# Characterizing Information in Physical Systems: from Biology to Black Holes

Lauren McGough

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Physics
Advisers: William Bialek and Herman Verlinde

June 2018

# Abstract

In this thesis, we use classical and quantum information theory to probe fundamental questions about living and nonliving physical systems, including developing embryos, conformal field theories, black holes, and holographic dualities. We begin by analyzing how spatially varying concentrations of four proteins in the early stage fruit fly embryo are able to encode enough information to specify a precise body plan for the developed adult fly.

We then transition to studying how information is tied to physics in nonliving systems, beginning with a conjecture on the structure of universal terms in the Rényi entropy in $3+1$-d CFT. In the following pair of chapters, we use "topological" entanglement in $\mathrm{AdS}_3/\mathrm{CFT}_2$ duality as a springboard for developing a precise correspondence between Liouville theory and the topological sector of $2+1$-d gravity.

Then, in the final chapter, we study the "information flow" among energy scales in a 2-d field theory constructed as a CFT deformed by the irrelevant dimension-4 operator $T\bar{T}$. We do so by constructing an explicit manifestation of holographic RG. By identifying the $T\bar{T}$ coupling with a hard cutoff in the bulk, we are able to exactly match the thermodynamics of a "black hole in a box" with the physics of an integrable field theory.

Throughout, our goal is to demonstrate that studying uncertainty with information theory, entanglement, and renormalization group flow allows us to organize the unknowns and thus obtain new methods for constraining, characterizing, and dualizing the system at hand. In the process, we learn fundamental properties we might not have otherwise known to study.

# Acknowledgements

I'd like to begin by thanking my two advisors, Herman Verlinde and Bill Bialek. Herman, thank you for the many stimulating discussions and for often demonstrating how to think "outside the box", connecting ideas nobody else has considered to current questions in the field.

Bill, thank you for bringing me into the exciting field that is biophysics, and teaching me some of the myriad ways we can think of biology using ideas from information theory and statistical mechanics. Thank you for encouraging me to begin to play with data. Most importantly, thank you for being supportive from the beginning, and for your constant positivity.

Next, I'd like to thank my additional collaborators. Jeongseog Lee and Ben Safdi, I learned so much and had a lot of fun figuring out every technical detail together, in addition to gaining experience doing numerical work early in my career. Mark Mezei, I am grateful for your willingness to answer questions with such clarity, and for your insistence on completing calculations with care and rigor. Thank you also to Steven Jackson, collaborator on our work on Liouville theory and $AdS_3/CFT_2$. And thanks to David Schwab (The Graduate Center, CUNY), Miles Stoudenmire (Simons Foundation), and Caroline Holmes for their earlier discussions on machine learning.

There have been several other faculty I'd like to recognize for their support in different ways. Shivaji Sondhi, thank you so much for your mentoring and for helping me learn how to ask questions until I have fully understood every aspect of a problem or calculation. Thank you for helping me learn to demand more rigor in my thought, and thank you for your invaluable advice on how to proceed at a time when I was unsure.

Josh Shaevitz, thank you for being an awesome experimental project advisor, making sure I got hands-on experience as well as ownership of a piece of a project, while keeping it fun. The experience in Matlab has also been indispensible going from high energy theory to biophysics, and knowing how to use epoxy will surely be useful at some point going forward. Thank you also for ensuring I had one of the best possible AI's with you, and thank you for your advice in my job search.

Ned Wingreen, thank you for helping me learn to give a talk both through instruction and through example via your amazingly clear and engaging presentations. Thank you also for your practical advice regarding interviews. Thank you also to biophysics professor Thomas Gregor for an informative discussion about previous work and experimental results, and for the fly embryo data on which my work is predicated.

There are several postdocs I'd like to recognize for enlightening discussions and mentoring. Then-PCTS postdoc Daniel Harlow, thank you for your engaging discussions from early in my first year

through your time as a postdoc. You are one of few people for whom no question is too elementary, and you strive for clarity in ways others do not. Moreover, thank you for being a mentor and for your ability to be among the best speakers at journal club when others would not. Thank you to postdoc Eric Perlmutter for many clear, interesting discussions. Thank you also to then-Lewis Sigler fellow Ben Machta and PCTS postdoc Pierre Ronceray for enlightening discussions.

I'd like to thank the people who patiently listened to me talk about my work and provided valuable feedback and discussion on numerous occasions. These include Yuval Elhanati, Amir Erez, Andreas Mayer, Farzan Beroz, Xiaowen Chen, Junyi Zhang, Bin Xu, Leenoy Meshulam, Benjamin Weiner, and the many attendees of the hep-th journal club and the condensed matter graduate student seminar, of which I was, oddly, at one point the most frequent presenter.

Special thanks to Ilya Belopolski, who not only heard both my biophysics and high energy work on numerous occasions, but whose enthusiasm and confidence is contagious. Thank you, Ilya, for asking questions even when they seem elementary, and thank you for being someone who is always up for a viewing party of important Kitaev lectures, or other celebrations of physics.

Similarly, thank you to Aris Alexandradinata, another person who is always interested in my work, whether it be about black holes or biophysics. Thank you for sharing your own exciting work, and thank you, both for always trying to understand, and for always helping me understand. Thank you for your unique enthusiasm and curiosity. And thank you for the space heater, which has gotten me through many days of poor climate control at Jadwin, and which is still going strong six years later.

Thank you to my office mates, especially: to Aaron Levy for clear, interesting discussions and friendship, and Sarthak Parikh, for your clarity and enthusiasm, and to Aitor Lewkowycz, for your inspiring work ethic and balance, and for co-organizing the hep-th journal club, especially through the moments of rushing to find a speaker.

To many other graduate students who have made my time here enjoyable and enriching, including close friends who followed from MIT, Lin Fei (and family HaoQi Li, Helen Fei, and Neal Fei) and Shawn Westerdale as well as former roommates Guangyong Koh and Katie Spaulding for putting up with my crazy schedule; Nikolay Dedushenko for, among other things, teaching me some Ukrainian and some physics; Ksenia Bulychevia, for our foray into machine learning; Anne Gambrel, Ed Young, Tom Hazard, Lucia Mocz, Vladimir Kirilin, Christian Jepsen, and Debayan Mitra for your friendship and fun times; and Matthew "Math" de Courcy-Ireland, for your close friendship. A special acknowledgement to Zach Sethna, who encouraged me to change fields and offered advice

when I did, as well as for enlightening discussions of physics and brainteasers.

To Darryl Johnson, who manages to brighten my day every time he says hello, and Kate Brosowsky, who has on numerous occasions gone out of her way to make my life (and the lives of all graduate students) easier or more pleasant.

Thank you to the NSF for funding three years of my graduate career (and likely the grants that came after).

Outside the department, and in no particular order: Many thanks to Esma Pasic-Filipovic, from whom I have learned so much, and whose lessons often served as an oasis during times of stress. To all the people at CLRA, thank you. The club has been such an integral part of my life throughout my PhD, and during times of fluctuating commitment, the many friendships I have made have not waned.

Many thanks to my mentors and inspirations from MIT and before. Mehran Kardar, your lectures on statistical mechanics and on stat mech applications to biophysics are a large reason why I'm doing what I'm doing, and I aspire to your level of organization and clarity while communicating the excitement of the subject. I might not be doing biophysics without the inspiring influence of Jeff Gore; I would not be doing physics without the encouragement and inspiration of my previous research advisor John McGreevy and my previous mentor Krishna Rajagopal; and I would not be doing research without my incredible mentor Todd Kemp, in mathematics. Thank you to Jonathan Farley for introducing me to higher level math and to the world of academia, and thank you to David Meyer for taking the time to teach it to me. Mr. (Ken) Panaro, thank you for inspiring me to pursue science. Thank you for your high standards, for your support, and for reminding me to see everything through to completion many years later.

There are too many people to acknowledge from MIT, but here are a few. Thank you Haofei Wei, Katie Puckett, Shaunak Kishore and Phil Tynan for your continuing friendship. Thank you to Maria Monks and everybody at Random Hall who served as an inspiration to me. Thank you to the support of everyone on MITLW rowing, but especially my teammates; Marie McGraw, for continuing friendship; and my coaches Claire Martin-Doyle and Amelia Patton. Thank you to Ken Fan, for all the support over the years, and for being an inspiration with your propensity for hard work.

Among the most important people I have to thank is Mrs. (Lyubov) Shlain. Mrs. Shlain, thank you for showing me that with work, I am capable of much more than I think. I wouldn't be here without your influence.

And, last but most importantly, a huge thank you to my closest support network and family. Zenab Tavakoli, thank you for being here after so so many years. Thank you for being excited about black holes and biophysics, and for letting me explain them to you. Thank you for your friendship and support. Thank you to Mallika Randeria, my running and rowing buddy, for all your support and friendship throughout graduate school. Whether it be running a half marathon, racing a 1k, writing a paper, or switching fields, you've been there. (Thank you also to your mom, condensed matter physicist Nandini Trivedi, for her wonderful discussions and advice!)

Thank you to Nancy and David Smith for your random notes just because you were thinking of us and wanted to let us know you care.

And of course, thank you Mom (Merry Smith), and Tessie McGough (and Jeff Smith!); the many phone calls and neverending support have been invaluable throughout whole process. Finally, thank you to my husband, high energy theorist Kenan Diab, for our many physics discussions, for your humor and optimism, and for your neverending love and support. Over the years, these have meant the most to me.

*If you know what you don't know, then what you don't know won't hurt you...*

---

*Dedicated to Grandpa "Doc" Phil McLaren.*

# Contents

# Chapter 1

# Introduction

Broadly speaking, one of the physicist's most basic goals is to measure properties of a physical system, and then use that information to infer fundamental principles underlying the system's behavior. Yet incomplete information is intrinsic to the study of physical systems. This is natural from even the most basic of perspectives: it is trivial to use a ruler and timer to predict the motion of an object falling a few feet to the ground, but any measurement will be limited by the ticks on the ruler and precision of the stopwatch. Measurements carry with them error bars that parametrize our ignorance.

Of course, measurement error bars are but one of many forms of unavoidable ignorance the physicist must accomodate, with different systems and different questions about those systems carrying with them diverse forms of built-in uncertainty. The coarse-graining of space by the finite ticks on a meter stick serve is reminiscent of the renormalization group (RG) perspective of quantum field theory which organizes physics according to the scale at which we describe it. Our inability to know the position and momentum of every water molecule in a cup of water leads to statistical uncertainty, which we model with statistical mechanics and thermodynamics. Our inability to simultaneously know the position and momentum of even a single electron with arbitrary precision is a consequence of quantum mechanics.

The notion that parametrizing uncertainty leads to new physics underlies much of the current research in diverse fields of theoretical and applied physics, including black hole thermodynamics, holography, entanglement, quantum phases of matter, information processing and many-body methods in biological systems, and even machine learning. In this thesis, we will reflect this diversity by studying embryonic development, entanglement, topological quantities, holographic renormaliza-

tion, and dualities in CFT and $2+1$-d black holes. One unifying theme among the chapters will be that our ignorance - whether it be represented by statistical noise in concentrations of proteins, quantum entanglement entropy, black hole entropy, or integrated out degrees of freedom - will guide us to universal phenomena, including precision in embryonic development, scaling features of Rényi entropy, an exact dictionary for $AdS_3/CFT_2$, and a CFT dual to a BTZ black hole in a box.

The structure of this thesis is the following. In the remainder of this section, I review known physics from an information-theoretic perspective with the goal of providing context to what follows. I proceed from classical information theory and the principle of maximum entropy, to thermodynamics, quantum information theory, the renormalization group, and lastly, a lightning overview of AdS/CFT's connection to entanglement and the renormalization group.

Following this overview, we will present five works:

- Chapter 2 will review the paper in progress, "Reproducibility in Development from Correlated Fluctuations," (with W. Bialek), in which we show that spatial correlations in the concentrations of morphogens like the gap gene proteins in the fruit fly can provide "missing information" necessary to specify a body plan in an early stage of development. We also address the apparent confusion between discrete identities of cells vs. continuous morphogen concentrations, and consider the potential for error correction mechanisms in a model of inference as a random field problem in a discrete gaussian model with correlations.

- Chapter 3 will review the paper, "Rényi Entropy and Geometry," (with J. Lee and B. Safdi), wherein we conjecture a new universal structure in the Rény entropy of CFTs in 4-d. We provide evidence for the structure by numerically computing Rényi entropy in massive free field theory in $2+1$-d.

- Chapter 4 is based on the paper "Bekenstein- Hawking Entropy as Topological Entanglement Entropy," (with H. Verlinde), which computes the "topological entanglement entropy" of the BTZ black hole. This matches the Bekenstein-Hawking entropy and has an interpretation as the geodesic length of the horizon.

- Chapter 5 is based on the paper "Conformal Bootstrap, Universality and Gravitational Scattering," (with H. Verlinde and S. Jackson), in which we propose a duality between the gravity-dominated regime of $AdS_3$ and the maximal solution to the Virasoro bootstrap constraints, Liouville theory.

- Chapter 6, based on the paper "Moving into the Bulk with $T\bar{T}$" (with H. Verlinde and M.

Mezei), proposes that a $T\bar{T}$ deformation of a 2-d CFT is dual to a bulk dual with a hard cut off at a finite radius. The duality proposal is strongly supported through thermodynamic computations made possible by the integrability of the $T\bar{T}$ deformation. These thermodynamic computations perfectly match computations performed in the context of black holes in a "cut off" AdS space with Dirichlet boundary conditions.

## 1.1 Shannon entropy and mutual information

The following material can be found in information theory textbooks such as [1, 2].

In order to say anything quantitative about the physical implications of having or not having information about a system, we would like to have a rigorous definition of information. We thus begin our discussion with a digression into ideas first introduced in the context of information theory. Namely, we start by defining Shannon information, Shannon entropy and mutual information.

The notion of having or not having information about some system comes hand-in-hand with having uncertainty about the system. Consider some classical random variable $x$ which takes on discrete values $\{x_i\}$ with respective probabilities $p(x_i)$. (Although continuous variables are also valid, we assume discreteness for convenience.) Denote the set of outcomes $\{x_i\}$ by $X$. We would like to define a property of the distribution which measures how much uncertainty we have about the value of $x$. Intuitively, if the distribution is heavily weighted on $x_j$ for a specific $j$, we would say we have very little uncertainty, whereas if it is very nearly uniform across the different values $x$ can take on, we would say we have near-maximal uncertainty.

Suppose we consider a collection $W$ of $N$ "words," with $N \gg (\min p(x_i))^{-1}$. Each word $w \in W$ consists of a consists of a string of $K$ "letters", each of which is an element of $X$ pulled independently from the distibution $p(x)$, $K > 2$. Consider specifying a specific word $w \in W$ by specifying each letter one by one. With probability $p(x_i)$, we will receive $x_i$ as the first letter, allowing us to narrow down the space of possible words from $W$ to $W_1 \equiv \{w \in W | w_1 = x_i\}$. The expected size $|W_i| \approx p(x_i)N$. If $p(x_i)$ is very small, the number of possible remaining possible words is small; if $p(x_i)$ is near 1, the number fo possible remaining words is still nearly $N$. The event $w_1 = x_i$ conveys more information if $p(x_i)$ is of low probability.

If the second letter is $x_j$, then, because the letters are independently distributed, the set of remaining words $\{w \in W_i | w_2 = x_j\} \equiv W_{ij}$ has expected size $|W_{ij}| \approx p(x_i)p(x_j)N$. The number of words remaining decreases in a multiplicative fashion as each successive letter is revealed.

Define the information of an event $A$ which has nonzero probability $p(A)$ to be given, in bits, by

$$I(A) = -\log_2 p(A) \tag{1.1}$$

and $I(A) = 0$ if $p(A) = 0$. This is always a nonnegative quantity, and it is additive upon seeing independent events. In the example above, $I(w_1 = x_i, w_2 = x_j) = I(w_1 = x_i) + I(w_2 = x_j)$.

The logarithmic information matches the intuition we developed relating the amount of information one has to the reduction in uncertainty. To see this, consider the physics definition of entropy of a set $T$, written in bits, $S(T) = -\log_2 T$. Here, the information conveyed by the event $w_1 = x_i$ is just the amount by which the entropy of possible outcomes decreases upon learning that information: $S(W_i) = S(W) - I(x_i)$. Further learning $w_2 = x_j$ gives $S(W_{ij}) = S(W_i) - I(x_j) = S(W) - I(x_i) - I(x_j)$.

We can study the expected value of the information over a distribution by considering the expected information conveyed by each outcome. This is known as the Shannon entropy. It's given by $\langle I(x_i)\rangle_p$,

$$S(p) = -\sum_{x_i \in X} p(x_i) \log_2 p(x_i)\,, \tag{1.2}$$

taking $0 \log 0 \equiv 0$ for events with zero probability. The logarithmic information and Shannon entropy are especially important because in [3], they were proven to be the unique quantities which satisfy just a few highly natural properties one might want a measure of information to hold, up to a constant factor which amounts to a choice of units.

There is another crucial measure of information which can be defined, this time between two *different* random variables. Let $X$ and $Y$ be two random variables with a joint distribution $p(x, y)$. The mutual information specifies how much information the value of $X$ gives about the value of $Y$ and vice versa. It is defined by

$$I(X;Y) = I(Y;X) = S(X,Y) - S(X|Y) - S(Y|X) \tag{1.3}$$

$$= S(X) - S(X|Y) \tag{1.4}$$

$$= S(Y) - S(Y|X) \tag{1.5}$$

If specifying the value of $X$ does not give any information about the value of $Y$, then $S(Y) = S(Y|X)$ and the shared information is zero. On the other hand, if the value of $X$ completely specifies the

value of $Y$, then $S(Y|X) = 0$ and $I(X;Y)$ is maximized. In an intermediate case, $I(X;Y)$ is positive but less than $S(Y)$. In the example specifying a word $w$, different positions in the word share no information because we've assumed them to be independent; thus the mutual information vanishes. This is the case whenever $X$ and $Y$ are independent random variables. On the other hand, if $x_i$ were *always* followed by $x_j$, then the two variables will maximize their mutual information. This quantity is an important measure for physical systems and is commonly used in both biophysics and high energy theory. One relevant application of the mutual information is to positional information in embryonic development, as in chapter 2.

Other measures of information exist; one such example is the Rényi entropy, which is similar to the Shannon entropy in certain respects. The $q$th Rényi entropy is given by

$$S^q(p) \equiv \frac{1}{1-q} \log \sum_{x_i \in X} p(x_i)^q \tag{1.6}$$

The Shannon entropy equals the limit $\lim_{q\to 1} S^q$. The Rényi entropies are in a sense "inferior" to the Shannon entropy because they does not satisfy certain crucial axioms, but they are often computable when the Shannon entropy is not. A common trick in statistical mechanics, quantum mechanics and quantum field theory is to compute the Rényi entropy at $q \neq 0$, which requires computing powers rather than logarithms, and then take a "limit" as $q \to 0$ in order to find the Shannon entropy (the "replica trick"). We will have much to say about Rényi entropy in chapter 3.

## 1.2 Information, statistical mechanics and thermodynamics

Statistical mechanics describes the microscopic origins of the collective thermodynamic properties of systems in thermal equilibrium. However, statistical mechanics can also be understood in the "inverse" perspective: as the study of "minimal information added," aka max entropy. It addresses the question: assuming that we can only measure thermodynamic quantities, how should we model microscopic behavior consistent with our measurements?

We often assume, in equilibrium thermodynamics, that the only information we have access to about a physical system is the total energy, and the spectrum of microscopic energies. The minimal information added distribution corresponds to the *entropy-maximizing* distribution. Specifically, given the constraint that the mean energy is fixed to be $E$, to model the microscopic distribution,

we extremize the function

$$F[\{p\}] = -\sum_i p_i \log p_i + \alpha \sum_j p_j - \beta \sum_k E(k) p_k \tag{1.7}$$

where we recognize the first term as the Shannon entropy of the distribution, $\alpha$ and $\beta$ as lagrange multipliers for normalization and energy, respectively, and $E(i)$ as the (known) energy of the $i$th microstate. Optimizing this gives the familiar Boltzmann distribution,

$$p_i = \frac{e^{-\beta E(i)}}{Z} \tag{1.8}$$

where $Z$ is the partition function and $\beta$ sets the average energy. Finding the distribution that maximizes the entropy conditioned on total energy gives the Boltzmann distribution, familiar from statistical mechanics. Similar derivations hold when there are more conserved quantities.

We've shown that we can define entropy in the sense of information theory and obtain the physics of statistical mechanics as an output. Suppose we'd like to model the microscopic behavior of a physical system with many degrees of freedom, where we know the spectrum of microstates but have incomplete information about the current state. By specifying the values of the quantities we do know, maximizing Shannon entropy can be interpreted as adding the least additional information to our model. Then statistical mechanics becomes intricately connected with information theory as the study of answering the question, "what should one assume when you lack complete information?"

## 1.3 Entanglement entropy

Entanglement entropy, a measure of entanglement and therefore levels of "quantumness," is nothing more than a generalization of the Shannon entropy to density matrices. Whereas Shannon entropy is a measure of classical probability distributions, its quantum generalization is an essential probe of physical properties of quantum systems in fields as diverse as quantum optics, quantum computation, condensed matter physics, and high energy theory.

Consider a quantum system with Hilbert space $\mathcal{H}$. Moreover, suppose $\mathcal{H}$ factories into a product of two sets of degrees of freedom, $A$ and $\bar{A}$. That is, $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$. Let the system be in some state $|\Psi\rangle$. For example, $\mathcal{H}$ may be the Hilbert space of a quantum field theory in the ground state with density matrix $\rho$. Take $A$ and $\bar{A}$ to represent a region of space and its complement, respectively.

Define the reduced density matrix

$$\rho_A = \mathrm{tr}_{\bar{A}} \rho \tag{1.9}$$

to be the "state" of the degrees of freedom inside $A$ upon tracing out degrees of freedom in $\bar{A}$. The entanglement entropy between $A$ and $\bar{A}$ is defined as the von Neumann entropy,

$$S_A = -\mathrm{tr}\left(\rho_A \log \rho_A\right) \tag{1.10}$$

For a diagonal matrix $\rho_A$, $S_A$ is nothing more than the Shannon entropy of the diagonal elements of $\rho_A$.

Much is known about entanglement in the ground state of quantum field theories. For example, the leading singular term (in terms of a short-distance cutoff) typically satisfies an area law; in $1+1$-d CFT it is proportional to the central charge of the CFT; and in $2+1$-d topological field theory there is a topological term calculable from properties of the gauge group of the field theory.

As in the classical system, there exists a generalization of entanglement entropy, the $q$th Rényi entropy of $\rho_A$, which will also be relevant:

$$S_A^q = \frac{1}{1-q} \log \mathrm{tr}\left(\rho_A^q\right) \tag{1.11}$$

Although they are less physical than the entanglement entropy in the same way the classical Rényi entropy is less physical than the Shannon entropy, the Rényi entropies are easier to compute and give more refined information about the entanglement structure of the field theory since collectively, they give the eigenvalues of the reduced density matrix. One common strategy for computing the entanglement entropy is to take a formal limit of Rényi entropy as $q \to 1$ (the replica trick).

## 1.4 Integrating out short-distance degrees of freedom

One form of "information loss" with deep but not transparent connections to entanglement is that of renormalization group flow from UV to IR. RG formalizes the fact that we are able to describe many macroscopic phenomena without knowing their microscopic details; that is, we may describe the motion of a bouncing ball without knowing the detailed interactions among its atoms.

There are two common pictures of RG flow: the Kadanoff real-space RG [4], and the Wilsonian momentum-space RG [5]. Real-space RG begins with the intuition that long-distance degrees of

freedom should be derived from the information in short-distance degrees of freedom through an averaging (smearing) procedure among degrees of freedom in real space (say, neighboring spins in a lattice). Performing this averaging procedure decreases the number of points in the lattice, hence the interpretation as "coarse-graining". Generically, the original Hamiltonian can be regrouped and rescaled until it has the same form as the original Hamiltonian, except the new Hamiltonian is in terms of the smeared variables with different values of the couplings. This procedure can be interpreted as the study of the system at different scales, and thus fixed points of RG are often conformal field theories that have scale (and Weyl) invariance.

Like real-space RG, Wilsonian RG is concerned with the question of obtaining long-distance behavior of a system from its short-distance microscopic behavior, but now, working in momentum space, integrating out high-energy degrees of freedom while keeping low-energy degrees of freedom. Working in momentum space is often more convenient for continuum QFT and for doing computations in field theory more generally.

In carrying out Wilsonian RG, we choose a momentum cutoff $\Lambda$ to be large compared to other scales of the problem. (We work in Euclidean signature so that the norm is always nonnegative.) We restrict ourselves to considering momentum modes $\vec{k}$ with norm less than $\Lambda$. Then, we choose a real number $b < 1$ and break our field, $\psi(\vec{k})$, into two pieces: $\psi(\vec{k})_<$, which agrees with $\psi(\vec{k})$ on $|\,\vec{k}\,| < b\Lambda$ and vanishes on $|\vec{k}| > b\Lambda$; and $\psi(\vec{k})_>$, which agrees with $\psi(\vec{k})$ for $b\Lambda < |\vec{k}| < \Lambda$ and vanishes otherwise. Note that $\psi(\vec{k}) = \psi(\vec{k})_< + \psi(\vec{k})_>$.

In order to define the renormalized partition function, we integrate over momentum modes with norm between $b\Lambda$ and $\Lambda$:

$$Z = \int D\psi e^{-S_E[\psi]} = \int D\psi_< \int D\psi_> \, e^{-S_E[\psi_<, \psi_>]} \tag{1.12}$$

$$\equiv \int D\psi_< \, e^{-\tilde{S}_E[\psi_<]} \tag{1.13}$$

where $\tilde{S}_E$ is defined such that

$$e^{-\tilde{S}_E[\psi_<]} = \int D\psi_> \, e^{-S_E[\psi_<, \psi_>]} \tag{1.14}$$

The action $\tilde{S}_E[\psi_<]$ is an effective action on the momentum modes with norm less than $b\Lambda$; for this reason, it is said to model the low-energy physics. This quantity represents a coarse-grained action in momentum space, just as we defined a coarse-grained Hamiltonian in position space using real-space RG.

To connect with the previous discussion we'd like to have an explicit connection between RG and more information-theoretic concepts like entropy. In recent years, it was shown that this is in fact the case in $1+1$, $2+1$ and $3+1$ dimensions. The connection is given by the so-called $c$, $F$ and $a$ theorems, respectively.

The motivating question is, can we formalize the idea that RG flow truly coarse-grains in an irreversible way? Is there a well-defined notion of whether one theory could ever be definitively "more UV" or "more IR" than another? One way to answer this question in the affirmitive would be to provide a quantity that is always monotonic under RG flow. Indeed, it has been shown in the $c$, $F$ and $a$ theorems that in $1+1$, $2+1$ and $3+1$ dimensions, one such quantity can always be derived from the entanglement entropy on a sphere, although these were not all originally proven in terms of entanglement [6, 7, 8, 9, 10, 11, 12]. The precise notion that RG "fuzzes" the microscopic information irreversibly relies on specific, unique properties of EE (here, strong sub-additivity), and we discover another form of "uncertainty" in physical systems which has information-theoretic underpinnings.

## 1.5 Gravity, a hologram of field theory's information content?

The AdS/CFT correspondence is perhaps the most dramatic convergence of the ideas discussed thusfar [13, 14, 15, 16]. In this section, we give a flavor for the statement and how ideas like thermodynamics, entropy and RG fit into the picture.

A conservative definition of AdS/CFT would be as a conjectured duality between conformal field theories in $d+1$ dimensions and asymptotically-AdS theories of gravity in $d+2$ dimensions, although more general examples of such dualities have been found. It is possible to relate quantities on the boundary field theory to corresponding quantities in the bulk gravity theory, and much work on the correspondence has gone into studying precise manifestations of this principle. One of the earliest such examples is the GKPW formula [14, 17], which relates CFT correlators on the boundary to the bulk partition functions:

$$\left\langle e^{-\int d^d x \phi_0(x)\mathcal{O}(x)} \right\rangle_{CFT} = \mathcal{Z}_{grav}\left(\phi(x,z)|_{z=0} = \phi_0(x)\right) \tag{1.15}$$

Here, $x$ gives the $(d+1)$ coordinates in the boundary CFT, the LHS gives the generating functional of correlation functions of the operator $\mathcal{O}(x)$, and the RHS represents the gravity partition function

integrating the bulk field $\phi(x,z)$ dual to $\mathcal{O}(x)$, restricted to have boundary condition $\phi_0(x)$ as $z$ approaches the boundary. The AdS/CFT correspondence is usually most conveniently defined in the limit where the boundary CFT is strongly coupled and the bulk string coupling $g_s$ is large: $N \to \infty$ and $g_s^2 N \to \infty$. It is in this limit that Einstein gravity can be used in the bulk, making many physical quantities computationally tractable.

The relationship between AdS/CFT and thermodynamics was evident from the outset as the notion of holography grew out of the observation that black hole entropy scales as the area of the event horizon, rather than the volume,

$$S_{BH} = \frac{A}{4G_N} \tag{1.16}$$

How could it be that the gravitational degrees of freedom are somehow encoded on the horizon? Naively, the scaling would be as the volume. In the black hole thermodynamics context the idea that gravitational degrees of freedom can be encoded on a "boundary" in one fewer dimension is manifest [18, 19, 20].

Consider the geometry of $d+1$-dimensional AdS in Poincare coordinates,

$$ds^2 = \frac{\ell_{AdS}^2}{z^2}(d\vec{x}^{\,2} - dt^2 + dz^2) \tag{1.17}$$

Here $\ell_{AdS}$ is a constant which dictates the curvature of the spacetime, $z$ represents the bulk radial direction and the conformal boundary lives at $z \to 0$. Consider a fixed-$z$ slice. It is merely Minkowski space with some scaling factor. As we approach the boundary $z \to 0$, the scaling factor blows up. This rescaling is reminiscent of coarse-graining the field theory upon going to a finite radius in the bulk, where the infinite scaling factor near the boundary corresponds to a UV fixed point. Although these arguments are merely suggestive, holographic RG has proven to be a rich subject [21, 22] and has taught us much about CFT, gravity, and the duality between them. We will discuss one such contribution in chapter 6.

If holographic RG indicates that the bulk "organizes" the field theory's degrees of freedom into different energy scales, boundary entanglement entropy indicates that the *field theory* organizes the *bulk* degrees of freedom through its spatial entanglement structure. That is, in 2006, Ryu and Takayanagi gave strong evidence the black hole thermodynamic intuition can be used to measure spatial entanglement entropy on the boundary [23, 24, 25, 26]. In particular, if $\Sigma$ is a boundary entangling surface (codimension 2 in the boundary), its entanglement entropy is given by the mini-

mum (extremal) area of a Euclidean-(Lorentzian-) signature bulk surface $\Sigma'$ (codimension 2 in the bulk) anchored at $\Sigma$,

$$S_\Sigma = \frac{A_{\Sigma'}}{4G_N} \tag{1.18}$$

The RT formula demonstrates that CFT EE is encoded in bulk areas. Moreover, bulk degrees of freedom are organized by regions enclosed by minimal (extremal) surfaces $\Sigma'$ defined by the entanglement structure of the boundary state. Which degrees of freedom are contained in some $\Sigma'$, and which are "missed" by the entanglement structure? How do the $\Sigma'$ relate to the regions of spacetime which are somehow "influenced" by a given region in the boundary? Much work has gone into studying these questions [27, 28, 29, 30, 31, 32, 33], although the tantalizing idea that gravity somehow emerges from entanglement is still out of reach.

The AdS/CFT correspondence is a surprising statement: somehow, quantum gravity hides in field theories, in one fewer dimension, without gravity. This is highly analogous to the thermodynamic origins of black hole entropy, and there are hints that it is deeply tied to the information-theoretic properties of the field theory; its behavior under coarse-graining and its entanglement structure are enough to encode a highly nontrivial theory of quantum gravity.

In summary, we've indicated that there are fundamental relationships between information theory and our foundational models of thermodynamics, quantum entanglement, renormalization group, and even gravity. We've found that there is a sense in which it is quantitatively true that these many different ways of parametrizing and learning from our ignorance are actually facets of the same measure of information content.

We will now proceed to discuss aspects of embryonic development, Rényi and entanglement entropy, AdS/CFT, holographic RG, and thermodynamics, in roughly that order.

# Chapter 2

# Reproducibility in development from correlated fluctuations

This chapter is based on a work in progress in collaboration with William Bialek [34].

In this chapter, we consider how the information specifying the body plan of a fruit fly is encoded in the embryo during the early stages of its development, from the formation of the egg through the first four hours of development. During this time, approximately 6000 nuclei form and migrate to the surface of the embryo. In the absence of cell membranes, these nuclei sense the local concentrations of morphogens which are present, and use this information to determine their unique identities which define their entire lineages up to the final form of the adult fly, from the head and through the segments of the thorax and abdomen. Although these morphogen concentrations provide information to the nuclei about their future development, the exact concentrations fluctuate from embryo to embryo, and thus it is not obvious that the information transmitted at this stage is enough to correctly specify a unique identity for each nucleus and thus produce reproducible body plans across different members of a species. At what stage does the system of development become precisely reproducible across embryos, and what are the physical properties of the biological "signals" which imply this reproducibility?

These questions are naturally stated in the context of information theory, which is concerned with quantifying the information that can be conveyed in systems subject to uncertainty or noise. For example, would like to understand the transmission of information about a body plan through morphogen concentrations, a noisy signal which varies somewhat from embryo to embryo. We choose to study fruit fly embryo development because current experiments are able to take quantitative

measurements which characterize the spatial dependence of morphogen concentrations in detail.

In the first few hours of fruit fly development, a network of four genes known as the gap genes expresses proteins whose spatially varying concentrations serve as a "map" of positions along the anteroposterior axis of the embryo. This map has spatial resolution of approximately 1% the embryo's length [35], less than the spacing between individual nuclei. One would like to say that this spatial resolution is enough to explain the high level of reproducibility of body plans between embryos, since each nucleus should be able to determine its fate with very high precision, but in fact this is not quite so; that is, 1% spatial resolution is not enough if nuclei are inferring their positions using independent information. Moreover, a description in terms of high levels of spatial resolution is hiding a confusion, that we are using information about continuous positions to describe an inherently discrete notion, that of cell fates.

We will address the issue of reproducible inference of a body plan by showing that when fluctuations in morphogen concentrations are correlated, the probability that nuclei infer their relative positions incorrectly decreases exponentially as the correlation length increases. This is relevant to the physical system because it is known that the fluctuations in concentrations of the gap genes are long-range spatially correlated [36]. We then address the distinction between discrete identities and continuous positions by showing that a nucleus which lives in a finite-length embryo optimizes the information it can extract from the gap gene protein concentrations by using a discrete prior over its spatial coordinate. This creates an "emergent" discreteness without needing to build a discrete lattice explicitly into the model [37, 38].

In Section 2.1, we review the biological mechanisms by which positional information and hence information about cell fates is set up in the early fly embryo. We also make the idea of "precision" in development more precise by providing examples where biological systems exist at or close to intrinsic physical limitations.

In Section 2.2, we begin by discussing precisely what is meant by nuclei "knowing" their positions in the embryo subject to some error. We review the results of [39] demonstrating that during cycle 14 of fly development, nuclei have enough information to infer their positions within standard deviation of less than one cell spacing, and they have a specific, computable decoding map allowing them to do so. We then review results demonstrating that this level of precision is not enough to specify a sufficiently long sequence of unique identities. This is quantified by defining positional information as the mutual information between the distribution of gap genes and positions along the anteroposterior axis of the embryo. We review the statement in [35] that there are not enough

bits to specify the entire sequence, experimentally. Moreover, we compute that two neighboring cells with gaussian distributed inferred positions have a nontrivial probability of inferring their relative positions incorrectly even if each knows its own position with high precision.

In Section 2.3, inspired by [36], which found long-range spatial correlations in the fluctuations of the gap gene protein concentrations, we demonstrate that spatial correlations in the gap gene fluctuations increase the information carried by the noise by explicitly computing the amount by which exponentially decaying correlations increase the information carried by the system. We also demonstrate that spatial correlations in the fluctuations of the gap gene concentrations can decrease the probability that neighboring cells infer their fates out of order when the fluctuations have exponentially decaying spatial correlations. Indeed, correlations exponentially decrease the average number of neighbor ordering errors as well as the probability of having at least one error in specifying a sequence of $N$ unique identities. Thus correlations help solve the problem of encoding an entire body plan in addition to identities of individual cells.

In Section 2.4, we address the question of why discrete cell fates emerge from continuous signals specified by continuous positional coordinates. We demonstrate that if cells in a developing embryo are optimal decoders of the local gap gene concentration profiles, discrete degrees of freedom emerge naturally as a result of the fact that the embryo has finite length. We show, however, that for all intents and purposes, the embryo does not "sense" this discreteness, as the information in the system is nearly identical to that of the analogous continuous system.

We end in Section 2.5 with a model of this inference problem using a statistical mechanical system of discrete degrees of freedom on a lattice subject to correlated noise – a discrete gaussian model with an added random field. We discuss the possibility that, depending on the correlation strength and hence the phase structure of the system, this may specify an error correcting code for embryo development.

Overall, we address the problem of reproducibility of a precise body plan specified in the early stages of embryo development by showing that spatial correlations in the fluctuations of the morphogen concentrations can significantly enhance the gap gene system's ability to reproducibly encode the body plan of the fruit fly. We also address the apparent contradiction between discrete identities vs. continuous positions by showing that an optimal decoder will use a discrete prior, but when the noise is small, the distinction between discreteness and continuity is less important. Finally, combining these ideas, we discuss the possibility that statistical mechanical systems of discrete degrees of freedom subject to correlated nose could perform perfect inference depending on the properties

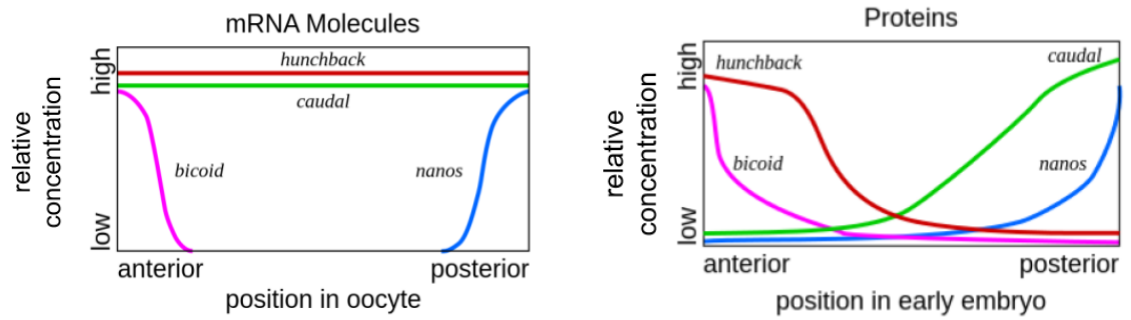



Figure 2.1: Concentrations of the maternal effect gene mRNA molecules and their corresponding translated proteins as a function of the coordinate along the anteroposterior axis (cartoon). In this cartoon, the values of the axes are not necessarily meant to be taken literally; for example, the relative amount of one mRNA molecule (or protein) to another is not necessarily to scale. [40, 41]

of the correlations.

## 2.1 The biology of early fruit fly development

In this section, we discuss the biological mechanisms by which a body plan is specified in the early stages of fruit fly embryonic development, with particular emphasis on the mRNA and proteins whose concentrations hold positional information. We also review the notion of discrete cell identities as well as concrete examples of reproducibility in development.

Most of the following discussion is based on material from [42].

The story of the highly reproducible patterns which characterize the adult fly's body plan begins in the earliest stages of development, even before fertilization, when the mother endows the egg with "maternal effect" mRNA molecules with particular spatial concentration profiles. These include *bicoid*, whose concentration is localized at the anterior pole before fertilization and, following fertilization, diffuses to create a decreasing gradient along the anteroposterior axis; *nanos*, localized at the anterior pole before fertilization and, following fertilization, diffuses to create a decreasing gradient along the anteroposterior axis; and *hunchback*, whose mRNA concentration begins as roughly constant throughout the egg. Following fertilization, these mRNA molecules are transcribed into the respective proteins, setting up corresponding nontrivial spatial concentration profiles defined by mutual transcriptional regulation as well as diffusion. Whereas Bicoid and Nanos proteins remain highest at the anterior and posterior poles, respectively, Hunchback is no longer evenly distributed throughout the egg; rather, Nanos represses its transcription, creating a Hunchback gradient which is higher in the anterior pole. Bicoid represses the transcription of Caudal, which develops a gradient

higher at the posterior pole. Note that these interactions among transcription factors indicate that these molecules do not necessarily all specify *independent* pieces of positional information. This is depicted in fig. 2.1.

The maternal effect genes serve as "initial conditions" for the embryo, defining the polarity. Their importance on the anteroposterior organization of the organism is illuminated by the effects of mutations on the body plan of the embryo; maternal mutations in bicoid can lead to missing head and thorax structures, for example – lethal mutations. Moreover, the surprisingly high degree of precision which characterizes our entire discussion begins at the level of these initial conditions: the *number* of *bicoid* mRNA molecules deposited by the mother varies by only $\sim 9\%$ among individuals [43], comparable to the $\sim 10\%$ reproducibility of the Bicoid gradient in the anterior half of the embryo [44].

After fertilization, the single original nucleus begins replicating without forming cell membranes, creating a so-called "syncytial blastoderm" of a single cytoplasm containing many nuclei. Each division defines a cycle of early development, indexed by the time preceding the division (such that the 10th cycle follows the 9th division, for example). The initial divisions are synchronous and occur every 9 minutes.

After around one hour - 7 divisions - the embryo contains 128 nuclei, and the vast majority of the nuclei are migrating toward the outer membrane of the embryo (with just a few remaining as yolk nuclei, less important for pattern formation, and 15 moving to the posterior pole, dividing on a different schedule). By cycle 10, essentially all of the nuclei have reached the membrane, forming a 2-d surface of $\sim 6000$ nuclei. Although the nuclei lack cell membranes, throughout their migration they are surrounded by structural molecules giving each nucleus its own individual "islet" within the common cytoplasm shared by all the nuclei.

Throughout this period, the positional information is primarily provided by proteins transcribed from maternal effect genes, like Bcd, Hb, Nanos, and others.The concentration profiles are produced both by passive mechanisms like diffusion and active mechanisms resulting from the regulation of one protein on another's transcription as well as dynamics maintaining mRNA gradients [45]. Key to the system is also that Bicoid regulates the transcription of Hunchback in a very precise manner, with a reduction in noise such that even individual nuclei can experience reproducible patterns of distinguishable levels of expression [44]. This is a manifestation of precision just downstream from the maternal effect genes and proteins.

During cycle 14, several crucial changes take place. First, the membrane surrounding the egg begins to fold in toward the cytoplasm, engulfing each nucleus to form individual cells, transforming

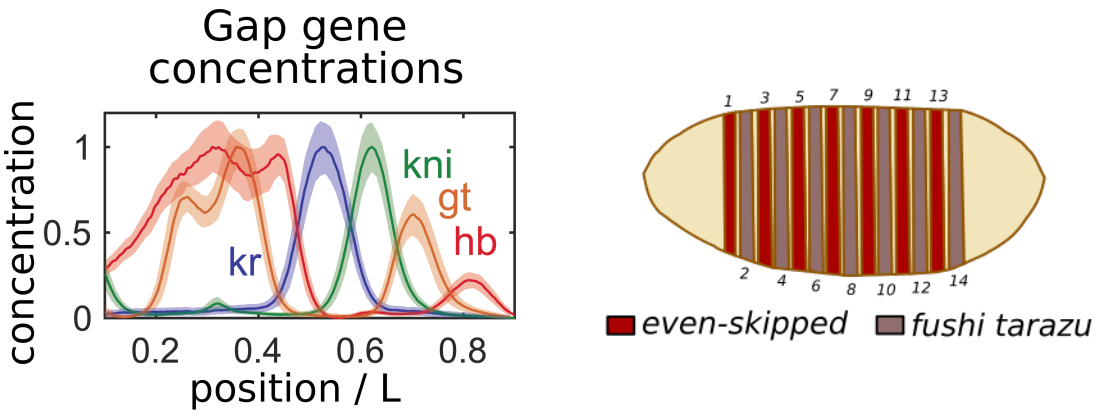



Figure 2.2: Concentrations of gap gene products along the anteroposterior axis [39] and a cartoon of the expression of two pair rule proteins [46], viewed as a slice through the embryo with stripes perpendicular to the anteroposterior axis. The $x$-axis of the left figure labels percent of the length along the anteroposterior axis, and the $y$-axis is in arbitrary intensity units. Each concentration has been normalized such that its maximal value along the length of the axis equals 1.

the syncytial blastoderm to a cellular blastoderm. During this process, which takes approximately 50 minutes and during which there are no nuclear divisions, new morphogens begin to play a critical role in determining the body pattern. Hunchback and Bicoid act as regulators, turning on the *gap gene network*, a gene network $\sim 6$ proteins (of which 4 are of relevance to our story), which is crucial for the specification of more refined positional information than just the polarity of the embryo, and which will be a main player in our discussion.

The gap gene network begins to specify the boundaries of regions along the length of the body of the fly, and its importance in specifying the body plan is well-known due to Nobel-prize-winning work by Wieschaus and Nüsslein-Volhard [47] and much subsequent work on the subject. The gap genes are named so because of the (fatal) effects of their mutation or deletion; namely, the cutting out of one or several contiguous segments in the larva (hence, a "gap"). For example, a hunchback mutation deletes the mesothoratic and metathoratic regions, and an embryo homozygous for a Kruppel mutation has no thorax or anterior abdomen. These "gaps" are distinctly discrete; this is a key clue that the information encoded by the gap genes is discrete.

Also during cycle 14, the gap genes activate one more gene network which further refines positional information. These are the pair rule genes, which are expressed in stripes along the length of the embryo, as shown in fig. 2.2; this pattern determines the organization of parasegments in the body plan, stripes and their correponding parasegments being another discrete characteristic of the fruit fly body plan. Pair rule mutants fail to express every other stripe, as indicated in names of pair rule genes like "even-skipped" and "odd-skipped".

The nontrivial positional information encoded in the maternal effect genes, gap genes, pair rule genes, and, indeed, potentially other analogous system of morphogens in different organisms, is a result of their nontrivial spatial concentrations. These different concentrations act as "coordinates" along the embryo, localizing points in space in a manner limited by noise and the amount of invertibility vs. redundancy in the mapping (for example, a uniform distribution contains no information about position, but a linear gradient with no noise contains perfect information about position along one axis). The extent to which concentrations of morphogens determine position, including the noise in this measurement, is a key determinant in the question of precision in development, and we'll revisit this in more quantitative detail later in this section.

The 14th cycle is especially important to a discussion of precision and information in embryo development because it seems to occur near the boundary of time at which nuclei do and do not take on "unique identities" (as determined by their fate). For example, the following experiment described in [42], indicates that individual nuclei do *not* take on well-defined physical identities by the 10th cycle of division, but they do specify unique compartments just after the 14th cycle of division. When an experimenter marks one nucleus by deleting a chromosome during the 10th cycle (i.e. during cleavage) and lets the embryo develop into an adult fly, the marked nuclei representing the lineage of the original nucleus could "colonise several different organs and germ layers" [42]. However, after marking a nucleus at the end of the 14th cycle, marked nuclei appear in only one compartment (a localized region) in the adult fly. One concludes that the individual nucleus in the 10th cycle did not have a lineage with a well-defined functional identity in the adult fly, but the 14th cycle nucleus did. This indicates that by the time the cellular blastoderm was fully formed, each nucleus had a well-defined identity, and perhaps that identity was a result of the pattern formation defined by the gap genes and pair rule genes.

Once cells form along the surface of the embryo, the embryo dramatically changes shape and undergoes complex dynamics as it develops into a larva. For example, the ventralic furrow forms at a highly well-defined location along the length of the embryo; subsequently, during gastrulation, the embryo folds onto itself to change topology and create two surfaces; followed by many further processes, leading to eventual hatching of the larva (1 day from fertilization), pupal stage ($\sim 8$ days from fertilization), and completion of an adult fly ($\sim 13$ days from fertilization).

Amazingly, from the initial mRNA molecule concentrations to the pair rule stripe locations to the ventral furrow and throughout the development process, a high level of precision is preserved at every step. As mentioned, the maternal effect mRNA *numerical* counts vary only by $\sim 8\%$

from embryo [43]; moreover, this leads to a $\sim 10\%$ reproducibility in the Bicoid profile [44]. Bicoid reproducibility translates to accuracy in the corresponding Hunchback readout; Hunchback then activates transcription of the gap gene proteins which, due to their spatial concentration profiles and their noise properties, together specify locations along the length of the embryo to a precision of $\sim 1\%$ [39]... only to have all of these and more precise elements lead to precision in the adult fly, such as wing patterns reproducible to half a cell length [48]. The propagation of precision all the way from maternal signals to adult body plan is incredible, and indicates that the processes and genetic networks which control development are tuned to minimize the additional incorporation of noise down to physical limits: precise signals in, precise signals out.

## 2.2 Ambiguous relative positions despite precise individual inference

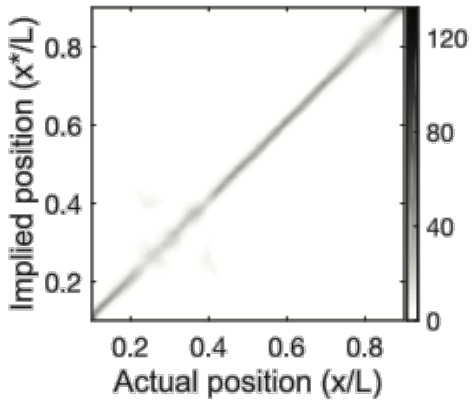



Figure 2.3: The distribution of inferred positions is very nearly the identity, with standard deviation $\sim 0.01L$, less than the spacing between nuclei [39].

In this section, we review previous results which showed that individual nuclei are able to determine their location in the cell to a very high accuracy. We begin with a result [39] which used measurements of gap gene product concentrations to derive an effective "decoder" which can be used to convert from gap gene concentrations to locations in the embryo with high accuracy. We then do a calculation to show that this level of accuracy is still not enough to specify the *relative* positions of nuclei unambiguously, and end with a previous result [35] which demonstrated experimentally that indeed, the high level of accuracy with which each nucleus can specify its position does not correspond to enough *positional information* to unambiguously assign to each nucleus a

unique identity.

