region for t_B obtaining, as imaged in figure 8.2. The expression to be evaluated is then

$$\delta_{\epsilon}^{2} S(\rho_{A} || \rho_{A}^{(0)})_{\phi} = \lim_{r_{0} \to 1} i \int \mathrm{d}\boldsymbol{\mu} \int_{-\infty}^{\infty} \frac{\mathrm{d}s}{4 \sinh^{2}\left(\frac{s+i\epsilon \operatorname{sgn}(\tau_{b}-\tau_{a})}{2}\right)} \int_{t_{B,*}^{\pm}}^{\infty} \mathrm{d}t_{B} \int_{\mathbb{H}^{d-1}} \mathrm{d}Y_{B} \omega_{\phi}(K_{E}, K_{+}-K_{-})$$

$$(8.166)$$
where

$$t_{B,*}^{\pm} = s - \ln\left(\alpha \pm \sqrt{\alpha^2 - 1}\right), \ \alpha = -\frac{r_0 Y_B \cdot Y_b}{\sqrt{r_0^2 - 1}} \tag{8.167}$$

corresponds to the singularities of the retarded propagator that occur at lightlike separation between the bulk and boundary points⁸. Note that we may pick either $t_{B,*}^+$ or $t_{B,*}^-$, since the propagator vanishes in between. Similarly we may exchange the *s* and t_B integrals, yielding

$$\delta_{\epsilon}^{2}S(\rho_{A}||\rho_{A}^{(0)})_{\phi} = \lim_{r_{0}\to 1} \int \mathrm{d}\boldsymbol{\mu} \int_{-\infty}^{\infty} \mathrm{d}t_{B} \int_{-\infty}^{s_{*}^{\pm}} \frac{\mathrm{d}s}{4\sinh^{2}\left(\frac{s+i\epsilon\mathrm{sgn}(\tau_{b}-\tau_{a})}{2}\right)} \int_{\mathbb{H}^{d-1}} \mathrm{d}Y_{B}\omega_{\phi}(K_{E},K_{+}-K_{-})$$

$$(8.168)$$

where

$$s_*^{\pm} = t_B + \ln\left(\alpha \mp \sqrt{\alpha^2 - 1}\right)$$
 (8.169)

With the integral on this form, we are ready to start working on integrating over s. Since K_E does not depend on s we only need to care about the retarded propagator. Therefore our current objective is to compute the integral

$$I(r_B, t_B, Y_B | Y_b) = i \int_{-\infty}^{s_*^{\pm}} \frac{\mathrm{d}s}{4\sinh^2(\frac{s+i\epsilon\mathrm{sgn}(\tau_b - \tau_a)}{2})} \left(K_+(r_B, t_B, Y_B | s, Y_b) - K_-(r_B, t_B, Y_B | s, Y_b) \right)$$

$$= i \lim_{\varepsilon \to 0^+} \int_{C_\varepsilon \cup C_{-\varepsilon}} \frac{\mathrm{d}s}{4\sinh^2(\frac{s+i\epsilon\mathrm{sgn}(\tau_b - \tau_a)}{2})} K_E(r_B, t_B, Y_B | is, Y_b) ,$$

(8.170)

where we have invoked the definition of the Wightman propagators as limits of the euclidean propagator and defined the curves

$$C_{\varepsilon} = (-\infty + i\varepsilon, s_{\ast}^{-} + i\varepsilon] , C_{-\varepsilon} = [s_{\ast}^{-} - i\varepsilon, -\infty - i\varepsilon) .$$
(8.171)

Using the 2π periodicity in euclidean of the hyperbolic black hole geometry we may move the curve $C_{-\varepsilon}$ to $C_{2\pi-\varepsilon}$.

To understand the value of this operation we shall consider the analytic structure of the integrand in $I(r_B, t_B, Y_B|Y_b)$, as imagine in figure 8.3. The function $\sinh\left(\frac{s+i\epsilon \operatorname{sgn}(\tau_a-\tau_b)}{2}\right)$ has a zero at $s = -i\epsilon \operatorname{sgn}(\tau_a - \tau_b) + N2\pi$ where N is any integer, so the integrand has (double) poles separated by a distance 2π on the imaginary s axis. The Euclidean bulk to boundary propagator has branch cuts at $\operatorname{Re}(s) \leq s_*^-$, $\operatorname{Im}(s) = 2\pi N$ and well as $\operatorname{Re}(s) \geq s_*^+$, $\operatorname{Im}(s) = 2\pi N$. This analytic structure is depicted in figure 8.3. We see that the displacement of $C_{-\varepsilon}$ to $C_{2\pi-\varepsilon}$ lets us complete the curve without including any branch cuts in its interior.

To simplify the analysis, it is good to think about the limit $r_0 \to 1$ and the integral over t_B right away. As $r_0 \to 1$ the t_B integral splits into two contributions, along the past and future horizons \mathcal{H}^- and \mathcal{H}^+ of the hyperbolic black hole.

The contribution from the past horizon may be evaluated by considering the limit $r_B \to 1$, $t_B \to -\infty$ with $\sqrt{r_B^2 - 1}e^{-t_B}$ fixed. In terms of light-cone coordinates

$$l_B^+ = \sqrt{r_B^2 - 1}e^{t_B} , \quad l_B^- = \sqrt{r_B^2 - 1}e^{-t_B}$$
(8.172)

we have $l_B^+ \to 0$ and l_B^- fixed.

⁸Since our result differs from [63], there is a short derivation in appendix



Figure 8.3: The analytic structure of the integrand in equation (8.170). The integrand has double poles on the imaginary s axis at $2\pi n - \epsilon$, and limits for the C_{ε} have to be takens such that $\varepsilon < \epsilon$. The Euclidean propagator has branch cuts on the real axis for timelike separated points, with the rest of the cuts due to periodicity of the geometry. Using the 2π periodicity in imaginary time of the hyperbolic black hole the integral over $C_{-\varepsilon}$ is equivalent to the integral over $C_{2\pi-\varepsilon}$. Notably, closing the curve with C_+ includes a simple pole in the interior. Completing the contour with $C_$ and C_+ is appropriate on \mathcal{H}_- and \mathcal{H}_+ respectively.

In this limit we may complete the curve with the vertical piece

$$C_{-} = [s_{*}^{-} + i\varepsilon, s_{*}^{-}i(2\pi - \varepsilon)] .$$
(8.173)

By Cauchy's theorem the integral around the closed contour is 0 since it encircles no poles and no branch cuts⁹. This tells us that I is minus the integral over C_{-} . Let us then consider the integral over C_{-} , dropping ϵ and ε (because they contribute no important behaviour) and adding in a single new regulator ε' to make sure our integration region doesn't intersect the branch cut.

$$I(C_{-}) = i \lim_{l_{B}^{+} \to 0} \int_{s_{*}^{-} + \varepsilon'}^{s_{*}^{-} + \varepsilon' + 2\pi i} \frac{\mathrm{d}s}{4\sinh^{2}(s/2)} K_{E}(r_{B}, it_{B}, Y_{B}|is, Y_{b}) .$$
(8.174)

Here, it is convenient to work in lightcone coordinates. Using translation invariance of K_E in Rindler time¹⁰ we can transform the integration variable to $w = e^{s-s_*}$ so that equation (8.174) becomes

$$I(C_{-}) = -i \lim_{l_{B}^{+} \to 0} \oint_{\Gamma} \mathrm{d}w \frac{e^{s_{*}}}{(w - e^{s_{*}})^{2}} K_{E}\left(\frac{l_{B}^{+}w}{e^{s_{*}}}, \frac{l_{B}^{-}e^{s_{*}}}{w}, Y_{B}|0, Y_{b}\right) , \qquad (8.175)$$

where Γ is the curve $|w| = 1 + \epsilon$ and there is some work involved in rewriting $\sinh^2(s)$ on a good form. We are interested in the limit of small l_B^+ , so we should make an expansion in powers thereof. The integral is about a closed curve, so we are also interested in the residues, i.e. terms proportional to 1/w. Let us first expand in powers of e^{-s_*}

$$I(C_{-}) = -i \lim_{l_{B}^{+} \to 0} \oint_{\Gamma} dw \left(\frac{e^{s_{*}^{-}}}{w^{2}} - \frac{2e^{2s_{*}^{-}}}{w^{3}} + \dots \right) K_{E} \left(\frac{l_{B}^{+}w}{e^{s_{*}^{-}}}, \frac{l_{B}^{-}e^{s_{*}^{-}}}{w}, Y_{B}|0, Y_{b} \right) .$$
(8.176)

⁹Technically we also need to close the contour at $\operatorname{Re}(s) = -\infty$, but the contribution vanishes since $\sinh \sim e^{\infty}$ in this limit.

 $^{{}^{10}}K_E(r_B, it_B, Y_B|is, Y_b) = K_E(r_B, it_Be^a, Y_B|is + ia, Y_b)$, time translation on the boundary is Rindler time translation in the bulk.

We see that contributions to the residue must come from powers of $l_B^+ w/e^{s_*^-}$, which looks promising for eventual convergence. To be completely sure, we should expand $e^{s_*^+}$ and $e^{-s_*^-}$ in powers of l_B^+ .

In terms of the lightcone coordinates we have

$$r_0^2 = 1 + l_B^+ l_B^-$$
, $e^{t_B} = \frac{l_B^+}{\sqrt{r_0^2 - 1}} = \sqrt{\frac{l_B^+}{l_B^-}}$, $e^{-t_B} = \sqrt{\frac{l_B^-}{l_B^+}}$. (8.177)

The α appearing in the definition of s_*^- can then be expressed as

$$\alpha = -\left(1 + \frac{1}{l_B^+ l_B^-}\right)^{1/2} Y_B \cdot Y_b$$

$$= -\frac{Y_B \cdot Y_b}{\sqrt{l_B^+ l_B^-}} \left(1 + \frac{l_B^+ l_B^-}{2}\right) + \mathcal{O}(l_B^+)^{3/2}$$
(8.178)

as well as

$$\sqrt{\alpha^{2} - 1} = \left[\left(1 + \frac{1}{l_{B}^{+} l_{B}^{-}} \right) (Y_{B} \cdot Y_{b})^{2} - 1 \right]^{\frac{1}{2}} \\
= \frac{Y_{B} \cdot Y_{b}}{\sqrt{l_{B}^{+} l_{B}^{-}}} \left[1 + \left(1 - \frac{1}{(Y_{B} \cdot Y_{b})^{2}} \right) l_{B}^{+} l_{B}^{-} \right]^{\frac{1}{2}} \\
= \frac{Y_{B} \cdot Y_{b}}{\sqrt{l_{B}^{+} l_{B}^{-}}} \left[1 + \frac{1}{2} \left(1 - \frac{1}{(Y_{B} \cdot Y_{b})^{2}} \right) l_{B}^{+} l_{B}^{-} \right] + \mathcal{O}(l_{B}^{+})^{3/2} .$$
(8.179)

Using this we can expand $e^{\pm s_*}$

$$e^{s_{*}} = e^{t_{B}} \left(\alpha + \sqrt{\alpha^{2} - 1} \right)$$

$$\cong \frac{Y_{B} \cdot Y_{b}}{l_{B}^{-}} \left[-\left(1 + \frac{l_{B}^{+} l_{B}^{-}}{2} \right) + 1 + \frac{1}{2} \left(1 - \frac{1}{(Y_{B} \cdot Y_{b})^{2}} \right) l_{B}^{+} l_{B}^{-} \right]$$

$$= \frac{Y_{B} \cdot Y_{b}}{l_{B}^{-}} \left[-\frac{1}{2(Y_{B} \cdot Y_{b})^{2}} l_{B}^{+} l_{B}^{-} \right]$$

$$= -\frac{l^{+}}{-2Y_{B} \cdot y_{b}} . \qquad (8.180)$$

Similarly we have

$$e^{-s_*^-} \cong -\frac{2Y_B \cdot Y_b}{l_B^+} + \mathcal{O}(1)$$
 (8.181)

Inserting into equation (8.176) we see that the factors w that can come from the propagator carry with them a leading $\mathcal{O}(1)$ term in l_B^+ . The lowest order term in the rest of the integrand is of order l_B^+ , and therefore the residue is proportional to $l_B^+ + \mathcal{O}\left((l_B^+)^2\right)$. Taking the limit $l_B^+ \to 0$ we conclude that

$$I(C_{-}) = 0 . (8.182)$$

This implies directly that $\lim_{l_0^+\to 0} I(r_B, t_B, Y_B|Y_b) = 0$ since the completed contour encircles no singularities. Note that this integrand would not have vanished in a limit as $l_B^- \to 0$, which is why we in the following need to consider a different contour.

Let us now consider the limit $l_B^- \to 0$, l_B^+ fixed, i.e. the future horizon \mathcal{H}^+ of the hyperbolic black hole. It is now more practical to extend the two horizontal contours up to s_*^+ and close the contour with the curve

$$C_{+} = (s_{*}^{+} + i\varepsilon, s_{*}^{+} + (2\pi - i\varepsilon)) .$$
(8.183)

The simultaneous extension of the horizontal contours does not contribute since they are not separated by a branch cut, which together with Im(s) periodicity makes them cancel as $\varepsilon \to 0$. Let us now check that C_+ vanishes as $l_B^- \to 0$, telling us that the integral over s on \mathcal{H}^+ is given by a residue.

Similarly to before we use Rindler time translation invariance to pick good coordinates. In this case we want to choose $w = e^{s-s_*^+}$ putting the integral along the vertical contour on the form

$$I(C_{+}) = i \lim_{l_{B}^{-} \to 0} \oint_{\Gamma} \mathrm{d}w \frac{e^{-s_{*}^{+}}}{(w - e^{s_{*}^{+}})^{2}} K_{E}\left(t \frac{l_{B}^{+}}{w e^{s_{*}^{+}}}, l_{B}^{-} w e^{s_{*}^{+}}, Y_{B}|0, Y_{b}\right) .$$
(8.184)

Then, using the partial results of the previous calculation we can write down

$$e^{s_{*}^{+}} = e^{t_{B}} \left(\alpha - \sqrt{\alpha^{2} - 1} \right)$$

$$\cong \frac{Y_{B} \cdot Y_{b}}{l_{B}^{-}} \left[-\left(1 + \frac{l_{B}^{+} l_{B}^{-}}{2} \right) - 1 - \frac{1}{2} \left(1 - \frac{1}{(Y_{B} \cdot Y_{b})^{2}} \right) l_{B}^{+} l_{B}^{-} \right]$$

$$= \frac{Y_{B} \cdot Y_{b}}{l_{B}^{-}} \left[-1 - \left(\frac{1}{2} - \frac{1}{(2Y_{B} \cdot Y_{b})^{2}} \right) l_{B}^{+} l_{B}^{-} \right]$$

$$= -\frac{Y_{B} \cdot Y_{b}}{l_{B}^{-}} \left[1 + \left(\frac{1}{2} - \frac{1}{(2Y_{B} \cdot Y_{b})^{2}} \right) l_{B}^{+} l_{B}^{-} \right]$$
(8.185)

and

$$e^{-s_*^+} \cong \frac{l_B^-}{2Y_B \cdot Y_b} \left[1 - \left(\frac{1}{2} - \frac{1}{(2Y_B \dot{Y}_b)^2}\right) l_B^+ l_B^- \right] .$$
(8.186)

From here it follows from the same argument as for the C^{-} case that the vertical contour vanishes.

Having proven that the vertical contour gives no contribution, we would now like to evaluate the full \boldsymbol{s} integral

$$\lim_{l_B^- \to 0} I(l_B^+, l_B^-, Y_B | Y_b) = i \lim_{l_B^- \to 0} \lim_{\varepsilon \to 0^+} \int_{C_\varepsilon \cup C_{-\varepsilon}} \frac{\mathrm{d}s}{4\sinh^2(\frac{s+i\epsilon \mathrm{sgn}(\tau_b - \tau_a)}{2})} K_E(l_B^+, l_B^-, Y_B | is, Y_b)$$

$$= -2\pi \partial_s K_E(l_B^+, 0, Y_B | \mathrm{sgn}(\tau_a - \tau_b)\epsilon, Y_b) .$$
(8.187)

Making the substitution $-2\pi\partial_s \to \mathcal{L}_{\xi_A}$ and inserting the result into equation (8.168) we have

$$\delta_{\epsilon}^{2} S(\rho_{A} || \rho_{A}^{(0)})_{\phi} = \int \mathrm{d}\boldsymbol{\mu} \int_{\mathcal{H}^{+}} \mathrm{d}t_{B} \mathrm{d}Y_{B} \omega_{\phi} (K_{E}(l_{B}^{+}, 0, Y_{B} || (\tau_{a} - \tau_{b}), Y_{a}), \mathcal{L}_{\xi_{A}} K_{E}(l_{B}^{+}, 0, Y_{B} || (sgn(\tau_{a} - \tau_{b})\epsilon, Y_{a}))$$

$$\equiv \int \mathrm{d}\boldsymbol{\mu} W_{\mathcal{H}^{+}} (K_{E}(l_{B}^{+}, 0, Y_{B} || (\tau_{a} - \tau_{b}), Y_{a}), \mathcal{L}_{\xi_{A}} K_{E}(l_{B}^{+}, 0, Y_{B} || (sgn(\tau_{a} - \tau_{b})\epsilon, Y_{b}))$$

$$(8.188)$$

where $W_{\mathcal{H}^+}$ is the integral of ω over \mathcal{H}^+ , to which the coordinates t_B and Y_B have been restricted. We can make the arguments of the propagators depend only on t_a and t_b respectively by rotating the lightcone coordinate $l_B^+ \to e^{-it_b} l_B^+$. The new lightcone coordinate can be rotated without crossing any branch cuts in the propagator [63], the upshot being that $K_E(l_B^+, 0, Y_B | (\tau_a -$ $(\tau_b), Y_a), \mathcal{L}_{\xi_A} K_E(l_B^+, 0, Y_B | \operatorname{sgn}(\tau_a - \tau_b)\epsilon, Y_a) \to K_E(l_B^+, 0, Y_B | \tau_a, Y_a), \mathcal{L}_{\xi_A} K_E(l_B^+, 0, Y_B | \tau_b + \operatorname{sgn}(\tau_a - \tau_b)\epsilon, Y_b).$ Remembering the definitions of $d\mu$ and $\delta\phi$ we find that

$$\delta_{\epsilon}^2 S(\rho_A || \rho_A^{(0)})_{\phi} = W_{\mathcal{H}^+}(\delta\phi, \mathcal{L}_{\xi_A}\delta\phi) , \qquad (8.189)$$

where we realized that $\delta\phi$ as defined in equation (8.140) onwards may be written

$$\delta\phi(l_B^+, Y_B) = \int d\tau dY \lambda(t, Y) \Omega^{d-\Delta}(\tau, Y) K_E(l_B^+, Y_B | \tau, Y) . \qquad (8.190)$$

The field $\delta\phi$ solves the vacuum Lorentzian equations of motion and has boundary behavior consistent with the sources λ_{α} . Therefore it is an on-shell scalar field perturbation in the same sense as the scalar field perturbations in the gravitational calculation.

Stress Tensor Deformations

We are now ready to attack stress tensor deformations, where most of the details of the previous calculation carry over. The main difference is that the vertical pieces of the contour will not vanish, giving rise to boundary terms. These boundary terms will be related to the boundary term $\sim \int d\chi$ of the gravitational calculation.

The stress energy tensor is a symmetric combination of scalars that transform under the Lorentz group, so all we need to do is let the source $\lambda_{\alpha} = \lambda_{(\alpha\beta)}$. Then from equation (8.151) we see that

$$\delta_{\epsilon}^{2}S(\rho_{A}||\rho_{A}^{(0)}) = -\int \mathrm{d}\boldsymbol{\mu}_{\alpha\beta\gamma\delta} \int \frac{\mathrm{d}s}{4\sinh^{2}\left(\frac{s+i\epsilon\mathrm{sgn}(\tau_{a}-\tau_{b})}{2}\right)} \left\langle \mathcal{T}\left[T^{\alpha\beta}(\tau_{a},Y_{a})T^{\gamma\delta}(\tau_{b}+is,Y_{b})\right] \right\rangle , \quad (8.191)$$

where we have defined

$$\int \mathrm{d}\boldsymbol{\mu} = \int_0^{2\pi} \mathrm{d}\tau_a \int_{\mathbb{H}^{d-1}} \mathrm{d}^{d-1} Y_a \int_0^{2\pi} \mathrm{d}\tau_b \int_{\mathbb{H}^{d-1}} \mathrm{d}^{d-1} Y_b \,\lambda_{\alpha\beta}(\tau_a, Y_a) \lambda_{\gamma\delta}(\tau_b, Y_b) \Omega^{\Delta-d}(\tau_a, Y_a) \Omega^{\Delta-d}(\tau_b, Y_b) \,\mathcal{I}_{\boldsymbol{\lambda}}^{\Delta-d}(\tau_b, Y_b) \,\mathcal{I}_{\boldsymbol{\lambda$$

As in the case of scalars, the stress tensor correlation function can be rewritten in terms of a symplectic flux

$$-\left\langle \mathcal{T}\left[T^{\alpha\beta}(\tau_a, Y_a)T^{\gamma\delta}(\tau_b + is, Y_b)\right] \right\rangle = \frac{\tilde{C}_T}{a^*} W^{\text{grav}}_{\mathcal{D}(A)}(K^{\alpha\beta}_{E;\mu\nu}, K^{\gamma\delta}_{R;\rho\sigma})$$
(8.193)

where $W_{\mathcal{D}(A)}^{\text{grav}}$ is the integral of the symplectic two form associated with metric perturbations (equation (8.110)) over the domain of dependence $\mathcal{D}(A)$ of the boundary region A. We have here defined $K_{E;\mu\nu}^{\alpha\beta}$ as the Euclidean bulk to boundary propagator, where $\mu\nu\ldots$ denote bulk indices and $\alpha\beta$ boundary indices.

The normalization factor \tilde{C}_T/a^* is needed because the flux ω^{grav} computes the stress tensor in terms of a^* as determined in terms of the auxiliary AdS radius by equation (7.59)

$$L_{\rm AdS} = \left(\frac{\Gamma(\frac{d}{2})}{\pi^{\frac{d}{2}}} 8\pi G_N a^*\right)^{\frac{1}{d-1}} , \qquad (8.194)$$

while \tilde{C}_T is proportional to the normalization of the two-point function as defined on the CFT side.

One additional issue arises in the case of the graviton propagators; they are not gauge invariant. Their analytic properties depend strongly on what gauge is chosen, so a good gauge is necessary to repeat the scalar field analysis without complication. With this in mind, a convenient gauge is the generalized de Donder gauge (also known as traceless transverse gauge)

$$\nabla^{(0)}_{\mu}h^{\mu\nu} = 0 , \quad g^{\mu\nu}_{(0)}h_{\mu\nu} . \qquad (8.195)$$

In this gauge the propagator can be written in terms of the embedding space of the auxiliary AdS space as $[77]^{11}$

$$K_E(X_B, Z_B | P, Z) = C \frac{(2(Z_B \cdot P)(Z \cdot X_B) - 2(P \cdot X_B)(Z_B \cdot Z))^2}{(-2P \cdot X_B)^{d+2}}$$
(8.196)

where the inner products are defined in the flat embedding space, Z, Z_B are auxiliary embedding space coordinates,

$$X_B = \left(r_B Y_B^I, \sqrt{r_B^2 - 1} \cos(\tau_B), \sqrt{r_B^2 - 1} \sin(\tau_B), r_B Y_B^m \right)$$
(8.197)

is the bulk point in the hyperbolic embedding of AdS (see equation (8.145)) and

$$P = (Y^I, \cos\tau, \sin\tau, Y^m) \tag{8.198}$$

is the boundary point.

In this gauge the Wightman propagators have no spacelike branch cuts, and the only poles are at $s_*^{\pm} + 2\pi i N$ where N is integer. This means that the possibility of performing the contour deformations of the previous section are left intact. Equation equation (8.193) is gauge invariant and may be verified in any gauge, but the following analysis depends on the gauge choice.

At this point, we have the expression

$$\delta_{\epsilon}^2 S(\rho_A || \rho_A^{(0)}) = -\frac{\tilde{C}_T}{a^*} \int \mathrm{d}\boldsymbol{\mu}_{\alpha\beta\gamma\delta} \int \frac{\mathrm{d}s}{4\sinh^2\left(\frac{s+i\epsilon\mathrm{sgn}(\tau_a-\tau_b)}{2}\right)} W_{\mathcal{D}(A)}^{\mathrm{grav}}(K_{E;\mu\nu}^{\alpha\beta}, K_{R;\rho\sigma}^{\gamma\delta}) \ . \tag{8.199}$$

We then repeat the scalar field analysis, pushing the integration region onto the horizon of the hyperbolic black hole. The goal is then to compute the ds integral by repeating the same techniques. The integral has three potential contributions:

- 1. The \sinh^2 double pole.
- 2. The vertical contour C_{-} at s_{*}^{-} near \mathcal{H}^{-} .
- 3. The vertical contour C_+ at s_*^+ near \mathcal{H}^+ .

The contribution from the double pole is entirely analogous to the scalar case, with the same ∂_s and Rindler-translation symmetry into \mathcal{L}_{ξ_A} resulting in the expression

$$\delta_{\epsilon}^2 S(\rho_A || \rho_A^{(0)})_{\text{grav}} \Big|_{pole} = \frac{\tilde{C}_T}{a^*} W_{\mathcal{H}^+}^{\text{grav}}(h, \mathcal{L}_{\xi_A} h) \equiv \frac{\tilde{C}_T}{a^*} \int_{\mathcal{H}^+} \omega^{\text{grav}}(h, \mathcal{L}_{\xi_A} h) , \qquad (8.200)$$

where

$$h_{\mu\nu}(l_B^+, Y_B) = \frac{1}{2} \int d\tau dY \lambda_{\alpha\beta}(\tau, Y) \Omega^{-2}(\tau, Y) K_{E;\mu\nu}^{\alpha\beta} .$$
(8.201)

¹¹We are using equation 60 of [77], which is written for general spin J using a particularly neat embedding space formalism.

Just like in the scalar case this metric perturbation solves the linearized Einstein field equations about the auxiliary AdS background given a Euclidean solution that matches onto the sources as boundary condition. Equivalently it can be seen as the vacuum Lorentzian solution with asymptotic behaviour determined by the expectation value of the boundary CFT stress energy tensor.

This CFT expression together with the scalar field term matches the gravitational result in Hollands-Wald gauge equation (8.124), given that the gravitational equations of motion are satisfied. We know that this can not be the whole story, since the de-Donder and Hollands-Wald gauges are not equivalent. Therefore we hope for the appearance of a boundary term localized on the RT surface. We will see that exactly such a term comes from the vertical contours.

Vertical Contours

We pick the integral along C_+ for explicits, using the same tricks as in the scalar case we write

$$I_{\mu\nu}(C_{+}) = i \lim_{l_{B}^{-} \to 0} \int_{s_{*}^{-} \varepsilon'}^{s_{*}^{-} \varepsilon' + 2\pi i} \frac{\mathrm{d}s}{4\sinh^{2}(s/2)} h_{\rho\sigma}(l_{B}^{+}e^{-s}, l_{B}^{-}e^{s}, Y_{B}) J^{\rho}{}_{\mu} J^{\sigma}{}_{\nu}$$

$$= i \lim_{l_{B}^{-} \to 0} \int_{\Gamma} \mathrm{d}w \frac{e^{-s_{*}^{+}}}{(w - e^{-s_{*}^{+}})^{2}} h_{\rho\sigma} \left(\frac{l_{B}^{+}}{e^{s_{*}^{+}}}, l_{B}^{-}we^{s_{*}^{+}}, Y_{B}\right) J^{\rho}{}_{\mu} J^{\sigma}{}_{\nu}$$
(8.202)

where Γ is defined by $|w| = 1 - \varepsilon'$ and translation invariance in Rindler time has been used to rewrite the bulk-to-boundary propagator as a bulk graviton. Compared to the scalar case the new feature is the appearance of a Jacobian $J^{\mu}{}_{\nu}$ due to the fact that the graviton itself transforms under a change of coordinates

$$J^{\overline{b}}_{\overline{a}} = \begin{bmatrix} \frac{e^{-s^{+}_{-}}}{w} & 0\\ 0 & we^{s^{+}_{*}} \end{bmatrix} , \quad J^{b}_{\ a} = \delta^{b}_{\ a} , \qquad (8.203)$$

where unbarred indices run along \mathbb{H}^{d-1} while the barred indices run over the lightcone directions l_B^{\pm} .

Now we should evaluate the w integral component for component. For example, we have

$$I_{--}(C_{+}) = i \lim l_{B}^{-} \oint_{\Gamma} \mathrm{d}w \left(\frac{e^{-s_{*}^{+}}}{w^{2}} + 2\frac{e^{-2s_{*}^{+}}}{w^{3}} + \dots \right) w^{2} e^{2s_{*}^{+}} h_{--} \left(\frac{l_{B}^{+}}{e^{s_{*}^{+}}}, l_{B}^{-} w e^{s_{*}^{+}}, Y_{B} \right) , \qquad (8.204)$$

where we should expand

$$h_{--}\left(\frac{l_B^+}{e^{s_*^+}}, l_B^- w e^{s_*^+}, Y_B\right) = h_{--}\left(0, 0, Y_B\right) + \frac{l^+ B}{w e^{s_*^+}} \partial_+ h_{--}\left(0, 0, Y_B\right) + l_B^- w \partial_- e^{s_*^+} h_{--}\left(0, 0, Y_B\right) \dots$$

$$(8.205)$$

Then, using the residue theorem we find

$$I_{--}(C_{+}) = -2\pi \left[2(h_{-})_{\tilde{A}} + l_{B}^{+}(\partial_{+}h_{--})_{\tilde{A}} + \mathcal{O}(l_{B}^{-}) \right] , \qquad (8.206)$$

where the \tilde{A} subscript tells us to evaluate h and it's derivatives at the undeformed RT surface $l_B^+ = l_B^- = 0$ (see figure 8.2). The reason l_B^+ is close to zero is because it is suppressed in

the propagator by a factor $e^{-s_*^+}$. Computing the rest of the components in the same way one finds

$$I_{-a}(C_{+}) = -2\pi (h_{-a})_{\tilde{A}} + \mathcal{O}(l_{B}^{-}) , \qquad (8.207)$$

$$I_{ab}(C_{+}) = -2\pi l_{B}^{-} (h\partial_{-}h_{ab})_{\tilde{A}} + \mathcal{O}((l_{B}^{-})^{2}) , \qquad (8.208)$$

$$I_{+-}(C_{+}) = \mathcal{O}(l_{B}^{-}) , \qquad (8.209)$$

$$I_{+a}(C_{+}) = \mathcal{O}((l_B^{-})^2) , \qquad (8.210)$$

$$I_{++}(C_{+}) = \mathcal{O}((l_B^{-})^3) .$$
(8.211)

The integral $I_{\mu\nu}(C_{-})$ is determined by an analogous computation. We see that in contrast to the scalar case we now have finite contributions, and these are localized at the RT surface $\mathcal{H}^{+} \cap \mathcal{H}^{-}$

Summarizing, the stress tensor contribution the Fisher information is given by

$$\delta_{\epsilon}^2 S(\rho_A || \rho_A^{(0)}) = \frac{\bar{C}_T}{a^*} \left[\int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A}) + \int_{\mathcal{H}^-} \omega_{\text{grav}}(h, -I(C_-)) + \int_{\mathcal{H}^+} \omega_{\text{grav}}(h, -I(C_+)) \right].$$
(8.212)

Matching Boundary Terms to Gravitational Result

In the gravitational result, being away from Hollands-Wald gauge gave rise to to the boundary term (equation (8.135))

$$\int_{\tilde{A}} \chi(h, [\xi_A, V]) = \int_{\Sigma_A} \omega^{\text{grav}}[h, \mathcal{L}_{[\xi_A, V]}] g^{(0)}$$
(8.213)

where V is a vector field satisfying equation (8.134), vanishing sufficiently fast at the boundary of AdS. We will now show that the boundary terms of the CFT calculation can be written on this form.

The first step is to use the de Donder gauge conditions equation (8.195) as well as $\xi_A = 2\pi (l_B^+ \partial_+ - l_B^- \partial_-)$ in lightcone coordinates to rewrite

$$I_{--}(C_{-}) = (\mathcal{L}_{\xi_{A}}h_{--})_{\tilde{A}} + 2\pi \frac{l_{B}^{+}}{2} \left(\nabla_{a}^{(0)}h^{a}_{-} - \frac{1}{2}\nabla_{-}^{(0)}h^{a}_{a}\right)_{\tilde{A}} + \mathcal{O}(l_{B}^{-})$$
(8.214)

$$I_{-a} = (\mathcal{L}_{\xi_A} h_{-a})_{\tilde{A}} + \mathcal{O}(l_B^-) .$$
(8.215)

The term in parentheses looks clearly related to the Hollands-Wald condition on V equation (8.134). This can be made exact, we make the guess that there exists a vector field $\mathcal{V}_{(+)}^a$ such that

$$\int_{\mathcal{H}^+} \omega^{\operatorname{grav}}(h, -I(C_+)) = \int_{\mathcal{H}^+} \omega^{\operatorname{grav}}(h, \mathcal{L}_{[\xi_A, \mathcal{V}_{(+)}]}g^{(0)}) .$$
(8.216)

To show that this is true, we make an ansatz for \mathcal{V}_+ and try to match $-I_{\mu\nu} = \mathcal{L}_{[\xi_A, V_{(+)}]} g_{\mu\nu}^{(0)}$. This turns out to be possible up to a term that vanishes when integrated over \mathcal{H}^+ . The required conditions on $\mathcal{V}_{(+)}$ turn out to be [63]

$$2\left(\partial_{-}\mathcal{V}_{(+)}^{+}\right)_{\tilde{A}} = (\mathcal{L}_{\xi_{A}}h_{--})_{\tilde{A}}$$
$$\left(\frac{1}{2}\partial_{a}\mathcal{V}_{(+)}^{+} + \frac{\delta_{ab}}{u^{2}}\partial_{-}\mathcal{V}_{(+)}^{b}\right)_{\tilde{A}} = (\mathcal{L}_{\xi_{A}}h_{-a})_{\tilde{A}}$$
$$\left(\nabla_{a}^{(0)}\nabla^{(0),a}\mathcal{V}_{(+),-} + [\nabla_{a}^{(0)},\nabla_{-}^{(0)}]\mathcal{V}_{(+)}^{a}\right)_{\tilde{A}} = \left(\nabla_{a}^{(0)}h^{a}_{-} - \frac{1}{2}\nabla_{-}^{(0)}h^{a}_{-}\right)_{\tilde{A}} .$$
(8.217)

These conditions are are exactly the $\overline{\alpha} = +$ Hollands-Wald gauge conditions, so we conclude that $\mathcal{V}^+_{(+)}\Big|_{\tilde{A}}$ and $\partial_-\mathcal{V}^+_{(+)}\Big|_{\tilde{A}}$ are exactly the +-component of the vector field V. By a similar argument one can define a vector field $\mathcal{V}^a_{(-)}$ on \mathcal{H}^- and make the replacement $-I(C_-) \to \mathcal{L}_{[\xi_A,\mathcal{V}_{(-)}]}g^{(0)}$.

At this point, the CFT result reads

$$\delta_{\epsilon}^{2}S(\rho_{A}||\rho_{A}^{(0)}) = \frac{\tilde{C}_{T}}{a^{*}} \left[\int_{\Sigma_{A}} \omega_{\text{grav}}(h,\mathcal{L}_{\xi_{A}}) + \int_{\mathcal{H}^{-}} \omega_{\text{grav}}(h,\mathcal{L}_{[\xi_{A},\mathcal{V}_{(-)}]}g^{(0)}) + \int_{\mathcal{H}^{+}} \omega_{\text{grav}}(h,\mathcal{L}_{[\xi_{A},\mathcal{V}_{(+)}]}g^{(0)}) \right] \\ = \frac{\tilde{C}_{T}}{a^{*}} \left[\int_{\Sigma_{A}} \omega_{\text{grav}}(h,\mathcal{L}_{\xi_{A}}) + \int_{\tilde{A}} \chi(h,[\xi_{A},\mathcal{V}_{(-)}]) + \chi(h,[\xi_{A},\mathcal{V}_{(+)}]) \right]$$
(8.218)

where we used that V vanishes at the boundary of AdS and that the past and future horizons share a boundary at $l_B^+ = l_B^- = 0$ which is precisely the RT surface \tilde{A} .

