INFORMATION THEORETIC INTERPRETATIONS OF RENORMALIZATION GROUP FLOW

Raymond Eveleth Fowler III

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Physics.

Chapel Hill
2019

Approved by:

Gerald Cecil

Louise Dolan

Joaquin Drut

Jonathan Heckman

Jack Ng

ABSTRACT

Raymond Eveleth Fowler III: Information Theoretic Interpretations of Renormalization Group Flow
(Under the direction Louise Dolan and Jonathan Heckman)

We interpret the renormalization group flow between quantum field theories as a communication channel problem, which allows us to quantify UV-IR mixing in terms of information theoretic quantities, i.e., we can quantify the information of the UV theory that remains accessible in the IR theory. Because of the AdS-CFT interpretation of the renormalization group flow, our interpretation applies to the AdS-CFT correspondence too.

In our interpretation, the UV variables are the input signal for the channel, the output variables are the IR variables resulting from the renormalization group flow, and the renormalization group transformation is viewed as the communication channel. To make this interpretation, we make use of the Kullback-Leibler (KL) divergence, which quantifies the information theoretic distance between two probability distributions. In order to use the KL divergence with quantum field theories, we study the probability distributions associated with Euclidean quantum field theories; the KL divergence thus computes the relative entropy between these Euclidean quantum field theories, as in statistical field theory. We then use the renormalization group and the techniques of effective field theory to find the probability distributions that we need to use with the KL divergence in order to measure the information lost upon performing the renormalization group transformation.

We study both lattice and continuum field theories and find the probability distributions needed to measure UV-IR mixing in both scenarios. We also find that the results of decimation on a lattice in the continuum limit matches the continuum results. We finally compute a number of explicit examples. For lattice examples, we primarily use the 1D and 2D Ising model and the 1D Ising model on a tree. We also sketch a calculation of some continuum models, including $\lambda\phi^4$ theory, the $T\bar{T}$ deformation to a 2D free fermionic system, and a Kaluza-Klein compactification of a scalar field theory. From our example calculations of the 2D Ising model, we find a regime where the KL divergence is negative, and we show

that this is due to UV effects that were not properly accounted for in the approximation that was used when decimating the 2D Ising model.

TABLE OF CONTENTS

LIST OF FIGURES viii

LIST OF ABBREVIATIONS xi

1 Introduction 1

2 Effective Field Theories in the Standard Model and Beyond 7

2.1 Standard Model–Overview 9

2.2 Excursis: Hierarchy Problem 14

2.3 Supersymmetry 16

2.4 String Compactification 18

2.5 Excursis: Kaluza-Klein Compactification 20

2.6 Intersecting Branes and Chiral Matter from M/F theory 21

2.7 Excursis: Warped Large Extra Dimensions 25

3 Renormalization Group–Overview 28

3.1 Kadanoff 28

3.2 Wilson 31

3.3 Wilson and the Callan-Symanzik Equation 33

3.4 Polchinski 34

3.5 AdS/CFT Holography 36

3.6 RG in General 37

4 Communication Channel 39

4.1 The KL Divergence 39

4.2 Channel Capacity and UV/IR Mixing 44

4.3 Communication Channel: RG Connection 46

4.4 Information Bottleneck 48
4.5 Quantum Relative Entropy 51

5 Decimation 53
5.1 Different Decimation Procedures 58
5.2 Thermodynamic Interpretation 59

6 Continuum 61
6.1 CFTs 61
6.2 Continuum RG Result 62
6.3 Fast and Slow Modes 65
6.4 Matching Lattice to Continuum: Part 1 68
6.5 Matching Lattice to Continuum: Part 2 71

7 Warm-up Example: Ising Model Perturbations 74
7.1 1D Perturbations 76

8 Example: Ising Model Decimation 84
8.1 1D Decimation 84
8.1.1 1D Decimation: Rescaling 91
8.2 2D Decimation 94
8.2.1 2D Decimation: Rescaling 106
8.2.2 2D Decimation: Mean Field Theory 110
8.3 (1 + epsilon) Dimensions 112

9 Channel Capacity Example: Ising Model 116
9.1 1D Ising Model on Tree 116

10 Continuum Examples 122
10.1 Scalar field 122
10.2 Irrelevant deformation 123
10.3 Kaluza-Klein tower of masses 125

11 Conclusion 127

APPENDIX A: MATHEMATICA CODE FOR 1D ISING MODEL FIGURES 130

APPENDIX B: MATHEMATICA CODE FOR 2D ISING MODEL FIGURES 137

REFERENCES 145

LIST OF FIGURES

1.1 A schematic representation of viewing the RG transformation as a noisy communication channel. The variables of the UV theory are the input. The RG transformation is the noisy channel that produces the output variables of the IR theory. 2

1.2 A schematic representation of viewing the RG transformation as communication through an information bottleneck. The variables of the UV theory are the input. The renormalization/coarse-graining step of the RG transformation performs the compression to $\tilde{X}$. The re-scaling step sends the compressed variables to the output IR theory. 3

2.1 A representation of effective theories of particle physics. We start with the Planck scale and go all the way down to the scale of the electron mass. At each scale, a new effective theory with its own fields becomes the useful description, especially as symmetries are broken with the lowering of the energy scale. 8

3.1 An example of grouping spins into blocks in a 2D lattice. In this particular case of blocking, the blocks are squares and contain nine spins; the center spin is chosen for the block, resulting in a lattice with fewer spins and a greater spacing between spins. Other blocking methods could be chosen, e.g., triangles, and different amounts of spins could be included in the blocks. 29

5.1 A 2D lattice of with sites labeled as black and white as a preparation to decimating the lattice. 53

5.2 A 2D lattice of white sites after decimating the black sites. The lattice is rotated by 45 degrees relative to its original orientation, as can be seen by simply rotating it by 45 degrees, and the nearest neighbors lie along the diagonals of the squares of the original lattice, resulting in a lattice spacing of $\sqrt{2}$ times the original lattice spacing. 54

5.3 A 1D chain of sites labeled as black and white as preparation for decimating the lattice. The lines show which sites are coupled to each other (nearest neighbors). 59

5.4 Result of decimation of white sites of 1D lattice before rescaling. Rescaling just changes the spacing between the sites to the original lattice spacing. 59

5.5 Result of decimating white sites of 1D lattice, then black sites, and forming a joint distribution. 59

8.1 The KL density for the 1D Ising model as a function of decimation step for fixed, low values of βJ. There is no critical value for the 1D Ising model, but we cover the parameter space that includes the critical value of the 2D Ising model. 86

8.2 The KL density for the 1D Ising model as a function of decimation step for fixed, higher values of βJ. We start to see the curves converge on the final curve shown in this figure. 87

8.3 The beta function of the KL density for the 1D Ising model as a function of decimation step for fixed, low values of βJ. These are the same values that were used for the KL density at low values of βJ..... 89

8.4 The beta function of the KL density for the 1D Ising model as a function of decimation step for fixed, higher values of βJ. These are the same values that were used for the KL density at higher values of βJ. 90

8.5 The KL density for the 1D Ising model as a function of decimation step for fixed, low values of βJ. The distribution q is formed from decimation plus rescaling. There is no critical value for the 1D Ising model, but we cover the parameter space that includes the critical value of the 2D Ising model. 92

8.6 The KL density for the 1D Ising model as a function of decimation step for fixed, high values of βJ. The distribution q is formed from decimation plus rescaling. We start to see the curves converge on the final curve shown in this figure. 93

8.7 The KL density for the 2D Ising model as a function of decimation step for fixed values of βJ above the critical value ≈ 0.50698. 96

8.8 The KL density for the 2D Ising model as a function of decimation step for fixed values of βJ below the critical value ≈ 0.50698. 97

8.9 The beta function of the KL density for the 2D Ising model as a function of decimation step for fixed values of βJ above the critical value ≈ 0.50698. 98

8.10 The beta function of the KL density for the 2D Ising model as a function of decimation step for fixed values of βJ below the critical value ≈ 0.50698. 99

8.11 The KL density for the 2D Ising model as a function of the coupling βJ for various decimation steps. Only a few steps are shown because the curves merge to the exact same curve around decimation step 8, i.e., the KL density stops significantly changing after each decimation step after around eight decimation steps. The critical value is ≈ 0.50698..... 101

8.12 The KL density for the 2D Ising model as a function of the coupling βJ after one decimation step. The critical value is ≈ 0.50698. 102

8.13 The KL density for the 2D Ising model as a function of decimation step for fixed values of βJ below the critical value of ≈ 0.50698. The distribution q is formed from decimation plus rescaling. There are similarities to the results when using decimation plus forming a joint distribution for values below the critical value. 108

8.14 The KL density for the 2D Ising model as a function of decimation step for fixed values of βJ above the critical value of ≈ 0.50698. The distribution q is formed from decimation plus rescaling. These are very different results from when using decimation plus forming a joint distribution: notably, the curves are non-negative in the present case. 109

LIST OF ABBREVIATIONS

ADE	An object that corresponds to the A, D, or E type of simply laced Dynkin diagrams
AdS	Anti-de Sitter space
CFT	Conformal Field Theory
EFT	Effective Field Theory
EW	Electroweak
h.c.	Hermitian Conjugate
GUT	Grand Unified Theory
IR	Infrared
KL	Kullback-Leibler
KK	Kaluza-Klein
LED	Large Extra Dimension
MERA	Multi-scale Entanglement Renormalization Ansatz
MSSM	Minimally Supersymmetric Standard Model
RG	Renormalization Group
SM	Standard Model
SUGRA	Supergravity
SUSY	Supersymmetry
QED	Quantum Electrodynamics
QCD	Quantum Chromodynamics
QFT	Quantum Field Theory
UNC	University of North Carolina at Chapel Hill
UV	Ultraviolet

CHAPTER 1
Introduction

Effective field theories provide a general framework for organizing and calculating physical phenomena. Different particles and physical phenomena become relevant for calculations and observation at different energy and distance scales. Accordingly, there are different effective field theories for the different scales at which physics occurs: the effective field theory at each scale allows for making calculations taking into consideration only those phenomena that are relevant to the scale of the problem, simplifying calculations. The effects of the particles and physics at higher energy and smaller distance scales (called ultraviolet physics) are incorprated into the coupling constants of the lower energy (called infrared physics) effective field theory, and when an effective field theory has divergences or broken symmetries, changing the scale of the theory to higher energies where new particles become relevant to interactions can remove or soften the divergences and restore the symmetries. An effective field theory that breaks down at some energy scale is then not a problem of the theory but a feature: it just means the effective theory is not the full theory, not containing all particles and their interactions. In the effective field theory framework, all theories in physics are effective theories, except for the fundamental theory at the highest of energy and smallest of distance scales.

String theory is a candidate for the fundamental theory. To produce lower energy effective theories, the extra dimensions of the strings need to be compactified, and the type of compactification made, i.e., the geometry of the extra dimensions that exist in the string theory, determines the type of effective theory that is produced. There are a large number of string vacua (some 10^{500} [1] or more if background fluxes are turned on [2]) that produce similar infrared (IR) physics in the effective field theories that we can access by experiment: the Standard Model. Although the IR physics generated by the string theories is the same, there should be some information about the string theory that remains in the IR theory, thereby allowing to determine the extent of distinguishability between the string theories (that generated the IR theory) from access of the IR physics.

A renormalization group (RG) flow is used to change the scale of a theory and generate effective field theories. In performing a renormalization group flow down from ultraviolet (UV) scales to IR scales, mixing between the UV and IR modes (UV-IR mixing) will occur. This gives rise to the possibility of detecting UV effects–and in particular, string compactification effects–while being able to only access the IR physics of the effective field theory.

Therefore, some questions arise: Is there a way to quantify this UV-IR mixing in order to distinguish between similar IR theories and thereby access information about the UV, and if so, how can this be done? It could also turn out that the measure used to distinguish the UV theories will show that classes of UV theories are indistinguishable: applying this to string theory means that classes of string theories could turn out to be indistinguishable according to this measure, thereby reducing the number of distinct string vacua.

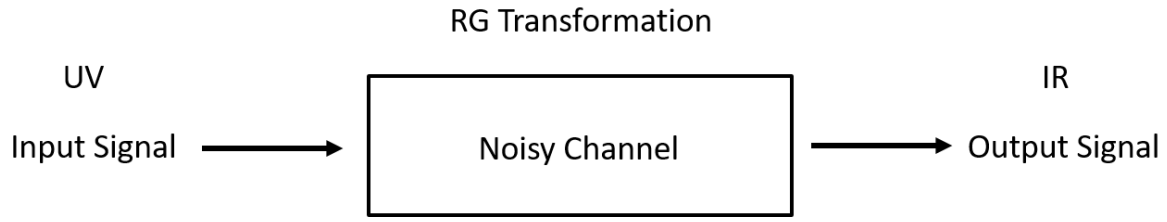

Figure 1.1: A schematic representation of viewing the RG transformation as a noisy communication channel. The variables of the UV theory are the input. The RG transformation is the noisy channel that produces the output variables of the IR theory.

Our answer is that we can think of the UV information accessible at the IR physics as a signal sent from the UV scale to the IR scale. The renormalization group transformation is then viewed as a noisy communication channel that the UV information traverses to the IR scale. We can thus evaluate the mutual information and find the channel capacity of the channel. See Figure 1.1 for a schematic representation of this idea.

It might also be possible to use the mutual information to calculate the information bottleneck [3]. The information bottleneck is an optimized intermediate set of variables $\tilde{X}$ that takes a signal X and compresses it to $\tilde{X}$ then sends it through a communication channel to give an output signal of Y. The intermediate variables $\tilde{X}$ are optimized to be as small as possible while preserving the maximum amount of information about Y in X. In our problem, the idea is to read the UV information (X) from the IR variables (Y) via the intermediate channel of the RG transformation ($\tilde{X}$ is our renormalized variables

before rescaling the lattice or momentum). See Figure 1.2 for a schematic representation of this idea. We explore this idea a little in this thesis to show how the bottleneck concept works in general and would work with our ideas, but we do not fully develop this point, and we ultimately conclude it is easier to just calculate the KL divergence and interpret our results in terms of Figure 1.1.

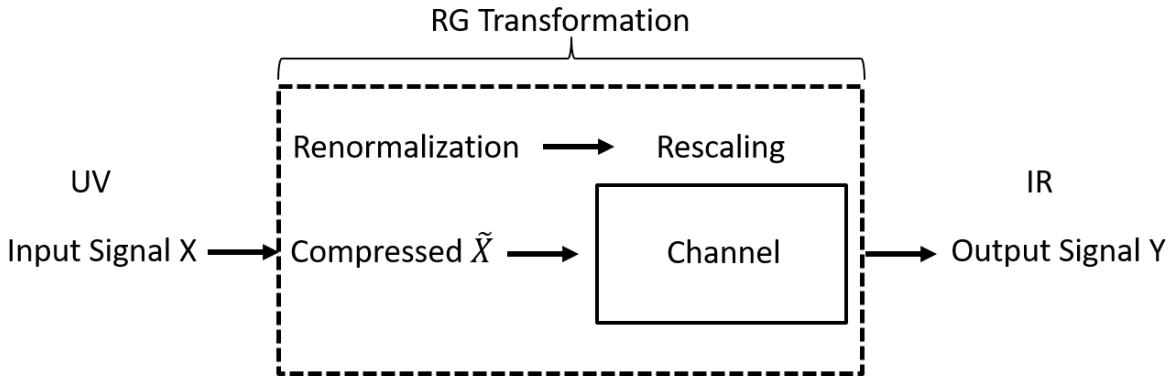

Figure 1.2: A schematic representation of viewing the RG transformation as communication through an information bottleneck. The variables of the UV theory are the input. The renormalization/coarse-graining step of the RG transformation performs the compression to $\tilde{X}$. The re-scaling step sends the compressed variables to the output IR theory.

In short, we can think of the renormalization group flow in terms of information theoretic quantities and thereby provide a direct, information theoretic measure of the UV information accessible at the IR scale. The renormalization group flow can also be thought of in terms of the AdS/CFT correspondence, where the energy scale associated with renormalization goes in the direction of the bulk AdS space as in, e.g., [4–7], so we are also providing an information theoretic interpretation of the AdS/CFT view of the renormalization group flow.

Thinking along these terms, we will make use of the Kullback-Leibler (KL) divergence as in [8, 9] to measure the information difference. For normalized probability distributions p and q, it is,

$$D(p||q) = \int \mathrm{D}\mu \; p \log \frac{p}{q}, \tag{1.1}$$

where the integration measure goes over all the space on which the probability distributions are defined. We will interpret p and q as normalized Boltzmann distributions for two different statistical (i.e., Euclidean) quantum field theories.

While progress has been made in calculating the UV information in the IR physics by using the geometric entanglement entropy [10–14], using the KL divergence has some advantages over the geometric entanglement entropy [15, 16]. The geometric entanglement entropy depends on the geometry of how the subspaces of the Hilbert space are formed: the KL divergence does not depend on the geometry. The geometric entanglement entropy also has difficulties being defined for gauge theories and dealing with extended objects like Wilson loops [17], while there is a possibility of using the KL divergence with gauge theories since all one needs is a well-defined probability distribution produced by the gauge theory's action. Furthermore, the interpretation of the renormalization group flow as a communication channel, thereby allowing for interpretation of the whole problem in information theoretic terms, has not yet been considered in uses of the geometric entanglement entropy.

There are other measures of information loss under the RG flow via the A and C theorems. In 2D, there is a number C that decreases monotonically under the RG flow and is equal to the central charge at fixed points of the RG flow [18]. Similarly, for other even dimensions, there is a number A [19–21]. For odd dimensions, sphere partition functions provide this monotonically decreasing quantity [22, 23]. Because these numbers decrease monotonically under the RG flow, they provide some measure of the degrees of freedom that are lost under the RG flow and thereby some measure of the information loss. However, there are two difficulties with using these quantities to quantify information loss. The first is that these theorems are only proven in 2D, 3D, and 4D: the rest is yet to be proven still. The second is that these numbers do not directly quantify the information loss: the numbers decrease with information loss, but unlike the KL divergence, they do not directly measure the information loss in terms of, e.g., bits.

We will proceed as follows. We first discuss effective field theories in Section 2. In that section, we show many examples of effective field theories, their problems, and how new UV physics continues to fix effective field theories as the scale goes to higher and higher energies. We start with the Standard Model, then discuss grand unified theories and supersymmetry, then finally string theory compactifications. At the end of the section, we discuss warped large extra dimensions, since it ties in nicely with an AdS/CFT interpretation of the renormalization group flow.

Having explained effective field theories and shown both their use and the kinds of UV physics that may be encoded in theories at the scale of the standard model, we then discuss the renormalization group flow in Section 3. We discuss the main RG procedures that we use and how to perform them. We show

the connection between the RG flow and AdS/CFT in Section 3.5 and close the section by discussing the RG procedure in general to prepare for discussing the information theoretic connection.

Next, in Section 4 we discuss the concept of a communication channel and the tools used for making calculations of quantities related to a communication channel. We first discuss the KL divergence in Section 4.1 and then subsequently discuss its use in calculating information-theoretic quantities, such as the mutual information and the channel capacity. After discussing the channel capacity, we make the connection to an information theoretic interpretation of the RG flow in Section 4.3. We then discuss the information bottleneck as another possible information theoretic quantity of interest to our problem and finish the section by discussing how our procedure with the KL divergence connects to the quantum theory.

In the discussion of the communication channel and the KL divergence, it is noted in the beginning of Section 4.2 that a difficulty in using the KL divergence to measure UV-IR mixing is that it naively requires the use of marginalizing distributions over IR modes, i.e., integrating out IR modes. Integrating out IR modes leads to non-local interactions in the effective field theory. Another difficulty that is noted in Section 4.1 is that the KL divergence is quadratic to leading order, which leads to integrated two-point functions and contact terms.

We therefore study these difficulties next and work to find a way around them. We do so by studying lattice theories first and then continuum theories. We consider lattice theories in Section 5, studying a particular RG procedure for general lattice field theories: the decimation procedure. The lattice theories are in position space and so do not have the conceptual difficulty of integrating out IR modes, and the lattice theories come with a natural regulator (the lattice spacing), so there is no need for concern about contact terms. We find that there are two different ways to perform the decimation procedure so as to calculate the mutual information between a UV theory and an IR theory resulting from the decimation procedure. At the end of the section, we show an interpretation of the mutual information in terms of thermodynamic quantities.

We then study continuum theories in Section 6, where we start with a theory, perform a renormalization group flow to a lower energy theory, and calculate the KL divergence between the lower energy theory and the original theory. As expected, contact terms are found, making the KL divergence generally dependent on the regularization scheme to regulate the UV divergence of the contact terms. Having calculated the continuum version of the earlier lattice calculations, we find the continuum limit of the

lattice theories and find that both decimation procedures produce the same result in the continuum limit. Finally, in Section 6.3, we find a way to define a UV completion for continuum theories when using the KL divergence with continuum theories. The UV completion removes the contact terms of our earlier result. We then find the energy scale dependence of the information of the UV theory in the IR theory.

Having discussed and done much theoretical formulation and general calculations, we then make concrete calculations for a variety of example models, starting with lattice models and then continuum models. As a warmup example, we first find the KL divergence between an Ising model and an Ising model with a perturbation in Section 7. We consider the general calculation for both 1D and 2D Ising models and then we make a specific calculation with the 1D model, where we also find the thermodynamic large N limit for the 1D model. We then make calculations for example decimated theories in Section 8. We study both 1D and 2D Ising models, along with a $(1+\varepsilon)$D model. We make progress towards an example calculation of the channel capacity in Section 9 by studying a 1D Ising model on a tree, which also provides an example of a more general block renormalization group procedure.

We then make progress towards calculations with a few example continuum theories in Section 10. Our examples are a $\phi\chi^3$ theory, a fermionic CFT with a $T\bar{T}$ deformation, and a Kaluza-Klein scalar field theory. In Section 11, we summarize our results and suggest future directions of research in our conclusion, and we give the Mathematica code that we used for our main calculations in the Appendices.

CHAPTER 2

Effective Field Theories in the Standard Model and Beyond

The idea of low energy physics being sensitive to the physics at higher energies is encapsulated in the framework and machinery of effective quantum field theories. Start with a quantum field theory defined by an action S_Λ with a cutoff Λ. An effective quantum field theory is produced by integrating out higher energy modes down to some lower energy scale Λ' to produce an action $S_{\Lambda'}$ [24]. The mathematics of this will be discussed in more detail with the discussion of the Renormalization Group in Section 3. See also [25] for another mathematical treatment of effective field theories.

The effective field theory is understood as capturing the physics that exists at the energy scale Λ', which provides a regulator for the theory [24]. The end result of the process of converting S_Λ to $S_{\Lambda'}$ is that higher energy variables are eliminated (they are integrated out), which then allows for describing the theory in terms of variables that exist at the lower energy scale. The effects of the higher energy variables are encoded in the coupling constants of the lower energy variables.

Thinking of effective actions in terms of Feynman graphs, consider a diagram at the original scale Λ that includes interactions of the higher energy particles with the lower energy particles. The effective action at scale Λ' will have intersecting lines in place of the loops and exchanges of the higher energy particles. Reversing the process by changing the scale back to the original Λ results in replacing the point where the lines intersect with loops and exchanges of the higher energy particles [25].

This Feynman graph picture of effective field theories helps conceptualize what an effective action is describing in terms of particles. There are high energy particles that cannot be detected (or have highly suppressed interactions) at lower energies. Thus, the exchange and interactions of these higher energy particles cannot be detected when computing the interactions of the lower energy particles. Instead, there is an effective interaction of the lower energy particles that remains.

The process of generating an effective field theory can be repeated to continually produce effective actions for lower and lower scales, producing correspondingly descriptions of physics at lower and lower energies. As the energy scale is lowered, the higher energy interactions are "blurred" out to produce their

collective effect at each energy scale. Taking this perspective of effective field theories, we see that all non-fundamental descriptions of physics are effective theories, and therefore, all physics is dependent on the energy scale of the problem. Furthermore, because physics can be described in terms of quantities that are detectable at a particular energy scale, one does not need the full, fundamental description of physics in order to make use of physics at a lower energy scale.

The dependence of physics on energy scale is in fact the usual procedure in physics. For the purposes of describing books sliding on ramps, one uses books, centers of mass, normal forces, and friction; one does not describe the purpose in terms of the molecular interactions that make up the friction, normal force, books, and ramps. However, if one wished to describe molecular interactions, one must use variables relevant to the molecular scale.

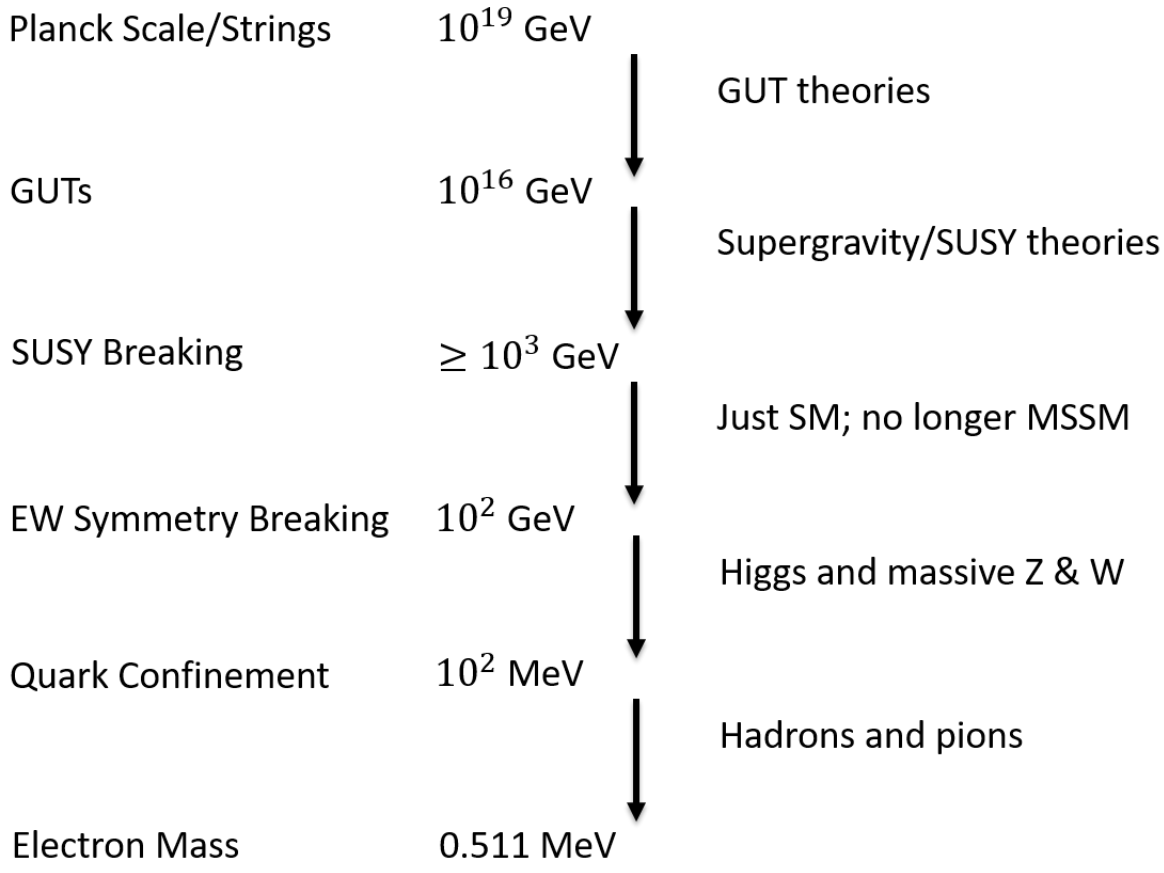

Figure 2.1: A representation of effective theories of particle physics. We start with the Planck scale and go all the way down to the scale of the electron mass. At each scale, a new effective theory with its own fields becomes the useful description, especially as symmetries are broken with the lowering of the energy scale.

As an example of an effective quantum field theory, consider the strong interaction [26, 27]. At a high energy scale (the QCD scale), we can describe our theory in terms of quarks and exchanges of gluons. For lower energies, spontaneous chiral symmetry breaking occurs, producing Goldstone modes. In this case, the Goldstone modes are pions. If we are not at a high enough energy scale where we need to describe physical objects and interactions in terms of quarks and gluons, the pion Lagrangian is easier to use for describing the lower energy effective interactions, thereby showing the utility of effective field theories. See Figure 2.1 for a graphical diagram of effective field theories as the energy scale is lowered; we will discuss a number of these effective theories in the sections that follow.

Effective field theories are not only useful for calculations, but they are also useful for removing UV divergences in a theory and for naturally having new particles at higher energies than the scale of a particular effective theory. Furthermore, because effective field theories encode the UV physics in the coupling constants of the lower energy theory, effective field theories provide a useful way to think about the problem of quantifying information about UV physics that is present in the IR physics. Furthermore, in the cases where a variety of different theories at a UV scale produce similar effective field theories at the IR scale, quantifying the information of the UV physics still present in the IR physics can be used to distinguish the lower energy effective theories from each other, providing a measure of the proximity of these quantum field theories to each other.

We shall now give an overview of a variety of quantum field theory models in the remainder of this section, showing how the idea of a fuller theory existing at a higher energy scale continues to solve the problems of the corresponding lower energy theories. In particular, we are especially interested in the Standard Model (SM) and its problems, since whatever UV physics exists needs to reproduce the SM at lower energies, so we will begin by discussing the Standard Model and its shortcomings.

2.1 Standard Model–Overview

As is well known, the Standard Model (SM) is a quantum field theory that accurately describes most of the known particles and forces. It covers a wide variety of scales from electrons at 0.511 MeV to the Higgs at 125 GeV to the top quark at 172 GeV. We will write out the Lagrangian density for the SM following [28] with some insights from the introductory chapter of [29]. The SM Lagrangian density

before electroweak symmetry breaking is,

$$\begin{aligned} L &= L_{kinetic} + L_h + L_{yuk} + L_{\nu R} \\ &= (L_{gauge} + L_{quark} + L_{lept}) + L_h + L_{yuk} + L_{\nu R}, \end{aligned} \tag{2.1}$$

where we have broken the Lagrangian down into kinetic terms, a piece involving both the kinetic and mass terms of the Higgs, and a piece involving kinetic and mass terms for right-handed neutrinos, supposing they exist. We will use the following conventions. We use the mostly minus metric $(+---)$, and we have assumed the various gauge fields transform as $\exp(-i\alpha(x)^a t^a)\phi(x)$, thereby giving a covariant derivative of the form $D_\mu = \partial_\mu + igA^a_\mu t^a$, where the t^a is the generator of the corresponding group symmetry. Notice that the convention for the field transformation is opposite that of Peskin and Schroeder (Eqs 15.21 and 15.42 in Peskin and Schroeder) [27], which results in a different convention for the covariant derivative (Eq 15.45 in Peskin and Schroeder). To switch between the conventions, plug in $-g$ wherever the couplings appear in the equations. However, as in Peskin and Schroeder, we will take the electron coupling to be $g = -|e|$.