In [35], it was shown that individual cells in a fruit fly embryo have enough information to specify their locations to within less than one cell spacing - 1% of the length $L$ of the embryo. Then, in [39], Petkova et. al. determined the "decoding map" which converts local gap gene protein concentrations into locations with spatial precision of $\sigma = 0.01L$. They did so experimentally by measuring the distribution $p(g_1, g_2, g_3, g_4|x)$ of concentrations $g$ of four gap gene proteins – bicoid, hunchback, giant, kruppel and knirps – and using Bayes' theorem to compute the conditional probability distribution $p(x|g_1, g_2, g_3, g_4)$, giving the "decoder" distribution of position after specifying a value for each of the four concentrations. They then used this decoder to construct a different probability distribution for each embryo $\alpha$, $p(x'|x)^\alpha$ of the distribution of locations $x'$ where a cell which is truly located at $x$ might "believe" itself to be,

$$p(x'|x)^\alpha = p(x'|g_1, g_2, g_3, g_4)|_{g_1(x)^\alpha\, g_2(x)^\alpha\, g_3(x)^\alpha\, g_4(x)^\alpha} \tag{2.1}$$

The concentrations are evaluated at the point $x$ in the embryo indexed by $\alpha$. The closer this map is to the identity, $p(x'|x) = \delta(x' - x)$, where $\delta$ is the dirac delta function (we've taken position as a continuous variable), the more specific the map specified by the gap genes. The average over the distributions they computed for their dataset is shown in fig. 2.3. It is very close to being the identity distribution - each cell "knows its location" to within 1% of the length of the embryo - less than the spacing between nuclei.

Surprisingly, specifying the location of each cell with an error that is less than the spacing between cells is not, on its own, enough to correctly specify the *relative ordering* of cells along the length of the embryo. This is a serious shortcoming of the model, assuming that the body plan depends on the relative positions of each of the possible cell identities. Suppose we wish to take seriously the idea that the gap gene concentrations specify a unique *identity* for each cell, and that this collection of identities in space determines a body plan relevant in future stages of development. We might make the reasonable assumption that the *overall* body plan depends on the *relative* pattern of the identities of the cells - or, in a one dimensional case, the ordering. The region measured a region with approximately 58 to 59 nuclei. Throughout, we will take as an approximation that we have 60 nuclei (or cells). In a system with 60 cells, there are 60! orderings, only one of which produces the correct body plan.

Our question then becomes: suppose we know that each cell individually knows its position to within 1% of the length of the embryo. If cells infer their positions using data which fluctuates

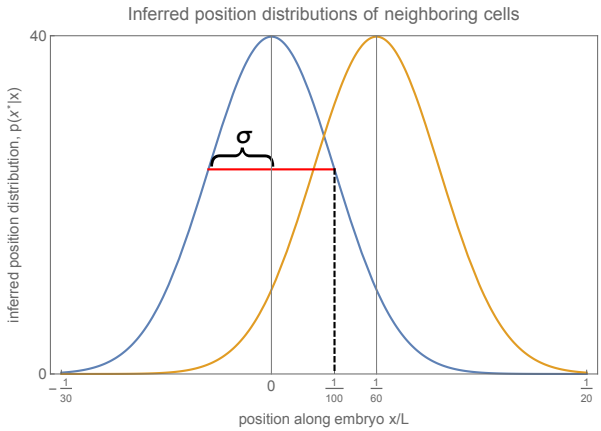



Figure 2.4: Here, the inferred positions of two neighboring cells in an embryo of length $L$ with 60 cells. The distributions are taken to be gaussian with standard deviation $\sigma = 0.01L$, as is observed in experiment. By assuming that the inferred position of each cell is independent, we find that the probability of the cells inferring their relative position incorrectly is around 12%, inconsistent with the reproducibility of embryo development seen in experiments.

independently at each point in space, what is the probability that they end up in the correct order overall? One might guess that, since the cells' inferences are well-localized around their true positions, they might infer their order correctly with very high probability, thus demonstrating that $\sigma \sim 0.01L$ is enough to produce a precise body plan, and not just individual positions. In fact, for this simple model of cells inferring their positions individually in the presence of spatially uncorrelated noise, this is not the case.

To see this, suppose that the cells' inferred positions are statistically independent. Denote $x_i$ to be the inferred position of the $i$th cell. If $p_A(x_1)$ denotes the distribution of positions for cell $A$, and $p_B(x_2)$ denotes the distribution of positions $x_2$ for cell $B$, we assume

$$p(A = x_1, B = x_2) = p_A(x_1)p_B(x_2)\,. \tag{2.2}$$

To model cells with $\sigma = 0.01L$, we take the length of the embryo $L = 1$, and model the distribution of each cell's inferred position as a gaussian around its mean $\mu_i = L/N = 1/60$,

$$p_i(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu_i)^2/2\sigma^2} \tag{2.3}$$

Suppose cells $A$ and $B$ are oriented as in fig. 2.4, and we want to know the probability that they infer their relative positions out of order. This occurs when the position of $A$ is greater than that of $B$, $x_A > x_B$. Consider the random variable equal to the difference in position between $B$ and $A$,

$\delta = x_B - x_A$. Its distribution is gaussian with mean $\sim L/N \sim 1/60$ and variance $2(0.01)^2$. We wish to compute the probability that $\delta < 0$:

$$\frac{1}{\sqrt{2\pi \cdot 2(0.01)^2}} \int_{\infty}^{0} dx\, e^{-(x-1/60)^2/2\cdot 2(0.01)^2} \simeq 0.119\ldots \tag{2.4}$$

For any two neighbors, there is a nearly 12% chance that they will infer their relative positions incorrectly. Considering we have an embryo with 60 cells, this seems at odds with the degree of reproducibility of development achieved in nature.

In fact, a model of 60 cells independently inferring their positions from concentrations uncorrelated in space with precision up to a scale $\sigma \sim 0.01L$ will never be able to encode the ordering of the cells, assuming their identities are distinct. This statement relies on a measurement of the positional information encoded by the spatial distributions of the gap genes. The claim is that the joint distribution of all four gap genes encodes only $\sim 4.3$ bits of positional information per nucleus, which is less than the 5.9 bits required to encode the order of 60 distinct cells; this was demonstrated in the paper [35].

To justify this claim, we must first define a quantity which we can use to rigorously measure this notion of "positional information" encoded by the gap genes. As was also done in [39], consider the distribution we have access to after taking measurements of concentrations along the anteroposterior axis of the embryo. Measurements consist of concentrations $g \equiv \{g_1, g_2, g_3, g_4\}$ as a function of $x$, and these can be combined across embryos to produce a distribution $p(g|x)$. We can also define the marginal distribution $p(x)$ over positions to be uniform across the embryo, $p(x) = 1/L$; this says that any given cell has a uniform prior about its location before making any measurements of the local concentrations. These two distributions allow us to define the joint distribution $p(g, x)$, and thus the marginal distribution $p(g)$.

The two quantities we have access to are position $x$ and concentrations $g$, and each carries information about the other. This is formalized in a quantity known as the mutual information, which uses the Shannon entropy to measure the amount of information one obtains about a variable upon knowing the value of another variable. We would like to know how much information about $x$, position, is defined upon knowing the values of the concentrations of the four gap genes proteins $\{g_1, g_2, g_3, g_4\}$ at the point $x$. This information can be written as

$$I_{g\to x} = S(X) - \langle S(X|g)\rangle_g \ , \tag{2.5}$$

where $S(X)$ is equal to the Shannon entropy of the distribution $p(x)$, and similarly for the second term, which is averaged over the distribution $p(g)$. Note that this quantity is an example of the mutual information defined in sec. 1.1.

To compute the mutual information from measurements, we need one more observation, since the above eq. 2.5 is expressed in terms of the distributions $p(x)$ and $p(x|g)$, but we measure $p(g|x)$. This is simple: Bayes' Theorem tells us that $p(x|g) = \frac{p(g|x)p(x)}{p(g)}$. Since we know everything on the right hand side from either measurements or assumptions, we can compute $p(x|g)$, and therefore compute $I_{g\to x}$ as above.

In practice, computing mutual information from data is a challenging task, given that one must properly model the probability distributions from data and other issues. Here, we will cite the result. It was found in [35] that the positional information held in the distribution of the four gap genes is equal $\sim 4.3$ bits. We have that $\log_2 60 \sim 5.9$ bits per nucleus are required to specify the relative positions of 60 nuclei, each with its own fate or identity. This demonstrates that the measured positional distributions, with inferences of positions treated as independent measurements, will not be able to encode the (one dimensional) body plan of the fruit fly at this stage. Are we missing biology (other morphogens, for example), or is our model lacking?

## 2.3 Spatially correlated noise reduces confusion

So far we have not modeled a mechanism by which different locations in space can have properties whose values are correlated. Intuitively, this makes it difficult to preserve information about the cells' relative positions, as seen in the previous example. In fact, such a mechanism is present in experiments: there are spatial correlations in the fluctuations of quantities which encode positional information at many different steps in the development process, such as the concentrations of the gap gene proteins [36].

The intuition that correlations may help preserve ordering can be made into a well-defined mathematical notion. We will consider two themes: first, that spatial correlations in noise increase the amount of information about relative positions which can be preserved, and second, that these correlations vastly decrease the probability of an error occuring.

We begin by comparing the information per nucleus transmitted when we have uncorrelated noise to the information per nucleus when a nontrivial covariance matrix is included.

Consider a model where positions within the embryo are modeled as real numbers $x \in (0, L)$. There are $N$ cells, each of which has a discrete position, such that the true positions of the cells are

a discrete set of real numbers $\{x_i\} \in (0, \ldots, L)$, $i = 1, \ldots, N$.

The gap gene concentrations encode each position $x_k$ to be "read out" by the nucleus within the cell located at $x_k$, as described previously, but the encoding is noisy, such that the value which is read out is not $x_k$, but a read-out value $y_k$ (this is the same inferred position we denoted as $x'$ previously; the notational change is for convenience). The value of $y_k$ is assumed to be given by $x_k$ plus an error term $\eta_k$. The value of $\eta_k$ is pulled from a gaussian distribution with mean 0 and standard deviation $\sigma$, and the noise at different points is taken to be independent. In equations, the noise distribution $p_N$ satisfies

$$p_N(\eta_1, \ldots, \eta_N) = \prod_{k=1}^{N} \frac{e^{-\eta_k^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \tag{2.6}$$

and the readout position $y_k$ is related to the true position $x_k$ via the relation

$$y_k = x_k + \eta_k \,. \tag{2.7}$$

Because of this relation, we can also write the noise as a conditional probability,

$$p(y_k|x_k) = \frac{e^{-(y_k - x_k)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \tag{2.8}$$

Again, independence of the noise allows us to write the conditional distribution for each position $k$ independently, as above, but this will no longer be true with correlated noise.

We wish to know how much information the gap gene encoding preserves about the input despite the fact that there is noise. We can measure this per cell by the mutual information between the possible input locations, which we denote as $X$, and the output space $Y$.

The mutual information $I(X;Y)$, can be written as $S(X) - S(X|Y)$, where $S$ is the Shannon entropy. If the input space $X$ consists of the possible input coordinates, it consists of the interval $(0, L)$. We make the assumption that the cell has no reason to favor any position over another before taking any measurements of morphogen concentrations. That is, the distribution over $X$ is uniform: $p(x) = 1/L$. [1] The entropy is thus

$$S(X) = \log_2 L \,. \tag{2.9}$$

[1] This uniformity assumption is essentially the same as the assumption we made in previous sections; it is reasonable from the point of view of cells with no information about their identity before taking "measurements" of local morphogen concentrations, but we will revisit it in later sections.

When the input prior is uniform, we can also easily determine $S(X|Y)$ using the conditional distributions for individual positions, $p(x_k|y_k) = \frac{p(x_k)p(y_k|x_k)}{p(y_k)}$. When $p(x_k)$ is uniform (for all $k$), $p(x_k|y_k) = p(y_k|x_k)$. To compute the mutual information recall that the differential entropy of a gaussian in one variable is equal to $\log \sqrt{2\pi e\sigma^2}$; that is,

$$-\int_{-\infty}^{\infty} d\eta \, \frac{e^{-\eta^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \log_2 \frac{e^{-\eta^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} = \log_2 \sqrt{2\pi e\sigma^2} \tag{2.10}$$

Thus

$$S(X|Y) = S(Y|X) = \log_2 \sqrt{2\pi e\sigma^2} \tag{2.11}$$

Taking the difference, we find the information transmitted per cell,

$$I(X;Y) = \log_2 L - \log_2 \sqrt{2\pi e\sigma^2}. \tag{2.12}$$

Observe that we can interpret this in terms of asking, how many cells can the system effectively distinguish within an interval of length $L$, despite the presence of noise? Whereas one might expect the answer to be $\sim L/\sigma$, above, we find that the entropy is in fact $\log_2 L/\sigma - \log_2 \sqrt{2\pi e}$. The "number of cells that can be distinguished" has entropy which is lower than expected by an amount $\log_2 \sqrt{2\pi e}$ – because of the gaussian noise, not only is the entropy divided by $\sigma$, but there is still "missing entropy" which decreases the effective number even further.

In the biological case, we can compute this information using data and compare to experiments. When $\sigma = 0.01L$, eq. 2.12 equals 4.6 bits – in particular, less than 5.9 bits, in agreement with the previous section.

One can instead consider the case where the noise is not uncorrelated, but has some correlation which falls off with distance. In this case, the $N$-dimensional *vector* $\vec{\eta}$ representing the noise over all $N$ positions is pulled from a *multivariate* distribution which depends on the correlation length and the specific functional form of the correlations.

Let the noise have a multivariate gaussian distribution with $N \times N$ covariance matrix $C \equiv \sigma^2 \overline{C}$ such that $\overline{C}$ is a matrix with all 1s on the diagonal. We can still compute the average mutual information $I(X;Y)$ per cell across the channel if we assume that the input prior over the space $(0, L)^N$ of positions of all $N$ cells is uniform. Recall that the differential entropy of a $N$-dimensional multivariate gaussian with nontrivial correlation matrix $C$ is given by $\log_2 \sqrt{(2\pi e)^N \det C}$. Using

an analogy to the previous calculation, and denoting the $N$-dimensional input and output spaces as $X_N$ and $Y_N$ respectively, we have

$$I(X;Y) = \frac{1}{N}(S(X_N) - S(X_N|Y_N)) \tag{2.13}$$

$$= \frac{1}{N}\left(N\log_2 L - \log_2\sqrt{(2\pi e)^N \det C}\right) \tag{2.14}$$

$$= \log_2\left(\frac{L}{(2\pi e)^{1/2}(\det C)^{1/2N}}\right) \tag{2.15}$$

$$= \log_2 L - \log_2\left(2\pi e(\det C)^{1/N}\right)^{1/2} \tag{2.16}$$

Observe that this is equivalent to the uncorrelated case, but with $\sigma^2 \to (\det C)^{1/N}$.

We can now compute the extra information transmitted by systems with spatially correlated noise as compared to systems with uncorrelated noise. Denoting by $I_0$ the information of an uncorrelated covariance matrix with covariance matrix equal to the identity times $\sigma^2$, $C_0 = \sigma^2\mathbb{I}$, we find that the *additional* information added by the correlations is given by

$$I_{extra} = I_c(X;Y) - I_0(X;Y) = -\frac{1}{N}\log_2\sqrt{\det\overline{C}} \tag{2.17}$$

Before doing any computations, we observe that $I_{extra}$ is always positive; i.e. $\det\overline{C} \leq 1$. This follows from the arithmetic mean-geometric mean (AM-GM) inequality, which states that for a list $(\alpha_1, \ldots, \alpha_N)$ of nonnegative numbers,

$$(\Pi_i\alpha_i)^{1/N} \leq \frac{1}{N}\left(\alpha_1 + \cdots + \alpha_N\right). \tag{2.18}$$

If $A$ is a matrix with eigenvalues $(\alpha_1, \ldots, \alpha_N)$, this is rephrased as

$$(\det A)^{1/N} \leq \frac{1}{N}(\mathrm{tr}A). \tag{2.19}$$

In our case, $\mathrm{tr}C_0 = \mathrm{tr}C = N\sigma^2$, and thus

$$(\det\overline{C})^{1/N} \leq \frac{1}{N}(N) = 1 \tag{2.20}$$

$$\det\overline{C} \leq 1 \tag{2.21}$$

necessarily. This just tells us that off-diagonal terms always serve to decrease the determinant when the diagonal values are fixed (and the eigenvalues are assumed positive); since correlations are just

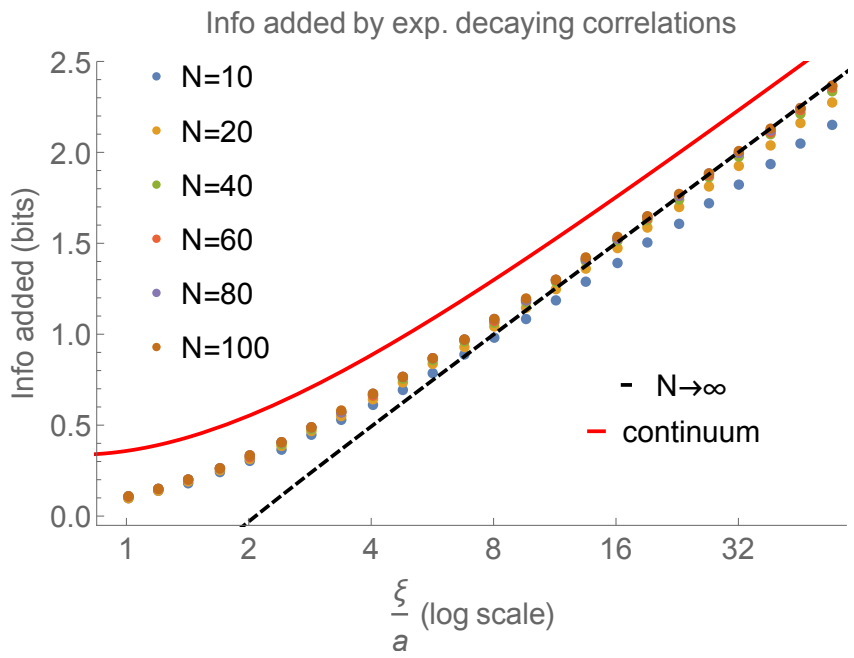



Figure 2.5: For systems with exponentially decaying correlations $C_{ij} = \exp(-|i-j|a/\xi)$, as the correlation length increases, so does the amount of additional information $-(1/2N)\log_2|C|$ each cell can know about its relative position. As $N \to \infty$, the information added approaches the black curve computed from the DTFT (eq. 2.27), but not the red curve computed from the FT.

off-diagonal terms in the correlation matrix, the result follows.

Not only can we prove that correlations always allow a system like the gap gene concentrations to carry more information, but we can compute precisely how this happens in models of physical systems. Suppose we have a 1-d lattice of $N$ points. We assign a random variable to each point such that the variables model noise pulled from multivariate gaussian distribution with correlation matrix with entries $C_{ij}$, where $i$ and $j$ index the $i$th and $j$th points respectively. We can compute the determinant of such a matrix (numerically) and therefore determine its associated $I_{extra}$ per cell, as in eq. 2.17. Such a computation is shown in fig. 2.5 for exponentially decaying correlations $C_{ij} = \exp(-|i-j|a/\xi)$. The plot shows $I_{extra}$ as a function of the correlation length $\xi/a$ on a log scale.

To carry out these computations, first, we directly compute the determinant of large correlation matrices $C_N$ with different correlation lengths to compare the amount of extra information as the correlation length increases.

In order to go from 4.6 bits of information per cell to 5.9 bits of information per cell, we need 1.3 bits of "extra" information. We compute that any system with correlation length of at least 12 lattice spacings will account for the necessary information, as one can see in fig. 2.5.

To complement the numerical results, we can do two different analytic calculations to model the system as $N$ becomes large. First, the approximation which preserves the discreteness of the system.

Let $C_N$ be the $N \times N$ correlation matrix of exponentially decreasing correlations with parameter $\xi/a$, the correlation length. We wish to compute $\log\det C_N$ as $N \to \infty$. Recall that log det of a matrix is equal to its tr log; it follows that we are in a good position if we can compute the eigenvalues of $C_N$ in a large-$N$ limit.

Observe that we have translation invariance such that $C_{ij}$ depends only on $|i-j|$. We might then think to use a fourier transform to diagonalize $C_N$. More specifically, this form of translation invariance means that $C_N$ is a so-called "Toeplitz matrix". Toeplitz matrices are convenient because given a sequence of $N \times N$ Toeplitz matrices such that each matrix agrees with the previous matrix except on the (new) $N$th row and columns, if the matrix entries $(C_N)_{i,j}$ fall off sufficiently quickly with increasing $|i-j|$, then there exists a well-defined large-$N$ limit in which we can use a form of fourier transform to compute functions of the eigenvalues in this limit.

Namely, let $C_N$ be the correlation matrix for $N$ cells on a lattice with lattice spacing $a$ and exponentially decaying noise with correlation length $\xi$,

$$C(|\, i-j\,|) = e^{-|\, i-j\,|a/\xi}, \tag{2.22}$$

Define $f_{DTFT}(\lambda)$ to be the so-called "discrete time fourier transform" (a technical term, although there is *no time dependence here*) [2]

$$f_{DTFT}(\lambda) = \sum_{n=-\infty}^{\infty} C(n)e^{in\lambda} \tag{2.23}$$

$$= \sum_{n=-\infty}^{\infty} e^{-|n|a/\xi}e^{in\lambda} \tag{2.24}$$

$$= -\frac{\sinh\left(\frac{a}{\xi}\right)}{\cos(\lambda) - \cosh\left(\frac{a}{\xi}\right)} \tag{2.25}$$

Taking $N$ progressively larger amounts to constructing a sequence of Toeplitz matrices which fall off exponentially as we go away from the diagonal. For sequences of Toeplitz matrices, the Szego theorem states that, if $\tau_{N,k}$ is the $k$th eigenvalue of $C_N$ and $g$ is a function which is continuous on

[2]The "discrete time fourier transform" differs from the "discrete fourier transform" because it is an *infinite* sum of samples resulting in a *continuous* function and *not* a finite series resulting in another finite series of the same length.

the image of $f(\lambda)$, we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}g(\tau_{N,k})=\frac{1}{2\pi}\int_0^{2\pi}d\lambda\; g\left(f_{DTFT}(\lambda)\right) \tag{2.26}$$

We'd like to compute the limit $\frac{1}{N}\log\det$ of $C_N$ as $N\to\infty$ in order to compute the information added, which is $-0.5$ times that quantity. Thankfully, as we observed previously, $\log\det=\mathrm{tr}\log$, and we can use Szego's theorem to compute the information in the infinite limit.

Taking $g(\cdot)$ to be $\log_2(\cdot)$, it follows that as long as $f_{DTFT}(\lambda)>0$ and $\tau_{k,N}>0$ the quantity $\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}\log C_N$ equals the integral of the log of $f_{DTFT}(\lambda)$; that is,

$$\frac{1}{2\pi}\int_0^{2\pi}d\lambda\;\log_2 f_{DTFT}(\lambda)=\frac{1}{2\pi}\int_0^{2\pi}d\lambda\;\log_2\left(-\frac{\sinh\left(\frac{a}{\xi}\right)}{\cos(\lambda)-\cosh\left(\frac{a}{\xi}\right)}\right) \tag{2.27}$$

$$=-\frac{a}{\ln(2)\xi}+2\log_2\left(e^{a/\xi}-1\right)+\log_2\left(\coth\left(\frac{a}{2\xi}\right)\right) \tag{2.28}$$

Here, $\ln(\cdot)$ refers to the natural logarithm. This is the limiting value of $\frac{1}{N}\mathrm{tr}\log C_N$ as $N$ approaches infinity, as a function of $\xi$, $-2$ times the $I_{extra}$ labeled as $N\to\infty$ in fig. 2.5.

A slightly different intuition for computing the eigenvalues of $C$ as $N$ (and, perhaps, $\xi$) become large would be to take a *continuum* limit and use the continuous fourier transform of the function $C(x)\propto e^{-ax/\xi}$. This intuition is especially natural for a physicist. Interestingly, a continuum limit does *not* limit toward the discrete result for *any* $N$ or $\xi$. This is demonstrated in red in fig. 2.5. The difference between the two approaches a constant value. This is somewhat surprising, so let's explore why it is that a continuum limit cannot reproduce the entropy of this system.

The DTFT is the fourier transform of an infinite series of equally-spaced (discrete) samples of some continuous function $F(x)$ which is not necessarily bandwidth-limited. The *result* of such a fourier transform is a *continuous, periodic* sum of the (continuous) fourier transform $\tilde{F}(k)$ of $F(x)$. If $F(x)$ is bandwidth-limited, it is possible to have exact reconstruction of $\tilde{F}(k)$ from one period of the periodic sum depending on the sampling rate, and therefore reconstruct $F(x)$, but this is not the case if $F(x)$ is not bandwidth-limited, as is the case here.

In our case, we can consider the discrete correlations mathematically as sampled points of the exponential function $(\xi/a)e^{-|x|}$ (the $\xi/a$ is useful for defining the sampling frequency properly). The samples are are taken at a rate $a/\xi$, and thus become more closely spaced as $\xi$ increases. In

equations, if $a$ is the lattice spacing and $\xi$ is the correlation length,

$$C(x) = \frac{\xi}{a} e^{-|x|} \tag{2.29}$$

$$f_{FT}(k) = \frac{2\xi}{a(1+(2\pi k)^2)} \tag{2.30}$$

$$C_n = e^{-a|n|/\xi} \tag{2.31}$$

$$f_{DTFT}(k = \lambda/2\pi) = \sum_{n=-\infty}^{\infty} C_n e^{2\pi i k} \tag{2.32}$$

$$\stackrel{\text{poisson}}{=} \sum_{n=\infty}^{\infty} f_{FT}(k - n\xi/a) \tag{2.33}$$

$$= \frac{\xi}{a} \sum_{n=-\infty}^{\infty} \frac{2}{1 + 4\pi^2 \left(k - n\xi/a\right)^2} \tag{2.34}$$

$$= -\frac{\sinh\left(\frac{a}{\xi}\right)}{\cos(2\pi k) - \cosh\left(\frac{a}{\xi}\right)} \tag{2.35}$$

We've denoted the (continuous) FT of $C(x)$ as $f_{FT}(k)$ (where $k$ is a linear frequency). We've gone to linear frequency space from angular frequency space for conventional reasons. One uses Poisson resummation to write the DFT in terms of the FT; this defines one of the most useful properties of the DTFT. In any case, this expression helps to clarify when the DTFT is a sometimes-good approximation of the continuous FT, and vice versa.

In analogy to eqs. 2.27, we can compute the entropy from the continuous fourier transform,

$$-\frac{a}{2\xi} \int_{-\xi/2a}^{\xi/2a} dk \, \log_2\left(\frac{2\xi}{a(4\pi^2 k^2 + 1)}\right) = -\frac{1}{2}\log_2\left(\frac{2\xi}{a(\pi^2(\xi/a)^2 + 1)}\right) + \frac{\tan^{-1}(\pi\xi/a)}{\pi(\xi/a)\ln(2)} - \frac{1}{\ln(2)} \tag{2.36}$$

The entropy as computed from the DTFT (eq. 2.27) and the FT (eq. 2.36) are plotted along with the entropy for finite $N$ in fig. 2.5; notably, the finite-$N$ results do not asymptote to eq. 2.36.

The FT is just the $n = 0$ mode of the periodic sum which defines the DTFT, so we can look at how the contributions of the $n \neq 0$ modes behave as a function of $\xi$ (for example) to understand if the $N \to \infty$ limit from the Toeplitz matrix property is equivalent to the continuous limit. We can look at both if the two *functions* $f_{FT}$ and $f_{DTFT}$ approach each other, and then look at if the *entropies* get close (or why they do not).

First, let's look at the difference between $f_{DTFT}(k)$ and $f_{FT}(k)$ (continuing to work in linear

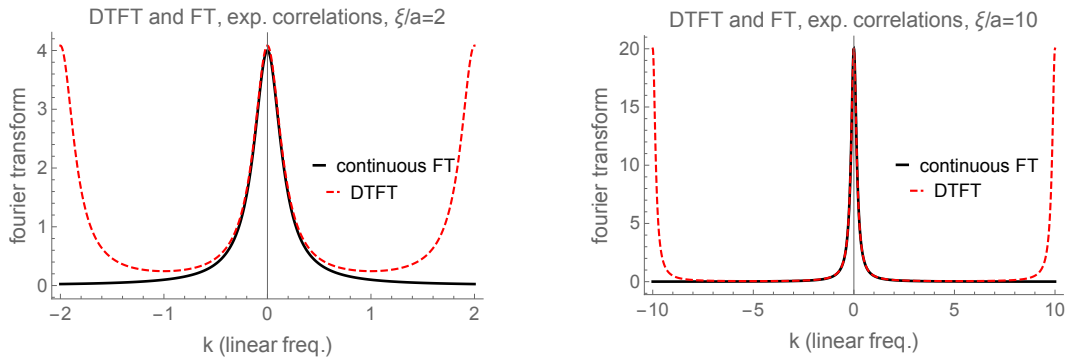



Figure 2.6: The discrete time fourier transform, which allows us to compute the information on the lattice, is a periodic sum of the continuous fourier transform, as depicted here. Because of aliasing, for small $\xi/a$, the two are significantly different even on the interval $(-\xi/2a, \xi/2a)$, but they quickly approach each other so as to be nearly indistinguishable on $(-\xi/2a, \xi/2a)$ for higher $\xi/a$.

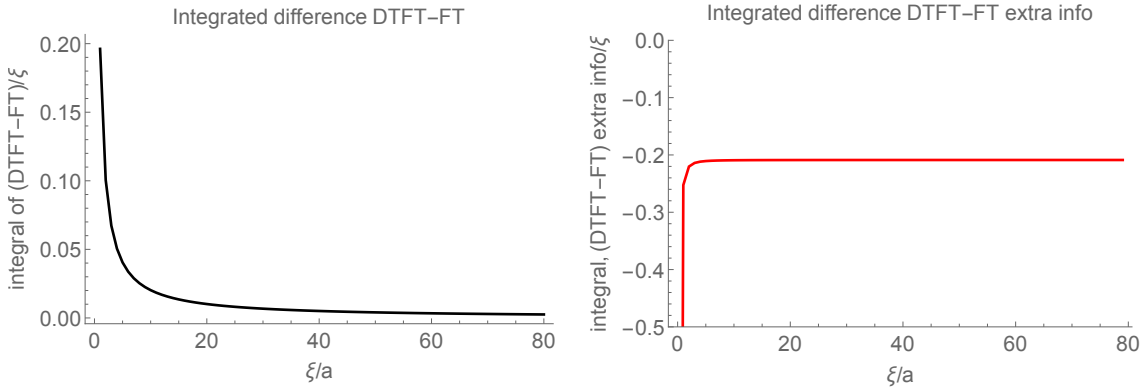



Figure 2.7: On the left, we see that the integrated difference $\frac{1}{\xi}\int_{-\xi/2a}^{\xi/2a}(f_{DTFT}(k) - f_{FT}(k))\,dk$ between the discrete and continuous fourier transforms decreases rapidly and continues to decrease toward 0 as $\xi/a$ increases. To the contrary, on the right, we see that the integrated difference between the discrete and continuous extra information, $\frac{1}{2\xi}\int_{-\xi/2a}^{\xi/2a}(\log_2 f_{DTFT}(k) - \log_2 f_{FT}(k))\,dk$ quickly approaches a nonzero limiting value of $\approx -0.2088\ldots$.
This entropy difference is the same as the nonzero gap we see between the black and the red curve in fig. 2.5.

frequencies),

$$f_{DTFT}(k) - f_{FT}(k) = -\frac{\sinh\left(\frac{a}{\xi}\right)}{\cos(2\pi k) - \cosh\left(\frac{a}{\xi}\right)} - \frac{2\xi}{a(1 + 4\pi^2\left(k - n\xi/a\right)^2)} \tag{2.37}$$

This difference is just equal to the contributions of the nonzero modes in the periodic sum of the continuous fourier transform which defines the DTFT. To measure how close the DTFT and the FT are in the relevant interval of $(-\xi/2a, \xi/2a)$, we can integrate this difference, which is always nonnegative, in the interval, and compare its magnitude as we change the correlation length $\xi/a$. In fig. 2.6, we have plotted the FT and DTFT for two values of $\xi$ to show the difference between a case where the non-zero modes contribute strongly, and one where they do not. (The case where the two differ strongly exemplifies what is known as aliasing, the contribution of higher modes, which can cause a problem for reconstructing the original continuous function from the samples.) We have also plotted the integrated value of the difference, eq. 2.37, divided by the correlation length $\xi$, in fig. 2.7, which shows that the two functions become very close and continue to approach each other as the correlation length becomes longer.

However, the integrated information tells something of a different story, as shown in fig. 2.7, where we've plotted the quantity

$$\Delta I_{extra} = \frac{a}{2\xi}\int_{-\xi/2a}^{\xi/2a} \log_2\left(-\frac{\sinh\left(\frac{a}{\xi}\right)}{\cos(2\pi k) - \cosh\left(\frac{a}{\xi}\right)}\right) - \log_2\left(\frac{2}{1 + 4\pi^2\left(k - n\xi/a\right)^2}\right) dk \tag{2.38}$$

as a function of $\xi/a$. Here, we find that the contributions from the higher modes do *not* approach zero, as they are within the log. The integrated difference approaches a constant, positive value. This difference accounts for the gap between the $N \to \infty$ curve and the "continuum" curve we plotted earlier. The continuum, "$n = 0$" approximation to the entropy is not a good approximation, and the periodic sum definition of the DTFT allows us to interpret why this is the case. It is even possible to write the limit of this difference analytically;

$$\lim_{\xi/a \to \infty} \Delta I_{extra} = -\frac{2(\log(\pi) - 1)}{\log(4)} \simeq -0.208801\ldots \tag{2.39}$$

To summarize, we've learned that if we take into account spatial correlations in the fluctuations of inferred positions (or gap gene concentrations directly), we obtain additional information which, in a system with exponentially decaying correlations, may provide the missing 1.3 bits of positional

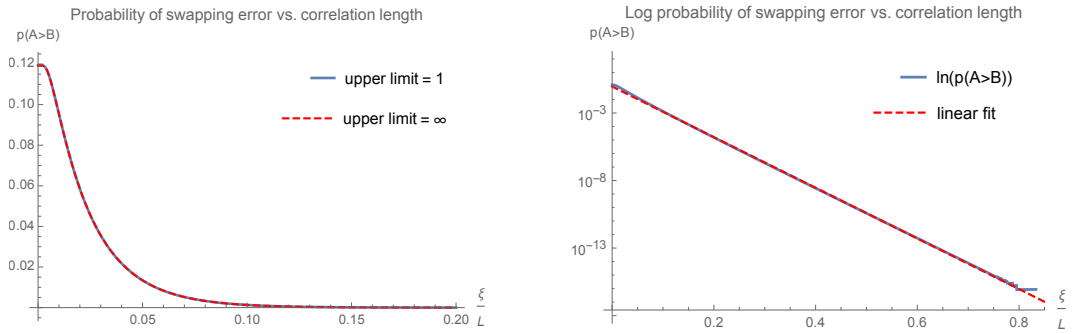



Figure 2.8: We compute the probability that the inferred position of cell $A$ is greater than the inferred position of cell $B$, $p(A > B)$, representing a "swapping error". Letting $N = 60$, we take each cell's position to be a gaussian with fixed variance $\sigma$ equal to $(0.01L)^2$, with the mean of $A$ centered at 0 and $B$ centered at $L/N$, and normalized covariance $v = \frac{\langle x_A x_B \rangle}{\sigma^2} = \exp(-\frac{1}{N(\xi/L)})$. We plot $p_{swap}(\xi) = \frac{1}{\sqrt{2\pi \cdot 2\sigma^2(1-v)}} \int_0^1 dx \exp(-\frac{(x+1/N)^2}{4\sigma^2(1-v)})$ since the difference in locations cannot be greater than $L$, and compare to the integral from 0 to infinity to show the two are effectively equal. Moreover, a log plot demonstrates that the decrease in probability is nearly exactly exponential as a function of $\xi/L$.

information observed when we assumed uncorrelated fluctuations, *provided* that $\xi/a \geq 12$. We can compare this to the results of [36], which looked at correlations in the gap gene expression levels. That paper found correlations in fluctuations of expression levels on the order of 20% of the length of the embryo up to the order of the scale of the whole length of the embryo. With $N = 60$ cells, $\xi/a = 12 \implies \xi \sim 20\%$ of the length of the embryo. This matches the correlations that were found among the gap gene protein concentration correlations. Perhaps the gap gene concentrations are transmitting just enough information to specify the identity of each cell, as a result of the spatial range of the correlations among their fluctuations? It is a highly interesting idea to pursue further with experimental data.

We've also learned that although there is a well-defined infinite discrete limit defined by the DTFT of the correlations, taking instead a continuous FT as a continuum limit for large-$N$, large-$\xi$ does not compute the eigenvalues of the discrete system accurately enough to compute the entropy. The DTFT calculation teaches us that the difference between the FT and the discrete-but-infinite limit is the contribution of higher modes in the periodic sum that defines the DTFT; whether or not we can use the FT to compute a discrete limit of a given quantity on the lattice will depend on how strongly those modes contribute to the calculation. This could have consequences for other computations one might want to perform using this common approximation.

We'd like to show that not only do spatial correlations increase information the cell has access to, but they allow lower probabilities of error (a more "operational" question). First, let's consider two

cells, $A$ and $B$, whose inferred position distributions are modeled as one-dimensional gaussians with standard deviations $\sigma_A = \sigma_B \equiv \sigma = 0.01L$ and means $x_A = 0$ and $x_B = 1/N = 1/60$. This is the same setting we encountered in sec. 2.2, but now, we will take the two particles' inferred positions to have a nontrivial joint distribution, rather than taking them to be independent. We define their covariance to be $\langle x_A x_B \rangle = (\sigma L)^2 \exp(-1/(N\xi))$, where $\xi$ is a correlation length which we will vary to obtain different correlation strengths. We assume an embryo with 60 cells, length normalized to $L = 1$, and standard deviation $0.01L \to 0.01$. We assume the joint distribution is a multivariate gaussian.

The probability that the two cells have inferred their relative positions incorrectly is equal to $p(x_A - x_B > 0)$. Define $y = x_A - x_B$. The variable $y$ is gaussian distributed with mean $0 - \langle x_B \rangle = -\frac{1}{N}$ and variance $2\sigma^2(1 - e^{-1/(N\xi)})$. This allows us to compute $p(x_A - x_B > 0)$, aka $p(\text{swap})$:

$$p(\text{swap}; \sigma, \xi, N) = \int_0^1 dy \, \frac{e^{-\frac{(y+\frac{1}{N})^2}{4(\sigma)^2(1-e^{-1/N\xi})}}}{\sqrt{2\pi \cdot (2\sigma^2(1 - e^{-1/N\xi}))}} \tag{2.40}$$

(Since $L = 1$, the upper limit is technically 1, but this is effectively the same as $\infty$ for such a small variance, as we show in fig. 2.8.)

It is instructive to plot $p(\text{swap})$ as a function of the correlation length $\xi$. This calculation is shown in fig. 2.8, which confirms that the probability of swapping effectively decays exponentially with increasing correlations in this regime. Now, not only can we say that correlations increase information; they also vastly decrease the probability of error in the simple case where two cells are inferring their positions in a correlated manner.

Evidently, fly embryos are composed of more than two cells, and thus we might want to know how errors "propagate" in a larger system with correlated errors among a greater number of cells. Do spatial correlations in noise prevent errors in larger systems? We can address this question with a simulation of the process wherein cells must infer their ordering from a signal which is a sum of the true signal and random gaussian noise which is correlated among sites. The computation is as follows. First, we fix $N$ (to be 40, 60, 80 or 100), $\sigma/a$ and $\xi/a$, $a$ being the lattice spacing. Then, we generate noise $\eta$ through the following recursion:

$$A = e^{-a/\xi} \tag{2.41}$$

$$\eta_{i+1} = A\eta_i + r \cdot (\sigma/a)\sqrt{1 - A^2}\,, \tag{2.42}$$

where $r$ is a number pulled from a gaussian distribution of zero mean and unit standard deviation.

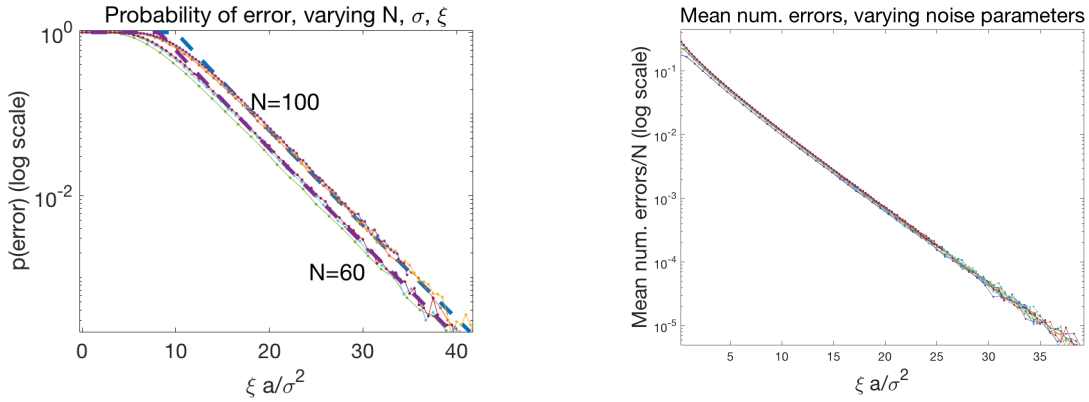



Figure 2.9: Probability that there exist neighboring cells out of order, and mean number of neighboring cells out of order divided by $N$, as a function of $\xi a/\sigma^2$ for values of $\sigma$ from $0.8/a$ to $1.4/a$, and values of $N = 60$ and 100 on the left, $40, 60, 80, 100$ on the right. The plot with probability of errors contains many thin lines consisting of data for different values of $\sigma$, as well as two thick dashed lines. These were computed by taking the independent-pair approximation discussed below, $Np = N \cdot \frac{1}{2}\mathrm{erfc}(\frac{1}{\sqrt{2\cdot 2\sigma^2(1-e^{-a/\xi})}})$ and substituting in either $N = 100$ or $N = 60$, along with the fixed value of $\sigma = 0.8a$. In the mean errors, the curves essentially completely collapse onto a single line on a log scale (although there is some residual $\sigma$ dependence).

The resulting noise satisfies $\langle\eta\rangle = 0$, $\langle\eta^2\rangle = \sigma^2/a^2$ and $\langle\eta_i\eta_j\rangle = (\sigma^2/a^2)e^{-|i-j|a/\xi}$. We let the vector $(1, 2, \ldots, N)$ represent the "input vector" of correct, discrete cell identities, compute the value of $\vec{x} = (1, 2, \ldots, N) + \vec{\eta}$, and determine how many neighbors are "out of order"; that is, satisfy $x_i > x_{i+1}$. This is defined as the "number of errors." We carry out this procedure $8 \times 10^4$ times, and determine the probability of having an error as well as the mean number of errors for various values of $N$, $\xi$, and $\sigma$. The results of the simulation are shown in fig. 2.9.

For analytical control, we can compute an estimate of the probability of producing a sequence of $N$ numbers in the correct order by supposing that each pair decides which order to be in independently of the other pairs. As before, the probability for a given pair to be out of order is given by

$$p(\text{out of order}) = \frac{1}{\sqrt{2\pi \cdot 2(\sigma^2/a^2)(1-e^{-a/\xi})}} \int_0^\infty dy\, e^{-\frac{(\frac{y}{a}+1)^2}{2\cdot 2(\sigma^2/a^2)(1-e^{-a/\xi})}} \tag{2.43}$$

$$= \frac{1}{2}\,\mathrm{erfc}\left(\frac{1}{\sqrt{4\sigma^2(1-e^{-a/\xi})}}\right) \tag{2.44}$$

The probability that there exists a pair which is out of order is equal to $1 - (1 - p(\text{out of order}))^N$ which, in the limit as $N \to \infty$, becomes $1 - (1 - Np)$, i.e.

$$p(\text{at least one error}) \simeq Np(\text{out of order}) = \frac{N}{2}\,\mathrm{erfc}\left(\frac{1}{\sqrt{4\sigma^2(1-e^{-a/\xi})}}\right) \tag{2.45}$$

We have plotted this approximation along with the numerics described previously in fig. 2.9, with high levels of success. The "independent-pair" approximation is an upper bound to the probability of error (adding more correlations makes an error less likely) which is quite tight, especially if we choose not to expand the exponential (at the cost of introducing a small amount of explicit $\sigma$ dependence). Moreover, we find a collapse after plotting lines with many different parameters as a function of the combination $\xi a/\sigma^2$. This indicates that although we were able to improve our bound some by not expanding in $a/\xi$, generically the expansion is a good assumption for large $\xi$, and the dependence of the probability and number of errors on just this parameter is impressive.

In summary, in this section, we have learned that spatially correlated noise can dramatically increase the amount of information that the gap genes convey about position along the anteroposterior axis of the embryo. Moreover, this translates into probabilities, decreasing the probability of error and correspondingly the mean number of errors as correlations increase. Computing these quantities first for two cells and then for many, we find exponential improvements when considered as a function of $\frac{\xi a}{\sigma^2}$. Moreover, we learn that increases in the variance can be offset by proportional decreases in the correlation length; increasing the correlation length by a multiplicative factor amounts to decreasing the variance.

## 2.4 Discreteness emerges naturally

Thusfar we have been discussing conveyance and inference of positions and fates in the presence of noise, but we've been largely agnostic about the question of how cells, discrete entities, can effectively read out continuous signals to infer a discrete identity. Is continuous positional information relevant, or is only information about discrete fates important, and how does one translate to the other? Here, we observe that if cells are optimal decoders and thus use the optimal prior to maximize the information they obtain from the gap gene signal, then discreteness is a mathematical inevitability. However, we also find that the system does not sense the discreteness in the regime where $\sigma \sim 0.01L$; the difference between a delta function and a very thin gaussian is not necessarily physically relevant, at least in terms of the amount of information each cell has access to in both cases. This somewhat reconciles the seeming challenge of clarifying what information is discrete and what information is continuous in the biological context.

As before, we parametrize the system with true positions and positions implied by the gap genes, which we think of input and output to a noisy channel with gaussian noise. Our input space $X = [0, 1]$ consists of coordinates along the length of the embryo; inputs $x \in X$ are transmitted

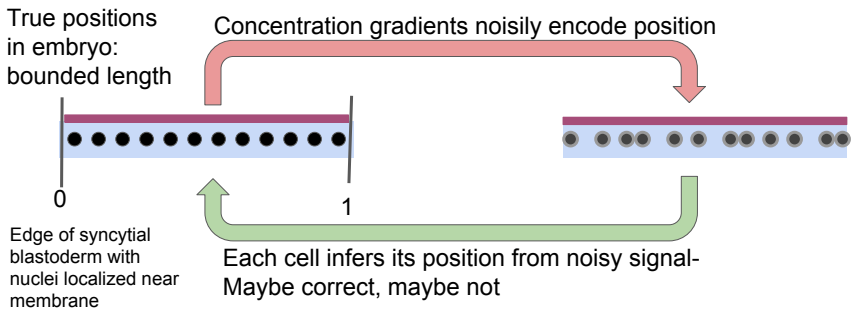



Figure 2.10: In the fly, true locations are bounded because the length of the embryo is finite. This implies the input prior must be discrete.

with noise $\eta$ pulled from a gaussian distribution $p(y|x)$ to produce an output $y = x + \eta$ in the output space $Y = \mathbb{R}$. The distribution of $\eta$ is derived from the fluctuations present in the gap gene concentrations. The cell uses a distribution $p(\text{input} = x^* |\ \text{output}= y)$ to determine its fate; this distribution is determined by the input prior $p_X(x)$ according to Bayes' theorem. The effective prior being used the biological system is unknown (and unmeasurable); in computing the decoder probability distributions previously, we assumed it was uniform.

Another possible choice for $p_X(x)$ derives from the assumption that embryos utilize the information from the gap genes in a way which maximizes the positional information; that is, they are optimal decoders. This assumption has some interesting physical consequences, if true.

The optimal prior $p_X(x)$ maximizes the mutual information between output and input variables [1],

$$I(X,Y) = \int dx\, p_X(x) \int dy\, p(y|x) \log \frac{p(y|x)}{p_Y(y)} \tag{2.46}$$

$$C = \max_{p_X(x)} I(X,Y) \tag{2.47}$$

In eqn. 2.47, the capacity $C$ is the maximum mutual information the channel can support between input and output given the noise in the transmission. By Shannon's noisy coding theorem [1, 2, 49], this is also equal to the maximum bits per symbol (rate) the channel can possibly transmit, error-free. The distribution $p_X(x) \equiv p_{opt}(x)$ which maximizes the mutual information is known as the *optimal input distribution* or the *optimal prior*.

It has been shown that fly embryos really do seem to*use* the positional information encoded in the spatial distributions of patterning genes early on in development [39]. This motivates the idea of studying optimal decoders in the context of development. Importantly, for channels with gaussian

noise, the optimal prior is not equal to the uniform distribution, which is the prior we have been using up until now. In the embryo, the input coordinates are bounded between 0 and 1 and the noise has finite standard deviation $\sigma$; in such circumstances, the optimal prior is not uniform; rather, it is nonzero on a *finite (hence, discrete) set* of points in $[0, 1]$.

This is a mathematical theorem about gaussian communication channels with an input space of real numbers bounded in some interval which was established in the 1970s and has been used in numerous applications since [37, 38]. The proof relies on the fact that an analytic function with infinitely many zeros in a finite interval must be identically zero on the complex plane (the complex identity theorm). To prove the result, define the Kubleck-Leibler divergence $D_{KL}(p(y|x)||p(y))$,

$$D_{KL}(p(y|x)||p(y)) = \int dy\, p(y|x) \log \frac{p(y|x)}{p_Y(y)} \tag{2.48}$$

We can think of the KL divergence as a function of the input coordinate $x$. Computing the capacity requires maximizing the mutual information subject to the constraints $\int_0^1 p_X(x) = 1$ and $p_X(x) \geq 0$. Observe that

$$I(X;Y) = \int dx p_X(x) D_{KL}(x) \tag{2.49}$$

Carrying out the optimization subject to the constraints implies that $p_X(x)$ is optimal if and only if

- $p_X(x_1) > 0$ and $p_X(x_2) > 0 \implies D_{KL}(x_1) = D_{KL}(x_2) \equiv D_{\max}$
- otherwise, $p_X(x) = 0$ and $0 < D_{KL}(x) < D_{\max}$

That is, $D_{KL}(x)$ achieves its maximum at points where $p_X(x)$ is nonzero.