Finally, we need only define a vector field V such that

$$V^{+}|_{\tilde{A}} = \mathcal{V}^{+}_{(+)}|_{\tilde{A}} , \quad \partial_{-}V^{+}|_{\tilde{A}} = \partial_{-}\mathcal{V}^{+}_{(+)}|_{\tilde{A}} V^{-}|_{\tilde{A}} = \mathcal{V}^{-}_{(-)}|_{\tilde{A}} , \quad \partial_{+}V^{-}|_{\tilde{A}} = \partial_{+}\mathcal{V}^{-}_{(-)}|_{\tilde{A}}$$
(8.219)

and note that the combination of the conditions equation (8.217) and their $\mathcal{V}_{(-)}$ analogs precisely imply the Hollands-Wald constraints on V. Thus, we can write the CFT result as

$$\delta_{\epsilon}^2 S(\rho_A || \rho_A^{(0)}) = \frac{\tilde{C}_T}{a^*} \left[\int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A}) + \int_{\tilde{A}} \chi(h, [\xi_A, V)] \right]$$
(8.220)

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

8.4.3 Combining the Results

We found our main gravitational and CFT results

$$\delta_{\epsilon}^{2}(E^{\text{grav}} - S^{\text{grav}}) = \int_{\Sigma_{A}} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_{A}}h) + \omega_{\phi}(\delta_{\epsilon}\phi, \mathcal{L}_{\xi_{A}}\phi) - 2\xi_{A}^{a}\left(\delta_{\epsilon}^{(2)}E_{ab}\right)\epsilon^{b} + \int_{\tilde{A}}\chi(h, [\xi_{A}, V])$$
$$\delta_{\epsilon}^{2}S(\rho_{A}||\rho_{A}^{(0)}) = \int_{\Sigma_{A}}\omega(\delta\phi, \mathcal{L}_{\xi_{A}}\delta\phi) + \frac{\tilde{C}_{T}}{a^{*}}\left[\int_{\Sigma_{A}}\omega_{\text{grav}}(h, \mathcal{L}_{\xi_{A}}) + \int_{\tilde{A}}\chi(h, [\xi_{A}, V])\right].$$

$$(8.221)$$

The CFT result carries through for an arbitrary ball shaped region A, in an arbitrary Lorentz frame. This is manifestly true of the gravitational result as well. Setting $\tilde{C}_T/a^* = 1$ and using that the results are supposed to agree we find

$$\int_{\Sigma_A} \xi_A^a \Big(\delta_\epsilon^{(2)} E_{ab} \Big) \epsilon^b = 0 , \qquad (8.222)$$

which is precisely the (integral of) the EFE to second order with stress tensor source terms. By arbitrariness of the choice of Σ_A , choice of the ball-shaped region and Lorentz frame this integral equation implies the Einstein equations locally to second order (for example this claim can be proven as we did in the first order case, by induction).

We have thus proven that the unique geometry that correctly computes the CFT ball entanglement entropies via the Ryu-Takayanagi formula satisfies the Einstein equations to second order given that the central charges of the CFT satisfy $\tilde{C}_T = a^*$. Since this calculation assumes no *conjectures* of the AdS/CFT correspondence it also serves as a check on the conjecture itself. In a general CFT we can not expect that $\tilde{C}_T \neq a^*$, so a generalization is necessary. Such a generalization was carried out in [78] by generalizing the gravitiational identities to higher curvature gravity. In [79] our choice of 'single trace' Euclidean operators as the CFT states is generalized. These more general nonlocal multi-trace operators are argued to capture more general entanglement structures, such as entanglement between two-points in the bulk, without spoiling the semiclassical geometry. We will learn of the importance of such a generalization to capture bulk quantum corrections in chapter 9.

Another very interesting direction is to take this computation to third order in perturbations. This will result in the presence of two new ingredients on the CFT side, namely three-point functions coming from terms proportional to $(\delta \rho)^3$ and operator product expansions coming from contact terms in the second order variations $\delta^2 \rho$. Some first steps on the third order problem will be taken in chapter 11. In principle, matching the gravitational result may put interesting restrictions on the forms of the OPE and three-point functions of the boundary CFT.

Chapter 9

Entanglement is Not Enough

In this chapter we will see that just entanglement is not enough to build up spacetime. This was hinted already by the fact that black hole interiors are not penetrated by any extremal surfaces. An additional troublesome fact is that entangled particle pairs (so-called EPR pairs) provide a counterexample to the fact that proximity and entanglement seem related as in the argument with mutual information and the geodesic approximation of section 7.4.

To resolve both of these at once we will turn our eyes towards black holes. More specifically, we will consider the holographic dual of the maximally extended AdS-Schwarzschild spacetime [67]. In 2001 Maldacena found that the holographic dual of the maximally extended AdS-Schwarzschild spacetime consists of two copies of the same CFT, living on the boundaries of each of the exterior spacetimes on the two sides of the central black hole. Moreover the state is entangled in such a way that the entropy of each of the two exterior spacetimes is cancelled. This construction relates to the seeming difference between the nature of the entanglement on the AdS and CFT sides of the AdS/CFT duality in section 7.1 and it turns out that both entropies were entanglement entropies the all along.

The significance of Maldacena's result in relation to the Ryu-Takayanagi formula was realized in $[7]^1$. Here Maldacena and Susskind point out two facts:

- 1. Entangled black holes share many properties with entangled particle pairs.
- 2. The maximally extended AdS-Schwarzschild spacetime is not time-translation symmetric, the black hole interior (and thus the length of the wormhole connecting the two exterior AdS spaces) grows. On the CFT side, this is captured by the fact that the relative phases between energy eigenstates are changing into an increasingly complex combination. In the mean time the throat area of the wormhole remains constant. While Ryu-Takayanagi accounts for the throat area, it can not account for the increasing length.

Based on these two points, they conjecture that

- 1. Entanglement = Wormholes (or ER = EPR)
- 2. The length of the wormhole is dual to some state-dependent quantity on the CFT side.

Additionally they argue that the uniqueness of the map between spacetimes and entangled states requires that this state dependent quantity is sensitive to the interior of the black hole. This

¹Technically Raamsdonk preceded this, since the disentangling experiment of section 7.4 is inspired precisely by [67]. The ER=EPR paper [7] can be seen as a continuation of this.

provides a resolution of both the EPR pair counterexample and the problem of describing black hole interiors holographically.

The holographic dual of the wormhole length is nowhere near as precisely understood as the Ryu-Takayanagi relation, but recent evidence points towards *computational complexity* [80]. Very recent papers by Raamsdonk et al. [81] and de Boer [82] et al. have also converged towards the microstate dependent picture, including tests of the complexity conjectures.

In this chapter, we will investigate all these things in some detail with the sections following essentially the narrative structure of the previous paragraphs.

9.1 Eternal Black Hole Equals Wormhole

Let us consider the Ads-Schwarzschild black hole in d + 1 dimensions. While this text in general has been very brief about general relativistic input, this particular example is both very important, remarkably tractable and incredibly interesting. This calculation is a messier AdS analog of the analysis in chapter 7 of [45].

The AdS-Schwarzschild metric is given by

$$ds^{2} = -f(r)dt^{2} + f^{-1}(r)dr^{2} + r^{2}d\Omega_{d-1}^{2} , \qquad (9.1)$$

where

$$f(r) \equiv 1 - \frac{\mu}{r} \tag{9.2}$$

and μ is related to the black hole mass. The black hole is called eternal because the metric is time independent, meaning the black hole will (classically) continue to exist for all eternity. The metric is singular at the center of the black hole, as well as at the roots of f(r) = 0. The larger root r_H of f(r) = 0 is the event horizon.

The black hole solves the vacuum equations of motion, so the curvature should be constant everywhere (except possibly at isolated points). This means that it is expected that the singularity of the metric at the horizon of the black hole might just be a coordinate singularity.

The singular behaviour is encoded in the form function f(r). A prototypical attempt at finding a coordinate system that removes the singularity at the horizon is

$$\rho = r - r_H \ . \tag{9.3}$$

Then, we see that

$$f(\rho) = 1 + \frac{(\rho + r_H)^2}{L^2} - \frac{\mu}{\rho + r_H}$$

= $1 + \frac{\rho^2 + 2r_H\rho + r_H^2}{L^2} - \frac{r_H}{\rho + r_H} \frac{\mu}{r_H}$
= $\mu + \frac{\rho^2 + 2r_H\rho + r_H^2}{L^2} + \frac{\rho}{\rho + r_H} \frac{\mu}{r_H} - \frac{\mu}{r_H}$
= $\frac{\rho^2 + 2r_H\rho}{L^2} + \frac{\rho}{\rho + r_H} \frac{\mu}{r_H}$, (9.4)

where in the last step we used that r_H is a root of the original f(r). We see that $f(\rho)$ is regular for all $\rho > -r_H$ including what used to be a singularity at the event horizon at $\rho = 0$. There is now only a singularity at $\rho = -\mu$ corresponding to r = 0. Note that the coordinate change cannot have changed how causality works, so if we try to compute timelike geodesics starting at $\rho < 0$, they will never pass $\rho = 0$ and exit the black hole. It was observed by Einstein and Rosen in 1935 that there is nothing preventing us from picking the coordinate

$$u^2 \equiv \rho = r - r_H , \qquad (9.5)$$

setting the form function to

$$f(u) = \frac{u^4 + 2r_H u^2}{L^2} + \frac{u^2}{u^2 + r_H} \frac{\mu}{r_H} .$$
(9.6)

For real u, these coordinates cover only the region outside of the black hole horizon. The metric under this change of coordinates, since 2udu = dr, is given by

$$ds^{2} = -f(u)dt^{2} + \frac{f^{-1}(u)}{4u^{2}}du^{2} + (u^{2} + \mu)^{2}d\Omega_{2}^{2}.$$
(9.7)

Let us think about the validity of our change of coordinates. It is clear that our coordinate u can take two values for each value of $r - r_H$ in the previous system of coordinates, so for the change of coordinates to be bijective we should have specified that we take the positive root of $u^2 = r - \mu$ to specify the coordinate u. There is nothing preventing us from analytically extending the domain of u from $[0,\infty]$ to $[-\infty,\infty]$ since the metric is completely regular for all such values. We expect the analytic continuation to be unique since u < 0 still has to solve the Einstein field equations, given the metric at the horizon of the black hole as a boundary condition.

This analytic extension could be seen as the opposite operation to the case where we make a change of coordinates such that the new coordinates do no cover the entire spacetime. Examples of such coordinate changes are when we went from a spacetime containing a causal diamond to a hyperbolic cylinder whose coordinates only covered the interior of the diamond in equation (4.183), as well as the change from ρ to u^2 that removed the black hole interior from our coordinate chart.

Since general relativity is supposed to be coordinate invariant, we should try to define what spacetime we are looking at in a coordinate invariant manner. It seems reasonable that if we start from some system of coordinates, the 'full' physical spacetime that we are describing should correspond to some 'maximal' analytic extension of the original one. It turns out that the correct prescription is to demand that the spacetime contains no geodesics that have endpoints except at curvature singularities. A spacetime that fulfills this can not be analytically extended any further.

Having extended the domain of u to $[-\infty, \infty]$ we should try to understand what the metric in equation (9.7) represents. The spacetime is symmetric under $u \to -u$, and we know that u > 0 represents AdS outside a black hole. The parity symmetry in u has a fixed point at u = 0, so we should investigate this region more closely. The radial part of the metric blows up at u = 0, while $f(u) \sim u^2$. For a photon moving in the radial direction we see that trying to approach u = 0 from some small, positive value u = R will take infinite coordinate time. We use $dt^2 = \frac{f^{-2}(u)}{2u^2} du^2$ to write

$$\int_{0}^{T} dt = \int_{R}^{\epsilon} du \frac{f^{-1}(u)}{2u} \sim \left(\frac{1}{\epsilon^{2}} - \frac{1}{R^{2}}\right) .$$
(9.8)

This diverges as $\epsilon \to 0$, so we conclude that crossing takes infinite coordinate time on both sides of the horizon. In this sense, the two copies of AdS are separated by a two-way event horizon, meaning nothing can move from one copy to the other in finite time. We will later find that this is due to the fact that the wormhole connecting the two horizons grows with time.

Finally, we note that for u = 0, the $d\Omega_2^2$ term in the metric describes a sphere of radius r_H . In these coordinates, the spacetime thus seems to consist of two copies of the AdS-Schwarzschild exterior, connected across a spherical two-way event horizon of radius r_H . This is an *Einstein-Rosen* bridge or wormhole, connecting two asymptotically AdS spacetimes to each other. We can bring this construction back to our single universe in the following way: the regions u > 0 and u < 0 describe two AdS-Schwarzschild solutions containing a black hole. We can make a bijective coordinate change to the radial coordinates r and r', corresponding to u > 0 and u < 0, both with metric as in equation (9.1). Then if we take the two black holes to be so far apart that we can neglect their attraction to each other, they still solve the equations of motion. Going back to u and identifying the event horizons, we have a spacetime containing two black holes a large distance apart that share a horizon. This seems to imply that there is a shortcut through spacetime across this horizon, but as we saw the horizon cannot be traversed in a time that looks finite to the exterior universe. Note that this solution is not exactly the same as a solution containing only one black hole is unstable to a black hole merger.

Let us finish with a comment about entanglement before moving on to the maximally extended black hole. Remember that in the Rindler spacetime of section 4.2.1 we had two complementary wedges that were causally disconnected. These wedges lived in a full Minkowski spacetime, and the local state on the two complementary wedges looked exactly the same, while the overall state was just the pure Minkowski vacuum. Here we have a similar situation. The parity operator J_{Ψ} that related the two wedges can readily be seen to be identifiable with $u \to -u$, and we know from section 4.2.3 that by the equivalence principle the near horizon region looks like the Rindler spacetime for an observer at fixed u.

Using our experience with axiomatic field theory in chapter 4, we can argue for the purity of the extended spacetime. The method is to make the ansatz that there exists a quantum wavefunction $|\Psi\rangle$ that respects the basic axioms of AQFT and checking that it leads to no inconsistency. We then consider two points:

- Both exterior spacetimes have Hilbert spaces with the same dimension since they are the same.
- Both exterior spacetimes are in thermal states, with the same spectrum. Their respective density operators are invertible, owing to the fact that $e^{-\beta H} > 0$.

These are exactly the properties of two entangled subsystems that are obtained by tracing out half of a cyclic, separating pure state $|\Psi\rangle \langle \Psi|$. We found these properties in the discussion following equation (4.25). Therefore, we conclude that the existence of a pure state $|\Psi\rangle$ is consistent with what we know of the quantum behaviour of the maximally extended Schwarzschild spacetime. The topic of [67] is determining the exact form of the relevant pure state.

In the case where we have put the two black holes in the same spacetime, we may intuit that the black holes being connected via a wormhole seems to correspond to the black holes being entangled with each other. This is the seed that leads to the ER=EPR conjecture, as we will discover soon.

Maximally Extended AdS-Schwarzschild Black Hole

In the previous section, we made a change of coordinates for the eternal AdS black hole and discovered that there lives a whole new spacetime on the other side of the black hole. The price we paid was that our coordinate u did not cover the interior of the black hole. We will now proceed to do two things in parallel, we will find coordinates that include both copies of AdS as well as the black hole interior, and we will analyze the causal structure of the two-sided black holes by considering the shape of lightcones according to various observers.

Let us begin by restating f(r):

$$f(r) \equiv 1 - \frac{\mu}{r} + \frac{r^2}{L^2} .$$
(9.9)



Figure 9.1: The opening angle of the lightcone at various radii, as measured by an observer staying at some constant radius, with proper time τ . According to an observer in the exterior universe radial motion becomes impossible at the event horizon. $d\tau/dr = 1$ corresponds to an angle of 90°, while $d\tau/dr = \infty$ corresponds to a closed up lightcone.

To understand the black hole spacetime it is instructive to consider objects falling radially into a black hole. Imagine that an observer stays still at some finite radius R and observes this object falling in. For such an observer, dt is directly proportional to proper time. This observer measures the shape of the lightcone of the infalling object in terms of their own proper time $d\tau$, which they know how to relate to to dt.

It is readily checked that equation (9.1) tells us that for light,

$$\frac{\mathrm{d}t}{\mathrm{d}r} = \pm f^{-1}(r) \ . \tag{9.10}$$

The two roots $\pm f^{-1}(r)$ tell us the shape of the future pointing lightcone of an object at radius r as measured by some observer sitting at a constant radius far away. The situation as seen by this stationary observer is illustrated in figure 9.1. Most importantly, as we approach the event horizon the lightcone closes up and radial motion becomes impossible.

Let us now consider the experience of an observer falling into the black hole. By the equivalence principle, there is a coordinate system that is locally flat along the trajectory of this observer. This means that we can pick coordinates such that the opening angle of the lightcone is always 90°. That is, we want a coordinate r^* such that $dt/dr^* = 1$ for all r^* . By inspecting the metric, this is achieved by an r^* that fulfills

$$dr^* = f^{-1}dr . (9.11)$$

The inverse form function f^{-1} makes for a rather unpleasant integrand, but we do not have to perform the integration explicitly. We note that f^{-1} diverges at $r = r_H$, so the new coordinates cannot cover $r > r_H$ and $r < r_H$ simultaneously. If we are to associate a finite value of r^* to coordinates at some finite distance outside the event horizon, we must put the event horizon at $r^* = -\infty$. This we can do since r^* is only defined up to a constant. Note that since there is an r^2 in f(r), the upper limit of r^* is finite. The metric is now of the form

$$ds^{2} = f(r(r^{*})) \left(-dt^{2} + dr^{*} \right) + r(r^{*})^{2} d\Omega_{2}^{2} .$$
(9.12)

An interesting note is that the metric is completely regular as $r^* \to -\infty$, since we have removed the divergent factor f^{-1} via the change of coordinates. Since $dt/dr^* = 1$ for lightlike radial trajectories,

this system of coordinates describes the lightcone of an observer at radius $r(r^*)$ as seen by that observer themself. We need some more work before these coordinates are useful, since they do not cover the black hole interior.

The next step is to pick lightcone coordinates², defined by

$$u = t + r^*$$
, $v = t - r^*$, (9.13)

in which metric becomes

$$ds^{2} = -f(r(u,v))dudv + r(u,v)^{2}d\Omega_{2}^{2}.$$
(9.14)

The coordinate u corresponds to outgoing null geodesics, and v corresponds to ingoing null geodesics. Since the lightcone coordinates are orthogonal, lines of constant u determine the direction of v and vice versa.

Before continuing, it would be nice to repeat figure 6.2 with the same radial coordinate, but with a time axis adapted to the infalling observer. This will show us that from the perspective of an infalling observer, the black hole horizon may be passed in finite time.

Let us define $t^* = u - r$. In the coordinate system (t^*, r) . The line of constant u will always point in a 45° angle inwards from the line of constant r. The line of constant u determines the direction of the ingoing null geodesic. To analyze the other side of the lightcone, i.e. the outgoing null geodesic, we need to relate lines of constant v to the radial coordinate r. We know how the lightcone behaves in terms of the old coordinate 2t = u + v. We have

$$2\frac{dt}{dr} = \pm 2f(r)^{-1} = \frac{du}{dr} + \frac{dv}{dr} = \frac{dv}{dr} + \frac{dt^*}{dr} + 1 .$$
(9.15)

We are looking for curves of constant v, so we can set dv = 0. We see that lines of constant v are given in these coordinates by

$$\frac{\mathrm{d}\tilde{t}}{\mathrm{d}r} = 2f^{-1}(r) - 1 , \qquad (9.16)$$

where we have picked the positive sign on f^{-1} since that is the solution corresponding to the outgoing null geodesic. It is readily seen that the outgoing null geodesic points straight up near the horizon since f^{-1} diverges, and that at very large radii it becomes quite wide, an artifact due to the asymptotic r^2 behaviour of the form function, close to the horizon t^* behaves more like a proper time coordinate. This is fine, since we are only interested in the experience of infalling observers near the horizon.

This all happens while the ingoing null geodesic stays at a constant angle, so in these coordinates, the lightcone tilts and narrows, but never closes completely. Crucially we see that at the event horizon, no timelike geodesics point out from the black hole, although they can point inwards. The new situation is illustrated in figure 9.2. Similarly, we may define a time coordinate $\tilde{t} = v + r$. Lines of constant v always point 45° outwards. The lines of constant u are determined in the same way as before, this time picking a negative sign on f^{-1} , so that

$$\frac{\mathrm{d}\tilde{t}}{\mathrm{d}r} = -2f^{-1}(r) + 1 \ . \tag{9.17}$$

By the same reasoning as before, the lightcone behaves as in figure 9.3. Notably, at the horizon, all future timelike geodesics point *away* from the singularity. In these coordinates, it seems we are not approaching the same object as before. In fact, we are approaching a *white hole*, whose event horizon cannot be entered, only exited.

²Historically u/v are called ingoing/outgoing Eddington-Finklestein coordinates.



Figure 9.2: The shape of lightcones at various radii r as measured by an observer falling into the black hole. The lightcone tilts and closes up, and at the horizon all timelike trajectories point towards the singularity at r = 0.



Figure 9.3: The shape of lightcones at various radii r as measured by an observer falling towards the event horizon. The lightcone tilts and closes up, and at the horizon all timelike trajectories point away the singularity at r = 0. The conclusion is that this cannot be the object that was approached by the observer using $\tilde{t} = u - r$ as a time coordinate. This object is in fact a *white hole*, where all timelike geodesics point away from the singularity.

The lightcone coordinates u and v can be put on a form that is readily extended to include the full spacetime. We can achieve this by picking the coordinates

$$\tilde{u} = e^{ku/2}, \quad \tilde{v} = -e^{-kv/2},$$
(9.18)

where k is a constant to be determined later. These coordinates are called the Kruskal-Szekeres lightcone coordinates. As defined, we have $\tilde{u} \in [0,\infty]$ and $\tilde{v} \in [-\infty, 0]$. The metric takes the form

$$\mathrm{d}s^2 = -f(r)\frac{1}{k^2\tilde{u}\tilde{v}}\mathrm{d}\tilde{u}\mathrm{d}\tilde{v} + r^2\mathrm{d}\Omega_2^2 , \qquad (9.19)$$

where we let the fact that r is a function of \tilde{u} and \tilde{v} remain implicit. Note that while the metric is singular at the event horizon $(r^* = -\infty \rightarrow \tilde{u} = \tilde{v} = 0)$, it is well defined for $\tilde{u} < 0$ and $\tilde{v} > 0$. We also know that the horizon singularity is just a coordinate singularity, so we can without trouble analytically extend our solution. The extension cannot continue beyond the curvature singularity at r = 0, because analyticity breaks down here independent of the choice of coordinates. This turns out to be the *maximal* extension of the AdS-Schwarzschild spacetime.

The coordinate r is implicitly defined by

$$\tilde{u}\tilde{v} = -e^{2kr^*(r)} . (9.20)$$

After extending the domains of \tilde{u} and \tilde{v} we see that the LHS is unchanged under the change $(\tilde{u}, \tilde{v}) \rightarrow -(\tilde{u}, \tilde{v})$. This means that the tilded coordinates cover the original black hole spacetime twice. To make the Kruskal-Szekeres coordinates properly extend to the interior of the black hole, we have to pick k in a specific manner. When the product of $\tilde{u}\tilde{v}$ is positive, equation (9.20) tells us that r^* should have an imaginary part $i\pi/k$. It should be possible to check by performing the integration and solving equation (9.11) explicitly for $r^*(r)$ that

$$\tilde{u}\tilde{v} = e^{2kr^*} \sim (r - r_H) , \qquad (9.21)$$

so that we are taken inside the event horizon exactly as the product $\tilde{u}\tilde{v}$ becomes positive. The constant k must be chosen so that the relation becomes exact to avoid a coordinate singularity at the horizon. It turns out that r^* is explicitly given by

$$r_* = \int \frac{1}{f} dr = \frac{(2r_+^2 + 4p) \arctan\left(\frac{r_H + 2r}{\sqrt{4p - r_H^2}}\right) + r_H \sqrt{4p - r_H^2} \ln\left(\frac{(r - r_H)^2}{r_+^2 + r_H r + p}\right)}{2c^2(2r_+^2 + p)\sqrt{4p - r_+^2}} + C , \qquad (9.22)$$

where $p = r_H^2 + L^2$ and C is an integration constant. The correct choice of k then end up as

$$k = \frac{2r_H^2 + p}{r_H L^2} . (9.23)$$

Using the known form of r_* and choosing C appropriately one can then show that

$$\lim_{r \to 0} \tilde{u}\tilde{v} = 1 ,$$
$$\lim_{r \to \infty} \tilde{u}\tilde{v} = -1 ,$$
$$\lim_{r \to r_H} \tilde{u}\tilde{v} = 0 .$$

That is, the horizon(s) lie at $\tilde{u} = 0$ and $\tilde{v} = 0$, the singularity lies on the curve defined by $\tilde{u}\tilde{v} = 1$ and the boundary of AdS lies at $\tilde{u}\tilde{v} = -1$.



Figure 9.4: Penrose diagram of the maximally extended AdS-Schwarzschild spacetime. The spacetime boundaries are indicated in grey and the event horizon with straight black lines.

The full spacetime can be split into four regions, F,P,L,R according to the signs of \tilde{u} and \tilde{v} as in figure 9.4. Figure 9.4 is a Penrose diagram of the spacetime, and we have denoted the boundaries at $\pm \infty$ of \tilde{u}, \tilde{v} as well as at r = 0. Note that r = 0 should correspond to a region where $\tilde{u}\tilde{v} = const$. We take the rotated $\tilde{u} - \tilde{v}$ axis to cross at zero. Here, R denotes the region corresponding to the black hole exterior covered by our old lightcone coordinates. This should be clear by considering that before the extension, the domain was $\tilde{u} \in [0, \infty]$ and $\tilde{v} \in [-\infty, 0]$.

The region F is the future interior of the eternal black hole, and the line of r = 0 is called the future timelike singularity. The shape of the r = 0 line is found by the following consideration: r depends directly on r^* , and r^* is constant for curves of constant $\tilde{u}\tilde{v}$.

Positive \tilde{v} and negative \tilde{u} cover the region L, and they correspond to the negative u solution of the wormhole coordinates of the previous section. Indeed, since we have a symmetry of the metric under $(\tilde{u}, \tilde{v}) \to -(\tilde{u}, \tilde{v})$, L and R should describe the same spacetime.

9.1.1 Holographic Dual of the Extended Spacetime

We just found out that the AdS-Schwarzschild solution is actually just an incomplete coordinate chart for a bigger system. The extended AdS-Schwarzschild spacetime contains two singularities, at the centers of black- and white holes respectively. In addition, it contains two copies of the AdS-Schwarzschild exterior. We will now consider the holographic dual of the maximally extended AdS-Schwarzschild spacetime.

We discovered that the thermodynamical properties of the spacetime matched our expectation for the behaviour of a bipartite system in a cyclic, separating pure state $|\Psi\rangle$. This state can be written as $|\Psi\rangle_L \otimes |\Psi\rangle_R$. The Hilbert space of a QFT living on a connected, continuous spacetime does not admit such a factorization. This leads us to consider a CFT defined on two disconnected spacetimes. These spacetimes must be the same to fulfill the cyclic condition for local operators. The CFT dual of the AdS_{d+1} -Schwarschild solution is a CFT on $\operatorname{S}^1 \otimes \operatorname{S}^{d-1}$. For a single copy of the exterior spacetime this CFT dual is uniquely determined, so we only have one possible candidate for the dual of the maximally extended AdS-Schwarzschild spacetime. The maximally extended spacetime must be dual to two copies of the same CFT on $\operatorname{S}^1 \otimes \operatorname{S}^{d-1}$ entangled in such a way that the overall state is pure.

The purification of a system using another copy of the same system is unique if the spectrum of the Hamiltonian is non degenerate. The unique state defined on subsystems L,R that gives the correct local density operator on L and R is given by

$$|\Psi\rangle_{LR} = \frac{1}{\sqrt{Z(\beta)}} \sum_{i} e^{-\frac{\beta}{2}H} |E_i\rangle_L \otimes |E_i\rangle_R \quad , \tag{9.24}$$

where $|E_i\rangle_L$ are eigenstates of H with energy E_i and $Z(\beta)$ is the partition function of a single copy of the system (necessary to normalize the wavefunction). Constructing the density operator we obtain

$$\rho_{LR} = \frac{1}{Z(\beta)} \sum_{i,j} e^{-\beta(E_i + E_j)/2)} |E_i, E_i\rangle \langle E_j, E_j| \quad .$$
(9.25)

If we trace out either subsystem we get the usual form of a thermal density matrix with standard normalization

$$\operatorname{Tr}_{L}[\rho_{LR}] = \frac{1}{Z(\beta)} \sum_{i} e^{-\beta E_{i}} |E_{i}\rangle\langle E_{i}| \quad .$$

$$(9.26)$$

The state described by equation (9.24) is called the *thermofield double*. This state has three interpretations, all of which we have seen at this point. The second and third interpretations are the ones of importance for the ER=EPR conjecture. In the following we will refer to the wormhole coordinates of the previous section.

• equation (9.24) is a representation of a single black hole in thermal equilibrium. Let us take R to be the single black hole. We then see the second spacetime as nothing but an auxiliary fictitious addition to simplify calculation. The time evolution of the full system is taken to be generated by the fictitious thermofield Hamiltonian

$$H_{tf} = H_R - H_L. (9.27)$$

The thermofield Hamiltonian generates forward time translation on the right side of the Penrose diagram and backwards translation on the left side. This "thermofield doubling" of the Hilbert space simplifies the calculation of correlation functions between operators in the right spacetime. This interpretation is the oldest one, and is the origin of the term "thermofield double". Once the maximally extended Schwarzschild solution was understood in the semiclassical sense, it was realized that this Hamiltonian corresponds to the quotient space maximally extended AdS-Schwarzschild/ \mathbb{Z}_2 where the \mathbb{Z}_2 identifies $(\tilde{u}, \tilde{v}) \sim -(\tilde{u}, \tilde{v})$.

• equation (9.24) represents two entangled spacetimes connected at the horizon of a black hole. Here, there is no quotient so time moves upwards on both sides of the Penrose diagram. In this interpretation, the Hamiltonian is written as

$$H = H_R + H_L \tag{9.28}$$

and the state described by equation (9.24) represents two black holes at some time t = 0. The state is not an eigenstate of this Hamiltonian, as its evolution is given by

$$|\Psi(t)\rangle_{LR} = \sum_{i} e^{-\beta E_n/2} e^{-2iE_n t} |\overline{n}, n\rangle \quad , \tag{9.29}$$

where $|\overline{n}\rangle$ is related to $|n\rangle$ by the exchange $\tilde{u} \leftrightarrow \tilde{v}$, which is a realization of the CPT symmetry of the extended spacetime. Since the full spacetime is pure, the two spacetimes must be entangled. In fact, they are *maximally* entangled in the sense that the entire entropy of the local states is due to entanglement with another system. The local density operators are still time independent, so the *amount* of entanglement is time translation invariant. The lack of time translation symmetry in the relative phases of the eigenstates has a counterpart on the gravitational side; starting with a time slice at t = 0 and evolving it upwards in the spacetime diagram 9.4 reveals that the interior of the black hole grows with time.

• equation (9.24) represents two far-separated, highly entangled black holes in a single spacetime. These black holes initially have identified event horizons, and are thus connected by a wormhole. AdS is unstable to black hole formation (in this case the merging of the two black holes) so the approximation does not hold for infinite time. Since we can place the black holes arbitrarily far apart, the approximation of two independent black hole can be made to hold for an arbitrary, large amount of time. Note that if we wish to apply the AdS/CFT duality, with a regularization capping off the radial coordinate at some finite distance ρ_0 , we would like the separation between the black holes d_{sep} to fulfill $d_{sep} \ll \rho_0$ so that the boundary CFT is far from both black hole singularities.

We conclude that the construction is sensible on the gravitational side for an arbitrarily large, finite time. The state also seems to fulfill the basic properties required for AdS/CFT duality, namely rotational symmetry in the large r limit and asymptotically AdS behaviour (see section 6.2.3).

On the CFT side we should understand how to handle the fact that the state describing the two black holes is supposed to factorize despite being defined on a single spacetime, as well as the finite entanglement of the black holes. Neither of these violate the axioms of local QFT if the the corresponding boundary states are smeared over the entire boundary spacetime. If we treat the two black holes independently of eachother, they should both be determined by a CFT state on the boundary of their past lightcone. Given that we know the Hamiltonian of the boundary theory, we can relate the two boundary states that determine the two black holes by a unitary time evolution. The entanglement of the two black holes then corresponds to entanglement entropy between a region and itself, so it is not subject to the entanglement divergences of local QFTs. This is illustrated in figure 9.5.

Another potential subtlety is the required periodicity of the Euclidean time circle. Near horizon analysis is still exactly the same, meaning the Euclidean time periodicity is $\tau \sim \tau + \frac{4\phi L^2 r_H}{(d+1)r_H^2 + (d-1)L^2}$ near the two black hole horizons, but it should not be periodic at the boundary. To resolve this we may think about the fact that these black holes are supposed to radiate. The radiation they emit must be entangled with the radiation from the other black hole in such away that their radiation interferes into a pure state for large radii. Since the entropies are constant, this should correspond to some stationary solution of the Einstein field equations where the radius of the Euclidean time circle is allowed to vary in space with a divergent boundary condition at infinity. It should be noted that letting the period of the Euclidean time circle become dynamic is very tricky and calls into question a number of uniqueness results from chapter 6.2.3.

A different interpretation of this is that "thermal equilibrium" does not make sense in this spacetime and we should not be using the periodicity in Euclidean time prescription. Indeed, when giving the two CFTs a common time, we observed that the pure entangled state was not an eigenstate of the Hamiltonian. However, since the state and subsystems keep constant



Figure 9.5: Two black holes in AdS, with the boundary of AdS represented as two black lines. By causality, the bulk black hole is completely determined by a boundary CFT state on its past lightcone. The dashed greyed lines represent the past lightcones of the black holes wrapping around the topologically cylindrical boundary of AdS. Since the gravitational state is pure in this construction, the two regions indicated by the dashed grey lines are related by unitary time evolution. We conclude that the entanglement of the black hole pair is dual to nonlocal entanglement between a region and itself on the CFT side.

entropy it is not clear that this is not thermal equilibrium. In addition to this, AdS is unstable to black hole formation so we might expect that given enough time any two $(large)^3$ black holes will merge, meaning there was never any equilibrium.

State of the Information Paradox

Way back in section 4.2.3 we encountered the *information paradox*, the essence of which is that gravity lets a system with no or little entropy evolve over time into a system with high entropy. This is in violation of the unitarity of quantum mechanics, since the von Neumann entropy is invariant under unitary transformations of the density operator.

This double black hole hints at a possible solution to the information paradox. Here, the radiation coming from the two black holes must be entangled in such a way that they interfere into a pure state at large radii, otherwise the boundary state cannot be pure. Based on this, we may speculate about the structure of unitary black hole formation.

For a single, one-sided black hole we might imagine a similar resolution, wherein the black hole becomes entangled with its own radiation. The problem with such a resolution is that the spacetime seems to stay impure for at least as long as it takes for the black hole to radiate away half its entropy. This time is called the *Page time*. Beyond this point, there is enough radiation in the exterior spacetime for an observer to determine the microstate of the remaining black hole by measuring the state of the radiation.

This hints that unitary black hole formation may require the emission of enough radiation for such an exterior-interior entanglement to be possible right away. This can be given a very interesting implication. Given the existence of a device that can collect all of this radiation and collapse it into a second black hole, we can create an entangled pair of black holes. We already know that black

³Small black holes would evaporate into a cloud of radiation before merging.

holes of this type are connected by a wormhole. If there was no geometrical connection between the radiated particles and the initial black hole, it interesting to ponder whether it makes sense for a wormhole to spontaneously form as you collect the entangled radiation. It was the conjecture of Maldacena and Susskind that this is not sensible, and we should think of each of the radiated particles as connected to the original black hole via a Planck-scale wormhole. This is the so-called ER=EPR conjecture.

9.2 Wormhole Equals Entanglement (ER = EPR)

We have in the previous section concluded that both on the AdS and CFT sides, we can put the two sides of the maximally extended spacetime into the same spacetime to obtain two distant regions connected by a wormhole. The CFT dual of this setup is some kind of maximally delocalized entanglement, and is therefore not described by an extremal surface that is connected to the boundary. When we have two disconnected spacetimes, there is a unique HRT surface not connected to the boundary, the wormhole event horizon, that computes the entropy of each of the CFTs on the copies of the AdS boundary. When we put the two wormholes in the same spacetime there is no obvious way of defining subregions to give boundary input to the HRT formula.

The ER=EPR conjecture of Susskind and Maldacena is that this type of nonlocal entanglement on the CFT side is actually dual to nontraversible wormholes. The area of the horizon at the throat of these wormholes is given by the usual entropy-area relationship

$$S = \frac{A_{WH}}{4G_N} . \tag{9.30}$$

We actually already outlined an argument for the ER=EPR conjecture in the paragraph preceding this section. In this section we would like to develop this a little further. This section focuses on a number of points, all heavily inspired by the original reference [7], with some points included from [83]:

- formation of maximally entangled black hole pairs,
- uniqueness relationship between wormholes and entangled states,
- further similarities between entanglement and wormholes,
- the AMPSS paradox, and its resolution via ER = EPR

The specific relationship between CFT states and distinct wormholes will be made clearer in section 9.3, where the complexity=volume and complexity=action conjectures are explored.