Because of the above gauge field conventions, we take the gauge fields to infinitesimally transform with a minus sign as $A^a_\mu t^a \rightarrow A^a_\mu t^a - \frac{1}{g}\partial_\mu \alpha^a t^a + i[\alpha^a t^a, A^b_\mu t^b]$, and we do the same with the finite transformation. The field strengths are then defined as $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_3 f_{abc} G^b_\mu G^c_\nu$. We take the commutator generators to go as $[t^a, t^b] = i f_{abc} t^c$.

Writing the Standard Model Lagrangian piece by piece we have,

$$L_{gauge} = -\frac{1}{4} G^a_{\mu\nu} G^{\mu\nu a} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W^A_{\mu\nu} W^{\mu\nu A}, \tag{2.2}$$

where G is the gluon field strength, B and W the field strengths for the B and W bosons. We have written the Lagrangian in terms of the components of the field strengths.

$$\begin{aligned} L_{quark} = &\sum_{N=1}^{3} i\bar{q}_L^{i,I,N} \gamma^\mu (\delta_{ij}\delta_{IJ}\partial_\mu + ig_3 G^a_\mu T^a_{ij}\delta_{IJ} + ig_2 W^A_\mu t^A_{IJ}\delta_{ij} + ig_1 B_\mu \frac{1}{6}\delta_{ij}\delta_{IJ}) q_L^{j,J,N} \\ &+ i\bar{u}_R^{i,N} \gamma^\mu (\delta_{ij}\partial_\mu + ig_3 G^a_\mu T^a_{ij} + ig_1 B_\mu \frac{2}{3}\delta_{ij}) u_R^{j,N} \\ &+ i\bar{d}_R^{i,N} \gamma^\mu (\delta_{ij}\partial_\mu + ig_3 G^a_\mu T^a_{ij} - ig_1 B_\mu \frac{1}{3}\delta_{ij}) d_R^{j,N}, \end{aligned} \tag{2.3}$$

where the g_3, g_2, g_1 are the gauge coupling constants of $SU(3)_C \times SU(2)_W \times U(1)_Y$, respectively. The color indices are $i = 1, 2, 3$, and the flavor indices are $I = 1, 2$. The $N = 1,2,3$ is the generation index. The summation over the color and flavor indices are implied by the summation convention. The q are the quark fields, and the u and d are the up and down quark fields. We are clearly using Dirac fermions (because of the presence of γ^μ) and so $\bar{q} = q^\dagger \gamma^0$, but we have broken the Dirac fermions into right and left components, e.g., $q = q_L + q_R$.

The covariant derivatives have been written out explicitly. The $t^A_{ij} = \frac{1}{2}\sigma^A_{ij}$, where the σ^A_{ij} are the Pauli matrices. The $T^a_{ij} = \frac{1}{2}\lambda^a_{ij}$, where the λ^a_{ij} are the Gell-Mann matrices. The weak hypercharge Y_W has already been evaluated in the Lagrangian, and the convention we use is $Q = T_3 + Y_W$ as in Peskin and Schroeder.

$$\begin{aligned} L_{lept} = & i\bar{\psi}^{I,N}_L \gamma^\mu (\delta_{IJ}\partial_\mu + ig_2 W^A_\mu T^A_{IJ} - ig_1 B_\mu \frac{1}{2}\delta_{IJ})\psi^{J,N}_L \\ & + i\bar{e}^N_R \gamma^\mu (\partial_\mu - ig_1 B_\mu) e^N_R, \end{aligned} \tag{2.4}$$

where the ψ are the lepton fields, and the ν and e are the neutrino and electron fields. The sum over the N generations is implied and will be assumed in the rest of the Lagrangian pieces.

$$L_h = \phi^\dagger (\partial^\mu - ig_2 W^{\mu A} t^A - ig_1 B^\mu \frac{1}{2})(\partial_\mu + ig_2 W^A_\mu t^A + ig_1 B_\mu \frac{1}{2})\phi - \lambda(\phi^\dagger \phi - \frac{\nu^2}{2})^2, \tag{2.5}$$

where ϕ is the complex Higgs doublet $\phi^T = (1/\sqrt{2})(\phi_1 + i\phi_2, \phi_0 + i\phi_3)$. Notice that this is both the kinetic term and mass term for the Higgs, whereas the previous Lagrangians only included the kinetic terms for the various particles. We take $\nu^2 = -\mu^2/\lambda$ and $\mu^2 < 0$ to achieve symmetry breaking. We then have the mass of the Higgs $m_h^2 = -2\mu^2 = 2\lambda\nu^2$. Upon spontaneous symmetry breaking from the Higgs, these interactions generate masses for the W and B bosons.

$$L_{yuk} = -y^e_{IJ}\bar{\psi}^{I,N}_L e^{J,N}_R \phi - y^d_{IJ}\delta_{ij}\bar{q}^{i,I,N}_L d^{j,N}_R \phi - y^u_{IJ}\delta_{ij}\bar{q}^{i,I,N}_L u^{j,N}_R i\sigma^2 \phi^* + h.c., \tag{2.6}$$

where the y^e, y^d, and y^u are the Yukawa coupling constants for the fields. Upon spontaneous symmetry breaking from the Higgs, these interactions generate mass terms for the quarks and leptons (except for the neutrino). The h.c. indicates to add the Hermitian conjugate.

Neutrinos are known to have mass and so should have a mass term. Supposing that the neutrinos are Dirac particles and that right handed neutrinos exist, the following piece must be added to the SM Lagrangian,

$$L_{\nu R} = i\bar{\nu}_R^N \gamma^\mu (\partial_\mu) \nu_R^N - y_{IJ}^\nu \bar{\psi}_L^{I,N} \nu_R^{J,N} i\sigma^2 \phi^* + h.c., \tag{2.7}$$

where we have both a dynamical term and a Yukawa interaction term. In this equation, the h.c. should only be that of the Yukawa interaction term.

Although very successful experimentally, the Standard Model has a number of shortcomings [29–31]. There is the hierarchy problem [30, 32–36]: the electro-weak force is stronger than gravity by a factor of 10^{24} without an explanation for this large difference within the Standard Model. There is the problem that the Standard Model only contains massless neutrinos, despite them having a small mass. The Standard Model does not have a dark matter sector, which is the most popular solution to explain such things as galaxy rotation curves and star formation. The Standard Model also cannot explain the observed baryon asymmetry: a vastly unequal amount of matter and anti-matter [37].

Another fundamental problem is that the Standard Model does not include a renormalizable theory of quantum gravity. There is also the related problem of requiring a regulator to be well-defined. The Standard Model cannot be put on a lattice, so there is no UV complete non-perturbative definition of the Standard Model. We see then that the Standard Model cannot be the full story for physics: there must be something beyond the Standard Model that explains and solves these difficulties.

Colliders such as the LHC can put bounds on new physics by studying Standard Model interactions. Even at currently accessible energy scales, there is in fact a tower of higher dimensional operators in the Standard Model made of Standard Model fields that go beyond the usual Standard Model interactions. The new physics can be parameterized in terms of the coefficients of these higher dimensional operators, and colliders seek to find and put bounds on the coefficients of these operators. Putting bounds on these operators then provides bounds on the kinds of new physics that would generate the coefficients of those operators.

The problem of not having a regulator–and new physics beyond the SM being needed–can be viewed as a feature of the SM not a problem: the Standard Model should be viewed as an effective field theory. It is then expected that the Standard Model will work up to some energy scale at which it breaks down (often taken to be the Planck scale) and new physics and fields appear. The energy scale at which the

Standard Model breaks down is therefore just the regulator of an effective field theory. This point of view also gives an understanding of why the Standard Model breaks down: beyond some energy scale, the lower energy fields, such as quarks and electrons, no longer provide an accurate description of the physics. The lower energy fields are a composite of higher energy fields and their interactions. This idea of the Standard Model as an effective field theory then points in the direction of there being new physics beyond the scale of the Standard Model.

A question arises as to whether the UV physics of the higher energy theory can fix some of the Standard Model's shortcomings. As a first step beyond the SM, there is the idea of Grand Unified Theories (GUT), in which the coupling constants of the fundamental forces approximately unify around 10^{15} GeV and a mechanism is provided for a small neutrino mass, among other nice features. A number of GUT candidates and their properties are discussed in [29].

Two popular GUT candidates are $SO(10)$ [38] and $SU(5)$ [39]. E_6 is another possibility [40]. Below the GUT scale of 10^{15} GeV, the larger symmetry of the GUT group is broken and breaks so as to produce the SM gauge group, along with some extra matter. There can be a number of different intermediate symmetry breakings before reaching the SM. Some examples of how these groups break to the SM are as follows [29],

$$\begin{aligned} E_6 &\to SO(10) \times U(1) \\ SO(10) &\to SU(5) \times U(1) \to G_{SM} \\ SU(5) &\to G_{SM}, \end{aligned} \tag{2.8}$$

where G_{SM} is the SM gauge group.

GUTs still leave many problems of the SM unsolved, such as the hierarchy problem and that of not including gravity. However, they provide a first look into possible UV physics solving problems with a lower energy effective field theory, and GUT groups are a motivation for constructing supersymmetry and string compactification models, as shall be seen.

Of the remaining problems of the SM not solved by GUT theories, the hierarchy problem has been a big motivator for finding new physics. The hierarchy problem can be reduced to a problem of the Higgs mass: it is the Higgs field that gives mass to particles at the electroweak scale and so the Higgs mass is responsible for the large difference between gravity and the electroweak scale. Quantum corrections to

the Higgs field mass cause the Higgs mass to have a quadratic UV divergence that should push the Higgs mass to the Planck scale: but the Higgs mass is small, giving a small electroweak scale.

New UV fields can remove a UV divergence and also provide more matter content (possibly providing dark matter candidates). Before discussing proposed UV physics beyond the SM, we will discuss the hierarchy problem to show the manner in which the UV divergence occurs in calculating the Higgs mass and why UV fields at a higher energy scale could be a solution to the problem.

2.2 Excursis: Hierarchy Problem

The hierarchy problem is not a problem particular to the Standard Model Higgs: it occurs with any fundamental scalar field with a quartic interaction, and so in particular, it occurs with the the SM Higgs field. We can compute the quantum 1-loop correction to a fundamental scalar field mass as follows. The exact two-point function for the scalar field (in Euclidean space) is [27],

$$\begin{aligned}\int d^4x \exp(ikx)\langle\phi(x)\phi(0)\rangle &= \frac{1}{k^2+m^2} + \frac{1}{k^2+m^2}\Pi(k^2)\frac{1}{k^2+m^2} \\ &+ \frac{1}{k^2+m^2}\Pi(k^2)\frac{1}{k^2+m^2}\Pi(k^2)\frac{1}{k^2+m^2} + ... \\ &= \frac{1}{k^2+m^2-\Pi(k^2)},\end{aligned} \tag{2.9}$$

where we have summed the geometric series and where $\Pi(k^2)$ (the self-energy) is the sum of all one-particle irreducible graphs. As can be seen, the self-energy gives the mass correction to the field. Working in perturbation theory, we can compute the self-energy to 1-loop and thereby compute the 1-loop correction to the scalar field mass. Using a cutoff Λ_0 to regulate the integral, the result is [27],

$$\begin{aligned}\Pi(k^2) &= -\lambda\int_{|p|\leq\Lambda_0}\frac{d^4p}{(2\pi)^4}\frac{1}{p^2+m^2} \\ &= \frac{-\lambda}{16\pi^2}\left(\Lambda_0^2 - m^2\ln\left(1+\frac{\Lambda_0^2}{m^2}\right)\right).\end{aligned} \tag{2.10}$$

We see that the loop correction grows quadratically with the cutoff. This means that the physical mass is [27],

$$m_{phys}^2 = m^2 - \Lambda_0^2\frac{\lambda}{16\pi^2} - O(\ln(\Lambda_0^2)), \tag{2.11}$$

so we see that the physical mass depends quadratically on the cutoff.

Not only the Higgs self-interaction, but the Higgs interaction with the Standard Model particles also give contributions to the Higgs mass that grow quadratically with the cutoff from the 1-loop diagrams [30]. Now, the physical mass of the Higgs is 125 GeV, and the cutoff is the scale of new physics. For a fundamental scalar field, the scale of new physics should go all the way to the Planck scale of 10^{19} GeV. Hence, the bare mass m^2 will need to undergo a fine cancellation with the quantum corrections in order to produce the relatively small Higgs mass.

We used a cutoff to compute the correction to the Higgs mass, but we could have used dimensional regularization. Using dimensional regularization we get [27],

$$\Pi(k^2) = \frac{-\lambda m^2}{16\pi^2}\left(\frac{2}{\varepsilon} - \gamma_E + 1 - \log\left(\frac{m^2}{4\pi}\right)\right) \tag{2.12}$$

No quadratic divergence appears. Instead, a divergence proportional to $1/\varepsilon$ appears, as it does for anything done with dimensional regularization. The ε is not a physical parameter and represents no scale, so it would appear there is no hierarchy problem. However, the quadratic divergence is just hidden in the dimensional regularization scheme, which treats all divergences the same. To see that the hierarchy problem still appears, consider a coupling of a light scalar field to a heavy massive scalar particle Φ with mass M at a high energy scale where new physics should occur: $\lambda\phi^2\Phi^2$. The one-loop Φ correction to the light ϕ mass is [27, 30],

$$\Pi(k^2) = \frac{-\lambda M^2}{16\pi^2}\left(\frac{2}{\varepsilon} - \gamma_E + 1 - \log\left(\frac{m^2}{4\pi}\right)\right), \tag{2.13}$$

with similar results for coupling to fermionic fields. We see then that we still have a quadratic dependence of the physical mass on the heavy particle with mass M. A finely tuned cancellation still needs to occur to produce the small mass of the light scalar field (and therefore also needed to produce the small Higgs mass).

Notice that the hierarchy problem occurs because (1) the scalar Higgs field is taken to be fundamental, and (2) there is no natural reason (like a symmetry or preferred regulator) to cancel the quadratic divergences. There have been a number of solutions proposed for the hierarchy problem that seek to address these causes. Technicolor takes the Higgs to be a composite particle instead of a fundamental particle, thereby allowing the Higgs field to have a naturally lower energy scale than the Planck scale

[33, 34, 41–43]. Another proposal is to make use of extra dimensions, see [44–46] and [47–49]. See [50] for another take on extra dimensions to solve the hierarchy problem that uses the extra dimension as a mere tool, leaving a 4d physical space-time.

Still another proposal is supersymmetry, where each bosonic field has a fermionic pair. This new symmetry can be used to cancel the quadratic divergences and give the Higgs field a naturally lower mass. We will now discuss supersymmetry as an example of UV physics that addresses the SM's shortcomings and should have information encoded about itself in the SM's couplings (as all effective theories have about their higher energy counterparts).

2.3 Supersymmetry

As just noted, supersymmetry (SUSY) is one proposal to address the problems of the SM, especially the hierarchy problem (as discussed above). This symmetry relates bosons to fermions as [29, 30],

$$\begin{aligned} Q|fermion\rangle &= |boson\rangle \\ Q|boson\rangle &= |fermion\rangle, \end{aligned} \tag{2.14}$$

where Q is the generator of the SUSY algebra and the details depend on the specifics of the SUSY theory. SUSY is classified by the number of generators: $\mathcal{N} = 1$ SUSY corresponds to having one generator, Q, etc.

The generator Q has the following anti-commutation relations (extensions of the SUSY algebra for multiple generators use a similar anti-commutation relation),

$$\begin{aligned} \{Q_\alpha, \bar{Q}_{\dot{\beta}}\} &= 2\sigma^\mu_{\alpha\dot{\beta}} P_\mu \\ \{Q_\alpha, Q_\beta\} &= 0, \end{aligned} \tag{2.15}$$

where the Q are anticommuting Weyl spinors and the P_μ is the spacetime momentum; the dotted indices indicate that the object transforms in the conjugate (right-handed) representation. The commutation relation between Q and P is zero, and the commutation relation between P's is also zero. Because the generators are related to the spacetime momentum, we thus see that SUSY is a spacetime symmetry, not merely an internal symmetry like isospin.

There are a variety of proposed SUSY extensions of the SM, with the minimal supersymmetric standard model (MSSM) among them. SUSY extensions of the SM are discussed in [29, 30] with some early SUSY extensions of the SM found in [51–54]. Early proposed SUSY scales can be found in [55].

Usually, we are interested in a minimal supersymmetric extension of the Standard Model, and we thereby choose $\mathcal{N} = 1$ SUSY [29]. Because SUSY has not been observed at the scale of the SM, the $\mathcal{N} = 1$ SUSY must also be broken at some energy scale above the Standard Model. In order for SUSY to result in a naturally small Higgs mass, the SUSY scale must be set around 1 TeV. However, SUSY still has not been observed despite probing higher energies, and there is increasingly less room for finding 1 TeV SUSY [56]. When we discuss string compactifications, it should be noted that the goal of compactification historically has been to preserve $\mathcal{N} = 1$ SUSY after the compactification is performed and to generate a GUT group that breaks to the SM, but because of the non-observence of 1 TeV SUSY, string compactification models will instead skip preserving SUSY at all and try to generate the Standard Model group directly from the compactification.

As noted earlier, SUSY addresses the hierarchy problem [55, 57]. This new proposed symmetry makes the Higgs mass naturally low because the fermionic loops (which contribute with opposite sign) cancel the paired bosonic loops to eliminate the quadratic divergences [29, 30, 58–61]. A specific example of this calculation to one loop can be seen in Ibañez with the Wess-Zumino model (a simple SUSY theory) [29].

SUSY also provides new (as yet) unobserved particles and thereby dark matter candidates [62]. Furthermore, it allows for a more exact gauge coupling unification by changing the running of the couplings of the Standard Model forces. The gauge couplings can now unify at 10^{16} GeV, instead of only approximately unifying [29, 39, 54, 55, 63–68].

When turned into a local symmetry, SUSY theories contain a graviton, thereby addressing the lack of gravity in the SM [69–73]. These supergravity theories can exist in a variety of dimensions, up to 11d [74] with $\mathcal{N} = 8$ supercharges [75], where the extra dimensions are compactified. $N = 8$ 4d supergravity [76, 77] is one of the SUGRAs of historical interest and can be obtained from compactifying 11d SUGRA. The 11d SUGRA ($\mathcal{N} = 1$) is also of interest because it is a unique theory by having the maximal number of dimensions without having fields of spin higher than 2 and having a single graviton [78] (unless one takes a 12d theory with two time-like dimensions [79]).

We thereby see how the UV physics of SUSY and SUGRA can address a number of the problems of the low energy SM. Because these theories produce the same low energy SM, as previously stated, these too should have information encoded about themselves in the lower energy SM theory. However, supersymmetry is not the fundamental theory. None of the SUGRA theories are known to be UV finite non-perturbatively, and they do not contain chiral fields, even when compactifying extra dimensions [29]. Hence, if they have any connection to the SM, these supergravity theories must also be low energy limits of a more fundamental theory that will allow for reduction to the SM.

2.4 String Compactification

String theory is one candidate for the fundamental physics at the Planck scale from which all the effective theories of SUGRA and the SM are generated. The quantum strings need to be supersymmetric and in 10D in order to produce fermions at low energies and in order for the string theory to be mathematically consistent. The 10d SUGRA models are still useful in the string framework because they provide the low energy effective field theory (i.e., massless tree level) to the 10d superstring theories.

String theory solves a number of problems that are in the SM and remain in the SUGRA theories. String theory solves the problem of quantizing gravity by always including a closed string in its spectrum (which corresponds to the spin 2 graviton). Unlike the SUGRA models, string theory is UV finite by introducing the string length as a smallest length scale regulator. By compactifying the extra dimensions and taking a low energy limit, string theory can produce quantum fields that resemble existing SM particles, including chiral fields. We will discuss how string theory compactifications can produce the SM gauge group in this chapter, following the discussions in [29] throughout. From our discussion, it will also become apparent that there are many different compactification schemes; counting the distinct compactifications produces the 10^{500} (or larger) number for the number of distinct string vacua (with some 10^{15} different F-theory compactifications that have been found that produce the chiral SM gauge group [80]).

String compactification takes the extra dimensions and rolls them up into a circle or puts them on another compact space. This is often done by factorizing the 10D space into $M_4 \times X_6$, where X_6 is a compact space and M_4 is the usual 4D Minkowski spacetime. Because the extra dimensions are curled up or compact, they are not observable at the energy levels experiments have observed.

As stated previously, when performing these compactifications, it is usually desirable to preserve SUSY (although sometimes desirable to reduce the amount of supersymmetry upon compactification) and then break SUSY at a lower scale after compactification. To preserve SUSY, Calabi-Yau manifolds are often used for the compactification space [81–84]. Orbifolds are also used as compactification spaces [85–88], getting rid of more supersymmetry and not requiring one to use a supergravity approximation to the string theory.

For making string compactifications in general, there are five consistent supersymmetric string theories. Type I, Type IIA, Type IIB, E_8 x E_8 heterotic, and $SO(32)$ heterotic. They differ in the gauge symmetries that they have, the charges and branes that exist in them, and whether they include open strings on the perturbative level. The five superstring theories are related by dualities, so they are not entirely different theories. They are in fact believed to be low energy limits of a more fundamental M-theory: a theory of 2d and 5d membranes. 11d SUGRA is then believed to be a low energy approximation (massless, tree-level) of M-theory, as the 10d SUGRA is for the 10d superstring theories.

M theory is defined as the strong coupling limit of Type IIA theory; in that limit, the Type IIA theory gains an 11th dimension [89]. F theory [90–93] has a similar relation to Type IIB theory, which can be seen as follows. F theory can be defined as M-theory compactified on a 2-torus in the limit of vanishing area. Compactifying M theory on one of the circles of the vanishing 2-torus produces Type IIA theory with a dimension compactified on a vanishing circle; T-duality then produces Type IIB theory compactified on a circle with infinite radius, i.e., the full Type IIB theory [94, 95]. The Type IIB theory comes with a varying axio-dilaton, and F theory then geometrically parameterizes the varying axio-dilaton by having two extra auxiliary dimensions via elliptic fibration over the Type IIB spacetime.

Heterotic string compactifications are one way to produce chiral fields [96, 97]. Other ways to obtain chiral fields include forming stacks of intersecting D6-branes in Type IIA theory and stacks of intersecting D7-branes in Type IIB theory [98–103]. (To get chiral matter in 4d from the intersecting branes in Type IIB theory, fluxes need to be turned on in the compactified space.) The intersecting D-brane models are of interest because M theory and F theory can also produce chiral matter via compactifications [104–107] and do so in a similar manner to the compactifications involving intersecting D-branes.

To show how M theory and F theory compactifications produce chiral matter, we will review how intersecting D-branes produce chiral matter and then lift the intersecting D-brane models to the M and F

theory limits [108]. However, first we will show a simple example of compactification that we also will use later: Kaluza-Klein reduction on a circle.

2.5 Excursis: Kaluza-Klein Compactification

For a simple concrete example of how compactification works, consider a Kaluza-Klein (KK) reduction of a 5d theory onto a circle [109, 110]. Start with a free scalar field theory in 5d (For this subsection and the remainder of this chapter, we switch to using the mostly plus metric convention $(-+++)$.),

$$S_{5d} = -\frac{1}{2}\int d^5x \partial_M \phi \partial^M \phi, \tag{2.16}$$

where M, N run from 0 to 4. Now put the 5th dimension (which we will label by y) on a circle S^1 and Fourier expand ϕ to get,

$$\phi = \sum_k \phi_k(x^\mu) \exp(iky/R), \tag{2.17}$$

where R in this equation is the radius of the S^1. Placing this back into the 5d action and integrating out y gives,

$$S_{4d} = -\frac{2\pi R}{2}\int d^4x \partial_\mu \phi_0 \partial^\mu \phi_0 - (2\pi R)\sum_{k=1}^{\infty}\int d^4x \partial_\mu \phi_k \partial^\mu \phi_k^* + \frac{k^2}{R^2}\phi_k \phi_k^*, \tag{2.18}$$

where $\phi_k^* = \phi_{-k}$. We see then that we have a 4d free scalar field theory for the zero modes and a tower of 4d massive scalar fields for the other Fourier modes. The massive modes are suppressed by $1/R$, so at energies that are much smaller than $1/R$, the massive fields cannot be observed, leaving just the massless scalar field from the zero modes. Choosing a smaller circle makes R small and so the non-zero modes become very massive and harder to observe at low energies.

To see how other kinds of particles can be produced from compactification, consider a 5d metric G_{MN}, where M, N again run from 0 to 4. The action is,

$$S_{5d} = \frac{M_5^3}{2}\int d^5x \sqrt{-G} R_{5d}, \tag{2.19}$$

where M_5 is the 5d Planck mass and $G = \det(G_{MN})$. Now put the 5th dimension on a circle again and Fourier expand G to get,

$$G_{MN} = \sum_k G_{MN}^k(x^\mu) \exp(iky/R), \tag{2.20}$$

where R in this equation is again the radius of the S^1. As before, this could be substituted back into the 5d action and integrated over y to get a 4d massless action plus a tower of massive 4d fields from the Fourier modes. The zero modes give us a scalar $G^0_{44} = \sigma$, a vector $G^0_{\mu 4} = A_\mu$, and a 4d graviton $G^0_{\mu\nu} = g_{\mu\nu}$. The zero mode action gives Einstein gravity along with electromagnetism and an extra scalar,

$$S^0_{4d} = M_5^3 \pi R \int d^4x \sqrt{-g} R_{4d}(g) - \frac{1}{6}\partial_\mu\sigma\partial^\mu\sigma - \frac{1}{4\exp(\sigma)}F^2_{\mu\nu}, \tag{2.21}$$

where $F_{\mu\nu}$ is the usual Maxwellian field strength.

A simple KK reduction is not a realistic compactification, but the calculation outlines how particles can be produced with more complicated compactifications and how the tower of unobserved particles are heavy when the compactified dimension is small (thereby making them unobservable at low energies). One ingredient of finding realistic compactifications is finding compactifications that both produce the correct SM gauge fields and chiral matter. Gauge groups in string theory can be produced from stacks of branes at the same location. N branes at the same position produce a $U(N)$ gauge group from open strings beginning and ending at the various branes in the stack. The $U(N)$ gauge group can decompose into the Standard Model gauge groups, along with some extra $U(1)$'s, as follows,

$$U(N) = SU(N) \times U(1). \tag{2.22}$$

The extra $U(1)$ field is usually assumed to be heavy and unobservable at low energies.

As for producing chiral matter in string theory, we will discuss how chiral matter can be generated from M and F theory, after first reviewing how chiral matter appears from intersecting D-brane models as a first step towards understanding the M and F theory compactifications.

2.6 Intersecting Branes and Chiral Matter from M/F theory

As has been stated, to understand how chiral matter can appear in M and F theory, first it is helpful to see how chiral matter is produced from intersecting brane models. For considering M theory compactifications, first consider D6-branes in Type IIA. A stack of $N + M$ parallel D6-branes have a $U(N + M)$ gauge group in the open string sector. Consider now unfolding the D6-branes so as to produce a stack of M D6-branes intersecting with N D6-branes at a point in the 10D space, i.e., rotate the

stack of M D6-branes by an angle θ so that they fill and intersect in the 4D space and thereby intersect at a point in the remaining six dimensions. This produces three intersection angles θ_1, θ_2, and θ_3 in 2-planes of the remaining six dimensions. The gauge group for the system is now $U(N) \times U(M)$.

Because we started with $(N+M)^2 = N^2 + M^2 + 2NM$ degrees of freedom, this same number of degrees of freedom must be found in the system after unfolding. The stack of N carries N^2, due to its $U(N)$ gauge group (from strings going between branes on the stack of N), and similarly the stack of M carries M^2. The remaining $2NM$ degrees of freedom can be found from the decomposition of $U(N+M)$ into $U(N) \times U(M)$: there are two bifundamentals in the adjoint of $U(N) \times U(M)$ that are not in the adjoint of $U(N+M)$. They are the $(N, \bar{M})$ and $(\bar{N}, M)$. Because these are created from strings going between the two stacks, the chiral matter is found at the intersection of the brane stacks so that the string tension is minimized. Hence, the chiral matter is found in the 4D intersection and has massless modes. Intuitively, the chiral matter has appeared because parity was broken in the 6D space: a preferred direction was established with the angles θ_i.

The SM group can be directly generated by having four stacks of intersecting branes. A stack of three branes gives the $SU(3)$, the stack of two gives the $SU(2)$, and the intersections produce chiral matter. Two stacks each with one brane are also used, which intersect with the stack of three and stack of two, in order to produce the $U(1)$ and the $SU(2)$ singlet field. However, note that there are some extra scalar fields that appear from the decomposition of the $U(3)$ and $U(2)$, which (as noted previously) are usually assumed to be too heavy to be observed at low energies.

In the M theory lift of the D6-branes, the D6-branes become geometrized to a purely metric background. The space transverse to the D6-branes asymptotically goes according to a multi-center 4D Taub-Nut geometry,

$$\begin{aligned} ds^2 &= V(x)dx^2 + V(x)^{-1}(dx^{10} + \omega \cdot dx)^2, \\ V(x) &= 1 + \sum_{a=1}^{N} \frac{1}{2|x - x_a|}, \\ \nabla \times \omega &= -\nabla V(x), \end{aligned} \tag{2.23}$$

where $x \in R^3$ parameterizes the 3D space that is transverse to the D6 branes, and ω is the 3D vector potential for N Dirac magnetic monopoles located at x_a. The D6 branes are located at the positions x_a in the transverse space R^3. The x^{10} is a periodic coordinate that we have labeled as x^{10} in anticipation of

its role in M-theory compactifications; it is the 11th dimension of M-theory: the direction along which the M-theory compactification circle S^1 lies.

The total 11d space breaks up as $M_7 \times X_4$, where X_4 has the Taub-Nut geometry (arising from the X_4 needing to preserve half the supersymmetries and therefore having $SU(2)$ holonomy) and is a fibration of S^1 (the M-theory compactification circle) over R^3: $X_4 = R^3 \times S^1$.

As can be seen from (2.23), intersecting branes are located at singularities of the Taub-Nut geometry x_a, which are places where the S^1 fibers shrink to zero size. Stacks of branes enhance the singularities. 2-cycles can be defined between the locations of non-intersecting branes, and the singularities from the non-intersecting branes pinch the 2-cycles. As branes approach each other to overlap, the 2-cycles vanish. M2 branes can be wrapped on these vanishing 2-cycles. The intersection pattern of the wrapped M2 branes then produces the Dynkin diagram of ADE singularities. ADE singularities are those that have an intersection pattern that is in the shape of the Dynkin diagram of the A, D, or E types of simply laced algebras (strictly speaking, the singularities need to be "blown up," and the blow ups of the singularities have the Dynkin diagram intersection pattern). ADE singularities can be described as the singularities of hypersurfaces of C^3, e.g., the A_N singularity is described by $y^2 + x^2 + z^{N+1} = 0$.