To see that the number of such points must be finite, consider the function $D_{KL}(x) - D_{\max}$. This function has a zero wherever $p_X(x) \neq 0$. It can also be extended to an analytic function on the complex plane, and therefore either has a finite number of zeros in a finite interval (such as $[0, 1]$), or is identitically equal to 0 over the complex plane (this follows from the Bolzano-Weierstrass theorem and the complex identity theorem). Since there must exist points where $p_X(x) > 0$, $D_{KL}(x)$ is not equal to zero everywhere, and therefore must only have a finite number of zeros. We conclude that if the input space of a gaussian channel in one dimension is a finite interval in $\mathbb{R}$, the optimal input distribution $p_{\text{opt}}(x)$ is nonzero on only a finite, discrete number of points (mass points). Points where $p_{\text{opt}}(x)$ is nonzero for noise with different values of $\sigma$ are plotted in fig. 2.11, as are the corresponding channel capacities.

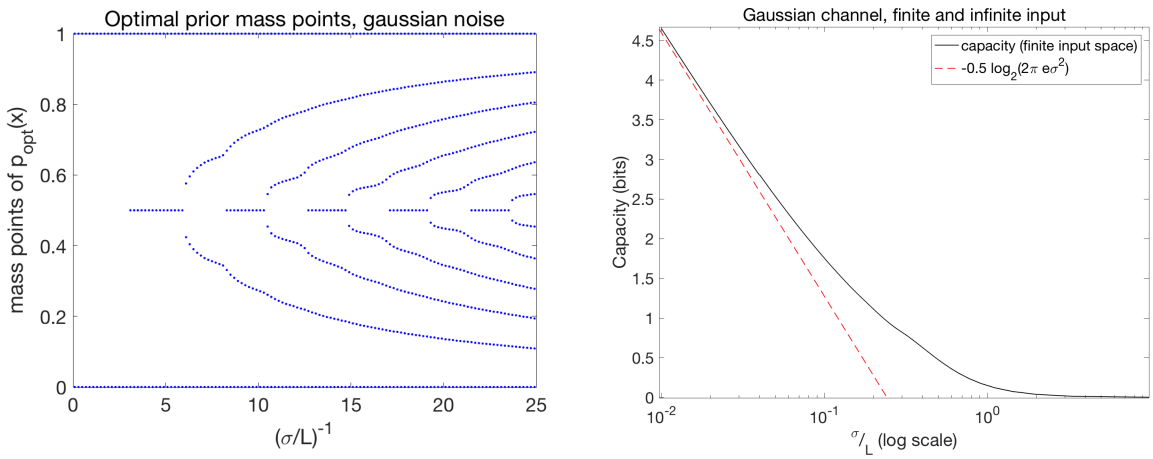



Figure 2.11: The left plot demonstrates the points at which the optimal distribution for a gaussian with inverse-standard deviation $(\sigma/L)^{-1}$ is nonzero (i.e. the "mass points" of the distribution)] The right hand plot demonstrates the channel capacities as a function of $\log(\sigma/L)$ for gaussian channels with bounded and unbounded input. The two approach each other as $\sigma/L$ becomes small, and agree to within $\sim 1\%$ for the physical value of $\sigma/L \sim 0.01$.

Biologically, the input space $[0, 1]$ represents the cell's coordinate along the length of the embryo. When a cell "reads" its position using the optimal prior, it will infer its position to be at one of a discrete set of choices closest to the output coordinate. Discreteness is not an assumption about the system - we have not imposed a lattice structure on $X$. Rather, it is an *output* which relies on the fact that positions are restricted to a finite region of space. Discrete fates and continuous positions coexist in the context of optimal decoders.

One might hope that discrete fates somehow automatically help the embryo encode a body plan more precisely than in the continuum system where gaussians have tails, but this is not the case. In fact, an optimal input prior supported on a discrete set of fates is barely distinguishable from a continuum system with very small error if the joint distribution contains no correlations. Consider the discrete case of transmitting positions from a lattice with $N$ points evenly spaced in the interval $[0, 1]$, a good analogy to the continuous case where $\sigma \ll 1$. As usual, the output is given by a lattice point plus a random sample from a gaussian distribution. Say that the cell infers the lattice point closest to the output position it received. The probability that a cell at lattice point indexed by 0 infers itself to be $i$ lattice spacings $a = 1/N$ away equals

$$p_i = \frac{1}{\mathcal{N}} \int_{x_i - 1/2N}^{x_i - 1/2N} dy\, e^{-\frac{y^2}{2\sigma^2}} \tag{2.50}$$

(In specifying the coordinates over the embryo, it is more symmetric to take coordinates on the embryo where $x_i \in (-1/2, 1/2)$ and distinguish between points on the left and the right.) The

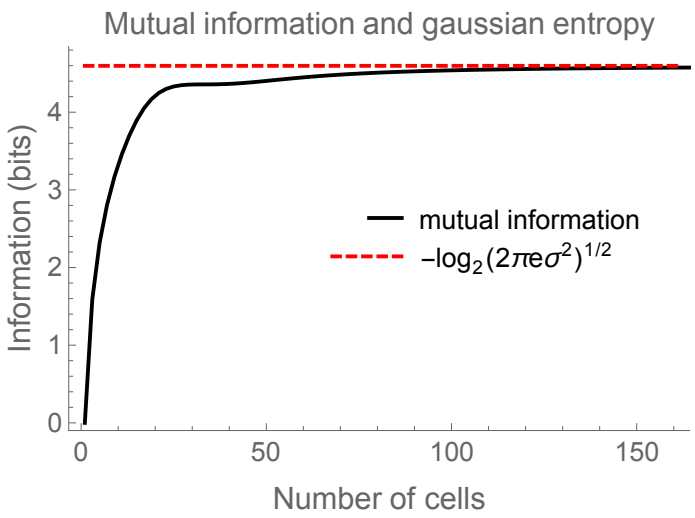



Figure 2.12: We define the discrete distribution in eq. 2.50, setting $\sigma = 0.01$. We then numerically plot the mutual information across the channel $\log N - \sum_i p_i \log p_i$ and compare to the entropy of the noise in the continuous sytem, which is $-\frac{1}{2}\log 2\pi e\sigma^2$ for small enough $\sigma$ (even in the case where the input is bounded, as we observed previously). We see that the mutual information in the discrete system converges to the continuous capacity quickly.

normalizing constant $\mathcal{N}$ normalizes the discrete distribution and approaches the continuous normalization as $\sigma \to 0$.

Can imposing discrete positions in the input allow the cell to transmit more information than inferring discrete positions as a result of optimizing? We expect that in the limit of many cells, the mutual information $S(X) - S(X|Y)$ will approach the information carried in the gaussian noise, which is $-\frac{1}{2}\log(2\pi e\sigma^2)$ when $\sigma \ll 1$. That is,

$$\log N - \sum_i p_i \log p_i \to -\frac{1}{2}\log(2\pi e\sigma^2) \tag{2.51}$$

$$p_i \log p_i \to \frac{1}{2}\log(2\pi e(\sigma N)^2) \tag{2.52}$$

We check: this is true just by approximating an integral with a sum using $p_i \simeq p(x_i)/N$,

$$-\sum_i \frac{1}{N}p(x_i)\log(\frac{1}{N}p(x_i)) \simeq \left[\int dx p(x)(\log(p(x))\right] - \log N \tag{2.53}$$

$$= -\frac{1}{2}\log(2\pi e(\sigma N)^2) \tag{2.54}$$

as needed. We conclude that with many degrees of freedom, the information of the discrete system approaches that of the continuum system, confirming our previous observation that discreteness is not too different from continuous-with-small-$\sigma$. We demonstrate this in fig. 2.4.

We thus have several calculations which point to similar conclusions: if the system is effectively inferring positions from gap gene proteins optimally, then discreteness is necessarily present in the system as an emergent fact rather than an input by-hand. However, the discreteness of the degrees of freedom does not, on its own, do much to explain the reproducibility in development of body plans in the presence of noise. Spatial correlations in gap gene concentrations are necessary in order to prevent cells from inferring their relative positions incorrectly.

Discrete models can have properties that continuous models lack, however. Next, we discuss how an intrinsically discrete (lattice) statistical mechanical system with spatial correlations might be able to model cells' inference of positions as an error-correcting code, depending on the statistical mechanical properties of the system.

## 2.5 Error-correcting codes from correlated, discrete systems

Thusfar, we've discussed spatial properties of noise (correlations) and properties of the underlying space itself (discreteness), but the two discussions have been largely separate. In particular, the discreteness of space seems unnecessary for precision in specifying the body plan. Correlations and discreteness actually go hand in hand when considered in the context of statistical mechanics and inference. Here, we describe a discrete gaussian model with correlated noise which could potentially specify an error correcting code for inference of a body plan.

We begin by reframing the inference problem at hand. Let $\vec{x} \equiv x$ denote the $N$-d vector of all input values to the channel (true positions), $\vec{y} \equiv y$ denote the $N$-d vector of all output values (readout positions), and $C$ denote the $N \times N$ covariance matrix of the noise. Consider the noise distribution

$$p(y|x) \propto e^{-(y-x)^t (C^{-1})(y-x)} \tag{2.55}$$

Since we have determined in sec. 2.4 that we are in a low-noise limit, we take the optimal prior to be discrete and approximately uniform, such that we have (schematically)

$$p(x|y) \propto \sum_{k=1}^{N} p(y|x)\delta(x - x_k) \tag{2.56}$$

(Here, each $x_k$ is an $N$-d vector representing one of the discrete points on which the input prior is nonzero.) We change variables to break down the error into an integer piece $h$ and a fluctuating

piece $\eta$:

$$y_i - x_i \equiv h_i - \eta_i \tag{2.57}$$

where $i$ indexes the spatial locations. We interpret the integer piece as the (integer, discrete) error made by the cell in inferring its (integer, discrete) identity, and the fluctuating $\eta$ to be an $N$-dimensional vector representing a random field pulled from a gaussian with the *same* correlation matrix $C$ as the original system. That is, the random field is correlated across different points in space.

Our $p(x|y)$ now looks like

$$p(x|y) = p(h;\eta) \propto e^{C_{ij}^{-1}h_ih_j + 2C_{ij}^{-1}h_i\eta_j} \equiv \frac{e^{-H_\eta}}{\mathcal{Z}} \tag{2.58}$$

We've defined a hamiltonian $H = C_{ij}^{-1}h_ih_j$ and its random-field-perturbed cousin $H_\eta = H + 2C_{ij}^{-1}h_i\eta_j$. Note that the integer error is correlated with the random field as well.

The fluctuations are thus pulled from a Boltzmann distribution with effective hamiltonian $H_\eta$, meaning that in the limit of large enough $N$ and strong enough correlations, they may be overwhelmingly likely to be in the ground state. If $\eta = 0$, for example, and $H$ is in an ordered phase, than the noise might be overwhelmingly likely to be in an all-0 state. That is, $h = 0$ and there is no error in the inference. Thus question number 1 we can ask: under what conditions does $H$ have an ordered phase which would guarantee zero error in the case of zero random field?

Turning on the random field $\eta$, if the ordered phase persisted, then the system would *remain* overwhelmingly likely to be in th $h = 0$ error-free state; however, in random field problems, whether or not the ordered phase structure is preserved can be a subtle question. Thus question number 2 becomes: Is the ordered phase preserved when $\eta \neq 0$?

If the answer to both of these questions is in the affirmative, then it would be accurate to describe the system as an *error-correcting code*: the properties of the covariance matrix $C$ between the fluctuations are such that the system has an ordered phase in which the indentification errors, $h$, are all vanishing. If, however, the answer is that such a situation *never* persists, then no amount of correlations is enough to guarantee an error-free encoding.

Inference with discrete degrees of freedom which are correlated across space has morphed into a statistical mechanics random field problem (the integer degrees of freedom make it a so-called discrete gaussian model). These are particularly rich questions because discrete gaussian models (with no

random field) are dual to a Coulomb gas. We thus expect the problem to be strongly dimension dependent. (For example, we expect no ordered phase in 1-d.) Moreover, in 2-d, Kosterlitz-Thouless-type transitions might play a role (indeed, these are related to so-called "roughening transitions" in discrete gaussian models known as solid-on-solid models). Does this problem relate to the generation of *topological* obstructions – in the gradient of the field, for example – which control the order?

The existence (or nonexistence) of an error-correcting code in this context is one of the most interesting questions raised by the work at hand.

## 2.6 Conclusions, looking forward

We began our story with the observation that although development of the fruit fly body plan is highly reproducible, even controlled at the precision of the single cell scale, the local values of the gap genes do not independently encode enough positional information to encode a unique identity for each of the $\sim 60$ cells along the length of the anteroposterior axis. This is true in spite of previous experimental evidence suggesting that cells develop unique, discrete identities which follow them throughout their entire lineages at approximately the time that the gap genes specify the positional information in the embryo.

In trying to determine where, mathematically, the discrete identities could be so reproducibly encoded, we found the language of mutual information, noise and optimality helpful. We discovered that the issue of discreteness is largely decoupled from that of reproducibility. Although discrete identitites emerge naturally for a finite system subject to noise which is optimally making use of the information available to it, this by no means guarantees that such a system will reproducibly infer *patterns* and relative degrees of freedom any better than a highly-localized-yet-continuous system.

Spatial correlations in the fluctuations of the gap gene concentrations (as characterized in [36]) provide the extra spatial information necessary cells at different spatial locations to infer their unique discrete identities correctly relative to one another even though each cell *independently* decodes its own identity. These correlations increase the positional information contained in the concentration profiles of the gap genes and vastly decrease the probability that neighboring cells will infer their relative locations incorrectly. This could therefore provide an information-theoretic model for the origins of reproducibility of a system with many degrees of freedom subject to noise and fluctuations.

Although fruit fly development has been the primary motivation for thinking about discreteness vs. continuity and reproducibility, the information theoretic understanding of discrete optimal input distributions, systems with spatial correlations, and the connection between noise and confusability

of neighbors apply broadly to systems governed by noisy spatial concentration profiles and systems where continuous signals end up producing discrete degrees of freedom. One reason information-theoretic models are useful is they are effective models which can be used even when "microscopic" molecular details like interactions or pathways are not well-understood; moreover, they generalize to different biological systems which share similar mathematical features.

That being said, there are now many opportunities for analyzing different kinds of data measuring spatial correlations, first in the spatial concentration profiles of the gap genes, and also in the concentrations of the pair rule gene products [50] and the fluctuations in the inferred positions of individual nuclei [50]. Indeed, these tools can help us analyze precisely how well relative positional information is specified by any spatial concentration profile.

One other interesting point is the observation that despite the nontrivial spatial dependence of the four gap genes and their noise profiles, they conspire to create a highly precise map of the embryo. This indicates a (delicate?) balance between noise and the functional form of the spatial dependence of the gap gene concentrations. Is this a general phenomenon? Are concentration profiles and the underlying noise evolved to be well-matched? How do correlations affect this matching? These are important questions to consider if we are to take the idea of optimal information encoding seriously, and they are approachable by data which currently exists.

By searching for positional information, we've thus discovered that correlations in the noise can carry the information about cell fates relative to one another, which is necessary to specify a body plan. Moreover, the phenomenon of unique, discrete cell identities may be less concerned with reproducibility than just required by the hypothesis of optimality. This leads to rich questions which connect the deep subject of information theory with fundamental questions about precision and differentiation in development using quantities which are experimentally accessible; namely, spatial correlations of measured protein concentrations. We expect this union of mathematical theory with biological applicability and experimental approachability to be a crucial component of a complete understanding of how the concentrations of many (or just a few!) biological molecules within an embryo can conspire to direct a dynamical process subject to noise which nonetheless reproduces, with very high precision, a fruit fly, capable of sight, smell, flight... and, of course, reproduction.

## Acknowledgements

We would like to acknowledge Thomas Gregor and Mariela Petkova for supplying data on which this work is based. We would also like to thank Farzan Beroz, Netta Haroush, Benjamin Machta, Henry

Mattingly, Andreas Mayer, Leenoy Meshulam, and Zachary Sethna for enlightening discussions.

# Chapter 3

# Rényi Entropy and Geometry

This chapter is adapted from the 2014 paper [51], written in equal collaboration with Jeongseog Lee and Benjamin R. Safdi.

In this chapter, we hypothesize a conjecture about the Rényi entropies of CFTs in $4d$. This research program contributes to the more general program of classifying quantum systems by universal aspects of their entanglement structure. Specifically, in an expansion in a short-distance cutoff, Rényi entropies of $4d$ CFT are believed to contain a universal term whose coefficient depends on only three theory-dependent functions of $q$, and geometric properties of the entangling surface. Although one of these three functions is notoriously hard to probe, in this paper we conjectured a relationship among the three which would make it possible to compute all three functions. Upon doing so, we also made contact with results on RG flows in free field theory by numerically confirming results related to the monotonicity of an entanglement-related quantity, $\mathcal{F}$, under RG flow.

For CFTs in even-dimensional spacetime, an expansion of the EE with respect to a short-distance cutoff $\epsilon$ in the field theory always contains a term proportional to $\log \epsilon$ as a result of the conformal anomaly. It is believed that this can always be written in terms of the conformal anomaly coefficients and the geometry of the entangling surface. For example, the entanglement entropy of an interval in 2d CFT takes the form

$$S_{EE} \sim \frac{c}{6} \log \epsilon \,. \tag{3.1}$$

In 4d, there is very strong evidence for Solodukhin's proposal that the entanglement entropy of

a region takes the form [52]

$$S_\Sigma = \alpha \frac{A_\Sigma}{\epsilon^2} + \left[ \frac{a}{180} \int_\Sigma E_2 + \frac{c}{240\pi} \int_\Sigma \left( \left( \mathrm{tr} k^2 - \frac{1}{2} k^2 \right) - C^{ab}{}_{ab} \right) \right] \log \epsilon + O(\epsilon^0) \,.$$

In this expression, $c$ and $a$ are the Weyl anomaly coefficients in 4d, normalized to equal one for the free scalar field; $\alpha$ is a nonuniversal coefficient; $A_\Sigma$ is the area of the entangling surface; $E_2$ is the 2d Euler density on $\Sigma$; $k^i_{ab}$ label the extrinsic curvatures, with $a, b$ local indices on $\Sigma$ and $i = 1, 2$ labels on the normal vectors; $C^{ab}{}_{ab}$ is the trace of the pullback of the spacetime Weyl tensor onto $\Sigma$. The take-away is that the coefficient of the $\log \epsilon$ term depends only on the coefficients $a$ and $c$, and integrals of local geometric quantities over the entangling surface.

The Rényi entropies, on the other hand, are not so well-understood, although they should possess similar expansions in terms of the cutoff. It is known that in 2d, the $q$th Rényi entropy of an interval in a CFT satisfies $S_q = \frac{c}{12}(1 + \frac{1}{q})$. The 4d analog of the $\log \epsilon$ term in the EE expansion contains new theory-dependent polynomials in $q$ which are not present in the case of entanglement entropy [53]. In 4d, the coefficient of the $\log \epsilon$ term is expected to take the form

$$S^q_\Sigma|_{\log \epsilon} = \left[ \frac{f_a(q)}{180} \int_\Sigma E_2 + \frac{f_b(q)}{240\pi} \int_\Sigma \left( \mathrm{tr} k^2 - \frac{1}{2} k^2 \right) - \frac{f_c(q)}{240\pi} \int_\Sigma C^{ab}{}_{ab} \right] \log \epsilon \,,$$

where $f_a(q), f_b(q)$ and $f_c(q)$ depend on the theory but not on the entangling surface.

In Section 3.1, we present a conjecture relating the structures $f_a(q), f_b(q)$ and $f_c(q)$ to each other. In the Section 3.2, we use a dimensional reduction to provide evidence for our conjecture by numerically computing Rényi entropies across a circle in free scalar and free fermion theories in $2 + 1$d. We also compute the CFT Rényi entropy and show that the numerical computations are consistent with analytical computations obtained by mapping to thermal free energy on $\mathbb{H}^2$. In Section 3.15, we compute numerical evidence for a cutoff-independent contribution to the area law in $2+1$d massive Rényi entropy. Appendices 3.A and 3.B contain details of the numerical computation.

## 3.1 Universal structure in Rényi entropy

As indicated above, in $3 + 1$d CFT, a $\log \epsilon$ term is present in the small-$\epsilon$ of the Rényi entropy. The coefficient of this term is parametrized by the geometry of the entangling surface and the conformal anomaly coefficients. For the entanglement entropy ($q = 1$), these are enough to determine the value of the coefficient, but for general $q$, there are three theory-dependent coefficients, $f_a(q), f_b(q)$ and

$f_c(q)$ below, which depend only on the Rényi parameter $q$ (and fully characterize the dependence on $q$). Recall the form of the coefficient of this expansion, where the various geometric quantities are consistent with the notation defined in the introduction:

$$S^q_\Sigma|_{\log\epsilon} = \left[\frac{f_a(q)}{180}\int_\Sigma E_2 + \frac{f_b(q)}{240\pi}\int_\Sigma \left(\mathrm{tr}k^2 - \frac{1}{2}k^2\right) - \frac{f_c(q)}{240\pi}\int_\Sigma C^{ab}{}_{ab}\right]\log\epsilon\,. \tag{3.2}$$

Computing the $q$th Rényi entropy therefore requires knowledge knowledge not merely of the entangling surace, but also of the coefficients $f_a(q), f_b(q), f_c(q)$. These are generally hard to compute. For a theory with $n_0$ massless scalars and $n_{1/2}$ massless fermions, $f_a(q)$ and $f_c(q)$ have been computed explicitly [54, 55, 56, 57, 58], with results

$$f_a(q) = n_0\frac{(1+q)(1+q^2)}{4q^3} + n_{1/2}\frac{(1+q)(7+37q^2)}{16q^3}\,, \tag{3.3}$$

$$f_c(q) = n_0\frac{(1+q)(1+q^2)}{4q^3} + n_{1/2}\frac{(1+q)(7+17q^2)}{16q^3}\,. \tag{3.4}$$

These can be probed directly by computing the RE across a spherical entangling surface ($f_a(q)$) and a cylindrical entangling surface ($f_c(q)$), since these single out different terms, geometrically.

The third term, $f_b(q)$, is notoriously difficult to compute, and, prior to this work, had never been probed in any theory. Here, we compute $f_b(q)$ numerically in free field theory, and observe that for both free scalar fields and free fermions, $f_b(q) = f_c(q)$. The robustness of this result across the spectrum of $q$s leads us to the following conjecture:

$$\textbf{Conjecture: } f_b(q) = f_c(q) \text{ in all } (3+1)\text{-dimensional CFTs.} \tag{3.5}$$

This conjecture, if true, would indicate that all three pieces of the universal $\log\epsilon$ coefficient in the CFT RE may be accessible through $f_a(q)$ and $f_c(q)$, for example via computations of RE across the sphere and the cylinder.

Our numerical computations of the RE in $3+1$d rely on dimensional reduction from $3+1$d to $2+1$d. Namely, consider an entangling surface $S^1\times S^1$ in a time slice of a $3+1$d spacetime. Take the radius $R$ of the first $S^1$ to be much less than the radius $L$ of the second $S^1$, $R \ll L$. Let our theory be that of a free massless scalar. Then we can fourier transform the field into modes according to their angular momentum in the $L$ direction,

$$\phi_n(x,y,z,t) = \phi(y,z,t)e^{2\pi ikx/L} \tag{3.6}$$

with $x \sim x + R$.

Observe that $\Box_4 \phi_n = \left(\Box_3 - (2\pi k/L)^2\right) \phi_n$, such that $\phi_n$ behaves like a massive free field in one fewer dimension. Moreover, since the $\phi_n$ are decoupled, the $3+1$d RE decomposes into a sum of REs of $2+1$d massive free fields along a circular entangling surface of radius $R$.

The $2+1$d RE of a free massive scalar or Weyl fermion can be expanded in powers of $mR$:

$$S^q = \alpha^q \frac{R}{\epsilon} + \beta^q \, 2\pi(mR) - \gamma^q + \sum_{n=0}^{\infty} \frac{C^q_{-1-2n}}{(mR)^{2n+1}} \,. \tag{3.7}$$

The term $\gamma^q$ is expected to be related to the topological entanglement entropy, and $\beta^q$ is a cutoff-independent contribution to the perimeter law. We will discuss $\beta^q$ in more detail later. The $C_{-1-2n}$ in the free massive $2+1$d theory is related to the massless RE across a $T^{2n+1} \times S^1$ entangling surface in $2n+4$d. (The volume of the $T^{2n+1}$ should be large compared to $R^{2n+1}$, where $R$ is the radius of the $S^1$.) Here, $n = 0$, so we are interested in the $C_{-1}$ term, which is known to satisfy [59, 60, 61]

$$\left(C^q_{-1}\right)^{\text{scalar}} = -\frac{\pi}{240} f_b(q) \qquad \left(C^q_{-1}\right)^{\text{fermion}} = -\frac{\pi}{480} f_b(q) \,. \tag{3.8}$$

This follows from explicitly carrying out the dimensional reduction and comparing coefficients [59, 60, 61].

## 3.2 Numerical Rényi entropy

In this section, we compute $C^q_{-1}$ by directly computing $S^q$ across circular entangling surfaces in massive free field theories in $2+1$d, and infer $f_b(q)$ using eq. 3.8. To carry out the numerical computations of $S^q$, we use a technique proposed in [62] which is a straightforward generalization of Srednicki's technique for computing EE across a circle in $2+1$d free field theory [63, 64, 62, 59]. This technique has recently been used in numerically computing EE across circular entangling surfaces in $2+1$d massive free field theory [59, 10, 61, 65].

We compute $S^q$ in both the scalar and fermion theories, in which the method is nearly identical. First, the scalar theory. Working in polar coordinates and discretizing the radial direction $r$. For each $n$, we can compute the discrete hamiltonian $H_n$, and using $H_n$, we compute the contribution $S^q_n$ of the $n$th mode to the Rényi entropy,

$$S^q = S^q_0 + 2 \sum_{n=1}^{\infty} S^q_n$$

This proceeds essentially by a direct diagonalization. The fermion case is analogous, except half-integer modes replace integer modes in the above expression. The details of the numerical methods are found in Appendix 3.A.

One point is worth noting. While we must cut off at finite $n$, we know the large-$n$ behavior of $S_n^q$ for both cases:

$$\text{Scalar:} \quad S_n^q \sim \begin{cases} \frac{1}{n^{4q}} & q < 1 \\ \frac{1}{n^4} & q \geq 1 \end{cases}, \qquad \text{Fermion:} \quad S_n^q \sim \begin{cases} \frac{1}{n^{2q}} & q < 1 \\ \frac{1}{n^2} & q \geq 1 \,. \end{cases} \tag{3.9}$$

From these, we learn that our numerical methods break down for $q < \frac{1}{4}$ for the scalar theory and $q < \frac{1}{2}$ for the fermion theory since the sums are no longer convergent. We restrict ourselves to $q$ above these values for this reason.

On to our computations of $C_{-1}$, the crucial issue at hand. We compute $S^q$ by splitting into angular momentum modes as described above. Then, we wish to cut off the cutoff-dependent perimeter law term not of interest to us. We thus compute the following quantity,

$$\mathcal{F}^q(mR) = -S^q(mR) + R\,\partial_R S^q(mR) \tag{3.10}$$

which is cutoff-independent and contains no perimeter law, as well as no constant term for these theories. It also approaches the renormalized EE as $q \to 1$.

It is straightforward to extract $C_{-1}$ from the $1/(mR)$ expansion of $\mathcal{F}^q(mR)$, as can be seen from a comparison with the expansion of $S^q$, eq. 3.9. We use our numerical computations of $S^q(mR)$ to compute $\mathcal{F}^q(mR)$, then fit to the form

$$\mathcal{F}^q(mR) \sim \frac{\tilde{C}^q_{-1}}{mR} + \frac{\tilde{C}^q_{-3}}{(mR)^3} \tag{3.11}$$

(cutting off after the first two terms of the $1/(mR)$ expansion). We then use $\tilde{C^q_{-1}}$ to approximate $C^q_{-1}$, restoring constants as necessary. The results of this computation for the scalar and fermion are found in fig. 3.1. The numerics match the analytical predictions of our conjecture to within 3% across a wide range of $q$s.

Another quantity that is easily extracted from these numerics is $S^q \equiv \mathcal{F}^q(0)$ (which now contains a nonzero $mR$-independent term). Massless free theories are conformal. It is possible to compute $S^q$ analytically by performing a conformal mapping from the Rényi entropy across the circle to the

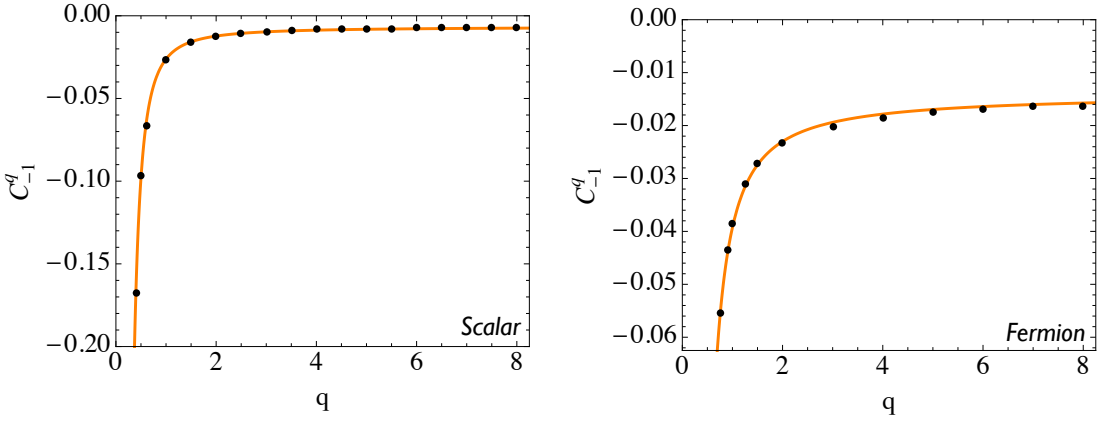



Figure 3.1: The coefficients $C^q_{-1}$ of the $1/(mR)$ term in the large $mR$ expansion of the Rényi entropy (see (3.7)) for the complex scalar (left) and Dirac fermion (right) theories. These coefficients are related to the $f_b(q)$ coefficients appearing in (3.2) in the $(3+1)$-dimensional Rényi entropy for the massless CFTs through (3.8). The orange curves show the predictions from our conjecture that $f_b(q) = f_c(q)$, with the $f_c(q)$ given in (3.3). The black points are the results of the numerical calculations. The numerical results agree with the analytic prediction to within 3% for the scalar and fermion theories for all $q$.

calculation of a thermal partition function on $\mathbb{H}^2$ [53, 58, 57]. Taking the radius of $\mathbb{H}^2$ to be $R$, the temperature is $1/(2\pi Rq)$. Defining $Z_q$ to be the Euclidean partition function on $S^1 \times \mathbb{H}^2$ (with the $S^1$ circumference equal to $2\pi Rq$), the Rényi entropy $S^q$ is written in terms of the free energy $F^q_{therm} = -\log|Z_q|$ as

$$S^q = \frac{qF^1_{\text{therm}} - \mathcal{F}^q_{\text{therm}}}{1-q} \tag{3.12}$$

Recent computations [58] on $S^1 \times \mathbb{H}^2$ imply that

$$\left(F^q_{\text{therm}}\right)^{\text{scalar}} = -\int_0^\infty d\lambda \tanh(\pi\sqrt{\lambda})\log(1-e^{-2\pi q\sqrt{\lambda}}) + q\frac{3\zeta(3)}{4\pi^2} \tag{3.13}$$

$$\left(F^q_{\text{therm}}\right)^{\text{Dirac}} = 2\int_0^\infty d\lambda\, \lambda \coth(\pi\lambda)\log(1+e^{-2\pi q\lambda}) + q\frac{\zeta(3)}{\pi^2} \tag{3.14}$$

These computations are highly nontrivial, and require regularizing the volume of $\mathbb{H}^2$ and the sum over eigenvalues, e.g. We numerically confirm these results by calculating $S^q$ over a wide range of $q$ and comparing to the analytical results in eqs. 3.13 and 3.14. The computations show excellent agreement for both the scalar and the fermion theory.

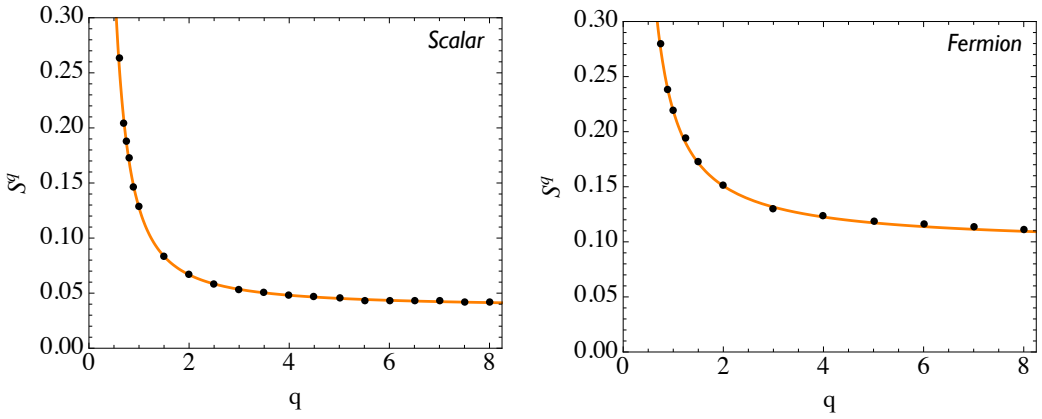



Figure 3.2: The massless Rényi entropies $S^q$ in the free complex scalar (left) and Dirac fermion (right) theories as functions of the Rényi parameter $q$. The orange curves are the analytic predictions coming from the mapping to $S^1 \times \mathbb{H}^2$ (see (3.13)). The black points are the results of the numerical computation. We find that the numerical results agree with the analytic predictions to within 2% for the scalar and fermion theories across all $q$.

## 3.3 Calculable contributions to the perimeter law

In the previous section, we computed the Rényi entropies $S^q(mR)$ and used the resulting quantity $\mathcal{F}^q(mR)$ to probe $C^q_{-1}$ and therefore $f_b(q)$ in the $3+1$d theory. Although $\mathcal{F}^q(mR)$ is a universal quantity which is sensitive to much of the interesting physics in $S^q(mR)$, it does not capture all we might want to learn about the RE. For example, essentially by construction, $\mathcal{F}^q(mR)$ can not probe the cutoff-independent piece of the perimeter law, as parametrized by $\beta^q$ in eq. 3.7. Here, we use our numerical computations to directly compute $\beta^q$ from $S^q(mR)$.

We are not the first to consider the possibility of a perimeter law in the entanglement and Rényi entropies. Reference [66] used the results of [67, 68, 69] to conjecture that $\beta^1 = -\frac{1}{2}$ for the massive real scalar and Dirac fermion theories. They assumed that for general entangling surfaces $\Sigma$, the $\beta^1$ term obeys a perimeter law,

$$S_\Sigma \supset \beta^1 \, m \, \ell_\Sigma \,, \tag{3.15}$$

with $\ell_\Sigma$ equal to the perimeter of $\Sigma$. Here, the authors calculated $\beta^1$ in a waveguide geometry, where the entanglement entropy can be computed explicitly using the heat kernel method.

Ref. [59] checked this prediction for $\beta^1$ using numerics in massive free field theory in flat spacetime and a circular entangling surface. This paper found perfect agreement with the previous study and thus provided evidence for the perimeter law scaling conjecture in eq. 3.15.

Moving away from $q = 1$, one might ask if a perimeter law scaling still holds for general Rényi

entropy, and if yes, can one calculate $\beta^q$ for certain entangling geometries. Reference [70] carried out such a computation in the waveguide geometry and found that

$$\beta^q = -\frac{1+q}{24q} \tag{3.16}$$

Note that the $-(1+\frac{1}{q})$ scaling resembles that of the $1+1$d RE. This is not an accident. The $\beta^q$ computation by heat kernel factors into an integral over $C_q \times S^1$, where $C_q$ is a 2d cone with deficit angle $2\pi(1-q)$. On the other hand, the $1+1$d Rényi entropy computation in the free theories may also be computed by an integral over the heat kernel on $C_q$. Thus it is not surprising that the two answers would be similar.

In this work, we numerically computed $\beta^q$ in flat spacetime with a circular entangling surface to see if 3.16 still holds. We did so by first fixing $q$, then directly computing $S^q(mR)$ for different values of $m$, and fitting these to the form

$$S^q = b_1^q(m)R + b_0^q(m) + b_{-1}^q(m)\frac{1}{R}\,. \tag{3.17}$$

In computing $\beta^q$, we are interested in the perimeter law $b_1(m)R$, but we must separate the universal $\beta^q$ from lattice-dependent contributions. In order to do so, we fit the set of $b_1^q(m)$ to the following form:

$$b_1^q(m) = \tilde{\alpha}_2^q\, m^2 + 2\pi\tilde{\beta}^q\, m + \tilde{\alpha}_0^q\,. \tag{3.18}$$

The term proportional to $m^2$ derives from finite lattice size, the $\tilde{\alpha}_0^q$ term derives from the nonuniversal short-distance cutoff term, and the term proportional to $m$ gives us the desired contribution.

Since the separation of $\beta^q$ is numerically sensitive, we improved our numerical precision by repeating the calculation using lattices of varying size, between $N = 200$ and $N = 350$. For each $q$, we fit the resulting $\tilde{\beta}^q$ data to $\tilde{\beta}_0^q + \tilde{\beta}_1^q/N + \tilde{\beta}_2^q/N^2$, and take $\tilde{\beta}_0^q$ as our approximation to $\beta^q$. Our numerical results are compared to the prediction eq. (3.16) in Fig. 3.3. We find that $\beta^1$ deviates from $-1/12$ by less than 0.1% in both the scalar and fermion theories. At large $q$ the deviation from (3.16) is $\sim$2%. We emphasize that the numerical results for $\beta^q$ are sensitive, at the few percent level, to the form of the fits in eqs. 3.17, 3.18, and the large-$N$ extrapolation. In particular, in Fig. 3.3 it may be seen that at large $q$ the fermion results are systematically above the scalar ones. This gap decreases with increasing lattice size $N$. While the large-$N$ extrapolation helps bring the

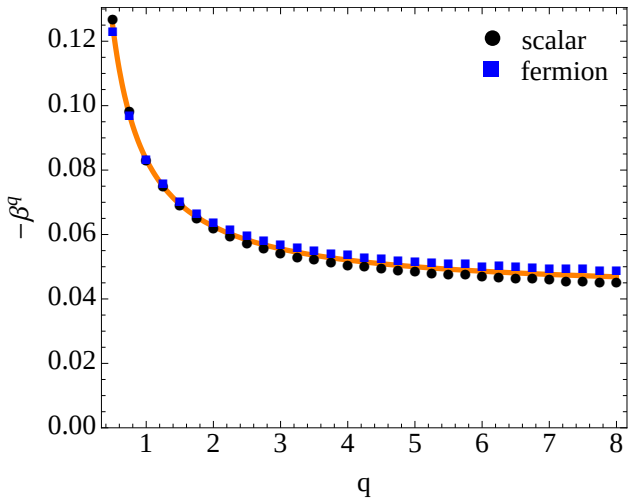



Figure 3.3: The coefficients $-\beta^q$ in the large-$mR$ expansion of the free-field Rényi entropy (3.7). An explicit computation of the $\beta^q$ in the wave-guide geometry [70] combined with the assumption that the $\beta^q$-term obeys the perimeter law, as in eq. (3.15), leads to the analytic prediction (3.16) (solid orange) for both the real scalar and Dirac fermion theories. Our numerical results for $\beta^1$ agree with (3.16) to better than 0.1%, while the $\beta^q$ at large $q$ agree with the analytic expression to within $\sim$2%.

two results into agreement, we are left with a small, residual error. Practical limitations prevent us from further increasing the lattice size.

The numerical procedures described in Sec. 3.2 were subject to less difficulty. One reason for this is that the calculations in Sec. 3.2 involved the renormalized entropies while the calculation of $\beta^q$ involves the Rényi entropies directly. The process of separating the non-universal, cutoff-dominated component of the area law from the universal component is non-trivial and introduces additional lattice sensitivity. A direct, analytic calculation of the $\beta^q$ for the circular entangling surface would be useful.

## 3.A The numerical technique

In this Appendix we give details of the Srednicki procedure for calculating Rényi entropy across a circular entailing surface in free 3-dimensional CFT [63, 64, 62, 59]. We begin with an account of the procedure for the scalar theory. For our purposes, the Hamiltonian is most conveniently expressed by expanding in modes of integer angular momentum $n$ and discretizing the radial direction into $N$ units. The method relies on the observation that the resulting Hamiltonian takes the form

$$H = \sum_n H_n\,, \qquad H_n = \frac{1}{2}\sum_i \pi_i^2 + \frac{1}{2}\sum_{ij}\phi_i K_n^{ij}\phi_j\,, \tag{3.19}$$

where $\pi_i$ is the conjugate momentum to $\phi_i$, and $i, j$ run from $1, \ldots, N$. The matrix $K_n^{ij}$ has non-zero elements [59]

$$K_n^{11} = \frac{3}{2} + n^2 + m^2\,, \qquad K_n^{ii} = 2 + \frac{n^2}{i^2} + m^2\,, \qquad K_n^{i,i+1} = K_n^{i+1,i} = -\frac{i+1/2}{\sqrt{i(i+1)}}\,. \tag{3.20}$$

For each $n$, the two-point correlators $X_{ij} = \langle \phi_i \phi_j \rangle$ and $P_{ij} = \langle \pi_i \pi_j \rangle$ are directly related to $K$,

$$X_n = \frac{1}{2}\left(K_n^{1/2}\right)\,, \qquad P_n = \frac{1}{2}\left(K_n^{-1/2}\right) \tag{3.21}$$

and are thus easily computed. We find the $q^{\text{th}}$ Rényi entropy across a circle of radius $R = r + \frac{1}{2}$ in lattice units by constructing the truncated matrices $X^r \equiv \left(X_{ij}\right)_{1\le i,j\le r}$ and $P^r \equiv \left(P_{ij}\right)_{1\le i,j\le r}$, for each $n$. The Rényi entropy is then given by (3.9) with

$$S_n^q = \frac{1}{1-q}\mathrm{tr}\log\left[\left(\sqrt{X_n^r P_n^r} + \frac{1}{2}\right)^q - \left(\sqrt{X_n^r P_n^r} - \frac{1}{2}\right)^q\right]\,. \tag{3.22}$$

The fermion computation is similar. We expand the fermion field in half-integer angular momentum modes, labeled by $n$, and write the Hamiltonian as $H = \sum_n H_n$. As before, the radial direction is discretized into $N$ units. For each $i = 1, \ldots, N$, we decompose the Dirac spinor $\psi_i$ as $\psi_i = (u_i, v_i)^T$. Each $H_n$ takes the form

$$H_n = \sum_{i,j=1}^{N} \begin{pmatrix} u_i^* & v_i^* \end{pmatrix} M_n^{i,j} \begin{pmatrix} u_j \\ v_j \end{pmatrix}\,, \tag{3.23}$$

where the $M_n^{ij}$ are known $(2\times 2)$ matrices with entries $\left(M_n^{ij}\right)_{\alpha\beta}$, $\alpha, \beta = 1, 2$. It is convenient to reorganize the data for each $n$ into a single $(2N \times 2N)$ matrix $\tilde{M}_n^{2k+\alpha-2,2l+\beta-2} \equiv \left(M_n^{kl}\right)_{\alpha\beta}$. The nonzero entries of the matrices $\tilde{M}_n$ are given by [59]

$$\begin{aligned}
&\tilde{M}_n^{kk} = (-1)^{k+1} m\,, && \tilde{M}_n^{1,2} = i\left(n + \tfrac{1}{2}\right)\,, && \tilde{M}_n^{2k-1,2k} = i\tfrac{n}{k}\,,\\
&\tilde{M}_n^{2k-1,2k+2} = -\tfrac{i}{2}\,, && \tilde{M}_n^{2k-1,2k-2} = \tfrac{i}{2}\,, && \tilde{M}_n^{2,1} = -i\left(n + \tfrac{1}{2}\right)\,,\\
&\tilde{M}_n^{2k,2k-1} = -i\tfrac{n}{k}\,, && \tilde{M}_n^{2k,2k-3} = \tfrac{i}{2}\,, && \tilde{M}_n^{2k,2k+1} = -\tfrac{i}{2}\,.
\end{aligned} \tag{3.24}$$

As in the scalar case, each $n$ is associated with a matrix of correlators $C_{ij} = \langle \psi_i \psi_j^\dagger \rangle$; in the fermion case, the $C_n$ are given by [59]

$$C_n = \Theta\left(-\tilde{M}_n\right)\,. \tag{3.25}$$

To compute the $q^{\text{th}}$ Rényi entropy across a circle of radius $R = r + \frac{1}{2}$, we define a truncated $C_n^r = \left(C_{ij}\right)_n^{1\le i,j\le r}$ for each $n$. The Rényi entropy $S^q(R)$ is then given by

$$S^q(R) = \sum_n S_n^q\,, \qquad S_n^q = \frac{1}{1-q}\mathrm{tr}\log\left[(1-C_n^r)^q + (C_n^r)^q\right]\,. \tag{3.26}$$

For both the scalar and fermion theories, we take our radial lattice to have $N = 200$ points. For the scalar case, we study entangling circles of radii $30 \le r \le 50$ in lattice units, and for the fermion we take $20 \le r \le 80$. We study the massive scalar with mass $m = 0.01k$, $0 \le k \le 8$, and the massive fermion with mass $m = 0.005k$, $0 \le k \le 20$. We compute the $S_n^q$ for $0 \le |n| \le 2000$ for the scalar and $\frac{1}{2} \le |n| \le \frac{4001}{2}$ for the fermion.

The primary sources of error in the numerical method are finite lattice size effects and finite angular momentum cutoff effects. We address the latter source of error by summing the asymptotic expansions of the $S_n$ at large $n$, given in Appendix 3.B, from the angular momentum cutoff to infinity. Finite lattice size effects are most pronounced for small angular momentum modes [71]. To adjust for this, for small $n$ we compute the $S_n^q$ on larger lattices and extrapolate to obtain the value as the lattice size approaches infinity. Specifically, for the scalar we carry out the corrections for angular momentum modes $n = 0, 1\ldots, 5$ on lattices of size $N = 200 + 10\cdot i$ for $i = 0, \ldots, 49$; for the fermion, modes $n = \frac{1}{2}, \frac{3}{2}, \ldots, \frac{15}{2}$ on lattices of size $N = 200 + 10\cdot i$, with $i = 0, 1, \ldots, 15$. Denoting the lattice size as $N$, we fit the resulting data to

$$S_n = a + \frac{b_1}{N^2} + \frac{b_2\log N}{N^2} + \frac{c_1}{N^4} + \frac{c_2\log N}{N^4} + \frac{d_1}{N^6} + \frac{d_2\log N}{N^6} \tag{3.27}$$

and extrapolate to $N \to \infty$ to obtain the lattice-size-corrected value.

## 3.B The $S_n$ at large $n$

It was shown in [65] how, in the scalar theory, one may determine the leading, relevant eigenvalues of $\sqrt{X_n^r P_n^r}$ that go into determining $S^q(R)$ (see (3.22)) in a $1/n$ expansion at large $n$. A key point in the derivation is that the matrices $K_n$ are diagonal to leading order in $1/n$. The result is that the matrix $\sqrt{X_n^r P_n^r} + \frac{1}{2}$ has one eigenvalue equal to [65]

$$1 + \frac{r^2(r+1)^2}{16n^4} + O(1/n^6)\,, \tag{3.28}$$

with all other eigenvalues equal to unity to higher order in $1/n$. Similarly, the leading eigenvalue (away from zero) of the matrix $\sqrt{X_n^r P_n^r} - \frac{1}{2}$ is [65]

$$\frac{r^2(r+1)^2}{16n^4} + O(1/n^6)\,. \tag{3.29}$$

Notice that the leading-order terms in the eigenvalues are independent of the mass $m$.

We may calculate the $1/n$ expansion of the $S_n^q$ by substituting the eigenvalues (3.28) and (3.29) into (3.22). The leading-order behavior of the expansion depends on whether or not $q > 1$. If $q > 1$, we find

$$S_n^q = \frac{q}{1-q}\frac{r^2(r+1)^2}{16n^4} + o(1/n^4)\,, \tag{3.30}$$

while if $q < 1$,

$$S_n^q = -\frac{1}{1-q}\left(\frac{r^2(r+1)^2}{16n^4}\right)^q + o(1/n^{4q})\,. \tag{3.31}$$

Note that when $q = 1$ there is $(\log n)/n^4$ term in the expansion, while this term is absent away from $q = 1$.

The fermion computation is similar. The matrices $\tilde{M}_n$ are block diagonal to leading order in $1/n$. The $2 \times 2$ blocks are indexed by $k = 1, \ldots, N$. To leading order in $1/n$, the $k^{\text{th}}$ block

$$\begin{pmatrix} 0 & \frac{in}{k} \\ -\frac{in}{k} & 0 \end{pmatrix}$$

is diagonalized by $V_k^{(1)} = (i, 1)$, with eigenvalue $\frac{n}{k}$, and $V_k^{(2)} = (-i, 1)$, with eigenvalue $-\frac{n}{k}$. Let $D = \text{Diag}\left(n, -n, \frac{n}{2}, -\frac{n}{2}, \ldots, \frac{n}{N}, -\frac{n}{N}\right)$ be the diagonal matrix of the eigenvalues of $\tilde{M}_n$. To construct the matrix $C_n$, we should evaluate $\Theta(-D) = \text{Diag}(0, 1, \cdots, 0, 1)$. At large $n$, this result does not receive $1/n$ corrections. From this discussion, it is straightforward to see that the eigenvalues of the reduced matrix $C_n^r$, with $r$ even, are equal to either 0 or 1 to leading order in $1/n$.

The eigenvectors may be corrected order by order in $1/n$. To next-to-leading order, we find that

$$\begin{aligned} V_k^{(1)} &= \left(0, \ldots, 0, -i\frac{\alpha_{k-1}}{2n}, \frac{\alpha_{k-1}}{2n}, i, 1 - \frac{km}{n}, \frac{i\alpha_k}{2n}, -\frac{\alpha_k}{2n}, 0, \ldots, 0\right)\,, \\ V_k^{(2)} &= \left(0, \ldots, 0, +i\frac{\alpha_{k-1}}{2n}, \frac{\alpha_{k-1}}{2n}, -i, 1 + \frac{km}{n}, -\frac{i\alpha_k}{2n}, -\frac{\alpha_k}{2n}, 0, \ldots, 0\right)\,, \end{aligned}$$

with $\alpha_k = k(k+1)/(2k+1)$ and the first non-zero entries above sitting $2k-3$ positions to the right. Only the eigenvectors $V^{(1)}_{r/2}$ and $V^{(2)}_{r/2}$ are relevant for determining the two leading eigenvalues of the the matrix $C^r_n$. A straightforward calculation shows that these eigenvalues are

$$1 - \frac{r^2(r+2)^2}{64(r+1)^2}\frac{1}{n^2}\,, \qquad \frac{r^2(r+2)^2}{64(r+1)^2}\frac{1}{n^2}\,. \tag{3.32}$$

All other eigenvalues are equal to either 0 or 1 to higher order in $1/n$. Moreover, the eigenvalues are independent of $m$ at order $1/n^2$. The $m$-dependence arises at higher order. Substituting these eigenvalues into (3.26), we may construct the large $n$ expansion of the $S^q_n$. When $q > 1$, we find

$$S^q_n = -\frac{q}{1-q}\frac{r^2(r+2)^2}{32(r+1)^2}\frac{1}{n^2} + o(1/n^2)\,, \tag{3.33}$$

while when $q < 1$,

$$S^q_n = \frac{2}{1-q}\left(\frac{r^2(r+2)^2}{64(r+1)^2}\frac{1}{n^2}\right)^q + o(1/n^{2q})\,. \tag{3.34}$$

# Chapter 4

# Bekenstein-Hawking Entropy as Topological Entanglement Entropy

This chapter is an edited version of [72], written with Herman Verlinde in 2013.