Formation of Entangled Black Holes in the Same Spacetime

Previously we made the suggestion that the formation of a single black hole requires the emission of enough radiation to purify the exterior state. This type of construction is not understood in detail, and not necessarily a correct understanding of what goes on.

One might worry that entangled state of the two black holes is very specific, meaning that the particular entangled state in equation (9.24) is very hard to reproduce. It is a result due to Strominger and Garfinkle that black hole pair creation in the presence of an electromagnetic field produces black holes in precisely these states [84].

Another argument comes from a limit of black hole fission. Black hole fission is a conjectured process in which a black hole splits into two daughter black holes. There are no classical dynamics

that describe such a process, but since there are no symmetries preventing the process it is possible quantum mechanically. This phenomenon is not quantitatively understood, but at least qualitatively there is no reason it not possible [83]. In fact a major motivation behind the well known weak gravity conjecture comes from the fact that we expect that black holes are able to split [85].

As an example a single black hole of mass M may split into two black holes of mass M/2. The Schwarzschild radius is proportional to the mass, and so the total area of the daughters is half that of the of the parent black hole. The process involves a decrease of entropy by -S/2 and is therefore very rare ($p(\text{event}) \sim e^{\Delta S}$). In fact the black hole is more likely to evaporate completely in the time it takes for it to split in half once, but nevertheless the process is possible.

Importantly these black holes before splitting may have a large behind-the-horizon region. This region exists either because they are part of a two-sided black hole, or because we are considering the maximal AdS-Schwarzschild/ \mathbb{Z}_2 one-sided black hole which has a bridge to nowhere behind its horizon [86]. When the black holes horizons split, we have no reason to expect the interior to split. The general idea of the splitting process is as follows. As a statistical fluctuation, the horizon is deformed into two almost disconnected lobes. At this point the deformed area is approximately half that of the original black hole, so the probability of such a fluctuation is roughly $e^{S/2}$ where S is the entropy of the original black hole. Classically the probability is zero, since the entropy carries an inverse power of \hbar .

We assume that the black hole degrees of freedom are well mixed at the time of this fluctuation, meaning the two lobes should be highly entangled. At this point a nonperturbative instanton allows for the complete splitting of the two black holes, after which the process proceeds classically and the black holes fly off on their merry way given that there is some repulsive force present, for example the black holes may carry electric charge. After splitting, each daughter black hole has a separate growing interior region behind the horizon. Due to entanglement the behind-the-horizon regions of the daughters are connected. More pictorially, the sequence of events (from left to right) is



where we have indicated the behind-the-horizon region as extending downwards.

More interesting is to consider the asymmetric fission of black holes. Consider a parent with mass M decaying into daughters with masses m and M - m. The entropy change is $\Delta S = -\frac{2m}{M}S$ to first order in m. If m is small this is clearly much more likely then the symmetric fission, since $p \sim e^{\Delta S}$. In fact remembering the definition of the Hawking temperature T_H we see that

$$e^{\Delta S} = e^{-\frac{2m}{M}S} = e^{-\frac{m}{T_H}}$$
(9.31)

just describes emission spectrum of small black holes via Hawking radiation. Again assuming that the black hole degrees of freedom were well mixed, the daughters are entangled and the post split classical evolution has the behind-the-horizons grow. In the limit of very small black holes there is an expectation that instead elementary particles are emitted due to consistency conditions related to the weak gravity conjecture. Since the entanglement characterized in the black holes by a growing behind-the-horizon region remains, it is reasonable view even particles as connected to the original black hole by some kind of limiting wormhole. Pictorially the situation has the following appearance after a few black holes have been radiated:



Uniqueness Relationship Between Wormholes and Entangled States

When the two halves of the thermofield double are given a common time, the state is not timetranslation invariant. Time evolution traces out a continuous family of states with the same entanglement entropies and energy spectra, but with different relative phases of the energy eigenstates. These states on the field theory side are distinct, so if there is a duality between entanglement itself and wormholes, the distinct states must correspond to distinct wormholes (with constant horizon area). We shall argue that time translation indeed takes us between distinct wormholes.

Consider the Penrose diagram of the maximally extended AdS-Schwarzschild solution in figure 9.4. To define a quantum state, we pick a spacelike slice such as t = 0. Hamiltonian formalism is only well defined in the boundary region, since the full spacetime has no forward timelike global translation symmetry. To translate the choice of boundary region on the CFT side to the bulk we may pick any spacelike surface that is anchored to the boundary spatial slice. The freedom of choice is due to a simple reason; two surfaces with the same domain of dependence have the same algebra of observables.

In figure 9.6 we indicate in grey the domains of dependence of various pairs of boundary times t_R, t_L on the boundary CFTs. These domains of dependence correspond to states at particular boundary times on the CFT side. The quantum algebra of observables in the bulk is determined by the causal structure of the domain of dependence of these boundary time slices. The claim is then that the boundary state at boundary times t_R, t_L is dual to some generally covariant functional of their domains of dependence. This region is referred to in the literature as a Wheeler-deWitt (WdW) patch.

The first two images are related by simultaneous time translation of both CFTs, while in the righthand image we have considered the case of only translating the left time forward by acting only with the unitary operator e^{iH_Lt} . The different regions illustrated are cut differently by the singularity⁴, so they are not related to each other by a change of coordinates. We conclude that a covariant functional of the WdW patch can distinguish between patches corresponding to different boundary states, giving a uniqueness relation between spacetime quantities and boundary states.

"Nonlocality" of Wormholes and Quantum Entanglement

Wormholes and entanglement share an important fundamental feature, namely a causality-preserving notion of nonlocality. For entanglement we realize that it does not enable superluminal communication because observers on two ends of an entanglement experiment need to exchange classical information to realize their states were entangled to begin with. In the case of wormholes, positive

⁴Notably the domain of dependence of the initial state at $t_L = t_R = 0$ is not cut by the singularity at all.



Figure 9.6: Bulk regions corresponding to Cauchy slices of the boundary spacetime. These regions are cut off by the singularity at the center of the black hole, represented by a thin curved line. The spacetime regions are cut differently by the singularity, ensuring that they are not related by a choice of gravitational gauge. We conclude that the wormhole described in each of these regions is distinguishable from wormholes described by the other regions. Similarly, on the CFT side the states corresponding to different choices of boundary time correspond to different entangled states, since the energy eigenfunctions have different relative phases.

energy conditions, which follow from demanding a CFT dual [16], imply that the wormhole necessarily has an event horizon that prevents causality violating shortcuts through spacetime. In this sense the two objects circumvent the problem of broken causality.

In chapter 7 we learned that the bulk dual of entanglement is area. Wormholes are by definition the only type of geometric object that both respects causality and is nonlocal. Thus, the only spacetime object that could conceivably be dual to an entangled particle pair in the bulk is a wormhole. This is a separate argument that this should hold for *any* EPR pair. As far as the boundary CFT is concerned, entangled black holes and entangled particle pairs in the bulk have a very similar type of nonlocal entanglement. It is therefore reasonable that the geometric dual of pair entanglement is a wormhole, even for individual particles. Note that the horizon area of the wormholes connecting an entangled pair with an entanglement entropy of one qubit is given by

$$A_{\rm WH} = \frac{4G_n\hbar}{c^3} , \qquad (9.32)$$

i.e. only four Planck areas. As such, the conjectured geometrical dual of individual entangled pairs is a highly quantum wormhole. It is possible that this wormhole can not be rigorously interpreted as a classical geometry unless there is a very large amount of entanglement, as in the case of a large black hole.

Neither wormholes nor entangled particle pairs can be created by so-called LOCC, standing for "Local Operations and Classical Communication". This means that two labs at spacelike separated locations A and B cannot use a combination of local operators on the form $\mathcal{O}_A \otimes \mathcal{O}_B$ to create this particular type of state.

On the quantum side this is well known. In fact, when defining a general quantum channel we demanded that classical communication cannot retroactively change the result of any quantum measurement. This directly implies that it cannot affect quantum entanglement. Local operators in A commute with any measurements performed on B, so they also cannot change the entanglement. This was one of the fundamental properties we proved about quantum channels.

On the gravitational side, we have suggested several for the production of wormholes. All of the

mechanisms we have proposed begin with a single black hole that goes on to radiate matter with which it is entangled. This is "nonlocal" in the sense that particles that are entangled must at one point have been part of the same black hole. Thus an operator that creates wormholes must create both systems 'A' (the black hole) and 'B' (the radiation) together. Such an operator is 'nonlocal' in the very specific sense that it has to act on both system A and B at the same time, but there is nothing preventing these systems from being in the same location at some point in time so that the relevant operator may be spacetime local.

Additional recent results are the equivalence between the no-cloning theorem on the CFT side and no-go theorems for topology change on the gravity side, as well as a conservation theorem for the throat area of the wormhole (corresponding to conservation of entanglement) [87].

To summarize, both entangled pairs and wormholes need to be created in a single place and then spatially separated. Neither of the two may be created without an operator that acts simultaneously on subsystems 'A' and 'B' in the preceding discussion. We conclude that both wormholes and entangled pairs are states that cannot be created by LOCC that violate locality but not causality. In essence, together with the Ryu-Takayanagi formula this is the basic motivation for the ER = EPRconjecture. A final argument comes from the resolution of the AMPSS paradox, as we will now detail.

AMPSS paradox and its resolution

The AMPSS paradox is named after the authors Almheiri, Marolf, Polchinski, Stanford and Sully [88, 89]. It details the tension between the following three axioms of black hole physics

- 1. Hawking radiation is in a pure state
- 2. The information carried by the radiation is emitted from the region near the horizon, and low energy effective field theory is valid beyond some microscopic distance from the horizon.
- 3. The geometry is smooth at the horizon, and nothing in particular happens to an infalling point observer.

Before AMPSS, these axioms were supposedly reconciled by "black hole complementarity" [90]. We shall give a bare-bones account of complementarity, AMPSS and its resolution by ER=EPR. Arguments that here seem a little thin are expanded in [90, 88, 89].

The basis of the axioms are (1) unitarity of quantum mechanics, (2) locality of physics and (3) the classical fact that the horizon is not distinguished by curvature. We can already anticipate that ER=EPR does away with (2) while we will see that AMPSS argued that the most elegant way to conserve quantum mechanics is by discarding (3). The most basic contradiction that arises from the axioms is that if the geometry is smooth, an infalling observer experiences nothing in particular and goes on to exist inside the horizon. This observer is then somehow also encoded in the later pure Hawking radiation, violating the no-cloning theorem. To arrive at black hole complementarity one notes that such a comparison is not consistent since an observer that can measure the Hawking radiation in the exterior does not have the black hole interior in its past lightcone. This means that no single observer is able to observe any cloning.

The case of an infalling observer is more difficult to reconcile with no cloning. We imagine an experiment as in figure 9.7 where an observer comes from O, and at A creates an entangled pair of particles b - c. particle b is sent together with a timed measuring device straight into the horizon. The measuring device measures the state of b at B and sends a signal in the outgoing lightcone direction. In the meantime the observer brings with them the other half of the entangled pair a and waits outside the horizon for a while before entering. At C the exterior observer measures Hawking



Figure 9.7: Observer etangling two spins b and c at A, one of which is sent together with a timed measurement apparatus to B at which point the result of a spin measurement is sent. The observer waits outside the horizon until the information about b is Hawking radiated. The observer intercepts the Hawking radiation at C before going into the black hole to intercept the message from B at D. In gray, we have illustrated what happens if the signal is sent too late from B.

radiation until information about the spin of b has been radiated. This observer then enters the black hole to receive a message from B. According to postulate (1) the Hawking radiation may (and eventually will) carry the same information as the message in the interior. This means that the spin c will appear entangled with both the message from B and the Hawking radiation, violating monogamy of entanglement. Looking more closely, one realizes that the exterior observer has to wait for a long time to measure the spin in the Hawking radiation. This means that the apparatus at B must send its message about the spin B very quickly after going through the horizon, or the message will hit the singularity before it can reach the infalling observer. It turns out that the required timescale demands that B send a signal (photon) with energy well above the Planck scale [90]. The conclusion is that pending a contradiction coming from a model that describes physics above the Planck scale, black hole complementarity resolves the information paradox.

AMPSS argues that even in the absence of trans-Planckian physics, complementarity is not enough. To illustrate the issue of AMPSS, we consider a black hole older than the Page time, with modes decomposed as

$$|\Psi\rangle = |\psi\rangle_{\text{interior}} \otimes |\psi\rangle_{\text{radiation}} . \tag{9.33}$$

Let us call the interior Hilbert space A and the radiation Hilbert space R. Let us also divide the radiation into modes R_A far away from the horizon and R_B very close to the horizon. Well after the Page time, the black hole degrees of freedom should be mostly entangled with the far away radiation R_A . This follows because at the Page time, all interior degrees of freedom are entangled with a finitely thick shell of radiation that is now moving away from the black hole.

This leads to a conflict. According to axiom 2, low energy effective field theory is valid near the horizon and the horizon geometry is also smooth. It follows from LQFT on a smooth background that there is UV (short range) entanglement across the horizon. This can only work if either $R_A = R_B$, which is extremely nonlocal, or if the horizon is not smooth. AMPSS argue that the latter is the only reasonable alternative, since the identification of early and late radiation implies that what happens to an observer jumping into the black hole depends strongly on what an eventual observer does to R_B very far away, violating locality. The conclusion of AMPSS was that they had

proved that the horizon of black holes can not be smooth as seen by any infalling observer. To salvage the smoothness axiom, AMPSS suggested that the infalling observer would disentangle the interior with the far radiation causing the release of a large amount of energy, resulting in the infalling observer being greeted by a 'firewall' of high energy modes as they entered the black hole.

Maldacena and Susskind in [7] instead took the position that AMPSS had proved that there must be some kind of short-cut between the far radiation and the interior of the black hole. Based on this clue that there should be a geometric shortcut between entangled radiation and the black hole, ER=EPR was formulated. In doing so they also illuminated some ambiguity regarding the description of black hole interiors in the geometry from entanglement program of Raamsdonk et al.

9.3 From Entanglement and Areas to Complexity and Volumes

In the previous section we discovered that the interior of black holes seems to be described holographically by the details of the boundary microstate. We will now make this relationship between the boundary CFT microstate and the bulk theory more precise via the complexity=volume (CV) and complexity=action (CA) conjectures. To do this we first need to define what we mean by complexity, volume and action. Complexity is a measure of how complicated a state is to construct using only "simple" operators and a reference state. By action we mean the gravitational Lagrangian integrated over a WdW patch, but it is not completely clear what gravitational action to pick. The volume in this scenario is the volume of a spacelike "nice slice" anchored at the boundaries of the spacetime. In making the relation more precise, we will specialize to the eternal AdS-Schwarzschild spacetime and consider the time evolution of the wormhole connecting the two black hole horizons. We will find that the growing length of the wormhole will have the behaviour expected of the complexity of a chaotic quantum state.

The material covered in this section is highly conjectural. This is because complexity is barely under control in the discrete case, and so far has no natural continuum definition. This means that the most sophisticated checks of the conjectures we will resort to theories of gravity that are dual to discrete matrix models on the boundary, such as the Sachdev-Ye-Kitaev (SYK) model.

It seems that complexity specifically may not be the ultimate answer, but it may point in the right direction by putting the focus on microstate dependence of the behind-the-horizon region of black holes.

This section follows closely the lecture notes by Susskind titled "Entanglement is Not Enough" [83] with some extra details harvested mainly from [91] and [92].

9.3.1 Computational Complexity

Computational complexity is a candidate dual of the length of the wormhole or behind-the-horizon regions of black holes. To make this relation believable we better define computational complexity. We will being by making clear the classical notion of complexity before moving on to the quantum case.

Computational complexity is a computer science tool that quantifies the difficulty of carrying out a task. To define computational complexity we need a number of basic ingredients: a system, a space of states, a definition of a "simple state", a definition of a "simple operator", and a task. Generally the task is to transform a simple state into some arbitrary state, using a sequence of simple operators. The complexity of an arbitrary state is then defined as the minimum number of simple operators required to compose the general operator taking a simple state to the arbitrary state.

Let us consider a system of K classical bits whose space of states can be represented by K

binary numbers (11010110...). Let us also make an identification under the \mathbb{Z}_2 operation that flips all bits. An obvious candidate for a simple state is (00000...), which is equivalent to the state with all ones. The simplest imaginable operation is the flip of a single bit $0 \leftrightarrow 1$.

The maximum complexity C of a state is clearly K/2 due to the identification under \mathbb{Z}_2 . Another fact is that almost all states have nearly maximal complexity, straightforwardly understood by realizing that that a state with complexity C is built by acting with C simple operators on the simple state. When building a state of complexity C one can make

$$N(\mathcal{C}) = \frac{K!}{(K-\mathcal{C})!} \tag{9.34}$$

distinct choices of what bit to act on. Like the maximum complexity the maximum entropy is also proportional to K as $S = K \log 2$. Notably complexity and entropy are very similar in the classical case.

There are two characteristic timescales related to complexity, the thermalization time and the Poincaré recurrence time. The thermalization time is approximately equal to the time it takes to reach maximum complexity. By the thermilization time $t_{\text{therm}} \sim t_{\text{comp}}$, we mean the time it takes for some dynamical rule to bring a simple state into a nearly maximally entropic/complex state. Typically t_{therm} is polynomial bounded in K. The Poincaré recurrence time is the time it takes for a generic state to evolve into the simple state via some dynamical rule. Since the simple state is unique among 2^K states, this time is of order 2^K .

To summarize the quantities of classical complexity, we have:

$$\begin{array}{ll} \mathcal{C}_{\max} = K/2 & \text{maximum complexity} \\ S_{\max} = K \log 2 & \text{maximum entropy} \\ t_{\text{therm}} \sim t_{\text{comp}} < K^p & \text{time to thermalize or get maximally complex} \\ t_{\text{rec}} = 2^K & \text{Poincaré recurrence time.} \end{array}$$

We can now turn to the quantum case. The analogous system is K qubits with states of the form

$$|\psi\rangle = \sum_{i}^{2^{K}} \alpha_{i} |i\rangle \quad . \tag{9.35}$$

The main difference to the classical case is that now we require 2^{K} numbers to specify the state.

The quantum definition of a simple state is a state with no entanglement. Further the simplest state should be a tensor product with all qubits in the same state. Identifying states under a global SU(2) rotation on all qubits, there is a unique simplest state

$$|0\rangle = |000\ldots\rangle \quad . \tag{9.36}$$

Simple operators are more subtle than in the classical case. We want operators that are unitary and can create entanglement. The latter is not possible unless our "simple operators" act on at least two qubits at a time. In computing simple unitary two-qubit operators are called gates, and a sequence of gates is called a quantum circuit.

The task is then to build an arbitrary unitary K-qubit operator u by acting with a sequence of gates g_i , i.e. $u = g_n \dots g_2 g_1$. The complexity of the operator u is the minimum number of gates required to construct it.

In the AdS/CFT case it is better to define each step as containing K/2 gates. This means that all qubits are involved in a gate at each step, more closely representing the parallel computing of Hamiltonian time evolution. The complexity is then defined as the number of steps including K/2 gates are required to construct u. We can also define the complexity of a state $|\psi\rangle$ by noting that any state can be written as

$$|\psi\rangle = u \left|0\right\rangle \tag{9.37}$$

for some unitary operator u. The complexity of $|\psi\rangle$ is the minimum complexity of any u that fulfills equation (9.37). One can also define the relative complexity between between two states $|A\rangle$ and $|B\rangle$ as the minimum complexity of the operator u that takes one to the other.

Because of the significantly larger state space the complexification and recurrence times are significantly longer, while the thermalization time is unchanged. The new list of characteristic quantities is

$C_{\max} = e^{K}$	maximum complexity
$S_{\max} = K \log 2$	maximum entropy
$t_{\rm therm} < K^p$	time to thermalize
$t_{\rm comp} = e^K$	time to get maximally complex
$t_{ m rec} = e^{e^K}$	Poincaré recurrence time.

Just like in the classical case almost all states are nearly maximally complex, i.e. exponentially complex. The given thermalization-, complexification- and recurrence times assume that the evolution of the system is given by a Hamiltonian that is a sum of simple Hermitian operators. Simple Hermitian operators involve only one or two qubits, so a simple Hamiltonian implements time evolution by quantum gates.

Note that "simple" operators involving only a few qubits are notoriously hard to define in field theory. To check proposed gravitational duals against field theory results we therefore need to resort to cases where the CFT side may be described by a discrete matrix model, such as the SYK model [92].

Here we see the key property of quantum chaotic systems, the complexification time is *much* longer than the thermalization time. This means that there is a quantum informational property that grow long after the system has reached thermal equilibrium. This is precisely the situation of the black hole, after creation the black hole has a behind-the-horizon region that grows with time (in fact, classically it grows forever). Apart from complexity no good properties of a quantum chaotic system grow after the thermalization time, so complexity provides a unique quantum mechanical dual to the length of the wormhole.

We can also motivate the time dependence of complexity, assuming that at every time step we act with some finite number of quantum gates on random qubits. The key points are that at each timestep, the space of more complex (entropic) states is much larger than that of less complex (entropic) states, so complexity should increase in time. Additionally there is a maximum of complexity that can be gained per iteration since we act with a finite number of gates. Together these imply that complexity will grow as fast as it can roughly in a linear fashion until maximum complexity is almost saturated. At a linear rate the complexification time is clearly of order the complexity which is why $t_{\text{comp}} = e^{K}$.

We can also guess the rate of growth of complexity by noting that it is extensive, meaning it depends on the number of active degrees of freedom in a system. Another extensive measure of the size of a system is the entropy S. The derivative of complexity with respect to time should have units of inverse time. To take into account the fact that extremal black holes are not expected to have a growing behind-the-horizon region the correct quantity with the desired dimension is the

temperature T, so we end up with

$$\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}t} \sim TS \ . \tag{9.38}$$

Finally note that due to the quantum recurrence theorem the state will quasiperiodically return arbitrarily close to the original state on timescales doubly exponential in K. The resulting picture is that of complexity growing linearly until saturation, with quasiperiodic dips whose sharpness is limited by TS being the maximum rate of complexification and decomplexification.

Geometrized Complexity

It is possible to recast complexity in the language of Riemannian geometry as was done by Nielsen et al. in [93]. The geometrized version of complexity gives the most natural argument for the time dependence of complexity, and makes clearer just how complicated the relationship between microstates and interior geometries may be. We comment on complexity as if it is dual to a geometrical quantity here, although we have not yet demonstrated the credibility of such a conjecture.

The principal observation is that the space of special unitary operators acting on K qubits is $SU(2^K)$. The special unitary group is a Lie Algebra, meaning it is associated with a manifold whose connection is related to the generators of the group. The action of a sequence of gates traces out a trajectory in $SU(2^K)$ consisting of discrete line segments, and the complexity is the minimum number of such intervals that connects the simple element to the arbitrary one. With suitable definitions, it is clear that the complexity corresponds to the length of some kind of shortest trajectory connecting the point I and u on the manifold $SU(2^K)$.

To bring this to Riemannian form, we need to do away with the requirement of straight line segments, so we define a continuous version of the quantum circuit that replaces the discrete trajectories by smooth curves. We then have to define an appropriate Riemannian metric such that the minimal geodesic connecting two points has approximately the same length as the corresponding set of adjoined straight line segments. Additionally the geometry is usually taken to be translation invariant due to the arbitrariness in the choice of reference state.

The sequence of gates is replaced by the Hamiltonian evolution

$$u = \mathcal{P}\left[e^{-i\int h(s)\mathrm{d}s}\right] \tag{9.39}$$

where \mathcal{P} is the path ordering operator and s parametrizes a path.

We define a Hamiltonian to be k-local if it is a sum of terms involving at most k qubits. The metric is defined in such a way that path segments built from k-local Hamiltonians are short while segments are long if built from hard Hamiltonians that contain terms involving more than k qubits. More explicitly, a cost parameter q is introduced that multiplies the metric in directions which are not generated by k-local Hamiltonians. Nielsen shows that if $q > 4^k$, the complexity becomes insensitive to the precise value of q. This is a general property of complexity, it is very strongly dependent on K but has a weak dependence on any regulator parameters.

In the following we assume that the Hamiltonian that performs time evolution is k-local. For t small compared to the inverse cost parameter 1/q and the eigenvalues of h(s), the growth of complexity is linear. In this case, the trajectory generated by e^{-iHt} is a geodesic since it generates straight motion in a k-local direction. For some period of time this remains the shortest geodesic, but it can not hold forever since complexity has an upper bound. We can see this explicitly by considering the case k = 2. In this case each gate is an element of SU(4) requiring the specification of 15 real parameters, while a general unitary in $SU(2^K)$ requires the specification of $4^K - 1$ parameters. This implies that there are states that can only be obtained by acting with a minimum

of $\frac{4^{K}-1}{15}$ gates, but it also implies that the maximum number of gates required to get to any state is approximately $\frac{4^{K}-1}{15}$. We conclude that the maximum complexity is of order 4^{K} .

Next we may consider the shortest geodesic between the identity operator and some operator u in the complexity geometry. If we set q = 1, any point in the geometry is at a distance of order one (since we can move using a K-local operator at no additional cost). It is clear that for general $q < 4^{K}$ the upper bound on the length of the geodesic is exactly q. By setting $q = 4^{K}$ we ensure that the shortest geodesic is precisely upper bounded by 4^{K} .

Finally we may consider a geodesic swept out by a k-local Hamiltonian $U(t) = e^{-iHt}$. Assuming that the Hamiltonian is chaotic the geodesic never closes so it is always increasing in length, but at some point it ceases to be the shortest possibility. This happens when the geodesic intersects a point on the *cut-locus*⁵ of the geometry, at a distance of at most 4^{K} . The reason this particular point is interesting is that it signals the breakdown of classical general relativity on the gravitational side.

The important property of note is the case of evolution by a bounded N-local operator h with $u = e^{ih\epsilon}$ where N > k and ϵ small. This leads to a situation in which the states $|0\rangle$ and $|\phi\rangle = u |0\rangle$ are nearly parallel but have completely different complexity. Equivalently, the classical geometry behind the horizon may be vastly different for two quantum states that are indistinguishable from the perspective of expectations of Hermitian operators.

9.3.2 The Geometric Dual of Complexity

There are, roughly speaking, two conjectured duals of complexity in the CFT. These are the volume of the interior wormhole measured in a particular covariant and spacelike way, and the Einstein-Hilbert action integrated over a WdW patch. To compute any of these quantities, it is best to keep things simple. In the case of black hole pair creation and radiation of particles we can not use a classical approximation in the bulk because topology is classically conserved. Instead we consider the simpler, completely classical case of the two sided AdS-Schwarzschild black hole which has the crucial property of a growing wormhole which we may compare to the boundary state complexity.

Gravitational Preliminaries

To associate the time evolution of the bulk spacetime to time evolution on the boundary we need to pick an appropriate foliation of the spacetime in terms of codimension 1 spacelike surfaces. A good foliation needs to obey a set of conditions

- 1. The slices should be Cauchy surfaces. Formally all geodesically complete light- and timelike curves must intersect each slice exactly once.
- 2. The slices must stay way from singularities and high-curvature regions. Such slices are referred to as "nice slices".
- 3. The slices should have a coordinate invariant definition, up to a global Lorentz transformation.
- 4. On the boundary, the slices should be ordinary constant-time surfaces.
- 5. The entire black hole exterior should be foliated by these slices. By virtue of condition 2. the far interior of the black hole can not be foliated.

 $^{{}^{5}}A$ cut locus is the set of points are connected by several minimal geodesics to a reference point. An archetypal example is the antipodal point on the sphere.

To pick a surface that satisfies the above conditions, we first pick an asymptotic Lorentz frame. For a given time t on the boundary in this frame we consider the set of spacelike surfaces anchored on a spatial sphere of very large radius. Among the possible slices there exists one that has maximal volume and as $r \to \infty$ this maximal slice defines a unique global spacelike surface for each asymptotic time t. Letting t vary between $\pm \infty$ foliates the entire exterior spacetime by these maximal slices. This particular prescription is good because the existence of the maximal slice is guaranteed [91].

Two-Sided AdS Black Hole

Let us for simplicity study the simplest wormhole spacetime: the maximally extended AdS-Schwarzschild solution. The metric is given by

$$ds^{2} = -f(r)dt^{2} + f(r)^{-1}dr^{2} + r^{2}d\Omega_{D-2} .$$
(9.40)

For the BTZ 2+1d case the function f(r) is given by

$$f(r) = \frac{1}{L^2} (r^2 - \mu^2) , \qquad (9.41)$$

where L is the AdS radius and μ is related to the black hole mass by

$$\mu^2 = 8G_N M L^2 . (9.42)$$

For D > 3 the form factor is instead given by

$$f(r) = \frac{r^2}{L^2} + 1 - \frac{\mu^2}{r^{D-3}}$$
(9.43)

and μ is related to the black hole mass by

$$\mu^2 = \frac{G_N M L^2}{\omega_{D-2}} , \qquad (9.44)$$

where ω_{D-2} is the volume of a D-2-sphere. The maximally extended spacetime, as we saw in section 9.1, contains strange objects such as white holes. In addition it is very unstable to perturbations in the lower half of the Penrose diagram of figure 9.4. In the present case it is fine to throw away the lower half and just imagine that the two sided spacetime was created at t = 0by some unspecified event. Thus we start with a wormhole of vanishing length that continuously grows.

The quantity of interest in the maximal slice is the volume. Since it is attached to a spherically symmetric shell at infinite radius it should have spherical symmetry. In the bulk this tells us that it should be parametrized as the radius as a function of the bulk time s. The volume is then

$$V = \omega_{D-2} \int ds \sqrt{|f(r)| r^{D-2}} = \int ds \sqrt{g(s)} , \qquad (9.45)$$

where the object in the square root is the induced metric on the spacelike slice and the determinant of the metric $g(r) \equiv |f(r)|r^{2(D-2)}$. To determine the maximal slice we need to specify boundary conditions and solve a variational equation with respect to r(t)

$$\delta_r V = 0 \ . \tag{9.46}$$



Figure 9.8: Nice slices for different choices of boundary times. The portion of the nice slice inside the black hole horizons grows with time, approaching the final slice at $r = r_f$.

There is a final slice at $t = \infty$ that stays at a finite radius r_f from the central singularity. This is found by taking the limit $t \to \infty$, where time translation invariance and rotational invariance together imply a surface of constant radius. The value of r_f is found by maximizing g(r) (solving $\partial_r g(r)|_{r=r_f} = 0$). The function g(r) vanishes at the horizon and r = 0 and increases monotonically for $r > \mu$. This is because of the usual zero at the horizon as well as the suppression of the central singularity by $r^{2(D-2)}$, corresponding to the fact that the volume of the D - 2-sphere shrinks at zero faster than the singularity grows. Note that we ignore the divergent part as $r \to \infty$, so we are looking for a local maximum of a regulated expression for the volume somewhere inside the black hole horizon.

In the BTZ case the final radius is straightforwardly obtained as

$$r_f = \frac{\mu}{\sqrt{2}} \ . \tag{9.47}$$

The volume of the final slice is given by $V_f = \omega_{D-2} r_f^{D-2} \sqrt{|f(r_f)|}$.

The maximal slices for finite t are more difficult since we lose the time translation invariance. These are obtained by solving the equation with boundary condition:

$$\delta_r \int ds \sqrt{|f(r(s))|} r(s)^{D-2} = 0 , \quad \lim_{s \to t} r(s) = \infty .$$
 (9.48)

To solve this, we can note that codimension one surfaces with D - 2-spherical symmetry are geodesics with respect to the metric $ds^2 = -r^{2(D-2)}f(r)ds^2 + r^{2(D-2)}f^{-1}dr^2$. To find the volume we integrate the length element times the volume of the D-2 sphere along this geodesic. In the following, let us parametrize the curves by $r(\lambda), s(\lambda)$ and denote derivatives with respect to λ by \dot{r}, \dot{s} .

The metric has a time-translation symmetry with a corresponding conserved quantity given by $E = r^{2(D-2)} f(r) \dot{s}$ and the geodesic equation gives the parametrization constraint $r^{2(D-2)} \dot{r}^2 =$ $f(r) + E^2 r^{(-2(D-2))}$. We pick the asymptotic Lorentz frame in which the two boundary times $t_L = t_R$. The volume of a slice is then given by the integral

$$V(E) = 2 \int_{r_{\rm turn}(R)}^{R} \frac{\mathrm{d}r}{\dot{r}} = 2 \int_{r_{\rm turn}(R)}^{R} \frac{r^{2(D-2)}\mathrm{d}r}{\sqrt{E^2 + r^{2(D-2)}f(r)}}$$
(9.49)
where r_{turn} is the turning point at which \dot{r} vanishes and R is a regulator for the AdS boundary. The factor of two comes from the $t_R \leftrightarrow t_L$ symmetry of the problem in the chosen frame. It turns out that for for large R one finds [91]

$$V(t_L, t_R) \sim V_f \frac{|t_L + t_R|}{L} + \mathcal{O}(1) = \omega_{D-2} r_f^{D-2} \sqrt{|f(r_f)|} |t_L + t_R|$$
(9.50)

where $\mathcal{O}(1)$ represents terms that do not scale with the volume of the slice. Equation equation (9.50) is dimensionally sound because the form factor has the form of distance²/L². Slices corresponding to different boundary times, with the final slice $r = r_f$ and r_{turn} denoted are illustrated in figure 9.8.

Complexity=Volume (CV) Conjecture

Let us compare the expressions for the complexity of the thermofield double and the volume of the nice slice in the maximally extended AdS-Schwarzschild spacetime. In the complexity case we found in equation (9.38) that

$$\dot{\mathcal{C}} \sim TS \quad \left(=\frac{aA}{8\pi G_N}\right) ,$$

$$(9.51)$$

where on the gravitational side the temperature T is the Hawking temperature determined by $T = a/2\pi$ and S is the entropy $A/4G_N$. We should relate this expression the volume of the nice slice to find the conjectured relation between complexity and volume.

First let us note that the quantity $\omega_{D-2}r_f^{D-2}$ is not quite equal to the horizon area of the black hole, in fact in the 2+1-dimensional case they differ by a factor of two. In more general dimension, we may specify to large black holes so that the intractable constant term can be ignored. Then the determinant g(r) takes the form

$$g(r) = \frac{1}{L^2} (\mu^2 Z - Z^2) , \quad Z \equiv r^{D-1} .$$
 (9.52)

The values of the black hole and final slice radii are given by $g(r_H) = 0$, $g'(r_f) = 0$ respectively, and we find

$$Z_h = \mu^2 , \quad Z_f = \mu^2/2 , \qquad (9.53)$$

so that

$$r_H = \left(\mu^2\right)^{1/(D-1)}$$
, $r_f = \left(\frac{\mu^2}{2}\right)^{1/(D-1)}$. (9.54)

The areas of the two surfaces are related by

$$\frac{A_h}{A_f} = 2^{(D-2)/(D-1)} , \qquad (9.55)$$

which interestingly is very weakly dependent on the dimension D and of order 1. The surface gravity a of the black hole is given by

$$a = \frac{f'(r_H)}{2} = \frac{D-1}{2L^2} \mu^{\frac{2}{D-1}} .$$
(9.56)

The volume of the nice slice in terms of these quantities grows according to equation (9.50) as

$$\dot{V} = 2\omega_{D-2}r_f^{D-2}\sqrt{|f(r_f)|} = \omega_{D-2}\frac{\mu^2}{2L} = \frac{1}{D-1}A_ha$$
(9.57)

where the preceding equations were used. Comparing to equation (9.51) we see that

$$\mathcal{C} \sim (D-1) \frac{V}{G_N L} , \qquad (9.58)$$

given that both the complexity and the volume of the wormhole are zero at t = 0. This is a basis for the complexity=volume (CV) conjecture.