To produce $\mathcal{N} = 1$ supersymmetry, a manifold with G2 holonomy is needed. To produce chiral matter, the manifold must be singular. As already noted, in M-theory, the singularities produced are of the ADE type. Although the A and D type can be recognized as resulting from the lift of overlapping D6 branes, the E type is a non-perturbative result.

Instead of considering two stacks of intersecting D6 branes (which are now part of the geometrical background), consider the intersection of two Taub-NUT geometries. The Taub-NUT geometries produce A_{N-1} and A_{M-1} singularities with an enhanced co-dimension 7 A_{N+M-1} singularity at the intersection. The singularities in the geometry can then be viewed as being unfolded in the same way as unfolding the stacks of D6 branes: $A_{N+M-1} \to A_{N-1} \times A_{M-1}$. When doing so, extra chiral degrees of freedom are found at the co-dimension 7 singularity.

This unfolding procedure can then be generalized to other G2 manifolds with other kinds of ADE singularities as an unfolding of $G \to \otimes_r G_r$, where G is the symmetry of the enhanced ADE singularities and G_r are the symmetries of the other singularities. These then produce other gauge groups with chiral matter at the locations of the enhanced co-dimension 7 singularities. An example of an E type singularity

generated this way is [29, 111],

$$\begin{aligned} E_6 &\to SO(10) \times U(1), \\ 27 &\to 16_1 + 10_{-2} + 1_4 \\ 78 &\to 45_0 + 1_0 + (16_{-3} + \bar{16}_3). \end{aligned} \tag{2.24}$$

Some examples of an $SU(5)$ GUT group generated this way are [29, 111],

$$\begin{aligned} SU(6) &\to SU(5) \times U(1), \\ 6 &\to 5_1 + 1_{-5} \\ 20 &\to 10_{-3} + \bar{10}_3 \\ 35 &\to 24_0 + 1_0 + (5_6 + \bar{5}_{-6}). \\ SO(10) &\to SU(5) \times U(1), \\ 10 &\to 5_2 + \bar{5}_{-2} \\ 16 &\to 10_{-1} + 1_{-5} + \bar{5}_3 \\ 45 &\to 24_0 + 1_0 + (10_4 + \bar{10}_{-4}). \end{aligned} \tag{2.25}$$

Chiral matter in F theory appears in a similar manner to how it appears in M theory. First, consider intersecting D7 branes in Type IIB and perform the same unfolding as with the intersecting D6 branes. The chiral matter will appear at the intersection. However, the D7 branes intersect in a 6d space, so the chiral matter is 6d. The 6d chiral matter cannot be dimensionally reduced to produce 4d chiral matter: only vector-like fermion objects appear upon reduction. To produce 4d chiral matter, a flux on the D7 branes must be turned on. A 6d chiral fermion contains both a left and right Weyl spinor. The flux ensures that only the massless zero mode of the left or right Weyl spinor survives the compactification to 4d, hence ensuring chirality in 4d.

In the F theory limit, the location of a 7 brane produces a singularity in the elliptic fiber; an enhanced singularity in the fiber appears for multiple 7 branes at the same location. The deformed fibers generate the Dynkin diagram of ADE symmetry groups. The unfolding perspective can be taken here too: the enhanced singularities of the fiber have extra chiral degrees of freedom when unfolding the singularity. The chiral matter is 6d here too, so a flux must be turned on for the 7 branes to generate 4d chiral matter.

In the internal space of the base, the 7 branes wrap 4-cycles and intersect in 2 cycles. The 7 branes additionally intersect in a 4d Minkowski space. The singularity intersection patterns and different symmetry groups thus depend on the intersection of their 4-cycles in the internal space. Gauge fields then live on the base space, and matter is located at the intersection curves on the base (hence the 2 cycles are called matter curves). A triple intersection produces Yukawa couplings. By this scheme, the SM group can be directly generated [80, 112].

F theory can also generate GUT groups. As one example, F theory can generate $SU(6)$ and $SO(10)$ groups just like M theory, so see again (2.25) for some examples of $SU(6)$ and $SO(10)$ breaking into $SU(5)$.

We have thereby seen how the fundamental theory (M theory) can produce chiral matter and the SM gauge group via compactifications. We have worked our way up to higher and higher energies through various effective theories to finally arrive at the proposed fundamental theory that should have some information encoded about itself in the lower energy theory of the SM. There are a number of different proposed UV physics that could produce the SM, and quantifying their information content in the SM could be used to distinguish between their effective field theories at the SM energy scale.

We will now discuss the Renormalization Group, which provides the mathematical tools for generating effective field theories, after a brief discussion of another possibility for higher energy physics via warped Large Extra Dimensions (LEDs). The possibility of warped LEDs provides another solution to the hierarchy problem, should 1 TeV SUSY not be found (although at this time, it appears the warped LED solution is highly constrained), and it provides a geometrical picture for the AdS/CFT correspondence that we will be using as an additional interpretive framework of our concepts and results.

2.7 Excursis: Warped Large Extra Dimensions

With branes, it is possible to have larger dimensions in a warped space, referred to as "warped Large Extra Dimensions," while remaining within experimental bounds. This scenario of warped LEDs can provide another possible solution to the hierarchy problem [29, 113, 114], as we shall see. However, it should be noted that in recent years, the possibility of large extra dimensions has been highly constrained by gravitational wave experiments [115–117] and by the Large Hadron Collider [118–120]. We will discuss warped LEDs in the context of the Standard Model fields, but it should be noted that warped LEDs can also be combined with GUT theories, e.g., [121].

Take a 5d spacetime with a finite 5th dimension with coordinate $y \in [0, L]$. One way to do this is to take S^1/Z_2 with S^1 radius L/π, so that $y \to -y$. Place a 3 brane at $y = 0$ and another 3 brane at $y = L$. The Standard Model is localized on the brane at L with tension T, and the Planck scale physics is on the $y = 0$ brane with tension $-T$. So we have Lagrangians $L_{brane} = \pm\sqrt{-g}T$. The space between the branes is the "bulk" space that traverses the energy scales between them.

Compactification of any extra dimensions produces a 4d graviton, as before, but the SM fields are localized on a brane at the start [122, 123]: the localization process must then be what produces the chiral fields. There is no tower of states for the localized Standard Model fields, since the compactification was performed in the bulk, so the extra dimensions can be larger and still be consistent with experimental bounds.

The warped metric in this scheme with an exponential warping factor goes as,

$$\begin{aligned} ds^2 &= \exp\left(-\frac{2|y|}{r}\right)\eta_{\mu\nu}dx^\mu dx^\nu + dy^2, \\ \Lambda &= \frac{-24M^3}{r^2}, \\ r &= \frac{24M^3}{T}, \end{aligned} \tag{2.26}$$

where M is the 5d gravity scale, Λ is the 5d bulk cosmological constant. Thinking of the hierarchy problem in terms of the strength of gravity in comparison to the other forces, gravity propagates throughout the full higher dimensional space, so its strength is diluted accross the extra dimensions. However, the strength of the Standard Model interactions is not diluted, since the interactions are localized to the SM brane.

More precisely, we have an exponential shift of the metric within the fifth dimension,

$$g_{\mu\nu}(y = L) = \exp(-2L/r)g_{\mu\nu}(y = 0), \tag{2.27}$$

so mass scales are exponentially suppressed on the SM brane compared to the Planck brane. Hence, the Higgs mass will also be exponentially suppressed as,

$$m^2_{Higgs} \approx \exp(-2L/r)M^2_{Planck}. \tag{2.28}$$

The exponential factor thus suppresses the quantum correction to the bare Higgs mass so as to produce the correct low energy value (with an appropriate positioning of the SM brane and Planck brane in the extra dimension). With $L/r \approx 16$, the Higgs mass is on the order of the TeV scale.

We have discussed various effective field theories and possible higher energy physics that produces the SM and addresses its shortcomings. We will now discuss the Renormalization Group as the mathematical machinery that is used to produce effective field theories.

CHAPTER 3
Renormalization Group–Overview

The discussion of effective field theories and integrating out variables to have an effective description at a lower energy scale is tied into the renormalization group. We here discuss the concept of the renormalization group and various procedures that are in use and that we will use to make calculations later.

The renormalization group procedure is a way to transform a description of a theory at higher energy physics into a description at lower energy physics. The procedure results in a loss in the number of degrees of freedom. There are a number of ways to perform this transformation. Three common procedures are those of Kadanoff [124, 125], Wilson [24, 126], and Polchinski [127–129].

3.1 Kadanoff

The Kadanoff procedure, often called the block-spin method, is a lattice procedure. It is also the primary method that we use since we make specific calculations involving various Ising models. The procedure is as follows. Start with a large lattice of spins, σ_r, where r is a vector index labeling the spin's position in the lattice. The action is $S[\sigma; K]$, where K are the coupling constants. The partition function is then,

$$Z = \sum_{\sigma} \exp(-S[\sigma; K]). \tag{3.1}$$

Next, group neighboring spins together. These spins form a block. The block can be formed in all sorts of ways. For example, if we grouped nearest neighbor spins together in a block around a spin at site r, then the grouped spins would consist of σ_r and the spins located one lattice unit away from it; since the nearest neighbors are in the block, the block will look like a diamond.

Next, replace the spins in the block with a single spin at site r that summarizes the spins in the block, σ'_r. There are a variety of ways to do this. One way is to take the average value of all the spins

in the block, σ'_r = average of all the spins in a block centered at σ_r and consisting of σ_r and all spins located one lattice unit away from σ_r. Another possibility is to take the majority spin in the block, aka the majority rule (e.g., if most of the spins were up, then the spin that replaces the block will also be up). So for example, σ'_r = the spin that occurred the most frequently in the block centered at σ_r. If a tie-breaker is needed, the new spin takes a random value of the allowed spin values. See Figure 3.1 for an example of blocking spins in a 2D lattice.

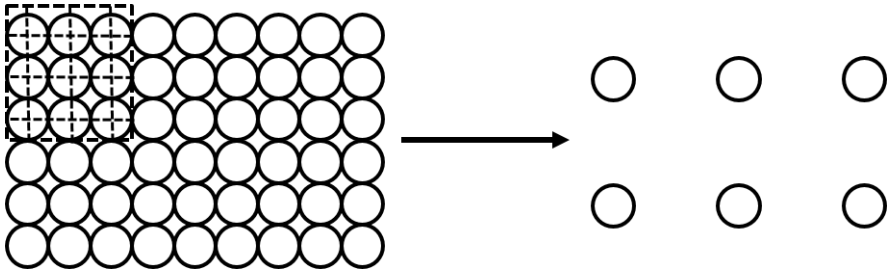

Figure 3.1: An example of grouping spins into blocks in a 2D lattice. In this particular case of blocking, the blocks are squares and contain nine spins; the center spin is chosen for the block, resulting in a lattice with fewer spins and a greater spacing between spins. Other blocking methods could be chosen, e.g., triangles, and different amounts of spins could be included in the blocks.

The replacing of the spins in the block by one spin can be described as a transformation $T[\sigma'|\sigma]$, which is normalized as,

$$\sum_{\sigma'} T[\sigma'|\sigma] = 1. \tag{3.2}$$

The transformation T is known as a block-spin transformation. Notice the similarity to a conditional probability distribution. This similarity will be exploited in later sections.

The new effective action is then defined as,

$$\exp(-S_{eff}[\sigma'; K']) = \sum_{\sigma} T[\sigma'|\sigma] \exp(-S[\sigma; K]), \tag{3.3}$$

where we have taken a further step and changed the coupling constants to K', re-writing the action in terms of local spin operators (to be explained momentarily). Note that the new partition function is the

same as the old partition function,

$$\begin{aligned} Z_{new} &= \sum_{\sigma'} \exp(-S_{eff}[\sigma'; K']) \\ &= \sum_{\sigma'} \sum_{\sigma} T[\sigma'|\sigma] \exp(-S[\sigma; K]) \\ &= \sum_{\sigma} \exp(-S[\sigma; K]) = Z. \end{aligned} \tag{3.4}$$

Because each block of spins was replaced with a single spin (known as a block spin), we now have a lattice of spins that have fewer spins than before. Because the block spins are located at the center of each block, the resulting spins are separated by a larger distance (from center to center), thereby indicating a change of scale to lower energy. Once the spins are replaced with new spin variables, the form of the action can be re-written in terms of the interactions between sites σ'. Hence, we have new couplings K' for these interactions. We now have a new action $S[\sigma'; K']$ that has the same form of interactions as the earlier action but with effective coupling constants and defined on a lattice with fewer sites. This effective action in general will also include higher order interactions than the original action, which can be understood to have had zero coupling constant in the original action.

Continuing our example of using a block centered at r, if the lattice originally had a spacing a, it now has a spacing $\sqrt{2}a$. The new lattice will be rotated relative to the original lattice. Since the new spins are separated by a larger distance, our description of the spin lattice can be understood as a lower energy effective description in terms of the new spin variables σ'_r.

Finally, the new lattice spacing is rescaled to be equivalent to the old spacing. This rescaling will also involve rescaling the spin operators so that the action remains invariant; however, the coupling constants will be unaffected by the rescaling operation. The rescaling is done so that the effective action looks like the original action but with the different coupling constants generated by the renormalization transformation, i.e., the new spins with new couplings are now described by the old lattice. The rescaling is necessary so that fixed points can be defined after iterating the Kadanoff procedure a number of times. However, it might not be possible to iterate the Kadanoff procedure without an approximation if there are new higher order interactions in the effective action; this is a mere practical problem with the procedure arising from the inability to exactly solve models with the higher order interactions.

Aside from a constant (i.e., spin-independent) shift in the action, the only thing that will change in the action is the coefficients in front of the spin operators in the action. The partition function is held fixed, so the physics is the same (alternatively, the constant shift in the action can be absorbed into the partition function, in which case the new partition function will differ from the old by an exponential factor of the constant shift).

3.2 Wilson

The Wilsonian method is a continuum procedure for renormalization. It is not really a different method from the Kadanoff procedure but should be thought of as a continuum limit of the Kadanoff procedure with a particular choice of transformation T. Historically, Wilsonian renormalization was used to find fixed points in an RG flow and thereby find universal behavior [126]. In this section, we will be following the treatment of [24].

Unlike the Kadanoff procedure that is used in position space, the Wilsonian renormalization occurs in momentum space. As with the Kadanoff procedure, one starts with an action $S[\Phi; K]$, where Φ are the original field variables (not necessarily scalar) and the K are the original coupling constants. Using the path integral representation, the partition function is,

$$Z = \int \mathrm{D}\Phi \ \exp(-S[\Phi; K]). \tag{3.5}$$

As before, a transformation must be chosen in order to define the effective action. For the Wilsonian method, variables that represent momenta that are higher than some scale Λ are integrated out, while the variables that represent momenta lower than the scale Λ are left for describing the physics at the lower scale Λ. An easy way to do this is to first define a cutoff for the momentum Λ_0. Then split the fields into fast (χ) and slow modes (ϕ) as,

$$\Phi(x) = \int_{p>\Lambda} \mathrm{d}^{\mathrm{D}}p \ \Phi(p)e^{ipx} + \int_{p<\Lambda} \mathrm{d}^{\mathrm{D}}p \ \Phi(p)e^{ipx} \equiv \chi(x) + \phi(x). \tag{3.6}$$

We can then integrate out the χ fields to produce the lower energy theory. The transformation looks like,

$$T[\phi|\Phi] = \int \mathrm{D}\chi \ \delta(\phi + \chi - \Phi). \tag{3.7}$$

We define an effective action as,

$$\begin{aligned}\exp(-S_{eff}[\phi;K']) &= \int \mathrm{D}\Phi\ T[\phi|\Phi]\exp(-S[\Phi;K]) \\ &= \int \mathrm{D}\chi\ \exp(-S[\phi+\chi;K]),\end{aligned} \tag{3.8}$$

where the K' are again the new coupling constants after re-writing the effective action in terms of local operators. Note that the effective action has a cutoff at the scale Λ because momentum between Λ and Λ_0 was integrated out. The effective action therefore describes the physics in terms of lower energy variables, i.e., the scale has been changed. The partition function is again the same,

$$\begin{aligned}Z_{new} &= \int \mathrm{D}\phi\ \exp(-S_{eff}[\phi;K']) \\ &= \int \mathrm{D}\phi\mathrm{D}\chi\ \exp(-S[\phi+\chi;K]) \\ &= \int \mathrm{D}\Phi\ \exp(-S[\Phi;K]) = Z.\end{aligned} \tag{3.9}$$

As with the Kadanoff procedure, the action must be re-written in terms of local interactions and effective coupling constants K'. When doing so, new higher order interactions will generally appear; indeed, all possible higher order terms consistent with the symmetries will appear and so will all lower order terms (including a constant operator-independent term: the identity operator). Also, to return the effective action to its original form and to be able to iterate the Wilsonian procedure to find fixed points, a rescaling of the momenta must be done to return the cutoff to its original value Λ_0. After rescaling, one could even relabel the fields ϕ as Φ so that the end result is an effective action that looks the same as the original action but with different couplings and additional interactions.

The Wilsonian procedure uses a specific transformation T that allows the procedure to continually be iterated, which is usually used to find fixed points in the RG flow. The rescaling does not affect the numerical values of the coupling constants, and the form of the action is the same, so there is again a flow in the coupling constants with a change in scale (understanding the new higher order interactions to have had coupling constant of zero in the original action) while the rest of the action remains the same: $S[\Phi;K] \to S[\Phi;K']$.

Also as with the Kadanoff procedure, a constant, operator-independent term will appear in the effective action. This term can be absorbed into the partition function or left in the effective action to keep the partition function fixed.

3.3 Wilson and the Callan-Symanzik Equation

The conceptual result of Wilsonian renormalization as changing the scale of a theory can be applied to the partition function. Doing so results in an example of a Callan-Symanzik equation. A Callan-Symanzik equation shows how observables change with scale to compensate for changes in the coupling constants of the action as the scale is changed. Callan-Symanzik equations are then useful for solving for the couplings and renormalization factors (like the wave-function renormalization factor) as a function of scale, and we will make use of relating observables to changes in scale when we make use of the KL divergence later.

Let us now see an example of a Callan-Symanzik equation. Let $Z_\Lambda(g(\Lambda))$ be the partition function at scale Λ with coupling constants $g(\Lambda)$. Because the partition function is invariant under scale changes, we must have [24, 27],

$$\Lambda \frac{d}{d\lambda} Z_\Lambda(g) = \left(\Lambda \frac{\partial}{\partial \Lambda}|_{g(\Lambda)} + \Lambda \frac{\partial g(\Lambda)}{\partial \Lambda} \frac{\partial}{\partial g}|_\Lambda \right) Z(g) = 0. \tag{3.10}$$

By definition,

$$\beta \equiv \Lambda \frac{\partial g(\Lambda)}{\partial \Lambda} = \frac{\partial g(\Lambda)}{\partial \log(\Lambda)} \tag{3.11}$$

is the beta function for the coupling constant. Likewise, the wavefunction renormalization factor, $\mathcal{Z}$, will reappear in the kinetic term after performing a renormalization group transformation, and its logarithmic scale derivative is also labeled as,

$$\gamma \equiv -\frac{1}{2} \Lambda \frac{\partial \log(\mathcal{Z}(\Lambda))}{\partial \Lambda} = -\frac{1}{2} \frac{\partial \log(\mathcal{Z}(\Lambda))}{\partial \log(\Lambda)}, \tag{3.12}$$

and γ is known as the anomalous dimension of the field.

Renormalization of correlation functions changes the correlation functions as,

$$\langle \phi(x_1) ... \phi(x_n) \rangle = \mathcal{Z}^{-n/2}(\Lambda) \langle \phi_0(x_1) ... \phi_0(x_n) \rangle, \tag{3.13}$$

where ϕ_0 is the bare field before redefining in terms of the canonically renormalized ϕ field. Correlation functions should also remain the same after a scale change, so the Callan-Symanzik equation for correlation functions Γ is,

$$\Lambda \frac{d}{d\Lambda}\Gamma(x_1, ..., x_n; g(\Lambda)) = \left(\Lambda \frac{\partial}{\partial \Lambda}|_{g(\Lambda)} + \beta(\Lambda)\frac{\partial}{\partial g} + n\gamma(\Lambda)\right)\Gamma(x_1, ..., x_n; g(\Lambda)) = 0. \tag{3.14}$$

Solving for the beta functions and anomalous dimension gives the partial differential equations that can then be used to solve for the couplings and renormalization factors in terms of scale as done in [24, 27].

3.4 Polchinski

The Polchinski renormalization approach is computationally the same as Wilson's, except instead of having a hard momentum cutoff, a smooth momentum regulator is used. The regulator is a function of the momentum and becomes large when the momentum is greater than the cutoff and does not affect the action for the momentum below the cutoff. For example, a cutoff function $K(p^2/\Lambda^2)$ is defined to be 1 for momentum below the scale Λ and rapidly and smoothly go to zero above the scale Λ [127]. However, Polchinski's approach differs in that the action remains defined in the continuum without having recourse to a lattice (as Wilson's approach does), and the change of the action is observed and computed as the energy scale changes: the action flows about in coupling space as renormalization is applied.

One way to derive Polchinski's equation is to make use of the fact that the scale derivative of the partition function is zero, as discussed earlier. One then finds a differential form of Wilson's effective action equation. This differential equation is known as Polchinski's Exact Renormalization Group Equation (ERGE), and it shows how the action changes with the scale Λ of the theory. The differential equation is [24],

$$-\frac{\partial S_{int}}{\partial \log \Lambda} = \int \mathrm{d^D}x \mathrm{d^D}y \left(\frac{\delta S_{int}}{\delta \phi(x)} D_\Lambda(x,y) \frac{\delta S_{int}}{\delta \phi(y)} - D_\Lambda(x,y)\frac{\delta^2 S_{int}}{\delta\phi(x)\delta\phi(y)}\right), \tag{3.15}$$

where S_{int} is the interaction piece of the action. $D_\Lambda(x,y)$ is the infinitesimal χ propagator resulting from lowering the energy scale from Λ' to $\Lambda = \Lambda' - \delta\Lambda$ and is equal to

$$D_\Lambda(x,y) = \frac{1}{(2\pi)^D}\frac{\Lambda^{D-1}\delta\Lambda}{\Lambda^2 + m^2}\int_{S^{D-1}} \mathrm{d}\Omega\ e^{i\Lambda\hat{p}\cdot(x-y)}. \tag{3.16}$$

This is the position space version of Polchinski's ERGE, which we will find useful later when using the KL divergence with continuum field theories. Polchinski originally derived a momentum space version of the ERGE, and we will now sketch a derivation of the momentum space ERGE, following Polchinski [127] and [129]. As Polchinski, we use an action with an external source $J(p)$, where $J(p) = 0$ for $p > \Lambda$ the cutoff. The action is,

$$S[\phi] = \frac{1}{2}\int \frac{\mathrm{d}^4 p}{(2\pi)^4}\ \phi(p)\phi(-p)p^2 K^{-1}(p^2/\Lambda^2) + S_{int}[\phi], \tag{3.17}$$

where we have placed possible mass interactions into the interaction part of the action, $S_{int}[\phi]$. The generating functional is then,

$$Z[J] = \int \mathrm{D}\phi\ \ \exp(-S[\phi] + J\cdot\phi). \tag{3.18}$$

The goal is for the generating functional to not change with scale. That is, we want to impose $\Lambda\,(dZ/d\Lambda) = 0$. Taking the derivative produces,

$$\Lambda\frac{dZ[J]}{d\Lambda} = \int \mathrm{D}\phi\ \left(\frac{-1}{2}\int \frac{\mathrm{d}^4 p}{(2\pi)^4}\ \phi(p)\phi(-p)p^2\Lambda\frac{\partial K^{-1}(p^2/\Lambda^2)}{\partial\Lambda} + \Lambda\frac{dS_{int}}{d\Lambda}\right)\exp(-S[\phi] + J\cdot\phi). \tag{3.19}$$

This equation ignores field independent terms that would appear.

A solution for $\Lambda\,(dS_{int}/d\Lambda)$ that satisfies $\Lambda\,(dZ/d\Lambda) = 0$ is Polchinski's ERGE,

$$\Lambda\frac{dS_{int}}{d\Lambda} = \frac{-1}{2}\int \frac{\mathrm{d}^4 p}{(2\pi)^4}\ \phi(p)\phi(-p)p^{-2}\Lambda\frac{dK(p^2/\Lambda^2)}{d\Lambda}\left(\frac{\delta S_{int}}{\delta\phi(p)}\frac{\delta S_{int}}{\delta\phi(-p)} - \frac{\delta^2 S_{int}}{\delta\phi(p)\delta\phi(-p)}\right). \tag{3.20}$$

Other exact renormalization group equations can be derived by having different momentum cutoffs, including a general cutoff. As already noted, a sharp cutoff corresponds to Wilson's method. Other effects, such as the effect of rescaling, can be included in the ERGE, for which we refer to the derivations and equations in [129].

We have now discussed the three main RG techniques: Kadanoff, Wilson, and Polchinski. We will now look into the AdS/CFT interpretation of the RG transformation that we will make use of from time to time in this thesis.

3.5 AdS/CFT Holography

The renormalization group flow can be understood and interpreted in terms of the AdS/CFT correspondence [4, 7, 130–133]. We will sketch how this can be done below.

The AdS/CFT correspondence states that a string theory on an asymptotically AdS space times a compact manifold M in $d+1$ dimensions is dual to a QFT in d dimensions at the boundary of the asymptotically AdS space. The string theory at strong coupling is dual to a weak coupling in the QFT, and weak coupling in the string theory is dual to a strong coupling in the QFT. This duality is supposed to be an exact equivalence, and it has been proven to be exact in some cases. The duality arises in the large N limit of the QFT.

On example of this correspondence is type IIB string theory in $AdS_5 \times S^5$ and $\mathcal{N} = 4$ Super Yang Mills theory in 4 dimensions. Following deBoer's summary [130], the Poincare patch of the AdS_5 is described by,

$$ds^2 = dr^2 + \exp(2r/L)\eta_{\mu\nu}dx^\mu dx^\nu, \tag{3.21}$$

where L is the radius of the AdS_5 and $r \to \infty$ is the boundary of the AdS_5. The radius $L = (g_{YM}^2 N)^{1/4}$. The low energy limit of the string theory with D3 branes is the Yang Mills theory, while the low energy limit of the Yang Mills theory gives rise to the $AdS_5 \times S^5$. The $\mathcal{N} = 4$ super Yang Mills is a CFT, and we can see indeed that the metric (3.21) is invariant under,

$$\begin{aligned} r &\to r + a, \\ \eta_{\mu\nu} &\to \eta_{\mu\nu} \exp(-2a/L). \end{aligned} \tag{3.22}$$

The metric $\eta_{\mu\nu}dx^\mu dx^\nu$ is to be identified with the metric on the boundary QFT, so we see that this metric is invariant under a change of scale. Indeed, we see that a change of scale compensates for moving radially through the AdS_5.

In general, the radial coordinate r is identified with the energy scale of the boundary theory, and therefore, r is related to the cutoff of the boundary theory. Truncating the AdS space with an IR cutoff at $r = r_0$ will correspond to a UV cutoff in the boundary theory. Thus, we have a relation between RG flow on the boundary theory and radial movement in the AdS space. However, it is not known what regularization or renormalization scheme this cutoff is to be identified with [130].

There are relations between the correlation functions of fields in the bulk Φ and correlation functions of operators in the boundary theory O. Working in the low energy limit of string theory, we use the supergravity approximation that has an asymptotically AdS ground state. The result for the bulk partition function is then [131],

$$\begin{aligned} Z_{bulk}[\phi_(0)] &= \int_{\Phi_{\partial AdS}=\phi_{(0)}} \mathrm{D}\Phi \ \exp(-S[\Phi]) \\ &= \langle \exp(-\int_{\partial AdS} \phi_{(0)} O)\rangle_{QFT}, \end{aligned} \tag{3.23}$$

where the last integral is intended to be over the boundary ∂AdS of the asymptotically AdS theory, the $\phi_{(0)}$ are the values of the boundary fields, and the expectation value is taken with the path integral on the boundary QFT. As part of the limits of the path integration, the $\phi_{(0)}$ becomes the boundary value of the bulk field Φ, and it becomes a source for the boundary operators. Correlation functions of the operator O can now be calculated by taking functional derivatives. In a leading order saddle point approximation, we have $S_{onshell}[\phi_{(0)}] = -W_{QFT}[\phi_{(0)}]$, where W is the generating function for the connected QFT graphs. Hence,

$$\begin{aligned} \langle O(x)\rangle &= \frac{\delta S_{onshell}}{\delta \phi_{(0)}(x)}|_{phi_{(0)}=0}, \\ \langle O(x_1)O(x_2)...O(x_n)\rangle &= (-1)^{n+1} \frac{\delta^n S_{onshell}}{\delta\phi_{(0)}(x_1)...\delta\phi_{(0)}(x_n)}|_{phi_{(0)}=0}. \end{aligned} \tag{3.24}$$

These expressions still require renormalization. The need to regulate the infinite spacetime volume of the AdS corresponds to the need for regulating the UV divergences in the QFT.

3.6 RG in General

In general, the renormalization procedure maps one set of operators $\{O\}$ at one energy scale Λ to another set of operators $\{O'\}$ that exist at a smaller energy scale Λ', while retaining the same form of the action in order to keep the partition function fixed: a transformation from one set of operators to a smaller set of operators followed by a rescaling of the points where the operators act. Since the partition function is fixed, the physical system is still the same, although it is described by different operators. The operators at the larger scale are composites of the operators at the smaller scale, e.g., a proton operator is described by a composite of quark operators at a smaller scale.

Furthermore, since one set of operators is mapped into another set, not all operators remain when the scale is changed under the RG flow. Operators that are highly suppressed when going to larger scales

under the RG flow are termed irrelevant operators. Operators that are present at the larger scale of the problem and that increase in strength with a larger scale are termed relevant operators. Operators that exhibit a different behavior to always increasing or always decreasing are termed marginal. As expected from our discussion of effective field theories, this allows physics at a larger scale to be independent of the operators from a smaller scale. So for example, it is possible to describe the physics of a proton at its own scale without having to refer to quarks because the quark operators are only relevant when we try to describe physics at their smaller scale.

Thinking of the RG transformation in this general setting lends itself to an information theoretic interpretation of the RG flow, which we will use to quantiy the information lost upon performing the RG transformation and to quantify the information that remains about the UV theory in the effective theory. We shall now turn to that subject.

CHAPTER 4
Communication Channel

Because irrelevant operators are suppressed and degrees of freedom are lost when performing a renormalization group transformation, information is lost as the renormalization group transformation is performed. We can view this loss of information as a communication channel problem. We start with one set of operators that have some amount of information. This information is sent down a noisy communication channel [134]: the renormalization group transformation is the noisy channel. We then end up with another set of operators: operators and couplings that are a result of the operators at the smaller scale, e.g., the proton is a result of the quark interactions. Our larger scale operators then must have some information about the smaller scale operators. Because the channel is noisy, some information is lost and so not all information about the smaller scale operators is present in the larger scale operators.