In this chapter, we continue to study aspects of quantum entanglement, this time in relation to thermodynamics of black holes in $AdS_3/CFT_2$. Namely, black holes in 2+1 dimensions enjoy long range topological interactions similar to those of non-abelian anyon excitations in a topologically ordered medium. Using this observation, we compute the topological entanglement entropy of BTZ black holes via the established formula $S_{\rm top} = \log(S_0^a)$, with $S_b^a$ the modular S-matrix of the Virasoro characters $\chi_a(\tau)$. We find a precise match with the Bekenstein-Hawking entropy. This result adds a new twist to the relationship between quantum entanglement and the interior geometry of black holes. We generalize our result to higher spin black holes, and again find a detailed match. We comment on a possible alternative interpretation of our result in terms of boundary entropy.

The close relation between black hole physics and thermodynamics provides crucial guidance to the search for consistent quantum theories that incorporate gravity. In particular, it indicates that pure quantum gravity – i.e. any attempt to directly quantize the Einstein lagrangian, without the addition of any matter degrees of freedom – is unlikely to give rise to a complete theory. Metric excitations alone seem insufficient to account for the microscopic entropy of black holes, quantified via the Bekenstein- Hawking formula [73, 74]

$$S_{\rm BH} = \frac{\rm Area}{4G_N}\,. \tag{4.1}$$

A more promising perspective is that general relativity represents a long range effective theory with dynamical rules that encode the quantum information flow of underlying elementary degrees of freedom. This point of view is supported by string theory realizations of black hole space-times, in which the B-H formula (4.1) has been successfully matched with the microscopic entropy of the constituent strings, D-branes and their excitations [75].

Another powerful diagnostic tool is the geometric entanglement entropy [76, 77], which has received much recent attention. Let $A$ denote a region of space, such as the interior of a black hole, and $B$ its complement, all of space outside of $A$. The density matrix associated with $A$ is $\rho_A = \mathrm{tr}_B\big(|\psi\rangle\langle\psi|\big)$, where $|\psi\rangle$ is typically taken to be the ground state of the system, and the trace is over all states of $B$. The von Neumann entropy

$$S_A = -\mathrm{tr}(\rho_A \log \rho_A)$$

quantifies the total entanglement between region $A$ and its complement $B$.

The importance of entanglement for the microscopic structure of space-time is only beginning to emerge. There are tantalizing hints of a deep connection, most notably the Ryu-Takayanagi formula [78, 79, 80] and the firewall debate [81, 82]. In this note, we study this relationship in $2+1$-d AdS spacetimes. Einstein gravity in $2+1$-d has special characteristics, akin to Chern-Simons (CS) gauge theories [83, 84] that capture the infrared properties of quantum critical systems with topological order [85, 86]. Massive spinning point particles and black holes enjoy long range interactions that generalize the braiding relations of particles with non-abelian statistics [87]. In addition, the system possesses a ground state degeneracy that is sensitive to the global space-time topology. In condensed matter systems, such as those exhibiting the fractional quantum Hall effect, these remarkable properties emerge because the ground state of the underlying medium is deeply entangled [88]. Quantum gravity in 2+1 dimensions should be thought of in the same way: as the effective theory that captures the topological Berry phases of the ground state wave function. It is through these topological interactions that the quantum order of the microphysical medium manifests itself.

Topological entanglement entropy provides a quantitative measure of this long range quantum order[85, 86]. Consider a region $A$ with disk-like topology and a smooth boundary of length $L$. In a gapped quantum many-body system, the geometric entanglement entropy of $A$ has the form

$$S_A \;=\; \alpha\, L + S_{\mathrm{top}} + \ldots \tag{4.2}$$

where ... indicate terms that vanish in the limit $L \to \infty$. The first term arises from short wavelength modes straddling the boundary of the entangling region. The pre-coefficient $\alpha$ is non-universal, and depends on the UV cut-off. The constant term $S_{\rm top}$ is the topological entanglement entropy; it represents a universal characteristic of the many-body vacuum state [85, 86]. In the above sign convention, it is typically $\leq 0$. It can be isolated from the length term by dividing the region $A$ into three or more segments and taking a suitable linear combination of the resulting entanglement entropies in which the boundary terms cancel. Since topological entanglement entropy survives in the long distance limit, $L \to \infty$, it can be calculated by means of the low energy topological field theory that describes the braiding properties of the quasi-particle excitations. In case the region $A$ contains a single excitation labeled by some charge $a$, one finds that [85, 86, 88]

$$S_{\rm top} = \log\big(d_a/\mathcal{D}\big) = \log(S_0^a). \tag{4.3}$$

Here $\mathcal{D}$ and $d_a$ are the quantum dimensions of the medium and the $a$ excitation, respectively. $S_0^a$ denotes a matrix element of the modular $S$-matrix of the 1+1-dimensional CFT that describes the edge excitations of the topologically ordered medium. The quantity $S_{\rm top}$ has the key property that it does not depend on the size or geometry of the region $A$.

Topological entanglement entropy and black hole entropy seem unrelated. The B-H formula of 2+1-D black holes [89, 90] relates the entropy to the length of the event horizon via

$$S_{\rm BH} \;=\; \frac{\text{Length}(\Gamma)}{4G_N}\,. \tag{4.4}$$

This looks similar to the non-universal length term in (4.2), except that the coefficient $\alpha$ is now a universal constant. Because of this similarity, many authors have suggested that the B-H formula may also have an interpretation as geometric entanglement entropy [76, 77]. There is growing evidence that this is indeed the case [78, 79, 80, 91]. This is an important insight. In particular, it indicates that black holes are typically in a near-maximally entangled state.

However, there is one unsatisfactory aspect to relating the length contributions in (4.4) and (4.2). Unlike the first term in (4.2), the B-H formula (4.4) is universal and robust. In this respect, black hole entropy seems more similar to the universal constant contribution in (4.2). Could it be that the 2+1-D black hole entropy (4.4) can be identified with the universal topological entanglement entropy associated with the black hole space-time?

At a first glance, this seems implausible: the B-H formula does not appear topological, for it

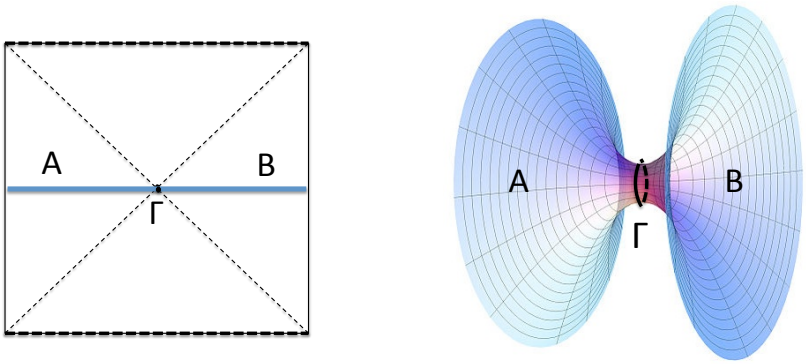



Figure 4.1: The black hole horizon forms a geodesic Γ. The entanglement entropy between the inside and outside regions $A$ and $B$ is equal to Length(Γ)/4.

is proportional to a length. How, then, could this be true? Fig. 1 shows a Penrose diagram of an eternal BTZ black hole of mass $M$ (and spin $J=0$) and a spatial slice with an Einstein-Rosen bridge connecting the two sides. The horizon is a geodesic: it has minimal length for the given topology of Γ. So we can view Length(Γ) as a common property of all loops with the same topology of Γ. In other words, Length(Γ) should not be viewed as a geometric property of a loop, but as a quantum number of the black hole state, determined by its mass $M$ and spin $J$.

Let us now view the black hole as a localized defect of a topological ordered system, and treat $M$ and $J$ in the same way as the charge label $a$ in (4.3). This interpretation is natural given that 2+1-D gravity can be written as a $\mathcal{G} = SL(2,\mathbb{R}) \times SL(2,\mathbb{R})$ Chern- Simons theory [83], in which the black hole state represents a heavy particle with a large $\mathcal{G}$ charge. The edge states of 2+1-D gravity are described by Liouville theory [92, 93, 94, 95, 96, 97], the universal conformal field theory associated with the Virasoro algebra. Although Liouville theory is a non-rational CFT with central charge $c = \frac{3}{2}\ell \gg 1$, it shares many features with rational CFTs [96, 97]. In particular, the Virasoro conformal blocks form a unitary representation of the modular and braid group, characterized by the quantum group $\mathcal{U}_q\big(\mathfrak{sl}(2,\mathbb{R}) \times \mathfrak{sl}(2,\mathbb{R})\big)$. This representation is infinite-dimensional, and modular and fusion relations are expressed as integrals rather than finite sums. Moreover, the spectral density of Liouville theory satisfies the Cardy formula in the heavy-operator regime. Partly as a result, one can identify analogs of quantum dimensions and of the modular S-matrix $S_b^a$.

We can thus apply the same formulas (4.3) to compute the topological entanglement entropy associated with the black hole excitation. Using the proper identification of a black hole of mass $M$ and $J$ with a superselection label $a$ of Liouville CFT, we find a precise match

$$S_{\rm BH} \;=\; \log\big(S_0^a\big), \qquad\qquad a = (M, J), \tag{4.5}$$

in some sense a consequence of the Cardy growth of Liouville theory's spectral density. We describe the details of this calculation in the following sections. To test the robustness of our result, we also consider the higher spin black holes, and find an encouraging match with known results.

This identification and interpretation of the Bekenstein-Hawking entropy as topological entanglement entropy raises many conceptual questions. Why does the computation of the topological entanglement entropy reproduce the microscopic entropy? What does our computation say about the applicability and validity of pure quantum gravity in 2+1 dimensions? What is entangled with what? What does the calculation imply for the firewall controversy? We address these questions in the concluding section.

## 4.1 BTZ Black Hole

We briefly summarize the main properties of the BTZ black hole [89, 90, 94, 95]. From now on we put $G_N = 1$, so $\ell$ denotes the $AdS_3$ curvature radius in Planck units.

$AdS_3$ can be identified with the universal covering space of the group $SL(2,\mathbb{R})$, and has isometry group $\mathcal{G} = SL(2,\mathbb{R}) \times SL(2,\mathbb{R})$. The BTZ black hole space-time is obtained by taking the quotient of $AdS_3$ with a hyperbolic group element $(h_+, h_-) \in \mathcal{G}$, acting via

$$g \sim h_+ g h_-, \qquad h_\pm = e^{\pi(r_+ \pm r_-)\sigma_3/\ell}. \tag{4.6}$$

The quotient describes a stationary and axially symmetric black hole with an outer event horizon at $r_+$ and an inner Cauchy horizon at $r_-$. The BTZ metric can be written as

$$ds^2 \;=\; -4\ell\big(\Delta_+ du^2 + \Delta_- dv^2\big) + d\rho^2 + \big(\ell^2 e^{2\rho} + \Delta_+\Delta_-\big) du dv. \tag{4.7}$$

The two radii $r_\pm$ and the constants $\Delta_\pm$ are related to the black hole mass and spin via

$$M = \frac{r_+^2 + r_-^2}{8\ell^2}, \qquad J = \frac{r_+ r_-}{4\ell}, \qquad \Delta_\pm = \frac{(r_+ \pm r_-)^2}{16\ell} = \frac{1}{2}\big(\ell M \pm J\big). \tag{4.8}$$

Einstein gravity in 2+1 dimensions can be formulated as a CS-type gauge theory by introducing the dreibein $e^a$ and spin connection $\omega^a$. The linear combinations $A^a_\pm = \omega^a \pm \frac{1}{\ell} e^a$ form two $SL(2,\mathbb{R})$ connections, in terms of which the torsion constraint and Einstein equation take the form of flatness

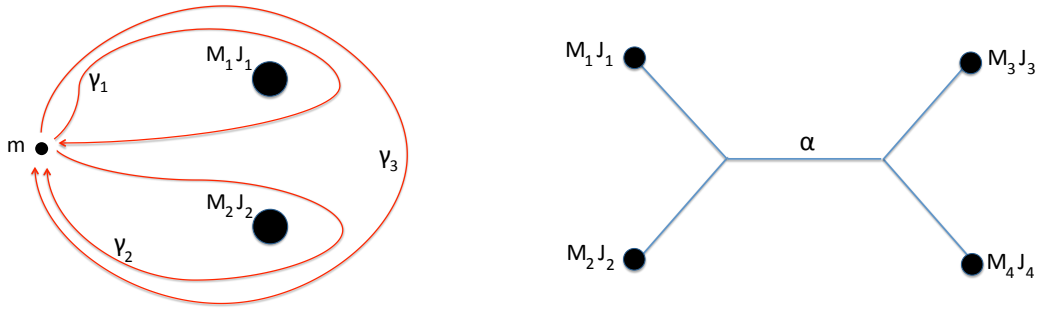



Figure 4.2: The classical space-time geometry is specified by the holonomies around the paths $\gamma_i$. In the quantum theory, states are identified with conformal blocks of 2D Liouville CFT. Although the left and right demonstrate different spacetimes, the conformal block on the right represents a spacetime with four defects labeled by the given conformal dimensions. The projected channel $\alpha$ selects a channel from the two OPEs, and thus specifies the holonomy of the paths around those defects.

constraints [83]. The group elements $h_\pm$ in Equation (4.6) coincide with the holonomies of $A^a_\pm$ around the black hole. In general, $SL(2,\mathbb{R})$ holonomies come in three types, depending on whether the conjugacy class of the group element is hyperbolic, parabolic, or elliptic. For a black hole, both holonomies are in a hyperbolic conjugacy class.

The Bekenstein-Hawking entropy of the BTZ black hole is equal to

$$S_{\rm BH} = \frac{2\pi r_+}{4}. \tag{4.9}$$

This formula has been reproduced in numerous dual CFT realizations of string theory on $AdS_3$ by counting the number of states at energy $M$ and with angular momentum $J$. Below we will give a new derivation and interpretation.

## 4.2 Quantum Geometry

Quantum geometry arises from quantizing the phase space of space-time geometries. As an example, Fig. 2 indicates the geometry of two BTZ black holes, specified by the $SL(2,\mathbb{R})\times SL(2,\mathbb{R})$ holonomies around the paths $\gamma_i$. These holonomies are determined, up to overall conjugation, by the mass, spin, center of mass energy and total angular momentum of the two black holes. This description generalizes to any number of point particles and black holes [83]. The space of $SL(2,\mathbb{R})$ holonomies is isomorphic to Teichmüller space, the space of constant negative curvature metrics on a 2-D surface. The phase space of 2+1-D Einstein gravity consists of two copies of Teichmüller space [92, 93].

The problem of quantizing Teichmüller space has been solved [96, 97, 98, 99]. It gives rise to a Hilbert space of states that can be identified with the linear space spanned by the chiral conformal

blocks of 2-D Liouville theory [96, 97, 98, 99]

$$S_L(\varphi) \;=\; \frac{1}{4\pi}\int d^2\xi\left[\frac{1}{2}(\partial\varphi)^2 + QR\varphi + \mu e^{b\varphi}\right], \qquad Q = b + b^{-1}. \tag{4.10}$$

This correspondence generalizes the well-known relationship between Chern-Simons theories and WZW conformal field theory [84]. The dictionary is analogous. The 2-D CFT describes the massless edge excitations at the boundary of the AdS space, and supports a unitary representation of the asymptotic symmetry group of the bulk theory. For pure $AdS_3$ gravity, this symmetry group takes the Virasoro algebra with central charge [100]

$$c = 1 + 6Q^2 = 3\ell/2\,. \tag{4.11}$$

States of 2+1-D gravity with particle and black hole excitations in the bulk are identified with the product of left and right conformal blocks of Liouville CFT with corresponding vertex operator insertions. They enjoy a $q$-deformed version of the monodromy properties of the classical geometry, governed by the non-compact quantum group $\mathcal{U}_q\big(\mathfrak{sl}(2,\mathbb{R})\times\mathfrak{sl}(2,\mathbb{R})\big)$ with $q = \exp(i\pi b^2)$. Liouville vertex operators take the general form $V_\alpha = e^{\alpha_+\varphi_+}\,e^{\alpha_-\varphi_-}$, and are in one-to-one correspondence with unitary highest weight representations of the left and right Virasoro algebra with conformal weights $\Delta_\pm = \alpha_\pm(Q-\alpha_\pm)$. The physical range of positive conformal weights splits into two separate regimes of Liouville momenta

$$\alpha_\pm \in [0, \tfrac{1}{2}Q] \,\cup\, \big(\tfrac{1}{2}Q + i\mathbb{R}^+\big). \tag{4.12}$$

The Liouville equation prescribes that the metric have constant negative curvature everywhere except at the location of the vertex operators. Vertex operators with real Liouville momentum in the interval $[0, \frac{1}{2}Q]$ create elliptic solutions, which are local cusps specified by a patching function in the elliptic conjugacy class of the isometry group $\mathcal{G}$. Vertex operators with complex momenta of the form $\frac{1}{2}Q + i\mathbb{R}^+$ create hyperbolic solutions, which are macroscopic holes in 2-D space identified with the spatial section of BTZ black hole geometries (as shown in Fig. 1 and Fig. 3.). We may parametrize the Liouville momenta in this range as

$$\alpha_\pm = \tfrac{1}{2}Q + ip_\pm, \qquad\qquad \Delta_\pm = p_\pm^2 + \tfrac{1}{4}Q^2. \tag{4.13}$$

These relations, combined with Equations (4.7) -(4.8), specify a precise dictionary between the

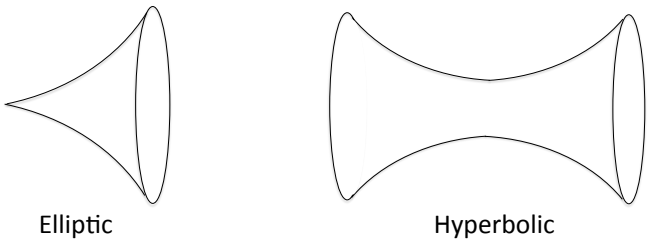



Figure 4.3: Liouville vertex operators fall into two classes. Those with $\Delta < \frac{1}{4}Q^2$ create *elliptic* solutions (punctures), those with $\Delta > \frac{1}{4}Q^2$ create *hyperbolic* solutions (macroscopic holes) [101] These are explicitly related to $SL(2,\mathbb{R})$ holonomies which specify the spacetime, as described in detail in chapter 5.

classical data of the BTZ black hole and the quantum data of Liouville theory. For later reference, we make note that in the semiclassical regime $p_\pm \gg b \gg 1$, the relations between the Liouville momenta $p_\pm$ and the conjugacy class of the holonomies $h_\pm$ in (4.6) simplify to

$$r_\pm = 4b(p_+ \pm p_-), \qquad\qquad b^2 = \ell/4\,. \tag{4.14}$$

Most of the above dictionary was known before the discovery of gauge/gravity duality. An important insight from AdS/CFT is that the bulk theory can not be pure gravity. Gravity in 2+1-D describes how massive localized excitations interact at long distances, but it does not specify the hyperfine structure of the excitation spectrum of the bulk string theory.

The situation in the 1+1-D boundary theory is analogous. Liouville theory has a continuous spectrum of conformal dimensions, and is therefore capable of describing any set of Virasoro representations. However, it does not prescribe the spectrum of some given unitary CFT. Liouville theory is similar to a non-compact space with a continuous spectrum of wave solutions; choosing a specific CFT realization of $\mathrm{AdS}_3$ is like putting the wave solutions in a finite box, so that the spectrum becomes discrete and countable.

## 4.3 Cardy growth and a universal regime of pure gravity

We claim that the description of pure gravity in terms of Liouville theory given above has a CFT interpretation whereby Liouville theory serves as a universal limit of unitary CFTs dual to any *particular* realization of bulk gravity. Crucial to this claim is the observation that in the regime of large conformal dimension, the (continuous, representation–theoretically complex) spectrum of Liouville theory exhibits exponential growth in agreement with the Cardy formula, which describes a regime of states in a CFT dual to black holes in the bulk.

First, a review of general properties of the spectra of CFTs dual to gravity theories via AdS/CFT. These CFTs are nonrational CFTs with $c >> 1$ which we further assume to be "typical," with no additional symmetries. Such CFTs have a spectrum divided into a sparse spectrum of light primary operators ($\Delta < c/12$) and a dense spectrum of heavy primary operators ($\Delta > c/12$)[102].

The best-understood CFTs are rational CFTs, which have finitely many primaries and therefore have entropy which mostly derives from the Virasoro symmetry. To the contrary, a general result in CFT, the Cardy formula, implies that for a *generic, nonrational* CFT, this is *not* the case. For $\Delta >> 1$, such CFTs have exponentially dense spectra that grow with the conformal dimension. This is a consequence of modular invariance of the CFT partition function. The Cardy formula states that [103]

$$S_{\text{Cardy}}\left(\Delta, \overline{\Delta}\right) = 2\pi\sqrt{\left(\Delta - \frac{c}{24}\right)\frac{c}{6}} + 2\pi\sqrt{\left(\bar{\Delta} - \frac{c}{24}\right)\frac{c}{6}}. \tag{4.15}$$

One possible point of confusion is that a typical nonrational CFT should have nondegenerate spectrum. What is the Cardy formula counting? One can think of it as an entropy in a microcanonical ensemble: it measures the number of primaries in a small interval around a primary with conformal dimensions $(\Delta, \overline{\Delta})$. Again, this is in the large-$c$, heavy-operator regime.

We emphasize that the Cardy entropy does not depend on detailed information about a specific CFT. This suggests that the regime of heavy operators describes *universal* physics in nonrational CFT. Moreover, famously, the Cardy formula matches the Bekenstein-Hawking formula for the entropy of a black hole dual to an operator with given $\Delta, \overline{\Delta}$.

The Cardy formula serves as an indication that there exists a regime of states in AdS/CFT whose properties do not depend on the specific details of the spectrum, but only on an overall "coarse-grained" picture. This "coarse-grained" property is reproduced in Liouville theory. Despite the fact that the Liouville spectrum derives from the representation theory of a noncompact, deformed quantum group and is highly nontrivial to compute, in the limit of large, real conformal dimension, it satisfies the Cardy growth.

In order to express this Cardy growth, we make use of some notation from section 4.2. Let $b$ be the deforming parameter such that $q = e^{i\pi b^2}$. Label representations of $U_q(\mathfrak{sl}_2(\mathbb{R}))$ by the parameter $\alpha$. We express the central charge $c$ and conformal dimensions $\Delta$ using variables

$$Q = b + b^{-1} \qquad c = 1 + 6Q^2 \qquad \Delta = j(Q - j) \tag{4.16}$$

For large$-c$, we take small-$b$, such that $Q \sim b^{-1}$. Since we will be in the heavy state regime, we take $\Delta > Q^2/4$. The spectrum of Liouville theory has two pieces: a discrete piece and a continuous piece. The heavy states correspond to the continuous part of the spectrum. Therefore, in this regime, the we can use a real variable $p$ to parametrize the representation labels,

$$j = \frac{Q}{2} + ip \qquad\qquad \Delta = \frac{Q^2}{4} + p^2. \tag{4.17}$$

Of interest to us is the density of the spectrum of Liouville theory. The density of the continuous spectrum of Liouville theory is known [98, 99], so we will write it down, first by expressing it in terms of representations $\alpha$ of $U_q(\mathfrak{sl}_2(\mathbb{R}))$ (a chiral half of Liouville theory).

Since the spectrum is continuous, we want a *measure* on the conformal dimensions. The measure $d\mu(\alpha)$ of $U_q(\mathfrak{sl}_2(\mathbb{R}))$ is expressed in terms of the density $\rho(\alpha)$ as

$$d\mu(\alpha) \equiv \rho(\alpha) d\alpha \tag{4.18}$$

$$\rho(\alpha) = 4 \sinh(2\pi b p) \sinh(2\pi b^{-1} p). \tag{4.19}$$

We can use the dictionary from equation 4.16 to rewrite this in terms of $\Delta$. Taking the large-$c$ limit, we find that

$$d\mu(\Delta) \sim e^{2\pi\sqrt{\frac{c}{6}(\Delta - \frac{c}{24})}} d\Delta. \tag{4.20}$$

This is in perfect agreement with one half of the Cardy formula in eq. 4.15. This measure is known as the Plancherel measure. Since the nonchiral Liouville theory requires two copies of this measure, we find perfect agreement with the expectation from CFT with discrete spectra.

In the following section, we will see that *this* identification between Liouville theory and pure gravity encapsulates the *topological* properties of the gravity theory. This is further demonstrated by the fact that the quantum dimension, and therefore the topological entanglement entropy, are deeply connected with - and in some sense, identified with - the Plancherel measure.

## 4.4 Quantum Dimension

An important ingredient of our story is the *quantum dimension* associated with a local excitation in a topological quantum field theory. We first recall the definition and properties of the quantum dimension of a topological QFT associated to a rational CFT. We then generalize to the case of interest, the non-rational $c > 25$ Virasoro CFTs.

The most physical definition of the quantum dimension is as follows. Let $\mathcal{H}_a(N)$ denote the Hilbert space of the 2+1-dimensional topological QFT spanned by all states that contain $N$ local excitations of charge $a$. It can be shown that the dimension of this Hilbert space grows exponentially at large $N$ according to

$$\dim \mathcal{H}_a(N) \propto \big(d_a\big)^N. \tag{4.21}$$

The number $d_a$ defines the quantum dimension of the excitation $a$.

Quantum dimensions are linked with the fusion algebra [104, 105, 106]. A local excitation with charge $a$ corresponds to a primary vertex operator $V_a$ in the CFT. The operator product of $V_a$ and $V_b$ can be expanded as a sum of operators $V_c$. For rational CFTs, the fusion coefficients $N_{ab}{}^c$ are integers that specify the multiplicity of $V_c$ in this expansion. The fusion algebra is commutative and associative, and admits a one-dimensional representation $d_a d_b = \sum_c N_{ab}{}^c d_c$. This relation can be used to prove the result (4.21).

Quantum dimensions can be thought of as the character of the superselection sector $\mathcal{H}_a$ associated with the primary vertex operator $V_a$. States in $\mathcal{H}_a$ are obtained by acting with symmetry generators on the primary state $|a\rangle = V_a|0\rangle$. The partition function

$$\chi_a(\tau) \;=\; \mathrm{tr}_{\mathcal{H}_a}\big(e^{i\pi\tau L_0}\big) \tag{4.22}$$

is called the character of the sector $\mathcal{H}_a$. The quantum dimension $d_a$ is obtained by taking the $\tau \to 0$ limit of the ratio of $\chi_a(\tau)$ with the identity character [104, 105, 106]

$$d_a \;=\; \lim_{\tau\to 0} \frac{\chi_a(\tau)}{\chi_0(\tau)}\,. \tag{4.23}$$

This definition naturally explains why the quantum dimensions generate the fusion algebra. It also allows us to re-express $d_a$ in terms of the modular $S$-matrix, which describes the transformation

properties of the characters under the modular transformation $\tau \to -1/\tau$

$$\chi_a\big(-1/\tau\big) \;=\; \sum_b S_a^b\, \chi_b(\tau). \tag{4.24}$$

Applying the modular transformation (4.24) to (4.23), and using that for $\tau \to 0$ the dominant term in the sum comes from the identity character, one finds that

$$d_a = \frac{S_0^a}{S_0^0}. \tag{4.25}$$

This formula for the quantum dimension holds for rational CFTs and plays a key role in the computation of topological entanglement entropy. We will use this connection momentarily.

First, we need to generalize the above formulas to the case of non-rational CFTs relevant to 2+1-D gravity. The modular geometry of Liouville theory is by now quite well-developed [96, 97, 98, 99], and many of the RCFT formulas have found direct non-rational analogs. There are two main differences. Since the spectrum of allowed conformal dimensions is continuous, modular transformations and fusion coefficients are no longer described by discrete sums and finite matrices but by integrals and continuous distributions. Another important difference is that the identity representation plays a rather distinct role. In spite of these dissimilarities, there still exists a natural analog of the notion of quantum dimension.

Let us follow the naive route and simply apply the formula (4.25). The $c > 25$ Virasoro characters for the continuous representations of conformal weight $\Delta > \frac{1}{4}Q^2$ are given by

$$\chi_p(\tau) = \frac{e^{i\pi\tau p^2}}{\eta(\tau)}, \qquad \Delta_p = p^2 + \tfrac{1}{4}Q^2, \tag{4.26}$$

where the Dedekind $\eta$-function $\eta(\tau) = q^{c/24}\prod_{n>0}(1-q^n)$ with $q \equiv e^{2\pi i\tau}$. The identity character

$$\chi_0(\tau) = \frac{e^{-i\pi\tau\frac{Q^2}{4}}(1-e^{i\pi\tau})}{\eta(\tau)}, \qquad \Delta = 0 \tag{4.27}$$

follows the following modular transformation property [96]

$$\chi_0\big(-1/\tau\big) \;=\; \int_0^\infty dp\; S_0^p\, \chi_p(\tau) \tag{4.28}$$

$$S_0^p \;=\; 2\sqrt{2}\sinh(2\pi bp)\sinh(2\pi b^{-1}p)\,. \tag{4.29}$$

Note that i) $S_0^p$ is not a matrix entry of a finite matrix, but a measure on the continuous series of Virasoro representations, ii) $S_0^p$ grows exponentially with $p$, and iii) the identity representation itself does not appear on the right-hand side of (4.28).

Boldly applying the formula (4.25), we find that, up to an irrelevant overall constant, the quantum dimension of the representation with Liouville momentum $\alpha = \frac{1}{2}Q + ip$ is given by

$$d(\alpha) \;=\; \sinh(2\pi bp)\sinh(2\pi b^{-1}p). \tag{4.30}$$

Observe that this is the Plancherel measure, which we have already seen in the context of the Cardy growth of the spectrum of Liouville theory in eq.4.18. We see that the Plancherel measure is also the most natural counterpart of the quantum dimension in the nonrational case.

## 4.5 Topological Entanglement Entropy

Gravity is topological in the sense that every observable must be coordinate invariant. In 2+1 dimensions, this topological nature is enhanced by the fact that there are no graviton excitations, and that the metric, outside of matter distributions, locally always looks the same. Our proposal is that from a microscopic perspective, these properties emerge because gravity is the long distance description of the highly entangled ground state of a topologically ordered system close to a quantum critical point. For analogous condensed matter systems, the natural diagnostic for the presence of topological order is the topological entanglement entropy introduced in [85, 86]. Let us briefly recall its definition.

To compute the topological entanglement entropy $S_{\rm top}$ of a disk-shaped region $R$ with the outside $D = R^c$, one first divides the interior of $R$ into three sectors $A$, $B$ and $C$. Let $S_A = -{\rm tr}\rho_A \log \rho_A$ denote the von Neumann entropy of the density matrix $\rho_A$ associated with subregion $A$, and analogously for $S_B$, $S_C$. Similarly, let $S_{AB}$ denote the entropy associated with $A \cup B$, etc. The topological entanglement entropy of $R$ is then defined as [85] $S_{\rm top} = S_A + S_B + S_C - S_{AB} - S_{AC} - S_{BC} + S_{ABC}$. This linear combination has the special property that all non-universal perimeter-law contributions cancel out. Moreover, any local deformation of the entangling boundary does not alter the final result.

Applying this definition to a topological field theory associated with a 2-D rational CFT, one finds [85, 86] that an empty region of space has $S_{\rm top}(0) = \log\big(1/D\big) = \log\big(S_0^0\big)$. Here $D$ is the

total quantum dimension of the medium, and related to the quantum dimension $d_a$ of individual excitations via $D^2 = \sum_a d_a^2$. In case the region contains a quasi- particle excitation of charge $a$, the topological entanglement entropy is given by the formula [88]

$$S_{\text{top}}(a) \;=\; \log\big(d_a/D\big) = \log\big(S_0^a\big). \tag{4.3}$$

Note that for a rational CFT the topological entanglement is always negative since $d_a \;<\; D$. By itself, this would not make sense, as the definition of the entanglement entropy is a manifestly positive quantity. Once we include the non-universal contribution proportional to the length $L$ of the entangling boundary, however, the total result is positive.

We make the assumption that the relationship between the quantum dimension and the topological entanglement entropy remains mostly unchanged in going from rational to non-rational CFTs. One important difference is that there no longer exists an analog of the total quantum dimension $D$, and hence there is no obvious notion of topological entanglement entropy of an empty region of space. However, there does exist a natural formula for the topological entanglement entropy of a black hole excitation. Applying the formula (4.3) to the hyperbolic Virasoro representation associated with a BTZ black hole, we find

$$S_{\text{top}}(M,J) \;=\; \log\big(S_0^{p_+} S_0^{p_-}\big)\,. \tag{4.27}$$

The relation between the mass $M$ and spin $J$ and the Liouville momenta is given in Equation (4.16) with $\Delta_\pm$ defined in (4.8). Plugging in the explicit modular S-matrix element (4.29) gives

$$S_{\text{top}}(M,J) \;=\; \log\Big(8\sinh(2\pi b p_+)\sinh(2\pi b^{-1}p_+)\sinh(2\pi b p_-)\sinh(2\pi b^{-1}p_-)\Big). \tag{4.28}$$

Note that unlike the rational CFT case, the right-hand side is positive. Moreover, it grows unboundedly for large $p_\pm$. In the limit where $\ell M \pm J$ and $b$ are all large, it reduces to

$$S_{\text{top}}(M,J) \;=\; 2\pi b(p_+ + p_-) \;=\; \frac{2\pi r_+}{4}, \tag{4.29}$$

which exactly matches the Bekenstein-Hawking entropy. Note that this has derived from the quantum dimension, which is equal to the Plancherel measure of eq. 4.18. We thus find that the Cardy growth of the Plancherel measure representing the continuous spectral density of Liouville theory encodes the topological properties of BTZ black holes in $2+1$-D. This is our main result.

## 4.6 Higher Spin Black Hole Entropy

As a test of our proposal, let us consider black holes in 2+1-D higher spin gravity [107, 108]. Luckily, all the necessary technology is available. Our presentation will be brief.

Higher spin gravity in 2+1 dimensions is a generalization of Einstein gravity in 2+1 dimensions that includes a collection of $n-2$ higher spin fields [107]. All the fields together can be assembled into a $SL(n,\mathbb{R}) \times SL(n,\mathbb{R})$ gauge connection $(A_+, A_-)$ with a Chern-Simons action. The generalized space-time geometry of a higher spin black hole is characterized by two $SL(n,\mathbb{R})$ holonomies

$$h_\pm = e^{2\pi(\lambda_+ \pm \lambda_-)/\ell}, \tag{4.30}$$

which generalize the $SL(2,\mathbb{R})$ holonomies (4.6) of the BTZ black hole. Here $\ell$ denotes the higher spin generalization of the $AdS_3$ radius, and $\lambda_+$ and $\lambda_-$ are diagonal elemenents of the $\mathfrak{sl}(n,\mathbb{R})$ Lie algebra. Higher spin black holes thus carry $2(n-1)$ quantum numbers, including the mass, angular momentum and $2(n-2)$ higher spin charges.

Extracting an actual space-time geometry from this general description of the higher spin black hole turns out to be a non-trivial task. In particular, there is no gauge-invariant notion of a 2+1-D space-time metric that can be used to compute a horizon area. As a result, there appears to be no immediately obvious higher spin generalization of the Bekenstein-Hawking formula. There are indeed various proposals [108, 109, 110].

A simple geometric proposal for a generalized Bekenstein-Hawking formula was put forward in [109, 110]. Let $e_i$ denote the simple roots of $\mathfrak{sl}(n)$ and $\langle\ ,\ \rangle$ denote the Cartan Killing form. The Weyl vector is defined as $\rho = \frac{1}{2}\sum_{e>0} e$. The higher spin generalization of the black hole entropy formula derived in [109, 110] is expressed in terms of the $SL(n,\mathbb{R})$ holonomies $h_\pm$ as

$$S_{\text{HSBH}} \ = \ \frac{2\pi}{4}\langle \rho, \lambda_+ \rangle \, . \tag{4.31}$$

This elegant proposal passes some non-trivial checks [109, 110] and appears well-motivated.

Can one reproduce the generalized B-H formula (4.31) by counting states in the dual CFT? This is a non-trivial task, since one needs a generalization of the Cardy formula that keeps track of conformal dimensions and all higher spin quantum numbers. This has not been done yet. We now give a simple derivation of (4.31) via the topological entanglement entropy (4.3).

2+1-D higher spin gravity is dual to 1+1-D conformal field theory with $W_n$ symmetry, the natural

higher spin generalization of Virasoro symmetry. The universal CFT with $W_n$ symmetry is $\mathfrak{sl}(n,\mathbb{R})$ Toda theory

$$S = \frac{1}{2\pi}\int d^2\xi \left[ \langle \partial\varphi, \bar{\partial}\varphi\rangle + R\langle Q,\varphi\rangle + \mu \sum e^{b\langle e_i,\varphi\rangle}\right], \qquad Q = 2(b+b^{-1})\rho. \tag{4.32}$$

Toda theory is a non-rational CFT with central $c = n-1+3\langle Q,Q\rangle$. As before, states of the 2+1-D higher spin theory with localized excitations are identified with the tensor product of left and right conformal blocks of the CFT. Black holes states correspond to vertex operators that create macroscopic holes in the generalized space time, with holonomies (4.30) in a hyperbolic conjugacy class of $SL(n,\mathbb{R})$. Their vertex operators $V = e^{\langle \alpha_+,\varphi_+\rangle}e^{\langle \alpha_-,\varphi_-\rangle}$ have Toda momenta $\alpha_\pm = \frac{1}{2}Q + ip_\pm$ and conformal weights $\Delta_\pm = \langle \alpha_\pm, Q-\alpha_\pm\rangle = \langle p_\pm, p_\pm\rangle + \frac{1}{4}\langle Q,Q\rangle$. The semi-classical relations (4.14) naturally generalize to

$$\lambda_\pm = 4b(p_+ \pm p_-), \qquad\qquad b^2\langle \rho,\rho\rangle = \ell/8. \tag{4.33}$$

Just like their BTZ counterparts, higher spin black holes can be viewed as macroscopic quasi-particle excitations with topological interactions. We can thus compute their topological entanglement entropy in the same way as before. The relevant modular S-matrix elements of $\mathfrak{sl}(n,\mathbb{R})$ Toda field theory was computed in [111]

$$S_0^p = \Xi \prod_{e>0} 4\sinh\big(\pi b\langle e,p\rangle\big)\sinh\big(\pi b^{-1}\langle e,p\rangle\big) \tag{4.34}$$

with $\Xi$ some irrelevant constant. Using the formula $S_{\text{top}} = \log\big(S_0^{p_+}S_0^{p_-}\big)$ and taking the semi-classical limit, we reproduce the result (4.31)

$$S_{\text{top}} \;=\; 2\pi b\big(\langle \rho,p_+\rangle + \langle \rho,p_-\rangle\big) \;=\; \frac{2\pi}{4}\langle \rho,\lambda_+\rangle. \tag{4.35}$$

## 4.7 Concluding Remarks

We have put forward a new interpretation of 2+1-D quantum gravity as the effective field theory that describes the long range properties of a highly entangled ground state. Moreover, we propose that Liouville theory describes this pure gravity regime as an approximation to any specific CFT, using the Cardy growth of the Liouville spectrum as evidence. In line with this interpretation, we have computed the topological entanglement entropy of a BTZ black hole. Our computation does

not make use of the Bekenstein-Hawking, Ryu-Takayanagi, or Cardy formulas. It is a new and independent derivation, yet yields a leading-order result that matches all three. Our result also raises a number of questions. We briefly comment on some of them.

*Does pure 2+1-D quantum gravity exist? What is its role?*

Via the identification with the space of left and right conformal blocks of 2-D Liouville theory, we have given a well-defined description of the Hilbert space of 2+1-D quantum gravity. Does this mean that pure 2+1-D quantum gravity exists as a UV complete theory? The answer is "No" [112]. The spectrum of Virasoro representations is continuous, and thus the level density of states of Liouville theory and pure 2+1-D gravity is strictly infinite. This is an unphysical situation. To get a well-behaved physical system, one needs to supply a specific holographic dual in the form of some unitary 2-D CFT. This CFT prescribes the allowed discrete spectrum of conformal dimensions, with a finite level density. In this note, we implicitly assumed that this CFT is maximally non-rational, i.e. that it does not have any other symmetries than conformal invariance. In this idealized case, once the spectrum of excitations is prescribed, 2+1-D gravity gives an accurate description of their long range interactions and assigns the correct quantum dimension to the black hole states.

*What does the topological entanglement entropy count?*

This is the most important question. It is natural to interpret $S_{\rm top}$ as the universal contribution to the entanglement across the black hole horizon. The fact that it saturates the B-H bound is consistent with the idea [79, 80] that entanglement is responsible for the continuity of space across the horizon. However, this interpretation immediately raises an important puzzle, closely related to the firewall paradox [81, 82].

According to the usual AdS/CFT dictionary, any typical CFT state with large enough energy describes a black hole in the bulk. The level density of the CFT indeed matches the microscopic B-H entropy. However, to write a state with entanglement entropy proportional to $S_{\rm BH}$, one needs to include *two* Hilbert space sectors each with entropy at least equal to $S_{\rm BH}$. The CFT seems to provide only one of these sectors. So where is the other sector?

Liouville vertex operators with momenta $\alpha = \frac{1}{2}Q + ip$ in fact create macroscopic holes in space, as indicated in Fig. 3. Based on the similarity with Fig. 1, it is tempting to identify both sides of the hyperboloid in Fig. 3 as the two sides of the eternal black hole solution. According to this interpretation, it seems that by acting with the vertex operator, one has created a completely new asymptotic region with its own holographic CFT dual. This could be where the other sector resides.

But how would one create such a second asymptotic region via gravitational collapse, i.e. by acting with operators on the vacuum of one single CFT? This is one version of the firewall question.

*Is there a firewall or fuzzball? Is $S_{\rm BH}$ a boundary entropy?*

In our view, if our proposal that the entanglement entropy of BTZ black holes saturates the B-H bound is correct, then there is no firewall. The state looks like an eternal black hole that realizes the balanced holography postulate put forward in [91]. The entanglement across the horizon is then sufficient to safeguard the continuity of space [79, 80].

There is, however, another possible interpretation[1] of our formula $S_{\rm BH} = \log S_0^a$ in terms of the Affleck-Ludwig boundary entropy [113]. Suppose that, instead of the hyperbolic solution of Fig. 3, we place a reflecting boundary at the location of the black hole horizon. A natural conformal boundary for Liouville CFT is the ZZ-boundary state $|\,ZZ\rangle$ [96]. Its overlap with the Ishibashi state $\|\,p\rangle\!\rangle$, the eigenstates with given Liouville momentum $\alpha = \frac{1}{2}Q + ip$, satisfies

$$\left|\Psi_{\rm ZZ}(p)\right|^2 = S_0^p, \qquad\qquad \Psi_{\rm ZZ}(p) = \langle\!\langle\, p \,\|\, ZZ\,\rangle. \tag{4.36}$$

This implies that the boundary entropy of the ZZ state in the sector with momentum $p$ is equal to $\log(S_0^p)$. Moreover, if we identify the topological entanglement entropy with the Bekenstein-Hawking entropy of the BTZ black hole, we obtain the very suggestive relation

$$Z_{\rm BH} = \left|\Psi_{\rm ZZ}(p_+, p_-)\right|^2 \tag{4.37}$$

with $Z_{\rm BH} = e^{S_{\rm BH}}$. Could it be that, instead of topological entanglement entropy, our formula is counting the boundary entropy of a reflecting boundary at the horizon? Or are both interpretations correct? Either way, we believe that finding the answer to these questions will shed important new light on the nature of the interior geometry of black holes.

## Acknowledgement

We thank Daniel Harlow, Tatsuma Nishioka, Per Kraus, Erik Verlinde, Nick Warner and Masahito Yamazaki for helpful discussions. The work is supported by NSF grant PHY-1314198.

[1] We thank Nick Warner for emphasizing this possible alternative interpretation.

# Chapter 5

# Conformal Bootstrap, Universality and Gravitational Scattering

This chapter fits into the $AdS_3/CFT_2$ theme of our discussion. Rather than quantifying *incomplete* information to learn new physics, here we are concerned with fleshing out an *exact* dictionary between Liouville theory on the boundary and a long-distance regime in which pure gravity dominates for a bulk theory containing a black hole. By analyzing the behavior of two particles interacting near a black hole horizon, we discover that the exchange matrix governing the scattering is exactly described by the braiding properties of the boundary CFT in the limit where its discrete spectrum is approximated by the continuous spectrum of Liouville theory. This demonstrates that scattering near a black hole horizon encodes *detailed information* about the boundary CFT. Perhaps counterintuitively, this implies that by considering this "thermodynamic limit" of $AdS_3/CFT_2$ we are able to construct an exact duality between scattering in the bulk and the Virasoro bootstrap on the boundary.

This chapter is based on results contained in [114] written with Steven Jackson and Herman Verlinde.

## 5.1 Introduction

The AdS/CFT correspondence is one of the deepest discoveries of modern high energy physics [13, 14, 17]. Since its discovery, a vast literature of work has documented progress toward using AdS/CFT to learn about the bulk and the boundary. Still, there remains much to learn about the dictionary

which connects the bulk theory and the boundary. For example, we would like to have a better understanding of quantum mechanical properties of the bulk gravity theory away from the asymptotic boundary.

Recently, Shenker and Stanford carried out an interesting computation of scattering in the vicinity of the black hole horizon where they demonstrated, among other things, an exponential dependence on the parameters dictating whether one particle reaches the asymtotic boundary as a result of the interaction [115]. This calculation was carried out in the context of classical gravity by observing a shift in the black hole horizon as a result of the interaction. One can go much further, however, with the observation that the holonomy picture of gravity in $2+1$d indicates a natural progression from the classical computation to a quantum computation using the braiding properties of a universal limit of irrational CFT on the boundary, Liouville theory. Indeed, the braid "matrix" of Liouville theory specifies the scattering matrix of particles in the bulk. We observe and test this correspondence, which contributes to our understanding of the $\mathrm{AdS}_3/\mathrm{CFT}_2$ dictionary away from the boundary.

We use this proposal to make a physical computation of the scattering matrix of massless particles in a black hole background analogous to the classical computation which had been carried out previously. In the semiclassical limit, the result we obtain is equal to the result obtained using classical Einstein gravity [116] as well as intuition derived from the scattering of wavepackets and Hamilton–Jacobi theory. Moreover, we will show that the scattering amplitude is encoded by the Lorentzian CFT exchange algebra [117], and therefore interactions near a black hole hold detailed dynamical information about the CFT. These facts correspond to highly nontrivial checks on our proposed correspondence between Liouville theory and gravity in the bulk.

We address this question by considering a fairly tractable toy model: that of quantum mechanics in the bulk of $\mathrm{AdS}_3/\mathrm{CFT}_2$. It is tractable because although the expressions for relevant quantities in Liouville theory are highly complicated, they are known [118]. We utilize topological structures (the braiding matrix, holonomy) on both sides of the correspondence to model the gravity-dominated regime of the bulk theory nonperturbatively using the CFT. The key to this match is the universal high-entropy limit of the CFT, with spectrum dominated by the Cardy entropy [103], as demonstrated in chapter 4, sec. 4.3. In this regime, the Cardy formula is equal to the the density of the continuous spectrum of Liouville theory [99]. We demonstrate that by making a continuity approximation to model the high-entropy regime of any large-$c$ 2d CFT using Liouville theory, we obtain a description of the dynamics of the gravity theory, with Virasoro conformal blocks serving as states

and the Virasoro modular geometry governing the dynamics of the theory.

The plan for this chapter is as follows. In section 5.2, we review the construction of locally $AdS_3$ spacetimes in terms of holonomy and $SL(2,\mathbb{R}) \times SL(2\mathbb{R})$. In section 5.3, we discuss two classical formulations of a scattering problem between infalling and outgoing particles near a black hole horizon. We compute time-dependent quantities which demonstrate exponential sensitivity to initial conditions, and conclude this process demonstrates the "butterfly effect," i.e. chaos. In section 5.4, we review the description of $AdS_3$ quantum mechanically using Teichmüller space and demonstrate that its Hilbert space is isomorphic to the linear space spanned by Virasoro conformal blocks - that is, it is associated with Liouville theory [119, 120, 97]. We explicitly demonstrate a mechanism for describing spacetimes using conformal blocks. Finally, in section 5.5, we demonstrate that scattering near a black hole is described by the braid matrix of Liouville theory, and moreover, the Lorentzian, nonchiral theory encodes detailed dynamical time-dependent information in the bulk. We end with a discussion in section 5.6, which provides a summary and gives some future prospects.

Appendix 5.A discusses the basics of 2d CFT and reviews our notation regarding the normalization of conformal blocks. Appendix 5.B details some technical expressions not given in the main text.

This chapter can be seen as building upon the framework which was proposed in Chapter 4.

## 5.2 $AdS_3$ and nontrivial holonomy

We begin with a review of classical $2+1$-d gravity with negative cosmological constant. We do so using the language of holonomies around topological defects, as this formalism makes the topology of the gravity theory manifest.

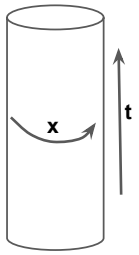



Figure 5.1: $AdS_3$ may be portrayed as a cylinder, with $r$ the radial direction and $t$ shown vertically.