Let us note that the relation between complexity and volume is not universal, as can be seen by considering the two-sided asymptotically Minkowski-Schwarzschild spacetime. In this case [83],

$$\begin{split} f(r) &= 1 - \frac{\mu}{r^{D-3}} \;, \\ g(r) &= \mu r^{D-1} - r^{2D-4} \;, \\ r_H &= \mu^{1/(d-3)} \;, \\ r_f &= \left(\frac{1}{2}\frac{D-1}{D-2}\right)^{1/D-3} \mu^{1/(d-3)} \\ \frac{A_h}{A_f} &= \left[\frac{1}{2}\frac{D-2}{D-1}\right]^{(D-2)/(D-3)} \;, \\ \kappa &= \frac{1}{2}f'_h = \frac{D-3}{2\mu^{1/(D-3)}} \;. \end{split}$$

Doing the same algebra as in the AdS-Schwarzschild case, identifying $\dot{\mathcal{C}} \sim TS$ one finds

$$C \sim \frac{D-3}{Gr_H} V \left[\frac{1}{2^{1/(D-3)}} \sqrt{\frac{D-3}{D-1}} \left(\frac{D-1}{D-2} \right)^{(D-2)/(D-3)} \right] .$$
(9.59)

The *D* dependence in the square brackets looks complicated but in the span $4 \le D \le \infty$ it only changes by a factor of $9\sqrt{3}/24 \approx 0.65$. Ignoring this factor the complexity takes the form

$$\mathcal{C} \sim (D-3) \frac{V}{Gr_H} \ . \tag{9.60}$$

Note that the inverse AdS radius has been replaced by the mass dependent Schwarzschild radius. To see that this implies that the CV conjecture is not universal consider the case of two black holes in Minkowski space and interior volumes V_1, V_2 . The total complexity becomes proportional not to $V_1 + V_2$ as it would in the AdS case, but rather $\frac{V_1}{\mu_1} + \frac{V_2}{\mu_2}$.

Note that non-universality does not just tell us that we have a problem in Minkowski. There is also trouble for small black holes in AdS, as these will behave approximately as regular Schwarzschild black holes.

Complexity=Action (CA) Conjecture

We see that while the CV conjecture points us in a productive direction, it is only valid for very large AdS black holes. In addition the conjecture has a degree of arbitrariness built into the choice of foliation and does not describe the deep interior of the black hole. To remedy these problems a more general conjecture has been proposed, that the complexity equals the bulk action of a WdW patch [92]. We will now follow the argument of [92] in formulating the complexity=action (CA) conjecture.

To motivate the generalized duality the first thing to note is that the worldvolume \mathcal{W} of the behind the horizon region is a tube of average length $\frac{1}{2}|t_R+t_L|$ that exists for a time $\sim L$. In terms of the worldvolume the CA duality can be thus be expressed as

$$\dot{C} \sim \frac{V}{G_N L} \sim \frac{|\mathcal{W}|}{G_N L^2}$$
 (9.61)

The cosmological constant Λ is proportional to $-1/L^2$, so the latter expression is roughly the classical gravitational action of the worldvolume \mathcal{W} . This is a clue that the CV conjecture may be a special case of a more general CA conjecture.

In the eternal black hole geometry, we define the WdW patch $\mathcal{W}(t_L, t_R)$ as the union of all spacelike surfaces anchored at at t_L, t_R as in figure 9.6. Note that the WdW patch is not a causal patch, and thus cannot be monitored by a single observer. This is perfectly consistent with the fact that complexity is not a quantum observable.

Next we define $\mathcal{A}_{\mathcal{W}}$ as the action obtained by integrating the bulk (gravitational) Lagrangian over \mathcal{W} , including suitable boundary terms on ∂W . The CA conjecture is then

$$\mathcal{C}(|\psi(t_L, t_R)\rangle) = \frac{\mathcal{A}_{\mathcal{W}}}{\pi\hbar} .$$
(9.62)

The factor of π is arbitrarily chosen so that $e^{i\mathcal{A}_{W}\hbar}$ changes by a sign as one unit of complexity is added. [92] do not assign any physical meaning to this factor, noting that hopefully a more universal continuum version of complexity in field theory may exist such that the prefactor can be unambiguously defined.

Test of CA Duality with Static Uncharged Black Holes

To check that the proposed CA duality applies to AdS black holes we want to find out what the growth rate of the WdW action is in time, hopefully finding a leading contribution constant in time. To simplify the problem somewhat, we will only compute the t_L derivative of the bulk action since this leaves us with half as many integrals to compute. The setup under consideration is shown in figure 9.9. Let us also state the Einstein-Hilbert action plus Hawking-Gibbons boundary term:

$$\mathcal{A} = \frac{1}{16\pi G_N} \int_{\mathcal{M}} \sqrt{|g|} (\mathcal{R} - 2\Lambda) + \frac{1}{8\pi G} \int_{\partial \mathcal{M}} \sqrt{|h|} K .$$
(9.63)

The action (and volume) of the WdW patch is divergent due to the usual infinite distance to the AdS boundary. This should be related to a UV divergence due to a large number of UV modes contributing to the complexity on the CFT side. The time derivative of the WdW action depends on the difference in the action of the two grey patches in figure 9.9. When computing the time derivative, the boundary divergences should cancel because the asymptotic AdS geometry is time translation invariant. In fact due to time translation symmetry of the exterior the entire contribution of the exterior region cancels.

The past interior region is important at early times, but at late times $t_L + t_R \gg \beta$ the contribution to the action vanishes exponentially. At late times the S^{D-2} component of the past interior has constant radius in the region still covered by the WdW patch (up to exponentially vanishing terms). At this point we may apply the Gauss-Bonnet theorem to the remaining two dimensions, concluding that the past interior only contributes a topological term which remains constant because topology is classically conserved.

What remains is to understand the future interior. We will refer to right side of figure 9.9. The contributions on the lightsheet B' and the removed lightsheet B cancel because they are related by



Figure 9.9: To the left: Penrose diagram for AdS black hole, as well as two WdW patches for two different left boundary times. As t_L increases the WdW patch loses a region in the past and gains a region in the future of t_L , both are light gray. To the right: future interior region of the black hole covered by the difference of the two WdW patches, with the boundary A at the horizon $r = r_H$.

time translation symmetry. Similarly the corner contribution at AB' is cancelled by the removed corner at AB and B'C cancels BC. The remaining contributions to the derivative of the action is given by boundary terms at $A(r = r_H)$ and C(r = 0) as well was the bulk of the region.

From the Einstein equations we know that the cosmological constant and the curvature scalars are given by

$$\Lambda = -\frac{(D-1)(D-2)}{2L^2} , \quad \mathcal{R} = \frac{2D}{D-2}\Lambda .$$
(9.64)

Inserting these solutions and evaluating the Einstein-Hilbert term in equation (9.63) on the bulk of the future interior patch we get the contribution

$$\frac{\mathrm{d}\mathcal{A}_{EH}}{\mathrm{d}t_L} = -\frac{\omega_{D-2}r_H^{D-1}}{8\pi G L^2} \ . \tag{9.65}$$

To compute the GH boundary term we use that the trace of the extrinsic curvature of a constant r surface is

$$K = \frac{1}{2}n^r \frac{\partial_r (r^{2(D-2)}f)}{r^{2(D-2)}}$$
(9.66)

where f(r) is the relevant form factor present in the metric. The boundary contributions are then

$$\frac{\mathrm{d}\mathcal{A}_{\partial\mathcal{M}}}{\mathrm{d}t_L} = \left[-\frac{D-1}{D-2}M + \frac{\omega_{D-2}r^{D-3}}{8\pi G} \left((D-2)(D-1)\frac{r^2}{L^2} \right) \right]_0^{r_H} \,. \tag{9.67}$$

Adding the two terms and using that $f(r_H) = 0$ we obtain the result

$$\frac{\mathrm{d}\mathcal{A}}{\mathrm{d}t_L} = 2M \ . \tag{9.68}$$

If we set

$$C = \frac{\mathcal{A}}{\pi\hbar} \tag{9.69}$$

we find that

$$\dot{\mathcal{C}} = \frac{2M}{\pi\hbar} \sim TS \ . \tag{9.70}$$

Where the last approximate equality comes from equations (4.171) and (4.174). In contrast to the CV calculation this one does not at all depend on the size of the black hole, and the duality seems universal. The specific factor and proper equality is based on the expectation [94] that the static black hole saturates a particular bound on the rate of change of complexity (equation (2.12) in [95]).

Further motivation by charged and spinning black holes computations are provided in [92], and in [95] the relationship between the SYK model and Jackiw-Teitelbom gravity provides further insight into the relation. Most importantly [95] finds that for the CA duality to hold, boundary terms that were previously thought to be physically meaningless have to be taken into account. This is put in a larger context of new physics fixing previous ambiguities in the action. Examples are:

- gravity depending on the VEV of the potential while non graviational physics does not,
- quantum corrections being proportional to the overall factor in front of the action, since the loop expansion carries powers of the action,
- the dynamics of quantum gravity depending on topological terms in the action through genus expansions.

This is in no way a proof of the CA conjecture, but it is a compelling parallel to previous extensions of theoretical physics.

9.4 Black Hole Microstates

The prior results in this chapter point towards the importance of the microscopic structure of the CFT state in a holographic understanding of the interior of black holes. Up until now we have studied the possibility of complexity encoding the interior geometry, but the description is not precise. There are two recent papers that provide different constructions of the behind-the-horizon region of the black hole spacetimes [82, 81]. We will in this chapter follow [81], but note that [82] find closely related results using a completely different approach.

We will see that in addition to an exploration of the physics of black hole interiors, one of the asymptotic AdS spaces is cut off by a sharp boundary: an end-of-the-world (ETW) brane or Randall-Sundrum II brane. It is conjectured that some CFT states correspond to a localized gravitational theory on these branes. These branes have in this scenario an inflating FLRW geometry, possibly described holographically by a CFT. In this way holography may describe the experimentally relevant asymptotically de Sitter FLRW cosmology.

For the obtained geometries it is possible to compute the volume and WdW actions of the bulk spacetime, testing the conjectures of the previous section on a significantly more interesting spacetime than just the maximally extended AdS-Schwarzschild solution. While this is interesting we will refer these results to the original papers [82, 81], and instead focus on the novel prospect of FLRW cosmology.

9.4.1 Boundary CFT (BCFT) Microstate Geometries

To consider black hole microstates containing the interior, we start with the maximally extended Schwarzschild spacetime, and therefore the thermofield double on the field theory side:

$$|\Psi_{TFD}^{\beta}\rangle = \frac{1}{Z_{\beta}} \sum_{i} e^{\frac{-\beta E_{i}}{2}} |E_{i}\rangle_{L} \otimes |E_{i}\rangle_{R} \quad . \tag{9.71}$$

We then consider projecting this state on a pure state $|B\rangle$ of the left CFT, resulting in the state

$$|\Psi_B^\beta\rangle = \frac{1}{Z_\beta} e^{\frac{\beta E_i}{2}} \langle B|E_i\rangle |E_i\rangle_R \quad . \tag{9.72}$$

We can see this state as the result of measuring the state of the left CFT. If the measurement corresponds to looking at high energy degrees of freedom, we expect that the geometry is only affected close to the boundary. This effect then propagates causally into the interior of the left asymptotic region. Importantly, this lets us keep a significant portion of the left side of the extended spacetime without a second CFT to populate the thermofield double. Based on this we should specifically consider states $|B\rangle$ that have no long-range entanglement.

By considering the state obtained by complex conjugating the coefficients, we obtain a state that we know how to construct via a path integral. To see this, let us compute

$$\begin{split} |\hat{\Psi}_{B}^{\beta}\rangle &= \frac{1}{Z_{\beta}} \sum_{i} e^{\frac{-\beta E_{i}}{2}} \langle E_{i} | B \rangle | E_{i} \rangle \\ &= \frac{1}{Z_{\beta}} \sum_{i} e^{\frac{-\beta E_{i}}{2}} | E_{i} \rangle \langle E_{i} | B \rangle \\ &= \frac{1}{Z_{\beta}} e^{\frac{-\beta H}{2}} | B \rangle \end{split}$$
(9.73)

which indeed looks the operator expression for a path integral on a strip in Euclidean time of width $\beta/2$. We can relate this back to our original state by remembering that the complex conjugation of the coefficients is the antilinear, antiunitary operator corresponding to time reversal. Thus we know that $|\hat{\Psi}_B^{\beta}\rangle(t) = |\Psi_B^{\beta}\rangle(-t)$ and for a time-reversal symmetric state the two are equivalent.

Since we have a neat representation of this via a Euclidean path integral on a finite Euclidean time interval of length $\beta/2$, we can translate this state into a geometry by a method similar to the one we used in section 8.4.2. Note that the strip is a manifold with a boundary so we need to supply boundary conditions for $|B\rangle (\pm \frac{\beta}{2})$. In general we get a family of distinct CFTs depending on the choice of boundary condition. Some boundary conditions are particularly nice because they preserve some of the conformal symmetry. In general the boundary states may have infinite energy, but Euclidean evolution to $\tau = 0$ kills off high-energy modes resulting in a state of finite energy.

The proposed AdS dual to a CFT with a boundary (BCFT) is a bulk spacetime that is cut off deep into the bulk by a brane anchored on the boundary of the CFT [96]. For simplicity, one may model this brane by a constant tension ETW-brane plus a Neumann boundary condition ensuring that no momentum/energy flows out of the universe. The difference between the usual and BCFT cases is illustrated in figure 9.10.

It turns out that several expected properties of the boundary CFT (such as entropy) are captured in the bulk by considering the relatively simple gravitational action

$$I_{\text{bulk}} + I_{\text{ETW}} = \frac{1}{16\pi G} \int_{M_{\text{AdS}}} \mathrm{d}^{d+1}x \sqrt{-g} (R - 2\Lambda) + \frac{1}{8\pi G} \int_{Q_{\text{ETW}}} \mathrm{d}^{d}\sqrt{-h} (K - (d - 1)T) , \quad (9.74)$$

where we have defined the dimensionless tension parameter T of the ETW-brane such that the stress energy tensor on the brane is given by

$$8\pi GT_{ab} = (1-d)Tg_{ab}/L \tag{9.75}$$

where L is the AdS radius. For a stronger argument for the sensibility of this construction we refer to [96, 81].



Figure 9.10: To the left, the usual AdS/CFT setup where the CFT lives on the $r \to \infty$ boundary of AdS. To the right we have a finite boundary spacetime, and the spacetime is cut off by the ETW-brane $Q_{\rm ETW}$. The ETW-brane can be seen as the holographic image of the CFT boundary on the boundary of AdS.

We are now ready to figure out the geometries associated to the Euclidean time-evolved states

$$|\Psi\rangle = e^{-\frac{\beta}{2}H} |B\rangle \quad . \tag{9.76}$$

The first step to note is that t = 0 correlation functions may be interpreted as a path integral on the Euclidean time interval $\tau \int \left[-\frac{\beta}{2}, \frac{\beta}{2}\right]$ (times the spatial S^{d-1}) with operator insertions at $\tau = 0$. To find a Lorentzian geometry that is dual to the state in question, we use the $\tau = 0$ slice to define a Cauchy surface for the Lorentzian solution, which will inherit time-reversal symmetry from the Euclidean geometry.

There will be two possible Lorentzian solutions corresponding to the desired initial conditions, but depending on the values of T and β one of them will have a smaller action and dominate the path integral. For some value of $\beta(T) > \beta^*(T)$ there will be a phase transition between the two modes analogous to the Hawking-Page transition of section 7.1.3. Specifically for $\beta(T)$ above some critical value the Euclidean geometry with the lower action is a portion of pure Euclidean AdS and the Lorentzian solution will be thermal AdS at low temperature $1/\beta$. When $\beta(T) < \beta^*(T)$ we instead find that the Lorentzian initial condition is one full exterior of the AdS-Schwarzschild solution plus a behind-the-horizon region terminating on the ETW-brane. The size of the behind-the-horizon region is not constant, and the ETW-brane geometry has an FLRW character.

Euclidean Geometries

The class of CFT states that we have chosen preserve the spherical symmetry of the AdS-Schwarzschild spacetime since they are created by a measurement on a single asymptotic boundary of the CFT which is (conformally) a sphere. The most general spherically symmetric metric in the present case is the Euclidean AdS_{d+1} -Schwarzschild metric given by

$$ds^{2} = f(r)d\tau^{2} + \frac{dr^{2}}{f(r)} + r^{2}d\Omega_{d-1}^{2} , \qquad (9.77)$$

where

$$f(r) = \frac{r^2}{L^2} + 1 - \frac{r_H^{d-2}}{r^{d-2}} \left(\frac{r_H^2}{L^2} + 1\right) , \qquad (9.78)$$

and r_H is the Schwarzschild radius. As usual the absence of a conical singularity determines the periodicity of the Euclidean time direction. In terms of the Schwarzschild radius the periodicity is

 $\tau \sim \tau + \beta_H$ where

$$\beta_H = \frac{4\pi r_H L^2}{(d-2)L^2 + dr_H^2} \,. \tag{9.79}$$

Note that the Euclidean periodicity of the black hole is not necessarily related to the width of the CFT. Essentially the CFT with a boundary and its ETW-brane projected into the bulk can be superimposed on any spacetime with $\beta_H > \beta$. The boundary CFT will depend implicitly on β_H because the equations of motion will relate β and β_H .

To discern between the two geometries we have to explicitly compute the actions and compare them. Comparing the results is an exercise in numerics (that was done in [81]), but there are some enlightening details that we should sort out before using the Euclidean solutions to determine the Lorentzian ones. Let us consider again our bulk and boundary actions and find some equations of motion. The bulk action together with the Gibbons-Hawking term on the asymptotic boundary is given by

$$I_{\text{bulk}} = \frac{1}{16\pi G} \int_{M_{\text{AdS}}} \mathrm{d}^{d+1} x \sqrt{-g} (R - 2\Lambda) + I_{\text{bulk}}^{\text{matter}} + I_{GH} \;. \tag{9.80}$$

The equations of motion are the usual Einstein field equations

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T^{\text{bulk}}_{\mu\nu} - \Lambda g_{\mu\nu} . \qquad (9.81)$$

The ETW-brane action is a Gibbons-Hawking term with a dynamical boundary metric

$$I_{\rm ETW} = \frac{1}{8\pi G} \int_{M_{\rm AdS}} \mathrm{d}^d y \sqrt{-h} K + I_{\rm ETW}^{\rm matter} , \qquad (9.82)$$

where y^a are coordinates intrinsic to the ETW-brane and K_{ab} is the extrinsic curvature formally given by

$$K_{ab} = e^{\mu}_a e^{\nu}_b \nabla_{\mu} \hat{n}_{\nu} , \qquad (9.83)$$

where the $e_a^{\mu} = \frac{\partial x^{\mu}}{\partial y^a}$ are a pullback onto the brane of the bulk coordinates. The equations of motion obtained by varying with respect to h are

$$K_{ab} - Kh_{ab} = 8\pi G T_{ab}^{\text{ETW}} , \qquad (9.84)$$

where the stress energy tensor of the ETW-brane is defined in the usual way with respect to the metric

$$T_{ab}^{\rm ETW} = \frac{2}{\sqrt{-h}} \frac{\delta I_{\rm ETW}^{\rm matter}}{\delta h^{ab}} .$$
(9.85)

Following [81] we will focus on ETW-branes with constant tension, defining

$$8\pi G T_{ab}^{\rm ETW} = (1-d)T h_{ab} , \qquad (9.86)$$

where the prefactor (1-d) is a matter of later convenience. The bulk stress tensor is set to zero as we are considering a vacuum solution. Contracting the equations of motion with the appropriate inverse metrics we find

$$\begin{aligned} R - 2\Lambda &= -2d \ , \\ K &= dT \ . \end{aligned} \tag{9.87}$$

Since we are considering spherically symmetric geometries, we also expect a spherically symmetric ETW-brane, so its embedding will be parametrized by $r(\tau)$ where r is the radial coordinate in the bulk and the AdS boundary is $r = \infty$.

Combining equations (9.87) and (9.84) we see that

$$K_{ab} = Th_{ab} {.} {9.88}$$

To compute h_{ab} we make the spherically symmetric ansatz that the ETW-brane has the coordinates $x^{\mu} = (\tau, r(\tau), \Omega_{d-1})$, so that the induced line element is

$$\mathrm{d}s^{2} = \frac{\partial x^{\mu}}{\partial y^{a}} \frac{\partial x^{\nu}}{\partial y^{b}} g_{\mu\nu} \mathrm{d}y^{a} \mathrm{d}y^{a} = \left[\dot{r}^{2} \frac{1}{f(r)} + f(r)\right] \mathrm{d}\tau^{2} + r^{2}(\tau) \mathrm{d}\Omega^{2} , \qquad (9.89)$$

where we have denoted $\dot{r} = \frac{\partial r}{\partial \tau}$ Since the embedding function is just a function of one variable, the unit normal is $n_{\mu} = (\dot{r}, 1, 0)/(1 + \dot{r}^2)$ which when projected onto Q retains only the τ component. Going to the definition (equation (9.83)) of K_{ab} and looking at the $\tau\tau$ component of equation (9.84) we have

$$\nabla_{\tau} \frac{\dot{r}}{\sqrt{1+\dot{r}^2}} = T \left[\left(\frac{\partial r}{\partial \tau} \right)^2 \frac{1}{f(r)} + f(r) \right]$$
(9.90)

Based on the results of [81] it is possible to find that $K_{\tau\tau} = \frac{f^2(r)}{Tr^2} 6$, which after rearranging implies that the embedding is determined by

$$\frac{\partial r}{\partial \tau} = \frac{f(r)}{Tr} \sqrt{f(r) - T^2 r^2} . \qquad (9.91)$$

Assuming symmetry about $\tau = 0$ (motivated by the expected Lorentzian time reversal symmetry) we are looking for solutions with $\frac{dr}{d\tau}|_{\tau=0} = 0$. We see that this tells us that

$$f(r_0) = T^2 r_0^2 . (9.92)$$

Together with the condition that r is infinite for $\tau = \pm \beta/2$, this means that $r_0 > r_H$ for all nonzero T. We can now separate differentials and integrate equation (9.91) with $r(\tau = 0) = r_0$ fixing the integration constant, obtaining

$$\int_{\tau_0}^{\beta/2} \mathrm{d}\tau = \int_{r_0}^r \mathrm{d}\hat{r} \frac{T\hat{r}}{f(r)\sqrt{f(\hat{r}) - T^2\hat{r}^2}} \,. \tag{9.93}$$

Given r_H , T and β there are two choices of τ_0 that respect time reversal symmetry, namely $\tau_0 = \beta_H/2$ and $\tau_0 = 0$. For the choice of $\tau_0 = \beta_H/2$ we need to pick the negative sign of the square root in the radial integral to respect $\beta < \beta_H$. Thus, we have

$$\left\{\tau_{0} = \frac{\beta_{H}}{2}\right\} \quad \frac{\beta}{2} = \frac{\beta_{H}}{2} - \int_{r_{0}}^{\infty} \mathrm{d}r \frac{Tr}{f(r)\sqrt{f(r) - T^{2}r^{2}}} , \qquad (9.94)$$

$$\{\tau_0 = 0\} \quad \frac{\beta}{2} = \int_{r_0}^{\infty} \mathrm{d}r \frac{Tr}{f(r)\sqrt{f(r) - T^2 r^2}} \,. \tag{9.95}$$

The general shape of the solutions is shown in figure 9.11. The most important feature is that the $\tau = 0$ slice which will serve as initial conditions for the Lorentzian spacetime solution contains regions on both sides of the black hole horizon, with the second region cut off by the ETW-brane at the finite radius r_0 .

⁶I would like to find this result myself, but at the current time I do not know how the given information can give an expression with the right powers of T. My main guess is that we should use the traced equation of motion and the fact that $\nabla_{\tau} n^{\tau} = g^{\tau \tau} \nabla_{\tau} n_{\tau}$ in some way. I have some other problems, such as the induced metric given after equation (145) in [81] not matching mine.



Figure 9.11: a) Schematic appearance of Q_{ETW} embedded in the full AdS space with Euclidean periodicity β_H . The slice at $\tau = 0$ is indicated in blue. Particularly interesting is the spacelike slice when we pick $\tau_0 = \beta_H/2$, as the inside of the ETW-brane contains spacetime regions on two sides of the event horizon but only has one asymptotic boundary on which there is a CFT. This spacelike slice (with the circle fulfilling the role of S^{d-1} is imaged in b). We see that the initial condition for our Lorentzian spacetime contains one full black hole exterior and part of the second, with the second exterior cut off by the ETW-brane. To see that the Euclidean AdS-Schwarzschild truncates at $r = r_H$, realize that $r = r_H$ is a single point in the $\tau - r$ plane by the fact that we can pick cylindrical coordinates about $r = r_H$.

Note that in requiring that $\frac{dr}{d\tau} = 0$ at a time reflection fixed point we have thrown away some solutions, namely the ones where the ETW-brane does not connect the boundaries at $\pm \beta/2$ to each other⁷. Depending on the parameter values r_H, T we find that τ_0 is completely determined. For a bit of solid footing, we shall cite the numerical results of [81] for various dimensions.

• For d = 2 one finds (i.e. a bulk AdS₃ background)

$$\frac{r_0}{r_H} = \frac{1}{\sqrt{1 - T^2}} , \qquad (9.96)$$

and the ETW-brane always reaches the boundary at antipodal points, meaning that $\beta = \beta_H/2$. This occurs regardless of the values of T and r_H , so there exist solutions with arbitrarily large regions on the other side of the horizon (r_0 goes to infinity as the tension goes to 1).

- For d=3 there is a maximum value of the tension beyond which no solutions exist with a connected ETW-brane, it is found to be at $T_{\text{max}} \approx .95635$. This implies that the maximum value of $\frac{r_0}{r_H} \approx 2.2708$, so the region on the other side of the horizon is of a size comparable to the black hole itself.
- For d=4 there is a maximum of $T^*_{\text{max}} \approx 0.79765$ obtained in a limit of large r_H , leading to a maximum ratio $\frac{r_0}{r_H} \approx 1.2876$. The maximum T^* is set by the fact that β is sent to negative values, corresponding to a self intersecting ETW-brane.

⁷The disconnected solutions are treated in [81], but are omitted for brevity here.

In the pure AdS case, $f(r) = r^2 + L^2$ and the brane embedding is best parametrized by $\tau(r)$ with $\tau(\infty) = \pm \beta/2$. This time the symmetry of the problem tells us that $\frac{\partial \tau}{\partial r} = 0$ at r = 0. In this case the geometry is cut off in the τ direction rather than the r direction, leaving nothing interesting in the $\tau = 0$ slice. The background geometry has the same form and we have the same equation relating τ to r

$$\tau(r) - \beta/2 = -\int_r^\infty \frac{Tr}{(r^2 + L^2)\sqrt{(1 - T^2)r^2 + L^2}} , \qquad (9.97)$$

where we inserted f(r). This can be integrated to find

$$\tau(r) - \beta/2 = -\operatorname{arcsinh}\left(\frac{T}{\sqrt{r^2 + 1}\sqrt{1 - T^2}}\right)$$
 (9.98)

The surface that starts from the negative $\beta/2$ boundary is found by multiply by an overall -1. We have a similar to the black hole here where if $\operatorname{arcsinh}\left(\frac{T}{\sqrt{r^2+1}\sqrt{1-T^2}}\right) > \beta/2$ we have two intersecting ETW branes.

Lorentzian Microstate Geometries

In figure 9.11 we have to the right the initial condition for our Lorentzian solutions, containing a behind-the-horizon region cut off at r_0 by the ETW-brane. For our purposes we are satisfied to know that there technically exist solutions for the ETW on a black hole background such that the cut-off action is smaller than the pure AdS phase. For a discussion on the conditions on the parameters for a black hole spacetime as well as the domain of validity of the constant tension brane model, see [81]. We will now look at the trajectory of the ETW-brane with its initial (t = 0position determined by the above Euclidean analysis.

To find the Lorentzian trajectory, we analytically continue the trajectory to find

$$t(r) = \int_{r_0}^r \mathrm{d}\hat{r} \frac{T\hat{r}}{f(\hat{r})\sqrt{T^2\hat{r}^2 - f(\hat{r})}} \ . \tag{9.99}$$

In d = 2 by inserting $f(r) = L^2 + r^2$ one sees

$$\cosh(tr_H)\sqrt{\frac{r^2}{r_H^2} - 1} = \frac{T}{\sqrt{1 - T^2}}$$
, (9.100)

while in general dimension some more work is necessary. A convenient first step is to write the equation for t(r) in terms of the proper time λ on the brane, related to the usual time coordinate by

$$\frac{\mathrm{d}t}{\mathrm{d}\lambda} = \gamma = \sqrt{\frac{f(r)}{f^2(r) - \dot{r}^2}} , \qquad (9.101)$$

where \dot{r} is the derivative with respect to λ . The equation of motion then takes the form

$$\dot{r}^2 + [f(r) - T^2 r^2] = 0 . (9.102)$$

This can be reexpressed in terms of $L = \log(r)$ as

$$\dot{L}^2 + V(L) = T^2 , \qquad (9.103)$$

where

$$V(L) = \frac{f(r)}{r^2} = 1 + e^{-2L} - e^{-d(L-L_H)} (1 + e^{-2L_H}) .$$
(9.104)

We see that the the ETW-brane trajectory is that of a particle with total energy T^2 in the potential V(L). W are interested in the d > 2 case, for which the trajectories are classified in figure 9.12. Note that most of the values of T in the figure do not correspond to the black hole phase, which ceases to exist for $T > T_*$. The trajectories labelled by a and b correspond to ETW-branes that start at r = 0 and expand to a maximum radius $(r = r_0)$ at time t = 0 to then shrink to r = 0 again. Thus these solutions represent an ETW-brane coming out of the past singularity, going outside the event horizon and then crashing into the future singularity. The infinitely expanding trajectories d, e and the trajectories for $1 < T < T_{\rm crit}$ are uninteresting because they correspond to values of T for which the black hole phase is not favored. We do not give an explicit value for $T_{\rm crit}$, but it is clear by inspection of the potential that it has a global maximum that will be exceeded for sufficiently large T. In d = 2 the form function is simple and there are only trajectories of types a and e.

We can also write down the worldvolume metric for the ETW-brane, where since $d\lambda$ is the proper time we get an FLRW-like metric

$$\mathrm{d}s^2 = -\mathrm{d}\lambda^2 + r^2(\lambda)\mathrm{d}\Omega^2 , \qquad (9.105)$$

with $r(\lambda)$ acting as the scale factor. The proper time that elapses along the entire trajectory is

$$\lambda_{tot} = 2 \int_0^{r_0} \frac{\mathrm{d}r}{\sqrt{T^2 r^2 - f(r)}} \ . \tag{9.106}$$

Computing the integral for variable r it is possible to solve for the scale factor $r(\lambda)$ in terms of the proper time λ . In [81] it is found that for d = 2

$$r(\lambda) = \frac{r_H}{1 - T^2} \cos\left(\lambda \sqrt{1 - T^2}\right) \tag{9.107}$$

while for d = 4,

$$r(\lambda) = \frac{1}{\sqrt{2(1-T^2)}} \left[\cos\left(2\lambda\sqrt{1-T^2}\right) \sqrt{1+4(1-T^2)r_H^2(1+r_H^2)} - 1 \right]^{1/2} .$$
(9.108)

The d = 3 result is expressed in terms of elliptic integrals.

9.4.2 Microstate Cosmology

We saw in the previous section that the ETW-brane worldvolume looks like a d-dimensional FLRW spacetime, with a specific potential coming from the higher dimensional (or CFT) physics. In general the physics on a d-dimensional brane embedded in a higher dimensional does not include d-dimensional gravity, since closed strings propagate in the extra direction. Despite this braneworld cosmology is extensively studied, and there are in fact mechanisms for effectively localizing gravity to the surface of a brane [97]. The localization mechanism which we will review now was introduced in [97] and is called the Randall-Sundrum II (RSII) brane-world scenario. It is obtained for a brane which cuts off the UV region of an AdS spacetime. While this is not exactly the same as for our AdS-Schwarzschild space, one can expect that for well chosen values of L, r_H and trajectory the physics should be sufficiently close to pure AdS for localization to occur. The following discussion closely follows [81].



Figure 9.12: Image of potential V(L(r)) as well as schematic trajectories for r(t) depending on the tension T.

The RSII model we have a d = 4-dimensional brane embedded in a 5d AdS space. The bulk metric takes the form

$$ds_5^2 = dz^2 + e^{-2A(z)}g_{\mu\nu}dx^{\mu}dx^{\nu} , \qquad (9.109)$$

with a 3-brane at z = 0 and a \mathbb{Z}_2 symmetry imposed about z = 0 (so that the spacetime is cut off at the brane, we have a reflective boundary condition). Choosing $A(z) \sim |z|/L$ the bulk spacetime is a slice of AdS₅ that is cut off by a 3-brane. By tuning the brane tension relative to the cosmological constant of the bulk AdS space one can obtain a flat induced metric $g_{\mu\nu} = \eta_{\mu\nu}$ on the brane. Within this setup, linearized Einstein gravity is reproduced on the 3-brane for distances much larger than the AdS radius. Specifically, the gravitational potential on the brane becomes

$$V(r) \approx \frac{GM}{r} \left(1 + \frac{2L^2}{3r^2} \right) . \tag{9.110}$$

The formal reason for the localization of gravity is that the warp factor $e^{-2|z|/L}$ quickly kills off perturbations away from the brane, trapping gravitational energy on the brane. The ansatz for the linearized field equations is on the form $h_{\mu\nu} = \epsilon_{\mu\nu}\psi(z)\phi(x^{\mu})$ where $\phi(x^{\mu})$ solves the massive (4d) Laplace equation. The Einstein equation reduces to a Schrödinger equation for $\phi(z)$, and the solution turns out to give a massless zero mode that is localized to the brane and reproduces the 4d Newtonian potential.

In spacetimes that are only approximately AdS cut off by a brane that is only approximately Minkowski, it is interesting to ask if the above construction survives. A detailed analysis is carried out for a brane at constant radius r_b in a bulk AdS black hole background in [98], with the key difference being the absence of a stable zero mode. This means that there is no stable 4d graviton in the nearly AdS case. By the expectation that gravitational localization depends locally on the bulk metric, there should be an effective Einstein gravity far away from the black hole where the spacetime looks sufficiently AdS and the brane looks sufficiently Minkowski, with corrections going as r_H/r_b . Note that while the ETW-brane in our holographic construction is does not exist at constant r_b the picture should be valid when $\dot{r}/r << 1/L$, i.e. when the brane is static on the AdS timescale.

In summary, the existing knowledge about braneworld (RSII) cosmology gives us two restrictions for gravity localized to the ETW-brane:

1. We need $r_b >> r_h$, which implies that we must have $r_0 >> r_H$. In all of our scenarios this means that $T \sim 1$

2. We need $\dot{r}/r \ll 1/L$, which should be achievable as long as the total proper time $\lambda_{\max} \gg L$.

To satisfy the first requirement we in all cases require $T > T_{\text{crit}}$ or at least $T \ge 1$. Similarly the condition that λ_{max} is large is only realized for $T \approx 1$, since the $f(r) \ln 1/\sqrt{f(r) - T^2 r^2}$ asymptotes to $\sim r^2$.

While the Lorentzian solutions make sense for these values of T, consistency of the Euclidean solution that defines the setup only holds for $T < T_* < 1$ in $d \ge 3$ and the largest realizable r_0 are of the same order as the black hole radius. We see that the naive construction with a constant tension brane is not enough to realize localized braneworld gravity in a holographic setting. Despite this it is interesting that a brane with an FLRW type geometry is realizable at all, and it is very interesting to see if a more sophisticated holographic setup could actually realize the FLRW cosmology.

9.4. Black Hole Microstates

Chapter 10

Beyond Holography, de Sitter

In the previous chapters we have discussed the reconstruction of the bulk spacetime from entanglement information via the AdS/CFT duality. Such an approach has one main weakness: according to observation the universe is de Sitter. This has been a big problem for all proposed theories of quantum gravity so far. For example string theory at times seems to conspire against de Sitter vacua after compactification from 10d down to 4d. Recently, it was even conjectured that de Sitter vacua are impossible in string theory [99], although the validity of this "swampland conjecture" is heavily disputed [100].

Some hints towards holographic constructions of non-AdS cosmologies were produced in chapter 9.4, but we saw that to realize a 4d FLRW cosmology with 4d gravity we need complicated ETW brane dynamics. Having to add complicated dynamics is a problem encountered in de Sitter constructions in string theory and is under poor control. This motivates a departure from holography in an attempt to formulate quantum gravity in a way that in a controlled manner may produce asymptotically de Sitter cosmology.

There is a question of what we should keep if we try to make such a departure, and given the contents of the rest of this text it should be no surprise that we suggest the Ryu-Takayanagi relation. The structure of this chapter will be as follows:

- We follow [10], arguing that any asymptotically AdS spacetime can be approximated arbitrarily closely by entangling a large number of BCFTs. Arguing from string theory that effective spacetime dimensionality does not need to be constant everywhere we loosen the constraints on the individual BCFTs. We conjecture that the boundary theory might not need to have a geometrical interpretation at all.
- Based on this conjecture, we consider a discrete realization of AdS/CFT in the form of the MERA tensor network. This necessitates an introduction to tensor networks. We find that some modification to MERA is necessary to accommodate the Bousso bound, bulk-boundary correspondence and the Ryu-Takayanagi formula simultaneously.
- We consider a de Sitter/MERA correspondence, noting compatibility with de Sitter causality, cosmic no-hair theorems and the CV/CA conjectures.