Understanding the RG transformation in terms of information lost in a communication channel raises a number of questions. How can we measure this information loss? In what sense is this information being lost? Having quantified it, what is the capacity of the communication channel? Is there an optimum theory or renormalization procedure we can choose that minimizes the information lost in the channel or maximizes the capacity? We discuss the possible answers to these questions in the next sections.

4.1 The KL Divergence

We first discuss the KL divergence [135–138] for normalized probability distributions p and q associated with different statistical field theories, which is the main tool we use in this paper to quantify information. In the case where $p = p(X, Y)$, we also define marginalized distributions as,

$$\begin{aligned} p_1(X) &\equiv \int \mathrm{D}\mu_2 \; p(X, Y) \\ p_2(Y) &\equiv \int \mathrm{D}\mu_1 \; p(X, Y), \end{aligned} \tag{4.1}$$

where $\mathrm{D}\mu_2$ indicates integrating over the space associated with the variable Y, resulting in integrating out the variable Y. The $\mathrm{D}\mu_1$ is defined in like manner.

The KL divergence is non-negative and results in a c-number, but it is not a metric: it is generally asymmetric. The KL divergence can be understood as the relative entropy of theory p relative to theory q. In the context of probability theory, the KL divergence is understood as the difference or distance (proximity) between probability distributions p and q, where q is a best guess distribution chosen to approximate or model p. The probability that the best guess distribution q is in fact the distribution p after N draws (at large N) is [137–139],

$$Pr(p|E) = e^{-ND_{KL}(p||q)}, \tag{4.2}$$

where E is a set of N independent events drawn from the distribution q.

The KL divergence can also be used to calculate the mutual information between random variables X and Y associated with the distributions $p(X,Y)$ and $q(X,Y)$. For marginalized distributions p_1 and p_2, the mutual information is defined as [137],

$$\begin{aligned} I(X;Y) \equiv D\left(p(X,Y)||p_1(X)p_2(Y)\right) &= \int \mathrm{D}\mu \; p \log \frac{p}{p_1 p_2} \\ &= \int \mathrm{D}\mu \; p \log p - \int \mathrm{D}\mu_1 \; p_1 \log p_1 - \int \mathrm{D}\mu_2 \; p_2 \log p_2, \end{aligned} \tag{4.3}$$

where advantage has been taken of the fact that the $\log p_{1,2}$ terms do not depend on the space in $\mu_{2,1}$, respectively. Hence, the space can be summed over for the distribution p, producing marginalized distributions multiplying the $\log$ terms.

Although the KL divergence is generally asymmetric, it is symmetric to lowest order. It is also quadratic to lowest order, producing the Fisher information metric, which thereby allows for an appropriate distance metric between probability distributions. The presence of the Fisher information metric to lowest order is also important because we could have considered more general alpha divergences, instead of considering the KL divergence (a special case of alpha divergences when $\alpha = 1$). However, these alpha divergences also all reduce to the Fisher information metric to lowest order, so we lose no opportunities by using the KL divergence in our measurements of proximity between distributions/field theories.

The claims that the KL divergence is quadratic to lowest order and produces the Fisher information metric to lowest order will be shown simultaneously below. Consider a distribution p that can be obtained from q by changing p's parameters, causing p itself to change by some (not necessarily infinitesimal) amount dp: $q = p + dp$. To lowest order, the KL divergence of q and p is then,

$$D(p||q) = \int \mathrm{D}\mu \; p \log \frac{p}{q} = \int \mathrm{D}\mu \; p \log \frac{p}{p+dp} = -\int \mathrm{D}\mu \; p \log(1 + \frac{dp}{p}), \tag{4.4}$$

where, as before, the $\mathrm{D}\mu$ just integrates over the space on which the probability distributions p and q are defined. We expand the logarithm about the parameter dp/p to the lowest order (which is quadratic) in dp to get,

$$D(p||q) = \int \mathrm{D}\mu \; -(pdp/p - p\frac{dp^2}{2p^2}) + O(dp^3) = \int \mathrm{D}\mu \; (-dp + p\frac{dp^2}{2p^2}) + O(dp^3), \tag{4.5}$$

where we simply define $dp^3 \equiv (dp)^3$. We now drop the dp in the integrand because it is assumed the probability distributions fall to zero towards the boundary of integration, which then gives,

$$D(p||q) = \int \mathrm{D}\mu \; p\frac{dp^2}{2p^2} + O(dp^3) = \langle \frac{dp^2}{2p^2} \rangle_p + O(dp^3), \tag{4.6}$$

where the integral becomes an expectation value with respect to p. We now put the result in terms of both q and p to remove the dependence on dp, and we also put the result in terms of $d \log p$ to show the quadratic symmetry of the result. We arrive at,

$$\begin{aligned} D(p||q) &= \frac{1}{2} \langle (\frac{q}{p} - 1)^2 \rangle_p + O(dp^3) \\ &= \frac{1}{2} \langle d\log p \, d\log p \rangle_p + O(dp^3). \end{aligned} \tag{4.7}$$

We see then the quadratic symmetry of the KL divergence to lowest order, and we see that the Fisher information metric appears as the lowest order term! Of course, if dp is small enough, then this would not only be the lowest order piece but also the only piece of the KL divergence needed for calculations to lowest order.

On the other hand, we will see the same result to lowest order for the KL divergence with p and q reversed,

$$\begin{aligned} D(q||p) &= \int \mathrm{D}\mu \; q \log \frac{q}{p} = \int \mathrm{D}\mu \; (p+dp) \log(1+\frac{dp}{p}) \\ &= \int \mathrm{D}\mu \; p\frac{dp}{p} + dp\frac{dp}{p} - p\frac{dp^2}{2p^2} + O(dp^3) = \int \mathrm{D}\mu \; p\frac{dp^2}{2p^2} + O(dp^3), \end{aligned} \tag{4.8}$$

where we have carried out many of the same steps as before, e.g., substituting for the value of q in terms of p, expanding the logarithm, and dropping the boundary dp term. We continue to simplify the result as,

$$\begin{aligned} D(q||p) &= \langle \frac{dp^2}{2p^2} \rangle_p + O(dp^3) = \frac{1}{2} \langle (\frac{q}{p} - 1)^2 \rangle_p + O(dp^3) \\ &= \frac{1}{2} \langle d\log p \, d\log p \rangle_p + O(dp^3), . \end{aligned} \tag{4.9}$$

where we have again placed the result in terms of q and p and in terms of $d\log p$. This is the exact same result as the KL divergence with p and q reversed in the arguments, so we see that the KL divergence is indeed both quadratic and symmetric in its arguments to lowest order.

When studying QFTs, it is easier to Wick rotate and study Euclidean QFTs. Hence, we use Boltzmann factors for the probability distributions (as in statistical field theory) that we will be measuring with the KL divergence. In terms of the Hamiltonian, the Boltzmann distributions go as,

$$p = \frac{e^{-\beta H_p}}{Z_p}, \tag{4.10}$$

where β is the inverse temperature and Z_p is the partition function. For general Boltzmann factors $e^{-\kappa S}$ (with a constant κ for generality and S some sort of an action), the KL divergence is,

$$\begin{aligned} D(p||q) &= \int \mathrm{D}\mu \; \frac{e^{-\kappa S_p}}{Z_p} \log \frac{e^{-\kappa S_p}/Z_p}{e^{-\kappa S_q}/Z_q} \\ &= \int \mathrm{D}\mu \; \frac{e^{-\kappa S_p}}{Z_p} \left(\log \frac{Z_q}{Z_p} + \kappa(S_q - S_p) \right) \\ &= \log \frac{Z_q}{Z_p} - \kappa \langle S_q - S_p \rangle_p, \end{aligned} \tag{4.11}$$

where $\langle\rangle_p$ again indicates that the expectation value is taken with respect to theory p. The partition function to lowest order is,

$$\begin{aligned} Z_q &= \int \mathrm{D}\mu \; e^{-\kappa S_q} = \int \mathrm{D}\mu \; e^{-\kappa S_p} e^{-\kappa(S_q - S_p)} \\ &= Z_p(1 - \kappa\langle(S_q - S_p)\rangle_p + ...). \end{aligned} \tag{4.12}$$

So to lowest order, the KL divergence of the general Boltzmann distributions we defined is (making use of our earlier result (4.7)),

$$D(p||q) = \frac{1}{2}\langle(\frac{q}{p} - 1)^2\rangle_p + O(dp^3) = \frac{1}{2}\langle(\frac{Z_p}{Z_q}e^{-\kappa(S_q - S_p)} - 1)^2\rangle_p + O(dp^3), \tag{4.13}$$

where we have used the general Boltzmann distributions for q and p. We now expand the exponential and the partition functions to lowest order (quadratic in κ), making use of our earlier result for Z_q (4.12) to get,

$$\begin{aligned} D(p||q) &= \frac{1}{2}\langle(\frac{1}{1 - \kappa\langle(S_q - S_p)\rangle_p}(1 - \kappa(S_q - S_p)) - 1)^2\rangle_p + O(dp^3, \kappa^3) \\ &= \frac{1}{2}\langle((1 + \kappa\langle(S_q - S_p)\rangle_p)(1 - \kappa(S_q - S_p)) - 1)^2\rangle_p + O(dp^3, \kappa^3). \end{aligned} \tag{4.14}$$

Multiplying through now gives (neglecting terms of higher order than κ^2),

$$\begin{aligned} D(p||q) &= \frac{1}{2}\langle(1 + \kappa\langle(S_q - S_p)\rangle_p - \kappa(S_q - S_p) - 1)^2\rangle_p + O(dp^3, \kappa^3) \\ &= \frac{1}{2}\langle(\kappa\langle(S_q - S_p)\rangle_p - \kappa(S_q - S_p))^2\rangle_p + O(dp^3, \kappa^3), \end{aligned} \tag{4.15}$$

where the 1's cancel in the second line. We then expand the square to get,

$$\begin{aligned} D(p||q) &= \frac{1}{2}\langle(\kappa\langle(S_q - S_p)\rangle_p)^2 + \kappa^2(S_q - S_p)^2 - 2\kappa^2\langle(S_q - S_p)\rangle_p(S_q - S_p)\rangle_p + O(dp^3, \kappa^3) \\ &= \frac{\kappa^2}{2}(\langle(S_q - S_p)^2\rangle_p - (\langle(S_q - S_p)\rangle_p)^2) + O(dp^3, \kappa^3) \\ &= \frac{\kappa^2}{2}(\langle(S_p - S_q)^2\rangle_p - (\langle(S_p - S_q)\rangle_p)^2) + O(dp^3, \kappa^3), \end{aligned} \tag{4.16}$$

where we have distributed the expectation value to the other terms in the last two lines, and we emphasize the symmetry between p and q in the KL divergence by rearranging the distributions in the last line.

When we put in operators for the actions S_q and S_p, we end up with integrated two-point functions. These two-point functions have an IR divergence due to a volume factor, and they have a UV divergence

due to contact terms [9]. The UV divergence will be discussed in Section 6.3. The IR divergence suggests that the physical quantity of interest may be the KL divergence per space-time volume, rather than the KL divergence alone.

Notice also from our result and from our discussion of the KL divergence that the KL divergence is general enough that any two (Euclidean) QFTs could be compared, so long as they have the same field content or descend from the same master theory [9] (e.g., maybe both are deformations of the same string compactification to lower energy effective field theories). The two theories do not necessarily have to be related by an RG transformation, and if we directly had the distributions for both an UV theory and an IR theory, we could compute the KL divergence without having to find the IR theory from the UV theory first. However, we will be turning our attention to UV and IR distributions that are related by a change of scale/RG flow so that we can try to isolate the UV information and interpret the RG flow as a communication channel.

Having discussed the KL divergence and its properties, we will discuss in the remainder of the section the uses of the KL divergence in quantifying information and possible tools we can use to quantify the information about UV physics accessible at IR scales.

4.2 Channel Capacity and UV/IR Mixing

One way to find the information about the UV theory in the IR theory is by directly using the mutual information (4.3). UV-IR mixing is then studied in a theory $p(UV, IR)$ by forming a joint distribution $q = p_{UV}p_{IR}$, where,

$$\begin{aligned} p_{UV} &= \int \mathrm{D}\mu_{IR}\; p(UV, IR) \\ p_{IR} &= \int \mathrm{D}\mu_{UV}\; p(UV, IR), \end{aligned} \tag{4.17}$$

but there is a conceptual difficulty in summing over the IR space: summing over the IR space means we are integrating out IR modes, which leads to non-local interactions in the effective field theory. So in Section 5, we study the QFT on a lattice in order to better understand UV/IR mixing. A lattice QFT avoids the problem of summing over IR modes by exploring the question in position space and providing a well-defined marginalization procedure via decimation.

Another quantity of interest that might be used for quantifying UV information in IR physics and is related to the KL divergence is the channel capacity [137], and we will discuss the channel capacity in

the rest of this subsection. The channel capacity measures the upper bound on how much information can pass through a communication channel: it is the maximum rate that information can be transferred through a communication channel with zero error. Channel capacity is defined as,

$$C = sup_{p_X(x)} I(X;Y), \tag{4.18}$$

where I is the mutual information between the X and Y random variables, and the supremum is taken over all possible probability distributions (produced by Lorentz invariant actions) for the random variable X.

In our case, we will take random variables ϕ (IR fields) and J (the usual field theory generator, which functions as a coupling constant for the linear ϕ term) as follows. Thinking in terms of string compactifications, the J coupling constant becomes a dynamical field variable with $p(J)$ being viewed as the random dynamics of moduli that cause J to fluctuate in value. In an AdS/CFT interpretation, the $p(J)$ corresponds to fluctuations in the couplings on the CFT boundary theory due to string fluctuations in the AdS space. The use of $p(J)$ in the relative entropy could also be viewed as a disorder average over J, where $p(\phi)$ is the result of the disorder average.

We see then from our discussion that J is a UV variable, so we can directly find the mutual information between the UV and IR variables. The mutual information between ϕ and J becomes,

$$\begin{aligned} p(\phi) &\equiv \int \mathrm{D}J\ p(\phi, J) = \int \mathrm{D}J\ p(\phi|J)p(J), \\ I(\phi, J) &= D_{KL}(p(\phi, J)||p(\phi)p(J)) = \int \mathrm{D}\phi\ p(\phi, J) \log \frac{p(\phi|J)}{p(\phi)}, \end{aligned} \tag{4.19}$$

where notice that $p(J)$ does not need to explicitly appear in the calculation at this stage, so one does not need to think in terms of integrating out IR modes, thereby removing the conceptual difficulty of the previous section. The action for $p(\phi|J)$ is of the following form,

$$S = S[\phi] + J \cdot O[\phi], \tag{4.20}$$

where $O[\phi]$ is some functional of ϕ fields that couples to J.

When calculating the channel capacity, which requires maximizing over all $p(J)$, it seems unavoidable that one will need to speak of $p(J)$. However, this still does not require one to think in terms of integrating

out IR modes because the $p(J)$ is simply just a given of the problem (rather then resulting from integrating out), and it describes the fluctuations of the coupling constant J.

Another quantity of interest for a communication channel that is related to channel capacity is the distortion rate. It is the minimum amount of information in a signal that needs to be sent so that the signal can be reconstructed up to some given distortion D. It is defined as the infimum of the mutual information over all possible conditional probability distributions, i.e., over all possible communication channels. We will discuss this more with the concept of the information bottleneck in Section 4.4. However, we will first draw a connection between a communication channel (and its properties, such as the channel capacity) and the RG transformation.

4.3 Communication Channel: RG Connection

We can use the channel capacity to understand our RG procedure as follows. As we have done before when discussing the RG transformation, a RG procedure can be encapsulated in a kernel (or transition or transformation matrix), which obeys the following property,

$$\int \mathrm{D}\phi' \; T(\phi', \phi) = 1, \tag{4.21}$$

where ϕ' are the new variables that result after an RG transformation and ϕ are the original variables. T applies the RG transformation as follows,

$$p(\phi') = \int \mathrm{D}\phi \; T(\phi', \phi) p(\phi), \tag{4.22}$$

which given (4.23) implies,

$$\int \mathrm{D}\phi' \; p(\phi') = 1. \tag{4.23}$$

For Boltzmann factors, T with the above properties preserves the partition function.

We can now interpret T as a conditional probability distribution,

$$p(\phi', \phi) \equiv p(\phi'|\phi) p(\phi) = T(\phi', \phi) p(\phi). \tag{4.24}$$

T is normalized as a conditional probability distribution, and T can be used to reproduce the marginal distributions of both ϕ and ϕ' in the manner of a conditional probability distribution, so T mathematically behaves as a conditional probability distribution. Physically, we see that T transforms the ϕ variables into ϕ', so it is a transition matrix, which is a conditional probability distribution over operators.

Defining then $T \equiv p(\phi'|\phi)$, we can make a connection to a communication channel: this conditional distribution T gives the probability of receiving variables ϕ' given source variables ϕ. T is then the distribution for the communication channel.

Furthermore, we can understand marginalizing distributions as being produced by the distribution $p(\phi'|\phi) = \delta(\phi' - \phi)$ (as one example; other, smoother choices could be made). That is, we have,

$$\begin{aligned} p_{UV} &= \int \mathrm{D}\mu_{IR}\ \delta(UV - IR)p(IR) \\ p_{IR} &= \int \mathrm{D}\mu_{UV}\ \delta(UV - IR)p(UV). \end{aligned} \tag{4.25}$$

We see then a connection between the RG transformation and a communication channel. We can view the IR variables as a signal from the UV variables sent through a communication channel defined by the RG transformation. In the case of a lattice with operators localized at points, we can make the same interpretation for decimation schemes and other block transformations.

Having interpreted the RG transformation as a communication channel, we can then ask questions about its properties. What is the channel capacity of this communication channel? What is the distortion rate? What distributions will maximize mutual information and minimize distortion rate?

We can make a guess at the distributions that make the channel operate at the channel capacity: because a CFT is invariant under the RG transformation, the CFT should be the optimal distribution so as to communicate at the channel capacity rate. A general QFT that has undergone an RG flow loses information after the transformation, but a CFT remains invariant, having the same information before and after. Hence, it must be the input that allows for the maximum amount of information to be communicated without error.

As for minimizing the distortion rate, we have discussed how the CFT has a dual ADS space and how the ADS space can be viewed as an RG flow by slices in the radial direction. So the dual AdS space can be viewed as a communication channel in the radial direction. Because the CFT does not change as it

passes through this channel, there is zero distortion and so the ADS channel should provide the optimal channel distribution that minimizes distortion rate.

These statements about AdS/CFT conceptually make sense, but we have not been able to prove them yet.

Having discussed the connection between the RG transformation and a communication channel with its properties, we will discuss another potential tool for quantifying information by making use of the KL divergence: the information bottleneck.

4.4 Information Bottleneck

The concept of the information bottleneck provides another way to think about our problem. We will be following the presentation in [3], which relies heavily on the presentation of [137] for the mutual information and distortion rate material.

Consider a signal $x \in X$ used to predict a signal $y \in Y$. The relevant information in x is defined as the amount of information about y in x. There is a certain minimal amount of information in the signal x that is needed to predict the signal y. Equivalently, there is a way to represent X in a short code so as to preserve the maximum information about Y. The information about Y in X is thus squeezed through the bottleneck of a set of codewords $\tilde{X}$. We thereby have for the information flow,

$$\begin{aligned} X &\to \tilde{X} \\ \tilde{X} &\to Y. \end{aligned} \tag{4.26}$$

Information about Y in X is represented by a limited set of codewords $\tilde{X}$ and then the limited set of codewords is used to predict Y with the goal of maximizing the amount of information about Y.

The mapping between elements of X and the codewords $\tilde{x} \in \tilde{X}$ can be represented by a probability density $p(\tilde{x}|x)$. The mapping thereby partitions X into blocks with each block associated to a codeword $\tilde{x}$ with probability

$$p(\tilde{x}) = \sum_x p(x)p(\tilde{x}|x). \tag{4.27}$$

Notice here that the probability $p(\tilde{x})$ is just one of the marginal distributions produced from the joint $p(x, \tilde{x})$.

To determine the quality of a signal, two things are needed: the rate of transmission and the accuracy of the transmission. The rate of transmission, i.e., the average number of of bits per message needed to specify an element in the codebook without confusion, is determined by the mutual information. The mutual information bounds from below the rate per element of X,

$$I(X;\tilde{X}) = \sum_{x\in X}\sum_{\tilde{x}\in\tilde{X}} p(x,\tilde{x}) \log\left(\frac{p(\tilde{x}|x)}{p(\tilde{x})}\right). \tag{4.28}$$

To determine the accuracy of a signal, the distortion function is used. The distortion function is supposed to be small, and so it gives a measure for the most relevant aspects of X. The expected distortion for the partitioning of X from $p(\tilde{x}|x)$ is,

$$\langle d(x,\tilde{x})\rangle = \sum_{x\in X}\sum_{\tilde{x}\in\tilde{X}} p(x,\tilde{x}) d(x,\tilde{x}). \tag{4.29}$$

As noted earlier, the rate distortion function $R(D)$ is given by the minimal achievable rate (i.e., the minimal mutual information) for a given distortion D,

$$R(D) \equiv \min I(X;\tilde{X}), \tag{4.30}$$

where the minimum is taken over $\{p(\tilde{x}|x) : \langle d(x,\tilde{x})\rangle \leq D\}$. The rate distortion function $R(D)$ captures the relation between rate and distortion. A larger rate R means there is a smaller achievable distortion D. The optimal distribution can be found, and it is,

$$p(\tilde{x}|x) = \frac{p(\tilde{x})}{Z(x,\beta)} \exp(-\beta d(x,\tilde{x})), \tag{4.31}$$

where Z is a normalization function (the partition function) and β is the Lagrange multiplier (and is positive) used to solve the constrained optimization problem for $R(D)$,

$$\frac{\delta R}{\delta D} = -\beta. \tag{4.32}$$

A difficulty with the usual distortion rate tools is that it is hard to find a correct and non-arbitrary distortion measure $d(x,\tilde{x})$. So Tishby et al. propose using an information bottleneck to determine the relevant

information content of a signal [3]. Access to $p(x, y)$ is assumed, like access to $p(x)$ is assumed for rate distortion theory. X and Y must have positive mutual information: the relevant information about Y is found in X. $\tilde{X}$ compresses X as much as possible but so as to preserve as much information about Y as possible. Hence, the following relation must hold,

$$I(\tilde{X}; Y) = \sum_y \sum_x p(y, \tilde{x}) \log \frac{p(y, \tilde{x})}{p(y)p(\tilde{x})} \leq I(X, Y), \tag{4.33}$$

since the compression cannot in general preserve more information about Y than what X has. The goal is then to keep fixed the amount of information about Y in X while minimizing the number of bits needed to represent that information in $\tilde{X}$. The method of Lagrange multipliers then gives the following functional to minimize,

$$L[p(\tilde{x}, x)] = I(\tilde{X}; X) - \beta I(\tilde{X}; Y). \tag{4.34}$$

It is then shown that the solution to the optimization problem is,

$$\begin{aligned} p(\tilde{x}|x) &= \frac{p(\tilde{x})}{Z(x, \beta)} \exp\left(-\beta \sum_y p(y|x) \log \frac{p(y|x)}{p(y|\tilde{x})}\right), \\ p(y|\tilde{x}) &= \frac{1}{p(\tilde{x})} \sum_x p(y|x)p(\tilde{x}|x)p(x). \end{aligned} \tag{4.35}$$

Because of the multiple appearances of $p(\tilde{x}, x)$, the solution for $p(\tilde{x}, x)$ and $p(\tilde{x})$ must be determined self-consistently.

Notice that this solution (4.35) is equivalent to,

$$\begin{aligned} p(\tilde{x}|x) &= \frac{p(\tilde{x})}{Z(x, \beta)} \exp(-\beta D_{KL}(p(y|x)||p(y|\tilde{x}))), \\ Z(x, \beta) &= \sum_{\tilde{x}} p(\tilde{x}) \exp(-\beta D_{KL}(p(y|x)||p(y|\tilde{x}))). \end{aligned} \tag{4.36}$$

So the KL divergence appears as the correct distortion measure for this problem of optimizing the information bottleneck.

The connection to our problem should be clear from the above discussion. With the stochastic association of blocks of X into elements of $\tilde{X}$, we here see another way to understand the renormalization group procedure. The X represents the UV variables. The $\tilde{X}$ represents the renormalized variables, e.g.,

in the Kadanoff procedure they are the block-spins that summarize the information about the spins in a block. The Y is then the variables after performing a rescaling, e.g., of the lattice in the Kadanoff procedure.

The KL divergence that we have been calculating can be understood as an effective distortion measure of passing information from the UV variables to the IR variables via renormalized variables: in particular, it can be understood as an effective distortion measure of passing information from string physics to SM physics as the scale is changed. The distribution $p(y|x)$ is the final distribution after renormalization and rescaling; $p(x)$ is the original UV distribution, giving $p(x,y) = p(y|x)p(x)$; $p(\tilde{x}|x)$ is the kernel T that transforms from the UV variables to the blocked variables, which allows us by (4.35) to find $p(y|\tilde{x})$ by finding $p(y,\tilde{x})$ and then marginalizing over y to get $p(\tilde{x})$. However, it is not clear how to find $p(y,\tilde{x})$, and it is not clear how to easily invert (4.35) to find $p(y|\tilde{x})$ directly. Due to these difficulties, we continue on with our procedure of using the KL divergence to directly quantify the information.

4.5 Quantum Relative Entropy

All of our results in discussing quantifying information have so far been for Euclidean QFTs. However, they are the same results for the quantum theory with commuting density matrices (i.e., there needs to be a basis in which both density matrices are diagonal), so long as the states have non-zero overlap and so long as there is some suitable notion of taking a $\log$ of a pure state. In such cases, the quantum relative entropy is exactly the KL divergence. See [140] for a review of basic quantum information theory.

Consider the representation of a density matrix for a theory at finite temperature,

$$\begin{aligned}\rho &= \sum_E \frac{e^{-\beta H}}{Z}|E\rangle\langle E| = \sum_E \int \mathrm{d}\phi \mathrm{d}\phi' \ \frac{e^{-\beta H}}{Z}|\phi\rangle\langle\phi||E\rangle\langle E||\phi'\rangle\langle\phi'| \\ &= \int \mathrm{d}\phi \mathrm{d}\phi' \ \frac{e^{-\beta H[\phi]}}{Z}|\phi\rangle\langle\phi||\phi'\rangle\langle\phi'| \\ &= \int \mathrm{d}\phi \ \frac{e^{-\beta H[\phi]}}{Z}|\phi\rangle\langle\phi|.\end{aligned} \tag{4.37}$$

We now find the relative entropy of density matrices ρ and σ. If there is a basis in which both of the density matrices are diagonal, the relative entropy can be evaluated as,

$$\begin{aligned} D_{KL}(\rho||\sigma) &= Tr(\rho \log \rho - \rho \log \sigma) \\ &= \int \mathrm{d}\phi \, \frac{e^{-\beta H_\rho[\phi]}}{Z_\rho} \log \frac{e^{-\beta H_\rho[\phi]}/Z_\rho}{e^{-\beta H_\sigma[\phi]}/Z_\sigma}, \end{aligned} \tag{4.38}$$

which is the earlier expression for the KL divergence. As is typical, diagonal density matrices reduce the quantum behavior to the classical: the relative entropy to the KL divergence. Also, note that the $\beta \to \infty$ limit computes the relative entropy between the ground states of the Hamiltonians. Because in this case, the contribution of other states is much smaller, the density matrices will automatically be diagonal. However, the density matrices will be pure states in this zero temperature limit (unless the ground state is degenerate), so care needs to be taken in using the KL divergence to evaluate the quantum relative entropy.

We can measure UV/IR mixing in a similar way to our method for classical distributions. Take a density matrix ρ and trace out some subset to get a density matrix ρ_A. Tracing out the complementary subset produces ρ_B. A joint distribution can then be formed $\rho_A \otimes \rho_B$ and compared to the original ρ using the quantum relative entropy. If the diagonal terms are small, we could approximate the matrices as diagonal, in which case the KL divergence would be measuring the quantum effects in the system.

The approximation of the matrices as diagonal can be understood in a MERA setup [141–143]. After producing a new Hamiltonian from tracing out the density matrix ρ, perform a disentangling step. Use a disentangler on the Hamiltonian to make it is as diagonal as possible in the original basis. Then a joint distribution can be formed and the process can repeated. The relative entropy in this case will then measure the complexity of the MERA circuit.

Having discussed the information theoretic interpretation of the RG flow and the methods that can be used to quantify information about UV physics in IR theories, we will look at a particular RG transformation that we will use in our example calculations: decimation.

CHAPTER 5
Decimation

We describe the decimation procedure, which is a specific and simple choice of Kadanoff block-spin renormalization (as we will soon explain). We will be making extensive use of the decimation procedure in our later calculations.

With regards to measuring the proximity between theories and quantifying information using the KL divergence, there are two distinct decimation procedures: one is to decimate, followed by a rescaling of the lattice, and the other is to create a joint distribution from the decimated theories. We describe the latter in this section and will show at the end Section 5.1 how these two procedures for decimation are distinct.

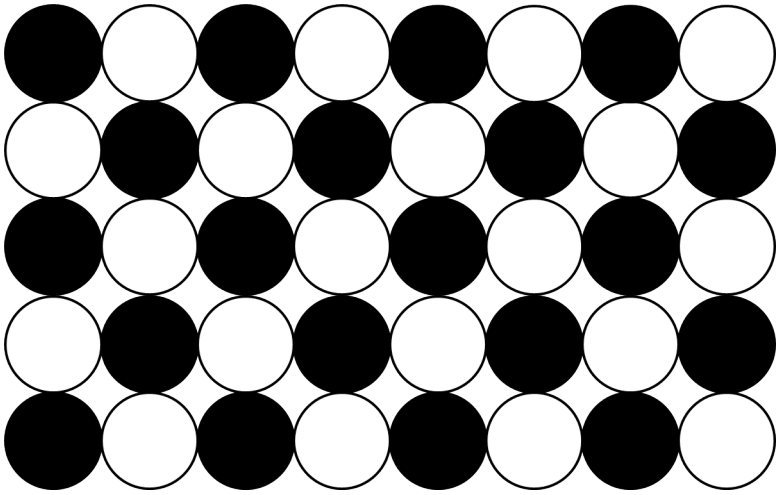

Figure 5.1: A 2D lattice of with sites labeled as black and white as a preparation to decimating the lattice.

Suppose we take a general lattice QFT and decimate it, where half the sites are labeled as "even" and half the sites are labeled as "odd." As an example in 2D with black and white sites playing the role of even and odd sites, see Figure 5.1. Creating a joint distribution by multiplying the decimated probability distributions, this decimated theory can be compared to the original, undecimated theory because the joint distribution is supported on the same space as the original, undecimated theory.