$AdS_3$, depicted in figure 5.1, is a fundamental object of study in our discussion. It is the maximally symmetric space of constant negative curvature in $2+1$-d, and it can be constructed via

an embedding as a hyperboloid in $\mathbb{R}_{2,2}$. Take $\mathbb{R}_{2,2}$ with metric

$$ds^2 = -dX_0^2 - dX_1^2 + dX_2^2 + dX_3^2.$$

Then $AdS_3$ is embedded as

$$X_0^2 + X_1^2 - X_2^2 - X_3^2 = \ell^2, \tag{5.1}$$

where $\ell$ is a length scale, the AdS radius of curvature.[1]

Observe that the embedding condition 5.1 for $AdS_3$ is equivalent to the condition

$$\det \begin{pmatrix} X_0 + X_2 & X_3 - X_1 \\ X_3 + X_1 & X_0 - X_2 \end{pmatrix} = \ell^2.$$

Taking units where $\ell = 1$[2], we see that $AdS_3$ can be identified with $SL(2,\mathbb{R})$. In this language, the metric becomes the group-invariant metric

$$ds^2 = -\ell^2 \operatorname{tr}\left(g^{-1}dg\right)^2.$$

The space of isometries of $AdS_3$ is $SL(2,\mathbb{R}) \times SL(2,\mathbb{R})$, where two group elements $h, \tilde{h} \in SL(2,\mathbb{R})$ act on $g \in SL(2,\mathbb{R})$ as $g \to hg\tilde{h}$.

The Einstein equation in $2+1$-d specifies that any spacetime has constant curvature. Therefore, in 2+1 dimensions, any asymptotically AdS space will be locally AdS everywhere. Despite the lack of local degrees of freedom, these spaces are hardly boring! They have global degrees of freedom. This becomes evident when we consider the action of isometries on $AdS_3$. We can construct nontrivial locally $AdS_3$ spaces by taking quotients by isometry groups,

$$g \sim hg\tilde{h} \tag{5.2}$$

for $g \in SL(2,\mathbb{R}) \sim AdS_3$, $(h, \tilde{h}) \in SL(2,\mathbb{R}) \times SL(2,\mathbb{R})$.

Choosing one element of the isometry group produces a spacetime with one defect, but we may create more defects by choosing more elements of $SL(2,\mathbb{R}) \times SL(2,\mathbb{R})$, which we will denote by $G$ in

[1] This embedding has closed timelike curves. Thus we need to unwrap the time direction by taking the universal cover. It is technically this space that we refer to as $AdS_3$. We will generally suppress this technicality.

[2] A note about units: we will generally use Planck units rather than taking $\ell = 1$; however, it is clearly still the case that $AdS_3 \simeq SL(2,\mathbb{R})$.

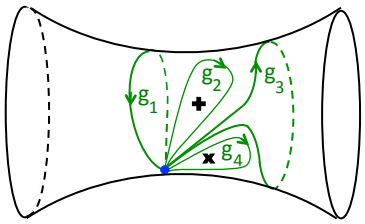



Figure 5.2: An example of a geometry with two asymptotic regions and two punctures which is parametrized by four holonomies $g_i \in SL(2,\mathbb{R})$ which satisfy $g_1 g_2 g_3 g_4 = 1$, modulo overall conjugation.

the following. In order to specify a spacetime more generally, fix a spacetime manifold of interest $\mathcal{M}$ with fundamental group $\Pi_1(\mathcal{M})$. Then to define the geometry, one must choose an element $g_i \in G$ for each nontrivial cycle $\gamma_i$ of $\mathcal{M}$. Let the group generated by the $\{g_i\}$ be denoted by $\mathrm{Hol}_*(\mathcal{M}, G)$. The spacetime is characterized by a homomorphism $\phi : \Pi_1(G) \to \mathrm{Hol}_*(\mathcal{M}, H)$. Gauge transformations act by conjugation of the elements of the holonomy group. The holonomies are thus defined only up to overall conjugation.

Physically, each nontrivial holonomy sources matter within the bulk theory. Of course, the properties of the matter determined by each holonomy can only depend on properties of the associated group element's conjugacy class. Define elliptic, parabolic, and hyperbolic conjugacy classes $\mathcal{C}$ of $\mathrm{SL}(2,\mathbb{R})$ to have $h \in \mathcal{C}$, $|\mathrm{tr}(h)| < 2, |\mathrm{tr}(h)| = 2$, and $|\mathrm{tr}(h)| > 2$ respectively. Geometrically, if $h$ and $\tilde{h}$ are both elliptical elements of $\mathrm{SL}(2,\mathbb{R})$, then they create a conical singularity on the spatial slice, whereas if they are both hyperbolic, they create a black hole. This is depicted in fig. 4.3 in chapter 4. To give more explicit expressions, for a black hole, if we denote the mass by $M$, the angular momentum by $J$, and the inner and outer horizons by $r_\pm$ respectively, we have

$$(h, \tilde{h}) = e^{\pm \pi \sigma_2 (r_+ + r_-)/\ell} \tag{5.3}$$

$$g \sim h g \tilde{h} \tag{5.4}$$

$$8\ell^2 M = r_+^2 + r_-^2 \tag{5.5}$$

$$4\ell J = r_+ r_- \tag{5.6}$$

and for a spinless point particle, we have $8\ell m < 1$, such that the conical deficit angle is given by

$$\frac{\theta}{2\pi} = \sqrt{1 - 8\ell m}. \tag{5.7}$$

We emphasize that in this picture, we can very explicitly specify the relationship between physical

parameters of the objects of interest and the topological holonomies that define the spacetime. Such a spacetime is demonstrated in fig. 5.2, which is directly relevant for our scattering discussion, as there are two punctures (particles) interacting in a black hole background (hyperbolic space).

In order to provide intuition for the upcoming connection with Teichmüller space, we will make a brief connection with the metric formulation of $2+1$-d gravity. The metric of a $2+1$-d spacetime can be expressed in terms of two $\mathrm{SL}(2,\mathbb{R})$ connections $A^a_\pm = e^a \pm \omega^a$ where $e^a$ is the dreibein and $\omega^a = \frac{1}{2}\omega^a_b \eta^b$ is the spin connection. These are the variables of interest in the Chern-Simons formulation of $2+1$-d gravity. The equations of motion dictate that these connections be flat. Their only degrees of freedom are global, and are in fact given by $\mathrm{SL}(2,\mathbb{R})$ holonomies such as $h, \tilde{h}$ in identifications of the form in equation 5.2.

## 5.3 Scattering in a black hole background

In this section, we review the computation, originally performed in [115], of the time of escape of an outgoing particle after interacting with an incoming particle near the horizon of a black hole. This computation was found to have an exponential dependence on the relevant energies, and demonstrates that the outgoing particle can get "pushed" behind the black hole horizon as a result of the interaction, never reaching the boundary. This calculation, along with previous suggestive work [121, 122], is indicative of the deep connection between gravitational scattering and dynamics in the CFT. Of course, this is the deep connection we will expand upon later in our discussion.

We will work in a black hole background in $2+1$-d. The metric of a nonspinning BTZ black hole of mass $M$ and Schwarzschild radius $R$ can be written as

$$
\begin{aligned}
ds^2 &= \frac{1}{f(r)} dr^2 - f(r) dt^2 + r^2 d\theta^2 \\
f(r) &= \frac{r^2 - R^2}{\ell^2} \\
R^2 &= 8M\ell^2.
\end{aligned}
$$

Its entropy is given by the Bekenstein-Hawking entropy $S_{BH} = \frac{2\pi R^2}{4}$, the geodesic length of its horizon. [89, 90, 94, 95].

Consider the Penrose diagrams depicted in figure 5.3. The figure on the left shows a double horizontal line at early time, depicting that we take as initial conditions the black hole masses $M, M+\omega$, the masses and spins (equal to zero) of particles $a$ and $b$, the time $t_b$ at which $b$ falls in,

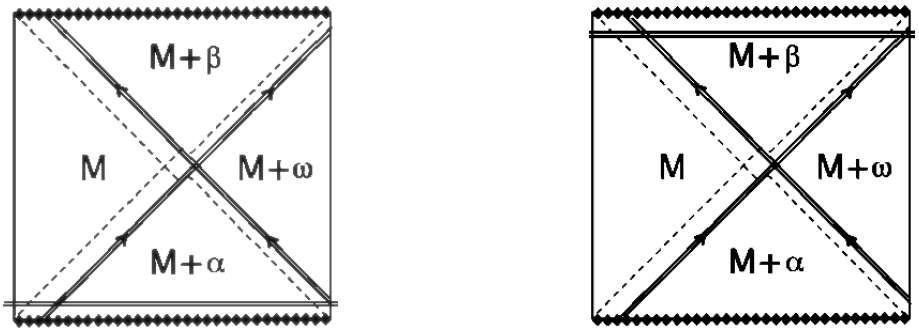



Figure 5.3: On the left, we fix initial conditions in the past, fixing the time $t_b$ as well as $M, \alpha, \omega$. The horizontal double lines indicate the setting of initial conditions at early time. We time evolve forward, measuring the new time $t'_a$ at which $a$ reaches the boundary. In the text, we denote $t_\alpha = t'_a - t_b$. The image on the right sets initial conditions in the future (see double horizontal lines), fixing $t_a, \beta, \omega, M$, and time evolves backwards. The new time at which $b$ falls in is $t'_b$, and the difference $t_\beta = t_a - t'_b$ in the text. These are slight modifications of a figure from [115].

and the energy of particle $a$ as initial conditions. The key physical idea that we are capturing is that particle $a$ and particle $b$ interact very close to the black hole horizon. As a result, this interaction has an exponentially large effect on boundary physics, and in particular, the time at which $a$ hits the boundary changes as a result of the interaction. We denote this new time $t'_a$ and define $t_\alpha = t'_a - t_b$, as shown.

The right image in figure 5.3 is similar except here, we fix the late-time conditions and time evolve backwards. In particular, we fix $t_a$, not $t_b$. In analogy to the previous case, the time at which $b$ fell in from the boundary changes as a result of the interaction with $a$. We denote this new time as $t'_b$ and define $t_\beta = t_a - t'_b$.

To clarify the remaining notation: $\alpha$ is equal to the energy of particle $a$ when it leaves the black hole, $\beta$ equals the energy of particle $b$ when it falls into the black hole, after interacting with $a$, $\omega$ equals the total energy, $M$ is the mass of the black hole without the presence of any energetic particle. In the quantum case, these are not simultaneously well-defined. We will explain this later in the discussion.

We begin with a classical computation relating $\alpha, \beta, \omega, t_\alpha, t_\beta$. It merely relies on the fact that spacetimes in $2+1$-d are pieces of AdS glued in the appropriate manner.

The following discussion largely follows the computation in [116, 114]. In order to construct the geometry, we will glue together four black hole geometries of the appropriate masses. This perspective makes it clear that the full geometry (and in particular the times $t_a, t'_a, t_b, t'_b$) is determined by $M, M+\alpha, M+\beta, M+\omega$.

To carry out the gluing, we impose two conditions. The first is just continuity of the spacetime at

the horizon. The second is that the two particle collide near the black hole horizon, which imposes the constraint that we are in the gravity-dominated regime and, on the CFT side, we can ignore contributions from descendants.[3] Let $r$ be the $r$-coordinate of the collision between the two particles. The gluing condition gives us the constraint

$$\left(r^2 - 8(M+\alpha)\ell^2\right)\left(r^2 - 8(M+\beta)\ell^2\right) = \left(r^2 - 8(M+\omega)\ell^2\right)\left(r^2 - 8M\ell^2\right). \tag{5.8}$$

Moreover, in the regime where the collision takes place very close to the horizon, the distance between $r$ and the two Schwarzschild radii $R_\alpha, R_\beta$ scales exponentially with $t_\alpha$ and $t_\beta$ [115]:

$$r^2 - 8(M+\alpha)\ell^2 = 4\ell^2 e^{-\kappa(t_\alpha - t_R)} \tag{5.9}$$

$$r^4 - 8(M+\beta)\ell^2 = 4\ell^2 e^{-\kappa(t_\beta - t_R)}. \tag{5.10}$$

Here, $\ell$ is the AdS radius and we define $\kappa$ to be the surface gravity and $t_R$ the time delay:

$$\kappa = R/\ell^2 \tag{5.11}$$

$$\kappa t_R = \log(R^2/\ell^2). \tag{5.12}$$

Together, these two equations imply the following relations:

$$\beta = \omega - \alpha + 2\alpha(\omega - \alpha)e^{\kappa(t_\alpha - t_R)} \tag{5.13}$$

$$\alpha = \omega - \beta + 2\beta(\omega - \beta)e^{\kappa(t_\beta - t_R)}. \tag{5.14}$$

Observe that for $t_\alpha$ sufficiently large, $\beta$ becomes larger than $\omega$, implying that there is no real solution for $t_\beta$. This means that the particle $b$ pushes particle $a$ behind the horizon. Equations 5.13, 5.14 will be the key classical results underlying the remainder of the discussion.

We now consider another semiclassical perspective in which particles $A$ and $B$ are spherical wavepackets $\phi_A$ and $\phi_B$ with well-localized energies and positions. This will be useful for comparison with later discussion. We will show that the Hamilton-Jacobi equations similarly imply an exponential relation between energy and time, demonstrating the same sensitivity to initial condi-

[3] Physically, this is true because the descendants derive solely from the Virasoro symmetry, which describes asymptotic symmetries near the boundary[100]. Moreover, the entropy of the black hole dominates the entropy generated by Virasoro symmetries in this regime.

tions implied by eqs. 5.13 and 5.14.

Take $\phi_A$ and $\phi_B$ to be second-quantized mode operators. We expect them not to commute, but rather to satisfy an exchange relation [123, 124]

$$\phi^B_{\omega-\alpha}(t_1)\phi^A_\alpha(t_0) = e^{\frac{i}{\hbar}S_{\alpha\beta}}\phi^A_{\omega-\beta}(\tilde{t}_0)\phi^B_\beta(\tilde{t}_1) \tag{5.15}$$

By similar computations to the previous section, we find relations for the time shifts,

$$\tilde{t}_0 - t_0 = -\frac{1}{\kappa}\log\left(\frac{\omega-\beta}{\alpha}\right) \tag{5.16}$$

$$\tilde{t}_1 - t_1 = -\frac{1}{\kappa}\log\left(\frac{\omega-\alpha}{\beta}\right) \tag{5.17}$$

Observe that, as before, the time delay $\tilde{t}_0 - t_0$ becomes infinite as $\beta$ approaches $\omega$.

We can solve for the dependence of the scattering phase $e^{\frac{i}{\hbar}S_\alpha}$ on $\alpha$ and $\beta$ using the Hamilton-Jacobi equations,

$$t_\alpha = t_0 - t_1 = -\frac{1}{\hbar}\frac{\partial S_{\alpha\beta}}{\partial\alpha} \tag{5.18}$$

$$t_\beta = \tilde{t}_0 - \tilde{t}_1 = \frac{1}{\hbar}\frac{\partial S_{\alpha\beta}}{\partial\beta}. \tag{5.19}$$

These expressions derive from the consistency of eqs. 5.15 upon varying both sides by $\alpha$ and $\beta$, respectively.

We can integrate eqs. 5.18 using eqs. 5.16 to solve for $S_{\alpha\beta}$. Writing $t_\alpha = t_0 - t_1$ and $t_\beta = \tilde{t}_1 - \tilde{t}_0$, using eqs. 5.16 to express in terms of $\omega, \alpha, \kappa$ and $\beta$, and introducing the parameter $t_R$ as in 5.11, we find

$$t_\alpha - t_R = \frac{1}{\kappa}\log\left(\frac{\alpha+\beta-\omega}{2\alpha(\omega-\alpha)}\right) \tag{5.20}$$

$$t_\beta - t_R = \frac{1}{\kappa}\log\left(\frac{\alpha+\beta-\omega}{2\beta(\omega-\beta)}\right) \tag{5.21}$$

Using eqs. 5.20 to integrate eqs. 5.18, we find that, provided that $\omega > \alpha, \beta > 0$, and $(\alpha+\beta) > \omega$,

$$\frac{1}{\hbar}S_{\alpha\beta} = \frac{1}{\kappa}\{\alpha\log\alpha + \beta\log\beta - (\omega-\alpha)\log(\omega-\alpha) - (\omega-\beta)\log(\omega-\beta) - (-\alpha+\beta-\omega)\log(\alpha+\beta-\omega)\} - (\alpha+\beta-\omega)t_R \tag{5.22}$$

The $\omega > \alpha, \beta > 0$ and $(\alpha+\beta) > \omega$ regime is precisely the regime where there exists a real trajectory.

We thus find that the gravity theory also demonstrates an exponential sensitivity in time dependence deriving from the phase shifts. We will see in the following sections that we can lift this to a description in terms of quantum geometry and nonrational conformal field theory to see that the exponential dependence derives from the representation theory of a noncompact quantum group, $U_q(SL_2(\mathbb{R}) \times SL_2(\mathbb{R}))$.

## 5.4 Teichmüller space and the Hilbert space of conformal blocks

We wish to claim that the scattering process we have described thusfar is intimately related to topological properties of both the spacetime and its corresponding conformal field theory. To do so, we discuss the deep connection between the geometry of constant negative curvature spaces and conformal field theory by looking at the properties of Teichmüller space. Teichmüller space is defined as the space of constant negative curvature metrics on a given Riemann surface (with a specified number of cusps and holes). It is related to $\mathrm{SL}(2,\mathbb{R})$ through uniformization, which states that any Riemann surface can be constructed as a quotient of the upper half plane by a discrete subgroup of $\mathrm{SL}(2,\mathbb{R})$ (the "Fuchsian group"). This Fuchsian group is precisely the holonomy group of the surface.

Consider a spacetime with 4 defects (cusps + holes). Our next step is to quantize Teichmüller space $T_{0,k}$ in order to study the Hilbert space of $2+1$-d gravity. The statement is that the resulting Hilbert space is spanned by these Virasoro conformal blocks with projected internal channels, For a review of CFT and conformal blocks in 2d, see appendix 5.A.

First recall that the space of conformal blocks is the span of a linear space of solutions to the conformal Ward identity [92, 93], which states that conformal blocks $\langle \mathcal{O}_1(z_1) \dots \mathcal{O}_4(z_4) \rangle$ satisfy

$$\langle T(x) \mathcal{O}_1(z_1) \dots \mathcal{O}_4(z_4) \rangle = \hat{T}(x) \langle \mathcal{O}_1(z_1) \dots \mathcal{O}_4(z_4) \rangle \tag{5.23}$$

where

$$\hat{T}(x) = \sum_{i=1}^{4} \frac{\Delta_i}{(x-z_i)^2} + \frac{1}{x-z_i} \frac{\partial}{\partial z_i}. \tag{5.24}$$

(More generally, "4" can be replaced with $k$, the number of defects.)

This is related to geometry as follows. We define an ansatz classical stress-energy tensor $T(x)$ to

be of the form

$$T(x) = \sum_{i=1}^{4} \frac{\Delta_i}{(x - z_i)^2} + \frac{c_i}{x - z_i} \tag{5.25}$$

where the $c_i$ are complex numbers ("accessory parameters") and the $\Delta_i$ are real numbers. We wish to construct an operator $\hat{T}$ to satisfy the expression

$$\langle T(x) \rangle = \frac{\langle T(x)\mathcal{O}(z_1)\dots\mathcal{O}(z_4)\rangle}{\langle \mathcal{O}(z_1)\dots\mathcal{O}(z_4)\rangle} \tag{5.26}$$

in a CFT of central charge $c >> 1$.

The remainder of the line of reasoning which connects the geometry of spaces of constant negative curvature with the Virasoro bootstrap is as follows. The derivation begins with classical stress-energy tensors of the form given in equation 5.25. The phase space of Teichmüller space on a surface with four defects can be identified with stress tensors $T(z)$ of the form 5.25 satisfying expression 5.26. Said phase space is parametrized by the coordinates of the defects, $z_i$, and the accessory parameters, $c_i$. So the space of these stress tensors can be identified with the space of constant negative curvature metrics on the surface.

After defining a symplectic form on phase space, the coordinates $z_i$ and the parameters $c_i$ are canonical conjugates, and so take $c_i = \frac{\partial}{\partial z_i}$. It follows that the stress tensor 5.25 takes a form that exactly matches that of equation 5.24. Equation 5.26 implies that $\langle \mathcal{O}_1(z_1)\dots\mathcal{O}_4(z_4)\rangle$ satisfies the conformal Ward identity. The quantum $T(z)$s are therefore in one-to-one correspondence with the solutions to the conformal Ward identity, and so the Hilbert space can be viewed as correlation functions $\langle \mathcal{O}_1(z_1)\dots\mathcal{O}_4(z_4)\rangle$ satisfying the conformal Ward identity - that is, Virasoro conformal blocks, whose associated CFT is Liouville theory.

The conclusion is that Virasoro conformal blocks give the Hilbert space resulting from quantizing Teichmüller space. We have also seen that Teichmüller space is isomorphic to $\mathrm{SL}(2,\mathbb{R})$, and that the space of constant $2+1$-d metrics can be described as the space of $\mathrm{SL}(2,\mathbb{R}) \times \mathrm{SL}(2,\mathbb{R})$ holonomies. After quantizing, there is a direct correspondence between locally $\mathrm{AdS}_3$ spacetimes and the space of Virasoro conformal blocks.

Next, we would like to say a few words about the physical interpretation of conformal blocks in terms of bulk geometry.

Recall that defining a $2+1$-d geometry with constant negative curvature amounts to choosing a holonomy group $H \subset \mathrm{SL}(2,\mathbb{R}) \times \mathrm{SL}(2,\mathbb{R})$, a subgroup of the isometry group of $\mathrm{AdS}_3$, generated

by group elements that define nontrivial holonomy. A geometry is then constructed as a quotient, $\mathrm{AdS}_3/H$.

Quantum mechanically, we expect that the nontrivial holonomy of a connection is roughly associated with a representation of a gauge group. In this case, after quantizing, the relevant structure is a quantum group, $U_q\left(\mathfrak{sl}_2(\mathbb{R}) \times \mathfrak{sl}_2(\mathbb{R})\right)$, whose representations are associated with primary operators in the spectrum of nonchiral Liouville theory.

Consider a geometry with $k$ defects (cusps + holes). We claim that each defect is associated with a representation as described above. If group elements are operators after quantization, what is group multiplication? The most natural thing is the tensor product, which is expanded as a direct sum of representations - that is, the OPE. We obtain an simple, if informal, interpretation of a geometry after quantization. A given (quantum) spacetime is equivalent to a conformal block with the Liouville operators associated with defects. It can be written as an expansion in terms of conformal blocks with the internal line projected to a given channel. *Each* of these *projected* conformal blocks can be thought of as a *classical* geometry, with the projected channels analogous the (classically) well-defined multiplication within the holonomy group.

This was briefly demonstrated in chapter 4, fig. 4.2.

CFT conformal blocks exhibit structure according to the conformal bootstrap and including braiding symmetries exhibited by operations like $F$ and $R$, which are reviewed in appendix 5.A. Of primary importance to us in the following sections is the topological, braiding properties of a 2-d CFT. What do the bootstrap constraints tell us about geometry? It would seem that they indicate that we cannot specify different channels independently, and dictate how to write one such specification in terms of others. In particular, they somehow indicate "quantum" relationships among "classical" geometries, if we take the geometries with projected inner channels to be classical geometries, and also allow us to change bases within this space given these mathematical relationships.

In summary, the spectrum of conformal dimensions defines the possible holonomies around the defects; the OPE defines how a given geometry is expanded in terms of classical geometries; the braiding and fusion indicate the interdependence of different classical interpretations. These interpretations not rigorous as presented here, given the many challenges with defining objects like superpositions of geometries, but they give a physical intuition for how Liouville theory and representation theory can correspond to a useful notion of quantum geometry.

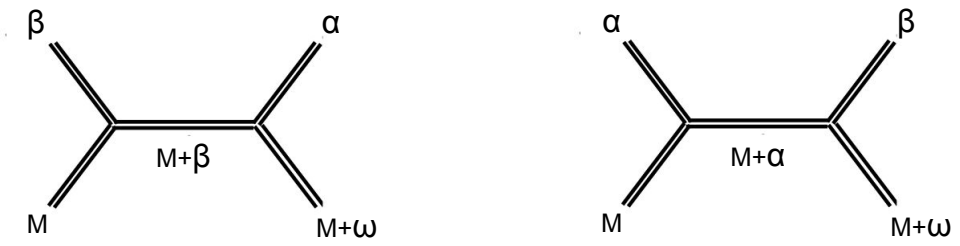



Figure 5.4: Conformal blocks representing the initial and final states whose overlap represents the scattering amplitude of two particles of energy $\alpha$ and $\beta$ and total energy $\omega$ near the horizon of a black hole of mass $M$. Compare to the classical spacetimes depicted in fig. 5.3.

## 5.5 Scattering, $\mathcal{R}$, and CFT exchange algebra

### 5.5.1 Braiding relations and scattering

We use our formalism relating states in the gravity theory to Virasoro conformal blocks to compute the scattering amplitude of interest using CFT tools. We then demonstrate strong evidence that this computation indeed has the correct physical interpretation by taking the classical $\hbar \to 0$ ($c \to \infty$) limit, whereby we obtain expressions eqs. 5.13 and 5.14 as needed.

The first step will be to convert the gravity problem to a problem in Liouville theory by converting the geometric calculation to a calculation of the overlap between two states represented as conformal blocks in the CFT. For notational convenience, we begin by using notation which applies to theories with discrete spectra; we then discuss the new challenges which arise when the spectrum is actually continuous. To this end, we write down two conformal blocks, one representing the initial state, with fixed energy $\beta$ of the incoming particle, and one representing the outgoing state, with fixed energy $\alpha$ of the incoming particle. We fix the mass $M$ of the black hole and the combined energy $\omega$ of incoming and outgoing particles. The classical gravity setup we intend to model and generalize is shown in fig. 5.3. The conformal blocks we will use to model the initial and final states are shown in fig. 5.4.

Recall that in the universal gravity regime modeled by Liouville theory, black holes and particles are represented by hyperbolic and elliptic operators acting on the spacetime, respectively. These create holes and punctures in the geometry. To create a black hole of mass $M$ and spin 0, we need an operator of left and right conformal dimensions

$$\Delta = \overline{\Delta} = \frac{M}{2}. \tag{5.27}$$

We take the energies $\alpha$ and $\beta$ to be the conformal dimensions of the operators dual to particles exiting and entering the black hole, respectively, as shown in the gravity theory in fig.5.3.

In CFT language, we would like to find the overlap between the initial and final states represented by conformal blocks as in fig. 5.4. Denote the two states as $\big|\beta\rangle$ and $\big|\alpha\rangle$. We know that there exists an $R$ matrix which allows us to expand the $\big|\alpha\rangle$ state in terms of other states,

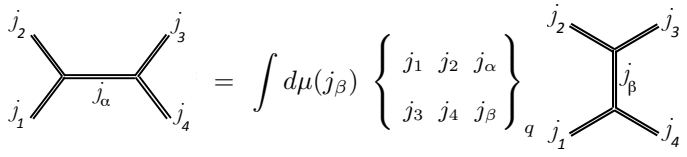
$$\big|\alpha\rangle = \mathcal{R}_{\alpha\gamma}\big|\gamma\rangle \tag{5.28}$$

by definition of the $\mathcal{R}$ matrix and the crossing equation. (Here, we've suppressed the $M$ and $M+\omega$ indices, which remain fixed.) The overlap between the two states $\big|\alpha\rangle$ and $\big|\beta\rangle$ is given by the relevant $\mathcal{R}$ matrix element,

$$\langle\beta\big|\alpha\rangle = \mathcal{R}_{\alpha\beta}. \tag{5.29}$$

In principle, this suggests a "recipe" for computing scattering amplitudes in the universal regime of 2 + 1d gravity, where braiding relations apply: define the appropriate conformal blocks, and compute the relevant element of the braiding matrix within the CFT. However, the situation is hardly so simple. First, as we indicated above, we've used notation relevant for theories with discrete spectra, but, as we know, Liouville theory has a continuous spectrum with density defined by the Plancherel measure. This means the $R$ "matrix" is not a matrix, but rather an operator; a generalization of the braiding matrices of CFTs with discrete spectra.

For Liouville theory, the operator analogous to the $\mathcal{R}$ matrix is given by the 6j symbol associated with the representation theory of $U_q(\mathfrak{sl}(2,(\mathbb{R}))$. Via the isomorphism between the tensor algebra of the Virasoro algebra and $U_q(\mathfrak{sl}(2,\mathbb{R}))$, one can show [120, 118] that duality transformation of the Virasoro conformal blocks are specified by the quantum 6j-symbols




(5.30)

where $d\mu(j_\beta)$ is the Plancherel measure as discussed previously. In other words, we can explicitly

write the $\mathcal{R}$ matrix as follows,

$$\mathcal{R}_{\alpha\beta}^{U_q(\mathfrak{sl}_2)} \;=\; e^{\pi i(\Delta_1+\Delta_3-\Delta_\alpha-\Delta_\beta)} \begin{Bmatrix} j_1 & j_2 & j_\alpha \\ j_3 & j_4 & j_\beta \end{Bmatrix}_q , \tag{5.31}$$

with:

$$q \;=\; e^{i\hbar} \;=\; e^{i\pi b^2}, \qquad\qquad \ell M_i \;= 2\Delta_i = \; 2j_i(Q-j_i). \tag{5.32}$$

Here $j_i \in [0, \frac{1}{2}Q] \cup \frac{1}{2}Q + i\mathbb{R}_+$ labels the representation of $U_q(\mathfrak{sl}(2,\mathbb{R}))$, and, as before, $Q$, $b$, $\hbar$ and $\ell$ are all related and determined by the central charge $c$ via

$$c \;=\; 1+6Q^2, \qquad Q \;=\; b+b^{-1}, \qquad \hbar = \frac{4\pi}{\ell}, \qquad c = \frac{3\ell}{2}. \tag{5.33}$$

Indeed, the quantum 6j symbols form a unitary matrix which partially defines the conformal bootstrap and encodes topological properties of Liouville theory and the Virasoro bootstrap. Complete expressions for the quantum 6j symbols are known [120]. They are expressed in terms of the double sine function and the quantum dilogarithm; the complete expressions for the quantum 6j symbols are given in appendix 5.B. Although the expressions are rather complicated, one significant point is that in the semiclassical limit $b \to 0$, eq. 5.64 simplifies and takes on a geometric interpretation,

$$\begin{Bmatrix} j_1 & j_2 & j_\alpha \\ j_3 & j_4 & j_\beta \end{Bmatrix}_q \;\simeq\; \exp\Big\{ \frac{i}{2\pi b^2} \mathrm{Vol}(\mathrm{T}) \Big\} \tag{5.34}$$

where $T$ is the hyperbolic tetrahedron with six dihedral angles specified by $\nu_i = 2\pi b^2 j_i$. The details of this calculation are found in [118]. An explicit expression for $\mathrm{Vol}(\mathcal{T})$ can be found in eq. 5.60 in appendix 5.B.

We'd like to use the quantum 6j symbols to compute the scattering amplitude from sec. 5.3 in the regime described by Liouville theory. In order to do so, we need to go from 6j symbol to braiding matrix. In the kinematical regime we are interested in, where we set the masses of particles $A$ and $B$ to 0 and define the energies $M$, $\omega$, $\alpha$ and $\beta$ as before, this relationship is explicitly given by the exact Liouville expression,

$$\mathcal{R}_{\alpha\beta} \;=\; e^{\pi i(\Delta_\omega-\Delta_\alpha-\Delta_\beta)} \begin{Bmatrix} j_M & j_\omega & 0 \\ j_\alpha & j_\beta & 0 \end{Bmatrix}_b . \tag{5.35}$$

Taking the semi-classical limit involves converting a question about scattering into a question about braiding and hence geometry. In the previous discussion, we learned that the dictionary is given by

$$\mathcal{R}^2_{\alpha\beta} \;=\; \exp\big(\tfrac{i}{\hbar}S_{\alpha\beta}\big), \qquad\qquad S_{\alpha\beta} = \mathrm{Vol}_{\mathcal{T}}(l_M, l_\omega, l_\alpha, l_\beta) \tag{5.36}$$

$$l_\omega = l_M + \hbar\omega/\kappa, \qquad\qquad l_\alpha = l_M + \hbar\alpha/\kappa, \qquad\qquad l_\beta = l_M + \hbar\beta/\kappa \tag{5.37}$$

$$\hbar = \pi b^2 = 4\pi/\ell, \qquad\qquad l_M/2\pi = R/\ell = \kappa\ell \qquad\qquad R^2 = 8M\ell^2. \tag{5.38}$$

In the $\hbar \to 0$ limit, $\mathcal{T}(l_M, l_\omega, l_\alpha, l_\beta)$ approaches an ideal tetrahedron for four identical and two trivial tetrahedral angles. Using the discussion in appendix 5.B to evaluate the volume of the hyperbolic tetrahedron, one can extract the leading order $\hbar \to 0$ limit to find that

$$\begin{aligned}\frac{1}{\hbar}S_{\alpha\beta} \;\simeq\; \frac{1}{\kappa}\Big\{ &\alpha\log\alpha + \beta\log\beta - (\omega-\alpha)\log(\omega-\alpha) - (\omega-\beta)\log(\omega-\beta) \\ &- (\alpha+\beta-\omega)\log\Big(2\sinh\big(\tfrac{\pi R}{\ell}\big)(\alpha+\beta-\omega)\Big)\Big\}.\end{aligned} \tag{5.39}$$

Observe that this precisely the scattering phase derived earlier from geometry in eq. 5.22, up to a shift in the time delay. Thus scattering near a black hole has given us an example where there indeed exists a gravitational regime of $\mathrm{AdS}_3/\mathrm{CFT}_2$ in which physical quantities can be computed using known results in Liouville theory.

We have learned that scattering and sensitivity to initial conditions in $2+1$d gravity are encoded in the rich topological braid group structure built in to the Virasoro bootstrap. In the end, the computation demonstrated that the scattering matrix for the given process is exactly determined by the braid matrix, and is given by the $\mathcal{R}$ matrix of the conformal bootstrap.

There is, however, another crucial observation to the story. Clearly the $\mathcal{R}$ matrix does not just describe the result of the given computation; it is intrinsic to the very definition of the CFT. This demonstrates that the near-horizon gravitational interactions contain detailed information about the gravity-dominated regime of irrational CFT, and vice versa: the CFT exchange algebra in this regime determines the near-horizon physics described here. This connection has been previously studied, see [121, 122, 123, 124], but the current perspective provides a very direct manifestation using the bootstrap. Indeed, this endows the gravitational scattering experiment described above with a powerful interpretation in terms of the CFT dictionary.

Given all of this generality, it is somewhat strange that thusfar our "scattering" did not explicitly involve Lorentzian time. We remedy this in the following section.

### 5.5.2 Exchange relations and Lorentzian time

In this section, we will demonstrate that the $\mathcal{R}$ matrix determines the Lorentzian exchange algebra for timelike separated operators. The arguments closely follow those in [117, 114].

Fix Lorentzian signature and choose lightcone coordinates $u = x+t, v = x-t$. Let $\hat{A}(0,0), \hat{B}(u,v)$ be two nonchiral operators in the CFT mapping the sector with energy $M$ to the sector with energy $M+\omega$. We will first show that the conformal bootstrap constraints are equivalent to the constraint of locality. Take $uv > 0$ and consider the commutator

$$\Big[\, \hat{B}(u,v), \hat{A}(0,0) \,\Big] = 0, \tag{5.40}$$

which must vanish for spacelike separation in a local theory.

Upon decomposing into chiral components, this becomes a nontrivial constraint. Namely, let $A_\alpha(0)$ be a map from the chiral sector with energy $M$ to that with $M+\omega$, and similarly for $\overline{A}_\alpha(0)$, $B_\beta(u)$, $\overline{B}_\beta(v)$. We write

$$\hat{A}(0,0) = \sum_\alpha \overline{A}_\alpha(0) A_\alpha(0) \tag{5.41}$$

$$\hat{B}(u,v) = \sum_\beta \overline{B}_\beta(v) B_\beta(u). \tag{5.42}$$

Define

$$\Psi_\alpha(u) = \langle M+\omega \big| B_{\omega-\alpha}(u) A_\alpha(0) \big| M \rangle \tag{5.43}$$

$$\Phi_\beta(u) = \langle M+\omega \big| A_{\omega-\beta}(0) B_\beta(u) \big| M \rangle. \tag{5.44}$$

The commutator between $\hat{A}$ and $\hat{B}$ is written in terms of chiral vertex operators using the following expressions:

$$\langle M+\omega \big| \hat{B}(u,v)\hat{A}(0,0) \big| M \rangle = \sum_\alpha \overline{\Psi}_\alpha(v)\Psi_\alpha(u) \equiv \big(\overline{\Psi}, \Psi\big) \tag{5.45}$$

$$\langle M+\omega \big| \hat{A}(0,0)\hat{B}(u,v) \big| M \rangle = \sum_\beta \overline{\Phi}_\beta(v)\Phi_\beta(u) \equiv \big(\overline{\Phi}, \Phi\big) \tag{5.46}$$

The requirement of locality translates to

$$\left(\overline{\Psi}, \Psi\right) = \left(\overline{\Phi}, \Phi\right). \tag{5.47}$$

But observe the form of $\Psi_\alpha$ and $\Phi_\beta$. Specifically, they can be related by an $\mathcal{R}$ matrix.

$$\Psi_\alpha(u) = \sum_\beta \mathcal{R}_{\alpha\beta}\Phi_\beta(u)$$
$$\overline{\Psi}_\alpha(v) = \sum_\beta \overline{\mathcal{R}}_{\alpha\beta}\overline{\Phi}_\beta(v).$$

Observe that $\overline{\mathcal{R}}^T = \mathcal{R}^\dagger$. Equation 5.47 implies that

$$\left(\overline{\Psi}, \Psi\right) = \left(\overline{\Psi}\mathcal{R}^\dagger, \mathcal{R}\Psi\right) = \left(\overline{\Phi}, \Phi\right) \tag{5.48}$$
$$\iff \mathcal{R}^\dagger\mathcal{R} = \mathrm{id} \tag{5.49}$$

In other words, locality of the exchange algebra is equivalent to unitarity of $\mathcal{R}$.

The spacelike separated case is interesting because it is ensures locality, but it is not essentially different from the Euclidean case. Timelike separation, however, provides a nontrivial exchange algebra which is fundamental to Lorentzian signature. As we will indicate presently, the $\mathcal{R}$ matrix does *not* cancel for timelike separated operators. Consider the setup above, but now take $uv < 0$ and restrict to the case where $\hat{A} = \hat{B}$. We are therefore considering

$$\left[\, \hat{A}(0,0), \hat{A}(u,v) \,\right]. \tag{5.50}$$

Observe that, in the notation of definitions 5.43, analytic continuation of $u \to -u$ sends $\Psi_\alpha(u)$ to $\Phi_\alpha(u)$. However, we must impose an $i\epsilon$ prescription in continuing through the lightcone. The usual $i\epsilon$ prescription is to approach the new value of $u$ via the upper half plane; thus, if $u > 0$, it approaches $-u$ through a counterclockwise route; if $u < 0$, it approaches clockwise. The two cases differ in terms of which $\mathcal{R}$ matrix is applied: if $\Psi$ transforms via $\mathcal{R}$ in the clockwise case, it will transform via $\mathcal{R}^{-1}$ in the counterclockwise case.

This fact holds more generally than just for $A = B$; it holds when they are unequal as well. We can use this to study the transformation properties of the original, nonchiral operators $\hat{A}(0,0), \hat{B}(u,v)$. Doing so just involves keeping track of the transformations of the various chiral pieces. In the end,

one obtains the expression

$$\hat{B}(u,v)\hat{A}(0,0) = \sum_{\omega,\beta,\delta} \mathcal{R}^{2\epsilon}_{\beta\delta}\overline{A}_{\omega-\beta}(0)A_{\omega-\delta}(0)\overline{B}_{\beta}(v)B_{\delta}(u) \tag{5.51}$$

The symbol $\epsilon = 1$ in the future lightcone and $\epsilon = -1$ in the past lightcone indicates whether the operator transforms with $\mathcal{R}$ or its inverse. Moreover, $\mathcal{R}^{2\epsilon}_{\beta\delta}$ is a shorthand for $\sum_{\alpha} \mathcal{R}^{\epsilon}_{\beta\alpha}\mathcal{R}^{\epsilon}_{\alpha\delta}$.

It follows that in Lorentzian signature, timelike separated, *nonchiral* operators have nontrivial exchange relations deriving from the conformal bootstrap. The conformal bootstrap acts nontrivially in Lorentzian signature. Specifically, the exchange relations are dominated by the same $\mathcal{R}$ matrix that served as the gravitational scattering matrix in the previous section! Understanding the near-horizon scattering physics is *equivalent* to understanding the exchange algebra of the Lorentzian CFT; thus the system is sensitive to time-dependence as well as braiding within the CFT. The two are linked by the $\mathcal{R}$ matrix.

## 5.6 Discussion

In the preceding discussions, guided by the notion of scattering close to a black hole horizon, we have developed a new aspect of the AdS/CFT dictionary in the case of $\mathrm{AdS}_3/\mathrm{CFT}_2$. Using the result from chapter 4 that the Plancherel measure of the spectrum of Liouville theory equals the Cardy formula in the high-energy regime, we have proposed a universal, continuum description of the high-entropy spectrum of nonrational CFTs. Passing to the continuum by using Liouville theory provides a powerful method for studying gravity-dominated physics in a $2+1$-d bulk gravity theory without knowing microscopic details of the complex, nonrational dual CFT. The primary tool linking the gravity description with the CFT is the Virasoro bootstrap, which allows us to identify states in the gravity theory with Virasoro conformal blocks.

We used Liouville theory to compute a scattering amplitude between two particles interacting near a black hole horizon. This computation relied on the representation-theoretic braiding structure which Liouville theory and its associated quantum group $U_q(\mathfrak{sl}(2,\mathbb{R}))$ exhibit. Despite the complicated expressions, however, in the $\hbar \to 0$ limit, the expression exactly matched computations from both classical gravity and from a Hamilton-Jacobi analysis using wavepackets. That the three computations agreed in the same limit indicates that the proposal of a braiding structure determining scattering dynamics in the bulk is physical. This serves as a powerful check of the formalism.

There are several interesting potential future directions. The two most interesting are likely to

be applications to black hole information, and perhaps generalizations to higher dimensions.

A high volume of research in recent years has focused on fundamental questions in black hole information theory. These questions largely reduce to asking, how much of the spacetime is the CFT able to probe? Does it know about physics behind the horizon? Our proposal concretely suggests a match between Virasoro conformal blocks and bulk physics in the gravity sector. Moreover, we have learned from our scattering example that near-horizon physics is deeply intertwined with the exchange algebra of the CFT. Perhaps this example suggests that there may be a sense in which black hole horizon physics dictates a universal regime which "averages" many irrational CFTs to produce something like Liouville theory in higher dimensions or other contexts. Our explicit proposal may shed light on what aspects of the spacetime are universal and which are not.

Of course, $\mathrm{AdS}_3/\mathrm{CFT}_2$ is very special on both sides of the duality. It would be somewhat disappointing if the results discussed here were true only because of its special properties without other applications. However, there are indications that this is not the case. First of all, the Cardy formula exists in all dimensions. Second, CFTs in all dimensions have exchange algebras thought to be connected with near-horizon gravitational interactions. What can we learn from these techniques in higher dimensions? Is there a universal CFT regime that describes gravity-dominated bulk physics? Is it possible to develop an explicit description of such a regime? It would be extremely interesting and instructive to do so.

By describing a universal regime of nonrational CFT with Liouville theory, we have obtained a novel, explicit description of quantum mechanics in the bulk of the gravity sector of AdS/CFT. A nontrivial physical check of the formalism demonstrates a tantalizing connection between near-horizon physics and the defining data of the boundary CFT. Perhaps the true remaining question is just, what else can we learn about the bulk using this explicit mapping between braid group and bulk physics?

## 5.A Brief review of 2-d CFT

Here, we review the basics of the conformal bootstrap, and specify notation and normalization conventions. For now, we ignore complications related to the continuous spectrum associated with the Virasoro bootstrap and assume a discrete spectrum. We address these issues in other parts of the text.

Consider a 2-d CFT, and adopt holomorphic spacetime coordinates $z, \bar{z}$. Recall that in 2-d, conformal transformations are defined by holomorphic functions: $z \to f(z)$, $\bar{z} \to \bar{f}(\bar{z})$, and therefore

the conformal group is infinite-dimensional. The conformal group is infinitesimally generated by the Virasoro algebra, with generators $\{L_n\}_{n\in\mathbb{Z}}$.

States in the CFT are organized into "towers," each defined by a primary operator $\mathcal{O}_i(z,\bar{z})$ with left- and right- conformal dimensions $\Delta_i, \overline{\Delta}_i$, respectively. The Hilbert space takes the form

$$\mathcal{H} = \oplus_{a,b} \left(\mathcal{H}_a \otimes \overline{\mathcal{H}}_b\right),$$

with the "bar" indicating the right-moving factor, and with $a, b$ labeling the left- and right- conformal dimensions of primaries. The notation is slightly misleading, as $a$ and $b$ are not independent. Quantization of spin implies that $a - b \in \mathbb{Z}$. Throughout our entire discussion, we restrict to the spinless sector, with $\Delta_a = \overline{\Delta}_b$, and therefore we abuse notation by writing $a$ and $\overline{a}$ as the left- and right- labels. The spinless assumption simplifies much of our discussion.

The spectrum of conformal dimensions is one of the defining properties of a CFT. Another is the coefficients of the operator product expansion. The OPE between two primary operators at different points is an expansion of the product of the two operators as a series of operators at one of the points:

$$\mathcal{O}_a(z,\bar{z})\mathcal{O}_b(0,0) = \sum_i \sum_{\{k\},\{\overline{k}\}} f^i_{ab} C_{\{k\},\{\overline{k}\}} z^{\Delta_i-\Delta_a-\Delta_b-\{k\}} \bar{z}^{\overline{\Delta}_i-\overline{\Delta}_a-\overline{\Delta}_b-\{\overline{k}\}} L_{\{k\}}\overline{L}_{\{\overline{k}\}}\mathcal{O}_i(0,0).$$

The sum over $i$ is a sum over primary operators, the $f^i_{ab}$ are OPE coefficients, and the $C_{\{k\},\{\overline{k}\}}$ are constants that depend only on Virasoro symmetry and not on the specific primary. The indices $\{k\}$ and $\{\overline{k}\}$ are sets of integers and they count descendants, formed by acting with Virasoro generators on primaries denoted $O_i$. (Subtracting $\{k\}$ is meant to denote subtracting the sum $\sum k$ and similarly for $\{\overline{k}\}$.)

We will define conformal blocks using the OPE as follows. Depict $\big|a,n\rangle$ as a shorthand for $\big|a,n,\overline{a},\overline{n}\rangle$, a descendant in the full CFT. Factor $f^a_{bc} \equiv \lambda^a_{bc}\overline{\lambda}^a_{bc}$ and $C_{mn} = \gamma_{mn}\overline{\gamma}_{mn}$. Let $\mathcal{O}_c(z,\bar{z})$ be a primary operator in Euclidean 2-d CFT, and use the state-operator correspondence to project out one term in the OPE:

$$\langle a,n\big|\mathcal{O}_c(z,\bar{z})\big|b,m\rangle = \lambda^c_{ab}\overline{\lambda}^c_{ab}\gamma_{mn}\overline{\gamma}_{mn} z^{\Delta_c-\Delta_a-\Delta_b+n-m}\bar{z}^{\overline{\Delta}_c-\overline{\Delta}_a-\overline{\Delta}_b+\overline{n}-\overline{m}}.$$

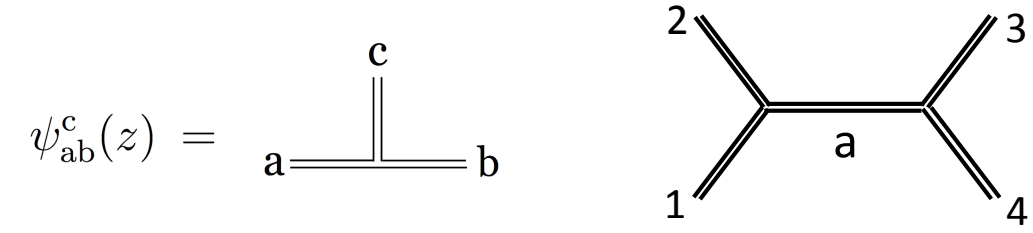



Figure 5.5: Left, a CVO; right, a conformal block $\Psi_a(w,z) = \langle 1 \big| \psi^2_{1a}(w) \psi^3_{a4(z)} \big| 4 \rangle$.

Next, define chiral objects

$$\langle a, n \big| \psi^c_{ab} \big| b, m \rangle(z) \equiv \lambda^c_{ab} \gamma_{mn} z^{\Delta_c - \Delta_a - \Delta_b + n - m}$$
$$\langle \overline{a}, \overline{n} \big| \overline{\psi}^c_{ab}(\bar{z}) \big| \overline{b}, \overline{m} \rangle \equiv \overline{\lambda}^c_{ab} \overline{\gamma}_{mn} \bar{z}^{\overline{\Delta}_c - \overline{\Delta}_a - \overline{\Delta}_b + \overline{n} - \overline{m}}$$

such that

$$\langle a, n, \overline{a}, \overline{n} \big| \mathcal{O}_c(z, \bar{z}) \big| b, m, \overline{b}, \overline{m} \rangle = \psi^c_{ab} \overline{\psi}^c_{ab}(\bar{z}).$$

Each piece of the OPE splits into chiral components. These objects are known as chiral vertex operators (CVOs), as denoted in fig. 5.5. Observe that, in our definition, chiral vertex operators know about the OPE coefficients. They therefore know of important defining info in the CFT- they are not theory-independent (as they would be in standard normalization that pulls the OPE coefficients out of the CVO).

The OPE coefficients cannot be arbitrarily specified. They are highly constrained by the demand that the OPE must be associative. For example, consider expanding the product

$$\mathcal{O}_1(x_1)\mathcal{O}_2(x_2)\mathcal{O}_3(x_3)\mathcal{O}_4(x_4). \tag{5.52}$$

We could expand in pairs as $\mathcal{O}_1(x_1)\mathcal{O}_2(x_2)$ and $\mathcal{O}_3(x_3)\mathcal{O}_4(x_4)$, or we could expand as $\mathcal{O}_1(x_1)\mathcal{O}_3(x_3)$, $\mathcal{O}_2(x_2)\mathcal{O}_4(x_4)$. Associativity dictates that these should give the same answer. It is instructive to write this condition by inserting the identity on both sides. Writing this schematically (but suggestively), we have

$$\sum_a \langle 1 \big| \mathcal{O}_2 \big| a \rangle \langle a \big| \mathcal{O}_3 \big| 4 \rangle = \sum_b \langle 1 \big| \mathcal{O}_3 \big| b \rangle \langle b \big| \mathcal{O}_2 \big| 4 \rangle.$$

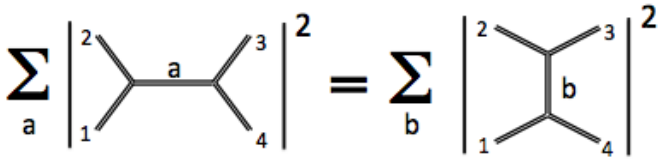



Figure 5.6: Associativity, the bootstrap, in theory-dependent normalization.