This chapter will be of lesser depth than the previous two, because Part III will focus on the continuation of the results in chapter 8.



Figure 10.1: Path integration region preparing an entangled state between two disconnected CFT's living on the manifolds \tilde{M}_i and \tilde{M}_j . To the right, we have denoted the geometry that computes correslation functions in the prepared state. Notably the complex conjugation in the bra acts as time reflection making the geometry symmetric under a reflection in Euclidean time. This time reflection symmetry is inherited by the Lorentzian geometry defined with $\cup_i \tilde{M}_i$ as initial condition for the Einstein equations. The boundary CFT geometries are connected in the bulk by the ETW brane associated with the BCFTs.

10.1 It From BC Bit

There have been numerous attempts at emerging spacetime from quantum many body systems [8, 101], in which the full system is represented in terms of "tensor networks". Our objective in this section is justify tensor network constructions from the AdS/CFT direction by arguing that we can in a consistent manner divide the CFT on the boundary of AdS into finite subregions while keeping the interior geometry largely unchanged. This section follows an essay by Raamsdonk [10].

Tensor network constructions rely on factorizability of the Hilbert space on which the dynamics are defined, so we need to consider bulk theories with many disconnected boundaries of finite extent. We will refer to these disconnected boundary-CFTs (BCFTs) as BC-bits in analogy with the basic unit of quantum information, the qubit.

We have already introduced the holographic machinery of BCFTs in section 9.4. Also in this case it is sufficient to set Neumann conditions on the boundary and terminate the bulk spacetime on an ETW brane of constant tension. In contrast to the previous case we have boundaries in the spatial directions. The natural way to prepare the disconnected CFTs in an entangled state at t = 0 is to connect them in Euclidean time, and then relate correlators to a path integral over two copies of the euclidean geometry glued together at $\tau = 0$ as in figure 10.1. We then use the $\tau = 0$ slice as initial conditions for the Lorentzian gravitational action, obtaining a solution for the spacetime geometry (which inherits a time reflection symmetry from the path integral construction).

There is nothing preventing the extension of this construction to many boundaries, and by entangling the many BC bits we can expect a connected bulk. What remains is to make sure that we obtain a bulk spacetime that closely approximates the holographic dual of a single CFT on the full boundary of the spacetime. To carry out this construction it is useful to make some definitions.

Let us call the spacetime that we are seeking to approximate X and its boundary M. We want to divide M into a set of simply connected pieces $\{\tilde{M}_i\} \in M$. The union $\cup_i \tilde{M}_i$ should be a large subset of M, and we take all of the boundaries of the \tilde{M}_i to be smooth (which will require that some small patches of M can not be covered). [10] refers to the smoothing of the edges as 'sanding' the edges. Furthermore let us call the Euclidean path integral that defines the entangled state H and its mirror image \overline{H} . The mirror image is necessary in the construction of the correlation function



Figure 10.2: In a) we define a disconnected set of boundaries at $\tau = 0$. In b) we denote the Euclidean boundary geometry that prepares a state in which the CFTs living on each disjoint piece are entangled. In c) we have the geometry that is used to compute correlation functions. In red are the boundaries that will extend as ETW branes into the bulk, cutting off parts of the interior AdS spacetime. If the separation between boundary regions is taken to be small compared to the size of the regions, the resulting geometry of the $\tau = 0$ slice is given by d). As time passes the rough boundary features propagate inwards causally as in e), so the resulting Lorentzian spacetime is only smooth inside a WdW patch corresponding t = 0 slice of the boundary.

as in the right of figure 10.1.

The prescription to pick the boundaries turns out to be quite simple, we want the \tilde{M}_i be separated by a distance that is small compared to the size of the individual regions. This requirement is related to the Hawking-Page type transition of section 9.4, when the boundaries are far separated the action of the ETW brane that connects the boundary to itself is smaller and a disconnected bulk extension is preferred over the ETW brane stretching between multiple boundaries. If we instead pick the \tilde{M}_i to be very close, we get an ETW extending only a short distance into the bulk before turning back to another region \tilde{M}_j on the boundary, thus barely modifying the defining euclidean geometry H. The steps of this construction are illustrated in figure 10.2.

Despite the small modification of the Euclidean geometry, the Lorentzian geometry receives some nonvanishing modifications. This is due to the fact that the defects near the boundary, however small, will propagate inwards causally as in figure 10.2 e). This means that the Lorentzian geometry only remains smooth in the interior of the WdW patch associated to the boundary at t = 0 and cannot be (analytically) continued outside it.

In the limit of very many BC bits we can expect many small modifications to the original boundary M, so the geometry is well described inside the WdW patch by $\{\tilde{M}_i\}$. When the individual bits are very small compared to M, they should individually carry almost no information about the geometry represented by the full set. In this sense the spacetime is almost completely characterized by only the entanglement information and it manifestly disconnects if we try to disentangle the individual bits. Another important point is that from the (Lorentzian) CFT perspective the individual bits have no intrinsic position relative to eachother, so it is all dimensions of the bulk that emerge purely from entanglement.

An interesting point to focus on is the fact that the individual bits carry almost no information about the interior geometry. This implies that if we replace a BC bit with something radically different while preserving the entanglement structure, we might encode the same gravitational physics. Naively this is too optimistic, if we replace the CFT on one BC bit with a different CFT of another dimensionality we clearly are not describing the same theory of gravity at least in the vicinity of the replaced BCFT. If the bulk is to be unaffected we must be describing a bulk theory in which things like the dimension and dynamics of spacetime change as we move around in the bulk. Such behaviour is included in string theory, in which different stable (or metastable) low energy theories of gravity can exist separated by a transition region¹. Therefore if we are optimistic, it is possible to replace the boundary BC bits with different objects while only preserving the entanglement structure without straying from known physics.

In the same sense we might consider replacing all of the boundary BC bits by a different type of object while only retaining the entanglement structure (i.e. the Ryu-Takayanagi formula plus some eventual subleading correction). Close to the boundaries of the WdW patch this drastically changes the geometry and may even result in a completely nongeometric description, but in the deep interior we have a transition towards a spacetime that very closely resembles the case of the usual holographic CFT. This is of course wildly speculative, but it is interesting that the relation between entanglement and geometry might not depend very much on what systems you are entangling. Being extremely optimistic this may point us towards a completely universal relationship between entanglement and geometry. It might event be that geometry emerges from an a priori completely nongeometric theory.

10.2 Cosmology/MERA correspondence

The AdS/MERA correspondence is a relation between discretized Anti-de Sitter and a specific tensor network called MERA, short for *multi-scale entanglement renormalization ansatz*. This tensor network is used as an ansatz for finding the ground state of many body systems in condensed matter theory, but with the Ryu-Takayanagi formula it has been related to AdS and more recently dS cosmology.

We will now proceed to introduce tensor networks in general and the specific network named MERA. We show that the general properties of MERA match CFT expectations under some conditions. We then demonstrate the basic elements of the AdS/MERA and dS/MERA correspondences. This entire section follows structurally chapter 2 of the PhD thesis of Chunjun Cao [68]. The de Sitter/MERA correspondence is a boundary-free correspondence between gravitational and quantum physics and inspires the non-holographic program of bulk entanglement gravity [8].

¹This is most often described in the context of compactifications of string theory. The parameters describing the internal compactification manifold in string theory are dynamical and may change values as a function of position (and time). Even more interestingly the extensively studied Calabi-Yau manifolds on which compactification is usually studied can even transition smoothly between topologies, resulting in radically different effective theories in different regions. The difficulty of finding de Sitter in string theory is related to the difficulty of finding solutions for which the exterior geometry has positive curvature with a stable interior geometry.

10.2.1 Tensor Networks and Their Relation To Quantum States

This short introduction to tensor networks and MERA is based on [102] and [68]. The first of these has lots of examples using basic tensor networks, and the second is the source for the AdS/MERA and dS/MERA sections. Tensor networks are a useful graphical representation of tensors and their contraction. The most useful feature of a tensor network is that it scales very well with the number of tensors involved and it can make symmetries obfuscated by index notation graphically obvious.

The basic idea is to represent a rank (a,b) tensor as a geometric shape with a + b legs that are somehow differentiated, usually by which direction they extend in. For example we rewrite the rank four tensor R according to

$$R^{\mu}{}_{\nu\rho\sigma} \longrightarrow \bigwedge^{\mathbf{R}} . \tag{10.1}$$

If the tensors have indices of more than one type, such as spinor indices, group indices, spacetime indices or transverse indices, this is indicated by having several types of leg. For example we could have drawn lines for spacetime indices, dashed lines for spinor indices and squiggly lines for adjoint indices.

Tensor operations can also be translated into the graphical language in a simple way. For example, we represent contraction by joining two external legs

$$R^{\mu}_{\mu\rho\sigma} = \bigcirc (10.2)$$

which generalizes in an obvious way to the contraction of two tensors

$$A^{\mu}{}_{\nu\rho}B^{\nu\rho} =$$
(10.3)

where we have taken the liberty of representing A and B as a circle and a square with nothing inscribed. Note that whether the contracted legs are raised or lowered on either component is conventionally unimportant since raised and lowered indices are usually related by a metric or its inverse, and you always contract raised indices with lowered indices or vice versa. In most applications of tensor networks the metric is flat Euclidean, although raised and lowered indices may carry a conjugation in a complex vector space.

We can also write down the tensor product

$$A^{\mu}{}_{\nu\rho}B^{\sigma\lambda} = \bigwedge \otimes \blacksquare \equiv R^{\mu}{}_{\nu\rho}{}^{\sigma\lambda} = \bigwedge \qquad (10.4)$$

where we have represented the new rank five tensor $R^{\mu}{}_{\nu\rho}{}^{\sigma\lambda}$ by an oval. Note that the ordering of the external legs may matter, it is often good for notation if this ordering is somehow reflected by the geometric representation of the tensor. For example an antisymmetric rank 3 tensor may be best represented by a triangle with the legs coming out of the corners.

A particular symmetry that is especially apparent in the tensor network notation is the cyclicity of the trace, proven as follows:

$$(10.5)$$

where we in the intermediate and final steps have noted that by exchanging the raised and lowered indices of a matrix we obtain the transpose, indicated with a T. Note that if we contract the indices in a complex vector space, we should replace the transpose with a dagger. This is because the inner product on complex vector spaces usually includes an equivalence under phase transformations meaning that in addition to being related by a flat metric to covariant components, contravariant components also need to be complex conjugated. We only considered the example of three distinct matrices, but the proof generalizes to any number by the fact that the multiplication of two matrices is a matrix.

MERA

Having established the basic notation we can now introduce the MERA tensor network. The MERA tensor network is an especially good ansatz for ground states in quantum critical systems, which are described by CFTs. We will demonstrate that the MERA wavefunction possesses basic properties expected of conformal field theories.

The MERA tensor network has three basic constituents. The first is the unitary disentangler



that take input states $i_1, i_2 \in V \otimes V$ and produces unentangled states (i.e. living in the $V \oplus V$ subspace). U is defined so that $U_{i_1i_2}^{j_1j_2}(U^{\dagger})_{k_1k_2}^{i_1i_2} = \delta_{k_1}^{j_1}\delta_{k_2}^{j_2}$. The second basic constituent is the coarse graining tensor



which takes states from the space $V \otimes V$ to $|0\rangle \otimes V$. W is taken to be an isometry (in the quantum channel sense) so we have $W_{i_1i_2}^j(W^{\dagger})_k^{i_1i_2} = \delta_k^j$ where we have performed an implied inner product of the $|0\rangle$'s. Note that by their definitions, W and U are quantum channels as can be checked by looking to equation (2.24).

Finally, we have the top tensor $t_{\mu\nu}$ defined by

$$\sum_{\mu\nu} t_{\mu\nu} t^*_{\mu\nu} . (10.8)$$

The top tensor serves to truncate the network at the top. The top tensor can have any number of legs, which will significantly affect the near-center geometry of the dual AdS to the MERA network.

By combining two $W_{i_1i_2}^j$ we can map the output of two disentanglers to the input of a single one. The idea is then to have a physical lattice (the 'boundary') at the bottom of a tensor network, and the tensor network itself is constructed as many layers of increasing coarse-graining of the boundary state. Going back to the name, we apply multiple scales of renormalization as the ansatz for the state of a quantum system. Let us write this down more explicitly



where we should extend the pattern to infinity or until we hit a periodic boundary condition [103]. In the case of periodic boundary conditions, the network is truncated at the top by the top tensor.

In principle we could generalize this to a MERA of degree k, in which case the coarse graining tensor W would take k inputs to one. Additionally we could have lattices that are more connected, corresponding to more legs on the disentangler U. We will stick to the 1d lattice and k = 2 for our examples. With periodic boundary conditions the network truncates with a top tensor at depth $\mathcal{O}(\log_k(\beta))$.

We can also think about the entanglement entropy between the physical sites at the bottom of this tensor network. We know that the von Neumann entropy is upper bounded by $\log(\text{Dim}(\mathcal{H}))$ for a system with Hilbert space \mathcal{H} . We know here that each leg represents an index living in a Hilbert space of some dimension \mathcal{X} , which we call the *bond dimension*. In principle we could construct systems in which this is not constant, but we abstain from such complications here. The entanglement entropy between N physical sites and the rest of the MERA is thus upper bounded by the minimum number of legs n in the tensor network that you have to cut to separate the the N sites and their complement in the following way

$$S_{\max} = n \log(\chi) , \qquad (10.10)$$

using the additivity of entropy under the tensor product of subsystems. We can illustrate the cut as follows



By inspecting the diagram at different scales it is possible to deduce that the number of cuts scales as $n \sim \log(N)$ [102]. Therefore the maximum entanglement entropy scales as

$$S_{\max} \sim \log(N) \log(\chi) \tag{10.12}$$

for subsystems of size N. Note that this matches the CFT expectation for 1 + 1-dimensional systems (equation (4.188), corroborating the claim that MERA captures the behaviour of systems at criticality.

We can also compute correlation functions. Local operators are inserted at a single site on the lattice, contracted with a $|0\rangle$ on the end of the closest entangler. The bra $\langle \psi |$ is transposed, which

means it is upside down in the tensor network notation, and the $|0\rangle$ s are replaced by $\langle 0|s$. We can then compute a correlator between two sites as follows (truncating the contracted $|0\rangle$ s)

$$\langle \psi | O(x_j) O(x_{j+N}) | \psi \rangle =$$

We see the beginnings of a particularly interesting pattern here, the two point function seems to (nearly) factorize. Note that the two operator insertions in equations (10.13) are connected by a coarse-grainer one layer up from what is displayed. If we were to separate the two insertions by a distance that is 2^p times larger, they would be connected by a coarse-grainer p layers deeper. Below this scale the correlators factors into $\mathcal{O}(p)$ applications of the same superoperator



where we have denoted the argument of S by a grey, upright rectangle. We can decompose S into Choi-Kraus operators according to

$$\mathcal{S}(O(x_i)) = \sum_k V^k O(x_i) V_k^{\dagger} . \qquad (10.15)$$

 \mathcal{S} is a completely positive and unital map by virtue of the properties of U and W, meaning that the matrices V_k in the decomposition have eigenvalues $\lambda < 1$. The correlation function between two operators separated by N sites is given by $\log N$ applications of \mathcal{S} . Considering O_1, O_2 as eigenoperators of \mathcal{S} we see that the two point function goes as

$$\langle \psi | O_1(x_j) O_2(x_{j+N}) | \psi \rangle \sim \frac{\langle \psi | O_1(x_j) O_2(x_{j+1}) | \psi \rangle}{|N|^{\Delta_1 + \Delta_2}} ,$$
 (10.16)

where $\Delta_i = \log \lambda_i$, the Δ_i are known as the scaling dimensions of the operators O_i and the λ_i are the corresponding eigenvalues of \mathcal{S} . The factor $\langle \psi | O_1(x_j) O_2(x_{j+1}) | \psi \rangle$ comes from the last coarsegrainer connecting the two correlators at depth log N. This is reminiscent of the behaviour of CFT two-point functions that go as $\langle \mathcal{O}(x) \mathcal{O}(y) \rangle \sim 1/|x-y|^{2\Delta}$ where Δ is the conformal weight

10.2.2 AdS/MERA Correspondence

We have seen that the MERA ansatz for the wavefunction reproduces expected CFT results for correlation functions and entanglement entropies on the lower boundary. We will now demonstrate that the 'bulk' of the MERA network, i.e. all the deeper layers can be mapped to a discretized anti-de Sitter geometry by assigning a geometric meaning to the contractions in the graph. This correspondence was first noted by Swingle in [104]. The section closely follows the presentation of Cao in [68], however I use less sophisticated inequalities, resulting in weaker bounds that nevertheless reflect the same conclusions.

Let us begin with a very general reason for why MERA realizes the correspondence. In Ad-S/CFT we can see the distance into the bulk as the renormalization scale of the CFT becoming a physical dimension. In the MERA ansatz we do the exact same thing, we have the physical lattice at the bottom "boundary", and we then organize correlations into layers of increasing length scale. Thus the MERA ansatz also naturally "grows an extra dimension" that can be associated to the energy scale, where increasing depth corresponds to deeper IR physics in the boundary theory. The top tensor naturally corresponds to r = 0 in AdS, while the boundary is located at $r \sim 1/\epsilon$ where ϵ is the lattice spacing that in QFT would correspond to a UV cutoff.

Geometrizing MERA

We can begin by endowing the graph with a geometry. To do this we impose a graph metric on the MERA network, and show that the discretized geometry reproduces AdS on large scales. To prepare for the introduction of a graph metric, we should pick equivalent reference points on the MERA. We mark these reference points as follows, adding some horizontal lines for reference



Next we realize that translation invariance on the boundary (by the lattice spacing) fixes the distance L_1 between any two boundary points. Each of the layers corresponding to a red horizontal line is also separated by a single distance scale L_2 , since they are separated by the same exact intermediate tensor. We implement this with the following translation from tensor component to a weighted graph



Thus the MERA can be represented as the following graph with two inequivalent distance scales L_1 and L_2 .



Here we have noted that to each point on the graph we assign a Hilbert space V with $\dim(V) = \tilde{\chi}$, where $\tilde{\chi}$ is not necessarily equal to the bond dimension.

Geodesic Between Boundary Points

From equation (7.38) we expect the minimal geodesic between two boundary points of AdS to be separated by

$$L_{\gamma} = 2L \ln\left(\frac{R}{\epsilon} \sin\frac{l}{R}\right) \,. \tag{10.17}$$

where R is the radius of the boundary CFT and and L is the AdS radius and we have redefined $l \rightarrow \pi l$ relative to section 7.2. In the case of R large, i.e. a nonperiodic boundary CFT we have

$$L_{\gamma} = 2L \ln\left(\frac{l}{\epsilon}\right) \,. \tag{10.18}$$

We would also like to consider a geodesic in AdS that stays at constant radius and goes between points on the boundary separated by a distance l. In terms of Fefferman-Graham coordinates we can express this condition as $z = z_0$, where z_0 is small. Such a geodesic can straightforwardly be be shown to have length

$$L_{z_0} = \frac{L}{z_0} l \ . \tag{10.19}$$

We would like to show that our invented MERA graph geometry reproduces equations (10.18) and (10.19). Since the boundary of the MERA has a definite lattice spacing, it should in the AdS picture be defined at a finite distance from the boundary $z = \epsilon$, where ϵ is the holographic UV cutoff. In the *m*:th layer of the MERA, the number of sites separating two points that are separated

by distance l on the boundary scales as $N_{\text{sites}} \sim \frac{l}{2^m \epsilon}$. The graph distance at constant z is L_1 times the number of separating nodes, so we find that

$$L_{z_0} = L_1 \frac{l}{z_0} \Big|_{z_0 = 2^m \epsilon} .$$
 (10.20)

Upon comparing with equation (10.19) we see that L_1 is to be identified with the AdS radius L. Note that for a MERA of degree k we replace $2 \rightarrow k$.

Next we should perform the computation corresponding to the minimal geodesic of the Ryu-Takayanagi chapter. We consider two sites on the boundary lattice of the MERA. If we take the two sites to be separated by $l >> \epsilon$, the shortest distance will be given by a path that is dominated by vertical segments. The depth at which the two points are connected is located at depth $m = \log_2 \frac{l}{\epsilon}$. The vertical distance has to be travelled twice (up from one endpoint and down to the second), so we find

$$L_{\gamma} = 2L_2 \log_2 \frac{l}{\epsilon} . \tag{10.21}$$

To match with the AdS result in equation (10.18) we see that we should set $L_2 = L \ln 2$. Notably, MERA only approximates AdS on distances larger than the AdS scale.

It is difficult say something about sub-AdS scale physics due to the fact that the MERA only contains super-AdS length scales. One is forced to conclude that localized physics must be encoded in the individual tensor factors².

Boundary Entanglement Entropies

In addition to recreating the large scale geometric features of AdS, we would like to replicate the entanglement structure. More specifically, we will demand that MERA correctly computes the CFT entanglement entropy of a line element, given by equation (4.188) as

$$S_{CFT} = \frac{c}{3} \ln\left(\frac{l}{\epsilon}\right) \,, \tag{10.22}$$

where the central charge is related to the AdS_3 scale by c = 3L/2G. This will give us a restriction on the bond dimension χ .

We have two expressions relating to the MERA entanglement entropy, namely the upper bound coming from the minimum number of bonds cut to separate a segment of the MERA from its complement as well as the Ryu-Takayanagi prescription applied to the MERA bulk geodesic. From these we get

$$S_{RT} = \frac{L_{\gamma}}{4G} = \frac{L}{2G} \ln\left(\frac{l}{a}\right) \tag{10.23}$$

$$S_{\text{MeraCuts}} \le \log(N) \log(\chi) = \log(\chi) \ln(2) \ln\left(\frac{l}{a}\right) . \tag{10.24}$$

Both of these can only be true if

$$\chi > e^{\frac{L}{2G}} = e^{c/3} , \qquad (10.25)$$

i.e. if χ is exponentially large in the central charge. We also know that semiclassical bulk gravity is associated with large c, so a holographic MERA must have large bond dimension χ .

 $^{^{2}}$ One might think that by writing the MERA coordinates as some function of the AdS coordinates instead of just equating them, one can bypass this problem. This apparent freedom turns out to be unphysical and you end up with the same MERA length scales in the end [68].

Bulk Entanglement Entropy

If the correspondence between AdS and MERA is real then it should also be possible to realize the holographic principle, namely that black holes maximize the entropy per volume. We assume that the black hole is well scrambled, so that the entropy of a black hole is proportional to the logarithm of the dimension of the bulk Hilbert space of its interior $\ln(\dim \mathcal{H}_B) = \frac{A}{4G}$.

In this section, we will leave the degree k arbitrary, but all images will be drawn with k = 2 for simplicity. To realize the Bousso bound in the MERA geometry, we need the dimension of the Hilbert space inside a ball with surface area A to satisfy

$$\ln(\dim \mathcal{H}_B) \le \frac{A}{4G} \ . \tag{10.26}$$

To realize a good definition of a ball we should map the MERA from a half-plane representation of AdS $(x, z), x \in [-\infty, \infty], z \in [0, \infty]$ to a radial one (ρ, θ) with $\rho \in [0,1], \theta \in [0,2\pi)$. This requires us to truncate the MERA at a finite depth M with a top tensor and impose periodic boundary conditions. The metric becomes

$$ds^{2} = \frac{L^{2}}{(1-\rho)^{2}} \left[d\rho^{2} + \left(\frac{d\theta}{2\pi}\right)^{2} \right] .$$
 (10.27)

To adapt the MERA to these coordinates we first remember that the first layer of the MERA lies at $z = \epsilon$. We can then pick a site, label it site number 0 and define its position as x = 0. The UV-most lattice is then situated at $(x,z) = (n\epsilon,\epsilon)$ where $n = 0, 1, \ldots (k^M - 1)$ where we have identified x = 0 and $x = k^M \epsilon$. We obtain a consistent embedding of the MERA graph by defining the coordinates ρ, θ in the following way

$$\rho = \frac{k^M \epsilon - z}{k^M \epsilon} \tag{10.28}$$

$$\theta = 2\pi \frac{x}{k^M \epsilon} . \tag{10.29}$$

If the top tensor is allowed to have an arbitrary number of legs T, the periodicity at the UVmost level becomes $Tk^{M-1}\epsilon$. This means that we should redefine the angular variable so that $\theta = 2\pi x/(Tk^{M-1}\epsilon)$. In this case the metric becomes

$$ds^{2} = \frac{L^{2}}{(1-\rho)^{2}} \left[d\rho^{2} + \frac{T^{2}}{k^{2}} \left(\frac{d\theta}{2\pi} \right) \right] .$$
 (10.30)

In the radial coordinates, the MERA with T = k = 2 can be graphically represented as in figure 10.3. where the thick lines are of length L_2 and the circular, thinner lines are of length L_1 between each node indicated by a grey dot. The grey diamond at the center is the top tensor. Note that the circles should be at radii $1 - 1/k^n$ where n is the n:th layer from the top tensor. The four radii in this image are thus $\rho = \{\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}\}$.

We are now ready to interpret (10.26) in terms of the MERA. Let us pick the ball B to be centered on $\rho = 0$. We take the ball to contain the top tensor, the sites at the top tensor's legs and the first n_B layers that come after. This region is indicated for $n_B = 1, T = 2, k = 2$ in figure 10.4. The boundary of the ball is a surface at constant ρ and the area (circumference) of its boundary is given by

$$A = Tk^{n_B}L_1 \tag{10.31}$$



Figure 10.3: MERA network superimposed on ball representation of Euclidean AdS space. The grey diamond in the center is the top tensor. The four circles denoted in the image are at $\rho = \left\{\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}\right\}$. The boundary at $\rho = 1$ is the boundary of AdS.

which using that $L_1 = L$ and $c = \frac{3R}{2G}$ tells us that

$$\frac{A}{4G} = \frac{Tk^{n_B}c}{6} \ . \tag{10.32}$$

We now need a prescription for assigning a dimension to the bulk Hilbert space of the region inside the ball. The simplest construction is to assign a Hilbert space factor V_{bulk} to each bulk vertex and a Hilbert space V_T to the top tensor³. The dimension of the V_{bulk} factors must be the same due to the symmetries of the MERA network, while the dimension of the top tensor Hilbert space is a free choice. The number of sites contained by B in addition to the top tensor is

$$N_B = T \sum_{i=0}^{n_B} k^i = T \left(\frac{k^{n_B - 1}}{k - 1} \right) .$$
 (10.33)

The Hilbert space inside the ball is therefore $\mathcal{H}_B = (V_{\text{bulk}})^{\otimes N_B} \otimes V_T$. To the spaces V_{bulk} and V_T we assign the dimensions $\tilde{\chi}, \tilde{\chi}_T$ respectively. It follows that $\ln(\dim(\mathcal{H}_B)) = N_B \ln(\tilde{\chi}) + \ln(\tilde{\chi}_T)$, which when used together with equations (10.26) and (10.32) tells us that

$$T\left(\frac{k^{n_B-1}}{k-1}\right)\ln(\tilde{\chi}) + \ln(\tilde{\chi}_T) \le \frac{Tk^{n_B}c}{6} .$$

$$(10.34)$$

If we can pick $\tilde{\chi}$ and $\tilde{\chi}_T$ to be anything, we can clearly satisfy this bound, but we have another property of AdS/CFT to reconcile, namely the dimensions of the bulk and boundary Hilbert spaces. The boundary Hilbert space has dimension $\chi^{N_{\text{boundary}}} = \chi^{Tk^M}$ determined by the bond dimensions. The number of bulk sites (apart from the top tensor) is given by $Tk^{N_{M-1}}$. By equating the dimensions of boundary and bulk we find

$$T\left(\frac{k^{M-2}}{k-1}\right)\ln(\tilde{\chi}) + \ln(\tilde{\chi}_T) = k^M \ln(\chi) . \qquad (10.35)$$

³This is a dangerous prescription since if we expect local QFT to hold in a weak gravity limit, a bulk-local factorization of Hilbert space is not expected to exist.



Figure 10.4: MERA network superimposed on ball representation of Euclidean AdS space. The region in red is the interior of an eventual black hole with event horizon denoted by the red circle. The gravitational analogy only makes sense when the black hole is close in size to the largest MERA layer it contains.

For the MERA to reproduce the Ryu-Takayanagi formula we saw that that χ must satisfy (10.25)

$$\ln \chi > \frac{c}{3} \ . \tag{10.36}$$

We can now combine the Bousso bound (equation (10.34)), the holography condition (equation (10.35)) and the Ryu-Takayanagi bound (equation (10.36)) to arrive at a contradiction in this simplest construction. We begin by inserting equation (10.36) into equation (10.34) to obtain

$$T\left(\frac{k^{n_B-1}}{k-1}\right)\ln(\tilde{\chi}) + \frac{1}{T}\ln(\tilde{\chi}_T) \le \frac{k^{n_B}}{2}\ln(\chi) .$$
 (10.37)

We then rearrange (10.35) to the following form

$$T\left(\frac{k^{M-2}}{k-1}\right)\ln(\tilde{\chi}) = k^M \ln(\chi) - \ln(\tilde{\chi}_T) , \qquad (10.38)$$

to see that (10.37) becomes

$$\frac{k^{n_B-1}}{k^{M-2}} \left(k^M \ln(\chi) - \ln \tilde{\chi}_T \right) + \frac{1}{T} \ln(\tilde{\chi}_T) \le \frac{k^{n_B}}{2} \ln(\chi) , \qquad (10.39)$$

$$k^{n_B+1}\ln(\chi) + \left(\frac{1}{T} - \frac{k^{n_B-1}}{k^{M-2}}\right)\ln(\tilde{\chi}_T) \le \frac{k^{n_B}}{2}\ln(\chi) , \qquad (10.40)$$

$$\left(\frac{1}{Tk^{n_B}} - \frac{1}{k^{M-1}}\right) \frac{\ln(\tilde{\chi}_T)}{\ln(\chi)} + 2k \le 1 .$$
 (10.41)

Since $k, T, \chi, \tilde{\chi}$ are positive integers, the following is true:

• If n_B is nearly M - 1 (the maximum size), the inequality can be satisfied. The inequality should hold independently of n_B , so this is not helpful.

• The bound can be satisfied for any $n_B < M$ if $T \ge k^{M-2}$ by tuning $\tilde{\chi}_T$. This is not sensible, since this means that the MERA network truncates with a top tensor after the first layer (or sooner).

Therefore, we conclude that the MERA without additional structure can not accommodate Ryu-Takayanagi, bulk-boundary correspondence and the Bousso bound. Since AdS does accommodate these there can be no self consistent AdS/MERA correspondence without additional structure.

As noted by Cao [68], a minimal generalization of the MERA would be a version in which the boundary state is a highly entangled state. This means that we may have non maximal entanglement in the bonds connecting to the boundary, and the holographic condition weakens to

$$T\left(\frac{k^{M-2}}{k-1}\right)\ln(\tilde{\chi}) + \ln(\tilde{\chi}_T) = \eta k^M \ln(\chi) , \quad \eta \in [0,1] , \qquad (10.42)$$

where η measures the average entanglement across the UV-most bonds. Then the bound in equation (10.41) is weakened to

$$\left(\frac{1}{Tk^{n_B}} - \frac{1}{k^{M-1}}\right)\frac{\ln(\tilde{\chi}_T)}{\ln(\chi)} + \eta 2k \le 1$$
(10.43)

Notably for $\eta \leq \frac{1}{4}$ the part in parentheses no longer has to be negative for k = 2, and it is in principle possible to reconcile holography, Ryu-Takayanagi and the Bousso bound⁴. This is interesting, and motivates further investigation of the possible correspondence between AdS and MERA-like tensor networks. This is outside the scope of this text, and we will instead proceed to consider the possibility of a de Sitter/MERA correspondence.

10.2.3 dS/MERA Correspondence

In addition to the AdS/MERA correspondence there is a proposed correspondence between de Sitter space and MERA, i.e. a dS/MERA correspondence. In this correspondence we make a completely different interpretation of the MERA, in which the fine graining direction (or UV direction) is taken to be the direction of increasing time. Thus taking the k = 2 MERA of the previous section, we are considering a dS₂/MERA₂ correspondence instead of an AdS₃/MERA₂ correspondence. Viewed in this way, we will see that the MERA captures the causal structure of de Sitter spacetime, where the UV direction represents time evolution away from the de Sitter throat. This section also follows the thesis of Cao [68] as well as the original paper upon which Cao bases his section [105]. The most notable property of this correspondence is the fact that it is not holographic.

Geometry of de Sitter and MERA

The de Sitter spacetime has constant, positive curvature. Its most notable property is the presence of a cosmological horizon for any given observer, which is consistent with the accelerating expansion of the universe. We will focus on dS_{1+1} , which has the standard representation

$$ds^{2} = L_{dS}^{2}(-dt^{2} + \cosh^{2} td\theta)$$
(10.44)

where t is non-periodic and θ has period 2π . It is instructive to think of de Sitter as an expanding ball that has radius L_{dS} at t = 0 and then expands for all time (in either direction from t = 0).

⁴One may interpret this subset of sufficiently entangled boundary states as a kind of "geometric subspace" of the full boundary theory. This is in line with the intuition we built about coherent CFT states in section 8.4.



Figure 10.5: The MERA tensor network, seen as a quantum circuit in which each layer performs a step of time evolution, admits a notion of causality. In the image we have illustrated the causal past (of well chosen points) at the bottom of the MERA, and the causal future of a point at the top. The past (future) domain of dependence of a set of points S is the set of all nodes such that they are connected to S by a set of future (past) directed edges. The causal past of a set of point in the UV-most layer of the MERA is known in the literature as a *causal cone*.

Setting $\cosh t = \sec \alpha$, we end up with the particularly convenient conformally flat representation of dS

$$ds^{2} = \frac{L_{dS}^{2}}{\cos\alpha} \left(-d\alpha^{2} + d\theta^{2} \right)$$
(10.45)

where α is restricted to $-\pi/2,\pi/2$. Because of this the Penrose diagram of de Sitter is best represented by a rectangle in the $\theta - \alpha$ plane on the following form



Where we have drawn out the causal diamonds of line elements at $\alpha = 0$. Note that these two causal patches are completely disjoint, and signals from either can never reach the other.

There is a consistent way to geometrically embed the MERA in the top half $(t \ge 0)$ of this diagram. To understand and anticipate this embedding we should first note that MERA admits a form of causal structure for flows in the IR direction. The basic elements of this causal structure is illustrated in figure 10.5, and in figure 10.6 complementary causal regions are chosen so that they match the Penrose diagram of de Sitter.

Noting the similarity in the causal structure we may think more closely about the correspondence. Starting at a time t = 0, the length of the constant time slice has doubled n times at the time $t_n = \operatorname{arccosh}(2^n)$. If we as in the AdS/MERA correspondence associate the bulk sites of MERA to local spacetime regions with the same dimension Hilbert space, it is natural to set the n + 1:th layer from the top of the MERA at time t^n . In figure 10.7 the first few layers of the identification is illustrated. The angles are chosen as

$$\theta_j^{(n)} = \frac{\pi}{2^{n+1}} \left(j + \frac{1}{2} \quad j = 0, \dots, 2^{n+2} - 1 \right) .$$
(10.47)



Figure 10.6: We set the MERA network to be periodic, with a period of four sites in the IR-most layer. Next we sort the MERA into complementary causal pasts and causal futures. Remembering that time runs upwards in (10.46), this matches the causal structure of the top half of the de Sitter Penrose diagram, with the UV-most sites close to $\alpha = \pi/2$ and the IR-most sites at $\alpha = 0$.



Figure 10.7: MERA imposed on the upper half of the de Sitter Penrose diagram. Two of four choices of causal patch are highlighted in grey. The toher two possible choices are related to the indicated patches via a translation by $\pi/2$ in the θ direction.

This particular angular alignment of the MERA is related to the so-called *static patch* of dS. Let us consider the upper half of the de Sitter Penrose diagram. The cosmological horizon of a static dS observer has constant proper radius, and the region of spacetime that can be in the causal past of this observer is given by the static patch. If we as in the AdS case identify a proper distance L_1 between nodes in the same layer, we see that the proposed imposition of MERA in figure 10.7 captures the constant proper radius property of the cosmological horizon for static observers at $\theta = \frac{\pi}{2}, \frac{3\pi}{2}$. By inspecting the $\alpha = 0$ time slice, the proper size of the static patch is πL_{dS} . Consistency between the MERA graph and de Sitter then tells us to identify $L_1 = \frac{\pi}{2}L_{dS}$. The conclusion is that each static patch of de Sitter coincides with a causal cone of the MERA network.

This description of dS is very coarse grained, we have only two sites per horizon volume with each layer separated by cosmological time scales. Like in the AdS case, the small scale physics must be somehow embedded in the individual sites. To have consistency with the Gibbons-Hawking entropy of the cosmological horizon $S_{\rm dS}$, the Hilbert space corresponding to a spatial slice of the static patch must satisfy $\ln \left[(\dim(\mathcal{H}_{\rm static})] \sim S_{\rm dS} \right]$. For our asymptotically de Sitter universe consistency sets $S_{\rm dS} \sim 10^{122}$.