Let p be the distribution of the original theory and let $q = p_{odd}p_{even}$ be the joint distribution resulting from multiplying the decimated theory, where $p_{odd/even}$ is the theory where the even/odd sites have been summed over,

$$p_{even} \equiv \int \mathrm{D}\mu_{odd}\ p. \tag{5.1}$$

In 2D, the lattice sites are rotated relative to the original lattice by 45 degrees after a decimation step, and the nearest neighbor sites are now a distance of $\sqrt{2}$ larger than previously, instead of the space between sites doubling (as in 1D models Figure 5.4). A second decimation step will rotate the lattice again back to its original orientation, and the lattice spacing will now be doubled from the original spacing. See Figure 5.2 to see how the sites are rotated after decimation.

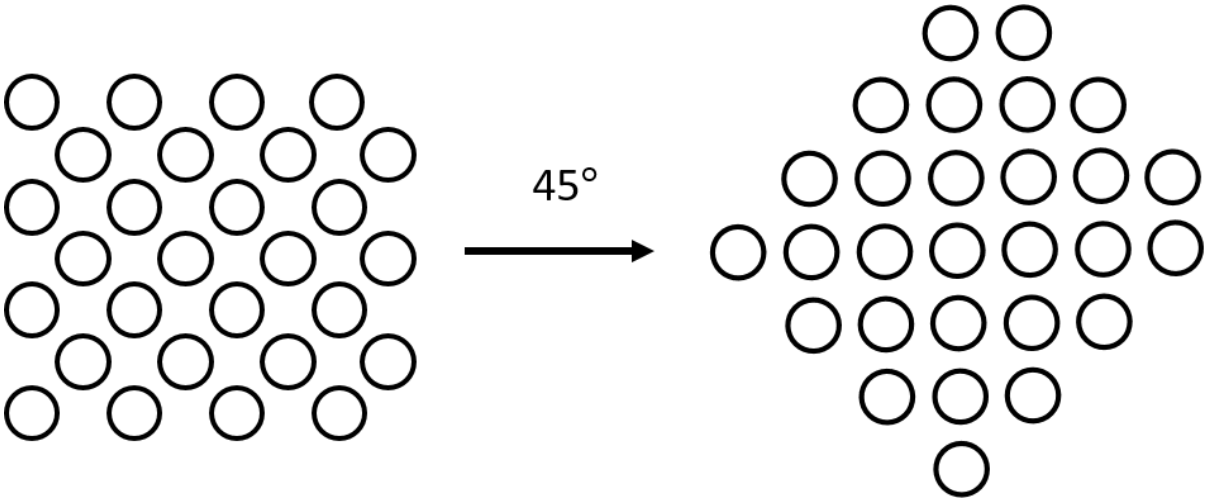

Figure 5.2: A 2D lattice of white sites after decimating the black sites. The lattice is rotated by 45 degrees relative to its original orientation, as can be seen by simply rotating it by 45 degrees, and the nearest neighbors lie along the diagonals of the squares of the original lattice, resulting in a lattice spacing of $\sqrt{2}$ times the original lattice spacing.

In terms of the Kadanoff block-spin procedure, the transformation $T[\sigma'|\sigma]$ for the decimation procedure is simply the delta function $\delta_{\sigma',\sigma}$, where σ' is an element of the sublattice of operators that we wish to remain after decimation, and σ is an operator in the whole, original lattice. In summing over σ, we sum over all the operators in the lattice, so this choice of T ensures that we sum over one sublattice of operators (e.g., the odd sites) while leaving the remaining operators σ' (e.g., the even sites) untouched. The blocks for the decimation procedure contain the two nearest-neighbor operators (one operator is on an even site, the other operator on an odd site) in each block; each of these blocks of two operators has one of these operators removed from the block while the other remains untouched.

For concreteness, let us take p to be a Boltzmann distribution with Hamiltonian H. We will absorb the inverse temperature β into the coupling constants of the Hamiltonian, so the Boltzmann factor becomes e^{-H}. The decimated Hamiltonians are defined by,

$$\begin{aligned} e^{-H_{even}} &\equiv \int \mathrm{D}\mu_{odd}\ e^{-H}, \\ e^{-H_{odd}} &\equiv \int \mathrm{D}\mu_{even}\ e^{-H}, \end{aligned} \tag{5.2}$$

where the new Hamiltonians $H_{even/odd}$ are of the same form as H up to an operator independent constant. We need this to be the case in order to iterate the decimation procedure, but as is well known from particular examples (such as the 2D Ising Model, Section 8.2), the Hamiltonian in general does not retain the same form (up to an operator independent constant) during the decimation procedure.

Letting Z be the partition function for theory p, notice that the partition function after decimation remains the same,

$$\begin{aligned} Z_{even} &= \int \mathrm{D}\mu_{even}\ e^{-H_{even}} \\ &= \int \mathrm{D}\mu_{even} \left(\int \mathrm{D}\mu_{odd}\ e^{-H} \right) \\ &= Z. \end{aligned} \tag{5.3}$$

As noted in the earlier section about the Kadanoff procedure (Section 3.1), we could choose to let the partition function change by an operator independent constant, while the Hamiltonian remains of the same form, or we can keep the partition function the same, while letting the Hamiltonian remain of the same form but shifted by an operator independent constant (i.e., let the Hamiltonian's zero of energy shift). We have chosen to keep the partition function the same.

We will now calculate the KL divergence for theory p and q where,

$$\begin{aligned} q = p_{even} p_{odd} &= (e^{-H_{even}}/Z_{even})(e^{-H_{odd}}/Z_{odd}) \\ &\equiv e^{-H_q}/Z_q. \end{aligned} \tag{5.4}$$

Note that,

$$\begin{aligned} Z_q &= \int \mathrm{D}\mu \; e^{-H_q} \\ &= \int \mathrm{D}\mu_{even}\mathrm{D}\mu_{odd} \; e^{-H_{even}} e^{-H_{odd}} \\ &= \left(\int \mathrm{D}\mu_{even} e^{-H_{even}}\right)\left(\int \mathrm{D}\mu_{odd} \; e^{-H_{odd}}\right) \\ &= Z_{even} Z_{odd} = Z^2. \end{aligned} \tag{5.5}$$

The KL divergence is then,

$$\begin{aligned} D(p||q) &= \int \mathrm{D}\mu \; p \log\left(\frac{p}{q}\right) = \int \mathrm{D}\mu \; p \log\left(\frac{p}{p_{odd} p_{even}}\right) \\ &= \int \mathrm{D}\mu \; p \left(\log p - \log p_{odd} - \log p_{even}\right) \\ &= \int \mathrm{D}\mu \; p \log p - \int \mathrm{D}\mu_{odd} \; p_{odd} \log p_{odd} - \int \mathrm{D}\mu_{even} \; p_{even} \log p_{even}, \end{aligned} \tag{5.6}$$

where we have p go to $p_{odd/even}$ because we can sum over the sites not found in $p_{odd/even}$ since the respective logarithms do not depend on those sites. This results in p going to $p_{odd/even}$ after this summation. In other words, $\langle\phi\rangle_{odd/even} = \langle\phi\rangle$, where we define expectation values for an operator ϕ as,

$$\begin{aligned} \langle\phi\rangle &= \int \mathrm{D}\mu \; p \, \phi, \\ \langle\phi\rangle_{odd/even} &= \int \mathrm{D}\mu_{odd/even} \; p_{odd/even} \, \phi. \end{aligned} \tag{5.7}$$

While we do not in principle need translation invariance to go through this decimation procedure and calculate the KL divergence, for the purposes of our computations, we will look at the case that our QFT is translation invariant. In that case, we get for the KL divergence,

$$D(p||q)_{tr} = \int \mathrm{D}\mu \;\; \left(p \log p - 2 p_{odd} \log p_{odd}\right). \tag{5.8}$$

Putting in the Boltzmann distributions for p and q, we get,

$$\begin{aligned} D(p||q) &= \int \mathrm{D}\mu\ p \log \frac{e^{-H} Z_{odd} Z_{even}}{Z e^{-H_{odd}} e^{-H_{even}}} = \int \mathrm{D}\mu\ p \log \frac{e^{-H} Z}{e^{-H_{odd}} e^{-H_{even}}} \\ &= \log Z + (-\langle H \rangle + \langle H_{odd} \rangle_{odd} + \langle H_{even} \rangle_{even}) \\ &= \log Z + (-\langle H \rangle + \langle H_{odd} \rangle + \langle H_{even} \rangle) \\ D(p||q)_{tr} &= \log Z + (-\langle H \rangle + 2\langle H_{odd} \rangle) \\ &= \log Z + (-E(\beta) + 2(E(\beta_{odd}) - C)) \end{aligned} \tag{5.9}$$

where C is the constant energy shift to the Hamiltonian after decimation, and for any coupling β and no constant energy shift in the Hamiltonian $E(\beta) \equiv \langle H(\beta) \rangle$ is the internal energy as a function of the coupling. Hence, in our case $E(\beta_{odd}) = \langle H_{odd} \rangle - C$.

Because we have made the assumptions needed to iterate the decimation procedure, we can generalize the above equation (5.9) to any number of decimation steps. That is, we can also calculate the relative entropy between the original theory and a theory decimated k times, forming the latter by a joint probability distribution from all possible decimations as was done in the above.

To do this, we define H_k to be the resulting Hamiltonian after decimating the original Hamiltonian k times. We define $Z_k = Z$ to be the partition function after decimating k times and g_i^k to be the coupling constants after decimating k times. We let C_k be the constant shift to the Hamiltonian that occurs with the kth decimation step and $C = \sum_{i=1}^{k} C_i$ becomes the total accumulated energy shift after k decimation steps. In the case where all the coupling constants are the same (and so can be pulled out of the Hamiltonian), we define $\beta_k = g_i^k$ to be the coupling constant at the kth decimation step and $\bar{H}_k$ as $H_k = \beta_k \bar{H}_k - C$.

We can then define p_k as the distribution for theory p decimated k times, and we can define the joint probability distribution as $q(k) = \prod_{j=1}^{2^k} p_j$ (each p_j is a copy of the theory that is decimated k times; there are 2^k such copies at the kth decimation step). We denote the expectation value with respect to the theory decimated k times as $\langle \rangle_k$.

If a method (such as a recursion relation) exists for finding the coupling constants at each decimation step, we can then find for the relative entropy between the original theory and a theory decimated k times

(assuming translation invariance),

$$
\begin{aligned}
D(p||q(k)) &= \int \mathrm{D}\mu \; p \log\left(\frac{p}{q(k)}\right) = \int \mathrm{D}\mu \; p \log\left(\frac{p}{\prod_{j=1}^{2^k} p_j}\right) \\
&= \int \mathrm{D}\mu \; p \left(\log p - \sum_{j=1}^{2^k} \log p_j\right) \\
&= \int \mathrm{D}\mu \; \left(p \log p - 2^k p_k \log p_k\right) \\
&= \int \mathrm{D}\mu \; p \log\left(\frac{e^{-H}\prod_{j=1}^{2^k} Z_j}{\prod_{j=1}^{2^k} e^{-H_j} Z}\right) \\
&= \log Z^{2^k-1} + \int \mathrm{D}\mu \; \left(p \log(e^{-H}) - p \log(e^{-2^k H_k})\right) \\
&= (2^k - 1)\log Z + \left(-\langle H\rangle + 2^k \langle H_k\rangle\right) \\
&= (2^k - 1)\log Z + \left(-E(\beta) + 2^k (E(\beta_k) - C)\right).
\end{aligned}
\tag{5.10}
$$

Following the same reasoning as above, we can calculate the relative entropy between a theory decimated k times and the same theory decimated l times (the following result does not depend on whether k or l is greater). The joint distributions are over the same space as the original, undecimated theory. The result is (assuming translation invariance),

$$
\begin{aligned}
D(q(l)||q(k)) &= \int \mathrm{D}\mu \; \left(2^l p_l \log p_l - 2^k p_k \log p_k\right) \\
&= (2^k - 2^l)\log Z + \left(-2^l \langle H_l\rangle + 2^k \langle H_k\rangle\right) \\
&= (2^k - 2^l)\log Z + \left(-2^l \left(E(\beta_l) - \sum_{i=1}^{l} C_i\right) + 2^k \left(E(\beta_k) - \sum_{i=1}^{k} C_i\right)\right)
\end{aligned}
\tag{5.11}
$$

which reduces to the earlier equation when $l = 0$, as expected.

Finally, if we wish to calculate the KL density then, we divide by N, where N is the total, original number of lattice sites (because multiplying the joint distributions makes theory q have N sites, just like theory p has N sites).

5.1 Different Decimation Procedures

Notice that this version of decimation we have described is distinct from the usual procedure where a theory is decimated and the lattice is rescaled. Consider a 1D chain of black and white sites with nearest

neighbor couplings, as in Figure 5.3. We decimate by summing over (and hence removing) the black sites or the white sites. Decimating once and rescaling corresponds to Figure 5.4, while decimating once and making a joint distribution of the black and white sites corresponds to Figure 5.5.

Different lattice sites are coupled together in these two different schemes, showing that they are distinct procedures. Notably, the joint distribution method no longer has nearest neighbor interactions, while the decimation-plus-rescaling continues to have them.

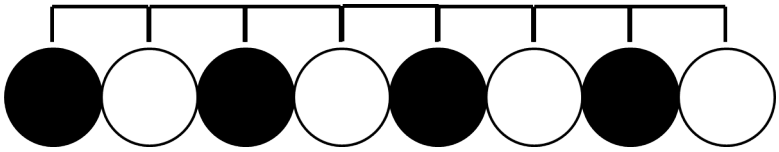

Figure 5.3: A 1D chain of sites labeled as black and white as preparation for decimating the lattice. The lines show which sites are coupled to each other (nearest neighbors).

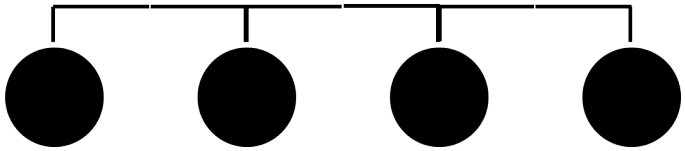

Figure 5.4: Result of decimation of white sites of 1D lattice before rescaling. Rescaling just changes the spacing between the sites to the original lattice spacing.

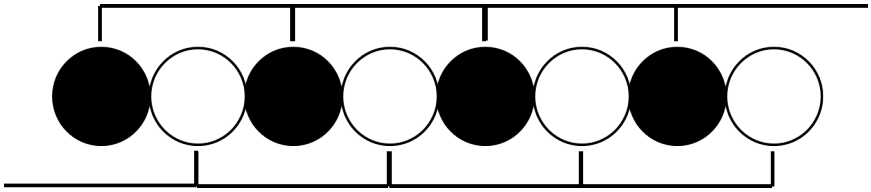

Figure 5.5: Result of decimating white sites of 1D lattice, then black sites, and forming a joint distribution.

5.2 Thermodynamic Interpretation

The relative entropy we have calculated can be written in terms of thermodynamic quantities (using F for the free energy and S for entropy of the whole system, and we will now separate out the inverse temperature β from the coupling to make the equations clearer). Taking the translation invariant case for

the entropy between a theory and the theory decimated k times results in,

$$\begin{aligned} D(p||q(k)) &= (2^k - 1)\log Z + \beta\left(-E(J) + 2^k(E(J_k) + C)\right) \\ &= -(2^k - 1)F/T + \left(-E(J)/T + 2^k(E(J_k)/T + C/T)\right), \end{aligned} \tag{5.12}$$

where we have put the KL divergence in terms of E and F (and also the constant shift C). Alternatively, we could, of course, put the KL divergence in terms of F and S as,

$$D(p||q(k)) = -(2^k - 1)F/T + \left(-F(J)/T + S(J) + 2^k(F(J_k)/T + S(J_k) + C/T)\right), \tag{5.13}$$

which we have arrived at by using standard thermodynamic identities.

CHAPTER 6
Continuum

Having investigated the RG flow with lattice theories in general, we will now investigate the RG flow for continuum theories and use the KL divergence to measure the UV-IR mixing. We expect the RG flow of the continuum theory to match both of the previously-used decimation procedures in the continuum limit, and we will show later in this section that this is in fact the case.

Conceptually, it is straightforward to see how the decimation plus rescaling the lattice spacing ought to correspond to the continuum renormalization plus rescaling the momentum modes of the Wilson procedure. As for the other decimation procedure of forming a joint distribution after decimation, suppose we have a continuous probability distribution that takes values p_i at i sites. Perform the RG flow by averaging nearest neighbors to produce a new probability distribution $\hat{p}_i = (p_i + p_{i+1})/2$. This new probability distribution is equivalent to the lattice case of averaging the nearest neighbors of the even and odd sites separately and then forming a joint distribution $p_{odd}p_{even}$. Hence, we expect the continuum renormalization group flow to also match to the procedure of decimating and forming a joint distribution.

Before we discuss the continuum case with the KL divergence, we first discuss some useful properties of conformal field theories that we will make use of in our continuum calculations. We then will make our calculations for the continuum case in Section 6.2 and run into problems with contact terms. We then resolve the problems with the contact terms by defining the UV completion for the models of interest in Section 6.3. In the last sections, we show how the lattice results match the continuum results that we find in Section 6.2.

6.1 CFTs

In our description of conformal field theories (CFTs), we follow the notation of [144]. CFTs are quantum field theories that are the same at all scales (invariant under a scale transformation) and are thereby also UV complete. In the context of renormalization flows, CFTs play a special role: as a result of their scale invariance, they are the fixed points of RG flows. CFTs also have relatively easy correlation

functions to compute: their one-point functions vanish, and their two-point functions take a simple form. For scalar primary operators of dimension Δ, the two-point functions go as,

$$\langle O(x)O(y)\rangle \sim \frac{1}{(x-y)^{2\Delta}}, \tag{6.1}$$

where Δ is the scaling dimension of the field $O(x)$. For two dimensions, the fundamental scalar field (called the "primary" field) is $\partial\phi$, while the fundamental fermion field remains ψ. In two dimensions, the propagators for the free fields go as,

$$\begin{aligned} \langle \phi(x)\phi(y)\rangle &\sim -\log(x-y)^2 \\ \langle \partial_x\phi(x)\partial_y\phi(y)\rangle &\sim -\frac{2}{(x-y)^2} \\ \langle \bar{\psi}(x)\psi(y)\rangle &\sim \frac{1}{(x-y)}. \end{aligned} \tag{6.2}$$

In two dimensions, the fields are often rewritten in terms of complex coordinates that separate into holomorphic and antiholomporhic parts. Aside from replacing x and y in our above expressions by the holmorphic coordinates z and w, the only propagator that changes is the logarithm, which becomes $\log(z-w)$. Also, the fermion propagator is generated by $\psi(z)\psi(w)$ in holomorphic coordinates. The antiholomorphic propagators are the same as the holomorphic but with antiholomorphic coordinates.

For a general $D \neq 2$ dimensions and for when Δ is the physical dimension of the field, we have for the propagators,

$$\begin{aligned} \langle \phi(x)\phi(y)\rangle &\sim \frac{1}{(x-y)^{D-2}} \\ \langle \psi(x)\psi(y)\rangle &\sim \frac{1}{(x-y)^{D-1}}. \end{aligned} \tag{6.3}$$

Because of these useful properties of CFTs, they will especially be useful in our continuum calculations that follow.

6.2 Continuum RG Result

In this section, we show the result of the KL divergence when making use of a general continuum renormalization procedure for a CFT (so the one point functions vanish and the propagator is simple). We will follow a Wilsonian procedure to perform the renormalization group transformation. First, we will integrate out momentum shells. In integrating out momentum shells, the coupling constants for the fields

will change, and we can find an RG flow for the trajectory of the couplings, as discussed in an earlier section. Next, we will rescale the momenta to put the whole theory on the same space as the original theory.

In this renormalization procedure, the partition function remains the same (as discussed in earlier sections) when integrating out momentum shells, so it is the same at all scales. This can be seen because we just need to perform the remaining integral over the lower energy momentum modes after performing the integral over part of the higher modes. This statement can be captured in a Callan-Symanzik equation for the partition function Z that depends on couplings g_i and scale Λ,

$$\frac{dZ(g)}{d\log\Lambda} = \left(\frac{\partial}{\partial\log\Lambda}\,|_{g_i} + \frac{\partial g_i(\Lambda)}{\partial\log\Lambda}\frac{\partial}{\partial g_i}\,|_{\Lambda}\right) Z_\Lambda(g) = 0, \tag{6.4}$$

which we showed earlier on. As also noted earlier, an operator-independent constant appears in the action after performing the RG transformation. This constant can either be absorbed into the partition function to keep the partition function fixed (as the Callan-Symanzik equation above assumes), or it can be put in the action, which means the partition function changes with scale. For our following calculation, we will allow the partition function to change and place the constant in the action.

Now take a theory at scale Λ and integrate out momentum shells to produce a theory at a lower but infinitesimally close scale $\Lambda' = \Lambda - \delta\Lambda$. A rescaling of the momenta will then allow the distributions to have the same support in momentum space, and it (combined with the RG flow) also makes the partition functions for the two theories have different values, since the new partition function is no longer related to the old one by simply integrating out momentum shells.

Let $Z_{\Lambda'}, S_{\Lambda'}$ be the resulting partition function and action after the RG transformation. For convenience, we define $\hat{g}(\Lambda) = g(\Lambda)\mathcal{Z}(\Lambda)$, where the g is the coupling as before and $\mathcal{Z}$ is the wavefunction renormalization. The difference in the actions for general dimension D is,

$$\begin{aligned}\langle S_{\Lambda'} - S_\Lambda\rangle_\Lambda &= \int \mathrm{d}^D x\ \Lambda^{D-\Delta}(\hat{g}(\Lambda') - \hat{g}(\Lambda))O(x) \\ &\approx -\delta\Lambda\Lambda^{D-\Delta}\frac{\partial\hat{g}(\Lambda)}{\partial\Lambda}\int \mathrm{d}^D x\ O(x)\end{aligned} \tag{6.5}$$

We then produce for the KL divergence between these theories, using (4.11),

$$\begin{aligned} D(p||q) &\approx \frac{(\hat{\beta}\delta \log \Lambda)^2}{2} \int \mathrm{d}^D x \mathrm{d}^D y \ \langle O(x)O(y)\rangle_\Lambda \\ &= \frac{(\hat{\beta}\delta \log \Lambda)^2}{2} Vol(M_D)\Omega(S_D) \int \mathrm{d}z \ z^{D-1-2\Delta}, \end{aligned} \tag{6.6}$$

where we have expanded the logarithm in the KL divergence and made use of the CFT properties to evaluate the one-point function (vanishes) and the two-point function that remains. We see that we are left here with an integrated two point function that becomes a contact term. To evaluate further, we make use of dimension regularization to get,

$$\begin{aligned} D(p||q) &= \lim_{\varepsilon \to 0} \frac{(\hat{\beta}\delta \log \Lambda)^2}{2} Vol(M_D)\Omega(S_D)\mu^\varepsilon \int \mathrm{d}z \ z^{D-\varepsilon-1-2\Delta} \\ &= \lim_{\varepsilon \to 0} \frac{(\hat{\beta}\delta \log \Lambda)^2}{2} Vol(M_D)\Omega(S_D)\mu^\varepsilon \Gamma(D-\varepsilon-2\Delta)\Gamma(-D+\varepsilon+2\Delta), \end{aligned} \tag{6.7}$$

where we used the following form of the beta function (the beta function related to products of the Γ functions; not the derivative of the coupling constants with respect to scale) to do the integral,

$$\beta(x,y) \equiv \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \int_0^\infty \mathrm{d}s \ \frac{s^{x-1}}{(1+s)^{x+y}}, \tag{6.8}$$

where $x = -y$ in our case.

We now take the $\varepsilon \to 0$ limit to find that the factors from the integral become,

$$\mu^\varepsilon \Gamma(D-\varepsilon-2\Delta)\Gamma(-D+\varepsilon+2\Delta) = \Gamma(D-2\Delta)\Gamma(2\Delta-D) \tag{6.9}$$

This result gives an infinite and scheme dependent answer to our KL divergence calculation for when $D-2\Delta$ is an integer. There is also both a finite scheme independent part and a finite scheme dependent part: the finite scheme independent part is negative. For when $D-2\Delta$ is a half-integer, the answer is finite and scheme independent.

To arrive at this answer, we had to use a regularization scheme because of the contact term in (6.6), which then results in a generally scheme dependent answer. The reason a contact term appears is due to a lack of clear definition of the UV theory. In our next section, we will therefore work to find a UV completion for calculations to measure UV-IR mixing with the KL divergence.

6.3 Fast and Slow Modes

As already noted, due to the problem of contact terms appearing in the integrated two-point function that appears at the lowest order term of the KL divergence, our final result is currently scheme dependent. The contact terms happened because of a lack of definition of the UV theory. We will again make use of integrating out momentum shells, following Skinner's presentation [24] of splitting the fields into fast (UV) and slow (IR) modes, in order to define what is meant by a UV completion of a model.

We split our fields into fast (χ) and slow modes (ϕ) as,

$$\begin{aligned}\Phi(x) &= \int_{p>\Lambda} \mathrm{d}^{\mathrm{D}}p\ \Phi(p)e^{ipx} + \int_{p<\Lambda} \mathrm{d}^{\mathrm{D}}p\ \Phi(p)e^{ipx} \\ &\equiv \chi(x) + \phi(x).\end{aligned} \tag{6.10}$$

We can then integrate out the χ fields to produce the IR theory, which has a finite two-point function. We will expand about $\phi = 0$ in order to produce a low energy approximation to the action of the IR and UV fields mixed together. Making use of the following definitions for the free part of the actions,

$$\begin{aligned}S_0[\phi] &\equiv \int \mathrm{d}^{\mathrm{D}}x\ \frac{1}{2}(\partial\phi)^2 + \frac{1}{2}m_\phi^2\phi^2 \\ S_0[\chi] &\equiv \int \mathrm{d}^{\mathrm{D}}x\ \frac{1}{2}(\partial\chi)^2 + \frac{1}{2}m_\chi^2\chi^2,\end{aligned} \tag{6.11}$$

the full action becomes,

$$\begin{aligned}S &= S_0[\phi] + S_0[\chi] + S[\phi,\chi], \\ &= S_0[\phi] + S_0[\chi] + \int \mathrm{d}^{\mathrm{D}}x\ L[0,\chi] + \frac{\partial L[\phi,\chi]}{\partial\phi}|_{\phi=0}\,\phi + \frac{1}{2!}\phi\frac{\partial^2 L[\phi,\chi]}{\partial\phi^2}|_{\phi=0}\,\phi + ...,\end{aligned} \tag{6.12}$$

where we took the lowest order piece of the higher order mixed action $S[\phi,\chi]$. We now integrate out the χ fields to produce an effective action,

$$\begin{aligned}S_{eff} &= -\log\int \mathrm{D}\chi\ \exp(-S_0[\phi]) \\ &\quad\times \exp(-S_0[\chi] - \int \mathrm{d}^{\mathrm{D}}x\ L[0,\chi] + \frac{\partial L[\phi,\chi]}{\partial\phi}|_{\phi=0}\,\phi + \frac{1}{2!}\phi\frac{\partial^2 L[\phi,\chi]}{\partial\phi^2}|_{\phi=0}\,\phi + ...) \\ &\equiv S_0[\phi] + \int \mathrm{d}^{\mathrm{D}}x\ J(x)\phi(x) + \int \mathrm{d}^{\mathrm{D}}x\ \phi(x)\frac{Q}{2}\phi(x) + ... \\ &\equiv S_0[\phi] + J\cdot\phi + \phi\cdot\frac{Q}{2}\cdot\phi,\end{aligned} \tag{6.13}$$

where we have defined various quantities in the last two lines to simplify the analysis. The $J(x)$ is the result of integrating out all the dependence on χ and represents the heavy mass of the χ fields to which the IR theory couples; the J can be understood as sourcing the ϕ fields. The Q is the result of integrating out the χ fields on the ϕ^2 term: this term must be included because it is consistent with the symmetries of the Lagrangian, but note that it will produce logarithmic divergences. We will show how the source terms and UV logarithmic divergence appears in a scalar ϕ^4 theory in Section 10.1. Also note that the Q may depend on higher derivative interactions, but we may neglect interaction terms higher than two because we are working with an effective field theory and those operators are irrelevant. The S_0 and L_0 refer to the free action and Lagrangian.

We can apply this expansion to the KL divergence so as to measure the UV information in the IR theory in two ways. First, one can take the KL divergence between theories with two different UV completions. The actions of interest are,

$$\begin{aligned} S_p &= S_0[\phi] + J_p \cdot \phi + \phi \cdot \frac{Q_p}{2} \cdot \phi, \\ S_q &= S_0[\phi] + J_q \cdot \phi + \phi \cdot \frac{Q_q}{2} \cdot \phi, \end{aligned} \tag{6.14}$$

where we assume $J_p - J_q \equiv \Delta J$ and $Q_p - Q_q \equiv \Delta Q$ are small so that the KL divergence can be calculated to lowest order. Using the earlier results for the KL divergence, the result to lowest order is,

$$D_{KL}(p||q) = \frac{1}{8} \int \mathrm{d}^{\mathrm{D}}x \mathrm{d}^{\mathrm{D}}y \ \langle \Delta J(x) \Delta J(y) \Delta Q \Delta Q \phi(x) \phi(y) \rangle_p, \tag{6.15}$$

where theory p could be redefined as free by moving the perturbation to the q action.

The second way to apply this expansion to the KL divergence is to compare the following two actions,

$$\begin{aligned} S_p &= S_0[\phi] + S_0[\chi] + S[0, \chi], \\ S_q &= S_0[\phi] + S_0[\chi] + S[0, \chi] + \frac{\delta S[0, \chi]}{\delta \phi} \cdot \phi + \frac{1}{2} \phi \cdot \frac{\delta^2 S[0, \chi]}{\delta \phi^2} \cdot \phi, \end{aligned} \tag{6.16}$$

which directly measures the UV/IR mixing between the theories. We used a free field action for ϕ in order to make the calculations easier, but in principle, interacting ϕ terms could be added.