Noticing that each of these objects are just terms in the OPE as above, we expand in terms of CVOs,

$$\sum_a \left(\psi^2_{1a}\psi^3_{a4}\right)\left(\overline{\psi}^2_{1\overline{a}}\overline{\psi}^3_{\overline{a}4}\right) = \sum_b \left(\psi^3_{1b}\psi^2_{b4}\right)\left(\overline{\psi}^2_{1\overline{b}}\overline{\psi}^3_{\overline{b}4}\right) \tag{5.53}$$

In fact, eq. 5.53 is one way to define a conformal block with internal channel projected out (to channel $a$ or $b$, for example, in the notation above). Notationally, we have

$$\Psi_a(w,z) = \langle 1\big|\psi^2_{1a}(w)\psi^3_{a4(z)}\big|4\rangle \tag{5.54}$$

and, pictorially, as in fig. 5.5.

Schematically, we can write $\mathcal{F}^{23}_{14} = \psi^2_{1a}\psi^3_{a4}$, so that each term in the sums above is of the form $|\,\mathcal{F}\,|^2$ for some external vertices. One way to express this pictorially is as shown in fig. 5.6. Note that here, as in our definition of CVOs, we absorb the OPE coefficients and any theory-dependent data into the definition of the conformal block, such that there appear to be no coefficients in the sum. We use this normalization throughout the chapter.

The associativity constraint in eq. 5.53 is known as the *conformal bootstrap*, and has recently been used with much success to numerically compute CFT quantities that would otherwise be intractable. Fig. 5.6 shows the bootstrap equation 5.53 in pictures (in our theory-dependent normalization).

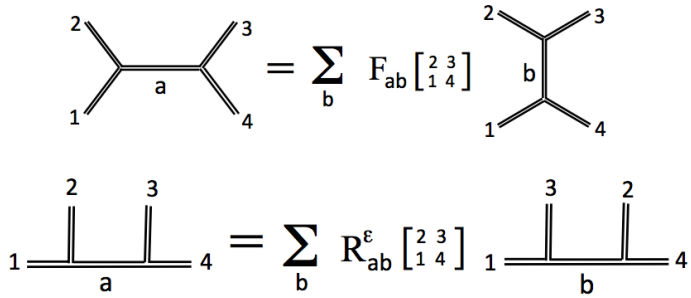



Figure 5.7: The $F$ move and the $R$ move, in pictures. These are change-of-basis matrices that are essential to the conformal bootstrap. The $\epsilon$ in the definition of $R$ equals $\pm 1$ and represents either $R$ or its inverse, depending on if 2 and 3 are exchanged "over" one another, or "under."

The conformal bootstrap equations endow the CFT with its topological properties. Consider the

four-point function $\langle \mathcal{O}_1(x_1)\mathcal{O}_2(x_2)\mathcal{O}_3(x_3)\mathcal{O}_4(x_4)\rangle$. In light of the bootstrap equations, it is natural to consider the $F$ matrix, a unitary matrix that carries out the change of basis shown in figure 5.7. A related "move" that performs a change of basis is given by the $R$ matrix, defined by the second expression in fig. 5.7. In equations, we have (here $\epsilon = \pm 1$ encodes the orientation of the braiding move)

$$\Psi_\alpha\left[\begin{smallmatrix} 2 & 3 \\ 1 & 4 \end{smallmatrix}\right](z) \;=\; \sum_{\mathrm{b}} \mathrm{F}_{\alpha\beta}\left[\begin{smallmatrix} 2 & 3 \\ 1 & 4 \end{smallmatrix}\right]\; \Psi_\beta\left[\begin{smallmatrix} 3 & 4 \\ 2 & 1 \end{smallmatrix}\right](1-z) \tag{5.55}$$

$$\Psi_\alpha\left[\begin{smallmatrix} 2 & 3 \\ 1 & 4 \end{smallmatrix}\right](z) \;=\; \sum_{b} \mathrm{R}^{\epsilon}_{\alpha\beta}\left[\begin{smallmatrix} 2 & 3 \\ 1 & 4 \end{smallmatrix}\right]\; \Psi_\beta\left[\begin{smallmatrix} 3 & 2 \\ 1 & 4 \end{smallmatrix}\right](1/z) \tag{5.56}$$

The fusion and R-matrix $\mathrm{F}_{\alpha\beta}$ and $\mathrm{R}_{\mathrm{ab}}$ are constant matrices independent of the coordinate $z$. They satisfy the following relation [125]

$$\mathrm{F}_{\alpha\beta}\left[\begin{smallmatrix} 2 & 3 \\ 1 & 4 \end{smallmatrix}\right] = e^{-\epsilon i\pi(\Delta_1+\Delta_3-\Delta_a-\Delta_b)}\mathrm{R}^{\epsilon}_{\alpha\beta}\left[\begin{smallmatrix} 2 & 4 \\ 1 & 3 \end{smallmatrix}\right] \tag{5.57}$$

Crossing symmetry is equivalent to the condition that crossing operations (5.55) and (5.56) expresses a unitary change of basis. This unitarity condition

$$\sum_{\gamma} \mathrm{F}_{\alpha\gamma}\,\mathrm{F}^*_{\mathrm{bc}} \;=\; \delta_{\alpha\beta} \qquad\qquad \sum_{\gamma} \mathrm{R}_{\alpha\gamma}\,\mathrm{R}^*_{\mathrm{bc}} \;=\; \delta_{\alpha\beta} \tag{5.58}$$

is one of the defining relations of the modular bootstrap. Other identities that $F$ and $R$ must satisfy, such as the Yang-Baxter equation and the pentagon equation, in addition to unitarity, serve to define a solution to the conformal bootstrap, and thus a local, unitary, 2-d CFT.

The braid group is intimately related to topology, not in $1+1$-d, but in $2+1$-d. This is a clue that the topological structure of a 2-d CFT does not manifest itself in a 2-d structure, but in a 3-d structure. This is indeed realized in beautiful works such as [84], where it is shown that rational 2-d CFT describes $2+1$-d Chern-Simons theory, a topological theory. The braid group then describes braiding of Wilson lines within the Chern-Simons bulk. In our discussion, braiding is crucial as well, but the gauge group is no longer compact, making the representation theory and thus CFT behavior more complex, but also richer.

## 5.B Expressions for volumes and 6j symbols

The Murakami-Yano formula [126] for the hyperbolic volume of a tetrahedron with dihedral angles $l_i$, $i = 1, .., 4, \alpha, \beta$ can be compactly written in terms of the auxiliary variables [127]

$$\begin{aligned} \mathcal{V}_1 &= l_\alpha + l_1 + l_2 & \mathcal{E}_1 &= l_1 + l_2 + l_3 + l_4 \\ \mathcal{V}_2 &= l_\alpha + l_3 + l_4 & \mathcal{E}_2 &= l_\alpha + l_\beta + l_1 + l_3 \\ \mathcal{V}_3 &= l_\beta + l_1 + l_4 & \mathcal{E}_3 &= l_\alpha + l_\beta + l_2 + l_4 \\ \mathcal{V}_4 &= l_\beta + l_2 + l_3 & \mathcal{E}_4 &= 0 \end{aligned} \tag{5.59}$$

as

$$\mathrm{Vol}\Big(T\Big[\begin{smallmatrix} 1 & 2 & \alpha \\ 3 & 4 & \beta \end{smallmatrix}\Big]\Big) = \sum_{i=1}^{4}\Big(\mathrm{Li}_2\big(z_+ e^{\mathcal{V}_i/2}\big) - \mathrm{Li}_2\big(z_- e^{\mathcal{V}_i/2}\big) - \mathrm{Li}_2\big(z_+ e^{\mathcal{E}_i/2}\big) + \mathrm{Li}_2\big(z_- e^{\mathcal{E}_i/2}\big)\Big), \tag{5.60}$$

where

$$\mathrm{Li}_2(z) = \int_z^0 dz\, \frac{\log(1-z)}{z} \tag{5.61}$$

denotes the dilogarithm function, and $z_+$ and $z_-$ are the two solutions to the quadratic equation (note that the $z^0$ and $z^4$ terms cancel out between both sides)

$$\prod_{i=1}^{4}\big(1 - z e^{\mathcal{V}_i/2}\big) = \prod_{i=1}^{4}\big(1 - z e^{\mathcal{E}_i/2}\big). \tag{5.62}$$

The above formula arises as the semi-classical limit of the quantum 6j symbols of $U_q(\mathfrak{sl}(2,\mathbb{R}))$, first found by Ponsot and Teschner [120]. It is expressed in terms of the double sine function $s_b(x)$ and quantum dilogarithm $e_b(x)$. These are defined as

$$s_b(z) = \prod_{m,n\geq 0} \frac{mb + nb^{-1} + \frac{Q}{2} - iz}{mb + nb^{-1} + \frac{Q}{2} + iz}, \qquad e_b(z) = e^{\frac{iQ}{2}z^2} s_b(z) \tag{5.63}$$

Both functions are symmetric under $b \to b^{-1}$. It is useful to introduce the double Sine function via $S_b(x) = s_b\big(ix - \frac{i}{2}Q\big)$. It satisfies $S_b(x) = S_{b^{-1}}(x)$ and $S_b(x)S_b(Q-x) = 1$. If we write $j_{ijk} = j_i + j_j + j_k$

and $j_{ijkl} = j_i + j_j + j_k + j_l$, then the quantum 6j symbols are given by the formula

$$\begin{aligned}\left\{ \begin{array}{ccc} j_1 & j_2 & j_\alpha \\ j_3 & j_4 & j_\beta \end{array} \right\}_b &= \Delta(j_\alpha, j_2, j_1)\Delta(j_4, j_3, j_\alpha)\Delta(j_\beta, j_3, j_2)\Delta(j_4, j_\beta, j_1) \\ &\quad \times \int_{\mathcal{C}} du\, S_b(u - j_{12\alpha})S_b(u - j_{\alpha 34})S_b(u - j_{23\beta})S_b(u - j_{1\beta 4}) \\ &\qquad \times S_b(j_{1234} - u)S_b(j_{\alpha\beta 13} - u)S_b(j_{\alpha\beta 24} - u)S_b(2Q - u)\end{aligned} \tag{5.64}$$

$$\Delta(j_3, j_2, j_1) = \Big(\frac{S_b(j_{123} - Q)}{\prod_{i=1}^3 S_b(j_{123} - 2j_i)}\Big)^{1/2}. \tag{5.65}$$

The integral is defined for $j_k \in Q/2 + i\mathbb{R}$ and a contour $\mathcal{C}$ which lies on the real axis in the interval $[\frac{3}{2}Q, 2Q]$ then proceeds to $2Q + i\infty$.

Defining the Plancherel measure via (with $j = \frac{Q}{2} + ip$)

$$d\mu(j) = \rho(j)\, dj \qquad\qquad \rho(j) \equiv 4\sinh(2\pi b p)\sinh(2\pi b^{-1} p). \tag{5.66}$$

the quantum 6j-symbols of $U_q(\mathfrak{sl}(2,\mathbb{R}))$ satisfy the required orthogonality conditions

$$\int d\mu(p) \left\{ \begin{array}{ccc} j_1 & j_2 & j_p \\ j_3 & j_4 & j_\alpha \end{array} \right\}_q^* \left\{ \begin{array}{ccc} j_1 & j_2 & j_p \\ j_3 & j_4 & j_\beta \end{array} \right\}_q = \frac{1}{\rho(j_\alpha)}\, \delta(j_\alpha - j_\beta), \tag{5.67}$$

and polynomial equations, such as the pentagon equation

$$\int d\mu(p) \left\{ \begin{array}{ccc} j_1 & j_2 & j_\alpha \\ j_3 & j_\beta & j_p \end{array} \right\}_q \left\{ \begin{array}{ccc} j_1 & j_p & j_\beta \\ j_4 & j_\epsilon & j_\delta \end{array} \right\}_q \left\{ \begin{array}{ccc} j_2 & j_3 & j_p \\ j_4 & j_\delta & j_\gamma \end{array} \right\}_q = \left\{ \begin{array}{ccc} j_\alpha & j_3 & j_\beta \\ j_4 & j_\epsilon & j_\delta \end{array} \right\}_q \left\{ \begin{array}{ccc} j_1 & j_2 & j_\alpha \\ j_\gamma & j_\epsilon & j_\delta \end{array} \right\}_q, \tag{5.68}$$

etc.

In the semi-classical limit $b \to 0$, the double Sine function behaves as

$$S_b\Big(\frac{\nu}{2\pi b}\Big) \;\simeq\; e^{-\frac{i}{2\pi b^2}(\frac{1}{4}\nu^2 - \frac{\pi}{2}\nu + \frac{1}{6}\pi^2)} \exp\Big(\frac{i}{2\pi b^2}\mathrm{Li}_2(e^{i\nu})\Big). \tag{5.69}$$

Upon inserting this result into the expression (5.64) for the quantum 6j-symbol, one can do the integral via the stationary phase approximation [118]. The stationary phase condition takes the

form (5.62). One finds that for small $b$

$$\left\{ \begin{array}{ccc} j_1 & j_2 & j_\alpha \\ j_3 & j_4 & j_\beta \end{array} \right\}_q \;\simeq\; \exp\Big\{ \frac{i}{2\pi b^2} \mathrm{Vol}\big(\mathrm{T}\big[ \begin{array}{ccc} 1 & 2 & \alpha \\ 3 & 4 & \beta \end{array} \big]\big) \Big\}, \tag{5.70}$$

where $T$ is the hyperbolic tetrahedron with dihedral angles $l_i = 2\pi b j_i$.

It is a relatively straightforward calculation to specialize the formula for the quantum 6j-symbol to the kinematic regime of interest, and verify that it gives the scattering phase $S_{\alpha\beta}$ of the gravitational shock wave interaction described in the Introduction. Setting the masses of the two particles A and B to zero, we start from the exact Liouville expression

$$\mathcal{R}_{\alpha\beta} \;=\; e^{\pi i(\Delta_\omega - \Delta_\alpha - \Delta_\beta)} \left\{ \begin{array}{ccc} j_M & j_\omega & 0 \\ j_\alpha & j_\beta & 0 \end{array} \right\}_b . \tag{5.71}$$

Applying the geometric dictionary and taking the semi-classical limit gives

$$\mathcal{R}^2_{\alpha\beta} \;=\; \exp\big(\tfrac{i}{\hbar} S_{\alpha\beta}\big), \qquad S_{\alpha\beta} = \mathrm{Vol}_{\mathcal{T}}(l_M, l_\omega, l_\alpha, l_\beta) \tag{5.72}$$

$$l_\omega = l_M + \hbar\omega/\kappa, \qquad l_\alpha = l_M + \hbar\alpha/\kappa, \qquad l_\beta = l_M + \hbar\beta/\kappa \tag{5.73}$$

$$\hbar = \pi b^2 = 4\pi/\ell, \qquad l_M/2\pi = R/\ell = \kappa\ell \qquad R^2 = 8M\ell^2. \tag{5.74}$$

We see that in the $\hbar \to 0$ limit, $\mathcal{T}(l_M, l_\omega, l_\alpha, l_\beta)$ approaches an ideal tetrahedron for four identical and two trivial tetrahedral angles. The Murakami-Yano formula simplifies for an ideal tetrahedron [126], because one of the solutions $z_\pm$ to (5.62) is equal to 1. Using this simplification, it becomes an easy calculation to extract the leading order $\hbar \to 0$ limit

$$\begin{aligned} \frac{1}{\hbar} S_{\alpha\beta} \;\simeq\; \frac{1}{\kappa} \Big\{ & \alpha \log\alpha + \beta\log\beta - (\omega-\alpha)\log(\omega-\alpha) - (\omega-\beta)\log(\omega-\beta) \\ & - (\alpha+\beta-\omega)\log\Big(2\sinh\big(\tfrac{\pi R}{\ell}\big)(\alpha+\beta-\omega)\Big)\Big\}. \end{aligned} \tag{5.75}$$

This formula matches with the scattering phase (6.22) derived from the shock wave geometry

# Chapter 6

# Moving into the bulk with $T\bar{T}$

This chapter is a lightly edited version of the paper [128] with M. Mezei and H. Verlinde.

In this chapter, we demonstrate a precise relationship between a $T\bar{T}$ deformation in a $2d$ CFT and a hard radial cutoff in an $AdS_3$ black hole bulk. We provide compelling support for our proposal by matching thermodynamic properties of the black hole to thermodynamic properties of CFTs with a $T\bar{T}$ deformation. Then we show that the proposal is consistent with a holographic renormalization group picture based on the Wheeler-de Witt equation.

Recent work by Zamolodchikov and others has uncovered a solvable irrelevant deformation of general 2D CFTs, defined by turning on the dimension 4 operator $T\bar{T}$, the product of the left- and right-moving stress tensor. We propose that in the holographic dual, this deformation represents a geometric cutoff that removes the asymptotic region of AdS and places the QFT on a Dirichlet wall at finite radial distance $r = r_c$ in the bulk. As a quantitative check of the proposed duality, we compute the signal propagation speed, energy spectrum, and thermodynamic relations on both sides. In all cases, we obtain a precise match. We derive an exact RG flow equation for the metric dependence of the effective action of the $T\bar{T}$ deformed theory, and find that it coincides with the Hamilton-Jacobi equation that governs the radial evolution of the classical gravity action in AdS.

## 6.1 Introduction and Summary

AdS/CFT duality is a powerful statement, thanks to the fact that one partner in the duality is a manifestly well defined quantum system with precise rules for computing correlation functions of local operators. Conformal field theory is by definition a UV complete framework, in which the rules of local quantum field theory apply at all energy scales. These statements remain true for relevant

or marginal deformations of CFTs that preserve the existence of a UV fixed point.

This virtue also has a flip side, as it makes AdS/CFT rather special. CFTs, or more generally, quantum field theories that are connected via RG flow to a UV fixed point, form a set of measure zero within the space of all effective QFTs. It is then natural to ask: can holography be extended to effective QFTs for which the UV behavior is *not* described by a CFT? In the context of $\mathrm{AdS}_3/\mathrm{CFT}_2$, this question has recently become more opportune, due to the discovery of Smirnov and Zamolodchikov [129] of a general class of exactly solvable irrelevant deformations of 2D CFT. Turning on an irrelevant coupling typically spoils the existence of a UV fixed point and destroys locality at some high cutoff scale. Properties of the deformed CFTs uncovered in [129], however, are found to be robust and largely decoupled from the question of their UV completeness.

In this paper we consider the simplest example of a solvable irrelevant deformation a 2D CFT, obtained by turning on a $T\bar{T}$ coupling

$$S_{\mathrm{QFT}} = S_{\mathrm{CFT}} + \mu \int d^2x \; T\bar{T} \,. \tag{6.1}$$

Here $T\bar{T}$ denotes the composite irrelevant (dimension 4) operator given by the product of the left- and right-moving components $T \equiv T_{zz}$ and $\bar{T} \equiv T_{\bar{z}\bar{z}}$ of the stress tensor, where we defined $z = x + i\tau$. Note that because $T\bar{T} = \frac{1}{8} T^{\alpha\beta} T_{\alpha\beta} - \frac{1}{16} (T^{\alpha}_{\alpha})^2$, the deformation preserves Lorentz invariance. By finite $\mu$ we mean that there is a one parameter family of theories defined by $dS^{(\mu)}_{\mathrm{QFT}}/d\mu = \int d^2x \, (T\bar{T})_{\mu}$, where the $\mu$ subscript of $T\bar{T}$ emphasizes that in this equation we have to use the stress tensor of $S^{(\mu)}_{\mathrm{QFT}}$. The deformation (6.1) is exactly solvable, in the sense that, even if the original 2D CFT itself has no extra symmetries other than Virasoro symmetry, the deformed theory possesses an infinite set of conserved charges and allows for exact computation of interesting physical quantities such as scattering phases, energy levels, and the thermodynamic equation of state [129, 130, 131]. Moreover, as we will see, there are several indications that the deformed CFT defined by (6.1) represents a consistent unitary quantum theory, with many interesting properties that are worth exploring.

We are interested in how the deformation (6.1) affects the standard holographic dictionary [13, 14] between CFT quantities and corresponding properties in AdS gravity . In the following we will argue that the coupling $\mu$ acts as a geometric cutoff that removes the asymptotic region of the AdS spacetime, and thereby places the QFT at a finite radial distance $r = r_c$ from the center of the bulk. We will test this proposal for the special subclass of quantities that can be created or measured by the stress tensor, or equivalently, by deformations of the metric.[1] Examples of such quantities are

[1]The background metric in the presence of the $T\bar{T}$ deformation is defined via the relation $\langle T_{\alpha\beta} \rangle$ =

signal propagation speeds, finite size effects, thermodynamic properties, and the Euclidean partition function $Z_{\rm QFT}(g,\mu)$ in a general background metric $ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta$.

Our concrete proposal is that the deformed CFT (6.1) is dual to the original gravitational theory (i.e. the gravity dual of the original CFT) living on a compact sub-region of AdS space-time

$$ds^2_{\rm AdS} \,=\, \frac{dr^2}{r^2} \,+\, r^2\, g_{\alpha\beta}dx^\alpha dx^\beta\,, \qquad\qquad r < r_c\,, \tag{6.2}$$

defined by restricting the radial coordinate to the finite interval $r < r_c$, with $r_c$ related to $\mu$ via

$$\frac{\mu \,=\, 16\pi G}{r_c^2 \,=\, \frac{24\pi}{c}\,\frac{1}{r_c^2}}\,. \tag{6.3}$$

Throughout the paper we set $\ell_{\rm AdS} = 1$, hence the Brown-Henneaux relation used in the above equation is $c = \frac{3}{2G}$ [100]. At large central charge $c$, we can identify

$$Z_{\rm QFT}(g_{\alpha\beta},\mu) = \exp\Big(-\frac{1}{16\pi G}\, S_{\rm cl}\big(r_c^2\, g_{\alpha\beta}\big)\Big)\,, \tag{6.4}$$

where $S_{\rm cl}(r_c^2 g_{\alpha\beta})$ is the classical action of the 3D gravity theory restricted to the region $r<r_c$, with Dirichlet boundary conditions $ds^2|_{r=r_c} = r_c^2 g_{\alpha\beta}dx^\alpha dx^\beta$ on the metric and $\phi_i|_{r=r_c} = 0$ on all bulk fields $\phi_i$. Here we assume that the classical matter fields do not contribute any stress-energy source.

The proposal has interesting implications for the holographic renormalization group program. In the formulation of [132] (see also [133, 134, 22, 135]) the CFT partition sum $Z_{\rm CFT}$ is identified with the gravity partition function in which the bulk path integral is cut into an IR and UV part via

$$Z_{\rm CFT}(\tilde g_{\alpha\beta},\epsilon) = \int Dg_{\alpha\beta}\Psi_{\rm IR}\big(r_c^2\, g_{\alpha\beta}\big)\ \Psi_{\rm UV}\left(r_c^2\, g_{\alpha\beta},\, \epsilon^{-2}\tilde g_{\alpha\beta}\right)\,. \tag{6.5}$$

Here $\epsilon$ denotes the short distance cutoff of the CFT. Here we have suppressed the integral over all matter fields: we assume that their saddle point value can be consistently set to zero. $\Psi_{\rm UV}$ is a path integral over metrics of the form (6.2) over the region $r_c < r < 1/\epsilon$ with prescribed boundary conditions, while $\Psi_{\rm IR}$ is an integral over all metrics in the region $r < r_c$ with boundary conditions that match those of $\Psi_{\rm UV}$. The IR wave- function $\Psi_{\rm IR}$ satisfies the Wheeler-DeWitt constraints, and via the holographic dictionary, is to be identified with a QFT path integral with a UV cutoff of size $1/r_c$. The UV wave-function $\Psi_{\rm UV}$ is related to the Wilsonian action by an functional Legendre transform, and is local on distance scales larger than $1/r_c$.

$\frac{2}{\sqrt{g}}\frac{\delta}{\delta g^{\alpha\beta}}\log Z_{\rm QFT}$, with $T_{\alpha\beta}$ the unique local conserved current associated with translation symmetry.

In this language our proposal states that

$$Z_{\rm QFT}\left(g_{\alpha\beta},\,\mu\right)=\Psi_{\rm IR}\left(r_c^2\,g_{\alpha\beta}\right) \tag{6.6}$$

with $\mu$ and $r_c$ related via (6.3). The full CFT partition function is insensitive to how we choose our renormalization scale, hence (6.5) is independent of $r_c$. Then the role of $\Psi_{\rm UV}$ is to undo the $T\bar{T}$ deformation of the CFT to get back the CFT result for the full partition function. It is also important to note that if the CFT has a large $N$ counting, where $c=O(N^2)$, the $T\bar{T}$ deformation is an irrelevant double trace deformation. There has been earlier speculation that the sharp radial cutoff in the bulk could be related to this kind of deformations [132].

The proposed dictionary is supported by several quantitative agreements between the two sides. We list three of them below.

1. *Deformation of the light cone.* The physical consequences of the $T\bar{T}$ deformation become most apparent by considering the system at finite temperature or in some eigenstate with finite energy density. In both cases, the stress-energy tensor has a non-zero expectation value. As pointed out by Cardy [136], this leads to a renormalization of the propagation speed $v_\pm$ of left- and right-moving massless excitations. In Minkowski space (6.1) takes the form

$$S_{\rm QFT}\,=\,S_{\rm CFT}-\mu\int d^2x\;T_{++}\,T_{--}\,, \tag{6.7}$$

where we defined $x^\pm=t\pm x$. Splitting off the expectation value from $T_{++}$ and $T_{--}$, the deformed action (6.7) acquires a linear term $-\mu\int d^2x\,[\langle T_{++}\rangle\;T_{--}+\langle T_{--}\rangle\;T_{++}]$, which has the same physical effect as a perturbation of the 2D metric of the form[2] $ds^2_{\rm CFT}\simeq -dx^+dx^- -\frac{\mu}{2}\,\langle T_{++}\rangle\,(dx^+)^2-\frac{\mu}{2}\,\langle T_{--}\rangle\,(dx^-)^2$. We see that the deformed CFT behaves like a gravitational theory in which stress-energy back reacts on the space-time geometry. The null directions of the effective metric are $dx^+=-\frac{\mu}{2}\,\langle T_{--}\rangle\,dx^-$ or $dx^-=-\frac{\mu}{2}\,\langle T_{++}\rangle\,dx^+$, and the propagation speed for left- and right-movers thus gets renormalized to

$$v_\mp\simeq 1+\mu\,\langle T_{\pm\pm}\rangle\;. \tag{6.8}$$

Note that for $\mu>0$, the deformation gives rise to superluminal propagation speeds, as the null energy $\langle T_{\pm\pm}\rangle$ is non-negative in states to which the above hydrodynamic argument applies.

[2]To derive this equation, we used the definition $S_{\rm QFT}\equiv-\frac{1}{2}\int d^2x g^{\alpha\beta}\,T_{\alpha\beta}$, and that $g^{\pm\pm}=-4g_{\mp\mp}$.

This effect has a natural interpretation in the gravity dual. A high energy CFT state is dual to a BTZ black hole geometry [137]. The propagation speed of metric perturbations of a BTZ black hole placed with Dirichlet walls at $r = r_c$ was analyzed by Marolf and Rangamani in [138]. Somewhat surprisingly, they found that these perturbations propagate at superluminal speed relative to the metric at the cutoff surface. Generalizing their derivation to the rotating case, one finds that the left- and right-moving propagation speeds are given by $v_\pm \simeq 1 + \frac{(r_+ \mp r_-)^2}{2r_c^2}$, with $r_+$ and $r_-$ the radius of the outer and inner horizon [89, 90]. Equating the renormalized velocities on both sides of the duality reproduces the standard result for the holographic stress-energy tensor [139] in the BTZ background, provided that $\mu$ and $r_c$ are related via (6.3).

*2. Deformed energy spectrum.* Another interesting physical quantity is the $\mu$ dependence of a given energy level $E_n(\mu, L)$ on a cylinder with circumference $L$.[3] Remarkably, this quantity can be computed exactly for any energy eigenstate of the perturbed CFT [129, 130, 131]. For a given CFT state with conformal dimension $(\Delta_n, \bar{\Delta}_n)$ one finds

$$E_n(\mu, L)L = \frac{2\pi}{\tilde{\mu}} \left(1 - \sqrt{1 - 2\tilde{\mu}\, M_n + \tilde{\mu}^2\, J_n^2}\right), \qquad \tilde{\mu} \equiv \frac{\pi\mu}{L^2}, \tag{6.9}$$

with $M_n = \Delta_n + \bar{\Delta}_n - \frac{c}{12}$, and $J_n = \Delta_n - \bar{\Delta}_n$. Note the right-hand side becomes imaginary above some critical conformal dimension (for fixed $\tilde{\mu} > 0$) or above some critical value of $\tilde{\mu}$ (for fixed $\Delta_n + \bar{\Delta}_n > \frac{c}{12}$). This behavior is called the 'shock singularity' in [129] and indicates the presence of a UV cutoff. We will summarize the derivation of the result (6.9) in section 6.2.4.

The analogous quantity to $E_n(\mu, L)$ on the gravity side is the quasi-local energy of a BTZ black hole of mass $M$ and angular momentum $J$ placed in a spatial region $r < r_c$, with Dirichlet boundary conditions $ds^2|_{r=r_c} = r_c^2\, dx^+ dx^-$. This quantity was computed in [140] (shortly before the discovery of AdS/CFT, so without any reliance on or reference to the holographic dictionary) by integrating the Brown-York stress-energy tensor over the boundary surface. This gravity result, given in equation (6.46), and the QFT result (6.9) precisely match, again provided we identify $\mu = \frac{24\pi}{c}\frac{1}{r_c^2}$.

The agreement between the energy spectra extends to a precise correspondence between all thermodynamic quantities, such as the equation of state, pressure, temperature, heat capacity, etc. Since the equations remain valid for finite values of $\tilde{\mu}$, this provides a new tool for studying bulk physics deep inside AdS. In particular, the 'shock singularity' of the deformed CFT (above which $E_n(\mu, L)$ becomes imaginary) is mapped to the singular properties (such as a diverging temperature and pressure) of the BTZ black hole inside a box $r < r_c$ as $r_c$ approaches the horizon. Studying the

[3] In studies of 2D CFT on a cylinder, it is customary to set $L = 2\pi$.

nature of this transition may give new insight into the physics of black hole horizons.

*3. Exact RG equation.* A key property, on which many of the exact results about the deformed theory (6.1) are based, is the following relation for the expectation value of the composite operator $T\bar{T}$

$$\left\langle T\bar{T}\right\rangle = \left\langle T\right\rangle \left\langle \bar{T}\right\rangle - \left\langle \Theta\right\rangle^2 . \tag{6.10}$$

Here $\Theta = T_{z\bar{z}} = \frac{1}{4}T^{\alpha}_{\alpha}$ denotes the trace of the stress tensor. This remarkable factorization property was first derived by Zamolodchikov in [141] and holds for any translation invariant, stationary state in any relativistic 2D QFT. Equation (6.10) in particular implies that the composite operator $T\bar{T}$ has exact scaling dimension 4, up to possible total derivative terms.[4] The absence of anomalous dimensions makes it possible that energy spectrum (6.9) is independent of the UV cutoff. (6.10) can be used to derive an RG equation for the partition function of the deformed CFT as follows.

The partition function $Z_{\rm QFT}(g,\mu)$ of the deformed CFT has a prescribed dependence on the 2D metric. Using that $\mu$ is the only scale in the problem, by taking the functional derivative of the partition function with respect to the scale factor of the metric to first order in $\mu$ we get

$$\left\langle \Theta\right\rangle = -\frac{c}{96\pi}R(g) - \frac{\mu}{2}\left\langle T\bar{T}\right\rangle . \tag{6.11}$$

The first term on the right-hand side is the trace anomaly of the CFT, the second term is a correction due to the $T\bar{T}$ deformation. Assuming that the metric is slowly varying, we combine (6.11) with the Zamolodchikov relation (6.10) to get

$$\left\langle \Theta\right\rangle = -\frac{c}{96\pi}R(g) - \frac{\mu}{2}\left(\left\langle T\right\rangle \left\langle \bar{T}\right\rangle - \left\langle \Theta\right\rangle^2\right) . \tag{6.12}$$

The above equations (6.10), (6.11) and (6.12) all hold to leading order in a derivative expansion.

The result (6.12) can be viewed as an exact RG equation of the $T\bar{T}$ deformed CFT. It holds for any 2D CFT, but acquires a special meaning for CFTs with gravity duals. To make its interpretation more evident, let us insert the holographic dictionary (6.3) and (6.4) into (6.12). This leads to a non-linear first order differential equation for the classical gravity action, given in equation (6.92), which coincides with the Hamilton-Jacobi (HJ) form of the holographic RG equation [132, 133, 22] that governs the radial evolution of the classical gravity action in AdS as a function of the cutoff $r_c$.

[4] Typically, a factorization property of this type is only exact in a strict large $N$ limit or for suitable protected operators in supersymmetric QFTs.

The results summarized above all have a common geometric origin. The HJ equation (6.92) is the classical limit of the Wheeler-DeWitt constraint that describes the radial evolution of a wave-function in 3D gravity. It has been known for some time that the partition function of a 2D CFT can be mapped, via a simple integral transform [92, 93, 142], to a wave-function that solves the WDW constraint of 3D gravity. From the CFT perspective, this integral transform looks like the $T\bar{T}$ deformation (6.1), rewritten in terms of a Gaussian integral over metric fluctuations. This exact result, stated in equations (6.93) and (6.94) in section 6.6, provided the initial inspiration for our conjectured interpretation of the $T\bar{T}$ deformation as moving the CFT into the bulk.

In the following sections we give some more detailed derivations of the above results. In section 6.2, we review the known exact results about the integrable $T\bar{T}$ deformation, including the presence of an infinite set of conserved charges, the energy spectrum (6.9) and the Zamolodchikov equation (6.10). We also highlight a relationship between the $T\bar{T}$ deformation and the Nambu-Goto action. In section 6.3, we review the computation of the quasi-local energy and thermodynamical properties of the rotating BTZ black holes with finite radial cutoff. In section 6.4 and the Appendix, we derive the renormalization of the propagation speed in CFT states dual to rotating BTZ black holes and a general class of bulk space-times giving space dependent stress tensor expectation values. In section 6.5, we look in more detail at the derivation of the exact RG equation (6.12) and its relation with the WDW equations of the 3D gravity theory. We end in section 6.6 with a discussion of various open questions.

## 6.2 $T\bar{T}$ Deformed CFT

In this section we will give an overview of some exact properties of the $T\bar{T}$ deformed CFT. More details can be found in the original papers [129, 130, 131]. These exact results are an important cornerstone of our general proposal. We will also address the question of UV completeness. As a concrete piece of evidence in favor, we point out that for the special case that the CFT has central charge $c = 24$, the $T\bar{T}$ deformation is exactly soluble and manifestly consistent – and in fact equivalent to the worldsheet theory of critical string theory [130, 131].

### 6.2.1 Integrability

The $T\bar{T}$ deformation is a special case of a more general class of irrelevant integrable deformations of CFTs introduced by Smirnov and Zamolodchikov (henceforth SZ) in [129]. For completeness, we

briefly state their main result. More details and derivations can be found in [129] (see also [130]). Integrable deformations of 2D CFTs are characterized by the existence of an infinite set of conserved higher spin charges $P_s$ and $\bar{P}_s$ of the form

$$P_s = \oint_C \big(T_s\, dz + \Theta_s\, d\bar{z}\big) \qquad \bar{P}_s = \oint_C \big(\bar{T}_s\, d\bar{z} + \bar{\Theta}_s\, dz\big) \tag{6.13}$$

where the current components $T_s$ and $\Theta_s$ are local operators of spin $s+1$ and $s-1$, respectively, and satisfy the current conservation identity

$$\partial_{\bar{z}} T_s = \partial_z \Theta_s\,, \qquad \partial_z \bar{T}_s = \partial_{\bar{z}} \bar{\Theta}_s\,. \tag{6.14}$$

The simplest integrals of motion with $s=1$ are the total left- and right-moving energy-momentum $P_+ = \oint_C (T\, dz + \Theta\, d\bar{z})$ and $P_- = \oint_C (\bar{T}\, d\bar{z} + \Theta\, dz)$. The conserved charges $P_s$ all commute with each other by virtue of the fact that their commutator with the currents yields a total derivative.

In the undeformed CFT, all $\Theta_s = \bar{\Theta}_s = 0$ and the currents $T_s$ and $\bar{T}_s$ are all chirally conserved. They are given by special composite operators, generically made up from the left- or right-moving stress tensor, respectively. For irrational CFTs, all currents $T_s$ and $\bar{T}_s$ are of this type. For CFTs with Kac-Moody or W-symmetries, there may be additional currents. For clarity, we emphasize that our notion of integrability does not automatically imply exact solvability: irrational CFTs with holographic duals are typically not exactly soluble. However, thanks to the infinite Virasoro symmetry, they posses and infinite set of conserved charges, and allow for integrable deformations that preserve an infinite subset of them.

The main results of SZ is that, in the neighborhood of any 2D CFT within the space of all 2D QFTs, there exists an infinite parameter family of integrable QFTs obtained by turning on an infinite set of irrelevant deformations of the form

$$S_{\rm QFT} = S_{\rm CFT} + \sum_s \mu_s \int d^2x\, X_s\,, \qquad X_s \equiv T_s \bar{T}_s - \Theta_s \bar{\Theta}_s\,. \tag{6.15}$$

Since $X_s$ has scaling dimension $2s+2$, these theories all become strongly coupled in the UV. Nonetheless, one can derive exact results about their symmetries, integrability, scattering phases and energy spectrum. In particular, SZ show that the conserved charges $P_s$ can be defined such that $\partial P_s/\partial \mu_{s'} + \big[P_s, \int d^2x\, X_{s'}\big] = 0$ for all $s$ and $s'$. This identity implies that all charges are preserved by the infinite set of deformations (6.15).

Our interest is in the special case that only the least irrelevant coupling $\mu = \mu_1$ of the lowest operator $X_1 = T\bar{T} - \Theta^2$ is non-zero. The integrability of this deformation will not be central to our story, except that it helps with some exact computations and gives some confidence that the deformed theory is well defined. Following SZ we often refer to the operator $X_1$ simply as $T\bar{T}$.

### 6.2.2 Zamolodchikov equation

The key result, from which many of the exact properties of the deformed CFT can be derived, is the Zamolodchikov equation (6.10) for the expectation value of the composite operator $T\bar{T}$, which holds for any translation invariant state in any 2D QFT with a local stress tensor. Here we briefly summarize its derivation [141]. Consider the difference of two point functions

$$\Xi(z,w) \;\equiv\; \big\langle T(z)\bar{T}(w)\big\rangle - \big\langle \Theta(z)\Theta(w)\big\rangle\,. \tag{6.16}$$

By taking two opposite limits, this function $\Xi(z,w)$ formally reduces to the expectation value of the composite operator $X_1 = T\bar{T} - \Theta^2$, or factorizes into the product of expectation values of individual stress tensor components

$$\begin{aligned} \lim_{w\to z} \Xi(z,w) &= \big\langle T\bar{T}\big\rangle - \big\langle\Theta^2\big\rangle, \\ \lim_{w\to\infty} \Xi(z,w) &= \langle T\rangle\langle\bar{T}\rangle - \langle\Theta\rangle^2\,. \end{aligned} \tag{6.17}$$

The second equality follows from the cluster property of local QFT. The first limit a priori needs to be taken with care, since the OPE between two operators generally becomes singular at short distance. The key insight, that relates the two limits and makes the first limit well behaved, is that the gradient of $\Xi(z,w)$ with respect to $z$ and $w$ identically vanishes. Using the conservation laws (6.14) and the fact that in a translation invariant state, the two point functions in (6.16) depend only on the coordinate difference $z-w$, one readily derives that

$$\big\langle \partial_{\bar{z}} T(z)\bar{T}(w)\big\rangle - \big\langle \partial_{\bar{z}}\Theta(z)\Theta(w)\big\rangle = -\big\langle \Theta(z)\partial_w\bar{T}(w)\big\rangle + \big\langle \bar{T}(z)\partial_w\Theta(w)\big\rangle = 0\,, \tag{6.18}$$

which shows that $\partial_{\bar{z}}\,\Xi(z,w) = 0$. In a similar way, one derives that $\partial_z\,\Xi(z,w) = 0$. Hence the function $\Xi(z,w)$ is a constant. This proves that the two right-hand sides in (6.17) are equal, leading to the relation (6.10).

Let us make two cautionary comments. First, as mentioned in the introduction, the relation

(6.10) suggests that the composite operator $T\bar{T}$ has exact scale dimension 4. This seems a surprisingly strong statement. However, since the derivation of (6.10) makes essential use of translation invariance, this property only holds at zero momentum. A more cautious and correct statement is that $T\bar{T}$ behaves as a local scaling operator with scale dimension 4 up to total derivative terms [141]. In later sections, we will also make use of (6.10) as a property that describes the behavior of more general states to leading order in a derivative expansion. A second related point is that in what follows, we will assume the result (6.10) remains valid for the $T\bar{T}$ deformed CFT at finite coupling $\mu$. Since the derivation outlined above applies to any QFT, this seems reasonable. However, as explained in [141], to avoid possible ambiguities related to the total derivative terms may require extra assumptions about the UV behavior of the QFT, which may not obviously hold for the $T\bar{T}$ deformed theory. We will ignore this subtlety in what follows.

### 6.2.3 2 Particle S-matrix

A useful perspective on the general class (6.15) of integrable deformations of a 2D CFT is to view them as a limit of integrable deformations of a massive 2D QFT. Massive 2D QFTs are uniquely characterized by the spectrum of stable particles and their S-matrix. The presence of an infinite set of higher spin charges $P_s$ implies that the S-matrix factorizes into 2-particle S-matrices $S_{ab}(\theta)$, which depend on the difference $\theta = \theta_a - \theta_b$ between the rapidities of particles $a$ and $b$. Any integrable 2D QFT admits an infinite parameter family of deformations, defined by multiplying each 2 particle S-matrix $S_{ab}(\theta)$ with a so- called CDD phase factor

$$S_{ab}(\theta) \to e^{i\delta_{ab}(\theta)} S_{ab}(\theta)\,. \tag{6.19}$$

The most general allowed phase takes the form of a sum $\delta_{ab}(\theta) = \sum_{s\in\mathcal{N}} \alpha_s \sinh(s\theta)$ of integer spin contributions. The deformation parameters $\alpha_s$ of the 2-particle S-matrix are in one-to-one correspondence with the deformation parameters $\mu_s$ of the CFT action (6.15). In case only the lowest spin deformation is turned on, the CDD phase factor simplifies to $\delta_{ab}(\theta) = -\frac{\mu}{4} m_a m_b \sinh(\theta)$, with $m_a$ and $m_b$ the mass of each particle. CFTs have only massless left- and right-moving excitations. To take the CFT limit, we thus boost particle $a$ and $b$ in opposite directions to the speed of light, while sending $m_a$ and $m_b$ to zero. In this limit the phase reduces to the product of the light-cone momenta of the two particles

$$\delta_{ab}(\theta) = -\frac{\mu}{8} m_a m_b e^{\theta_a - \theta_b} = \frac{\mu}{4} p_a^+ p_b^-\,. \tag{6.20}$$

In the unperturbed CFT left- and right-moving excitations pass through each other. The 2-particle S- matrix of the $T\bar{T}$ deformed CFT thus takes the simple form

$$S_{ab} = e^{i\mu\, p_a^+ p_b^-/4}\,. \tag{6.21}$$

This scattering phase (6.21) is a toy version of the forward scattering amplitude of two highly boosted particles in 3+1-D Einstein gravity, as first studied by 't Hooft [143]. It describes the effect of a gravitational shockwave caused by the stress-energy of one particle on the trajectory of the other particle. To make this physical interpretation explicit, consider a localized right-moving excitation $A(x^+)$ with small light-cone momentum $p_+$ and a left-moving mode $B_{p_-}$ in a momentum eigenstate with large light-cone momentum $p_-$. The 2-particle S-matrix (6.21) expresses the property that the operators $A$ and $B$ do not commute, as they would in the undeformed CFT, but satisfy a non-trivial exchange relation of the form

$$A(x^+)\, B_{p_-} \,=\, B_{p_-}\, A\left(x^+ - \frac{\mu p_-}{4}\right)\,. \tag{6.22}$$

This exchange relation exhibits the effect of a gravitational shockwave created by the energetic left-moving mode $B$ on the position of the right-moving mode $A$.

Introducing the scattering phase (6.21) has many interesting consequences. We briefly mention two of these.

*Ground state energy.* Via the thermodynamic Bethe ansatz, one can derive the ground state energy on a spatial circle with period $L$ as a function of $\mu$. One finds that [131]

$$E_0 = \frac{2L}{\mu}\left(1 - \sqrt{1 + \frac{\pi c \mu}{6L^2}}\right)\,. \tag{6.23}$$

Note that this formula has a square root singularity and becomes imaginary for $-\frac{\pi c\mu}{6} > L^2$. Since we may interpret the period of the circle as an inverse temperature $L$, this indicates that for $\mu < 0$, the theory breaks down above a critical temperature $T_{\text{H}} = \sqrt{\frac{-6}{\pi c \mu}}$. This breakdown is called the Hagedorn transition in [130, 131]. As we will see shortly, for the connection with holography we are interested in the opposite regime $\mu > 0$.

*Lyapunov behavior.* Equation (6.22) is similar to the shockwave interaction that gives rise to the chaotic dynamics of black hole horizons. It seems plausible, therefore, that the $T\bar{T}$ perturbation can be used to make the Lyapunov growth of out of time ordered (OTO) correlation functions [144, 145]

at finite temperature more manifest.

### 6.2.4 Energy spectrum

The generalization of the exact result (6.23) to arbitrary energy eigenstates, quoted in the introduction, can easily be derived from the Zamolodchikov relation (6.10) as follows. Consider the deformed CFT on a spatial circle parametrized by a angular coordinate $\theta$ with period $L$. Let $\big|n\rangle$ denote an energy and momentum eigenstate in CFT. Its energy $E_n$ and momentum $P_n$ take the general form

$$E_n = \frac{\mathcal{E}_n\big(\mu/L^2\big)}{L}\,, \qquad\qquad P_n = \frac{2\pi J_n}{L}\,, \qquad J_n \in \mathcal{Z}\,. \tag{6.24}$$

In the CFT limit, we have $E_n^{(\mathrm{CFT})} L = 2\pi\big(\Delta_n + \bar\Delta_n - \frac{c}{12}\big)$ and $J_n = \Delta_n - \bar\Delta_n \in \mathcal{Z}$. Since energy and momentum eigenstates are stationary and translation invariant, the Zamolodchikov relation (6.10) applies. It is useful to rewrite it as

$$\begin{aligned}\langle n\big|T\bar T\big|n\rangle &= \langle n\big|T\big|n\rangle\langle n\big|\bar T\big|n\rangle - \langle n\big|\Theta\big|n\rangle\langle n\big|\Theta\big|n\rangle \\ &= -\frac{1}{4}\Big(\langle n|T_{\tau\tau}|n\rangle\langle n|T_{xx}|n\rangle - \langle n|T_{\tau x}|n\rangle\langle n|T_{\tau x}|n\rangle\Big)\,,\end{aligned} \tag{6.25}$$

where we used the definitions and simple algebra. The stress tensor components have physical meaning as the energy density, pressure and momentum density. Hence we can express the right hand side in terms of physical properties of the spectrum:

$$\langle n\big|T_{\tau\tau}\big|n\rangle = \frac{E_n}{L}\,, \qquad \langle n\big|T_{xx}\big|n\rangle = \frac{\partial E_n}{\partial L}\,, \qquad \langle n\big|T_{\tau x}\big|n\rangle = \frac{iP_n}{L}\,. \tag{6.26}$$

The $i$ in the last formula follows from $T_{\tau x} = i(T - \bar T)$. The left-hand side of (6.25) represents the $\mu$ dependence of the energy $E_n$, via the relation[5]

$$\frac{\partial E_n}{\partial \mu} = L\,\langle n\big|T\bar T\big|n\rangle\,. \tag{6.27}$$

The relation (6.25) thus combines into the following differential equation for $E_n$

$$0 = 4\frac{\partial E_n}{\partial \mu} + E_n\frac{\partial E_n}{\partial L} + \frac{P_n^2}{L}\,. \tag{6.28}$$

[5]To see this, note that in euclidean signature $H_{\mathrm{int}} = \int d\theta\,\mathcal{L}_{\mathrm{int}}$ . This gives $\frac{\partial}{\partial\mu}\langle H\rangle = \int d\theta\,\langle T\bar T\rangle = L\,\langle T\bar T\rangle$.

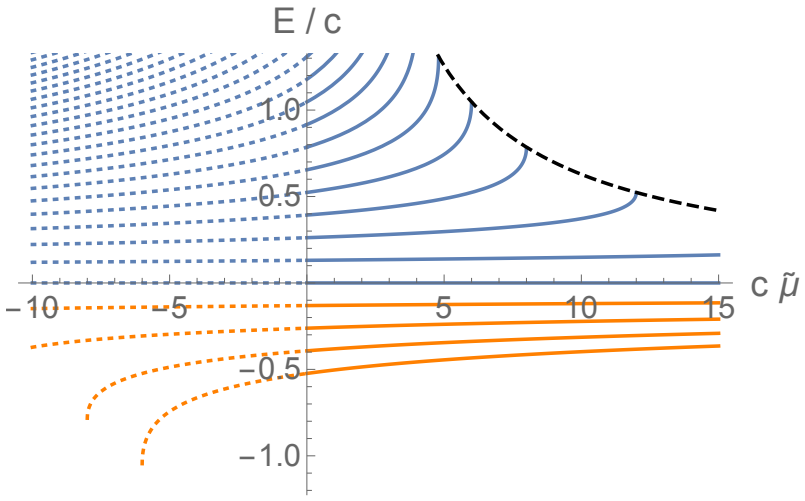



Figure 6.1: The energy levels $E_n$ at $L = 2\pi$ and $J = 0$ as a function of $\mu$ for different values of $E(0) = \Delta_n + \bar{\Delta}_n - \frac{c}{12}$. States with $E(0) > 0$ that correspond to black holes in holographic CFTs are plotted in blue, while low-lying states are plotted in orange. For $\mu > 0$ that is the relevant regime in our study we used solid lines, while for $\mu < 0$ the spectrum is plotted with dotted lines. The levels exhibit a square root singularity at the critical value $\mu E(0) = 2\pi$. This indicates that, for given $\mu$, the energy spectrum of the deformed CFT is bounded by $E < \frac{8}{\mu}$, indicated on the plot by a dashed black line.