Cosmic No Hair

In the gravitational literature, cosmic no-hair theorems say that spacetimes with a positive cosmological constant (with reasonable matter content) always asymptote to de Sitter [106]. Quantum versions of the no-hair theorems assert that the states of quantum field tend to the vacuum configuration at late times. If the dS/MERA correspondence is a good correspondence, there should be a MERA analog of the no-hair theorems. More explicitly this means that when flowing towards the UV in the MERA, the state should become closer and closer to some asymptotic state ρ^* , given any input state ρ_0 .

In section 2.3.7 we showed that given that there is a state ρ^* that is left invariant by a quantum channel \mathcal{N} , monotonicity of relative entropy implies that all other states approach ρ^* under the repeated application of \mathcal{N} .

To get at the character of the fixed state ρ^* , let us look more closely at the quantum channel that takes us from one cosmic horizon interior to the next. Each layer comes down to the application of the same operator represented by the following tensor network

The legs leaving the static patch exit the observable universe, so they should be traced over. By raising and lowering some indices we can rewrite the tensor network according to

$$\mathcal{N}(\rho)^c{}_d = \left(\begin{array}{c} \rho \\ \rho \\ \end{array} \right) \quad . \tag{10.49}$$

We can write this out algebraically using the definitions in equations (10.6) and (10.7) as

$$(\mathcal{N}(\rho))^{c}{}_{d} = U^{a}{}_{b}{}^{c}{}_{d}W^{j}{}_{ak}(W^{\dagger}){}_{i}{}^{bk}\rho^{i}{}_{j} \ . \tag{10.50}$$

By decomposing $U^{a}{}_{b}{}^{c}{}_{d}$ and $W^{j}{}_{ak}(W^{\dagger})_{i}{}^{bk} = \mathbf{W}^{j}{}_{a}{}^{b}{}_{i}$ into Choi-Kraus form equation (10.50) may be rewritten in matrix form

$$\mathcal{N}(\rho) = \sum_{i} \sum_{j} U_{i} W_{j} \rho W_{j}^{\dagger} U_{i}^{\dagger} . \qquad (10.51)$$

In this form, using that

$$\sum_{i} U_i U_i^{\dagger} = \mathbf{1} , \quad \sum_{j} W_j W_j^{\dagger} = \mathbf{1}$$
(10.52)

it is clear that the identity operator is an eigenoperator of \mathcal{N} with eigenvalue 1. This tells us that \mathcal{N} necessarily has the eigenvalue 1 in its spectrum, meaning there exist states that are preserved by \mathcal{N} . This proves that MERA admits a cosmic no-hair theorem in which the end state is the maximally mixed state $\rho^* \sim \frac{1}{e^{S_{dS}}} \mathbf{1}$.

It is interesting to ask what other states ρ^* are possible, but it turns out that the general problem is unsolved [68]. Here we are satisfied just saying that MERA admits a no-hair theorem, consistent with our knowledge of de Sitter cosmology.



Circuit Complexity, CV/CA Conjectures

We saw in the previous chapter two conjectures relating computational complexity to geometrical quantities. These were the Complexity=Volume (CV) and Complexity=Action (CA) conjectures. We will here focus on a possible relation to the CA conjecture, but with a pure de Sitter spacetime there is no way to differentiate the two. The CA conjecture says explicitly that the computational complexity C is related to the Einstein-Hilbert action $S_{\rm EH}$ evaluated on a Wheeler-de Witt patch by

$$\mathcal{C} = qS_{\rm EH} , \qquad (10.53)$$

where in the black hole example of section 9.3.2 we picked $q = 1/\hbar\pi$. The relation we test in this section is fundamentally different to 9.3.2 since it is not holographic, we have only a de Sitter bulk.

In the MERA there is a straightforward way to provide a lower and upper bound for the complexity of the state at time t. For the definition of complexity we choose the reference state to be the initial state of the MERA. We choose the set of simple gates to be the set $\{U, W\}$ of disentanglers and coarse-grainers from which we construct the MERA. The first observation is that since our time direction moves in the UV direction of the MERA, the disentanglers in fact entangle factors of the input state with extra $|0\rangle$ s that should be seen as part of the initial reference state. For a k-nary MERA each coarse-grainer generates a k times as many entangled factors as it takes inputs. In the j:th layer there are k^{j-1} coarse-grainers acting on k^{j-1} inputs, meaning that at time t the number of entangled factors is

$$N(t) = \sum_{j}^{t} k^{j} \tag{10.54}$$

where we take the time to be discretized into MERA time steps. To entangle N(t) states with k-local operators you need to act at least N(t)/k times, giving a lower bound on the complexity of the state at time t. An upper bound is trivially provided by the MERA itself, in which we at time t have acted with 2N(t) operators. Thus we have

$$\frac{1}{k}N(t) \le \mathcal{C}(t) \le 2N(t) . \tag{10.55}$$

The only important feature of the complexity of the MERA is that equation (10.54) is a geometric sum meaning it will be exponential in time. Changing the reference state or gates will change the prefactors in equation (10.55), but not the time dependence.

The Einstein Hilbert action evaluated on the patch $0 \le t' \le t$ for D-dimensional de Sitter is given in the coordinates of equation (10.45) by [68]

$$S_{\rm EH} = \frac{1}{16\pi G} \int_0^t dt' \int d\Omega_{D-1} \sqrt{-g} R$$

= $\frac{R l_{\rm dS}^D V_{D-1}}{16\pi G} \int_0^T \cosh^{D-1}(t')$
= $\frac{R l_{\rm dS}^D V_{D-1}}{16\pi G} \frac{1}{(D-1)2^{D-1}} e^{(D-1)t} + \text{subleading terms} ,$ (10.56)

where $R = D(D-1)/l_{dS}^2$ is the Ricci scalar and V_{D-1} is the volume of the unit D-1 sphere. This indeed matches the scaling behaviour of the MERA if we pick D = 2, so the dS/MERA correspondence is consistent with the CA conjecture. Note that in the MERA, we already know that the volume of the k:th layer goes as N(k), so MERA is also consistent with he CV conjecture.

Part III Original Work
Chapter 11

Gravity Emerging to Third Order

In this chapter we will try to generalize the emergence of EFE to second order in perturbations as found in section 8.4 to third order in perturbations. This should necessitate a plethora of new physics input such as the form of three point functions and operator product expansions in the CFT and potential derivative and quantum corrections to the gravitational theory, a second Hollands-Wald gauge condition and higher order terms in the gravitational analysis. What we have time to accomplish in this section is the explicit form of the correlation functions that are to be understood on the CFT side.

11.1 Gravitational Side

The only generalization necessary on the gravitaional side is the details of the second order Hollands-Wald gauge condition. This also includes the necessity of generalizing the condition on the vector field V that lets us understand metric perturbations that are not in Hollands-Wald gauge.

We extend the calculation of equation (8.89) by adding a second order metric perturbation, which will give us a second Hollands-Wald (HW) gauge condition. This should be implemented together with the first Hollands-Wald gauge. Terms of order δX^2 do not contribute because the extrinsic curvature vanishes to first order in metric perturbations by the first gauge condition. The starting point is then

$$\delta_G(\delta_X \delta_G A[G, X]) = \delta_G \left(\int \sqrt{-g} \left(\frac{1}{2} \nabla_{\overline{\alpha}} \delta G^{\alpha}{}_{\alpha} - G^{\alpha\beta} \nabla_{\beta} \delta G_{\overline{\alpha}\alpha} \right) \delta X^{\overline{\alpha}} \right) , \qquad (11.1)$$

where we use that the variations commute. Here, we are once again remembering that the barred indices are orthogonal to the extremal surface, while unbarred indices are parallel. Note that to this order the ∇_{α} are not invariant under perturbations so we get three nontrivial variations to consider

$$\delta G^{\alpha\beta} = -G^{\alpha\delta}G^{\beta\gamma}\delta G_{\delta\gamma} ,$$

$$\delta \Gamma^{\alpha}{}_{\beta\gamma} = G^{\alpha\delta} \left(\nabla_{(\beta}\delta G_{\gamma)\delta} - \frac{1}{2}\nabla_{\delta}\delta G_{\beta\gamma} \right) ,$$

$$\delta_G \sqrt{-g} = \frac{1}{2}\sqrt{-g}G^{\alpha\beta}\delta G_{\alpha\beta} \equiv \sqrt{-g}\delta G^{\alpha}{}_{\alpha} .$$

The variation of $\sqrt{-g}$ results only in multiplying the original Hollands-Wald gauge by the trace of the metric perturbation, so it is zero by imposing the first HW gauge. The $\nabla_{\overline{\alpha}} G^{\alpha}{}_{\alpha}$ term thankfully

gives no Christoffels. The variation of the first term is then

$$\delta_{G}(\nabla_{\overline{\alpha}}\delta G^{\alpha}{}_{\alpha}) = \delta_{G}(\nabla_{\overline{\alpha}}G^{\alpha\beta}\delta G_{\alpha\beta})$$
$$= \nabla_{\overline{\alpha}}\left(-G^{\alpha\delta}G^{\beta\gamma}\delta G_{\delta\gamma}\delta G_{\alpha\beta}\right) + \nabla_{\overline{\alpha}}\delta^{2}G^{\alpha}{}_{\alpha}$$
$$= -2\delta G^{\alpha\beta}\nabla_{\overline{\alpha}}\delta G_{\alpha\beta} + \nabla_{\overline{\alpha}}\delta^{2}G^{\alpha}{}_{\alpha} .$$
(11.2)

Let us now turn to the variation of the second term:

$$\begin{split} \delta_{G}(G^{\alpha\beta}\nabla_{\beta}\delta G_{\alpha\overline{\alpha}}) &= -G^{\alpha\gamma}G^{\beta\delta}\delta G_{\gamma\delta}\nabla_{\beta}\delta G_{\overline{\alpha}\alpha} + G^{\alpha\beta}\nabla_{\beta}\delta^{2}G_{\overline{\alpha}\alpha} \\ &\quad -G^{\alpha\beta}\left[\delta\Gamma^{\gamma}{}_{\beta\overline{\alpha}}\delta G_{\gamma\alpha} + \delta\Gamma^{\gamma}{}_{\beta\alpha}\delta G_{\overline{\alpha}\gamma}\right] \\ &= -G^{\alpha\gamma}G^{\beta\delta}\delta G_{\gamma\delta}\nabla_{\beta}\delta G_{\overline{\alpha}\alpha} + G^{\alpha\beta}\nabla_{\beta}\delta^{2}G_{\overline{\alpha}\alpha} \\ &\quad -G^{\alpha\beta}G^{\gamma\delta}\left[\left(\nabla_{(\beta}\delta G_{\overline{\alpha})\delta} - \frac{1}{2}\nabla_{\delta}\delta G_{\beta\overline{\alpha}}\right)\delta G_{\gamma\alpha} + \left(\nabla_{(\beta}\delta G_{\alpha)\delta} - \frac{1}{2}\nabla_{\delta}\delta G_{\beta\alpha}\right)\delta G_{\overline{\alpha}\gamma}\right] \,. \end{split}$$

Using metric compatibility of the unperturbed ∇_{β} with the unperturbed metric we can make some simplifications by contracting indices:

$$\delta_{G}(G^{\alpha\beta}\nabla_{\beta}\delta G_{\alpha\overline{\alpha}}) = -\delta G_{\alpha}{}^{\beta}\nabla_{\beta}\delta G_{\overline{\alpha}}{}^{\alpha} + \nabla_{\alpha}\delta^{2}G_{\overline{\alpha}}{}^{\alpha} - \nabla_{(\alpha}\delta G_{\overline{\alpha})}{}^{\delta}\delta G_{\delta}{}^{\alpha} + \frac{1}{2}\nabla_{\delta}\delta G^{\alpha}{}_{\overline{\alpha}}\delta G^{\delta}{}_{\alpha} - \nabla_{\alpha}\delta G^{\alpha}{}_{\delta}\delta G_{\overline{\alpha}}{}^{\delta} + \frac{1}{2}\nabla_{\delta}\delta G^{\alpha}{}_{\alpha}\delta G_{\overline{\alpha}}{}^{\delta} .$$
^(11.3)

Writing everything out at once, we have

$$\begin{split} \delta_{G}(\delta_{X}\delta_{G}A[G,X]) &= \int \sqrt{-g} \Bigg[-\nabla_{\overline{\alpha}}\delta G^{\alpha\beta}\delta G_{\alpha\beta} + \frac{1}{2}\nabla_{\overline{\alpha}}\delta^{2}G^{\alpha}{}_{\alpha} + \delta G_{\alpha}{}^{\beta}\nabla_{\beta}\delta G_{\overline{\alpha}}{}^{\alpha} - \nabla_{\alpha}\delta^{2}G_{\overline{\alpha}}{}^{\alpha} \\ &+ \nabla_{(\alpha}\delta G_{\overline{\alpha})}{}^{\delta}\delta G_{\delta}{}^{\alpha} - \frac{1}{2}\nabla_{\delta}\delta G^{\alpha}{}_{\overline{\alpha}}\delta G^{\delta}{}_{\alpha} + \nabla_{\alpha}\delta G^{\alpha}{}_{\delta}\delta G_{\overline{\alpha}}{}^{\delta} - \frac{1}{2}\nabla_{\delta}\delta G^{\alpha}{}_{\alpha}\delta G_{\overline{\alpha}}{}^{\delta} \Bigg] \delta X^{\overline{\alpha}} \\ &= \int \sqrt{-g} \Bigg[-\frac{1}{2}\nabla_{\overline{\alpha}}\delta G^{\alpha\beta}\delta G_{\alpha\beta} + \frac{1}{2}\nabla_{\overline{\alpha}}\delta^{2}G^{\alpha}{}_{\alpha} + \frac{1}{2}\delta G_{\alpha}{}^{\beta}\nabla_{\beta}\delta G_{\overline{\alpha}}{}^{\alpha} - \nabla_{\alpha}\delta^{2}G_{\overline{\alpha}}{}^{\alpha} \\ &+ \frac{1}{2}\nabla_{\alpha}\delta G_{\overline{\alpha}}{}^{\delta}\delta G_{\delta}{}^{\alpha} + \nabla_{\alpha}\delta G^{\alpha}{}_{\delta}\delta G_{\overline{\alpha}}{}^{\delta} - \frac{1}{2}\nabla_{\delta}\delta G^{\alpha}{}_{\alpha}\delta G_{\overline{\alpha}}{}^{\delta} \Bigg] \delta X^{\overline{\alpha}} \\ &= \int \sqrt{-g} \Bigg[\frac{1}{2}\nabla_{\overline{\alpha}}\delta^{2}G^{\alpha}{}_{\alpha} - \nabla_{\alpha}\delta^{2}G_{\overline{\alpha}}{}^{\alpha} \\ &+ \delta G_{\alpha}{}^{\beta}\nabla_{\beta}\delta G_{\overline{\alpha}}{}^{\alpha} + \nabla_{\alpha}\delta G^{\alpha}{}_{\delta}\delta G_{\overline{\alpha}}{}^{\delta} - \frac{1}{2}\nabla_{\delta}\delta G^{\alpha}{}_{\alpha}\delta G_{\overline{\alpha}}{}^{\delta} - \frac{1}{2}\nabla_{\overline{\alpha}}\delta G^{\alpha\beta}\delta G_{\alpha\beta} \Bigg] \delta X^{\overline{\alpha}} . \end{split}$$

$$(11.4)$$

Pending further simplifications, the vanishing of the expression in parentheses is the new Hollands-Wald gauge condition.

Graviational Constraints away from Hollands-Wald Gauge

In the Hollands-Wald analysis of section 8.4.1, the nonperturbative gravitational identity is given by equation (8.123) as

$$\delta_{\epsilon} \left(\delta E^{\text{grav}} - \delta S^{\text{grav}} \right) = \int_{\Sigma_{A}} \omega_{\text{grav}}(G, \delta_{\epsilon} G, \mathcal{L}_{\xi_{A}} G) + \omega_{\phi}(\phi, \delta_{\epsilon} \phi, \mathcal{L}_{\xi_{A}} \phi) + \int_{\Sigma_{A}} \mathcal{G} , \qquad (11.5)$$

and the perturbative metric ansatz by

$$G(\epsilon) = G_{AdS} + \epsilon G^{(1)} + \epsilon^2 G^{(2)} + \epsilon^3 G^{(3)} + \dots$$
 (11.6)

It is straightforward to write down the third order perturbative identity, which is given by

$$\delta^{(3)}(E^{\text{grav}} - S^{\text{grav}}) = \int_{\sigma_A} \left[\omega_{\text{grav}}(G^{(2)}, \mathcal{L}_{\xi_A} G^{(1)}) + \omega_{\text{grav}}(G^{(1)}, \mathcal{L}_{\xi_A} G^{(2)}) + \omega_{\phi}(\delta \phi^{(2)}, \mathcal{L}_{\xi_A} \phi^{(1)}) + \omega_{\phi}(\delta \phi^{(2)}, \mathcal{L}_{\xi_A} \phi^{(1)}) - 2\xi^a E^{(3)}_{ab} \epsilon^b \right],$$
(11.7)

where we have used $\mathcal{L}_{\xi_A} G^{(0)} = 0$ and that the first- and second order equations of motion are satisfied and denoted the third order Einstein equations sourced by the scalar field stress tensor by $E_{ab}^{(3)}$. We do not need to know the explicit form of this, since the point is to show that the CFT expression reproduces everything but the equations of motion term.

The last thing we need from the gravitational side of things is the expression away from Hollands-Wald gauge as well as the equations of motion for the vector field V. To generalize the reasoning from equation (8.126) onwards, it seems sufficient to consider

$$\gamma = h + \mathcal{L}_V G(\epsilon) , \qquad (11.8)$$

where we let γ be an (up to) second order perturbation that satisfies the second order Hollands-Wald gauge and h a general (up) second order perturbation. In this case, assuming that the firstand second order Einstein equations are satisfied the rest of the argument is unmodified up to and including the final result in equation (8.133)

$$\int_{\Sigma_A} \omega_{\text{grav}}(\gamma, \mathcal{L}_{\xi_A}, \gamma) = \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A}h) + \int_{\tilde{A}} \chi(h, [\xi_A, V]) , \qquad (11.9)$$

and all that is really needed is the addition of the term localized to the Ryu-Takayanagi surface.

To obtain the constraints on the vector field V we require that $G = h + \mathcal{L}_V G$ satisfies equation (11.4). This means that we should split γ into

$$\gamma^{(1)} = h_{ab}^{(1)} + 2\nabla^{(0)}_{(a} V_{b)}^{(1)}$$
(11.10)

and

$$\gamma^{(2)} = h^{(2)} + 2\nabla^{(0)}_{(a}V^{(2)}_{b)} + 2\Gamma^{c}_{ab;(1)}V^{(1)}_{c}$$

The first order variation of the Christoffel is

$$\delta\Gamma^{c}{}_{ab} = G^{cd}_{(0)} \left(\nabla_{(a}\gamma^{(1)}_{b)d} - \frac{1}{2}\nabla_{d}\gamma^{(1)}_{ab}\right)$$

which means that

$$\gamma^{(2)} = h^{(2)} + 2\nabla^{(0)}_{(a}V^{(2)}_{b)} + 2\nabla_{(a}\gamma^{(1)}_{b)c}V^{c}_{(1)} - \nabla^{(0)}_{c}\gamma^{(1)}_{ab}V^{c}_{(1)}$$

= $h^{(2)} + 2\nabla^{(0)}_{(a}V^{(2)}_{b)} + 2\nabla_{(a}h^{(1)}_{b)c}V^{c}_{(1)} + 6\nabla^{(0)}_{(a}\nabla^{(0)}_{b}V^{(1)}_{c}V^{c}_{(1)}$
 $- \nabla^{v}_{c}(0)h^{(1)}_{ab}V^{c}_{(1)} - 2\nabla^{v}_{c}(0)\nabla^{(0)}_{(a}V^{(1)}_{b}V^{c}_{(1)}$. (11.11)

We should insert these objects instead of δG and $\delta^2 G$ in equation (11.4), an exercise that we shall leave for future work.

11.2 CFT Side

Here we perform the first few steps of a third-order CFT computation. We begin by finding the third order relative entropy, and use results due to Lashkari et al. [107] to rewrite the relative entropy in terms of correlation functions. We conclude with some remarks on how to reinterpret these correlation functions as time-ordered to prepare for a future explicit computation using symplectic forms in an auxiliary AdS spacetime.

Relative Entropy to Third Order

The relative entropy is

$$S(\rho||\rho_0) = \text{Tr}[\rho \ln \rho - \rho \ln \rho_0] .$$
 (11.12)

For perturbations, we take $\rho(\epsilon) = \rho_0 + \epsilon \delta \rho + \epsilon^2 \delta^2 \rho + \epsilon^3 \delta^3 \rho + \mathcal{O}(\epsilon^4)$. The last terms will not contribute, since the second order perturbation to the relative entropy is proportional only to the square of the first order state perturbation.

The third order relative entropy is given by

$$\delta^3 S(\rho||\rho_0) = \frac{\mathrm{d}^3}{\mathrm{d}\epsilon^3} \mathrm{Tr}\left[\rho(\epsilon) \ln \rho(\epsilon) - \rho(\epsilon) \ln \rho_0\right]\Big|_{\epsilon=0} .$$
(11.13)

Let us leave the $\epsilon = 0$ implicit, denote $\frac{d}{d\epsilon}\rho \equiv \dot{\rho}$ and in the following computation:

$$\delta^{3}S(\rho||\rho_{0}) = \frac{1}{6} \frac{\mathrm{d}^{2}}{\mathrm{d}\epsilon^{2}} \operatorname{Tr} \left[\dot{\rho} \ln \rho + \rho(\dot{\ln \rho}) - \dot{\rho} \ln \rho_{0} \right]$$

$$= \frac{1}{6} \frac{\mathrm{d}}{\mathrm{d}\epsilon} \operatorname{Tr} \left[\ddot{\rho} \ln \rho + \dot{\rho}(\dot{\ln \rho}) - \ddot{\rho} \ln \rho_{0} \right]$$

$$= \frac{1}{6} \operatorname{Tr} \left[\ddot{\rho} \ln \rho + 2\ddot{\rho}(\dot{\ln \rho}) + \dot{\rho}(\ddot{\ln \rho}) - \ddot{\rho} \ln \rho_{0} \right]$$

$$= \frac{1}{6} \operatorname{Tr} \left[2\ddot{\rho}(\dot{\ln \rho}) + \dot{\rho}(\ddot{\ln \rho}) \right]$$
(11.14)

where in the first line we used that $\text{Tr}[\rho(\ln \rho)] = \text{Tr}[\dot{\rho}]$ vanishes to all orders in ϵ and in the last step we took the limit $\epsilon \to 0$ for the first term. To make sense of the logarithmic terms, we consider the identity from equation 8.60:

$$-\ln(\rho) = \int_0^\infty \frac{\mathrm{d}s}{s} \left(e^{-s\rho} - e^{-s} \right) \ . \tag{11.15}$$

We can combine this with the Baker-Campbell-Hausdorff related identity

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}e^{\rho(\epsilon)} = \int_0^1 \mathrm{d}y e^{(1-y)\rho} \frac{\mathrm{d}\rho}{\mathrm{d}\epsilon} e^{y\rho} \ . \tag{11.16}$$

Notably, to zeroth order in ϵ this tells us that

$$(\ln \rho) = \int_0^\infty ds \int_0^1 dy e^{-(1-y)\rho_0 s} \delta \rho e^{-y\rho_0 s}$$
(11.17)

related to the prior result in equation (8.61) by the change of variables x = 1 - y. For the second order perturbation of the logarithm, $\ddot{\rho}$, one can reuse the BCH related identity and evaluate the resulting double integrals. The result is rather messy as shown in Appendix A.3, but it possesses interesting cyclic symmetry. The presence of this symmetry is the observation that is at the center of a more sophisticated analysis carried out in [107], in which they find that

$$(\ln \rho)|_{\epsilon=0} = Q_1 + Q_2$$

= $-\frac{\pi}{2} \int_{-\infty}^{\infty} \frac{\mathrm{d}t}{\cosh^2(\pi t)} \boldsymbol{\delta}(t) - \frac{\pi}{4} \lim_{\epsilon \to 0} \int_{-\infty}^{\infty} \frac{\mathrm{d}t_1 \mathrm{d}t_2}{\cosh(\pi t_1) \cosh(\pi t_2)} g_{\epsilon}(t_2 - t_1)[\boldsymbol{\delta}(t_1), \boldsymbol{\delta}(t_2)]$
(11.18)

where $\boldsymbol{\delta}(t) = (\rho_0)^{-it} \boldsymbol{\delta}(\rho_0)^{it}$ and

$$g_{\varepsilon}(t) = \frac{i}{4} \left[\frac{1}{\sinh(\pi(t - i\varepsilon))} + \frac{1}{\sinh(\pi(t + i\varepsilon))} \right] , \qquad (11.19)$$

$$\boldsymbol{\delta} = \frac{\boldsymbol{\alpha}}{1 - \boldsymbol{\alpha}/2} \;, \tag{11.20}$$

$$\boldsymbol{\alpha} = 1 - (\rho_0)^{-\frac{1}{2}} \rho(\epsilon)(\rho_0)^{-\frac{1}{2}} . \tag{11.21}$$

Here the Q_i are coefficients in a series representation of a finite difference $\rho_0 - \rho(\epsilon)$ defined in equation (1.1) of [107]. Something noteworthy about the expansion is that is not an expansion in powers of ϵ , but the lowest power of ϵ present in Q_j is ϵ^j since α is zero at zeroth order. Since we are only interested of terms up to second order in ϵ , we can truncate the series with coefficient Q_j at j = 2.

We can rewrite the first order perturbation of the logarithm in terms of the same expansion, and it is simply

$$(\ln \rho)|_{\epsilon=0} = Q_1 = -\frac{\pi}{2} \int_{-\infty}^{\infty} \frac{\mathrm{d}t}{\cosh^2(\pi t)} \boldsymbol{\delta}(t) , \qquad (11.22)$$

where we will keep only the first order in ϵ part of $\delta(t)$, which is in contrast to the corresponding term in equation (11.18) where we keep only the second order part of $\delta(t)$. This is exactly the type of structural connection found between these terms in Appendix A.3.

It would be nice to express δ to first and second order in ϵ . First we realize that

$$\boldsymbol{\alpha} = -\rho_0^{-\frac{1}{2}} \epsilon \delta \rho \rho_0^{-\frac{1}{2}} - \rho_0^{-\frac{1}{2}} \epsilon^2 \delta^2 \rho \rho_0^{-\frac{1}{2}} .$$
 (11.23)

Inserting into the definition of δ we find

$$\begin{split} \boldsymbol{\delta} &= -\left(\rho_0^{-\frac{1}{2}}\epsilon\delta\rho\rho_0^{-\frac{1}{2}} + \rho_0^{-\frac{1}{2}}\epsilon^2\delta^2\rho\rho_0^{-\frac{1}{2}}\right)\left(1 - \frac{1}{2}\rho_0^{-\frac{1}{2}}\epsilon\delta\rho\rho_0^{-\frac{1}{2}}\right) + \mathcal{O}(\epsilon^3) \\ &= -\epsilon\left(\rho_0^{-\frac{1}{2}}\delta\rho\rho_0^{-\frac{1}{2}}\right) + \epsilon^2\left(\frac{1}{2}\rho_0^{-\frac{1}{2}}\delta\rho\rho_0^{-1}\delta\rho\rho_0^{-\frac{1}{2}} - \rho_0^{-\frac{1}{2}}\delta^2\rho\rho_0^{-\frac{1}{2}}\right) + \mathcal{O}(\epsilon^3) \;. \end{split}$$
(11.24)

We see that keeping only the first order part of $\boldsymbol{\delta}$ in equation (11.22) we recover the usual Fisher information as represented in equation (8.71) by substituting $s \to s - i\pi$. At this point we know the essentials about the relative entropy perturbation, so we should move on to the state.

Adding second order state perturbations

We are specifically considering the perturbation of a local density operator associated with the boundary ball-shaped region A. In section 8.4.2 we wrote down the first order state perturbation around the CFT vacuum local density operator on a ball shaped region (equation (8.140))

$$\delta \rho = \int \mathrm{d}^d x \lambda_\alpha(x) \rho_A^{(0)} \mathcal{O}_\alpha(x) \;. \tag{11.25}$$

Sticking to the "coherent" CFT state, the second order state perturbation is given by

$$\delta^2 \rho = \frac{1}{2} \int \mathrm{d}^d x_1 \mathrm{d}^d x_2 \lambda_\alpha(x_1) \lambda_\beta(x_2) \rho_A^{(0)} \mathcal{O}_\alpha(x_1) \mathcal{O}_\beta(x_2) , \qquad (11.26)$$

where $\rho_A^{(0)}$ is the unperturbed local state. The operators $\mathcal{O}(x)$ are time dependent, and are defined by modular time evolution as

$$\mathcal{O}_{\alpha}(x) \equiv \mathcal{O}_{\alpha}(\tau, \vec{x}) \equiv e^{\tau H_A} \mathcal{O}_{\alpha}(0, \vec{x}) e^{-\tau H_A} \frac{\Omega^{\Delta}(\tau, \vec{x})}{\Omega^{\Delta}(0, \vec{x})} , \qquad (11.27)$$

with Ω a necessary conformal factor. The next step is to classify the set of distinct terms that appear in Tr $\left[2\ddot{\rho}(\ln(\rho)) + \dot{\rho}(\ln(\rho))\right]$ written as correlation functions. The first step is writing out all of the terms that are to be rewritten explicitly. For cleanliness, let us drop the zero superscript on the zeroth order density operator, denoting it ρ_A :

$$\begin{aligned} \operatorname{Tr}[2\ddot{\rho}(\ln(\dot{\rho}))|_{\epsilon=0} &= \frac{\pi}{2} \int \mathrm{d}^{d}x_{1} \mathrm{d}^{d}x_{2} \mathrm{d}^{d}x_{3} \lambda_{\alpha}(x_{1}) \lambda_{\beta}(x_{2}) \lambda_{\gamma}(x_{3}) \\ &\times \int_{-\infty}^{\infty} \mathrm{d}s \frac{1}{\cosh^{2}(\pi s)} \operatorname{Tr}\left[\rho_{A} \mathcal{O}_{\alpha}(x_{1}) \mathcal{O}_{\beta}(x_{2}) \rho_{A}^{-is} \rho_{A}^{-\frac{1}{2}} \rho_{A} \mathcal{O}_{\gamma}(x_{3}) \rho_{A}^{-\frac{1}{2}} \rho_{A}^{is}\right] \\ \operatorname{Tr}[\dot{\rho}(\ln(\ddot{\rho}))|_{\epsilon=0} &= \frac{\pi}{4} \int \mathrm{d}^{d}x_{1} \mathrm{d}^{d}x_{2} \mathrm{d}^{d}x_{3} \lambda_{\alpha}(x_{1}) \lambda_{\beta}(x_{2}) \lambda_{\gamma}(x_{3}) \\ &\left(\int Q_{1}\right) :\times \left\{-\frac{1}{2} \int_{-\infty}^{\infty} \mathrm{d}s \frac{1}{\cosh^{2}(\pi s)} \operatorname{Tr}\left[\rho_{A} \mathcal{O}_{\alpha}(x_{1}) \rho_{A}^{-is} \rho_{A}^{-\frac{1}{2}} \rho_{A} \mathcal{O}_{\beta}(x_{2}) \mathcal{O}_{\gamma}(x_{3}) \rho_{A}^{-\frac{1}{2}} \rho_{A}^{is}\right] \\ &+ \int_{-\infty}^{\infty} \mathrm{d}s \frac{1}{\cosh^{2}(\pi s)} \operatorname{Tr}\left[\rho_{A} \mathcal{O}_{\alpha}(x_{1}) \rho_{A}^{-is} \rho_{A}^{-\frac{1}{2}} \rho_{A} \mathcal{O}_{\beta}(x_{2}) \mathcal{O}_{\gamma}(x_{3}) \rho_{A}^{-\frac{1}{2}} \rho_{A}^{is}\right] \\ &\left(\int Q_{2}\right) :+ \frac{1}{2} \int_{-\infty}^{\infty} \frac{\mathrm{d}s_{1} \mathrm{d}s_{2} \, g_{\varepsilon}(s_{2} - s_{1})}{\cosh(\pi s_{1}) \cosh(\pi s_{2})} \operatorname{Tr}\left[\rho_{A} \mathcal{O}_{\alpha}(x_{1}) \left[\rho^{-is_{1}} \rho^{-\frac{1}{2}} \rho_{A} \mathcal{O}_{\beta} \rho^{-\frac{1}{2}} \rho^{is_{1}}, \rho^{-is_{2}} \rho^{-\frac{1}{2}} \rho_{A} \mathcal{O}_{\gamma} \rho^{-\frac{1}{2}} \rho^{is_{2}}\right]\right] \right\} \end{aligned}$$

This leaves us with four terms that should be rewritten as time-ordered correlation functions. This entails the substitutions $s_i \to s_i \pm (\frac{i}{2} - \epsilon)$, $\rho_A = e^{-H_A}$ and inserting the definition of the time dependent conformal operators resulting in some conformal factors. The shift results in $\cosh(\pi s) \to \sinh(\pi(s - \epsilon))$ by using the imaginary periodicity of the hyperbolic functions. The resulting correlation functions are

$$\operatorname{Tr}[2\ddot{\rho}(\operatorname{ln}(\dot{\rho}))|_{\epsilon=0} = \frac{\pi}{2} \int \mathrm{d}^{d} x_{1} \mathrm{d}^{d} x_{2} \mathrm{d}^{d} x_{3} \lambda_{\alpha}(x_{1}) \lambda_{\beta}(x_{2}) \lambda_{\gamma}(x_{3}) A \qquad (11.30)$$

$$A = \int_{-\infty}^{\infty} \mathrm{d} s \frac{1}{\sinh^{2}(\pi(s-\epsilon))} \operatorname{Tr}\left[e^{-H_{A}}\mathcal{O}_{\alpha}(x_{1})\mathcal{O}_{\beta}(x_{2})e^{iHs}e^{(\frac{1}{2}\mp\frac{1}{2})H_{A}}e^{-H_{A}}\mathcal{O}_{\gamma}(x_{3})e^{(\frac{1}{2}\pm\frac{1}{2})H_{A}}e^{-iHs}\right]$$

as well as

$$\operatorname{Tr}[\dot{\rho}(\ln\ddot{(}\rho))|_{\epsilon=0} = \frac{\pi}{4} \int d^{d}x_{1} d^{d}x_{2} d^{d}x_{3}\lambda_{\alpha}(x_{1})\lambda_{\beta}(x_{2})\lambda_{\gamma}(x_{3}) \times (B+C+D)$$
(11.31)

$$B = -\frac{1}{2} \int_{-\infty}^{\infty} ds \frac{1}{\sinh^{2}(\pi(s-\epsilon))} \operatorname{Tr}\left[e^{-H_{A}}\mathcal{O}_{\alpha}(x_{1})e^{iHs}e^{(\frac{1}{2}\pm\frac{1}{2})H_{A}}e^{-H_{A}}\mathcal{O}_{\beta}(x_{2})\mathcal{O}_{\gamma}(x_{3})e^{(\frac{1}{2}\pm\frac{1}{2})H_{A}}e^{-iHs}\right]$$

$$C = \int_{-\infty}^{\infty} ds \frac{1}{\sinh^{2}(\pi(s-\epsilon))} \operatorname{Tr}\left[e^{-H_{A}}\mathcal{O}_{\alpha}(x_{1})e^{iHs}e^{(\frac{1}{2}\pm\frac{1}{2})H_{A}}e^{-H_{A}}\mathcal{O}_{\beta}(x_{2})\mathcal{O}_{\gamma}(x_{3})e^{(\frac{1}{2}\pm\frac{1}{2})H_{A}}e^{-iHs}\right]$$

$$D = \frac{1}{2} \int_{-\infty}^{\infty} \frac{ds_{1}ds_{2}g_{\varepsilon}(s_{2}-s_{1})}{\sinh(\pi(s_{1}-\epsilon_{1}))\sinh(\pi(s_{2}-\epsilon)_{2})}$$

$$\operatorname{Tr}\left[e^{-H_{A}}\mathcal{O}_{\alpha}(x_{1})\left[e^{-iH_{A}s_{1}}e^{(\frac{1}{2}\pm\frac{1}{2})H_{A}}e^{-H_{A}}\mathcal{O}_{\beta}e^{(\frac{1}{2}\pm\frac{1}{2})H_{A}}e^{iH_{A}s_{1}}, e^{-iH_{A}s_{2}}e^{(\frac{1}{2}\pm\frac{1}{2})H_{A}}e^{-H_{A}}\mathcal{O}_{\gamma}e^{(\frac{1}{2}\pm\frac{1}{2})H_{A}}e^{iH_{A}s_{2}}\right]$$

where we have taken the liberty of labelling the terms for later reference and kept the limit $\epsilon \to 0^+$ implicit. In the following, let us define

$$\int \mathrm{d}\boldsymbol{\mu} = \int \mathrm{d}^d x_1 \mathrm{d}^d x_2 \mathrm{d}^d x_3 \lambda_\alpha(x_1) \lambda_\beta(x_2) \lambda_\gamma(x_3) \ . \tag{11.32}$$