We now find the beta function for the KL divergence to lowest order. To do this, simply take the derivative of the expression we found earlier (4.16) to produce,

$$\frac{\partial}{\partial \log \Lambda} D(p||q) = \kappa^2 \left(\langle (S_p - S_q) \frac{\partial (S_p - S_q)}{\partial \log \Lambda} \rangle_p - \langle (S_p - S_q) \rangle_p \langle \frac{\partial (S_p - S_q)}{\partial \log \Lambda} \rangle_p \right), \tag{6.17}$$

where the derivative of the exponential is of higher order $O(\kappa^3)$ because it produces another κ. To measure UV/IR mixing, we use,

$$\begin{aligned} S_p &= S_0[\phi] + S_0[\chi] \\ S_q &= S_0[\phi] + S_0[\chi] + S_{int}[\phi, \chi] \\ \frac{\partial}{\partial \log \Lambda} D(p||q) &= \kappa^2 \left(\langle S_{int} \frac{\partial S_{int}}{\partial \log \Lambda} \rangle_0 - \langle S_{int} \rangle_0 \langle \frac{\partial S_{int}}{\partial \log \Lambda} \rangle_0 \right), \end{aligned} \tag{6.18}$$

where the end result is to lowest order. We can now evaluate the derivatives of S_{int} using Polchinski's equation [24], which gives the RG equation for the effective action. To use Polchinski's equation, we define the effective action as,

$$e_{\Lambda}^{-S_{eff}[\phi]} \equiv \int \mathrm{D}\mu_{p>\Lambda} \; e^{-S_{\Lambda_0}[\phi+\chi]}, \tag{6.19}$$

where Λ_0 is a reference scale at which the original action is defined, Λ is the scale at which the effective action is defined, and the measure is to be understood as integrating over all momenta that are greater than Λ (i.e., integrating out the χ field). The effective action can be broken into free and interacting pieces. Denote the interacting piece as S_{Λ}^{int}. Polchinski's equation is,

$$\begin{aligned} -\frac{\partial S_{int}}{\partial \log \Lambda} &= \int \mathrm{d}^{\mathrm{D}} x \mathrm{d}^{\mathrm{D}} y \; \left(\frac{\delta S_{int}}{\delta \phi(x)} D_{\Lambda}(x,y) \frac{\delta S_{int}}{\delta \phi(y)} - D_{\Lambda}(x,y) \frac{\delta^2 S_{int}}{\delta \phi(x) \delta \phi(y)} \right), \\ D_{\Lambda}(x,y) &= \frac{1}{(2\pi)^D} \frac{\Lambda^{D-1} \delta \Lambda}{\Lambda^2 + m^2} \int_{S^{D-1}} \mathrm{d}\Omega \; e^{i\Lambda \hat{p} \cdot (x-y)}. \end{aligned} \tag{6.20}$$

The $D_{\Lambda}(x,y)$ is the infinitesimal χ propagator resulting from lowering the energy scale from Λ' to $\Lambda = \Lambda' - \delta\Lambda$. This equation, then, gives an expression for the derivatives of S_{int} in (6.18).

This procedure can also be done for the exact KL divergence on these two theories. It is,

$$\frac{\partial}{\partial \log \Lambda} D(p||q) = \kappa^2 \left(\frac{\partial \log Z_0}{\partial \log \Lambda} - \kappa \langle S_{int} \frac{\partial S_0}{\partial \log \Lambda} \rangle_0 + \langle \frac{\partial S_{int}}{\partial \log \Lambda} \rangle_0 \right), \tag{6.21}$$

where the derivative of Z_q vanishes because it is the full partition function and so is scale-invariant. To calculate the entire expression requires evaluating the derivative of the free part of the action S_0 with respect to scale $\partial S_0/\partial \log \Lambda$. However, we do not currently know what this derivative is.

6.4 Matching Lattice to Continuum: Part 1

We now show in this and the next section that the earlier lattice results match the continuum limit that we found. In both cases, we take a lattice theory p on a square lattice with constant interactions and decimate k times. We will assume that the operators are localized at points of the lattice. Gauge theories would correspond to links of the lattice, but there is hope that our procedure can be extended to lattice gauge theories because defining the decimation procedure just requires dividing a lattice into two sublattices.

As previously noted, there are two procedures for decimation: decimation plus rescaling or decimation plus multiplying marginalized distributions. We will here study decimation plus rescaling. We therefore associate with the rescaled lattice a distribution q with coupling constant $\lambda(k)$ and constant shift $c(k)$. Notice that the rescaling does not affect the coupling constant, so theory p could just as well be another decimated version of an original theory (say, l times) and q another decimated version (k times) of that same theory.

We now take a limit where we go one decimation step downwards, so that the coupling constants are close to each other. We will also make use of the vector $v \in L$, where L is the lattice, in order to label the operators at the lattice sites. We then get,

$$\begin{aligned} Z_q(k = l-1) &= \int \mathrm{D}\mu \; exp[-K(\lambda(k)\sum_v O_v + c(k))] \\ &= \int \mathrm{D}\mu \; exp[-K(\lambda(l)\sum_v O_v + c(l))] exp[K(\partial_k\lambda(l)\sum_v O_v + \partial_k c(l))] \\ &= Z_p(1 + K\langle \partial_k H_p\rangle_p + \frac{K^2}{2}\langle(\partial_k H_p)^2\rangle_p + ...) \end{aligned} \tag{6.22}$$

The resulting KL divergence is (to lowest order in K),

$$\begin{aligned} D(p||q) &= \log\left(\frac{Z_q}{Z_p}\right) + K\langle(H_q - H_p)\rangle_p \\ &\approx \log\left(1 + K\langle\partial_k H_p\rangle_p + \frac{K^2}{2}\langle(\partial_k H_p)^2\rangle_p\right) - K\langle\partial_k H_p\rangle_p \\ &\approx \frac{K^2}{2}\left(\langle(\partial_k H_p)^2\rangle_p - (\langle\partial_k H_p\rangle_p)^2\right). \end{aligned} \tag{6.23}$$

To show that this lattice result is the same as the continuum result, we need to rewrite our result for the KL divergence in terms of the lattice spacing at the kth decimation step, so that we can take a continuum limit by taking $a \to 0$. To rewrite the KL divergence, we note that $Vol/a(k)^D = N/2^k = Vol/(a^D 2^k)$, where D is the dimension of our model, Vol its volume, a the original lattice spacing, and $a(k)$ the lattice spacing at decimation step k. Hence, we have $a(k) = a2^{k/D}$.

We now use the chain rule to rewrite the derivatives in terms of the lattice spacing, resulting in,

$$\begin{aligned} \frac{\partial}{\partial k} &= \frac{\partial a}{\partial k}\frac{\partial}{\partial a} \\ &= a(k)\log 2^{1/D}\frac{\partial}{\partial a}. \end{aligned} \tag{6.24}$$

Rewriting the KL divergence that we just calculated (6.23) in terms of a, we get,

$$\begin{aligned} D(p||q) &= \frac{K^2}{2}\left(\langle(\log 2^{1/D} a\partial_a H_p)^2\rangle_p - (\langle\log 2^{1/D} a\partial_a H_p\rangle_p)^2\right) \\ &= \frac{(K\log 2^{1/D})^2}{2}\left(\langle(\partial_{\log(a)}\lambda(a)\sum_v O_v + \partial_{\log(a)}c(a))^2\rangle_p\right. \\ &\left. - (\langle\partial_{\log(a)}\lambda(a)\sum_v O_v + \partial_{\log(a)}c(a)\rangle_p)^2\right), \end{aligned} \tag{6.25}$$

where we have inserted the operators on the lattice for the Hamiltonians. We now recognize that $\partial_{\log(a)}\lambda(a) \equiv \beta(a)$ is in fact the beta function of the coupling constants $\beta(a)$. Using this and expanding

out the squares gives,

$$\begin{aligned} D(p||q) &= \frac{(K\log 2^{1/D})^2}{2}\left(\langle(\beta(a)\sum_v O_v + \partial_{\log(a)}c(a))^2\rangle_p - (\langle\beta(a)\sum_v O_v + \partial_{\log(a)}c(a)\rangle_p)^2\right) \\ &= \frac{(K\log 2^{1/D})^2}{2}(\langle(\beta(a)\sum_v O_v)^2 + 2\beta(a)\sum_v O_v\partial_{\log(a)}c(a)\rangle_p \\ &- (\langle\beta(a)\sum_v O_v\rangle_p)^2 - 2\langle\beta(a)\sum_v O_v\rangle_p\partial_{\log(a)}c(a)). \end{aligned} \tag{6.26}$$

Note that the operator-independent constants in the Hamiltonians cancel out. Cancelling them and simplifying the result gives,

$$\begin{aligned} D(p||q) &= \frac{(K\log 2^{1/D})^2}{2}\left(\langle(\beta(a)\sum_v O_v)^2\rangle_p - (\langle\beta(a)\sum_v O_v\rangle_p)^2\right) \\ &= \frac{(K\beta(a)\log 2^{1/D})^2}{2}\left(\langle(\sum_v O_v)^2\rangle_p - (\langle\sum_v O_v\rangle_p)^2\right). \end{aligned} \tag{6.27}$$

Notice that the $\beta(a)$ function for the coupling constants is in terms of the lattice spacing. Because the lattice spacing is the inverse of the energy, just change β to $-\beta$ to put the result in terms of the energy scale, instead of the lattice spacing. Of course, flipping the sign on the β does not change the answer because it is squared.

This result for the KL divergence matches the general answer derived in (4.16). Also, notice that this matches the continuum answer we derived earlier (6.6) (which used a CFT, so the one-point functions vanished) when the lattice spacing is taken to 0 or to some UV cutoff, with the log factor playing the role of some volume factor (or maybe it is a lattice artifact). Indeed, it would seem the $\log 2^{1/D}$ matches the $\delta\log\Lambda$ exactly, which we will now show.

We note that our new lattice spacing after one decimation step is $a(1) = a2^{1/D}$. The $\log 2^{1/D}$ factor describes the difference between the new lattice spacing and the old,

$$\log 2^{1/D} = \log\frac{a(1)}{a}. \tag{6.28}$$

This $\log 2^{1/D}$ factor is a fixed constant ratio between our new spacing after one decimation step and our old. The factor is always the same because the decimation procedure is discrete and increases the spacing by a constant factor after each decimation step. However, because we are interested in the continuum

limit, we can take our decimation step k to be a continuous quantity, as we have in fact assumed in our earlier derivation (where we are taking derivatives with respect to k and assume the new coupling constants to be infinitesimally near the old ones). Having a continuous RG procedure then allows the new spacing to be infinitesimally close to the original spacing, since the 2^k in the $a(k) = a2^{k/D}$ can now be made infinitesimally small. We can quantify the infinitesimal change in spacing as $a(k) = a + \delta a$. The result is,

$$\begin{aligned} \log 2^{1/D} &= \log \frac{a + \delta a}{a} = \log \left(1 + \frac{\delta a}{a}\right) \\ &\approx \frac{\delta a}{a} = \delta \log a, \end{aligned} \tag{6.29}$$

where we have expanded the logarithm and have taken only the first term because it is infinitesimally small as we take k continuous and $k \to 0$ to get the continuum RG limit. Placing this result into our expression for the KL divergence (6.27) then gives something that looks exactly like our earlier continuum result with cutoff Λ but instead in terms of the lattice spacing a. If we convert the lattice spacing to an energy scale, we will indeed produce (in the continuum limit) the $\delta \log \Lambda$ of our earlier continuum result,

$$\begin{aligned} \log 2^{1/D} &= \log \frac{a(1)}{a} = \log \frac{\Lambda}{\Lambda'} \\ &= \log \frac{\Lambda}{\Lambda - \delta\Lambda} = -\log \frac{\Lambda - \delta\Lambda}{\Lambda} \\ &= -\log \left(1 - \frac{\delta\Lambda}{\Lambda}\right) \approx \frac{\delta\Lambda}{\Lambda} = \delta \log \Lambda, \end{aligned} \tag{6.30}$$

where we have been able to expand the logarithm to lowest order because we are interested in the continuum RG limit where $\delta\Lambda$ is infinitesimally small compared to Λ.

6.5 Matching Lattice to Continuum: Part 2

Having observed the result with one lattice decimation procedure (decimation plus rescaling), we will now see the result when we use the decimation plus multiplication of marginalized distributions procedure. Let p be some theory and q be the theory resulting from decimating p and then forming a joint distribution. We will now find sublattices (e.g., black and white sites; or odd and even sites) in order to perform the decimation procedure and take b and w to be elements of these sublattices. The coupling constants should be close to each other in this case too, so following the procedure in the previous section,

to lowest order we have,

$$D(p||q) \approx \frac{K^2}{2}\left(\langle(\sum_v \lambda(k)O_v - \sum_w \lambda(l)O_w - \sum_b \lambda(l)O_b)^2\rangle_p - (\langle\sum_v \lambda(k)O_v - \sum_w \lambda(l)O_w - \sum_b \lambda(l)O_b\rangle_p)^2\right), \tag{6.31}$$

where we have plugged in our operators for our Hamiltonians. Notice that this time we have operators defined over sublattices b and w. We now take the coupling constants to be infinitesimally near each other, so we get,

$$\begin{aligned} D(p||q) &\approx \frac{K^2}{2}\left(\langle(\sum_v(\lambda(a) - \frac{\partial a}{\partial k}\partial_a\lambda(a))O_v - \sum_w \lambda(a)O_w - \sum_b \lambda(a)O_b)^2\rangle_p \right. \\ &\left. - (\langle\sum_v(\lambda(a) - \frac{\partial a}{\partial k}\partial_a\lambda(a))O_v - \sum_w \lambda(a)O_w - \sum_b \lambda(a)O_b\rangle_p)^2\right) \\ &= \frac{K^2}{2}\left(\langle(\sum_v(\lambda(a) - \beta(a)\log 2^{1/D})O_v - \sum_w \lambda(a)O_w - \sum_b \lambda(a)O_b)^2\rangle_p \right. \\ &\left. - (\langle\sum_v(\lambda(a) - \beta(a)\log 2^{1/D})O_v - \sum_w \lambda(a)O_w - \sum_b \lambda(a)O_b\rangle_p)^2\right), \end{aligned} \tag{6.32}$$

where we have substituted $\beta(a)$ for the derivatives of the coupling constants, as before. We now expand the squared expressions and simplify, as we did in the previous section. However, make use of $\sum_v O_v = \sum_b O_b + \sum_w O_w$ (which holds because the expression on the right hand side is just summing over all the operators on the lattice) in order to achieve the cancellations that we had when expanding the squared expressions last time. The result is,

$$\begin{aligned} D(p||q) &= \frac{K^2}{2}\left(\langle(\sum_v(\lambda(a) - \beta(a)\log 2^{1/D})O_v - \sum_v \lambda(a)O_v)^2\rangle_p \right. \\ &\left. - (\langle\sum_v(\lambda(a) - \beta(a)\log 2^{1/D})O_v - \sum_v \lambda(a)O_v\rangle_p)^2\right) \\ &= \frac{K^2}{2}\left(\langle(\sum_v \beta(a)\log 2^{1/D}O_v)^2\rangle_p - (\langle\sum_v \beta(a)\log 2^{1/D}O_v\rangle_p)^2\right) \\ &= \frac{(K\beta(a)\log 2^{1/D})^2}{2}\left(\langle(\sum_v O_v)^2\rangle_p - (\langle\sum_v O_v\rangle_p)^2\right). \end{aligned} \tag{6.33}$$

The final result here is exactly the same result for the decimation plus rescaling procedure! Therefore, both decimation procedures produce the same result in the continuum RG limit and match the continuum result that we calculated earlier.

CHAPTER 7

Warm-up Example: Ising Model Perturbations

We now will start making calculations of the KL divergence with example models: starting with decimation and lattice models, then continuum models. We want to perform decimation with an Ising model, but as a warm-up, we will first consider adding a perturbation to the Ising model with no magnetic field. For both 1D and 2D Ising models, the KL divergence is calculated between a theory and the same theory with a "perturbation" (it will turn out that our result is exact, so the "perturbation" does not have to be small but could be any size). Allow the Ising model to have spins with the usual two values: 1 and -1. The Hamiltonian produced by the perturbation δ is,

$$\begin{aligned} H_0 &= \sum_i \sigma_i \sigma_{i+1} \\ H' &= H_0 - \delta \sigma_I \sigma_J, \end{aligned} \tag{7.1}$$

where H_0 is the Hamiltonian for the unperturbed Ising model, and the perturbation connects the sites I and J. In light of (4.11), the partition functions of both theories are needed. The partition function, Z_{pert}, for the perturbed theory can be written in terms of the partition function for the unperturbed theory, Z_0, and it is,

$$\begin{aligned} Z_{pert} &= \int \mathrm{D}\sigma \; e^{-\beta H_0 + \beta\delta\sigma_I\sigma_J} \\ &= \int \mathrm{D}\sigma \; e^{-\beta H_0} \sum_{m=0}^{\infty} \left(\frac{(\beta\delta)^{2m}}{(2m)!} + \frac{(\beta\delta)^{2m+1}\sigma_I\sigma_J}{(2m+1)!} \right) \\ &= \int \mathrm{D}\sigma \; \frac{e^{-\beta H_0}}{Z_0} Z_0 \left(\cosh(\beta\delta) + \sigma_I\sigma_J \sinh(\beta\delta) \right) \\ &= Z_0 \left(\cosh(\beta\delta) + \langle \sigma_I\sigma_J \rangle_0 \sinh(\beta\delta) \right), \end{aligned} \tag{7.2}$$

where the correlation function is taken with respect to the unperturbed theory and the series expansion was broken into odd and even parts in order to make use of $(\sigma_I\sigma_J)^n = \sigma_I\sigma_J$ for n odd and $(\sigma_I\sigma_J)^n = 1$ for n even. Note that the partition function is written in terms of the correlation function $\langle \sigma_I\sigma_J \rangle_0$. It turns out that the entire KL divergence can be written in terms of the correlation function $\langle \sigma_I\sigma_J \rangle_0$.

Now in general, the KL divergence becomes (taking p to be the theory without the perturbation and q to be the theory with the perturbation),

$$\begin{aligned} D(p||q) &= \log \frac{Z_q}{Z_p} + \beta \langle H_0 - \delta \sigma_I \sigma_J - H_0 \rangle_p \\ &= \log \frac{Z_q}{Z_p} - \beta\delta \langle \sigma_I \sigma_J \rangle_p, \end{aligned} \tag{7.3}$$

where use was made of (4.11). Plugging in the partition function for the perturbed theory q that was previously calculated, we get,

$$\begin{aligned} D(p||q) &= \log \frac{Z_p \left(\cosh(\beta\delta) + \langle \sigma_I \sigma_J \rangle_p \sinh(\beta\delta)\right)}{Z_p} - \beta\delta \langle \sigma_I \sigma_J \rangle_p \\ &= \log \left(\cosh(\beta\delta) + \langle \sigma_I \sigma_J \rangle_p \sinh(\beta\delta)\right) - \beta\delta \langle \sigma_I \sigma_J \rangle_p. \end{aligned} \tag{7.4}$$

Going in the other direction, calculating $D(q||p)$, we get,

$$\begin{aligned} D(q||p) &= \log \frac{Z_p}{Z_q} + \beta \langle H_0 - H_0 + \delta \sigma_I \sigma_J \rangle_q \\ &= \log \left(\frac{Z_p}{Z_p \left(\cosh(\beta\delta) + \langle \sigma_I \sigma_J \rangle_p \sinh(\beta\delta)\right)} \right) + \beta\delta \int \mathrm{D}\sigma \ \frac{e^{-\beta H_q}}{Z_q} \sigma_I \sigma_J \\ &= -\log \left(\cosh(\beta\delta) + \langle \sigma_I \sigma_J \rangle_p \sinh(\beta\delta)\right) \\ &+ \beta\delta \int \mathrm{D}\sigma \ \frac{e^{-\beta H_p + \beta\delta\sigma_I\sigma_J}}{Z_p} \frac{\sigma_I \sigma_J}{\cosh(\beta\delta) + \langle \sigma_I \sigma_J \rangle_p \sinh(\beta\delta)}. \end{aligned} \tag{7.5}$$

We now make use of the series summation performed in calculating the perturbed partition function in (7.2): the exponential with the δ perturbation can be resummed into a series of $\cosh$ and $\sinh$. This results in,

$$\begin{aligned} D(q||p) &= -\log \left(\cosh(\beta\delta) + \langle \sigma_I \sigma_J \rangle_p \sinh(\beta\delta)\right) \\ &+ \beta\delta \int \mathrm{D}\sigma \ \frac{e^{-\beta H_p}}{Z_p} \frac{\left(\cosh(\beta\delta) + \sigma_I \sigma_J \sinh(\beta\delta)\right) \sigma_I \sigma_J}{\cosh(\beta\delta) + \langle \sigma_I \sigma_J \rangle_p \sinh(\beta\delta)} \\ &= -\log \left(\cosh(\beta\delta) + \langle \sigma_I \sigma_J \rangle_p \sinh(\beta\delta)\right) + \beta\delta \int \mathrm{D}\sigma \ \frac{e^{-\beta H_p}}{Z_p} \frac{\sigma_I \sigma_J \cosh(\beta\delta) + \sinh(\beta\delta)}{\cosh(\beta\delta) + \langle \sigma_I \sigma_J \rangle_p \sinh(\beta\delta)} \\ &= -\log \left(\cosh(\beta\delta) + \langle \sigma_I \sigma_J \rangle_p \sinh(\beta\delta)\right) + \beta\delta \frac{\langle \sigma_I \sigma_J \rangle_p \cosh(\beta\delta) + \sinh(\beta\delta)}{\cosh(\beta\delta) + \langle \sigma_I \sigma_J \rangle_p \sinh(\beta\delta)}, \end{aligned} \tag{7.6}$$

where in the first line we made use of the fact that $\sigma_I^2 = 0$ regardless of the index.

We see from our results of the KL divergence in both directions that if the two-point function is known, then the KL divergence can be calculated exactly for the Ising model with a perturbation.

7.1 1D Perturbations

Having considered the perturbed Ising model in general, we now consider a 1D Ising model with a perturbation. The model has N sites and periodic boundary conditions, and the spins can again take the two values of 1 and -1. We first perturb the spin site at the boundary by a small amount δ, producing a Hamiltonian,

$$H' = H_0 - \delta\sigma_N\sigma_1, \tag{7.7}$$

where H_0 is the usual 1D Ising model Hamiltonian with periodic boundary conditions. In this case, the last site is perturbed, but if a different site was perturbed, the indices could be relabeled to make the last site the perturbed one. Using the method of transfer matrices, we define matrix components of matrices P and Q as,

$$\begin{aligned} Q_{1,1} = Q_{-1,-1} &= e^{\beta(J+\delta)}, \\ Q_{1,-1} = Q_{-1,1} &= e^{-\beta(J+\delta)}, \\ P_{ij} &= Q_{ij}|_{\delta=0}. \end{aligned} \tag{7.8}$$

This choice of transfer matrix elements makes the partition function reduce to

$$Tr[P^{N-1}Q]. \tag{7.9}$$

It turns out that both Q and P have the same eigenvectors ($\frac{1}{\sqrt{2}}(1, +/-1)$) and so can be simultaneously diagonalized. Thus, the trace is just (using Mathematica to calculate the eigenvalues/eigenvectors and sum them),

$$\begin{aligned} &\lambda_{p1}^{N-1}\lambda_{q1} + \lambda_{p2}^{N-1}\lambda_{q2} \\ &= \left(e^{-\beta J} + e^{\beta J}\right)^{N-1}\left(e^{-\beta(J+\delta)} + e^{\beta(J+\delta)}\right) \\ &+ \left(e^{\beta J} - e^{-\beta J}\right)^{N-1}\left(e^{\beta(J+\delta)} - e^{-\beta(J+\delta)}\right) \\ &= 2^N(\cosh^{N-1}(\beta J)\cosh(\beta(J+\delta)) + \sinh^{N-1}(\beta J)\sinh(\beta(J+\delta))), \end{aligned} \tag{7.10}$$

where the λ's are the eigenvalues of P and Q. The relative entropy between the theory with and the theory without the perturbation δ can now be calculated. Observe that the relative entropy is just the expectation value of the perturbation subtracted from the logarithm of the ratio of the partition functions,

$$\begin{aligned} D(p||q) &= \sum_{\sigma_i} \frac{1}{Z_p} e^{-\beta H_p} \log\left(\frac{e^{-\beta H_p} Z_q}{e^{-\beta H_q} Z_p}\right) \\ &= \sum_{\sigma_i} \frac{1}{Z_p} e^{-\beta H_p} \left(\beta(H_q - H_p) + \log\left(\frac{Z_q}{Z_p}\right)\right) \\ &= \log\left(\frac{Z_q}{Z_p}\right) - \sum_{\sigma_i} \beta\delta\sigma_N\sigma_1 \frac{1}{Z_p} e^{-\beta H_p} \\ &= \log\left(\frac{Z_q}{Z_p}\right) - \beta\delta\langle\sigma_N\sigma_1\rangle_p, \end{aligned} \tag{7.11}$$

where in the last two lines, we took p to be the unperturbed theory and q to be the perturbed theory. $D(p||q)$ is first calculated, followed by $D(q||p)$, although they should give the same result. The correlation function is (without loss of generality, we assume $n < m$, so take $m = N + 1$ by the periodic boundary condition),

$$\begin{aligned} \langle\sigma_n\sigma_m\rangle_p &= \frac{1}{Z_p} \sum_{a,b} \lambda_{pa}^{N+n-m} \lambda_{pb}^{m-n} \langle a|\tau_3|b\rangle\langle b|\tau_3|a\rangle, \\ \langle\sigma_N\sigma_1\rangle_p &= \frac{1}{Z_p} \sum_{a,b} \lambda_{pa}^{N-1} \lambda_{pb} \langle a|\tau_3|b\rangle\langle b|\tau_3|a\rangle, \end{aligned} \tag{7.12}$$

where $|a\rangle$ and $|b\rangle$ are the normalized eigenvectors of the transfer matrix for theory p. This equation assumes that the site on the left is smaller is less than the site on the right, i.e., that $N < 1$. This is works because the periodic boundary condition makes $1 = N+1$ and taking the absolute value at an appropriate place will make the result independent of which site is less than the other. Noting that $\langle 1|\tau_3|1\rangle = 0$ (and so $\langle\sigma_N\rangle = \langle\sigma_1\rangle = 0$), we get,

$$\begin{aligned} \langle\sigma_N\sigma_1\rangle_p &= \frac{1}{Z_p}(\lambda_{p1}^{N-1}\lambda_{p1}|\langle 1|\tau_3|1\rangle|^2 + \lambda_{p2}^{N-1}\lambda_{p2}|\langle 2|\tau_3|2\rangle|^2 \\ &\quad + \lambda_{p1}^{N-1}\lambda_{p2}|\langle 1|\tau_3|2\rangle|^2 + \lambda_{p2}^{N-1}\lambda_{p1}|\langle 1|\tau_3|2\rangle|^2) \\ &= \frac{\lambda_{p1}^{N-1}\lambda_{p2}|\langle 1|\tau_3|2\rangle|^2 + \lambda_{p2}^{N-1}\lambda_{p1}|\langle 1|\tau_3|2\rangle|^2}{\lambda_{p1}^N + \lambda_{p2}^N} \\ &= \left(\frac{\lambda_{p2}}{\lambda_{p1}}\right) \frac{\lambda_{p1}^N + \lambda_{p2}^N \left(\frac{\lambda_{p1}}{\lambda_{p2}}\right)^2}{\lambda_{p1}^N + \lambda_{p2}^N} |\langle 1|\tau_3|2\rangle|^2, \end{aligned} \tag{7.13}$$

where we now simplify our answer in terms of a correlation length $\frac{1}{\zeta}$ to get,

$$\begin{aligned}
\langle \sigma_N \sigma_1 \rangle_p &= |\langle 1|\tau_3|2\rangle|^2 e^{-1/\zeta} = |\langle 1|\tau_3|2\rangle|^2 e^{-1/\xi} e^{-1/\chi}, \\
\frac{1}{\zeta} &= \frac{1}{\xi} + \frac{1}{\chi}, \\
\frac{1}{\xi} &\equiv -\log(\frac{\lambda_{p2}}{\lambda_{p1}}), \\
\frac{1}{\chi} &\equiv -\log \frac{\lambda_{p1}^N + \lambda_{p2}^N \left(\frac{\lambda_{p1}}{\lambda_{p2}}\right)^2}{\lambda_{p1}^N + \lambda_{p2}^N},
\end{aligned} \tag{7.14}$$

where note that we have defined $1/\zeta$ to be the negative logarithm of the combination of eigenvalues that multiplies the $|\langle 1|\tau_3|2\rangle|^2$, and we have the usual correlation length from the 1D Ising model when there is no perturbation $1/\xi$. Notice that the combination of eigenvalues splits into a product of two fractions. Hence, we finally find that (noting that $\langle 1|\tau_3|2\rangle = 1$, since the eigenvectors are normalized),

$$\begin{aligned}
\langle \sigma_N \sigma_1 \rangle_p &= |\langle 1|\tau_3|2\rangle|^2 e^{-1/\zeta} = e^{-1/\zeta}, \\
\frac{-1}{\zeta} &= \log\left(\frac{\sinh(\beta J)}{\cosh(\beta J)}\right) + \log\left(\frac{\cosh(\beta J)^N + \sinh(\beta J)^N \left(\frac{\cosh(\beta J)}{\sinh(\beta J)}\right)^2}{\cosh(\beta J)^N + \sinh(\beta J)^N}\right) \\
&= \log(\tanh(\beta J)) + \log\left(\frac{\cosh(\beta J)^N + \sinh(\beta J)^N \coth(\beta J)^2}{\cosh(\beta J)^N + \sinh(\beta J)^N}\right).
\end{aligned} \tag{7.15}$$

In the limit of $N >> 1$, the only thing that changes is the ζ, so we find that (since $\lambda_{p1} >> \lambda_{p2}$ in this limit),

$$\begin{aligned}
\frac{-1}{\chi}|_{N>>1} &= \log\left(\frac{\lambda_{p1}^N}{\lambda_{p1}^N}\right) = 0 \\
\rightarrow \frac{1}{\zeta} &= \frac{1}{\xi} = -\log\left(\lambda_{p2}/\lambda_{p1}\right) \\
&= \log\left(\frac{\cosh(\beta J)}{\sinh(\beta J)}\right) = \log(\coth(\beta J)) \\
\rightarrow \langle \sigma_N \sigma_1 \rangle_p &= e^{-1/\xi} = \tanh(\beta J),
\end{aligned} \tag{7.16}$$

We hence see that in general (with $N >> 1$) for an expectation value taken with respect to theory p, $1/\xi = -\log(\lambda_{p2}/\lambda_{p1})$. Subtracting from the logarithm of the partition functions, the exact result for the

relative entropy is (Note that the following result does not require δ to be small.),

$$\begin{aligned} D(p(J)||q(J,\delta)) &= \log\left(\frac{Z_q}{Z_p}\right) - \beta\delta e^{-1/\zeta} \\ &= \log\left(\frac{(\cosh^{N-1}(\beta J)\cosh(\beta(J+\delta)) + \sinh^{N-1}(\beta J)\sinh(\beta(J+\delta)))}{(\cosh^N(\beta J) + \sinh^N(\beta J)}\right) \\ &- \beta\delta e^{-1/\xi} e^{-1/\chi} \\ &= \log\left(\frac{(\cosh^{N-1}(\beta J)\cosh(\beta(J+\delta)) + \sinh^{N-1}(\beta J)\sinh(\beta(J+\delta)))}{(\cosh^N(\beta J) + \sinh^N(\beta J)}\right) \\ &- \beta\delta\left(\tanh(\beta J)\frac{\cosh(\beta J)^N + \sinh(\beta J)^N \coth(\beta J)^2}{\cosh(\beta J)^N + \sinh(\beta J)^N}\right), \end{aligned} \tag{7.17}$$

where the leading order behavior for small perturbations (δ small) is quadratic as expected,

$$D(p(J)||q(J,\delta)) = 2(\beta\delta\,\mathrm{csch}(2\beta J))^2 \frac{\cosh^{2N}(\beta J)\sinh^2(\beta J) - \sinh^{2N}(\beta J)\cosh^2(\beta J)}{(\cosh^N(\beta J) + \sinh^N(\beta J))^2}. \tag{7.18}$$