As remarked in [129, 130], this equation is formally identical to the forced inviscid Burgers equation. Given that $E_n$ and $P_n$ have the form (6.24) and using the CFT value as initial condition, it is not hard to check that the solution to (6.28) is given by

$$\begin{aligned} E_n(\mu, L)L \equiv \mathcal{E}(\tilde{\mu}) &= \frac{2\pi}{\tilde{\mu}}\left[1 - \sqrt{1 - 2\tilde{\mu}M_n + \tilde{\mu}^2 J_n^2}\,\right] \\ M_n = \Delta_n + \bar{\Delta}_n - \frac{c}{12}\,, &\qquad J_n = \Delta_n - \bar{\Delta}_n\,, \qquad \tilde{\mu} \equiv \frac{\pi\mu}{L^2}\,. \end{aligned} \tag{6.29}$$

This relation reduces to the usual CFT value at $\tilde{\mu} \to 0$, and to the formula (6.23) for $\Delta_n = \bar{\Delta}_n = 0$.

The spectrum as a function of $\tilde{\mu}$ takes the form plotted in Fig. 6.1 for $\Delta_n = \bar{\Delta}_n$. With an eye towards large $c$ CFTs, we have scaled the energies and $\tilde{\mu}$ by $c$. The lowest energy level plotted is the ground state. The state with $\Delta_n + \bar{\Delta}_n = \frac{c}{12}$ (corresponding to the $M = 0$ BTZ black hole) has zero energy independent of $\mu$.

### 6.2.5 Thermodynamics

The formula (6.29) has a nice scaling form and does not depend on the UV cutoff. It can be read as describing the $\mu$ dependence of an energy level $E_n$ at fixed $L$, or as the variation of the energy under an adiabatic change in the circumference $L$ at fixed $\mu$. Note that $E_n(\mu, L, \Delta_n, \bar{\Delta}_n)$

is a monotonic function of $\Delta_n$ and $\bar{\Delta}_n$, so energy levels indeed do no cross as we vary $\mu$ or $L$. Hence the entropy remains $\mu$ independent and at high energy is given by the Cardy formula $S = 2\pi\sqrt{\frac{c}{6}(\Delta_n - \frac{c}{24})} + 2\pi\sqrt{\frac{c}{6}(\bar{\Delta}_n - \frac{c}{24})}$ [103].

Equation (6.29) exhibits a square root singularity for some critical value of $\tilde{\mu}$, called the 'shock singularity' in [129].[6] When $\Delta_n + \bar{\Delta}_n > \frac{c}{12}$, the singularity occurs for positive value of $\mu$. Its appearance indicates the presence of a high energy cutoff, and seems to suggest that for given $\tilde{\mu}$ the spectrum of the $T\bar{T}$ QFT truncates above a critical value for the conformal dimension. The scale at which the spectrum truncates is where a naive analysis would have predicted locality to break down.[7] Hence the deformed CFT on a cylinder has only a finite number of quantum states. Since the Cardy entropy monotonically grows with energy, we can also interpret this truncation as a bound on the total entropy of the system. The energy and entropy bound take the form

$$E < E_{\max} = \frac{2L}{\mu}\,, \qquad\qquad S < S_{\max} = L\sqrt{\frac{c}{6\pi\mu}}\,. \tag{6.30}$$

We will see that, on the gravity side, the state with maximal energy and entropy that saturates this bound corresponds to a maximal size black hole, that still fits inside the cutoff AdS space-time.

From now on we specialize to the non-rotating case $J = 0$. The equation of state of the deformed CFT defines a relation between the energy $E$, the circumference $L$, and the entropy $S$

$$EL - \frac{\mu}{4}E^2 = \frac{3S^2}{2\pi c}\,. \tag{6.31}$$

For $\mu = 0$, this reduces to the usual Cardy formula. From (6.31) we can derive other thermodynamic quantities, such as temperature and pressure, via the first law

$$dE = TdS - pdL \tag{6.32}$$

where the derivatives are taken while keeping the dimensionful coupling constant $\mu$ fixed. We thus obtain the following relations between the entropy density $s = S/L$, energy density $\rho = E/L$,

[6] In the application of the Burgers equation to fluid mechanics, $E(\mu, L)$ represents the fluid velocity, with $\mu =$ time and $L =$ position. The square root singularity then corresponds to the formation of a shock wave.

[7] The dimensionless coupling constant, that indicates the scale at which the theory may break down, is the product of $\mu$ and the energy density, $\tilde{\mu}M$. This becomes $O(1)$ where (6.29) has the square root singularity.

pressure $p$, and temperature $T$

$$\begin{aligned} \rho/\rho_{\text{H}} &= 1-\sqrt{1-s^2/s_{\text{H}}^2}\,, & p &= \frac{\rho}{1-\rho/\rho_{\text{H}}}\,, \\ p/\rho_{\text{H}} &= \sqrt{1+T^2/T_{\text{H}}^2}-1\,, & T/T_{\text{H}} &= \frac{s/s_{\text{H}}}{\sqrt{1-s^2/s_{\text{H}}^2}}\,. \end{aligned} \tag{6.33}$$

Here we introduced the critical values

$$\rho_{\text{H}} = \frac{2}{\mu}\,, \qquad s_{\text{H}} = \sqrt{\frac{2\pi c}{3\mu}}\,, \qquad T_{\text{H}} = \frac{\rho_{\text{H}}}{s_{\text{H}}} = \sqrt{\frac{6}{\pi c\mu}}\,. \tag{6.34}$$

Note that all these critical values diverge for $\mu \to 0$. One easily verifies that (6.33) reduces to the standard CFT relations in this limit. From equation (6.33) we verify the standard relation $p = -\rho + sT$, and obtain the free energy as a function of the temperature [131]

$$F = E - TS = \frac{2L}{\mu}\left(1-\sqrt{1+T^2/T_{\text{H}}^2}\right)\,. \tag{6.35}$$

The propagation speed $v_s$ of sound waves will play a central role in the comparison between the deformed CFT and gravity. For the non-rotating case $J = 0$, we can compute $v_s$ via

$$v_s = \sqrt{\frac{\partial p}{\partial \rho}} = \frac{1}{1-\rho/\rho_{\text{H}}} = \frac{1}{\sqrt{1-2\tilde{\mu}M}}\,. \tag{6.36}$$

with $\tilde{\mu} = \frac{\pi\mu}{L^2}$. We observe that for $\mu > 0$ sound waves propagate at superluminal speeds, and moreover that the temperature, pressure, and the sound speed all diverge at a critical values for the energy and entropy density. Near this critical value, the compressibility and the heat capacity of the system both go to zero. This singular behavior is another indication that the deformed CFT has a UV cutoff. As we will see in the following sections, the superluminal sound speed and the divergence of pressure and temperature all have a direct physical interpretation in the gravity dual description.

### 6.2.6 Equivalence to Nambu-Goto

There exists an instructive relationship between the deformed CFT and the Nambu-Goto (NG) string. This relationship is most explicitly understood for the case that the CFT has central charge $c = 24$, where it can be shown to be a direct equivalence with the worldsheet theory of critical string theory. Moreover, this reformulation makes manifest that, in this special case, the $T\bar{T}$ deformed

CFT represents a well defined, unitary and exactly soluble quantum system. This observation could help alleviate some possible worries the reader may have about the UV completeness of the theory.

Starting with some general CFT with $c = 24$, we define the deformed theory by adding two free massless scalar fields $X^+$ and $X^-$. The total action reads

$$S_{\rm QFT} \,=\, S_{\rm CFT} \,+\, \frac{1}{2\mu}\int d^2x\, \partial_\alpha X^+ \partial^\alpha X^- \,. \tag{6.37}$$

In the analogy with a string worldsheet theory, the free fields play the role of light-cone target space coordinates, whereas the CFT represents some general (abstract) 24-dimensional target space. Note that the kinetic term of the scalars $X^\pm$ has the opposite sign to the usual NG string. Just like one would in the NG formulation of string theory, we supplement the free field equation of motion $\partial_u\partial_v X^\pm = 0$ with the Virasoro conditions

$$-\partial_u X^+\partial_u X^- + \mu T^{\rm CFT}_{uu} = 0\,, \qquad -\partial_v X^+\partial_v X^- + \mu T^{\rm CFT}_{vv} = 0\,. \tag{6.38}$$

Here $u$ and $v$ denote the light-cone coordinates on the worldsheet. The constraints (6.38) implement gauge invariance under arbitrary conformal transformations $(u, v) \to (\tilde{u}(u), \tilde{v}(v))$. We can use this invariance to choose special worldsheet coordinate $(x^+, x^-)$ such that[8]

$$\partial_+ X^+ = \partial_- X^- = 1\,. \tag{6.39}$$

This gauge choice is analogous to the light-cone gauge in string theory. It identifies the worldsheet light-cone coordinates with the respective chiral halves of the target space light-cone coordinate fields: $X^+(x^+, x^-) = x^+ + \tilde{X}^+(x^-)$ and $X^-(x^+, x^-) = x^- + \tilde{X}^-(x^+)$. The other chiral halves of the light-cone fields are determined by integrating the Virasoro conditions

$$\begin{aligned} -\partial_- X^+ + \mu\, T^{\rm CFT}_{--} = 0\,, \qquad & \qquad X^+ = x^+ + \mu\int^{x^-} T^{\rm CFT}_{--}\,, \\ \Longrightarrow & \\ -\partial_+ X^- + \mu\, T^{\rm CFT}_{++} = 0\,, \qquad & \qquad X^- = x^- + \mu\int^{x^+} T^{\rm CFT}_{++}\,. \end{aligned} \tag{6.40}$$

Following the GGRT treatment of the NG string [146], equations (6.40) provide the quantum definition of the light-cone coordinate fields. The self-consistency of this identification at the full quantum level is well established for $c = 24$, and forms the basis of the no ghost theorem for the critical NG

[8] We thank Juan Maldacena for a helpful discussion on the usefulness of this particular light-cone gauge.

string [147, 148]. For our context, this theorem provides a direct proof that the $T\bar{T}$ deformed CFT with $c = 24$ is a well defined unitary quantum theory, in which all Hilbert states have positive norm.

The above reformulation provides an exact non-perturbative definition of the $T\bar{T}$ deformation of a CFT with $c = 24$. There are several ways to see that the two systems are indeed equivalent. A first simple check is that the leading order interaction term in the NG action indeed takes the form of a $T\bar{T}$ interaction [131, 130]. Conversely, if we insert equations (6.39) and (6.40) into (6.37), it becomes equal to the action (6.7). Note that the mapping between the two actions involves a coordinate transformation from the worldsheet coordinates $(u, v)$ to dynamical target space coordinates $(X^+, X^-)$ that back react to stress-energy. From equation (6.40) it is also clear why the above coupling of the CFT to the dynamical light-cone coordinates leads to non-trivial scattering phase and exchange relations between left and right-movers of the form (6.21) and (6.22) [149]. Another direct check is that the energy levels in Nambu-Goto theory are precisely of the form (6.29).[9]

## 6.3 Gravitational Energy and Thermodynamics

In this section we will make a comparison between the thermodynamic quantities of the $T\bar{T}$ deformed CFT and those of a BTZ black hole in a region of AdS with finite radial cutoff $r = r_c$. We will summarize the derivation of the total quasi-local gravitational energy of the black hole as a function $r_c$ [140], and show that it agrees with the QFT result (6.29). The metric of a BTZ black hole with mass $M$ and angular moment $J$ (with left- and right inverse temperature $\beta_\pm$) can be written as

$$\begin{aligned} ds^2 &= -f^2(r)dt^2 + f^{-2}(r)dr^2 + r^2\left(d\theta - \omega(r)\,dt\right)^2 \\ f^2(r) &= r^2 - 8GM + \frac{16G^2J^2}{r^2}\,, \qquad \omega(r) = \frac{4GJ}{r^2}\,, \\ M &= \frac{r_+^2 + r_-^2}{8G}\,, \qquad J = \frac{r_+ r_-}{4G}\,, \qquad \beta_\pm = \frac{2\pi}{r_+ \mp r_-}\,. \end{aligned} \tag{6.41}$$

The gravitational action for a 3D space-time with negative cosmological constant $\Lambda = -1$ and

[9] As we will see, however, the sign of $\mu$ required for our holographic interpretation is *opposite* to the standard sign in the Nambu-Goto string. This seems puzzling in view of the result of [150] that in non-linear QFTs with NG or DBI actions, only one sign of the higher derivative coupling leads to a consistent, causal theory. Our proposed resolution is that the notion of causality depends on how one defines the detectors that measure signals. There are two types: a) detectors anchored to the fixed coordinates $(u, v)$, or b) detectors attached to the dynamical coordinates $(X^+, X^-)$. Standard NG theory is causal with respect to detectors of type a), while the holographic theory is causal for detectors of type b). We'll return to this point in the next sections.

with a time-like boundary $B$ is

$$S = \frac{1}{16\pi G}\int d^3x\,\sqrt{-g_3}\,(R+2) - \frac{1}{8\pi G}\int_B d^2x\,\sqrt{-g}(K+1)\,, \tag{6.42}$$

where $K$ denotes the extrinsic curvature of the boundary. For simplicity, we omit all contributions due to matter fields: we assume that the classical matter sources can all be self-consistently turned off. The classical solutions of (6.42) are therefore locally pure $AdS_3$. We are interested in all solutions in the neighborhood of a general rotating BTZ black hole solution with a static boundary $B = \{r = r_c\}$. We impose Dirichlet boundary conditions, by fixing the form of the boundary metric

$$ds^2\big|_B = g_{\alpha\beta}dx^\alpha dx^\beta = -N^2dt^2 + e^{2\varphi}(d\theta - \omega dt)^2\,. \tag{6.43}$$

By solving the bulk equation of motion with given boundary condition, the action (6.42) becomes a functional $S[g]$ of the boundary metric. Note that in (6.42) we have included a boundary cosmological constant, tuned such that it cancels the leading volume divergence of the bulk action.

The gravitational energy of the black hole space-time is defined in terms of the variation of the action functional $S[g]$. In the parametrization (6.43), this variation takes the general form [140]

$$\delta S = \int_B d^2x\,\pi^{\alpha\beta}\delta g_{\alpha\beta} = \int_B d^2x\,\sqrt{-g}\,\big(-\epsilon\delta N - j\delta\omega + p\,\delta\varphi\big)\,. \tag{6.44}$$

The quantities $\epsilon$, $j$ and $p$ can respectively be interpreted as gravitational energy density, momentum density, and pressure, as measured on the boundary $B$. The total energy is thus obtained by integrating the energy density over the spatial section of $B$

$$E = \oint d\theta\, e^{\varphi}\epsilon\,. \tag{6.45}$$

We now want to apply this formalism to compute the quasi-local energy of a rotating BTZ black hole (6.41) with mass $M$ and angular momentum $J$ inside a Dirichlet wall at $r_c$. Following [140] we get

$$E = \frac{r_c}{4G}\left[1 - \sqrt{1 - \frac{8GM}{r_c^2} + \frac{16G^2J^2}{r_c^4}}\right]\,. \tag{6.46}$$

In the limit $r_c \to \infty$, this formula reduces to $E = M$. If we multiply (6.46) by the circumference

of the circle $2\pi r_c$, we obtain a formula for the dimensionless quantity $\mathcal{E} = 2\pi r_c E$, which perfectly matches with the result (6.29) obtained on the QFT side, provided we identify

$$M = M_n = \Delta_n + \bar{\Delta}_n - \frac{c}{12}\,, \qquad J = J_n = \Delta_n - \bar{\Delta}_n\,, \qquad \tilde{\mu} = \frac{4G}{r_c^2} = \frac{6}{c}\,\frac{1}{r_c^2}\,. \tag{6.47}$$

The first two identifications are completely standard in AdS/CFT. We again recover our proposed identification between the deformation parameter $\tilde{\mu}$ the cutoff radius $r_c$ (see footnote 3).

Note that the definition of the quasi-local energy only makes use of the intrinsic geometric properties of the cutoff space time and its boundary metric, and does not make any reference to the coordinate system of the asymptotic AdS observer. The thermodynamic quantities do not change under the change of coordinates in (6.43), $t' = at$, $\theta' = \theta + bt$, which respects the periodicity $\theta \sim \theta + 2\pi$. In other words, the formula (6.46) represents the total energy as measured in the canonical time coordinate of an observer living on the boundary $B$, in which the boundary metric has the form (6.43) with lapse $N = 1$. This is the right definition for comparison with the definition of energy in the deformed CFT. As explained in the previous section, due to the combined effect of the $T\bar{T}$ interaction and turning on a finite temperature, the metric (as defined as in footnote 1) gets renormalized relative to the metric of the undeformed CFT. The total energy of the QFT is defined as the integral of $\langle T_{00}\rangle = \frac{2i}{\sqrt{g}}\frac{\delta}{\delta g^{00}}\log Z_{\rm QFT}$, defined using the renormalized metric. This is the quantity that matches between both sides. Combined with the known correspondence between the Bekenstein-Hawking entropy

$$S = \frac{\pi r_+}{2G} \tag{6.48}$$

and the Cardy entropy of the CFT, this match between the energy spectra establishes a complete correspondence between the thermodynamical properties of the deformed CFT and the BTZ black hole with a sharp radial cutoff.

The holographic duality gives a new physical perspective on the square root singularity (6.9) in the energy levels of the $T\bar{T}$ deformed CFT. On the AdS side, this singularity in the gravitational energy (6.46) occurs when the Dirichlet wall approaches the event horizon of the BTZ black hole. For given $\mu$ and $r_c$, the critical behavior indicates an upper bound on the total energy and entropy

$$E < E_{max} = \frac{r_c}{4G}\,, \qquad S < S_{\rm max} = \frac{\pi r_c}{2G}\,, \tag{6.49}$$

which are saturated in the limit that the BTZ black hole completely fills out the space-time inside the wall at $r = r_c$. While $E$ and $S$ both remain finite in the limit, the temperature $T$ and pressure $p$ both diverge at this critical value. To compute both quantities, let us rewrite the QFT equation of state (6.31) in bulk notation

$$E r_c - 2GE^2 = \frac{GS^2}{2\pi^2} \,. \tag{6.50}$$

Using the first law (6.32), and the fact that the length of the Dirichlet wall is $2\pi r_c$, we derive that

$$\begin{aligned} T &= \frac{r_+}{2\pi\,(r_c - 4GE)} = \frac{r_+}{2\pi\sqrt{r_c^2 - r_+^2}} \,, \\ p &= \frac{E}{2\pi(r_c - 4GE)} = \frac{E}{2\pi\sqrt{r_c^2 - r_+^2}} \,. \end{aligned} \tag{6.51}$$

The divergence in the temperature is explained by the usual Tolman relation, or equivalently, by the fact that a static observer at the boundary must in fact undergo a uniform acceleration equal to $a = r_+/\sqrt{g_{00}}$, with $g_{00} = r_c^2 - r_+^2$, in order to stay at constant radius $r = r_c$. This acceleration diverges at the horizon, and via the Unruh effect, the static observer thus sees a black hole atmosphere as an incompressible fluid with a diverging temperature and pressure at the horizon.

From second equation in (6.51) we directly compute the sound speed via

$$v_s = \sqrt{\frac{\partial p}{\partial \rho}} = \frac{1}{1 - 4GE/r_c} = \frac{1}{\sqrt{1 - r_+^2/r_c^2}} \,, \tag{6.52}$$

which should be compared with the QFT result (6.36) obtained earlier. We now see more directly that the divergent sound propagation speed for $r_c = r_+$ arises from the incompressibility of the near horizon black hole atmosphere. We will study the propagation speed more closely in the next section.

## 6.4 Signal Propagation Speed

In this section we review Cardy's argument [136] that in the $T\bar{T}$ deformed CFT at finite temperature signals propagate faster or slower than the speed of light. We then consider the rotating state, and in the Appendix the case in which the expectation values of the stress tensor are arbitrary functions of the light-cone coordinate. We then compare with the prediction from our proposed holographic

dictionary, and find a precise match.

### 6.4.1 Propagation speed from QFT

To linear order in the deformation parameter $\mu$, one can regard the deformed CFT (6.1) as an undeformed CFT coupled to a Gaussian random background metric. The metric fluctuations represent two spin 2 Hubbard-Stratonovich (HS) fields $f, \bar{f}$:

$$S_{\rm QFT} = S_{\rm CFT} + \int d^2x \, \left[ -\frac{f\bar{f}}{\mu} + \bar{f}T + f\bar{T} \right] . \tag{6.53}$$

This HS representation can be extended to finite values of the deformation parameter $\mu$ via the more exact formula (6.93) quoted in the introduction. In this section we restrict ourselves to the linearized regime of small $\mu$. The action (6.53) describes the CFT in the random metric $ds^2 = dzd\bar{z} + \frac{\bar{f}}{2}dz^2 + \frac{f}{2}d\bar{z}^2$, or after analytic continuation to Lorentzian signature via $z \to x^-$, $\bar{z} \to -x^+$,

$$ds^2 = -dx^+dx^- + \frac{f}{2}(dx^+)^2 + \frac{\bar{f}}{2}(dx^-)^2 \, . \tag{6.54}$$

For later reference, note that to linearized order we may write this metric as $ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta$ with

$$g_{\alpha\beta} = \eta_{ab} v^a_\alpha v^b_\beta \, , \qquad\qquad v^\pm_\alpha \equiv e^\pm_\alpha + f^\pm_\alpha \, , \tag{6.55}$$

where $e^\pm_\alpha dx^\alpha = dx^\pm$ specifies the background metric and $f^+_\alpha dx^\alpha = -\frac{1}{2}\bar{f}dx^-$ and $f^-_\alpha dx^\alpha = -\frac{1}{2}fdx^+$ denotes the Gaussian fluctuation. The notation (6.55) will be useful later on, as it will allow us to vary the background metric $g_{\alpha\beta} = \eta_{ab}e^a_\alpha e^b_\beta$ independently from the fluctuating metric (6.55). Indeed, it will be important to keep track of the distinction between the two metrics and their corresponding light cones. The null geodesics of the fluctuating metric are specified by

$$v^+_\alpha dx^\alpha \equiv dx^+ - \frac{f}{2}dx^- = 0 \, , \qquad \text{or} \qquad v^-_\alpha dx^\alpha \equiv dx^- - \frac{\bar{f}}{2}dx^+ = 0 \, . \tag{6.56}$$

Let us consider a class of quantum states with non-zero expectation value of the stress tensor. Examples of such states are a thermal density matrix with finite inverse temperature $\beta$, or a semi-classical coherent state, in which the left- and right-moving component of the stress tensor have some general position dependent expectation value $\langle T_{++}(x^+)\rangle$ and $\langle T_{--}(x^-)\rangle$. In the above HS

representation of the deformed CFT, the fluctuating fields $f, \bar{f}$ attain the saddle point value[10]

$$\langle f \rangle = -\mu \langle T_{++} \rangle , \qquad \langle \bar{f} \rangle = -\mu \langle T_{--} \rangle . \tag{6.57}$$

Again we see that the presence of stress-energy affects the effective space-time geometry of the deformed CFT. This leads to a renormalization of the propagation speed of the left- and right-moving degrees of freedom. We can compute this effect using the null geodesics (6.56). At the stationary point of the random metric (6.54), the propagation speeds are renormalized to

$$v_+ \simeq 1 + \mu \langle T_{++} \rangle , \qquad v_- \simeq 1 + \mu \langle T_{--} \rangle . \tag{6.58}$$

This is the result that we argued for in the introduction (6.8). For the special case of a thermal state with left and right-moving inverse temperature $\beta_\pm$ we have (c.f. [136])

$$\langle T_{\pm\pm} \rangle = \frac{\pi c}{12 \beta_\pm^2} \qquad \Longrightarrow \qquad v_\pm \simeq 1 + \frac{\pi c \mu}{12 \beta_\pm^2} . \tag{6.59}$$

For $\mu > 0$, equation (6.59) and (6.36) represent superluminal speeds, while for $\mu < 0$ the renormalized speed is subluminal. The microscopic explanation of this effect is as follows [136]. The $T\bar{T}$ term for $\mu < 0$ leads to an attractive interaction and a positive time delay whenever a left- and right-moving particle collide. At finite temperature, the particles scatter off of a sea of quasi-particles, and the resulting time delay reduces the propagation speed. For $\mu > 0$, on the other hand, the repulsive inter-particle interaction leads to a time advance, and the accumulative effect of the scattering enhances the propagation speed. It is important to emphasize, however, that this speed is superluminal only relative to the fixed background metric, and that physics remains causal relative to the fluctuating metric (6.54). The UV limit of the $T\bar{T}$ deformed theory with $\mu > 0$ does not define a usual local CFT, but nonetheless behaves like a causal theory similar to 2D quantum gravity.

### 6.4.2 Propagation speed from thermodynamics

We can also compute the propagation speed using the thermodynamic equation of state. As we have seen in subsection 6.2.5, by combining the exact formula (6.29) for the $\mu$ dependence of the energy eigenvalues with the Cardy entropy formula of the CFT, we can derive the exact $\mu$ dependence of all

[10] Note that by going to Lorentzian signature we introduce a sign, $\langle T \rangle = -\langle T_{--} \rangle$, $\langle \bar{T} \rangle = -\langle T_{++} \rangle$.

thermodynamic quantities, including the speed of sound (6.36). Here we generalize the discussion to the rotating case.

In the rotating system, the equation of state (6.50) extends to a relation between the energy $E$, entropy $S$, radius $L$, and angular momentum $J$ of the form

$$EL - \frac{\mu}{4}\left(E^2 - \frac{4\pi^2 J^2}{L^2}\right) = \frac{3\,S^2}{2\pi c} + \frac{\pi^2 c J^2}{3\,S^2}\,. \tag{6.60}$$

The first law of thermodynamics generalizes to

$$dE = TdS - pdL + \Omega dJ \tag{6.61}$$

where $\Omega$ is to be identified with the angular rotation speed, possibly up to a constant shift. Here the derivatives are taken with $\mu$ fixed.

Let us first consider the system in the large $L$ limit, while keeping the energy density $\rho = E/L$, (angular) momentum density $j = J/L$ and entropy density $s = S/L$ fixed. We can then drop the term proportional to $J^2/L^2$ in (6.60), since it becomes small relative to the other terms. In this sense, we can think of this term as a finite size correction. The pressure and angular chemical potential $\Omega$ in the large $L$ limit can be expressed as

$$\bar{p} = \frac{\rho}{1 - \mu\,\rho/2}\,, \qquad\qquad \bar{\Omega} = \frac{2\pi^2 c j}{3\,s^2(1 - \mu\rho/2)}\,. \tag{6.62}$$

We notice that the relationship between $p$ and $\rho$ is identical to the non-rotating case. In particular, we deduce that the speed of sound in this limit is still equal to $\bar{v} = \sqrt{\frac{\partial\bar{p}}{\partial\rho}} = \frac{1}{1-\mu\rho/2}$. This speed is the same in both directions. This seems surprising, since a priori one would be inclined to interpret $\bar{\Omega}$ as the angular rotation speed of the QFT fluid. However, the corresponding linear speed $\bar{u} = \frac{L}{2\pi}\,\bar{\Omega}$ diverges in the large $L$ limit. So we will instead interpret $\bar{\Omega}$ as an off-set that has to be subtracted from $\Omega$ in order to get the physical angular velocity. So we will apply the redefinition $\Omega_{\rm new} = \Omega_{\rm old} - \bar{\Omega}$.

With this new definition, let us include the finite size term $J^2/L^2$. The pressure $p$ and angular velocity $\Omega$ at finite $L$ are given by

$$p = \frac{\rho - 2\pi^2\mu\,j^2/L^2}{1 - \mu\,\rho/2}\,, \qquad\qquad \Omega = -\frac{2\pi^2\mu\,j}{L^2(1 - \mu\rho/2)}\,. \tag{6.63}$$

The first relation reduces to $p = \rho$ at $\mu = 0$, as it should.

We would like to extract the propagation speeds from the two formulas (6.63). First we note that

$$\left(\frac{\partial p}{\partial \rho}\right)_L = \frac{1-\pi^2\mu^2 j^2/L^2}{(1-\mu\rho/2)^2} = v_+ v_- \tag{6.64}$$

with

$$v_\pm = \frac{1 \pm \pi\mu j/L}{1-\mu\rho/2} = \frac{1\pm\tilde{\mu}J}{\sqrt{1-2\tilde{\mu}M+\tilde{\mu}^2J^2}}\,, \tag{6.65}$$

where $\tilde{\mu} = \frac{\pi\mu}{L^2}$. Here in the second step we used equation (6.29).

It is reasonable to interpret the quantities $v_\pm$ as the left- and right-moving signal propagation speeds in the rotating deformed CFT. This interpretation is supported by the fact that the angular rotation frequency $\Omega$ and the left- and right velocities $v_\pm$ are related via

$$v_\pm = \bar{v} \mp \frac{\Omega L}{2\pi} \tag{6.66}$$

with $\bar{v} = 1/\sqrt{1-2\tilde{\mu}M+\tilde{\mu}^2J^2}$. As we will see shortly, the formula (6.65) for the propagation speeds agrees with the renormalized velocity computed via the gravity dual.

### 6.4.3 Propagation speed from gravity

The renormalization of the propagation speed has a direct interpretation in the dual gravity theory as the statement that metric perturbations of a BTZ black hole with Dirichlet boundary conditions at the fixed $r = r_c$ surface travel at superluminal speeds relative to the fixed boundary metric. The idea of the following calculation was introduced by Marolf and Rangamani in [138].

Consider a BTZ black hole of mass $M$ and angular momentum $J$, with the space-time metric is given in equation (6.41), surrounded by a Dirichlet wall at $r = r_c$. Now consider a fluctuation in the location of the boundary surface of the form $r_c \to r_c + \delta r(t,\theta)$. The Dirichlet boundary condition requires that the induced metric on the perturbed boundary surface remains flat. Computing the Ricci scalar of the induced metric on the perturbed boundary surface and expanding to linear order in the perturbation, one deduces that $\delta r(t,\theta)$ satisfies the linear wave equation

$$R\left(r_c + \delta r(t,\theta)\right) = -\frac{2}{r_c f^2(r_c)}\left[-\partial_t^2 + \partial_\theta^2\right]\delta r(t,\theta) = 0\,. \tag{6.67}$$

We see that, perhaps somewhat expectedly, the fluctuation $\delta r(t,\theta)$ describes a wave propagating along light-like trajectories $dt = \pm d\theta$, as measured in the coordinate system anchored to asymptotic infinity. These light-like trajectories are superluminal relative to the metric on the cutoff surface itself. To compute their speed as seen by an observer on the cutoff surface $r = r_c$, let us introduce coordinates $t_c$, $\theta_c$, so that the induced metric on surface is proportional to the standard flat metric $ds^2|_{r=r_c} = -dt_c^2 + d\theta_c^2$. In the $t$, $\theta$ coordinates the induced metric on the wall is

$$ds^2|_{r=r_c} = -f^2(r_c)dt^2 + r_c^2\left(d\theta - \omega(r_c)\,dt\right)^2 = r_c^2\left(-dt_c^2 + d\theta_c^2\right)\,. \tag{6.68}$$

The change of coordinates that preserves the $2\pi$ periodicity of $\theta$ is:

$$dt = \frac{r_c}{f(r_c)}\,dt_c\,, \qquad d\theta = d\theta_c - \Omega(r_c)dt_c\,, \qquad \Omega(r_c) \equiv -\frac{r_c\omega(r_c)}{f(r_c)}\,. \tag{6.69}$$

The quantity $\Omega(r_c)$ is the rotation speed due to the frame dragging effect of the rotating black hole as experienced at the cutoff surface $r = r_c$. It should be compared with the thermodynamic quantity $\Omega$ given in equation (6.63). The propagation trajectories are

$$dt = \pm d\theta \qquad \Longrightarrow \qquad \frac{r_c}{f(r_c)}\,dt_c = \pm\left(d\theta_c - \Omega(r_c)dt_c\right)\,. \tag{6.70}$$

We read off that the left- and right-moving parts of the wave $\delta r(t_c,\theta_c)$ propagate with velocity

$$v_\pm = \frac{r_c}{f(r_c)} \mp \Omega(r_c) = \frac{1 \pm \frac{4GJ}{r_c^2}}{\sqrt{1 - \frac{8GM}{r_c^2} + \frac{16G^2J^2}{r_c^4}}}\,. \tag{6.71}$$

This result precisely matches with the signal propagation speed (6.65) computed from the thermodynamics of the deformed QFT, provided we identify $\tilde{\mu} = \frac{4G}{r_c^2}$, which via the relations $\tilde{\mu} = \frac{\pi\mu}{L^2} = \frac{\mu}{4\pi}$ (setting $L = 2\pi$) and $G = \frac{3}{2c}$ reproduces the identification $\mu = \frac{24\pi}{c}\frac{1}{r_c^2}$ announced in the introduction. Specializing to the $J = 0$ case, equation (6.71) matches with (6.36). In the limit of large $r_c$, from (6.71) we obtain $v_\pm \simeq 1 + 2\pi^2/r_c^2\beta_\pm^2$, which coincides with the field theory result (6.59).

We generalize the analysis of signal propagation speed to a more general class of states in which the expectation value of the stress tensor has some arbitrary position dependence. The computation is presented in the Appendix, and we get the following linearized result

$$v_\pm \simeq 1 + \frac{16\pi G}{r_c^2}\left\langle T_{\pm\pm}(x^\pm)\right\rangle. \tag{6.72}$$

The agreement between this result and the propagation speed (6.58) computed in the deformed CFT is evidence that our proposed holographic dictionary extends to localized stress- energy perturbations.

From the gravity side, it still seems somewhat unsettling that the fluctuations of the boundary surface propagate at speeds that appear to violate boundary causality. So some clarifying comments may be in order. First we note that the speed (6.71) is equal to the inverse of the blackening factor, that relates the light-cone at $r = r_c$ to light-cone at asymptotic infinity in AdS. In other words, the speed (6.71) coincides with the light propagation speed at the asymptotic AdS-boundary. Somehow, the cutoff AdS space-time inside the Dirichlet wall has memory of the asymptotic light-cone, even though the asymptotic region is no longer there. A partial explanation for this phenomenon is that the propagating fluctuation described by (6.67) represents a boundary graviton mode. In spite of its name, a boundary graviton is not literally localized at the boundary of AdS, but instead represents a non-local geometric degree of freedom, encoded in the diffeomorphism that relates the uniformizing coordinate systems at the UV and IR boundaries of the AdS space-time. This diffeomorphism and the boundary graviton modes are topological excitations, in the sense that their propagation speed is insensitive to the introduction of the Dirichlet wall. Hence a bulk observer in AdS can not detect these graviton modes as superluminal localized excitations that violate local micro-causality constraints.

To gain further insight, it is instructive to view the propagation velocity from the perspective of information spreading in the QFT, as the speed by which a small perturbation in an equilibrium thermal state delocalizes throughout the system [151]. Suppose we act with a light local operator $\mathcal{O}(x, t = 0)$ on the thermal state. The strongly coupled QFT dynamics delocalizes the perturbation over a region $\Sigma(t)$ with radius $R(t)$ that grows linearly with time with the butterfly velocity $v_B$ [116, 144]. To determine $v_B$ we look for the smallest region that contains sufficient information to reconstruct $\mathcal{O}(x, 0)$. In the gravity dual, we can evaluate $v_B$ via the holographic postulate that the QFT state inside the boundary subregion $\Sigma(t)$ completely describes the bulk subregion $B_\Sigma$ contained within the Ryu-Takayanagi (RT) [152, 23] minimal surface associated with $\Sigma(t)$. The thermal state after acting with $\mathcal{O}(x, 0)$ is described by a BTZ black hole with a small particle that falls towards the horizon. The smallest boundary subregion $\Sigma(t)$ that contains the information created by $\mathcal{O}(x, 0)$ after time t is simply the smallest region such that the corresponding RT surface still contains the bulk particle [151].

This situation is depicted in figure 6.2. As time passes, the particle falls exponentially slowly

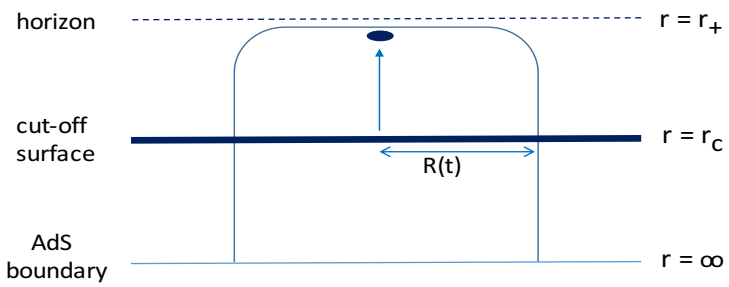



Figure 6.2: As a localized wave approaches the horizon, the minimal RT surface that contains the excitation at time $t$ extends along the horizon over a distance $R(t)$ that grows linearly in time.

towards the horizon. The minimal RT surface that contains the particle at some late time $t$ must follow a path that stays exponentially close to the horizon over a distance $R(t)$, that is roughly equal to the size of the associated boundary subregion. The size of the boundary subregion grows linearly in time as $R(t) \sim v_B t$. The butterfly velocity $v_B$ is equal to the speed of light as measured on the asymptotic AdS boundary at $r = \infty$. This means that the subregion on the cutoff surface at $r = r_c$ grows with superluminal speed compared to the light speed measured at $r = r_c$. This speed is equal to the signal propagation speed found in (6.71).[11] This is further holographic evidence that the effective signal propagation speed in the finite temperature QFT on the cutoff surface is superluminal.

## 6.5 Exact Holographic RG

In this section we present some more details about the relationship between the $T\bar{T}$ deformed theory and the holographic RG. On the QFT side, we use the Zamolodchikov equation to derive an exact RG equation for the scale dependence of the partition function. We then show that this RG equation is identical to the Hamilton-Jacobi equation that expresses the dependence of the bulk gravity action with a radial cutoff on the radial location of the boundary. Finally, we present a more precise definition of the $T\bar{T}$ deformed theory in terms of a suitable Hubbard-Stratonovich transformation, which has been shown to act as an intertwining map between the Weyl anomaly equation of a 2D CFT partition function and the Wheeler-DeWitt equation in 3D gravity [92, 93, 142]. This correspondence further substantiates our interpretation of the coupling constant $\mu$ as the radial location in the bulk.

[11] This result comes from comparing the opening angle of the light cone on the cutoff surface and asymptotically in the coordinates (6.41). The computation is logically different from the one used to derive (6.71), but they give the same result.

### 6.5.1 Zamolodchikov and Wilson-Polchinski

The $T\bar{T}$ deformed conformal field theories are interesting and special, because they allow for an exact study of their renormalization group flow. The $T\bar{T}$ interaction term introduces an effective UV cutoff scale, parametrized by the irrelevant coupling $\mu$. Hence if we consider the deformed CFT on a 2D space-time with arbitrary metric $ds^2 = g_{\alpha\beta}\, dx^\alpha dx^\beta$, the partition function and all other quantities will acquire a non-trivial dependence under Weyl rescalings. The goal of the exact renormalization group is to give a complete description of this scale dependence.

Metric variations of QFT quantities are governed by the action of the stress-energy tensor. Variations of the Weyl factor are generated, on the one hand, by inserting the trace of the stress tensor $\Theta = \frac{1}{4} T^\alpha_\alpha$, and on the other hand by varying the $T\bar{T}$ coupling $\mu$, which amounts to an insertion of the composite operator $T\bar{T}$. Our strategy is to write this relation, in combination with the Zamolodchikov equation, in the form of an exact RG equation. In this subsection, we will work to leading order in the coupling $\mu$ and to leading order in a derivative expansion. A more complete statement, which does not rely on these simplifying assumptions, is presented in subsection 6.5.3.

Using conformal perturbation theory, we can formally write the partition function $Z_{\rm QFT}(g, \mu)$ of the deformed CFT in a background metric $g$ as

$$Z_{\rm QFT}(g,\mu) = Z_{\rm CFT}(g) \left\langle \exp\big(-\mu \int d^2x\, T\bar{T}\big) \right\rangle_{\rm CFT} . \tag{6.73}$$

Here $T\bar{T}$ is short-hand for the combination

$$T\bar{T} = \frac{1}{8}\sqrt{g}\, g^{\alpha\gamma} g^{\beta\delta} \left( T_{\alpha\beta} T_{\gamma\delta} - \frac{1}{2} T_{\alpha\gamma} T_{\beta\delta} \right) . \tag{6.74}$$

In equation (6.73), we made the assumption that turning on the $T\bar{T}$ coupling does not generate other non-zero couplings. This assumption is reasonable as long as the background metric $g_{\alpha\beta}(x)$ varies slowly with $x$.

To extract the behavior of the partition function under Weyl transformations, we separate out the scale factor, and parametrize the 2D metric via

$$ds^2 = e^{2\varphi(x)}\, \hat{g}_{\alpha\beta} dx^\alpha dx^\beta , \tag{6.75}$$

where $\hat{g}_{\alpha\beta}$ specifies a unit determinant metric. We would like to make the dependence on the Weyl factor explicit. The undeformed CFT partition function transforms as $Z_{\rm CFT}(g) = e^{A(\varphi,\hat{g})} Z_{\rm CFT}(\hat{g})$

where the prefactor $e^{A(\varphi,\hat{g})}$ accounts for the scale dependence due to the trace anomaly

$$\frac{\delta A(\varphi,\hat{g})}{\delta\varphi} = -\frac{c}{24\pi} e^{-2\varphi}\, R(\varphi,\hat{g})\,, \tag{6.76}$$

where $c$ is the central charge of the CFT. The Weyl transformation rule of the total QFT partition function (6.73) then takes the form

$$Z_{\rm QFT}(e^{2\varphi}\hat{g},\mu) = e^{A(\varphi,\hat{g})}\, Z_{\rm CFT}(\hat{g}) \left\langle \exp\Big(-\mu \int d^2x\, e^{-2\varphi}\, T\bar{T}\Big) \right\rangle, \tag{6.77}$$

where $T\bar{T}$ now is short hand for the combination (6.74), with $g_{\alpha\beta}$ replaced by $\hat{g}_{\alpha\beta}$. Here we assumed that the expectation value can be defined such there no other sources of scale dependence. This assumption is strictly valid only to leading order in the coupling $\mu$.

Next we take the $\varphi$ variation on both sides. By definition, the variation of the left-hand side gives the expectation value of the trace of the stress tensor

$$\langle\,\Theta\,\rangle \equiv -\frac{e^{-2\varphi}}{4}\,\frac{\delta \log Z_{\rm QFT}}{\delta\varphi}\,. \tag{6.78}$$

The $\varphi$ dependence on the right-hand side has been made explicit through the anomaly factor and the rescaling of the coupling $\mu$ to $e^{-2\varphi}\mu$. Equating the $\varphi$ variation on both sides gives

$$\big\langle\Theta\big\rangle = -\frac{c}{96\pi}\, R(\varphi,\hat{g}) - \frac{\mu}{2}\big\langle T\bar{T}\big\rangle\,. \tag{6.79}$$

This equation can be viewed as the defining relation of the partition function of the deformed CFT.

As our final step, we use the Zamolodchikov relation (6.10), which holds to leading order in the derivative expansion and to leading order in the coupling $\mu$. Setting $\hat{g}_{\alpha\beta} = \eta_{\alpha\beta}$, we obtain

$$\langle\Theta\rangle = -\frac{c}{96\pi}\, R(\varphi) - \frac{\mu}{2}\big(\,\langle T\rangle\,\langle\bar{T}\rangle - \langle\Theta\rangle\,\langle\Theta\rangle\,\big)\,. \tag{6.80}$$

This is our proposed form of the exact renormalization group equation of the $T\bar{T}$ deformed CFT. Although we only derived it in the perturbative low energy regime, we will provide concrete evidence that for holographic CFTs, it continues to hold at finite values of $\mu$ and arbitrary $\hat{g}_{\alpha\beta}$.

Suppose we define an effective action $S_{\rm cl}(g,\mu)$ via[12]

$$Z_{\rm QFT}(g,\mu) \equiv \exp\left(-\frac{c}{24\pi} S_{\rm cl}(g,\mu)\right) . \tag{6.81}$$

With this definition, we can rewrite equation (6.80) as[13]

$$\frac{\delta S_{\rm cl}}{\delta\varphi} = -e^{2\varphi}\, R(\varphi,\hat{g}) - \frac{c\mu}{24\pi}\, e^{-2\varphi} \left( \hat{g}^{\alpha\gamma}\hat{g}^{\beta\delta}\, \frac{\delta S_{\rm cl}}{\delta \hat{g}^{\alpha\beta}}\, \frac{\delta S_{\rm cl}}{\delta \hat{g}^{\gamma\delta}} - \frac{1}{8}\left(\frac{\delta S_{\rm cl}}{\delta\varphi}\right)^2 \right) . \tag{6.82}$$

This an exact flow equation for the effective action $S_{\rm cl}(g,\mu)$ analogous to the Wilson- Polchinski exact RG equation. It has been recognized for some time [133] as we will now show, this equation precisely agrees the Hamilton-Jacobi equation that describes the radial dependence of the classical action in 3D gravity on an AdS space-time with a radial cutoff, provided we set $\mu = \frac{24\pi}{c}$.

### 6.5.2 WDW and Hamilton-Jacobi

We give a brief review of the holographic RG and its relation with the Wheeler-DeWitt and Hamilton-Jacobi equations. As we have done throughout this paper, we will concentrate on the dynamics of the bulk metric only, and assume that all other bulk matter fields are in their vacuum configuration and do not contribute any stress-energy or higher curvature corrections. Hence we will assume that the bulk is describe by pure Einstein gravity.

The holographic correspondence relates renormalization group flow in the CFT to radial evolution in the AdS space-time. The idea is to describe this evolution via a Hamiltonian formalism in which the radial direction plays the role of a euclidean time. We start by writing the 3D metric in the ADM parametrization corresponding to a foliation of the 3D manifold by constant $r$ slices

$$ds^2 = N^2 dr_c^2 + g_{\alpha\beta}\left(dx^\alpha + N^\alpha dr\right)\left(dx^\beta + N^\beta dr\right) \tag{6.83}$$

Here $N$ denotes the lapse, $g_{\alpha\beta}$ the metric on a radial slice, and $N^\alpha$ the shift vector. Next we write

[12] Here the normalization is chosen with an eye towards gravity where the prefactor would equal $1/16\pi G$.

[13] We start the rewriting of (6.80) by decomposing the stress tensor and metric variations into traceless and trace parts $T_{\alpha\beta} = \hat{T}_{\alpha\beta} + \frac{g_{\alpha\beta}}{2} T^\gamma_\gamma$, and $\delta g^{\alpha\beta} = e^{-2\varphi}\left(\delta\hat{g}^{\alpha\beta} - 2\hat{g}^{\alpha\beta}\,\delta\varphi\right)$ . This leads to the relation:

$$\delta \log Z_{\rm QFT} \;=\; \int d^2x\, \sqrt{g}\left[\frac{1}{2}\, e^{-2\varphi}\, \hat{T}_{\alpha\beta}\, \delta\hat{g}^{\alpha\beta} - T^\alpha_\alpha \delta\varphi\right] .$$

Equation (6.82) follows by combining this relation with the fact that

$$\langle T\rangle\, \langle \bar{T}\rangle \;=\; \tfrac{1}{8} g^{\alpha\gamma} g^{\beta\delta}\, \langle \hat{T}_{\alpha\beta}\rangle \langle \hat{T}_{\gamma\delta}\rangle .$$

the 3D Einstein action, including the boundary action given in (6.42), in the ADM decomposition

$$S_{\rm grav} = \int d^3x \left(\pi^{\alpha\beta}\dot{g}_{\alpha\beta} - N^\alpha H_\alpha - NH\right). \tag{6.84}$$

where the dot indicates derivative with respect to the radial coordinate $r$, $H_\beta = 2\nabla^\alpha \pi_{\alpha\beta}$ are the generators of 2D diffeomorphisms along the slice,

$$H = 2\pi^\alpha_\alpha + \frac{1}{\sqrt{g}}\left(\pi^{\alpha\beta}\pi_{\alpha\beta} - (\pi^\alpha_\alpha)^2\right) - \sqrt{g}R \tag{6.85}$$

denotes the ADM Hamiltonian, and $R$ is the two dimensional Ricci scalar of $g_{\alpha\beta}$.[14] Here the variable $\pi_{\alpha\beta}$ denotes the canonically conjugate variable to the metric $g_{\alpha\beta}$, and $R$ is the scalar curvature on the radial slice. The shift and lapse functions are Lagrange multipliers enforcing the the momentum and Hamiltonian constraints $H_\alpha = H = 0$.

Now let us define $S_{\rm cl}(g)$ as the value of the total 3D action $S_{\rm grav}$ evaluated on the classical background geometry with boundary values at $r = r_c$ given by

$$ds^2|_{r=r_c} = g_{\alpha\beta}(x)dx^\alpha dx^\beta\,. \tag{6.87}$$

The boundary values of all other bulk fields besides the metric are set to zero. We assume that their bulk dynamics can be consistently decoupled from the bulk dynamics of the metric.[15]

The Hamilton-Jacobi equation is a functional differential equation that governs how the on-shell value of the bulk action $S_{\rm cl}(g)$, defined as in (6.42), depends on the boundary value of the metric. It can most easily be derived by first consideingr the semi-classical partition function of the bulk theory with the same given boundary conditions. In the saddle point approximation

$$Z_{\rm grav}(g) = \exp\left(-\frac{1}{16\pi G}\,S_{\rm cl}(g)\right)\,. \tag{6.88}$$

By letting the radial direction play the role of time, we are led to interpret this partition function as

[14] This form of the ADM Hamiltonian follows from the more standard expression

$$H_{\rm ADM} = \frac{1}{\sqrt{g}}\left(\tilde{\pi}^{\alpha\gamma}\tilde{\pi}_{\alpha\beta} - (\tilde{\pi}^\alpha_\alpha)^2\right) - \sqrt{g}\,(R+2) \tag{6.86}$$

via the replacement $\tilde{\pi}_{\alpha\beta} = \pi_{\alpha\beta} - \sqrt{g}\,g_{\alpha\beta}$. This shift incorporates the extra boundary cosmological constant in (6.42), and is designed to cancel out the constant vacuum energy term while replacing it by a linear term proportional to $\pi^\alpha_\alpha$. This is a key step in the holographic renormalization procedure and for rewriting the radial evolution as an RG flow.

[15] Note that, without loss of generality, we have set $r_c = 1$ compared to equation (6.2). The radial AdS direction is uniquely parametrized by the Weyl factor of the metric.

a wave-functional of the boundary metric $g$. Accordingly, it must solve the gravitational analogue of the Schrödinger equation, commonly known as the Wheeler-DeWitt constraint

$$H_{\rm wdw} Z_{\rm grav}(g) = 0\,, \tag{6.89}$$

where $H_{\rm wdw}$ defines a functional differential operator, given by replacing in the classical ADM Hamilton (6.85) the momentum variables $\pi^{\alpha\beta}$ by the functional derivative with respect to the metric

$$\pi^{\alpha\beta} = -\frac{1}{\kappa}\frac{\delta}{\delta g_{\alpha\beta}}, \qquad\qquad \kappa \equiv \frac{1}{16\pi G}\,. \tag{6.90}$$

The WDW constraint (6.89) reduces to the Hamilton-Jacobi equation upon inserting (6.88) and taking the limit $\kappa \to \infty$.