We will now proceed to give the + and - versions of each of the terms A, B, C, D. In anticipation of time-ordering we expect the sign choices to correspond to different operator orderings. We then hope to be able to pick \pm and integration intervals such that the correlation functions are always ordered with the rightmost operator inserted at the earliest times. This is supposed to result in correlation functions that are time-ordered with respect to the integrands $t_i \subset x_i$, at which point we can apply the conventional CFT machinery to extract some information about the relative entropy to third order:

$$(A)^{+} = \frac{\pi}{2} \int d\boldsymbol{\mu} \int \frac{ds}{\sinh^{2}(\pi(s-\epsilon))} \frac{\Omega(x_{3})}{\Omega(x_{3}-is)} \operatorname{Tr} \left[e^{-H_{A}} \mathcal{O}_{\gamma}(x_{3}-is) \mathcal{O}_{\alpha}(x_{1}) \mathcal{O}_{\beta}(x_{2}) \right] ,$$

$$(A)^{-} = \frac{\pi}{2} \int d\boldsymbol{\mu} \int \frac{ds}{\sinh^{2}(\pi(s-\epsilon))} \frac{\Omega(x_{3})}{\Omega(x_{3}-is)} \operatorname{Tr} \left[e^{-H_{A}} \mathcal{O}_{\alpha}(x_{1}) \mathcal{O}_{\beta}(x_{2}) \mathcal{O}_{\gamma}(x_{3}-is) \right] ,$$

$$(11.33)$$

$$(B)^{+} = -\frac{1}{2}(C)^{+} = \frac{\pi}{8} \int d\boldsymbol{\mu} \int \frac{ds}{\sinh^{2}(\pi(s-\epsilon))} \frac{\Omega(x_{1})}{\Omega(x_{1}+is)} \operatorname{Tr} \left[e^{-H_{A}} \mathcal{O}_{\beta}(x_{2}) \mathcal{O}_{\gamma}(x_{3}) \mathcal{O}_{\alpha}(x_{1}+is) \right] ,$$

$$(B)^{-} = -\frac{1}{2}(C)^{-} = \frac{\pi}{4} \int d\boldsymbol{\mu} \int \frac{ds}{\sinh^{2}(\pi(s-\epsilon))} \frac{\Omega(x_{1})}{\Omega(x_{1}+is)} \operatorname{Tr} \left[e^{-H_{A}} \mathcal{O}_{\alpha}(x_{1}+is) \mathcal{O}_{\beta}(x_{2}) \mathcal{O}_{\gamma}(x_{3}) \right] ,$$

$$(11.34)$$

The *D* term has four possible sign structures coming from the four different ways of choosing ϵ_1 and ϵ_2 . We will use ++ to denote the positive sign for the substitution with both, and +- to denote a positive sign for ϵ_1 and negative for ϵ_2 . We then have

$$(D)^{++} = \frac{\pi}{8} \int d\mu \int ds \operatorname{Tr} \left[e^{-H_A} \mathcal{O}_{\alpha}(x_1) \left[e^{-H_A} \mathcal{O}_{\beta}(x_2 - is_1) e^{H_A}, e^{-H_A} \mathcal{O}_{\gamma}(x_3 - is_2) e^{H_A} \right] \right] , = \frac{\pi}{8} \int d\mu \int ds \operatorname{Tr} \left[e^{-H_A} \left[\mathcal{O}_{\beta}(x_2 - is_1), \mathcal{O}_{\gamma}(x_3 - is_2) \right] \mathcal{O}_{\alpha}(x_1) \right] (D)^{--} = \frac{\pi}{8} \int d\mu \int ds \operatorname{Tr} \left[e^{-H_A} \mathcal{O}_{\alpha}(x_1) \left[\mathcal{O}_{\beta}(x_2 - is_1), \mathcal{O}_{\gamma}(x_3 - is_2) \right] \right] ,$$
 (11.35)

$$(D)^{+-} = \frac{\pi}{8} \int d\mu \int ds \operatorname{Tr} \left[e^{-H_A} \mathcal{O}_{\alpha}(x_1) \left[e^{-H_A} \mathcal{O}_{\beta}(x_2 - is_1) e^{H_A}, \mathcal{O}_{\gamma}(x_3 - is_2) \right] \right] , (D)^{-+} = \frac{\pi}{8} \int d\mu \int ds \operatorname{Tr} \left[e^{-H_A} \mathcal{O}_{\alpha}(x_1) \left[\mathcal{O}_{\beta}(x_2 - is_1), e^{-H_A} \mathcal{O}_{\gamma}(x_3 - is_2) e^{H_A} \right] \right] ,$$

where we defined

$$\int d\mathbf{s} = \int \frac{ds_1 ds_2 g_{\varepsilon}}{\cosh(\pi(s_1 - \epsilon_1)) \cosh(\pi(s_2 - \epsilon_2))} \frac{\Omega(x_2)}{\Omega(x_2 - is_1)} \frac{\Omega(x_3)}{\Omega(x_3 - is_1)} .$$
(11.36)

Something to note is that the commutators are especially unpleasant to handle for the +- and -+ cases, as we will get several occurrences of the operator $e^{\pm H_A}$ that are not easily removed.

Incomplete Time-Ordering and Operator Product Expansions

Partial time-ordering for equation (11.33) is simply obtained by picking the +-labelled substitution when $t_3 > t_2 > t_1$ and the --labelled substitution when $t_2 > t_1 > t_3$. The question is then to understand whether the time-orderings

$$t_1 > t_2$$
$$t_1 > t_3 > t_2$$
$$t_2 > t_3 > t_1$$

can or need to naturally appear with the associated operators correspondingly shuffled around. To complete the time ordering it seems that the only solution is to turn the pair of operators that is kept fixed under the choice +- in the *s* substitution into a single operator or a sum of single operators. That is, we reexpress the two operators as an operator product expansion

$$\mathcal{O}_{\alpha}(x_1)\mathcal{O}_{\beta}(x_2) = \sum_{i=-k}^{\infty} c_i \mathcal{O}_{\alpha\beta}^i(x_2)(x_1 - x_2)^i \equiv \Xi_{\alpha\beta}(x_2) , \qquad (11.37)$$

where k is chosen such that \mathcal{O}^k respects the relevant unitarity bounds of the CFT given the Lorentz structure of the indices α, β . Validity of the equality between the *unordered* operators \mathcal{O} and the OPE on the operator level is guaranteed in Euclidean field theory, and since the fused operators are purely Euclidean this should not be a complication.

Given the OPE, equations (11.33), (11.34), (11.35) describe time-ordered correlation functions between Ξ and a third operator \mathcal{O}_{γ} for appropriate choices of when to use the + and – prescriptions. That is, we are considering correlation functions in the integrand on the form $\langle \mathcal{T}(\mathcal{O}_{\gamma}(x_3)\Xi(x_2))\rangle$. Note that there is no guarantee that the integration over $\int d\mu$ depends only on the residues of the OPE, so we may need the entire expansion.

For the case of \mathcal{O} chosen as conformal primary real scalar fields of weight Δ , the only part of the OPE that is nonzero in the correlation function is

$$\langle \phi(x_3) \Xi_{\phi}(x_2) \rangle \sim \frac{c_{\Delta}}{(x_1 - x_2)^{\Delta}} \langle \phi(x_3) \phi(x_2) \rangle$$
 (11.38)

due to the fact that scalar field correlators are only nonzero when both scalars have the same conformal weight as we saw in section 3.4.3. This should in principle reduce to the second order case plus an extra singular term $\frac{1}{(x_2-x_3)^{\Delta}}$ that needs to be dealt with when integrating the $d\mu$ part. To further this analysis, the next step is to understand stress-tensor and scala field three-point OPEs, so that the relevant correlators may be evaluated.

Chapter 12

Summary & Conclusions

In this chapter, we conclude this thesis part by part. We summarize what was discovered, and whether the parts did what they set out to do. We detail some possible continuations of each part.

Preliminaries

In *Part I: Preliminaries* we set out to give a comprehensive review of the fundamental theory necessary to understand the material covered in Part II. A main consideration was the attempt to give a very detailed account of every step along the way, with the aim of making the material more accessible to early graduate students.

Part I started from elementary quantum mechanics, defining and developing the density operator formalism and its relation to Shannon's theory of information. Most importantly we learned that all entropy can be explained as entropy of entanglement with a purifying system. We then derived the path integral formulation of quantum mechanics, which we then generalized to quantum fields. Our principal discovery was the path integral generalization of Noether's theorem in the Ward identities, which allowed us to understand conformal field theory without ever writing down a Lagrangian. We went on to generalize the density operator formalism to the case of quantum fields, discovering the divergent entanglement of the QFT vacuum with itself. With the help of string theory we then discovered the AdS/CFT correspondence, a deep connection between D-dimensional non-gravitational physics and D+1-dimensional gravitational physics. We generalized the idea of the entanglement entropy of fields to the entanglement entropy of gravity via this correspondence, discovering that the very presence of entanglement is the glue that builds geometry.

Some things were not done in quite as much detail as everything else, most notably the Lewkowycz-Parrikar proofs of the RT and HRT formulas. This was because the necessary machinery of general relativity and hypersurfaces had not been introduced. As such an obvious extension of Part I would be the addition of a chapter on specialized topics in general relativity focusing on hypersurfaces and Iyer-Wald formalism. This would also be of benefit for Part II, where some sections require the introduction of this machinery on the fly.

For the purpose of extending the analysis in chapter 11 it would also be useful to add to the CFT section a passage about the conformal bootstrap and higher point functions.

Another interesting thing to add would be the theory of covariant phase space methods for computing correlation functions to the end of the AdS/CFT chapter. This theory would put the second order CFT analysis of section 8.4.2 on much more solid footing. As it is now the covariant phase space formalism was introduced ad-hoc and is just verified to reproduce the usual correlation functions.

In principle, another interesting extension would be the addition of an entire "Quantum correc-

tions to Ryu-Takayanagi" track. Such a track would include an introduction to supergravity, some details about higher derivative gravity and how the HRT formula is affected by such a modification. One could then explore quantum corrected variants of the material in chapter 8. In contrast to other suggestions this would entail a genuine expansion of the scope of the thesis, rather than an addition for the sake of completion.

Recent Developments

The objective of *Part II: Recent Developments* was to give a comprehensive review of the current state of emergent gravity, with a focus on the entanglement=geometry viewpoint. The intention was for this presentation to be significantly more digestible than the original papers (even if many of these were very high in detail!).

In chapter 8 we began by realizing that in wanting to recover a dynamical notion of geometry, we should first understand the dynamics of entanglement. This was characterized by the entanglement first law and the perturbative relative entropy identities between entanglement entropy and the CFT modular energy. These two quantities were related holographically to the areas of HRT surfaces and the gravitational modular energy, respectively. We showed that the consistency of the relation required that the gravitational theory was uniquely described by the Einstein field equations up to second order in perturbations about pure AdS.

In chapter 9 we discovered that the interior of black holes, which was not described by entanglement, could be described by the microscopic structure of the CFT state. This was achievable via the CV/CA conjectures as well as by preparation of very particular states as in section 9.4.

In chapter 10 we discovered that the geometrical interpretation of the boundary was not necessarily important to the geometry of the bulk spacetime. This allowed us to play with non-geometric ideas such as tensor network cosmology, which we showed to satisfy several consistency conditions.

Like in the case of Part I, the level of depth and detail was not as constant as desired. Chapter 8 concludes with a detailed analysis of results as recent as 2017, referring to 2019 papers for some generalization of the end result. In contrast, chapter 9 does not bring the CA and CV dualities to the present day in the same level of detail. For chapter 9 to match chapter 8 in detail, I think a concluding section explaining the details of *Complexity of Jackiw-Teitelboim gravity* [95] would be a very good extension. Similarly a section on *Bulk Entanglement Gravity* [8] to finish off chapter 10 would go a long way towards consistency of depth.

Original Work

The goal of *Part II: Original Work* was to find interesting results that exist nowhere else in the literature. The problem that was approached was the extension of section 8.4 to third order in perturbations. What we concluded was that the third order analysis may as expected probe requirements on the operator product expansion imposed by holography. In addition we found that for primary scalars the conformal dimension of the operators becomes relevant, indicating that we should consider quantum effects on the gravity side. Progress towards quantum corrected variants of the Iyer-Wald formalism has been made very recently in [79].

To continue the analysis, one should write down the vector field that interpolates between general and Hollands-Wald gauges, understand the field theory correlators for the $A^{\pm}, B^{\pm}, D^{\pm}$ -terms in terms of symplectic forms in an auxiliary AdS space and carry out the necessary residue/branch cut analysis as in section 8.4.1.

In conclusion we have made some progress on the third order problem, and both the gravitational and CFT sides have some clear avenues of development. With a clearer understanding of the higher order conformal operators and quantum corrected Iyer-Wald formalism, verification of the emergence of gravity from entanglement to third order in perturbations seems to be within reach.

Appendices

Appendix A

Miscellaneous Proofs and Calculations

A.1 Vanishing of equation (8.131)

Let us write out explicitly the proof that equation (8.131) vanishes:

$$\begin{split} \chi(\dot{\gamma}, V)_{\tilde{A}} &\sim \epsilon_{+-} \left(\nabla_{[-}\dot{\gamma}_{+]}\overline{\alpha}V^{\overline{\alpha}} - \nabla_{\overline{\alpha}}\dot{\gamma}^{\overline{\alpha}}_{[+}V_{-]} + \nabla_{[+|}\dot{\gamma}_{\overline{\alpha}}^{\overline{\alpha}}V_{[-]} \right) \\ &= \frac{1}{2} \epsilon_{+-} \left(\nabla_{-}\dot{\gamma}_{++}V^{+} + \overline{\nabla_{-}}\dot{\gamma}_{++}V^{+} - \overline{\nabla_{+}}\dot{\gamma}_{-+}V^{+} - \nabla_{+}\dot{\gamma}_{--}V^{-} \right) \\ &- \nabla_{+}\dot{\gamma}^{+}_{+}V_{-} - \overline{\nabla_{-}}\dot{\gamma}_{-+}V_{-} + \overline{\nabla_{+}}\dot{\gamma}_{+}^{+}\nabla_{+} - \nabla_{-}\dot{\gamma}_{--}V_{+} \\ &+ \nabla_{+}\dot{\gamma}_{+}^{+}V_{-} + \nabla_{+}\dot{\gamma}_{-}^{+}V_{-} - \overline{\nabla_{-}}\dot{\gamma}_{+}^{+}V_{+} - \nabla_{-}\dot{\gamma}_{--}V_{+} \\ &= \frac{1}{2} \epsilon_{+-} \left(- \nabla_{+}\dot{\gamma}_{--}V^{-} \\ &- \overline{\nabla_{+}}\dot{\gamma}_{+}^{+}V_{-} + \overline{\nabla_{+}}\dot{\gamma}_{-}^{+}V_{-} - \overline{\nabla_{+}}\dot{\gamma}_{-}^{+}V_{+} \right) \\ &= \frac{1}{2} \epsilon_{+-} \left(- \nabla_{+}\dot{\gamma}_{--}V^{-} \\ &- \overline{\nabla_{+}}\dot{\gamma}_{+}^{+}V_{-} + \overline{\nabla_{+}}\dot{\gamma}_{-}^{+}V_{-} - \overline{\nabla_{+}}\dot{\gamma}_{-}^{+}V_{+} \right) \\ &= 0 \; . \end{split}$$

A.2 Derivation of Null Singularities of Wightman Propagators

In section 8.4.2 we consider the null singularities of the propagator

$$K_{\pm} \sim \lim_{\varepsilon \to 0^+} \frac{1}{\left(-2rY_B \cdot Y_b - 2\sqrt{r^2 - 1}\cosh(s - t_B \pm i\varepsilon)\right)^{\Delta}}$$

For example, we may look for singularities in s_B given s. This means we are looking to solve

$$\frac{rY_B \cdot Y_b}{\sqrt{r^2 - 1}} + \cosh(s - t_B) = 0 \ .$$

This is solved by $t_B = s - \operatorname{arccosh}\left(\frac{-rY_B\dot{Y}_b}{\sqrt{r^2-1}}\right)$. Defining the argument of the arccosh to be α , we use that

$$\operatorname{arccosh}(\alpha) = \ln\left(\alpha \pm \sqrt{\alpha^2 - 1}\right)$$
.

Note that we are close to the negative real axis in α , so in principle we may want to be careful of branch cuts in the logarithm. This should not be an issue, as we will exponentiate α in all applications. It now follows that the singularities are at

$$t_B = s - \ln\left(\alpha \pm \sqrt{\alpha^2 - 1}\right) \,.$$

Notably this is not the same as [63], although our sign on the logarithmic term is necessary to replicate their expansions.

A.3 Naive Third Order Perturbation of Relative Entropy

Here I present the naive approach to the computation of the logarithm to second order. While the calculation itself is not soluble, it exhibits some of the symmetry noted in [107], so it remains interesting. The second derivative of the logarithm of an operator, denoted $\ddot{\rho}$ has us compute

$$\begin{aligned} \ddot{(\ln \rho)} &= \int_0^\infty \mathrm{d}s \int_0^1 \mathrm{d}y \left(\frac{\mathrm{d}}{\mathrm{d}\epsilon} e^{-(1-y)\rho s} \frac{\mathrm{d}\rho}{\mathrm{d}\epsilon} e^{-y\rho s} + e^{-(1-y)\rho s} \frac{\mathrm{d}^2 \rho}{\mathrm{d}\epsilon^2} e^{-y\rho s} + e^{-(1-y)\rho s} \frac{\mathrm{d}\rho}{\mathrm{d}\epsilon} \frac{\mathrm{d}}{\mathrm{d}\epsilon} e^{-y\rho s} \right) \\ &= \int_0^\infty \mathrm{d}s \int_0^1 \mathrm{d}y \left(-(1-y)s \int_0^1 \mathrm{d}z e^{-(1-z)(1-y)\rho s} \frac{\mathrm{d}\rho}{\mathrm{d}\epsilon} e^{-z(1-y)\rho s} \frac{\mathrm{d}\rho}{\mathrm{d}\epsilon} e^{-y\rho s} \right) \\ &+ e^{-(1-y)\rho s} \frac{\mathrm{d}^2 \rho}{\mathrm{d}\epsilon^2} e^{-y\rho s} \\ &- ys \int_0^1 \mathrm{d}z e^{-(1-y)\rho s} \frac{\mathrm{d}\rho}{\mathrm{d}\epsilon} e^{-(1-z)y\rho s} \frac{\mathrm{d}\rho}{\mathrm{d}\epsilon} e^{-zy\rho s} \right) . \end{aligned}$$
(A.1)

The zeroth order in ϵ term is read off to be

$$(\ddot{\ln \rho}) = \int_0^\infty ds \int_0^1 dy \left(-(1-y)s \int_0^1 dz e^{-(1-z)(1-y)\rho_0 s} \delta\rho e^{-z(1-y)\rho_0 s} \delta\rho e^{-y\rho_0 s} + 2e^{-(1-y)\rho_0 s} \delta^2 \rho e^{-y\rho_0 s} -ys \int_0^1 dz e^{-(1-y)\rho_0 s} \delta\rho e^{-(1-z)y\rho_0 s} \delta\rho e^{-zy\rho_0 s} \right).$$
(A.2)

Evaluating $\mathbf{Tr}[2\ddot{\rho}(\ln \rho)]$

We now have integrals we wish to compute. This is simplified by picking a basis in which ρ_0 is diagonalized and writing the integrand in component form. Let us start with the term that requires only two integrals:

$$\operatorname{Tr}[2\ddot{\rho}(\ln\dot{\rho})]\Big|_{\epsilon=0} = 4\int_0^\infty \mathrm{d}s \int_0^1 \mathrm{d}y \operatorname{Tr}\left[\delta^2 \rho e^{-(1-y)\rho_0 s} \delta \rho e^{-y\rho_0 s}\right]$$
(A.3)

By picking a basis in which ρ_0 is diagonalized we can express the trace in terms of the eigenvalues of ρ_0 , ρ_a as

$$(\delta^2 \rho)^b{}_a e^{-(1-y)\rho_a s} (\delta \rho)^a{}_b e^{-y\rho_b s} = e^{y(\rho_a - \rho_b)s - \rho_a s} (\delta^2 \rho)^b{}_a (\delta \rho)^a{}_b , \qquad (A.4)$$

where sums over a and b are implied. For $\rho_a = \rho_b$ the integration over y is trivial, while in the case $\rho_a \neq \rho_b$ it is elementary. We find that

$$\begin{aligned} \operatorname{Tr}[2\ddot{\rho}(\dot{\ln\rho})]\Big|_{\epsilon=0} &= 4\int_{0}^{\infty} \mathrm{d}s \bigg[\sum_{\rho_{a}=\rho_{b}} e^{-\rho_{a}s} (\delta^{2}\rho)^{b}{}_{a} (\delta\rho)^{a}{}_{b} + \sum_{\rho_{a}\neq\rho_{b}} e^{-\rho_{a}s} \frac{1}{(\rho_{a}-\rho_{b})s} \left(e^{(\rho_{a}-\rho_{b})s} - 1\right) (\delta^{2}\rho)^{b}{}_{a} (\delta\rho)^{a}{}_{b} \\ &= 4\int_{0}^{\infty} \mathrm{d}s \bigg[\sum_{\rho_{a}=\rho_{b}} e^{-\rho_{a}s} (\delta^{2}\rho)^{b}{}_{a} (\delta\rho)^{a}{}_{b} + \sum_{\rho_{a}\neq\rho_{b}} \frac{e^{-\rho_{b}s} - e^{-\rho_{a}s}}{(\rho_{a}-\rho_{b})s} (\delta^{2}\rho)^{b}{}_{a} (\delta\rho)^{a}{}_{b}\bigg] \end{aligned} \tag{A.5}$$

The first term is simple to integrate, while for the second we reuse equation (11.15):

$$\int_{0}^{\infty} \mathrm{d}s \sum_{\substack{\rho_{a} \neq \rho_{b}}} \frac{e^{-\rho_{b}s} - e^{-s} - e^{-\rho_{a}s} + e^{-s}}{(\rho_{a} - \rho_{b})s} (\delta^{2}\rho)^{b}{}_{a}(\delta\rho)^{a}{}_{b}$$

$$= \sum_{\substack{\rho_{a} \neq \rho_{b}}} \frac{\ln \rho_{a} - \ln \rho_{b}}{(\rho_{a} - \rho_{b})} (\delta^{2}\rho)^{b}{}_{a}(\delta\rho)^{a}{}_{b} .$$
(A.6)

And we see finally that

$$\frac{1}{4} \text{Tr}[2\ddot{\rho}(\ln\rho)] = \sum \frac{1}{\rho_a} (\delta^2 \rho)^b{}_a (\delta\rho)^a{}_b + \sum_{\rho_a \neq \rho_b} \frac{\ln\rho_a - \ln\rho_b}{(\rho_a - \rho_b)} (\delta^2 \rho)^b{}_a (\delta\rho)^a{}_b .$$
(A.7)

Finally, we put (A.7) on a form that is related to time-ordered correlators by using equation (8.67)

$$\frac{1}{4} \int \mathrm{d}s \frac{e^{\frac{isx}{2\pi}}}{1 + \cosh(s)} = \frac{x}{2(e^{\frac{x}{2}} - e^{-\frac{x}{2}})} , \qquad (A.8)$$

with $x = \ln\left(\frac{\rho_a}{\rho_b}\right)$. The result is

$$\operatorname{Tr}[2\ddot{\rho}(\ln\rho)] = 2 \int_{-}^{\infty} \mathrm{d}s \frac{1}{1+\cosh(s)} \frac{1}{\sqrt{\rho_a \rho_b}} \left(\frac{\rho_a}{\rho_b}\right)^{\frac{is}{2\pi}} (\delta^2 \rho)^b{}_a(\delta\rho)^a{}_b$$

$$= 2 \int_{-\infty}^{\infty} \mathrm{d}s \frac{1}{1+\cosh(s)} \operatorname{Tr}\left[\delta^2 \rho \rho_0^{-\frac{1}{2}+\frac{is}{2\pi}} \delta \rho \rho_0^{-\frac{1}{2}-\frac{is}{2\pi}}\right] ,$$
(A.9)

which by substituting $s \to s \pm i\pi(1-\epsilon)$ and using $\cosh(s+i\pi) = -\cosh(s)$ and $\cosh(s) - 1 = 2\sinh^2(s/2)$ gives us

(+):
$$\operatorname{Tr}[2\ddot{\rho}(\ln\rho)] = -2\int_{-\infty}^{\infty} \mathrm{d}s \frac{1}{2\sinh^2(\frac{s+i\epsilon}{2})} \operatorname{Tr}\left[\rho_0^{-1}\rho_0^{\frac{is}{2\pi}}\delta\rho\rho_0^{-\frac{is}{2\pi}}\delta^2\rho\right]$$
, (A.10)

$$(-): \quad \operatorname{Tr}[2\ddot{\rho}(\ln\rho)] = -2\int_{-\infty}^{\infty} \mathrm{d}s \frac{1}{2\sinh^2(\frac{s-i\epsilon}{2})} \operatorname{Tr}\left[\rho_0^{-1}\delta^2\rho\rho_0^{\frac{is}{2\pi}}\delta\rho\rho_0^{-\frac{is}{2\pi}}\right] . \tag{A.11}$$

Then, by introducing the modular Hamiltonian $H \equiv -\ln(\rho_0)$ we see that depending on the choice of sign we get two operator orderings of the correlation function between $\delta\rho(s/2\pi) \equiv e^{-\frac{is}{2\pi}H} \delta\rho e^{\frac{is}{2\pi}H}$ and $\delta^2\rho(0)$ evaluated in a thermal ensemble with respect to H.

Evaluating the second row of $\text{Tr}[\dot{\rho}(\ln \rho)]$ in equation (A.2)

We will now turn to a second term with only two integrals to perform. We consider

$$\operatorname{Tr}[\dot{\rho}(\ln\rho)] \supset 2\int_0^\infty \mathrm{d}s \int_0^1 \mathrm{d}y \operatorname{Tr}[\delta\rho e^{-(1-y)\rho_0 s} \delta^2 \rho e^{-y\rho_0 s}] .$$
(A.12)

This is just $\frac{1}{2}$ times equation (A.3) with $\delta \rho$ and $\delta^2 \rho$ exchanged, so we know how to write down the correct result:

$$(+): \quad \mathrm{Tr}[\dot{\rho}(\ln \rho)] \supset -\int_{-\infty}^{\infty} \mathrm{d}s \frac{1}{2\sinh^2(\frac{s+i\epsilon}{2})} \mathrm{Tr}\left[\rho_0^{-1}\rho_0^{-\frac{is}{2\pi}} \delta^2 \rho \rho_0^{\frac{is}{2\pi}} \delta \rho\right] , \tag{A.13}$$

$$(-): \quad \operatorname{Tr}[\dot{\rho}(\ln \rho)] \supset -\int_{-\infty}^{\infty} \mathrm{d}s \frac{1}{2\sinh^2(\frac{s-i\epsilon}{2})} \operatorname{Tr}\left[\rho_0^{-1}\delta\rho\rho_0^{-\frac{is}{2\pi}}\delta^2\rho\rho_0^{\frac{is}{2\pi}}\right] . \tag{A.14}$$

By introducing the modular Hamiltonian $H \equiv \ln(\rho_0)$ we realize that the essential difference to the previous case is that $\delta^2 \rho(s/2\pi) \equiv e^{-\frac{is}{2\pi}H} \delta^2 \rho e^{\frac{is}{2\pi}H}$ is the operator that undergoes Lorentzian time evolution by the modular Hamiltonian.

Evaluating the first row of $\text{Tr}[\dot{\rho}(\ln \rho)]$ (equation (A.2))

We consider

$$\operatorname{Tr}[\dot{\rho}(\ln \rho)] \supset -2\int_0^\infty \mathrm{d}s \int_0^1 \mathrm{d}y(1-y)s \int_0^1 \mathrm{d}z \operatorname{Tr}[\delta\rho e^{-(1-z)(1-y)\rho_0 s}\delta\rho e^{-z(1-y)\rho_0 s}\delta\rho e^{-y\rho_0 s}] . \quad (A.15)$$

Let us diagonalize ρ_0 , calling the eigenvalues ρ_a and passing to tensor notation

$$Tr[...] = \delta \rho^{a}{}_{b} e^{-(1-z)(1-y)\rho_{b}s} \delta \rho^{b}{}_{c} e^{-z(1-y)\rho_{c}s} \delta \rho^{c}{}_{a} e^{-y\rho_{a}s} = e^{-s(y(\rho_{a}-\rho_{b})+zy(\rho_{b}-\rho_{c})+z(\rho_{c}-\rho_{b})+\rho_{b})} \delta \rho^{a}{}_{b} \delta \rho^{b}{}_{c} \delta \rho^{c}{}_{a}$$
(A.16)

Beginning with the z integration, we find

$$-s \int_{0}^{1} dz \operatorname{Tr}[\dots] = -se^{-\rho_{b}s} \sum_{\rho_{b}=\rho_{c}} e^{-sy(\rho_{a}-\rho_{b})} \delta\rho^{a}{}_{b} \delta\rho^{b}{}_{c} \delta\rho^{c}{}_{a} + e^{-\rho_{b}s} \sum_{\rho_{b}\neq\rho_{c}} \frac{e^{-s(y(\rho_{a}-\rho_{b})+(1-y)(\rho_{c}-\rho_{b}))} - e^{-sy(\rho_{a}-\rho_{b})}}{(1-y)(\rho_{c}-\rho_{b})} \delta\rho^{a}{}_{b} \delta\rho^{b}{}_{c} \delta\rho^{c}{}_{a} .$$
(A.17)

We see that the factor of (1-y) cancels the one present in the integrand of (A.15) in the last term. Let us now turn to the integration of these sums one at a time

$$\int_{0}^{1} dy (1-y) [\langle \rho_{b} = \rho_{c} \rangle] = \sum_{\rho_{a} = \rho_{b} = \rho_{c}} \frac{1}{2} s e^{-\rho_{b} s} \delta \rho^{a}{}_{b} \delta \rho^{b}{}_{c} \delta \rho^{c}{}_{a} + e^{-\rho_{b} s} \sum_{\rho_{a} \neq \rho_{b} = \rho_{c}} \frac{e^{-(\rho_{a} - \rho_{b})s} - 1}{\rho_{a} - \rho_{b}} \delta \rho^{a}{}_{b} \delta \rho^{b}{}_{c} \delta \rho^{c}{}_{a} - e^{-\rho_{b} s} \sum_{\rho_{a} \neq \rho_{b} = \rho_{c}} \left(\left[\frac{y e^{-(\rho_{a} - \rho_{b})ys}}{(\rho_{a} - \rho_{b})} \right]_{0}^{1} + \left[\frac{e^{-(\rho_{a} - \rho_{b})ys}}{s(\rho_{a} - \rho_{b})^{2}} \right]_{0}^{1} \right) \delta \rho^{a}{}_{b} \delta \rho^{b}{}_{c} \delta \rho^{c}{}_{a} .$$
(A.18)

The first term in the last row of equation (A.18) evaluates to 0 for x = 0 and cancels the $e^{(\rho_a - \rho_b)}$ term of the preceding row for x = 1. Explicitly, we end up with (remembering that we have yet to include the factor 2)

$$\operatorname{Tr}[\dot{\rho}(\ln \rho)] \supset \int_{0}^{\infty} \mathrm{d}s \sum_{\rho_{a}=\rho_{b}=\rho_{c}} se^{-\rho_{b}s} \delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a} -2\sum_{\rho_{a}\neq\rho_{b}=\rho_{c}} \left(\frac{1}{s} \frac{e^{-\rho_{a}s}-e^{-\rho_{b}s}}{(\rho_{a}-\rho_{b})^{2}} -\frac{e^{-\rho_{b}s}}{\rho_{a}-\rho_{b}}\right) \delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a} .$$
(A.19)

The first line cancels to a similar term coming from the third row of equation (A.2), so we ignore it. The second row is evaluated by applying equation (11.15) plus elementary integration and we have

$$\operatorname{Tr}[\dot{\rho}(\ln \rho)] \supset -2 \sum_{\rho_a \neq \rho_b = \rho_c} \left(\frac{\ln \rho_b - \ln \rho_a}{(\rho_a - \rho_b)^2} - \frac{\rho_b^{-1}}{(\rho_a - \rho_b)} \right) \delta \rho^a{}_b \delta \rho^b{}_c \delta \rho^c{}_a .$$
(A.20)

Having arrived at an algebraic expression, let us now consider the y integration in the case when $\rho_b \neq \rho_c$:

$$\frac{1}{2} \int_{0}^{1} dy (1-y) [\langle \rho_{b} \neq \rho_{c} \rangle] = \int_{0}^{1} dy e^{-\rho_{b}s} \sum_{\rho_{b} \neq \rho_{c}} \frac{e^{-s(y(\rho_{a}-\rho_{b})+(1-y)(\rho_{c}-\rho_{b}))} - e^{-sy(\rho_{a}-\rho_{b})}}{(\rho_{c}-\rho_{b})} \delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a}$$

$$= \int_{0}^{1} dy \sum_{\rho_{b} \neq \rho_{c}} \frac{e^{-s(y(\rho_{a}-\rho_{c})+\rho_{c})} - e^{-s(y(\rho_{a}-\rho_{b})+\rho_{b})}}{(\rho_{c}-\rho_{b})} \delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a}$$

$$= \sum_{\rho_{b} \neq \rho_{a}=\rho_{c}} \left(\frac{e^{-s\rho_{c}}}{\rho_{c}-\rho_{b}} + \frac{e^{-s\rho_{a}} - e^{-s\rho_{b}}}{s(\rho_{a}-\rho_{b})(\rho_{c}-\rho_{b})} \right) \delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a}$$

$$- \sum_{\rho_{a}=\rho_{b} \neq \rho_{c}} \left(\frac{e^{-s\rho_{a}} - e^{-s\rho_{c}}}{s(\rho_{a}-\rho_{c})(\rho_{c}-\rho_{b})} + \frac{e^{-s\rho_{a}} - e^{-s\rho_{c}}}{\rho_{c}-\rho_{b}} \right) \delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a}$$

$$+ \sum_{\rho_{a} \neq \rho_{b} \neq \rho_{c}} \left(\frac{e^{-s\rho_{a}} - e^{-s\rho_{b}}}{s(\rho_{a}-\rho_{b})(\rho_{c}-\rho_{b})} - \frac{e^{-s\rho_{a}} - e^{-s\rho_{c}}}{s(\rho_{a}-\rho_{c})(\rho_{c}-\rho_{b})} \right) \delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a}$$

$$(A.21)$$

We here have three distinct cases, with slightly different s integration. Note that we have no terms that (naively) diverge faster than 1/s, so direct application of (11.15) will give all the algebraic answers. We have after the s integration

$$\operatorname{Tr}[\dot{\rho}(\ln \rho)] \supset 2 \sum_{\rho_b \neq \rho_a = \rho_c} \left(\frac{\rho_c^{-1}}{\rho_c - \rho_b} + \frac{\ln \rho_b - \ln \rho_a}{(\rho_a - \rho_b)(\rho_c - \rho_b)} \right) \delta \rho^a{}_b \delta \rho^b{}_c \delta \rho^c{}_a - 2 \sum_{\rho_a = \rho_b \neq \rho_c} \left(\frac{\ln \rho_c - \ln \rho_a}{(\rho_a - \rho_c)(\rho_c - \rho_b)} + \frac{\rho_b^{-1}}{\rho_c - \rho_b} \right) \delta \rho^a{}_b \delta \rho^b{}_c \delta \rho^c{}_a + 2 \sum_{\rho_a \neq \rho_b \neq \rho_c} \left(\frac{\ln \rho_b - \ln \rho_a}{(\rho_a - \rho_b)(\rho_c - \rho_b)} - \frac{\ln \rho_c - \ln \rho_a}{(\rho_a - \rho_c)(\rho_c - \rho_b)} \right) \delta \rho^a{}_b \delta \rho^b{}_c \delta \rho^c{}_a .$$
(A.22)

This, together with equation (A.20) constitutes the nonzero terms of the first row of equation (A.2).