For $N >> 1$, we have,

$$\begin{aligned} D(p(J)||q(J,\delta)) &= \log\left(\frac{Z_q}{Z_p}\right)|_{N>>1} - \beta\delta e^{-1/\xi} \\ &= \log\left(\frac{\cosh^{N-1}(\beta J)\cosh(\beta(J+\delta))}{\cosh^N(\beta J)}\right) - \beta\delta e^{-1/\xi} \\ &= \log\left(\frac{\cosh(\beta(J+\delta))}{\cosh(\beta J)}\right) - \beta\delta\tanh(\beta J), \end{aligned} \tag{7.19}$$

with leading order,

$$D(q(J,\delta)||p(J)) = \frac{(\beta\delta)^2}{2}\mathrm{sech}^2(\beta J). \tag{7.20}$$

Computing the relative entropy the other way gives a similar result. We have for the relative entropy,

$$\begin{aligned} D(q||p) &= \sum_{\sigma_i} \frac{1}{Z_q} e^{-\beta H_q}\left(\beta(H_p - H_q) + \log\left(\frac{Z_p}{Z_q}\right)\right) \\ &= -\log\left(\frac{Z_q}{Z_p}\right) + \beta\delta\langle\sigma_N\sigma_1\rangle_q. \end{aligned} \tag{7.21}$$

The correlation function is,

$$\begin{aligned}
\langle\sigma_n\sigma_m\rangle_q &= \frac{1}{Z_q}\sum_{a,b}\lambda_{pa}^{N+n-m-1}\lambda_{qa}\lambda_{pb}^{m-n}\langle a|\tau_3|b\rangle\langle b|\tau_3|a\rangle, \\
\langle\sigma_N\sigma_1\rangle_q &= \frac{1}{Z_q}\sum_{a,b}\lambda_{pa}^{N-1}\lambda_{qb}\langle a|\tau_3|b\rangle\langle b|\tau_3|a\rangle \\
&= \frac{\lambda_{p1}^{N-1}\lambda_{q2}|\langle 1|\tau_3|2\rangle|^2 + \lambda_{p2}^{N-1}\lambda_{q1}|\langle 1|\tau_3|2\rangle|^2}{\lambda_{p1}^{N-1}\lambda_{q1} + \lambda_{p2}^{N-1}\lambda_{q2}} \\
&= \left(\frac{\lambda_{q2}}{\lambda_{q1}}\right)\frac{\lambda_{p1}^{N-1}\lambda_{q1} + \lambda_{p2}^{N-1}\lambda_{q2}\left(\frac{\lambda_{q1}}{\lambda_{q2}}\right)^2}{\lambda_{p1}^{N-1}\lambda_{q1} + \lambda_{p2}^{N-1}\lambda_{q2}}|\langle 1|\tau_3|2\rangle|^2 \\
&= |\langle 1|\tau_3|2\rangle|^2 e^{-1/\zeta} = |\langle 1|\tau_3|2\rangle|^2 e^{-1/\xi}e^{-1/\chi},
\end{aligned} \tag{7.22}$$

where the first correlation function equation is one for which the spin site is not the perturbed one (site N), and the second correlation function equation is for the case of interest, where site N is correlated. Notice that this case is the same as before, except that ζ has changed, this time defining $1/\xi \equiv -\log\left(\frac{\lambda_{q2}}{\lambda_{q1}}\right)$. We thus get for the relative entropy,

$$\begin{aligned}
D(q(J,\delta)||p(J)) &= -\log\left(\frac{(\cosh^{N-1}(\beta J)\cosh(\beta(J+\delta)) + \sinh^{N-1}(\beta J)\sinh(\beta(J+\delta)))}{\cosh^N(\beta J)+\sinh^N(\beta J)}\right) \\
&+ \beta\delta\tanh(\beta(J+\delta)) \\
&\times\left(\frac{\cosh^{N-1}(\beta J)\cosh(\beta(J+\delta)) + \sinh^{N-1}(\beta J)\sinh(\beta(J+\delta))\coth(\beta(J+\delta))^2}{\cosh^{N-1}(\beta J)\cosh(\beta(J+\delta)) + \sinh^{N-1}(\beta J)\sinh(\beta(J+\delta))}\right),
\end{aligned} \tag{7.23}$$

which is to leading order the same as calculated in the other direction,

$$D(p(J)||q(J,\delta)) = 2(\beta\delta\,\mathrm{csch}(2\beta J))^2\frac{\cosh^{2N}(\beta J)\sinh^2(\beta J) - \sinh^{2N}(\beta J)\cosh^2(\beta J)}{(\cosh^N(\beta J)+\sinh^N(\beta J))^2}. \tag{7.24}$$

And the results for $N >> 1$ are,

$$\begin{aligned}
\langle\sigma_N\sigma_1\rangle_q|_{N>>1} &= \left(\frac{\lambda_{q2}}{\lambda_{q1}}\right) = \tanh(\beta(J+\delta)) \\
D(q(J,\delta)||p(J)) &= -\log\left(\frac{\cosh(\beta(J+\delta))}{\cosh(\beta J)}\right) + \beta\delta\tanh(\beta(J+\delta)),
\end{aligned} \tag{7.25}$$

with leading order piece the same as calculated before,

$$D(q(J,\delta)||p(J)) = \frac{(\beta\delta)^2}{2}\operatorname{sech}^2(\beta J). \tag{7.26}$$

All of these results are independent of N for large N and do not blow up. As an explanation for why this is the case, consider the usual scaling argument for a free field propagator. We have that $\phi(x) \sim (1/L)^{(D-2)/2}$, where D is the dimension. Hence, the two-point function goes as,

$$\langle\phi(x)\phi(y)\rangle \sim \frac{1}{(x-y)^{D-2}}. \tag{7.27}$$

Taking $D = 1$ gives no singularity, so there are no contact terms in 1D. We are working with a 1D Ising model, so we can expect from this analogy that we have no contact terms in our result. If there was a contact term though, the lattice spacing would provide a natural regulator to prevent the results from blowing up.

Having considered a perturbation on the boundary of the 1D Ising model, now consider the case where all the spin sites are perturbed by different perturbations $\delta J_{ij}\sigma_i\sigma_j$, but the spin sites still interact in a nearest neighbor fashion, so $j = i + 1$. In this case, again using periodic boundary conditions, the partition function is simply,

$$\begin{aligned} Z_q &= Tr[Q_1 Q_2 ... Q_N], \\ Q_{i11} &= Q_{i,-1,-1} = e^{\beta(J+\delta J_{i,i+1})}, \\ Q_{i,1,-1} &= Q_{i,-1,1} = e^{-\beta(J+\delta J_{i,i+1})} \\ \rightarrow Z_q &= \lambda_{1,1}\lambda_{2,1}...\lambda_{N,1} + \lambda_{1,2}\lambda_{2,2}...\lambda_{N,2}, \end{aligned} \tag{7.28}$$

having simultaneously diagonalized the P matrices as done before. The eigenvalues are,

$$\begin{aligned} \lambda_{i,1} &= 2\cosh(\beta(J + \delta J_{i,i+1})), \\ \lambda_{i,2} &= 2\sinh(\beta(J + \delta J_{i,i+1})), \end{aligned} \tag{7.29}$$

keeping with the convention from before where the $\cosh$ is labeled as the first eigenvalue. Note that this method is completely general, i.e., the partition function for spin sites, even without periodic boundary

conditions, is the same as just calculated except an subscript j is used instead of $i+1$. Hence, there seems to be no advantage to calculating the relative entropy resulting from nearest neighbor interactions first. Instead, the general calculation will be made, which will contain the nearest neighbor interactions as a special case.

First, the relative entropy $D(p(J)||q(J,\delta J_{ij}))$ will be calculated. The correlation function for one pair of sites i' and j' is,

$$\begin{aligned}
\langle\sigma_{i'}\sigma_{j'}\rangle_p &= \frac{1}{Z_p}\sum_{a,b}\lambda_{pa}^{N+i'-j'}\lambda_{pb}^{j'-i'}\langle a|\tau_3|b\rangle\langle b|\tau_3|a\rangle \\
&= \frac{1}{Z_p}\left(\lambda_{p1}^N\frac{\lambda_{p2}}{\lambda_{p1}}^{j'-i'}+\lambda_{p2}^N\frac{\lambda_{p1}}{\lambda_{p2}}^{j'-i'}\right)|\langle 1|\tau_3|2\rangle|^2 \\
&= \frac{\lambda_{p2}}{\lambda_{p1}}^{j'-i'}\frac{\lambda_{p1}^N+\lambda_{p2}^N\frac{\lambda_{p1}}{\lambda_{p2}}^{2j'-2i'}}{\lambda_{p1}^N+\lambda_{p2}^N}|\langle 1|\tau_3|2\rangle|^2.
\end{aligned} \tag{7.30}$$

We again define a correlation length $1/\zeta$, where $1/\xi$ is the same as in (7.16). This gives,

$$\begin{aligned}
\langle\sigma_{i'}\sigma_{j'}\rangle_p &= |\langle 1|\tau_3|2\rangle|^2 e^{-1/\zeta(j',i')} = e^{-|j'-i'|/\xi}e^{-1/\chi(j',i')}, \\
\frac{1}{\chi(j',i')} &= -\log\frac{\lambda_{p1}^N+\lambda_{p2}^N\frac{\lambda_{p1}}{\lambda_{p2}}^{2j'-2i'}}{\lambda_{p1}^N+\lambda_{p2}^N}.
\end{aligned} \tag{7.31}$$

Now, since the difference in the actions is the sum of all the perturbations, the relative entropy becomes (sum over all pairs of sites on i' and j'; implicit product on i from 1 to N),

$$\begin{aligned}
D(p(J)||q(J,\delta J_{ij})) &= \log\left(\frac{Z_q}{Z_p}\right)-\beta\sum_{i',j'}\delta J_{i',j'}\langle\sigma_{i'}\sigma_{j'}\rangle_p \\
&= \log\left(\frac{Z_q}{Z_p}\right)-\beta\sum_{i',j'}\delta J_{i',j'}e^{-1/\zeta(i',j')} \\
&= \log\left(\frac{\cosh(\beta(J+\delta J_{i,j}))+\sinh(\beta(J+\delta J_{i,j}))}{\cosh^N(\beta J)+\sinh^N(\beta J)}\right) \\
&\quad -\beta\sum_{i',j'}\delta J_{i'j'}e^{-|j'-i'|/\xi}e^{-1/\chi(j',i')} \\
&= \log\left(\frac{\cosh(\beta(J+\delta J_{i,j}))+\sinh(\beta(J+\delta J_{i,j}))}{\cosh^N(\beta J)+\sinh^N(\beta J)}\right) \\
&\quad -\beta\sum_{i',j'}\delta J_{i'j'}\left(\tanh(\beta J)^{|j'-i'|}\frac{\cosh(\beta J)^N+\sinh(\beta J)^N\coth(\beta J)^{2|j'-i'|}}{\cosh(\beta J)^N+\sinh(\beta J)^N}\right),
\end{aligned} \tag{7.32}$$

where each i in the $\log$ term could have a different j, i.e., each site i is coupled to distinct site j's; each i site is not necessarily coupled to the same j site. So the log term has a sum of products over the paired i and j sites, while the correlation function is a sum over all the paired i and j sites (labeled by dummy indices i' and j').

For $N >> 1$, the correlation function becomes,

$$\begin{aligned}\langle \sigma_i \sigma_j \rangle_p|_{N>>1} &= \frac{1}{Z_p} \sum_{a,b} \lambda_{pa}^{N+i'-j'} \lambda_{pb}^{j'-i'} \langle a|\tau_3|b\rangle\langle b|\tau_3|a\rangle \\ &= |\langle 1|\tau_3|2\rangle|^2 e^{-|j'-i'|/\xi} = e^{-|j'-i'|/\xi},\end{aligned} \tag{7.33}$$

and the relative entropy is (implicit product over i from 1 to N and sum again on i' and j'),

$$\begin{aligned}D(p(J)||q(J,\delta J_{ij}))|_{N>>1} &= \log\left(\frac{Z_q}{Z_p}\right) - \beta \sum_{i',j'} \delta J_{i'j'} e^{-|j'-i'|/\xi} \\ &= \log\left(\frac{Z_q}{Z_p}\right) - \beta \sum_{i',j'} \delta J_{i'j'} \tanh^{|j'-i'|}(\beta J) \\ &= \log\left(\frac{\cosh(\beta(J+\delta J_{ij})) + \sinh(\beta(J+\delta J_{ij}))}{\cosh^N(\beta J) + \sinh^N(\beta J)}\right) \\ &\quad - \beta \sum_{i',j'} \delta J_{i'j'} \tanh^{|j'-i'|}(\beta J).\end{aligned} \tag{7.34}$$

CHAPTER 8

Example: Ising Model Decimation

For both 1D and 2D Ising models, the relative entropy is calculated between a theory decimated k times and the original, undecimated theory. Also, the relative entropy is calculated between a theory decimated k times and the same theory decimated l times. The exact expressions for these relative entropies that are true for any dimension and a general lattice QFT (with Boltzmann distributions) were calculated in Section 5, so we directly use these expressions with the Ising models. As in those expressions, to make the calculations easier, we will absorb the constant factor of the partition function into the Hamiltonian (an example of the shift in the Hamiltonian is in (8.1)), which makes the partition function for both theories the same at all decimation steps.

8.1 1D Decimation

Unlike with the Hamiltonian of the perturbed theory in Section 7, there will be no periodic boundary conditions, but instead there will be a chain with sites going from -2^N+1 to 2^N. Using and generalizing the results of [125] for 1 to 3 decimations, the Hamiltonian after k decimations is,

$$\begin{aligned} -\beta H_k &= \sum_{i=1}^{N/2^k} c_k + \beta_k \sigma_{2^k(i+1)} \sigma_{2^k i} (N\ even) \\ &= \sum_{i=1}^{(N-1)/2^k} c_k + \beta_k \sigma_{2^k(i+1)} \sigma_{2^k i} (N\ odd), \end{aligned} \tag{8.1}$$

where the recursion relations for the β_k and the c_k are,

$$\begin{aligned}
\beta_k &\equiv g^k(\beta J) = g(g(g(...g(\beta J))))(k\ times), \\
&= g(\beta J) = \frac{1}{2}\log\cosh(2\beta J), \\
c_k &= \sum_{m=0}^{k-1} 2^{k-1-m} h(g^m(\beta J)), \\
h(g^m(\beta J)) &\equiv h(\beta_m) = \frac{1}{2}\log(4\cosh(2\beta_m)), \\
h(g^0(\beta J)) &\equiv h(\beta J) = \frac{1}{2}\log(4\cosh(2\beta J)).
\end{aligned} \tag{8.2}$$

We already have $\log Z$, since Z was calculated earlier in (7.10) (where $\delta = 0$ in this case), and (8.2) gives the relation between β and β_{odd} (just perform the decimation once). Putting this into the equation (5.9) for relative entropy in the translation invariant case and then calculating with Mathematica yields (to leading order in the limit of large N with small and large β),

$$\begin{aligned}
D(p||q)|_{\beta,N\to\infty} &= N\beta J(3 - e^{-\beta J}) - \log(2^{N-1}) \\
D(p||q)|_{\beta\to 0,N\to\infty} &= \log(2^N) + \log(1 + (\beta J)^N) + \frac{N(\beta J)^{N+1}}{1 + (\beta J)^N}.
\end{aligned} \tag{8.3}$$

Using periodic boundary conditions, we can calculate the relative entropy between the 1D Ising model and the 1D Ising model decimated k times as a function of k, using (5.10) for the translation invariant case. We will treat βJ as one constant and will take the number sites $N = 2^{19}$. We calculate the relative entropy density and plot the relative entropy density versus decimation step, Figures 8.1 and 8.2.

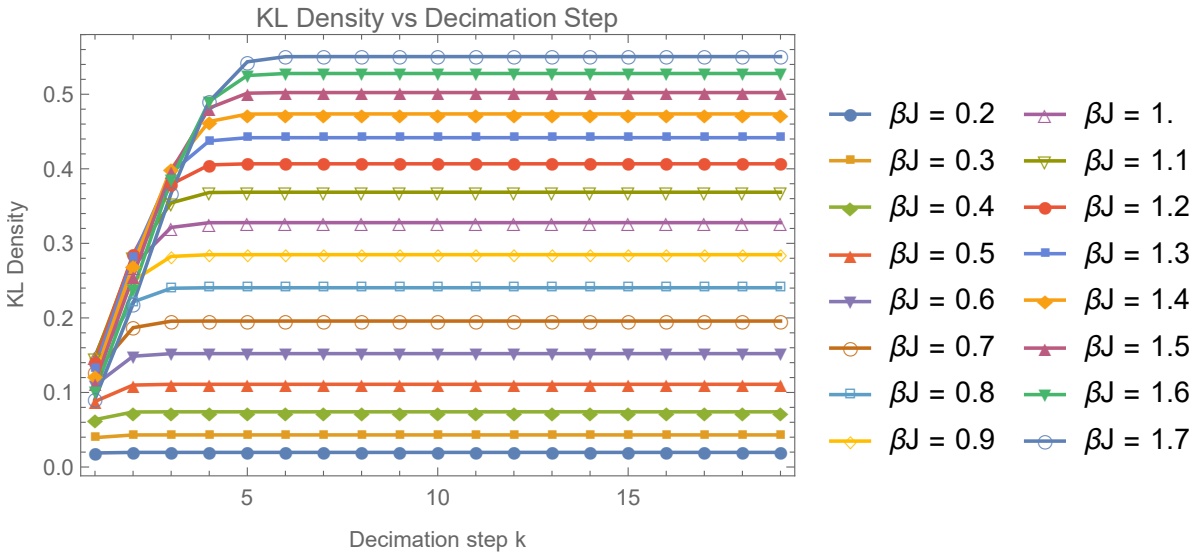

Figure 8.1: The KL density for the 1D Ising model as a function of decimation step for fixed, low values of βJ. There is no critical value for the 1D Ising model, but we cover the parameter space that includes the critical value of the 2D Ising model.

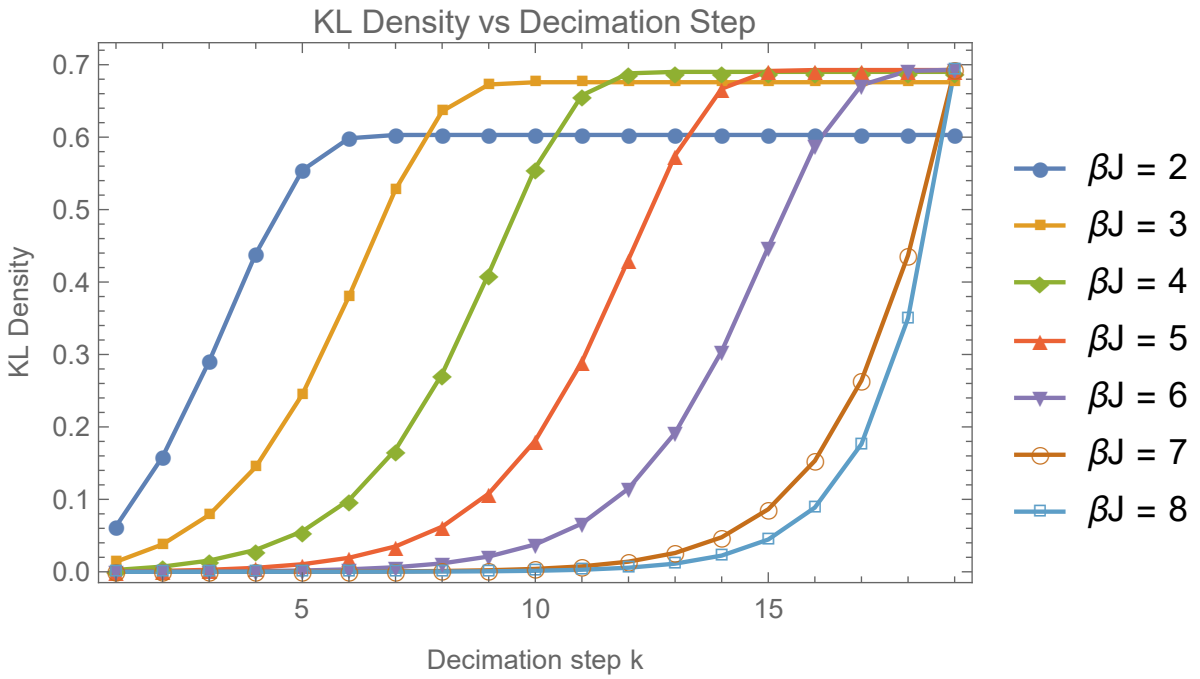

Figure 8.2: The KL density for the 1D Ising model as a function of decimation step for fixed, higher values of βJ. We start to see the curves converge on the final curve shown in this figure.

We can also calculate the beta function of the KL divergence density by performing,

$$\beta_{KL} = \frac{\partial}{\partial k} D(p||q) = \frac{D(p||q(k+1)) - D(p||q(k))}{2^k}, \tag{8.4}$$

where we divide by 2^k because the KL divergence density is a dimensionful quantity. By (6.24), this expression for the beta function of the KL divergence is correct, except for needing to divide by $\log 2^{1/D} = \log 2$. The missing numerical factor occurs due to calculating with respect to the decimation step instead of the energy scale or lattice spacing directly.

Our results for the beta function of the KL density follow in Figures 8.3 and 8.4.

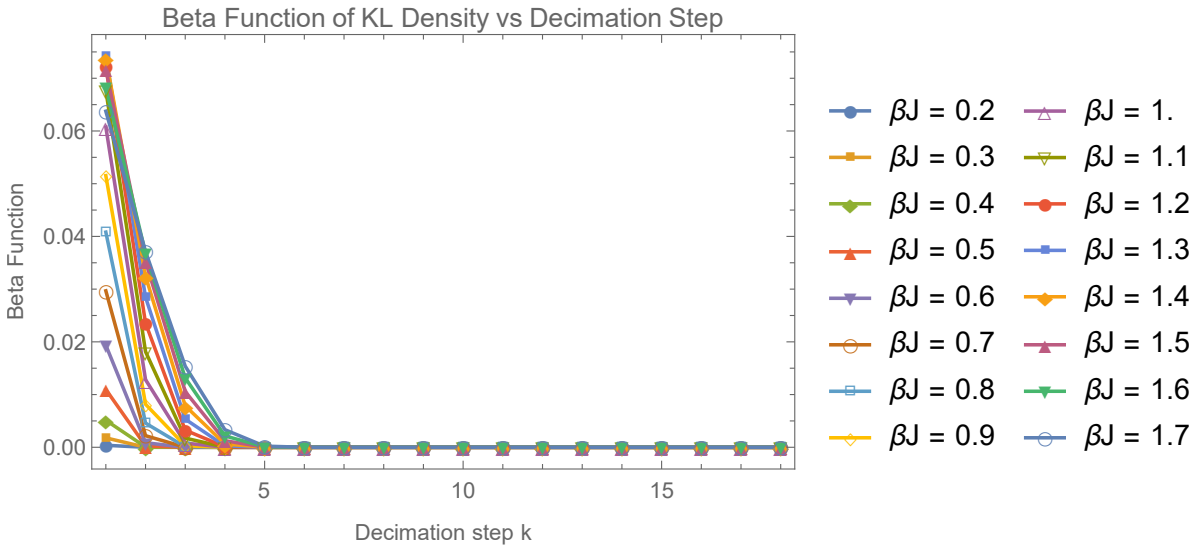

Figure 8.3: The beta function of the KL density for the 1D Ising model as a function of decimation step for fixed, low values of βJ. These are the same values that were used for the KL density at low values of βJ.

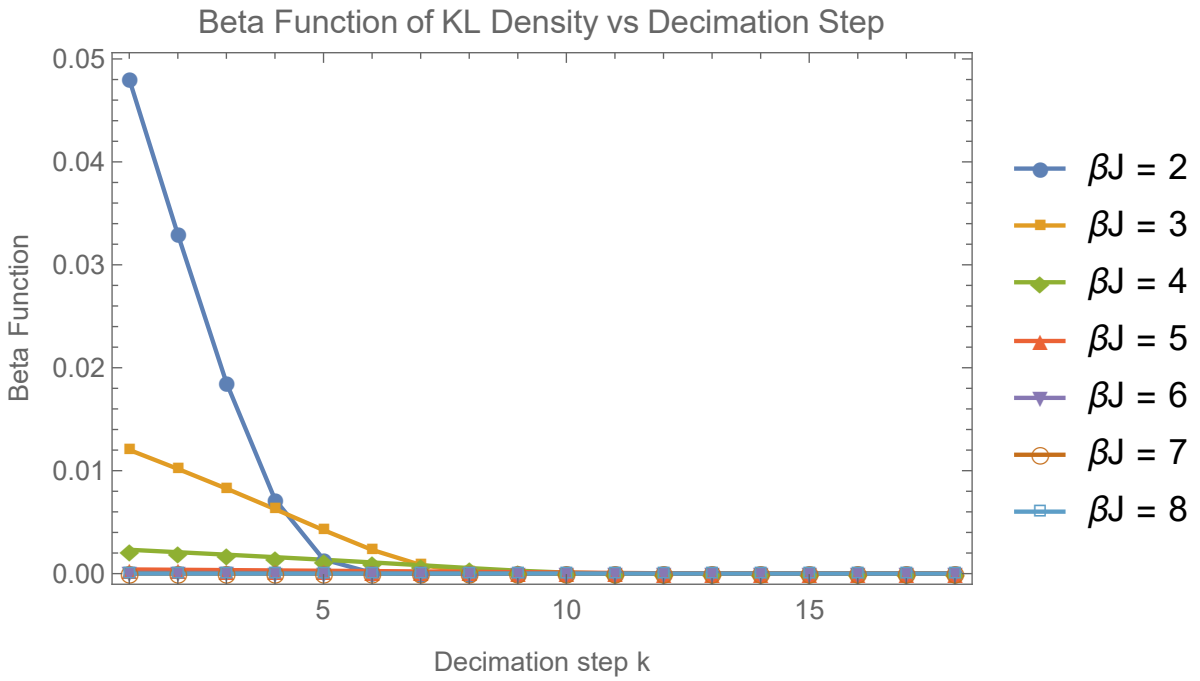

Figure 8.4: The beta function of the KL density for the 1D Ising model as a function of decimation step for fixed, higher values of βJ. These are the same values that were used for the KL density at higher values of βJ.

8.1.1 1D Decimation: Rescaling

We will now calculate the KL divergence for a 1D Ising model by using the method of decimation plus rescaling. In doing this, we start with one probability distribution with a set of couplings β, $p[H(\beta)]$. After decimation and rescaling, we are left with a probability distribution $p[H(\beta_k)]$ that has the same form and spin operators but with new coupling constants β_k and a spin independent constant in the Hamiltonian. The KL divergence is simply then,

$$\begin{aligned} D(p||q) &= (-\langle H\rangle + \langle H_k\rangle) \\ &= (-E(\beta) + (E(\beta_k) - C)) \\ &= \log Z_k - \log Z + (-E(\beta) + E(\beta_k))\,, \end{aligned} \tag{8.5}$$

where all the quantities β_k, C, etc., are defined as previously in the 1D Ising model calculated with decimation plus forming a joint distribution. The partition function disappears because it was held fixed. We have $C = \log Z - \log Z_k$ as the spin-independent constant that keeps the decimated partition function fixed. The only difference between this case and the previous is that the 2^k factors have disappeared from our final result (although the 2^k factors in the other case cancelled with the number of sites at each decimation step $N/2^k$).

Our results for the KL density when using decimation with rescaling are shown below in Figures 8.5 and 8.6. There are many similarities to the earlier results that used decimation plus forming a joint distribution.

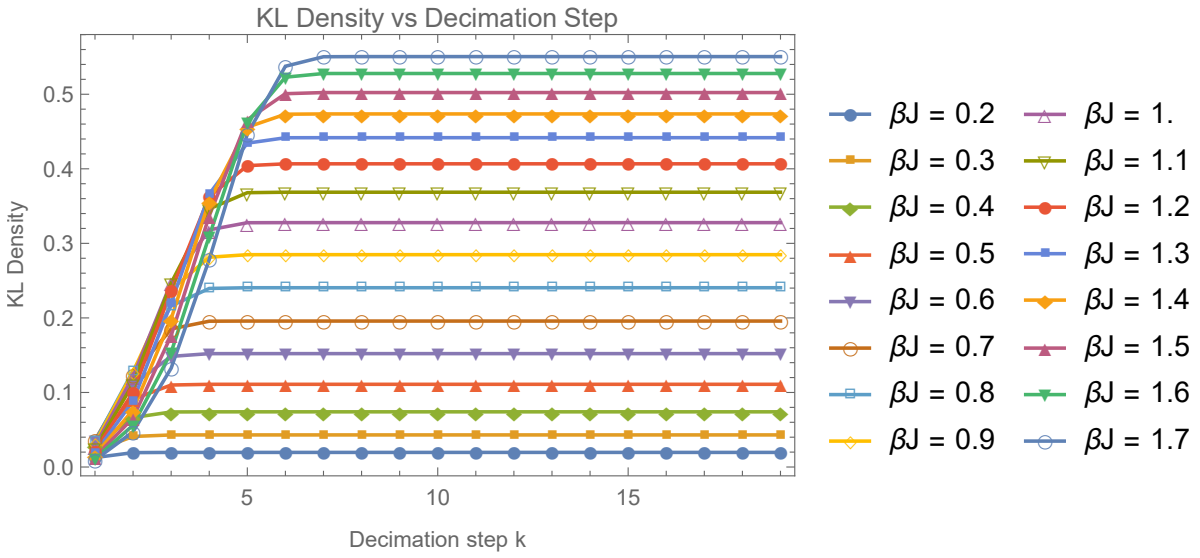

Figure 8.5: The KL density for the 1D Ising model as a function of decimation step for fixed, low values of βJ. The distribution q is formed from decimation plus rescaling. There is no critical value for the 1D Ising model, but we cover the parameter space that includes the critical value of the 2D Ising model.