Let us again separate out the scale factor and parametrize the 2D metric as in (6.75). Since $\varphi$ is a monotonic function of the radial coordinate $r$, it serves as a good parametrization for the bulk radial direction. In terms of these variables, the WDW Hamiltonian (6.85) takes the form[16]

$$\begin{aligned} H_{\rm wdw} &= 2\pi + e^{-2\varphi}\left(\hat{\pi}^{\alpha\beta}\hat{\pi}_{\alpha\beta} - \frac{1}{2}\pi^2\right) - e^{2\varphi}\,R\,, \\ \hat{\pi}_{\alpha\beta} &= -\frac{1}{\kappa}\frac{\delta}{\delta\hat{g}^{\alpha\beta}} \qquad\qquad \pi = \frac{1}{2\kappa}\frac{\delta}{\delta\varphi}\,, \end{aligned} \tag{6.91}$$

where indices are now contracted with $\hat{g}$. Note the factor of $-1/2$ in the equation relating $\pi$ to $\delta/\delta\varphi$, which can be easily determined by acting on $\sqrt{g}$ with both operators. Inserting (6.88) into (6.89) and taking the $\kappa \to 0$ limit yields the HJ-equation

$$\frac{\delta S_{\rm cl}}{\delta\varphi} + e^{-2\varphi}\left(\hat{g}^{\alpha\gamma}\hat{g}^{\beta\delta}\,\frac{\delta S_{\rm cl}}{\delta\hat{g}^{\alpha\beta}}\,\frac{\delta S_{\rm cl}}{\delta\hat{g}^{\gamma\delta}} - \frac{1}{8}\left(\frac{\delta S_{\rm cl}}{\delta\varphi}\right)^2\right) + e^{2\varphi}\,R(\hat{g},\varphi) = 0\,. \tag{6.92}$$

Via holography, this equation acquires the meaning of an exact RG equation for the effective action of a CFT with a UV cutoff, defined by integrating out all CFT degrees of freedom above a scale associated to the value of the Weyl factor $e^{\varphi}$. Up to now, however, it has not been clear what this holographic UV cutoff exactly looks like from the QFT perspective. The precise match with the exact RG equation (6.80)-(6.82) is strong evidence that this preferred holographic UV cutoff is given by the $T\bar{T}$ deformation.

[16]Here we use the decomposition $\pi_{\alpha\beta} = e^{2\varphi}\left(\hat{\pi}_{\alpha\beta} + \frac{1}{2}\hat{g}_{\alpha\beta}\,\pi\right)$ with $\pi = \pi^{\alpha}_{\alpha}$.

### 6.5.3 WDW from Hubbard-Stratonovich

The above results all have the following common geometric origin. Early studies of the modular geometry of the conformal block in 2D CFT revealed a deep connection with quantum states of 3D gravity. Based on this, it has been known for some time that the partition function of a 2D CFT can be mapped to a solution of the WDW equation of 3D gravity via an integral transform [92, 93, 142]. From the CFT perspective, this transform looks like a $T\bar{T}$ deformation, rewritten in terms of a Gaussian integral over metric fluctuations.

To write the integral transform, it is convenient to parametrize the metric by means of a zweibein $e^a = e^a_\alpha dx^\alpha$ via $g_{\alpha\beta} = \delta_{ab} e^a_\alpha e^b_\beta$. The CFT partition function $Z_{\rm CFT}(e)$ is a reparametrization and local Lorentz invariant functional, with scale dependence fixed by the trace anomaly. Suppose we now define the $T\bar{T}$ deformed theory such that its partition function $Z_{\rm QFT}(e)$ is obtained from the CFT partition function via (c.f. equation (6.55))

$$Z_{\rm QFT}(e) = \int \mathcal{D}f \; e^{\frac{2}{\mu}\int f^+ \wedge f^-} \; Z_{\rm CFT}(e+f)\,. \tag{6.93}$$

It is not self-evident that this definition of the deformed CFT is equivalent to the one we used so far, $dS^{(\mu)}_{\rm QFT}/d\mu = \int d^2x \; (T\bar{T})_\mu$,. However, as we show below, they both lead to the same detailed match with 3D gravity, which suggests that the two definitions do coincide for holographic CFTs with large central charge. It was shown by Freidel in [142], based on earlier work [92, 93], that this integral transform acts like an intertwining map between the trace anomaly and the WDW equation

$$\Big(\frac{\delta}{\delta\varphi} - \frac{c}{24\pi}\, e^{-2\varphi}\, R(g)\Big)\, Z_{\rm CFT}(g) = 0 \qquad \Longrightarrow \qquad H_{\rm wdw}\, Z_{\rm QFT}(g) = 0\,, \tag{6.94}$$

where $H_{\rm wdw}$ given in equation (6.85) and (6.90), and where the central charge $c$ and the Newton constant are related to $\mu$ via [142]

$$c = 1 + \frac{24\pi}{\mu}\,, \qquad\qquad \frac{1}{16\pi G} \equiv \kappa = \mu^{-1}\,. \tag{6.95}$$

The constraint equations (6.94) both hold locally at every point in 2D space-time.

The integral transform (6.93) can be transferred to inside the CFT functional integral. This

yields the following formula for the action of the deformed theory

$$S_{\rm QFT}(e) = \min_f \left( S_{\rm CFT}(e+f) \, - \, \frac{2}{\mu} \int f^+ \wedge f^- \right) . \tag{6.96}$$

Given that for small fluctuations $S_{\rm CFT}(e+f) = S_{\rm CFT}(e) - \int \left( f^+_\alpha \, T^\alpha_+ + f^-_\alpha \, T^\alpha_- \right)$ with $T^\alpha_\pm = T^\alpha{}_\beta e^\beta_\pm = -\frac{1}{e} \frac{\delta S_{\rm CFT}}{\delta e^\pm_\alpha}$, this looks like a Hubbard-Stratonovich representation described in section 6.4.1 of the $T\bar{T}$ deformed theory (6.1) as a Gaussian integral over a fluctuating metric with deformation parameter $\mu$. In the large $c$ limit, the integral transform (6.93) can be performed via a semi-classical approximation, and thus amounts to performing a Legendre transformation. For small stress-energy fluctuations, the formula (6.96) for the QFT action then reduces to $S_{\rm QFT} = S_{\rm CFT} + \mu \int d^2x \, T\bar{T}$.

In the semi-classical large $c$ limit, Freidel's result amounts to the statement that

$$S_{\rm grav}(e) = \min_f \left( S_{\rm CFT}(e+f) \, - \, 2\kappa \int f^+ \wedge f^- \right) . \tag{6.97}$$

where $S_{\rm grav}(e) = \kappa S_{\rm cl}(e)$ is the classical bulk gravity action with boundary conditions $g_{\alpha\beta} = \delta_{ab} \, e^a_\alpha e^b_\beta$ at the cutoff surface $r = r_c$. The derivation of this universal result relies on the fact that the metric dependence of the CFT action is fixed by the conformal anomaly and given by the Polyakov action

$$S_{\rm CFT}(e) = \frac{c}{192\pi} \int d^2x \, R \, \Box^{-1} R \, . \tag{6.98}$$

The precise equality

$$S_{\rm grav}(g) = S_{\rm QFT}(g) \tag{6.99}$$

between the bulk gravity action (6.97) and the metric dependence of the effective action (6.96) of the deformed CFT guarantees that all correlation functions of the stress-energy tensor in the deformed 2D CFT exactly match with those obtained from holography. The result (6.99) generalizes the known match between the holographic and CFT conformal anomalies [153] to the new situation, where the boundary is placed at a finite distance from the center of the bulk.

We refer to [142] for a detailed derivation of the result (6.94). Here we just add a short comment about its underlying intuition. The integral transform (6.94) has a geometrical significance as an gluing operation that combines together the wave-functions of two chiral gravity theories (given by the chiral conformal blocks of the CFT) into a single non-chiral wave-function (given by the partition

function $Z_{\rm QFT}$). It has long been known [92, 93] that the conformal Ward identities satisfied by a chiral conformal block in 2D CFT are identical to the physical state conditions on wave-functions of chiral gravity, provided one uses a holomorphic polarization on its phase space – that is, provided one writes the chiral wave-function in terms of complex variables analogous to the coherent state basis of a harmonic oscillator. In short, CFT conformal blocks, when viewed as functionals of the corresponding chiral zweibein $e^+$ (or $e^-$) are coherent states of chiral 3D gravity [92, 93]. Gluing the chiral coherent state wave-functions together into a real solution of the non-chiral WDW equation requires performing an integral transform, analogous to the integral transform that rewrites the coherent state basis of a harmonic oscillator into a wave-function in the position representation. In the first order formulation of 3D gravity, this integral takes the form of a Gaussian integral given in (6.94).

The integral transform (6.94) gives a well-controlled definition of the $T\bar{T}$ deformed CFT at large central charge. With this definition, the result by Freidel provides an independent derivation of the Zamolodchikov formula (6.10) and the corresponding exact RG equation. Moreover, it implies that for large $c$, the all $n$-point connected correlation functions of the stress tensor computed in the QFT are identical to the correlation functions computed via pure 3D gravity

$$\frac{(-2)^n}{\sqrt{g_1}\sqrt{g_2}\cdots\sqrt{g_n}}\frac{\delta S_{\rm grav}(g)}{\delta g_1^{\alpha\beta}\delta g_2^{\beta\gamma}\ldots\delta g_2^{\rho\sigma}} \quad = \quad \Big\langle T_{\alpha\beta}(x_1)T_{\beta\gamma}(x_2)\ldots T_{\rho\sigma}(x_n)\Big\rangle^{\rm conn}_{\rm QFT(g)}$$

with $g_i^{\alpha\beta} = g^{\alpha\beta}(x_i)$, etc. This relation looks perhaps more miraculous than it really is. The right-hand side is fixed by the conformal anomaly and Ward identities, and depends only on one single dimensionless number, the central charge $c$. Similarly, the correlation functions of boundary gravitons in 3D gravity are fixed by the AdS analogue of soft-graviton theorems and only depend on the ratio of the AdS scale and the Planck scale. Still, this result is a useful extension of the standard AdS/CFT dictionary, that may open up new ways of probing the gravitational bulk physics.

## 6.6 Conclusion

In this paper, we studied the class of 2D effective QFTs defined by turning on an irrelevant $T\bar{T}$ deformation in a general 2D CFT. We proposed that in the holographic dual, the deformation corresponds to introducing a rigid cutoff surface that imposes Dirichlet boundary conditions at a finite radial location $r = r_c$ in the bulk. As a check of the duality, we have shown that the energy spectrum, thermodynamic properties, propagation speeds, and the metric dependence of

the partition function agree on both sides. This correspondence is largely explained by the precise identification between the 2D conformal Ward identities and the physical state conditions in 3D gravity.

There are many open questions. It will be important to establish whether the $T\bar{T}$ deformation indeed produces a well defined unitary quantum system. We have seen that for CFTs with $c = 24$, the deformation is equivalent to the Nambu-Goto formulation of the string worldsheet theory on some general target space. The NG theory is soluble and appears to be a well defined deformation of the CFT for both choices of sign, including the one that leads to superluminal propagation speeds relative to the non-dynamical background metric. This indicates that the $T\bar{T}$ deformation is also consistent for large $c$ CFTs, but a general proof is not yet available.

It is natural to ask whether some of our results can be extended to higher dimension. The main catalyst our story, the Zamolodchikov equation (6.10), looks like a large $N$ factorization property. So it seems plausible that an analogous equation can be derived in large $N$ CFTs in higher dimensions. However, since conformal symmetry is less restrictive for $d > 2$, it is not clear if such an equation can be used to derive analogous unique flow equations for the energy levels and the partition function. Even so, it would be instructive to explore what double trace technology can teach us about the $T\bar{T}$ deformed theory. Because the stress tensor is normalized such that its two-point function is $\langle T\,T\rangle_{\rm CFT} = O(N^2)$, in order for the double trace coupling $\mu$ to appreciably influence the dynamics, and to preserve the structure of the large-$N$ expansion, it has to be $\mu = O(1/N^2)$ or parametrically larger. In the regime where $r_c/\ell_{\rm AdS} = O(1)$, $\mu$ is indeed of this order. To explore sub-$\ell_{\rm AdS}$ scale physics we cannot rely on a perturbative expansion in $1/N$, and non- perturbative methods are needed. We have seen that, in two dimensions and for correlation functions of the stress tensor, such non-perturbative methods are indeed available.

For most of our computations, we have restricted our attention to long distance properties of the $T\bar{T}$ QFT. Indeed, it is not clear whether it is possible to define true local operators, that probe or excite the QFT at arbitrarily short distance scales. We have seen that turning on the $T\bar{T}$ interaction leads to fluctuations in the effective metric that grow large in the UV. The randomness of the dynamical UV metric complicates the task of finding a precise holographic map analogous to the standard GKPW dictionary QFT and gravity observables. Still, it would be worthwhile to study the properties of localized probes in the QFT, other than stress tensors, and investigate whether is it possible to compute correlation function at sub-AdS distances, as measured at the cutoff surface.

If we define the dimensionless coupling as the ratio

$$\bar{\mu} \equiv \text{dimensionless coupling} = \frac{\mu}{\Delta\theta^2} = \frac{24\pi}{c r_c^2}\frac{1}{\Delta\theta^2} = \frac{24\pi}{c}\frac{1}{d^2}$$

$$d \equiv \text{distance scale in AdS units} = r_c \Delta\theta\,,$$

we see that the $T\bar{T}$ interaction and the associated metric fluctuations remain small all the way down to the short distance scale $d_{\text{planck}} = \sqrt{\frac{24\pi}{c}}\ell_{\text{AdS}}$. So in this sense, we should be able to use the $T\bar{T}$ QFT to probe bulk physics at sub-AdS distance scales. The key questions, however, are how to extend our calculations to general operators $\mathcal{O}$ and how the bulk physics in this regime is affected by the presence of the cutoff surface. Assuming that the cutoff surface continues to behave like a Dirichlet wall, correlation functions at this short distance scale should behave similar correlation functions in a gravitational theory in flat space.

## Acknowledgements

We thank Vijay Balasubramanian, Clay Cordova, Xi Dong, Raphael Flauger, Tom Hartman, Sung-Sik Lee, Hong Liu, and Juan Maldacena for helpful discussions and comments. The research of M.M. was supported in part by the U.S. Department of Energy under grant No. DE- SC0016244. The research of H.V. is supported by NSF grant PHY-1620059.

## 6.7 Propagation speed in general backgrounds

In this appendix we generalize the analysis of signal propagation speed presented in section 6.4.3 to a more general class of states. These resemble Bañados geometries, except that the induced metric is flat on the Dirichlet wall at $\rho = \rho_c$. To impose this, we start with the metric Ansatz:

$$\begin{aligned} ds^2 =& \frac{d\rho^2}{\rho^2} - \rho_c^2 dx^+ dx^- \\ &+ (\rho^2 - \rho_c^2) h^{(1)}(x^+, x^-)_{\alpha\beta} dx^\alpha dx^\beta + \left(\frac{\rho_c^4}{\rho^2} - \rho_c^2\right) h^{(2)}(x^+, x^-)_{\alpha\beta} dx^\alpha dx^\beta\,, \end{aligned} \tag{6.100}$$

where we used the property of 3D gravity that expansions in $\rho$ terminate after a couple of orders. Note that setting $\rho = \rho_c$ eliminates the second line in (6.100) and the metric on the Dirichlet wall is $ds^2|_{\rho_c} = -\rho_c^2 dx^+ dx^-$. Plugging into Einstein's equations, we get that $h^{(1,2)}_{\alpha\beta}$ can be parametrized

by two functions:

$$h^{(1)}_{\alpha\beta} = \frac{c-1}{4}\, M_{\alpha\beta}\,, \quad h^{(2)}_{\alpha\beta} = \frac{c+1}{4}\, M_{\alpha\beta}\,, \quad M_{\alpha\beta} \equiv \begin{pmatrix} a & -c \\ -c & b \end{pmatrix}\,, \quad c \equiv \sqrt{ab+1}\,, \tag{6.101}$$

where the functions $a(x^+, x^-)$, $b(x^+, x^-)$ satisfy

$$\partial_- a + \partial_+ c = 0\,, \qquad \partial_+ b + \partial_- c = 0\,, \qquad c = \sqrt{ab+1}\,, \tag{6.102}$$

which is a set of coupled nonlinear PDEs.[17] These can be solved in a series form:

$$\begin{aligned} a(x^+, x^-) &= \epsilon\, A'(x^+) - \frac{\epsilon^2}{2} A''(x^+) B(x^-) + O(\epsilon^3) \\ b(x^+, x^-) &= \epsilon\, B'(x^-) - \frac{\epsilon^2}{2} A(x^+) B''(x^-) + O(\epsilon^3)\,, \end{aligned} \tag{6.105}$$

hence $c = O(\epsilon^2)$. Of course, this is locally just $AdS_3$ in complicated coordinates.

Now consider a fluctuation in the location of the Dirichlet wall $\rho_c \to \rho_c + \delta\rho(x^+, x^-)$, and require that the resulting metric stays flat

$$0 = R\left(\rho_c + \delta\rho(t,\theta)\right) = -\frac{8}{r_c^3}\left[\partial_+\partial_- + \frac{\epsilon}{2}\left(A'(x^+)\,\partial_-^2 + B'(x^-)\,\partial_+^2\right)\right]\delta\rho(x^+, x^-) + O(\epsilon^2, \delta\rho^2)\,. \tag{6.106}$$

Assuming that $A'(x^+)$ and $B'(x^-)$ are slowly varying, we get the corrected propagation speeds to be:

$$v_+ = 1 - \epsilon\, B'(x^-)\,, \qquad v_- = 1 - \epsilon\, A'(x^+)\,. \tag{6.107}$$

[17] It is instructive to write down the planar BTZ black hole in this parametrization, which takes the form:

$$ds^2 = \frac{d\rho^2}{\rho^2} - \rho_c^2\, dx^+ dx^- - \frac{2\rho_c^6}{(\rho_c^4-1)^2}\left[(\rho^2 - \rho_c^2) + \left(\frac{1}{\rho^2} - \frac{1}{\rho_c^2}\right)\right](dx^+ \quad dx^-)\begin{pmatrix} d & \frac{\rho_c^4+1}{2\rho_c^2} \\ \frac{\rho_c^4+1}{2\rho_c^2} & \frac{1}{d} \end{pmatrix}\begin{pmatrix} dx^+ \\ dx^- \end{pmatrix} \tag{6.103}$$

where we chose $\rho_c > 1$, and $d > 0$ is a parameter characterizing the solution. In the parametrization (6.101) the solution corresponds to

$$a = -\frac{2\rho_c^2}{\rho_c^4 - 1}\, d\,, \qquad b = -\frac{2\rho_c^2}{\rho_c^4 - 1}\,\frac{1}{d}\,. \tag{6.104}$$

To $O(\epsilon)$ the metric (6.100) takes the form

$$ds^2 = \frac{d\rho^2}{\rho^2} - \rho_c^2\, dx^+ dx^- + \left(\frac{\rho_c^4}{\rho^2} - \rho_c^2\right)\left(\frac{\epsilon}{2} A'(x^+)(dx^+)^2 + \frac{\epsilon}{2} B'(x^-)(dx^-)^2\right) + O(\epsilon^2)\,. \qquad (6.108)$$

From the behavior of the metric near $r = r_c$ and the usual definition of the holographic stress tensor [139], we deduce the following expectation values in the dual field theory:

$$\left\langle T_{++}(x^+)\right\rangle = -\frac{\rho_c^2}{16\pi G}\,\epsilon\, A'(x^+), \qquad \left\langle T_{--}(x^-)\right\rangle = -\frac{\rho_c^2}{16\pi G}\,\epsilon\, B'(x^-)\,. \qquad (6.109)$$

We combine this equation with (6.107) to obtain (6.72). We note that there is an intriguing connection between this equation and the Nambu-Goto string: if we rename $X^- \equiv -\epsilon B$, $X^- \equiv -\epsilon A$, we obtain the Virasoro conditions (6.40).

# Chapter 7

# Conclusions

Ultimately, all of our discussions have been motivated by the ubiquity of uncertainty in physical systems and by the ways in which different forms of uncertainty play crucial roles in the behavior of those systems.

We began by studying how a biological system, fruit fly embryonic development, can operate at the physical limits of precision despite the noise present in the signals driving its behavior. We found that morphogen concentrations cannot specify the body plan to such a precision as seen in nature if the fluctuations of the concentrations act independently at different points in space. Spatial correlations in these flucuations could vastly improve the concentration profiles' ability to communicate precise cell identities, giving the embryo the ability to develop a body plan with precision because of specific properties of the noise.

We then moved to a completely different setting where studied a different measure of information, the quantum Rényi entropy of relativistic conformal QFTs. Here, we used the fact that certain quantum field theories - CFTs in $3+1$d - have quantum uncertainty in the form of Rényi entropy which is parametrizable in terms of the geometry of the entangling surface specifying which degrees of freedom we're considering. This chapter was concerned with understanding the properties of the uncertainty as we developed and gave evidence for a conjecture which would contribute to our understanding of the Rényi entropy across a general entangling surface for these theories.

The following two chapters led us to consider uncertainty in spacetimes with black holes. There, we learned that mathematical properties of $AdS_3$ gravity and Liouville theory, the conformal field theory which is its dual in a certain sense, imply a form of entropy which is topological and a generalization of an entropy commonly studied as a kind of "order parameter" in the context of quantum

matter. Surprisingly, this topological entanglement entropy is equal to the area-law Bekenstein-Hawking entropy of the black hole, indicating a connection between black hole states and topological properties of the underlying theories. Moreover, this connection inspired us to view Liouville theory as a thermodynamic approximation to $AdS_3/CFT_2$ - an approximation which showed that chaos and thermodynamics hide detailed information about topology and scattering in the CFT and gravity theories. In other words, we found that uncertainty in the form of statistical approximations and exponential sensitivity to initial conditions are intricately connected to *exact* properties which describe physics in the gravity theory and the CFT.

Finally, we considered RG flow in the context of AdS/CFT, whereby the value of a coupling in the field theory - here, the coupling of the $T\bar{T}$ perturbing operator - controls the energy scale of the CFT, and thus controls the uncertainty caused by integrating out certain high energy modes. In this, we were able to describe this form of holographic RG completely and thus construct an exact match with a dual theory of "black hole thermodynamics in a box." We quantitatively described the process of integrating out degrees of freedom in a specific integrable field theory amenable to calculation, and as a result, were able to better understand an example of holography with a boundary theory perturbed by an an *irrelevant* operator.

By characterizing uncertainty and its physical consequences – both those to be overcome and those to be taken advantage of – we have learned new aspects of systems as diverse as embryonic development, relativistic quantum field theory, black holes in $2+1$d, and AdS/CFT. Indeed, uncertainty in a system often characterizes crucial physical properties and is associated with specific order parameters which describe that system. And still, the questions along these lines are seemingly unending, and they seem to probe deep aspects of our physical universe. On the one hand, how can biological systems overcome noise and stochasticity to thrive in a wide variety of environments? On the other hand, how is it that entanglement in field theory appears to be so fundamentally related to the emergence of geometry in gravitational theories? And in between, how do we measure topological entanglement entropy, or get around the decoherence caused by quantum entanglement with the environment?

We are led to conclude that although noise and uncertainty are ubiquitous at every scale and in every context of the physicist's studies, when we attempt to *characterize* our ignorance which is due to the intrinsic uncertainty of the *system*, we end up *reducing* our *own* ignorance which is not due to intrinsic uncertainty of the system, in ways we could not have done without methodically confronting the uncertainty to begin with.

# Bibliography


[1] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, New York, NY, USA, 2002.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[3] C. E. Shannon, *A mathematical theory of communication*, *SIGMOBILE Mob. Comput. Commun. Rev.* **5** (Jan., 2001) 3–55.

[4] L. P. Kadanoff, *Scaling laws for ising models near $T_c$*, *Physics Physique Fizika* **2** (Jun, 1966) 263–272.

[5] K. G. Wilson, *The renormalization group: Critical phenomena and the kondo problem*, *Rev. Mod. Phys.* **47** (1975) 773.

[6] A. B. Zamolodchikov, *Irreversibility of the Flux of the Renormalization Group in a 2D Field Theory*, *JETP Lett.* **43** (1986) 730–732.

[7] Z. Komargodski and A. Schwimmer, *On Renormalization group flows in four dimensions*, `arXiv:1107.3987`.

[8] Z. Komargodski, *The constraints of conformal symmetry on RG flows*, *Journal of High Energy Physics* **7** (July, 2012) 69, [`arXiv:1112.4538`].

[9] H. Casini and M. Huerta, *On the RG running of the entanglement entropy of a circle*, `arXiv:1202.5650`.

[10] H. Liu and M. Mezei, *A Refinement of entanglement entropy and the number of degrees of freedom*, *JHEP* **1304** (2013) 162, [`arXiv:1202.2070`].

[11] I. Klebanov, S. Pufu, and B. Safdi, *F-Theorem without supersymmetry*, *JHEP* **1110** (2011) 038, [arXiv:1105.4598].

[12] H. Casini and M. Huerta, *Renormalization group running of the entanglement entropy of a circle*, *Phys. Rev. D* **85** (June, 2012) 125016, [arXiv:1202.5650].

[13] J. M. Maldacena, *The Large N limit of superconformal field theories and supergravity*, *Int. J. Theor. Phys.* **38** (1999) 1113–1133, [hep-th/9711200]. [Adv. Theor. Math. Phys.2,231(1998)].

[14] S. S. Gubser, I. R. Klebanov, and A. M. Polyakov, *Gauge theory correlators from noncritical string theory*, *Phys. Lett.* **B428** (1998) 105–114, [hep-th/9802109].

[15] E. Witten, *Anti-de Sitter space, thermal phase transition, and confinement in gauge theories*, *Adv. Theor. Math. Phys.* **2** (1998) 505–532, [hep-th/9803131].

[16] O. Aharony, S. S. Gubser, J. M. Maldacena, H. Ooguri, and Y. Oz, *Large N field theories, string theory and gravity*, *Phys. Rept.* **323** (2000) 183–386, [hep-th/9905111].

[17] E. Witten, *Anti-de Sitter space and holography*, *Adv. Theor. Math. Phys.* **2** (1998) 253–291, [hep-th/9802150].

[18] L. Susskind, *The World as a hologram*, *J. Math. Phys.* **36** (1995) 6377–6396, [hep-th/9409089].

[19] G. 't Hooft, *The Holographic principle: Opening lecture*, *Subnucl. Ser.* **37** (2001) 72–100, [hep-th/0003004].

[20] R. Bousso, *The Holographic principle*, *Rev. Mod. Phys.* **74** (2002) 825–874, [hep-th/0203101].

[21] I. Heemskerk, J. Penedones, J. Polchinski, and J. Sully, *Holography from Conformal Field Theory*, arXiv:0907.0151.

[22] T. Faulkner, H. Liu, and M. Rangamani, *Integrating out geometry: Holographic Wilsonian RG and the membrane paradigm*, *JHEP* **08** (2011) 051, [arXiv:1010.4036].

[23] S. Ryu and T. Takayanagi, *Aspects of Holographic Entanglement Entropy*, *JHEP* **08** (2006) 045, [hep-th/0605073].

[24] V. E. Hubeny, M. Rangamani, and T. Takayanagi, *A Covariant holographic entanglement entropy proposal*, *JHEP* **07** (2007) 062, [arXiv:0705.0016].

[25] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *JHEP* **1308** (2013) 090, [arXiv:1304.4926].

[26] X. Dong, A. Lewkowycz, and M. Rangamani, *Deriving covariant holographic entanglement*, *JHEP* **11** (2016) 028, [arXiv:1607.0750].

[27] E. Mintun, J. Polchinski, and V. Rosenhaus, *Bulk-Boundary Duality, Gauge Invariance, and Quantum Error Corrections*, *Phys. Rev. Lett.* **115** (2015), no. 15 151601, [arXiv:1501.0657].

[28] R. Bousso, B. Freivogel, S. Leichenauer, V. Rosenhaus, and C. Zukowski, *Null Geodesics, Local CFT Operators and AdS/CFT for Subregions*, *Phys. Rev.* **D88** (2013) 064057, [arXiv:1209.4641].

[29] B. Czech, J. L. Karczmarek, F. Nogueira, and M. Van Raamsdonk, *The Gravity Dual of a Density Matrix*, *Class. Quant. Grav.* **29** (2012) 155009, [arXiv:1204.1330].

[30] J. Cotler, P. Hayden, G. Salton, B. Swingle, and M. Walter, *Entanglement Wedge Reconstruction via Universal Recovery Channels*, arXiv:1704.0583.

[31] X. Dong, D. Harlow, and A. C. Wall, *Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality*, *Phys. Rev. Lett.* **117** (2016), no. 2 021601, [arXiv:1601.0541].

[32] A. Almheiri, X. Dong, and D. Harlow, *Bulk Locality and Quantum Error Correction in AdS/CFT*, *JHEP* **04** (2015) 163, [arXiv:1411.7041].

[33] I. A. Morrison, *Boundary-to-bulk maps for AdS causal wedges and the Reeh-Schlieder property in holography*, *JHEP* **05** (2014) 053, [arXiv:1403.3426].

[34] L. McGough and W. Bialek, *Reproducibility from correlated signals in embryo development*, *to appear* (Aug, 2018).

[35] J. O. Dubuis, G. Tkačik, E. F. Wieschaus, T. Gregor, and W. Bialek, *Positional information, in bits*, *Proceedings of the National Academy of Sciences* **110** (2013), no. 41 16301–16308.

[36] D. Krotov, J. O. Dubuis, T. Gregor, and W. Bialek, *Morphogenesis at criticality*, *Proceedings of the National Academy of Sciences* **111** (2014), no. 10 3683–3688, [http://www.pnas.org/content/111/10/3683.full.pdf].

[37] J. G. Smith, *The information capacity of amplitude- and variance-constrained scalar gaussian channels*, *Information and Control* **18** (April, 1971) 203–219.

[38] H. H. Mattingly, M. K. Transtrum, M. C. Abbott, and B. B. Machta, *Rational ignorance: simpler models learn more from finite data.*, *CoRR* **abs/1705.01166** (2017).

[39] M. D. Petkova, G. Tkačik, W. Bialek, E. F. Wieschaus, and T. Gregor, *Optimal decoding of information from a genetic network*, *ArXiv e-prints* (Dec., 2016) [arXiv:1612.0808].

[40] l. C. C. C. Wikimedia Commons, *Drosophila embryogenesis, mrna distributions*, 2014.

[41] l. C. C. C. Wikimedia Commons, *Drosophila embryogenesis, protein gradients*, 2014.

[42] D. Busson, *Peter a. lawrence (1992); the making of a fly: The genetics of animal design, blackwell scientific publications, oxford, 229 pp. 16.95. isbn: 0632300488*, *Journal of Evolutionary Biology* **6** no. 4 609–610, [https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1420-9101.1993.6040609.x].

[43] M. D. Petkova, S. C. Little, F. Liu, and T. Gregor, *Maternal origins of developmental reproducibility*, *Current Biology* **24** 2018/04/05 1283–1288.

[44] T. Gregor, D. W. Tank, E. F. Wieschaus, and W. Bialek, *Probing the limits to positional information*, *Cell* **130** (2007), no. 1 153 – 164.

[45] S. C. Little, G. Tkaik, T. B. Kneeland, E. F. Wieschaus, and T. Gregor, *The formation of the bicoid morphogen gradient requires protein movement from anteriorly localized mrna*, *PLOS Biology* **9** (03, 2011) 1–17.

[46] b. R. G. h. CC BY-SA 4.0 (https://creativecommons.org/licenses/by sa/4.0)], from Wikimedia Commons, *Pair-rule gene*, 2014.

[47] C. Nüsslein-Volhard and E. Wieschaus, *Mutations affecting segment number and polarity in drosophila*, *Nature* **287** (10, 1980) 795 EP –.

[48] L. Abouchar, M. D. Petkova, C. R. Steinhardt, and T. Gregor, *Precision and reproducibility of macroscopic developmental patterns*, *ArXiv e-prints* (Sept., 2013) [arXiv:1309.6273].

[49] C. E. Shannon, *Communication in the presence of noise*, *Proc. Institute of Radio Engineers* **37** (1949), no. 1 10–21.

[50] T. Gregor and M. Petkova. Personal communication; in progress.

[51] J. Lee, L. McGough, and B. R. Safdi, *Rnyi entropy and geometry*, *Phys. Rev.* **D89** (2014), no. 12 125016, [arXiv:1403.1580].

[52] S. N. Solodukhin, *Entanglement entropy, conformal invariance and extrinsic geometry*, *Phys.Lett.* **B665** (2008) 305–309, [arXiv:0802.3117].

[53] L.-Y. Hung, R. C. Myers, M. Smolkin, and A. Yale, *Holographic Calculations of Renyi Entropy*, *JHEP* **1112** (2011) 047, [arXiv:1110.1084].

[54] H. Casini and M. Huerta, *Entanglement entropy for the n-sphere*, *Phys.Lett.* **B694** (2010) 167–171, [arXiv:1007.1813].

[55] D. Fursaev and G. Miele, *Finite temperature scalar field theory in static de Sitter space*, *Phys.Rev.* **D49** (1994) 987–998, [hep-th/9302078].

[56] L. De Nardo, D. V. Fursaev, and G. Miele, *Heat kernel coefficients and spectra of the vector Laplacians on spherical domains with conical singularities*, *Class.Quant.Grav.* **14** (1997) 1059–1078, [hep-th/9610011].

[57] D. V. Fursaev, *Entanglement Renyi Entropies in Conformal Field Theories and Holography*, arXiv:1201.1702.

[58] I. R. Klebanov, S. S. Pufu, S. Sachdev, and B. R. Safdi, *Renyi Entropies for Free Field Theories*, *JHEP* **1204** (2012) 074, [arXiv:1111.6290].

[59] M. Huerta, *Numerical Determination of the Entanglement Entropy for Free Fields in the Cylinder*, *Phys.Lett.* **B710** (2012) 691–696, [arXiv:1112.1277].

[60] I. R. Klebanov, T. Nishioka, S. S. Pufu, and B. R. Safdi, *On Shape Dependence and RG Flow of Entanglement Entropy*, *JHEP* **1207** (2012) 001, [arXiv:1204.4160].

[61] B. R. Safdi, *Exact and Numerical Results on Entanglement Entropy in (5+1)-Dimensional CFT*, *JHEP* **1212** (2012) 005, [arXiv:1206.5025].

[62] H. Casini and M. Huerta, *Entanglement entropy in free quantum field theory*, *J.Phys.A* **A42** (2009) 504007, [arXiv:0905.2562].

[63] M. Srednicki, *Entropy and area*, *Phys.Rev.Lett.* **71** (1993) 666–669, [hep-th/9303048].

[64] I. Peschel, *Letter to the editor: Calculation of reduced density matrices from correlation functions*, *Journal of Physics A Mathematical General* **36** (Apr., 2003) L205–L208, [cond-mat/0212631].

[65] I. R. Klebanov, T. Nishioka, S. S. Pufu, and B. R. Safdi, *Is Renormalized Entanglement Entropy Stationary at RG Fixed Points?*, *JHEP* **1210** (2012) 058, [arXiv:1207.3360].

[66] M. P. Hertzberg and F. Wilczek, *Some Calculable Contributions to Entanglement Entropy*, *Phys.Rev.Lett.* **106** (2011) 050404, [arXiv:1007.0993].

[67] S. N. Solodukhin, *Entanglement entropy of black holes*, *Living Rev.Rel.* **14** (2011) 8, [arXiv:1104.3712].

[68] D. V. Fursaev, *Spectral geometry and one loop divergences on manifolds with conical singularities*, *Phys.Lett.* **B334** (1994) 53–60, [hep-th/9405143].

[69] D. V. Fursaev and S. N. Solodukhin, *On the description of the Riemannian geometry in the presence of conical defects*, *Phys.Rev.* **D52** (1995) 2133–2143, [hep-th/9501127].

[70] A. Lewkowycz, R. C. Myers, and M. Smolkin, *Observations on entanglement entropy in massive QFT's*, *JHEP* **1304** (2013) 017, [arXiv:1210.6858].

[71] R. Lohmayer, H. Neuberger, A. Schwimmer, and S. Theisen, *Numerical determination of entanglement entropy for a sphere*, *Phys.Lett.* **B685** (2010) 222–227, [arXiv:0911.4283].

[72] L. McGough and H. Verlinde, *Bekenstein-Hawking entropy as topological entanglement entropy*, *Journal of High Energy Physics* **11** (Nov., 2013) 208, [arXiv:1308.2342].

[73] S. W. Hawking, *Particle Creation by Black Holes*, *Commun. Math. Phys.* **43** (1975) 199–220. [,167(1975)].

[74] J. D. Bekenstein, *Black holes and entropy*, *Phys. Rev.* **D7** (1973) 2333–2346.

[75] A. Strominger and C. Vafa, *Microscopic Origin of the Bekenstein-Hawking Entropy*, *Phys. Lett.* **B379** (1996) 99–104, [hep-th/9601029].

[76] L. Bombelli, R. K. Koul, J. Lee, and R. D. Sorkin, *A Quantum Source of Entropy for Black Holes*, *Phys. Rev.* **D34** (1986) 373–383.

[77] C. Callan and F. Wilczek, *On geometric entropy*, *Physics Letters B* **333** (1994), no. 1–2 55 – 61.

[78] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, *Phys.Rev.Lett.* **96** (2006) 181602, [hep-th/0603001].

[79] M. Van Raamsdonk, *Comments on quantum gravity and entanglement*, arXiv:0907.2939.

[80] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, *Fortsch. Phys.* **61** (2013) 781–811, [arXiv:1306.0533].

[81] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, *Black Holes: Complementarity or Firewalls?*, *JHEP* **02** (2013) 062, [arXiv:1207.3123].

[82] S. L. Braunstein, S. Pirandola, and K. yczkowski, *Better Late than Never: Information Retrieval from Black Holes*, *Phys. Rev. Lett.* **110** (2013), no. 10 101301, [arXiv:0907.1190].

[83] E. Witten, *(2+1)-Dimensional Gravity as an Exactly Soluble System*, *Nucl. Phys.* **B311** (1988) 46.

[84] E. Witten, *Quantum field theory and the Jones polynomial*, *Commun. Math. Phys.* **121** (1989) 351.

[85] A. Kitaev and J. Preskill, *Topological entanglement entropy*, *Phys.Rev.Lett.* **96** (2006) 110404, [hep-th/0510092].

[86] M. Levin and X.-G. Wen, *Detecting topological order in a ground state wave function*, *Phys.Rev.Lett.* **96** (2006) 110405, [cond-mat/0510613].

[87] C. Nayak, S. H. Simon, A. Stern, M. Freedman, and S. Das Sarma, *Non-Abelian anyons and topological quantum computation*, *Reviews of Modern Physics* **80** (July, 2008) 1083–1159, [arXiv:0707.1889].

[88] S. Dong, E. Fradkin, R. G. Leigh, and S. Nowling, *Topological entanglement entropy in Chern-Simons theories and quantum Hall fluids*, *Journal of High Energy Physics* **5** (May, 2008) 16, [arXiv:0802.3231].

[89] M. Banados, C. Teitelboim, and J. Zanelli, *The Black hole in three-dimensional space-time*, *Phys. Rev. Lett.* **69** (1992) 1849–1851, [hep-th/9204099].

[90] M. Banados, M. Henneaux, C. Teitelboim, and J. Zanelli, *Geometry of the (2+1) black hole*, *Phys. Rev.* **D48** (1993) 1506–1525, [gr-qc/9302012]. [Erratum: Phys. Rev.D88,069902(2013)].

[91] E. Verlinde and H. Verlinde, *Passing through the Firewall*, arXiv:1306.0515.

[92] H. L. Verlinde, *Conformal Field Theory, 2-D Quantum Gravity and Quantization of Teichmuller Space*, *Nucl. Phys.* **B337** (1990) 652–680.

[93] H. L. Verlinde and E. P. Verlinde, *CONFORMAL FIELD THEORY AND GEOMETRIC QUANTIZATION*, in *Trieste School and Workshop on Superstrings Trieste, Italy, April 3-14, 1989*, pp. 422–449, 1989.

[94] S. Carlip, *Conformal field theory, (2+1)-dimensional gravity, and the BTZ black hole*, *Class. Quant. Grav.* **22** (2005) R85–R124, [gr-qc/0503022].

[95] Y.-j. Chen, *Quantum Liouville theory and BTZ black hole entropy*, *Class. Quant. Grav.* **21** (2004) 1153–1180, [hep-th/0310234].

[96] A. B. Zamolodchikov and A. B. Zamolodchikov, *Liouville field theory on a pseudosphere*, hep-th/0101152.

[97] J. Teschner, *Liouville theory revisited*, *Class. Quant. Grav.* **18** (2001) R153–R222, [hep-th/0104158].

[98] J. Teschner, *On the relation between quantum Liouville theory and the quantized Teichmuller spaces*, *Int. J. Mod. Phys.* **A19S2** (2004) 459–477, [hep-th/0303149].

[99] J. Teschner, *Nonrational conformal field theory*, arXiv:0803.0919.

[100] J. D. Brown and M. Henneaux, *Central Charges in the Canonical Realization of Asymptotic Symmetries: An Example from Three-Dimensional Gravity*, *Commun. Math. Phys.* **104** (1986) 207–226.

[101] N. Seiberg, *Notes on quantum Liouville theory and quantum gravity*, *Prog. Theor. Phys. Suppl.* **102** (1990) 319–349.

[102] T. Hartman, C. A. Keller, and B. Stoica, *Universal Spectrum of 2d Conformal Field Theory in the Large c Limit*, *JHEP* **09** (2014) 118, [arXiv:1405.5137].

[103] J. L. Cardy, *Operator Content of Two-Dimensional Conformally Invariant Theories*, *Nucl. Phys.* **B270** (1986) 186–204.

[104] E. P. Verlinde, *Fusion Rules and Modular Transformations in 2D Conformal Field Theory*, *Nucl. Phys.* **B300** (1988) 360–376.

[105] J. L. Cardy, *Boundary Conditions, Fusion Rules and the Verlinde Formula*, *Nucl. Phys.* **B324** (1989) 581–596.

[106] R. Dijkgraaf and E. Verlinde, *Modular invariance and the fusion algebra*, *Nuclear Physics B - Proceedings Supplements* **5** (1988), no. 2 87 – 97.

[107] M. R. Gaberdiel and R. Gopakumar, *Minimal Model Holography*, *J. Phys.* **A46** (2013) 214002, [arXiv:1207.6697].

[108] M. Ammon, M. Gutperle, P. Kraus, and E. Perlmutter, *Black holes in three dimensional higher spin gravity: A review*, *J. Phys.* **A46** (2013) 214001, [arXiv:1208.5182].

[109] J. de Boer and J. I. Jottar, *Thermodynamics of higher spin black holes in $AdS_3$*, *JHEP* **01** (2014) 023, [arXiv:1302.0816].

[110] A. Perez, D. Tempo, and R. Troncoso, *Higher spin gravity in 3D: Black holes, global charges and thermodynamics*, *Phys. Lett.* **B726** (2013) 444–449, [arXiv:1207.2844].

[111] N. Drukker, D. Gaiotto, and J. Gomis, *The Virtue of Defects in 4D Gauge Theories and 2D CFTs*, *JHEP* **06** (2011) 025, [arXiv:1003.1112].

[112] A. Maloney and E. Witten, *Quantum Gravity Partition Functions in Three Dimensions*, *JHEP* **02** (2010) 029, [arXiv:0712.0155].

[113] I. Affleck and A. W. W. Ludwig, *Universal noninteger 'ground state degeneracy' in critical quantum systems*, *Phys. Rev. Lett.* **67** (1991) 161–164.

[114] S. Jackson, L. McGough, and H. Verlinde, *Conformal Bootstrap, Universality and Gravitational Scattering*, *Nucl. Phys.* **B901** (2015) 382–429, [arXiv:1412.5205].

[115] S. H. Shenker and D. Stanford, *Black holes and the butterfly effect*, *JHEP* **03** (2014) 067, [arXiv:1306.0622].

[116] S. H. Shenker and D. Stanford, *Black holes and the butterfly effect*, *JHEP* **03** (2014) 067, [arXiv:1306.0622].

[117] K.-H. Rehren, *Locality of conformal fields in two dimensions: Exchange algebra on the light-cone*, *Communications in Mathematical Physics* **116** (1988), no. 4 675–688.

[118] J. Teschner and G. Vartanov, *6j symbols for the modular double, quantum hyperbolic geometry, and supersymmetric gauge theories*, *Lett. Math. Phys.* **104** (2014) 527–551, [arXiv:1202.4698].

[119] A. B. Zamolodchikov and A. B. Zamolodchikov, *Structure constants and conformal bootstrap in Liouville field theory*, *Nucl. Phys.* **B477** (1996) 577–605, [hep-th/9506136].

[120] B. Ponsot and J. Teschner, *Liouville bootstrap via harmonic analysis on a noncompact quantum group*, hep-th/9911110.

[121] T. Dray and G. 't Hooft, *The Effect of Spherical Shells of Matter on the Schwarzschild Black Hole*, *Commun. Math. Phys.* **99** (1985) 613–625.

[122] T. Dray and G. 't Hooft, *The Gravitational Shock Wave of a Massless Particle*, *Nucl. Phys.* **B253** (1985) 173–188.

[123] Y. Kiem, H. L. Verlinde, and E. P. Verlinde, *Black hole horizons and complementarity*, *Phys. Rev.* **D52** (1995) 7053–7065, [hep-th/9502074].

[124] K. Schoutens, H. L. Verlinde, and E. P. Verlinde, *Quantum black hole evaporation*, *Phys. Rev.* **D48** (1993) 2670–2685, [hep-th/9304128].

[125] G. W. Moore and N. Seiberg, *Polynomial Equations for Rational Conformal Field Theories*, *Phys. Lett.* **B212** (1988) 451–460.

[126] J. Murakami and M. Yano, *On the volume of a hyperbolic and spherical tetrahedron*, *Communications in Analysis and Geometry* **13** (2005), no. 2 379–400.

[127] N. A. Nekrasov, A. A. Rosly, and S. L. Shatashvili, *Darboux coordinates, Yang-Yang functional, and gauge theory*, *Theor. Math. Phys.* **181** (2014), no. 1 1206–1234. [Erratum: Theor. Math. Phys.182,no.2,368(2015)].

[128] L. McGough, M. Mezei, and H. Verlinde, *Moving the CFT into the bulk with $T\bar{T}$*, *ArXiv e-prints* (Nov., 2016) [arXiv:1611.0347].

[129] F. A. Smirnov and A. B. Zamolodchikov, *On space of integrable quantum field theories*, arXiv:1608.0549.

[130] A. Cavaglià, S. Negro, I. M. Szécsényi, and R. Tateo, *$T\bar{T}$-deformed 2D Quantum Field Theories*, *JHEP* **10** (2016) 112, [arXiv:1608.0553].

[131] S. Dubovsky, R. Flauger, and V. Gorbenko, *Solving the Simplest Theory of Quantum Gravity*, *JHEP* **09** (2012) 133, [arXiv:1205.6805].

[132] I. Heemskerk and J. Polchinski, *Holographic and Wilsonian Renormalization Groups*, *JHEP* **06** (2011) 031, [arXiv:1010.1264].

[133] J. de Boer, E. P. Verlinde, and H. L. Verlinde, *On the holographic renormalization group*, *JHEP* **08** (2000) 003, [hep-th/9912012].

[134] E. P. Verlinde and H. L. Verlinde, *RG flow, gravity and the cosmological constant*, *JHEP* **05** (2000) 034, [hep-th/9912018].

[135] D. Harlow and D. Stanford, *Operator Dictionaries and Wave Functions in AdS/CFT and dS/CFT*, arXiv:1104.2621.

[136] J. Cardy, *Quantum Quenches to a Critical Point in One Dimension: some further results*, *J. Stat. Mech.* **1602** (2016), no. 2 023103, [arXiv:1507.0726].

[137] M. Banados, C. Teitelboim, and J. Zanelli, *The Black hole in three-dimensional space-time*, *Phys. Rev. Lett.* **69** (1992) 1849–1851, [hep-th/9204099].

[138] D. Marolf and M. Rangamani, *Causality and the AdS Dirichlet problem*, *JHEP* **04** (2012) 035, [arXiv:1201.1233].

[139] V. Balasubramanian and P. Kraus, *A Stress tensor for Anti-de Sitter gravity*, *Commun. Math. Phys.* **208** (1999) 413–428, [hep-th/9902121].

[140] J. D. Brown, J. Creighton, and R. B. Mann, *Temperature, energy and heat capacity of asymptotically anti-de Sitter black holes*, *Phys. Rev.* **D50** (1994) 6394–6403, [gr-qc/9405007].

[141] A. B. Zamolodchikov, *Expectation value of composite field T anti-T in two-dimensional quantum field theory*, hep-th/0401146.

[142] L. Freidel, *Reconstructing AdS/CFT*, arXiv:0804.0632.

[143] G. 't Hooft, *Graviton Dominance in Ultrahigh-Energy Scattering*, *Phys. Lett.* **B198** (1987) 61–63.

[144] D. A. Roberts, D. Stanford, and L. Susskind, *Localized shocks*, *JHEP* **03** (2015) 051, [arXiv:1409.8180].

[145] J. Maldacena, S. H. Shenker, and D. Stanford, *A bound on chaos*, *JHEP* **08** (2016) 106, [arXiv:1503.0140].

[146] P. Goddard, J. Goldstone, C. Rebbi, and C. B. Thorn, *Quantum dynamics of a massless relativistic string*, *Nucl. Phys.* **B56** (1973) 109–135.

[147] R. C. Brower, *Spectrum-generating algebra and no-ghost theorem for the dual model*, *Phys. Rev. D* **6** (Sep, 1972) 1655–1662.

[148] J. Scherk, *An Introduction to the Theory of Dual Models and Strings*, *Rev. Mod. Phys.* **47** (1975) 123–164.

[149] K. Schoutens, H. L. Verlinde, and E. P. Verlinde, *Quantum black hole evaporation*, *Phys. Rev.* **D48** (1993) 2670–2685, [hep-th/9304128].

[150] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, *Causality, analyticity and an IR obstruction to UV completion*, *JHEP* **10** (2006) 014, [hep-th/0602178].

[151] M. Mezei and D. Stanford, *On entanglement spreading in chaotic systems*, arXiv:1608.0510.

[152] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, *Phys. Rev. Lett.* **96** (2006) 181602, [hep-th/0603001].

[153] K. Skenderis and S. N. Solodukhin, *Quantum effective action from the AdS / CFT correspondence*, *Phys. Lett.* **B472** (2000) 316–322, [hep-th/9910023].