Evaluating the third row of $\mathbf{Tr}[\dot{\rho}(\ddot{\ln \rho})]$ (equation (A.2))

We consider

$$\operatorname{Tr}[\dot{\rho}(\ln \rho)] \supset -2\int_0^\infty \mathrm{d}s \int_0^1 \mathrm{d}y \, ys \int_0^1 \mathrm{d}z \operatorname{Tr}[\delta \rho e^{-(1-y)\rho_0 s} \delta \rho e^{-(1-z)y\rho_0 s} \delta \rho e^{-zy\rho_0 s}] \,. \tag{A.23}$$

We pass to eigenvalues and tensor notation for the trace

$$\operatorname{Tr}[\ldots] = \delta \rho^{a}{}_{b} e^{-(1-y)\rho_{b}s} \delta \rho^{b}{}_{c} e^{-(1-z)y\rho_{c}s} \delta \rho^{c}{}_{a} e^{-zy\rho_{a}s}$$

$$= e^{-s(y(\rho_{c}-\rho_{b})+yz(\rho_{a}-\rho_{c})+\rho_{b})} \delta \rho^{a}{}_{b} \delta \rho^{b}{}_{c} \delta \rho^{c}{}_{a} .$$
(A.24)

We then perform the z integration, noting that only $\rho_a = \rho_c$ needs to be treated as a special case. In the following, we have suppressed the $\delta \rho^a{}_b \delta \rho^b{}_c \delta \rho^c{}_a$ that should be present in every sum.

$$-ys \int_{0}^{1} dz \operatorname{Tr}[\dots] = -yse^{-\rho_{b}s} \left[\sum_{\rho_{a}=\rho_{c}} e^{-sy(\rho_{c}-\rho_{b})} + \sum_{\rho_{a}\neq\rho_{c}} e^{-sy(\rho_{c}-\rho_{b})} \frac{e^{-sy(\rho_{a}-\rho_{c})}-1}{-sy(\rho_{a}-\rho_{c})} \right]$$

$$= e^{-\rho_{b}s} \left[\sum_{\rho_{a}=\rho_{c}} -yse^{-y(\rho_{c}-\rho_{b})s} + \sum_{\rho_{a}\neq\rho_{c}} e^{-sy(\rho_{c}-\rho_{b})} \frac{e^{-sy(\rho_{a}-\rho_{c})}-1}{(\rho_{a}-\rho_{c})} \right].$$
(A.25)

Let us begin with the y integration of the $\rho_a=\rho_c$ term

$$2\int_{0}^{1} \mathrm{d}y[\rho_{a}=\rho_{c}] = \left[\sum_{\rho_{a}=\rho_{b}=\rho_{c}} -se^{-\rho_{b}s}\delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a} + 2\sum_{\rho_{a}=\rho_{c}\neq\rho_{b}}e^{-\rho_{b}s}\left(\left[\frac{ye^{-y(\rho_{c}-\rho_{b})s}}{\rho_{c}-\rho_{b}}\right]_{0}^{1} + \left[\frac{e^{-y(\rho_{c}-\rho_{b})s}}{s(\rho_{c}-\rho_{b})^{2}}\right]_{0}^{1}\right)\delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a}\right].$$
(A.26)

Notably the $\rho_a = \rho_b = \rho_c$ part of the integrand exactly cancels the corresponding term in equation (A.19). We are left to integrate

$$\operatorname{Tr}[\dot{\rho}(\ln \rho)] \supset \int_{0}^{\infty} \mathrm{d}s \, 2 \sum_{\rho_{a}=\rho_{c}\neq\rho_{b}} \left(\frac{e^{-\rho_{c}s}}{\rho_{c}-\rho_{b}} + \frac{e^{-\rho_{c}s}-e^{-\rho_{b}s}}{s(\rho_{c}-\rho_{b})^{2}} \right) \delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a}$$

$$= 2 \sum_{\rho_{a}=\rho_{c}\neq\rho_{b}} \left(\frac{\rho_{c}^{-1}}{\rho_{c}-\rho_{b}} - \frac{\ln\rho_{c}-\ln\rho_{b}}{(\rho_{c}-\rho_{b})^{2}} \right) \delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a}$$
(A.27)

Now we turn to the $\rho_a \neq \rho_c$ term

$$2\int_{0}^{1} dy [\rho_{a} \neq \rho_{c}] = 2\int_{0}^{1} dy e^{-\rho_{b}s} \sum_{\rho_{a} \neq \rho_{c}} \frac{e^{-sy(\rho_{a}-\rho_{b})} - e^{-sy(\rho_{c}-\rho_{b})}}{(\rho_{a}-\rho_{c})} \delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a}$$

$$= 2\sum_{\rho_{a}=\rho_{b} \neq \rho_{c}} \left(\frac{e^{-\rho_{b}s}}{\rho_{a}-\rho_{c}} + \frac{e^{-\rho_{c}s} - e^{-\rho_{b}s}}{s(\rho_{a}-\rho_{c})(\rho_{c}-\rho_{b})}\right) \delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a}$$

$$= 2\sum_{\rho_{a} \neq \rho_{b}=\rho_{c}} \left(\frac{e^{-\rho_{a}s} - e^{-\rho_{b}s}}{(\rho_{a}-\rho_{b})(\rho_{a}-\rho_{c})} + \frac{e^{-\rho_{b}s}}{\rho_{a}-\rho_{c}}\right) \delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a}$$

$$+ 2\sum_{\rho_{a} \neq \rho_{b} \neq \rho_{c}} \left(\frac{e^{-\rho_{c}s} - e^{-\rho_{b}s}}{s(\rho_{a}-\rho_{c})(\rho_{c}-\rho_{b})} - \frac{e^{-\rho_{a}s} - e^{-\rho_{b}s}}{(\rho_{a}-\rho_{b})(\rho_{a}-\rho_{c})}\right) \delta\rho^{a}{}_{b}\delta\rho^{b}{}_{c}\delta\rho^{c}{}_{a} .$$
(A.28)

By applying the usual identity from equation (11.15) evaluating the s integral is straightforward and one finds

$$\operatorname{Tr}[\dot{\rho}(\ddot{\ln\rho})] \supset 2 \sum_{\rho_a = \rho_b \neq \rho_c} \left(\frac{\rho_b^{-1}}{\rho_a - \rho_c} + \frac{\ln\rho_b - \ln\rho_c}{(\rho_a - \rho_c)(\rho_c - \rho_b)} \right) \delta\rho^a{}_b \delta\rho^b{}_c \delta\rho^c{}_a - 2 \sum_{\rho_a \neq \rho_b = \rho_c} \left(\frac{\ln\rho_b - \ln\rho_a}{(\rho_a - \rho_b)(\rho_a - \rho_c)} + \frac{\rho_b^{-1}}{\rho_a - \rho_c} \right) \delta\rho^a{}_b \delta\rho^b{}_c \delta\rho^c{}_a + 2 \sum_{\rho_a \neq \rho_b \neq \rho_c} \left(\frac{\ln\rho_b - \ln\rho_c}{(\rho_a - \rho_c)(\rho_c - \rho_b)} - \frac{\ln\rho_b - \ln\rho_a}{(\rho_a - \rho_b)(\rho_a - \rho_c)} \right) \delta\rho^a{}_b \delta\rho^b{}_c \delta\rho^c{}_a .$$
(A.29)

Collecting terms and going to correlation functions

Let us write out the first and third rows of $\text{Tr}[\dot{\rho}(\ln \rho)]$ to zeroth order in epsilon in one place. Adding together equations (A.20), (A.22), (A.27) and (A.29) we find no cancellations, and the end result is

$$\begin{aligned} \operatorname{Tr}[\dot{\rho}(\ln\rho)] \supset &= -4 \sum_{\rho_a = \rho_b \neq \rho_c} \left(\frac{\ln \rho_a - \ln \rho_c}{(\rho_a - \rho_c)^2} - \frac{\rho_a^{-1}}{(\rho_a - \rho_c)} \right) \delta\rho^a{}_b \delta\rho^b{}_c \delta\rho^c{}_a \\ &- 4 \sum_{\rho_a \neq \rho_b = \rho_c} \left(\frac{\ln \rho_b - \ln \rho_a}{(\rho_b - \rho_a)^2} - \frac{\rho_b^{-1}}{(\rho_b - \rho_a)} \right) \delta\rho^a{}_b \delta\rho^b{}_c \delta\rho^c{}_a \\ &- 4 \sum_{\rho_a = \rho_c \neq \rho_b} \left(\frac{\ln \rho_c - \ln \rho_b}{(\rho_c - \rho_b)^2} - \frac{\rho_c^{-1}}{(\rho_c - \rho_b)} \right) \delta\rho^a{}_b \delta\rho^b{}_c \delta\rho^c{}_a \\ &+ 2 \sum_{\rho_a \neq \rho_b \neq \rho_c} \left(\frac{\ln \rho_a + \ln \rho_b - 2 \ln \rho_c}{(\rho_a - \rho_c)(\rho_c - \rho_b)} + \frac{\ln \rho_b - \ln \rho_a}{\rho_a - \rho_b} \left(\frac{1}{\rho_c - \rho_b} + \frac{1}{\rho_c - \rho_a} \right) \right) \delta\rho^a{}_b \delta\rho^b{}_c \delta\rho^c{}_a , \end{aligned}$$
(A.30)

where we have used the equalities in the sums to make apparent a cyclic symmetry in the first three terms. More specifically, if you let $a \rightarrow b, b \rightarrow c$ and $c = \rightarrow a$ the first row in (A.30) becomes the second, the second becomes the third and the third becomes the first.

To bring this into the form of integrals of correlation functions one would like to make use of

$$\frac{1}{2} \int \mathrm{d}s \frac{e^{\frac{isx}{2\pi}}}{1 + \cosh(s)} = \frac{x}{e^{x/2} - e^{-x/2}} = \left\langle \ln\left(\frac{\rho_i}{\rho_j}\right) \right\rangle = \sqrt{\rho_i \rho_j} \frac{\ln \rho_i - \ln \rho_j}{\rho_i - \rho_j} \tag{A.31}$$

together with an identity of the form

$$\frac{1}{\rho_i - \rho_j} = \sum_a f_a(\rho_i) g_b(\rho_j) \tag{A.32}$$

to return the expression to the form of a trace of linear operators. I am unable to find an identity of the desired type, and I suspect that a better approach is that of [107] in which the (hidden) symmetries of $(\ln \rho)$ are used to find a result on the desired form.

Bibliography

- Jacob D. Bekenstein. "Black Holes and Entropy". In: *Phys. Rev. D* 7 (8 Apr. 1973), pp. 2333–2346. DOI: 10.1103/PhysRevD.7.2333. URL: https://link.aps.org/doi/10.1103/PhysRevD.7.2333.
- [2] Raphael Bousso. "A Covariant entropy conjecture". In: JHEP 07 (1999), p. 004. DOI: 10. 1088/1126-6708/1999/07/004. arXiv: hep-th/9905177 [hep-th].
- Juan Maldacena. "The Large N Limit of Superconformal field theories and supergravity". In: (1998). eprint: 9711200v3 (arXiv:hep-th). URL: https://arxiv.org/pdf/hep-th/ 9711200.pdf.
- [4] Shinsei Ryu and Tadashi Takayanagi. "Holographic Derivation of Entanglement Entropy from the anti-de Sitter Space/Conformal Field Theory Correspondence". In: *Phys. Rev. Lett.* 96 (18 May 2006), p. 181602. DOI: 10.1103/PhysRevLett.96.181602. URL: https: //link.aps.org/doi/10.1103/PhysRevLett.96.181602.
- [5] Veronika E. Hubeny, Mukund Rangamani, and Tadashi Takayanagi. "A Covariant Holographic Entanglement Entropy Proposal". In: (May 2007). DOI: 10.1088/1126-6708/ 2007/07/062. arXiv: 0705.0016. URL: http://arxiv.org/abs/0705.0016%20http: //dx.doi.org/10.1088/1126-6708/2007/07/062.
- [6] Erik Verlinde. "Emergent Gravity and the Dark Universe". In: (2016). arXiv: arXiv:1611.
 02269v2. URL: https://arxiv.org/pdf/1611.02269.pdf.
- Juan Maldacena and Leonard Susskind. "Cool horizons for entangled black holes". In: Fortsch. Phys. 61 (2013), pp. 781-811. DOI: 10.1002/prop.201300020. arXiv: 1306.0533 [hep-th].
- [8] ChunJun Cao and Sean M. Carroll. "Bulk entanglement gravity without a boundary: Towards finding Einstein's equation in Hilbert space". In: *Phys. Rev. D* 97 (8 Apr. 2018), p. 086003. DOI: 10.1103/PhysRevD.97.086003. URL: https://link.aps.org/doi/10.1103/PhysRevD.97.086003.
- [9] Steven B. Giddings. "Quantum-first gravity". In: (Mar. 2018). arXiv: 1803.04973. URL: http://arxiv.org/abs/1803.04973.
- [10] Mark Van Raamsdonk. "Building up spacetime with quantum entanglement II: It from BCbit". In: (2018). arXiv: 1809.01197 [hep-th].
- [11] Maximilian Schlosshauer. "Decoherence, the measurement problem, and interpretations of quantum mechanics". In: (2005). arXiv: 0312059v4 [arXiv:quant-ph]. URL: https:// arxiv.org/pdf/quant-ph/0312059.pdf.
- [12] Phuc Nguyen et al. "Entanglement of purification: from spin chains to holography". In: (). URL: https://arxiv.org/pdf/1709.07424.pdf.

- [13] Mark M. Wilde. "From Classical to Quantum Shannon Theory". In: (June 2011). DOI: 10. 1017/9781316809976.001. arXiv: 1106.1445. URL: http://arxiv.org/abs/1106.1445% 20http://dx.doi.org/10.1017/9781316809976.001.
- [14] Fotini Markopoulou. "Towards Gravity from the Quantum". In: (2006). arXiv: 0604120 [arXiv:hep-th]. URL: http://cds.cern.ch/record/942758/files/0604120.pdf.
- [15] Mark Van Raamsdonk. "Lectures on Gravity and Entanglement". In: (2016). arXiv: arXiv: 1609.00026v1. URL: https://arxiv.org/pdf/1609.00026.pdf.
- [16] Dominik Neuenfeld, Krishan Saraswat, and Mark Van Raamsdonk. Positive gravitational subsystem energies from CFT cone relative entropies. 2018. arXiv: arXiv:1802.01585v2. URL: https://arxiv.org/pdf/1802.01585.pdf.
- [17] Elliott H Lieb and Mary Beth Ruskai. Recherche Coopérative sur Programme n o 25. URL: http://www.numdam.org/item?id=RCP25%7B%5C_%7D1973%7B%5C_%7D%7B%5C_%7D19%7B% 5C_%7D%7B%5C_%7DA5%7B%5C_%7D0.
- [18] Alexander Müller-Hermes and David Reeb. "Monotonicity of the Quantum Relative Entropy Under Positive Maps". In: (Dec. 2015). DOI: 10.1007/s00023-017-0550-9. arXiv: 1512. 06117. URL: http://arxiv.org/abs/1512.06117%20http://dx.doi.org/10.1007/ s00023-017-0550-9.
- [19] Edward Witten. "A Mini-Introduction To Information Theory". In: (May 2018). arXiv: 1805.11965. URL: http://arxiv.org/abs/1805.11965.
- [20] Michael Edward Peskin and Daniel V. Schroeder. An Introduction To Quantum Field Theory. Westview Press, 1995, p. 864. ISBN: 0201503972. URL: https://books.google.se/books/ about/An%7B%5C_%7DIntroduction%7B%5C_%7Dto%7B%5C_%7DQuantum%7B%5C_%7DField% 7B%5C_%7DTheory.html?id=i35LALN0GosC%7B%5C&%7Dredir%7B%5C_%7Desc=y.
- [21] A. Zee. Quantum field theory in a nutshell. Princeton University Press, 2010, p. 576. ISBN: 9780691140346. URL: https://press.princeton.edu/titles/9227.html.
- [22] Tom Hartman. Lecture notes on path integrals, states, and operators in QFT. URL: http: //www.hartmanhep.net/topics2015/4-pathintegrals.pdf.
- [23] Mario Flory, Robert C. Helling, and Constantin Sluka. "How I Learned to Stop Worrying and Love QFT". In: (Jan. 2012). arXiv: 1201.2714. URL: http://arxiv.org/abs/1201.2714.
- [24] Martin Ammon and Johanna Erdmenger. Gauge/Gravity Duality. Cambridge: Cambridge University Press, 2015. ISBN: 9780511846373. DOI: 10.1017/CB09780511846373. URL: http: //ebooks.cambridge.org/ref/id/CB09780511846373.
- [25] David J. Gross and Frank Wilczek. "Ultraviolet Behavior of Non-Abelian Gauge Theories". In: *Physical Review Letters* 30.26 (June 1973), pp. 1343-1346. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.30.1343. URL: https://link.aps.org/doi/10.1103/PhysRevLett.30.1343.
- [26] A. Gorsky, A. Mironov, A. Morozov, and T. N. Tomaras. "Is the Standard Model saved asymptotically by conformal symmetry?" In: (Sept. 2014). DOI: 10.1134/S1063776115030218. arXiv: 1409.0492. URL: http://arxiv.org/abs/1409.0492%20http://dx.doi.org/10. 1134/S1063776115030218.
- [27] Joshua D. Qualls. "Lectures on Conformal Field Theory". In: (2015). arXiv: 1511.04074 [hep-th].

- [28] S. W. Hawking. "Particle creation by black holes". In: Communications in Mathematical Physics 43.3 (Aug. 1975), pp. 199–220. ISSN: 1432-0916. DOI: 10.1007/BF02345020. URL: https://doi.org/10.1007/BF02345020.
- [29] J Eisert, M Cramer, and M B Plenio. Area laws for the entanglement entropy-a review. arXiv: arXiv:0808.3773v4. URL: https://arxiv.org/pdf/0808.3773.pdf.
- [30] R. F. (Raymond Frederick) Streater and A. S. (Arthur Strong) Wightman. PCT, spin and statistics, and all that, p. 207. ISBN: 9780691070629. URL: https://press.princeton.edu/ titles/6938.html.
- [31] R. Haag and B. Schroer. "Postulates of Quantum Field Theory". In: Journal of Mathematical Physics 3.2 (Mar. 1962), pp. 248–256. ISSN: 0022-2488. DOI: 10.1063/1.1703797. URL: http://aip.scitation.org/doi/10.1063/1.1703797.
- [32] Edward Witten. Notes On Some Entanglement Properties Of Quantum Field Theory. 2018. arXiv: arXiv:1803.04993v6. URL: https://arxiv.org/pdf/1803.04993.pdf.
- [33] Ko Sanders. "On the Reeh-Schlieder Property in Curved Spacetime". In: (Jan. 2008). DOI: 10.1007/s00220-009-0734-3. arXiv: 0801.4676. URL: http://arxiv.org/abs/0801.4676%20http://dx.doi.org/10.1007/s00220-009-0734-3.
- [34] Brian Swingle and Mark Van Raamsdonk. "Universality of Gravity from Entanglement". In: (2014). arXiv: 1405.2933 [hep-th].
- [35] Detlev Buchholz and Hanns Schulz-mirbach. "Haag Duality in Conformal Quantum Field Theory". In: 2 (Jan. 1990), pp. 105–125.
- [36] H Araki et al. Recherche Coopérative sur Programme n o 25 HUZIHIRO ARAKI Inequalities in Von Neumann Algebras Les rencontres physiciens-mathématiciens de Strasbourg-RCP25, 1975, tome 22 « Exposés de.
- [37] Joseph J. Bisognano and Eyvind H. Wichmann. "On the duality condition for quantum fields". In: *Journal of Mathematical Physics* 17.3 (Aug. 1976), p. 303. ISSN: 00222488. DOI: 10.1063/1.522898. URL: http://scitation.aip.org/content/aip/journal/jmp/17/3/ 10.1063/1.522898.
- [38] Mukund Rangamani and Tadashi Takayanagi. "Holographic Entanglement Entropy". In: Lect. Notes Phys. 931 (2017), pp.1–246. DOI: 10.1007/978-3-319-52573-0. arXiv: 1609.
 01287 [hep-th].
- [39] Horacio Casini, Marina Huerta, and Robert C. Myers. "Towards a derivation of holographic entanglement entropy". In: JHEP 05 (2011), p. 036. DOI: 10.1007/JHEP05(2011)036. arXiv: 1102.0440 [hep-th].
- [40] H. Casini and Marina Huerta. "On the RG running of the entanglement entropy of a circle". In: *Phys. Rev.* D85 (2012), p. 125016. DOI: 10.1103/PhysRevD.85.125016. arXiv: 1202.5650 [hep-th].
- [41] Pasquale Calabrese and John Cardy. "Entanglement entropy and conformal field theory". In: J. Phys. A42 (2009), p. 504005. DOI: 10.1088/1751-8113/42/50/504005. arXiv: 0905.4013 [cond-mat.stat-mech].
- [42] Mukund Rangamani and Tadashi Takayanagi. "Holographic Entanglement Entropy". In: Lect. Notes Phys. 931 (2017), pp.1–246. DOI: 10.1007/978-3-319-52573-0. arXiv: 1609. 01287 [hep-th].

- [43] Barton Zwiebach. A First Course in String Theory, Second Edition. Cambridge: Cambridge University Press, 2009. ISBN: 9780521880329.
- [44] David Tong. "String Theory". In: (2009). arXiv: 0908.0333 [hep-th].
- [45] Sean M. Carroll. "Lecture notes on general relativity". In: (1997). arXiv: gr-qc/9712019 [gr-qc].
- [46] Mans Henningson and Konstadinos Sfetsos. "Spinors and the AdS / CFT correspondence". In: *Phys. Lett.* B431 (1998), pp. 63–68. DOI: 10.1016/S0370-2693(98)00559-0. arXiv: hep-th/9803251 [hep-th].
- [47] Edward Witten. "Anti-de Sitter space and holography". In: Adv. Theor. Math. Phys. 2 (1998), pp. 253-291. DOI: 10.4310/ATMP.1998.v2.n2.a2. arXiv: hep-th/9802150 [hep-th].
- [48] Daniel Z. Freedman and Antoine Van Proeyen. Supergravity. Cambridge, UK: Cambridge Univ. Press, 2012. ISBN: 9781139368063, 9780521194013.
- [49] M. Henningson and K. Skenderis. "The Holographic Weyl anomaly". In: JHEP 07 (1998), p. 023. DOI: 10.1088/1126-6708/1998/07/023. arXiv: hep-th/9806087 [hep-th].
- [50] Igor R. Klebanov and Edward Witten. "AdS / CFT correspondence and symmetry breaking". In: *Nucl. Phys.* B556 (1999), pp. 89–114. DOI: 10.1016/S0550-3213(99)00387-9. arXiv: hep-th/9905104 [hep-th].
- [51] Peter Breitenlohner and Daniel Z. Freedman. "Stability in Gauged Extended Supergravity". In: Annals Phys. 144 (1982), p. 249. DOI: 10.1016/0003-4916(82)90116-6.
- [52] C.Robin Graham and John M Lee. "Einstein metrics with prescribed conformal infinity on the ball". In: Advances in Mathematics 87.2 (1991), pp. 186-225. ISSN: 0001-8708. DOI: https://doi.org/10.1016/0001-8708(91)90071-E. URL: http://www.sciencedirect. com/science/article/pii/000187089190071E.
- [53] Sebastian de Haro, Sergey N. Solodukhin, and Kostas Skenderis. "Holographic reconstruction of space-time and renormalization in the AdS / CFT correspondence". In: Commun. Math. Phys. 217 (2001), pp. 595–622. DOI: 10.1007/s002200100381. arXiv: hep-th/0002230 [hep-th].
- [54] Pietro Antonio Grassi and Pierre Vanhove. "Higher-loop amplitudes in the non-minimal pure spinor formalism". In: JHEP 05 (2009), p. 089. DOI: 10.1088/1126-6708/2009/05/089. arXiv: 0903.3903 [hep-th].
- [55] Lars Brink, Olof Lindgren, and Bengt E. W. Nilsson. "The Ultraviolet Finiteness of the N=4 Yang-Mills Theory". In: *Phys. Lett.* 123B (1983), pp. 323–328. DOI: 10.1016/0370– 2693(83)91210–8.
- [56] David Simmons-Duffin. "The Conformal Bootstrap". In: Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings (TASI 2015): Boulder, CO, USA, June 1-26, 2015. 2017, pp. 1–74. DOI: 10.1142/9789813149441_ 0001. arXiv: 1602.07982 [hep-th].
- [57] J. D. Brown and Marc Henneaux. "Central charges in the canonical realization of asymptotic symmetries: an example from three-dimensional gravity". In: *Comm. Math. Phys.* 104.2 (1986), pp. 207–226. URL: https://projecteuclid.org:443/euclid.cmp/1104114999.
- [58] Leonard Susskind and Edward Witten. "The Holographic bound in anti-de Sitter space". In: (1998). arXiv: hep-th/9805114 [hep-th].

- [59] Ingemar Bengtsson. "ANTI-DE SITTER SPACE". In: (1998). URL: http://3dhouse.se/ ingemar/Kurs.pdf.
- [60] Aitor Lewkowycz and Juan Maldacena. "Generalized gravitational entropy". In: JHEP 08 (2013), p. 090. DOI: 10.1007/JHEP08(2013)090. arXiv: 1304.4926 [hep-th].
- [61] Aitor Lewkowycz and Onkar Parrikar. "The holographic shape of entanglement and Einstein's equations". In: JHEP 05 (2018), p. 147. DOI: 10.1007/JHEP05(2018)147. arXiv: 1802.10103 [hep-th].
- [62] Veronika E. Hubeny. "Extremal surfaces as bulk probes in AdS/CFT". In: JHEP 07 (2012),
 p. 093. DOI: 10.1007/JHEP07(2012)093. arXiv: 1203.1044 [hep-th].
- [63] Thomas Faulkner et al. "Nonlinear Gravity from Entanglement in Conformal Field Theories". In: JHEP 08 (2017), p. 057. DOI: 10.1007/JHEP08(2017)057. arXiv: 1705.03026 [hep-th].
- [64] Aron C. Wall. "Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy". In: *Class. Quant. Grav.* 31.22 (2014), p. 225007. DOI: 10.1088/0264-9381/31/22/225007. arXiv: 1211.3494 [hep-th].
- [65] Xi Dong, Aitor Lewkowycz, and Mukund Rangamani. "Deriving covariant holographic entanglement". In: JHEP 11 (2016), p. 028. DOI: 10.1007/JHEP11(2016)028. arXiv: 1607. 07506 [hep-th].
- [66] Bartlomiej Czech, Joanna L. Karczmarek, Fernando Nogueira, and Mark Van Raamsdonk. "Rindler Quantum Gravity". In: *Class. Quant. Grav.* 29 (2012), p. 235025. DOI: 10.1088/ 0264-9381/29/23/235025. arXiv: 1206.1323 [hep-th].
- [67] Juan Martin Maldacena. "Eternal black holes in anti-de Sitter". In: JHEP 04 (2003), p. 021.
 DOI: 10.1088/1126-6708/2003/04/021. arXiv: hep-th/0106112 [hep-th].
- [68] ChunJun Cao. "Towards a Theory of Quantum Gravity Through Geometrization of Quantum Mechanics". PhD thesis. Caltech, 2018-05-21. URL: https://thesis.library.caltech. edu/10860/.
- [69] David D. Blanco, Horacio Casini, Ling-Yan Hung, and Robert C. Myers. "Relative Entropy and Holography". In: JHEP 08 (2013), p. 060. DOI: 10.1007/JHEP08(2013)060. arXiv: 1305.3182 [hep-th].
- [70] Nima Lashkari, Michael B. McDermott, and Mark Van Raamsdonk. "Gravitational dynamics from entanglement 'thermodynamics'". In: JHEP 04 (2014), p. 195. DOI: 10.1007/ JHEP04(2014)195. arXiv: 1308.3716 [hep-th].
- [71] Nima Lashkari and Mark Van Raamsdonk. "Canonical Energy is Quantum Fisher Information". In: JHEP 04 (2016), p. 153. DOI: 10.1007/JHEP04(2016)153. arXiv: 1508.00897
 [hep-th].
- [72] Stefan Hollands and Robert M. Wald. "Stability of Black Holes and Black Branes". In: *Commun. Math. Phys.* 321 (2013), pp. 629–680. DOI: 10.1007/s00220-012-1638-1. arXiv: 1201.0463 [gr-qc].
- [73] J. Lee and Robert M. Wald. "Local symmetries and constraints". In: J. Math. Phys. 31 (1990), pp. 725–743. DOI: 10.1063/1.528801.
- [74] Antony J. Speranza. "Entanglement entropy of excited states in conformal perturbation theory and the Einstein equation". In: JHEP 04 (2016), p. 105. DOI: 10.1007/JHEP04(2016) 105. arXiv: 1602.01380 [hep-th].

- [75] Donald Marolf et al. "From Euclidean Sources to Lorentzian Spacetimes in Holographic Conformal Field Theories". In: JHEP 06 (2018), p. 077. DOI: 10.1007/JHEP06(2018)077. arXiv: 1709.10101 [hep-th].
- [76] Cedomir Crnkovic. "Symplectic Geometry of the Covariant Phase Space, Superstrings and Superspace". In: Class. Quant. Grav. 5 (1988), pp. 1557–1575. DOI: 10.1088/0264-9381/ 5/12/008.
- [77] Miguel S. Costa, Vasco Gonçalves, and João Penedones. "Spinning AdS Propagators". In: JHEP 09 (2014), p. 064. DOI: 10.1007/JHEP09(2014)064. arXiv: 1404.5625 [hep-th].
- [78] Felix M. Haehl, Eliot Hijano, Onkar Parrikar, and Charles Rabideau. "Higher Curvature Gravity from Entanglement in Conformal Field Theories". In: *Phys. Rev. Lett.* 120.20 (2018), p. 201602. DOI: 10.1103/PhysRevLett.120.201602. arXiv: 1712.06620 [hep-th].
- [79] Felix M. Haehl et al. "Nonlocal multi-trace sources and bulk entanglement in holographic conformal field theories". In: (2019). arXiv: 1904.01584 [hep-th].
- [80] Leonard Susskind. "Computational Complexity and Black Hole Horizons". In: Fortsch. Phys. 64 (2016). [Fortsch. Phys.64,24(2016)], pp. 44–48. DOI: 10.1002/prop.201500093,10.1002/prop.201500092. arXiv: 1403.5695 [hep-th].
- [81] Sean Cooper et al. "Black Hole Microstate Cosmology". In: (2018). arXiv: 1810.10601 [hep-th].
- [82] Jan De Boer et al. "Probing typical black hole microstates". In: (2019). arXiv: 1901.08527 [hep-th].
- [83] Leonard Susskind. "Entanglement is not enough". In: Fortsch. Phys. 64 (2016), pp. 49–71.
 DOI: 10.1002/prop.201500095. arXiv: 1411.0690 [hep-th].
- [84] David Garfinkle and Andrew Strominger. "Semiclassical Wheeler wormhole production". In: Phys. Lett. B256 (1991), pp. 146–149. DOI: 10.1016/0370-2693(91)90665-D.
- [85] Nima Arkani-Hamed, Lubos Motl, Alberto Nicolis, and Cumrun Vafa. "The String landscape, black holes and gravity as the weakest force". In: JHEP 06 (2007), p. 060. DOI: 10.1088/1126-6708/2007/06/060. arXiv: hep-th/0601001 [hep-th].
- [86] Leonard Susskind and Ying Zhao. "Switchbacks and the Bridge to Nowhere". In: (2014). arXiv: 1408.2823 [hep-th].
- [87] Grant Newton Remmen. "Defining Gravity: Effective Field Theory, Entanglement, and Cosmology". PhD thesis. Caltech, Pasadena (main), 2017-05-30. DOI: 10.7907/Z90R9MD1. URL: http://resolver.caltech.edu/CaltechTHESIS:04072017-131612641.
- [88] Ahmed Almheiri, Donald Marolf, Joseph Polchinski, and James Sully. "Black Holes: Complementarity or Firewalls?" In: JHEP 02 (2013), p. 062. DOI: 10.1007/JHEP02(2013)062. arXiv: 1207.3123 [hep-th].
- [89] Ahmed Almheiri et al. "An Apologia for Firewalls". In: JHEP 09 (2013), p. 018. DOI: 10. 1007/JHEP09(2013)018. arXiv: 1304.6483 [hep-th].
- [90] Leonard Susskind and Larus Thorlacius. "Gedanken experiments involving black holes". In: *Phys. Rev.* D49 (1994), pp. 966-974. DOI: 10.1103/PhysRevD.49.966. arXiv: hep-th/9308100 [hep-th].
- [91] Douglas Stanford and Leonard Susskind. "Complexity and Shock Wave Geometries". In: *Phys. Rev.* D90.12 (2014), p. 126007. DOI: 10.1103/PhysRevD.90.126007. arXiv: 1406.2678 [hep-th].

- [92] Adam R. Brown et al. "Complexity, action, and black holes". In: *Phys. Rev.* D93.8 (2016),
 p. 086006. DOI: 10.1103/PhysRevD.93.086006. arXiv: 1512.04993 [hep-th].
- [93] Mark R Dowling and Michael A Nielsen. *The geometry of quantum computation*. Tech. rep. 2006. URL: www.qinfo.org/people/nielsen.
- [94] Juan Maldacena, Stephen H. Shenker, and Douglas Stanford. "A bound on chaos". In: JHEP 08 (2016), p. 106. DOI: 10.1007/JHEP08(2016)106. arXiv: 1503.01409 [hep-th].
- [95] Adam R. Brown et al. "Complexity of Jackiw-Teitelboim gravity". In: Phys. Rev. D99.4 (2019), p. 046016. DOI: 10.1103/PhysRevD.99.046016. arXiv: 1810.08741 [hep-th].
- [96] Tadashi Takayanagi. "Holographic Dual of BCFT". In: *Phys. Rev. Lett.* 107 (2011), p. 101602.
 DOI: 10.1103/PhysRevLett.107.101602. arXiv: 1105.5165 [hep-th].
- [97] Lisa Randall and Raman Sundrum. "An Alternative to compactification". In: Phys. Rev. Lett. 83 (1999), pp. 4690-4693. DOI: 10.1103/PhysRevLett.83.4690. arXiv: hep-th/ 9906064 [hep-th].
- [98] Sanjeev S. Seahra, Chris Clarkson, and Roy Maartens. "Delocalization of brane gravity by a bulk black hole". In: *Class. Quant. Grav.* 22 (2005), pp. L91–L102. DOI: 10.1088/0264-9381/22/16/L02. arXiv: gr-qc/0504023 [gr-qc].
- [99] Georges Obied, Hirosi Ooguri, Lev Spodyneiko, and Cumrun Vafa. "De Sitter Space and the Swampland". In: (2018). arXiv: 1806.08362 [hep-th].
- [100] Renata Kallosh, Andrei Linde, Evan McDonough, and Marco Scalisi. "dS Vacua and the Swampland". In: (2019). arXiv: 1901.02022 [hep-th].
- [101] ChunJun Cao, Sean M. Carroll, and Spyridon Michalakis. "Space from Hilbert Space: Recovering Geometry from Bulk Entanglement". In: *Phys. Rev.* D95.2 (2017), p. 024031. DOI: 10.1103/PhysRevD.95.024031. arXiv: 1606.08444 [hep-th].
- [102] Jacob C. Bridgeman and Christopher T. Chubb. "Hand-waving and Interpretive Dance: An Introductory Course on Tensor Networks". In: J. Phys. A50.22 (2017), p. 223001. DOI: 10.1088/1751-8121/aa6dc3. arXiv: 1603.03039 [quant-ph].
- [103] G. Vidal. "Entanglement Renormalization". In: *Phys. Rev. Lett.* 99.22 (2007), p. 220405.
 DOI: 10.1103/PhysRevLett.99.220405. arXiv: cond-mat/0512165 [cond-mat].
- Brian Swingle. "Entanglement Renormalization and Holography". In: *Phys. Rev.* D86 (2012), p. 065007. DOI: 10.1103/PhysRevD.86.065007. arXiv: 0905.1317 [cond-mat.str-el].
- [105] Ning Bao, ChunJun Cao, Sean M. Carroll, and Aidan Chatwin-Davies. "De Sitter Space as a Tensor Network: Cosmic No-Hair, Complementarity, and Complexity". In: *Phys. Rev.* D96.12 (2017), p. 123536. DOI: 10.1103/PhysRevD.96.123536. arXiv: 1709.03513 [hep-th].
- [106] Robert M. Wald. "Asymptotic behavior of homogeneous cosmological models in the presence of a positive cosmological constant". In: *Phys. Rev.* D28 (1983), pp. 2118–2120. DOI: 10. 1103/PhysRevD.28.2118.
- [107] Nima Lashkari, Hong Liu, and Srivatsan Rajagopal. "Perturbation Theory for the Logarithm of a Positive Operator". In: (2018). arXiv: 1811.05619 [hep-th].