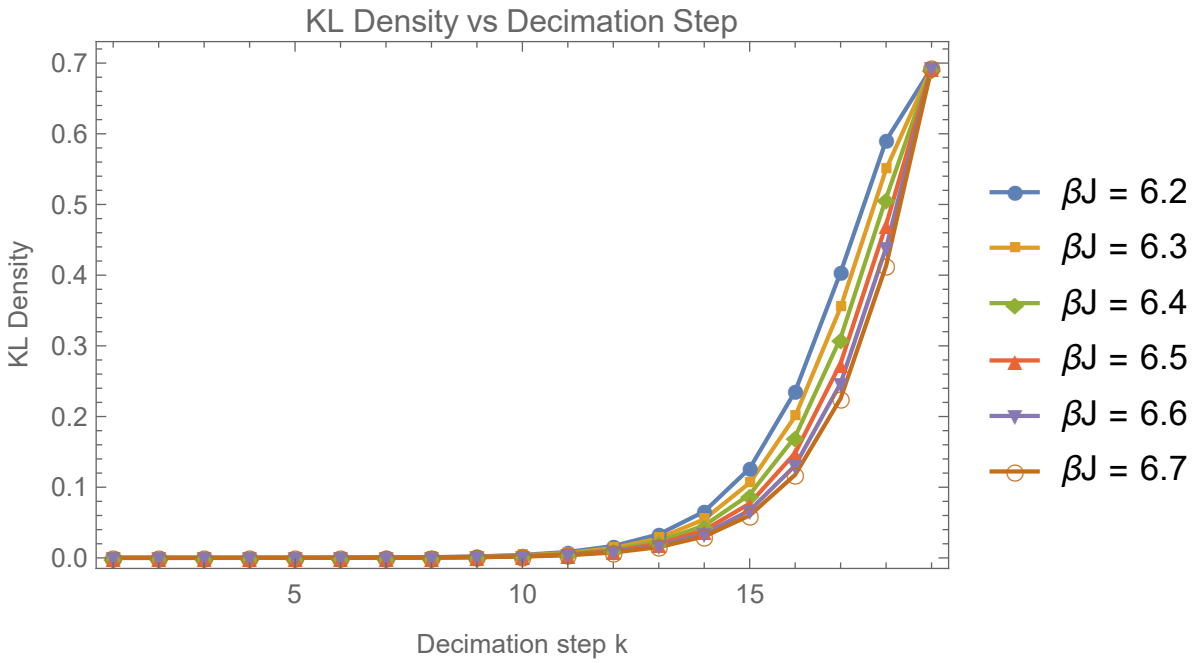

Figure 8.6: The KL density for the 1D Ising model as a function of decimation step for fixed, high values of βJ. The distribution q is formed from decimation plus rescaling. We start to see the curves converge on the final curve shown in this figure.

8.2 2D Decimation

We now calculate the relative entropy for 2D Ising models. We have a square, zero-field 2D Ising model on a plane (no periodic boundary conditions). The model has the same coupling strength in both directions: $-H = \beta \sum_{n.n.} \sigma_i \sigma_j$, where β is the coupling as before in Section 5. Decimating once produces a Hamiltonian of the form,

$$-H_1 = K_1 \sum_{n.n.} \sigma_i \sigma_j + L_1 \sum_{n.n.n} \sigma_i \sigma_j + M_1 \sum_{sq.} \sigma_i \sigma_j \sigma_k \sigma_l + C. \tag{8.6}$$

Note that each nearest neighbor pair is a nearest neighbor pair for two different lattice sites that are summed over when decimating. The Hamiltonian we have written is the final result for the decimation process, rather than the result for decimating one lattice site. Hence, the coupling K_1 is what appears in the partition function, rather than something half its value $K_1/2$.

This Hamiltonian is not of the same form we started with, so in order to iterate the decimation procedure, we will drop the quartic terms in the Hamiltonian that arise from decimation. This approximation is justified because the quartic term is irrelevant at long distances. However, as will be seen, this approximation slightly changes the location of the fixed point in the RG flow. We will also approximate the next-nearest neighbor interactions by re-writing them as nearest neighbor interactions and rewriting K_1 as $\beta_1 = K_1 + L_1$.

By the Kramers-Wannier duality between low and high temperature Ising models [145], the critical point is determined by,

$$e^{-2K_c} = \tanh K_c, \tag{8.7}$$

which produces a critical value of $K_c = (1/2)\log(1+\sqrt{2}) = 0.4407$. This agrees with the value Onsanger found in his exact solution for the 2D Ising model. Our approximation slightly changes the value of this fixed point to $\beta J = 0.50698$. This is calculated by finding the fixed point of the coupling constant, using the recursion relation (8.9).

We have in this approximation (the below is recorded in Pathria's *Statistical Mechanics* [146]),

$$\begin{aligned}
\log Z_{approx} &= N \log\left(\sqrt{2}\cosh(2\beta J)\right) + \frac{N}{\pi}\int_0^{\pi/2} \mathrm{d}\phi \ \log\left(1+\sqrt{1-\kappa^2\sin^2\phi}\right), \\
\kappa &= \frac{2\sinh(2\beta J)}{\cosh^2(2\beta J)}, \\
E_0 = E_{approx} = \langle H\rangle &= -NJ\coth(2\beta J)\left(1+\frac{2}{\pi}K(\kappa)(2\tanh^2(2\beta J)-1)\right),
\end{aligned} \tag{8.8}$$

and the recursion relation for the coupling constant is (from [125]),

$$\beta_{even} = \beta_{odd} = (3/8)\log(\cosh(4\beta J)), \tag{8.9}$$

where K is the complete elliptic integral of the first kind. Running through numerical values of βJ, we can calculate the relative entropy density and beta function of the relative entropy density for the 2D model as was done with the 1D model (number of sites is now $N = 2^{19}2^{19}$). Because there are no periodic boundary conditions, translation invariance is an approximation due to the large number of lattice sites.

We show the plots of the relative entropy density and the beta function of the relative entropy density in Figures 8.7-8.9.

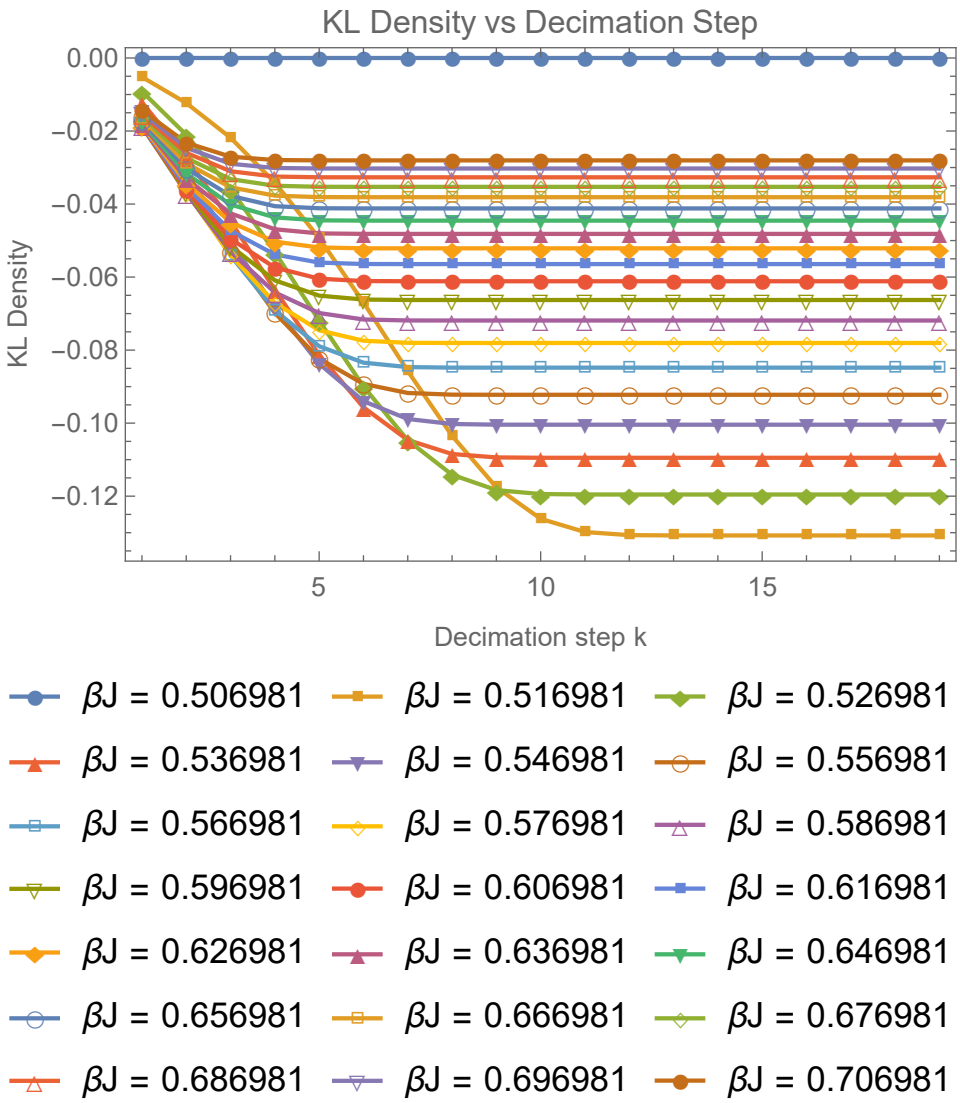

Figure 8.7: The KL density for the 2D Ising model as a function of decimation step for fixed values of βJ above the critical value ≈ 0.50698.

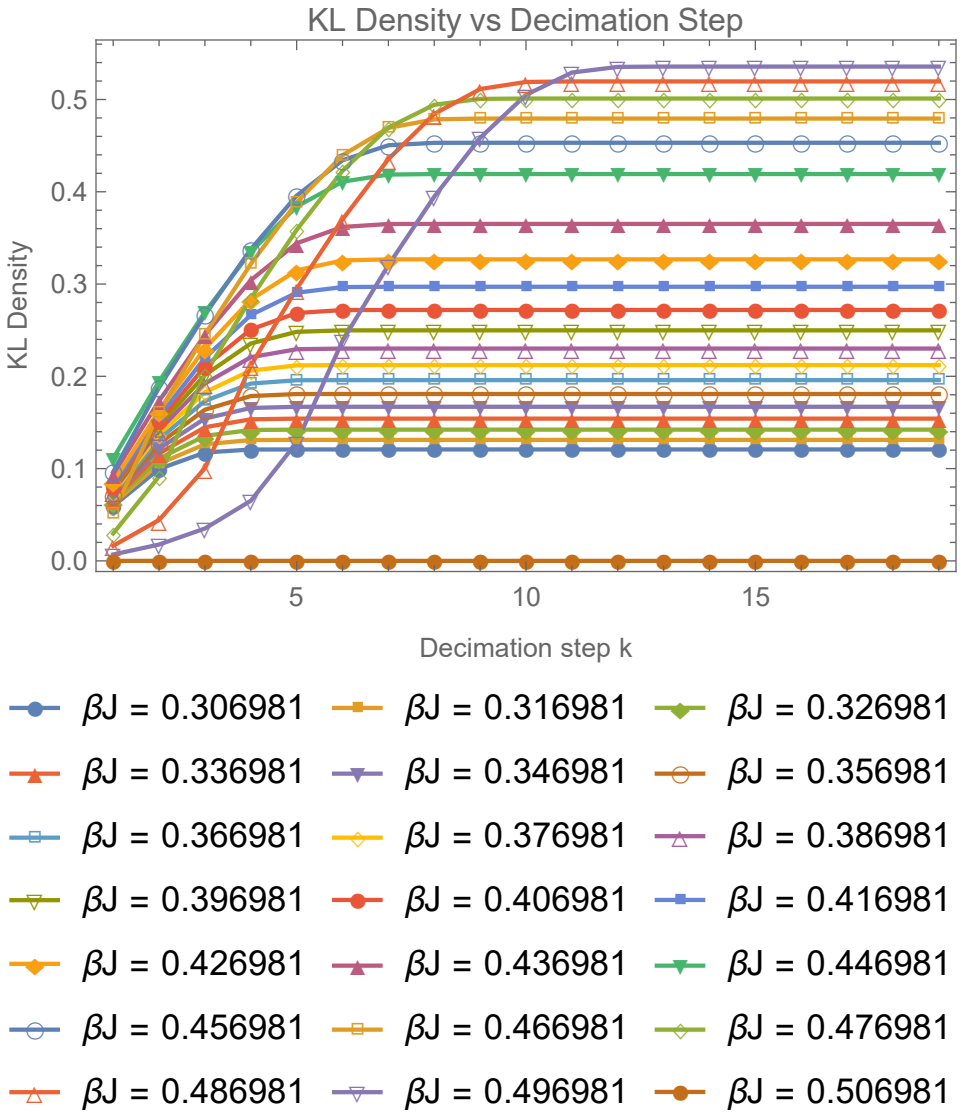

Figure 8.8: The KL density for the 2D Ising model as a function of decimation step for fixed values of βJ below the critical value ≈ 0.50698.

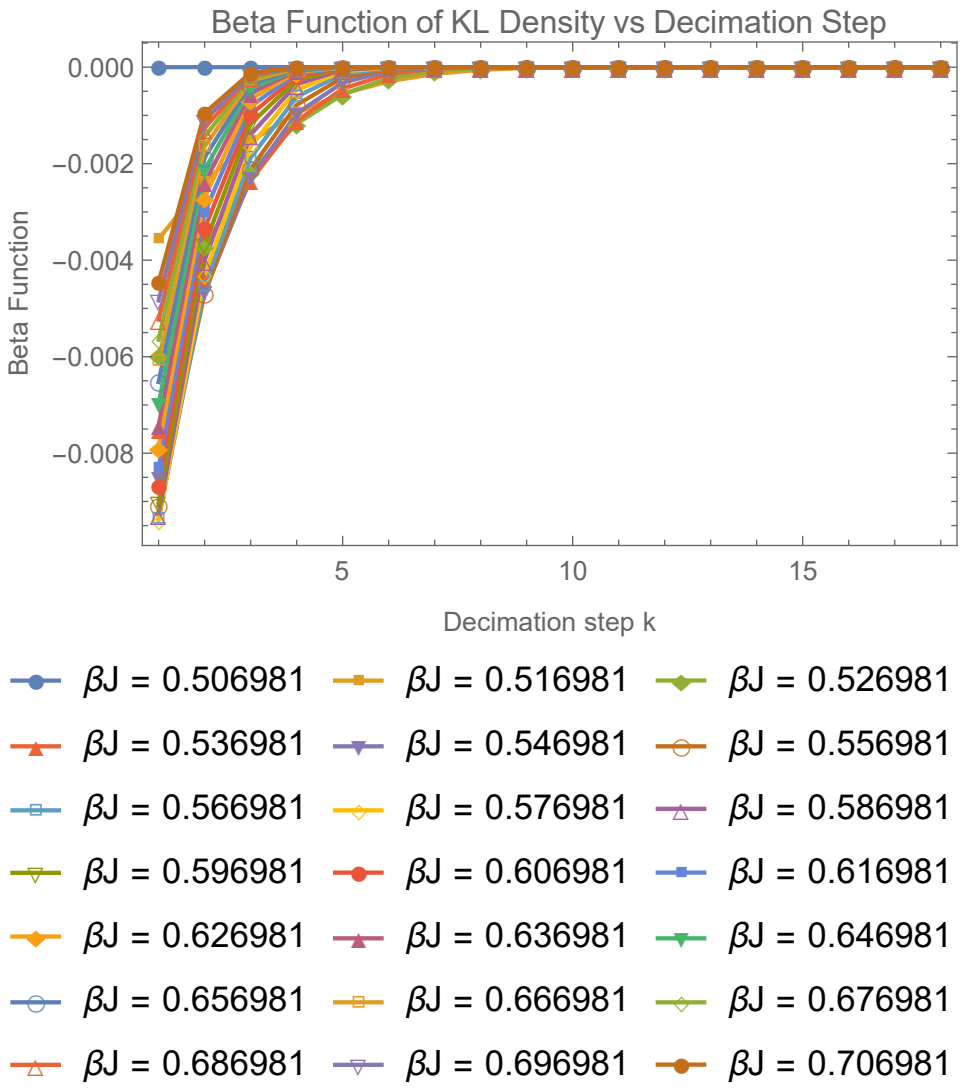

Figure 8.9: The beta function of the KL density for the 2D Ising model as a function of decimation step for fixed values of βJ above the critical value ≈ 0.50698.

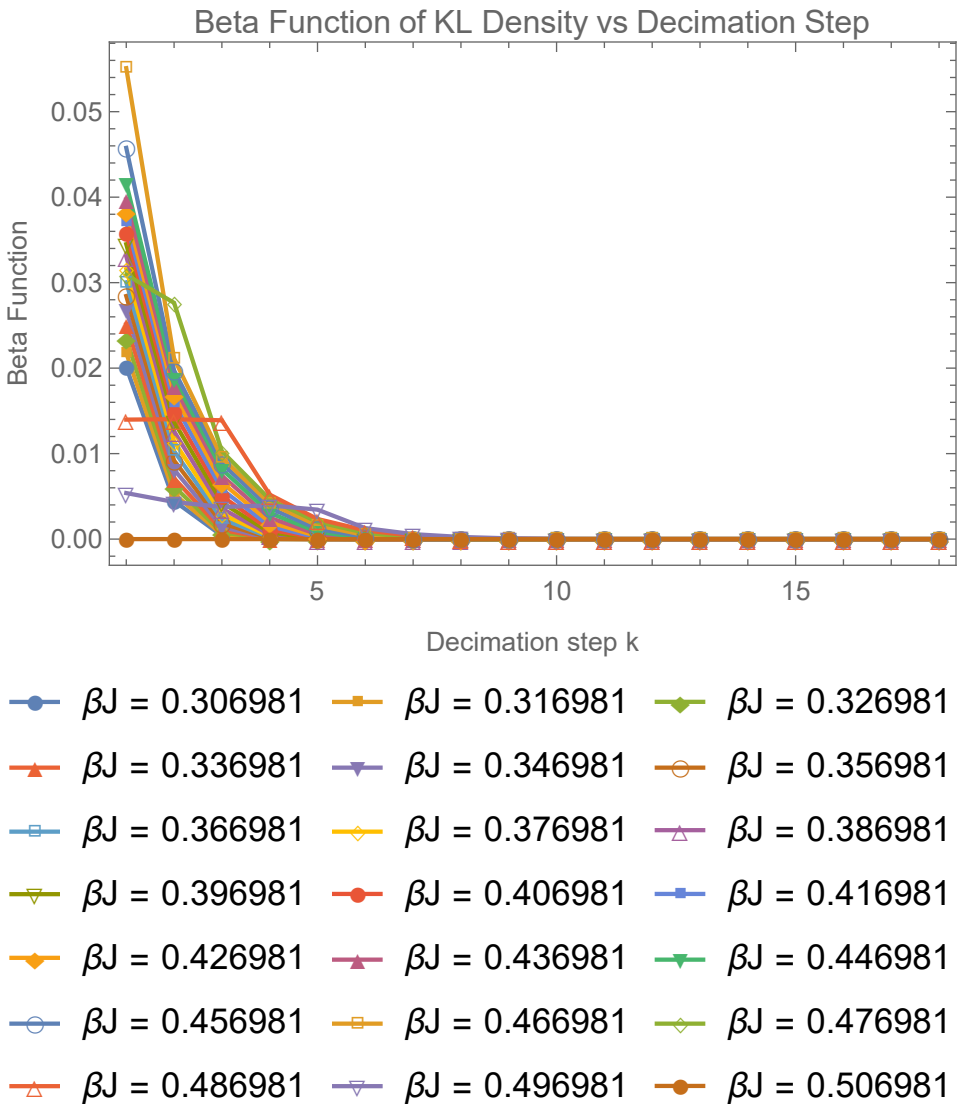

Figure 8.10: The beta function of the KL density for the 2D Ising model as a function of decimation step for fixed values of βJ below the critical value ≈ 0.50698.

As can be seen, in calculating the relative entropy density, we find the KL divergence is negative above the critical value (Figure 8.7), although it eventually drops to zero for large enough values of βJ or decimation step (above the critical point, βJ_k increases with decimation step). See the plots of the relative entropy density versus βJ in Figures 8.11 - 8.12. There is a spike near the critical point, which is due to the growth of the correlation length near the critical point. The growth in correlation length indicates that there is a greater sensitivity to all the operators that are brought down by the RG flow; the theories are thereby more distinct. Hence, there is a spike before being forced down to zero (at the critical point, the theories are invariant under RG flow, so the KL divergence for them must be zero).

The KL divergence is negative above the critical point because of neglecting the quartic and higher order irrelevant operators, as will be seen below.

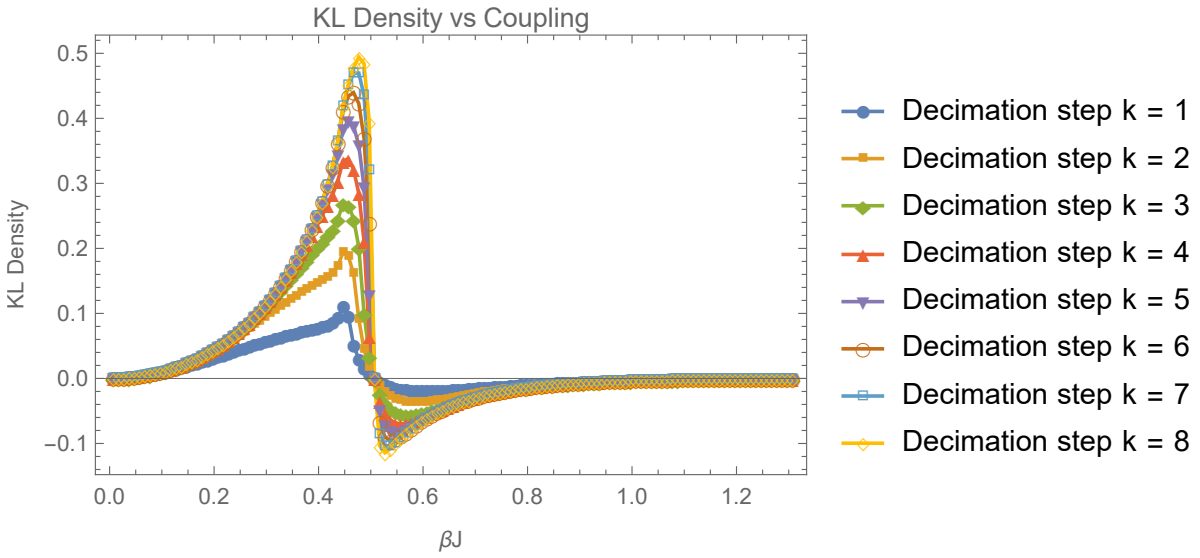

Figure 8.11: The KL density for the 2D Ising model as a function of the coupling βJ for various decimation steps. Only a few steps are shown because the curves merge to the exact same curve around decimation step 8, i.e., the KL density stops significantly changing after each decimation step after around eight decimation steps. The critical value is ≈ 0.50698.

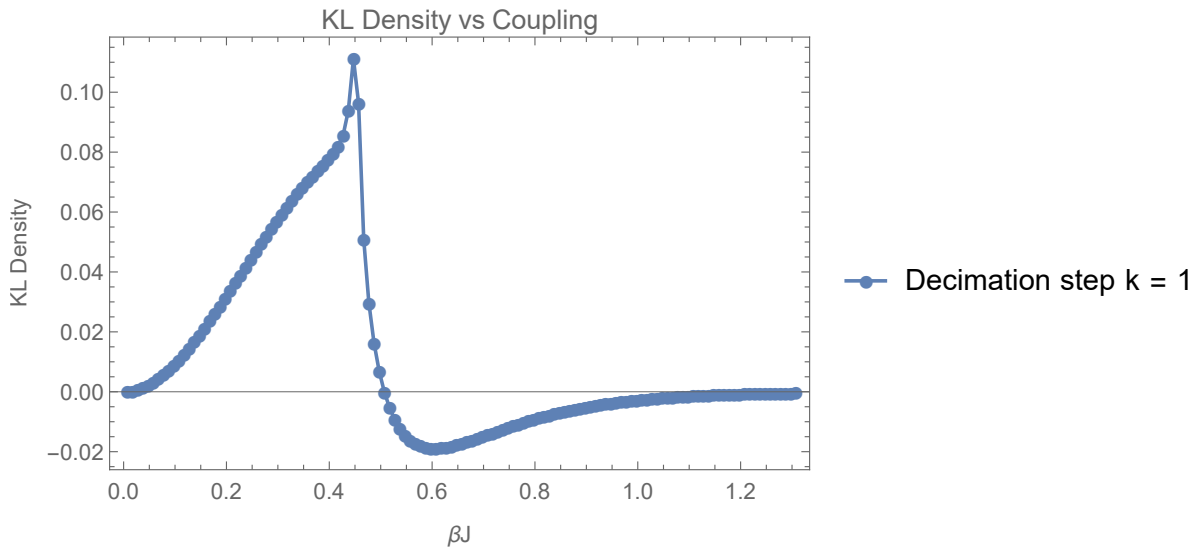

Figure 8.12: The KL density for the 2D Ising model as a function of the coupling βJ after one decimation step. The critical value is ≈ 0.50698.

First, note that we get negative values above the critical point even after one decimation step (Figure 8.12), where we only have a quartic term to drop. So the negative values cannot be blamed on the quartic interaction or higher interactions changing the recursion relation for the coupling constant: the only approximation for the recursion relation made at this step is combining the nearest and next-nearest neighbor interactions.

Next, the negative values are not a result of the distributions not being properly normalized. The exact thing that is plotted is,

$$\begin{aligned} D_{KL} =& 2^k \log Z_{approx}(N/2^k, J_k) - \log Z_{approx}(N, J) \\ &+ 2^k E_{approx}(N/2^k, J_k) - E_{approx}(N, J). \end{aligned} \tag{8.10}$$

The partition function and average energies come from two normalized distributions. One of them is a normalized distribution for a 2D square Ising model with N sites and coupling J. The other is also a square 2D Ising model with $N/2^k$ sites and coupling constant J_k. The procedure we use in general to form the distributions for the KL divergence is to 1) decimate the Hamiltonian, 2) drop the quartic and higher interactions, 3) define the partition function in terms of the decimated Hamiltonian. We can see that step 3) produces a partition function given by the approximation recorded in Pathria's book by inspecting the resulting partition function. Our distributions are therefore automatically normalized, and in particular, they are normalized distributions when using these approximations recorded in Pathria's book as done above in (8.10).

If we wished, we could have kept the quartic interaction as a perturbation around the approximation recorded in Pathria's book. The linear terms would have cancelled, leaving a lowest order positive contribution at quadratic order that (8.10) drops. However, this cannot explain the negative values because the distributions in (8.10) are properly normalized: even if the KL divergence was wildly inaccurate from dropping the quartic interaction contribution at quadratic order, the KL divergence should still be positive everywhere.

The reason for the negative values instead comes from approximating the expectation value for the decimated Hamiltonian. This is the only other place where an approximation has been made (the thermodynamic limit is also an approximation, but we used large N in our calculations for the Figures, and higher order corrections should be small, being suppressed by factors of N), and this approximation

can result in negative values even for normalized distributions because the approximation is made to the KL divergence, not the distributions. So we end up with an approximate KL divergence instead of the exact KL divergence between approximated distributions.

The exact expectation value is with respect to the undecimated theory, but by using the approximation recorded in Pathria's book to evaluate the expectation value, we are assuming that we can evaluate the expectation value with respect to the decimated theory and get the same results. This is not the case: evaluating the expectation value with respect to the decimated theory requires us to drop quartic interactions.

Defining $H_k(\sigma^2)$ to be the decimated Hamiltonian after dropping the quartic and higher interactions, leaving just the quadratic and lower; H_{2D} to be the Hamiltonian of the square 2D Ising model, which is the Hamiltonian approximation Pathria uses to calculate the expectation values recorded in his book; $H_k(\sigma^4)$ to be the part of the decimated Hamiltonian that includes all the quartic and higher interactions $H_k = H_k(\sigma^2) + H_k(\sigma^4)$; C the constant energy shift resulting from decimation; we can see that,

$$\begin{aligned}
\langle H_k(\sigma^2)\rangle = \langle H_{2D}(J_k, N_k)\rangle &= \frac{1}{Z}\int [d\sigma] H_{2D}(J_k, N_k)\exp(-H_{2D}(J, N)) \\
&= \frac{1}{Z}\int [d\sigma^k] H_{2D}(J_k, N_k)\exp(-H_{2D}(J_k, N_k))\exp(-H_k(\sigma^4))\exp(C) \\
&= \frac{1}{Z_k}\int [d\sigma^k] H_{2D}(J_k, N_k)\exp(-H_{2D}(J_k, N_k))\exp(-H_k(\sigma^4)),
\end{aligned} \tag{8.11}$$

where $[d\sigma^k]$ indicates integrating over the spin variables that appear at the kth decimation step. From this equation, we see that we must drop the quartic and higher interactions in order for this expression to be equivalent to taking an expectation value with respect to the decimated distribution. Note that the quartic and higher terms need to be dropped from the partition function too to evaluate with respect to the undecimated theory,

$$\begin{aligned}
Z_k &= \int [d\sigma^k]\exp(-H_{2D}(J_k, N_k))\exp(-H_k(\sigma^4)) \\
&= Z_k(\sigma^2)\frac{1}{Z_k(\sigma^2)}\int [d\sigma^k]\exp(-H_{2D}(J_k, N_k))(1 - H_k(\sigma^4) + ...) \\
&\equiv Z_{2D}(J_k, N_k)\frac{1}{Z_{2D}(J_k, N_k)}\int [d\sigma^k]\exp(-H_{2D}(J_k, N_k))(1 - H_k(\sigma^4) + ...) \\
&= Z_{2D}(J_k, N_k)(1 - \langle H_k(\sigma^4)\rangle_k + ...),
\end{aligned} \tag{8.12}$$

where Z_{2D} is the partition function for the square 2D Ising model. An expression for its logarithm $\log Z_{approx} \equiv \log Z_{2D}$ can be found in Pathria's book. Our final result to lowest order is,

$$\begin{aligned}
\langle H_{2D}(J_k, N_k)\rangle &= \frac{1}{Z_k}\int [d\sigma^k] H_{2D}(J_k, N_k)\exp(-H_{2D}(J_k, N_k))\exp(-H_k(\sigma^4)) \\
&= \frac{1}{Z_{2D}(J_k, N_k)(1-\langle H_k(\sigma^4)\rangle_k + ...)} \\
&\times \int [d\sigma^k] H_{2D}(J_k, N_k)\exp(-H_{2D}(J_k, N_k))(1 - H_k(\sigma^4) + ...) \\
&= \frac{1}{Z_{2D}(J_k, N_k)}\int [d\sigma^k] H_{2D}(J_k, N_k)\exp(-H_{2D}(J_k, N_k)) \\
&\times (1 - H_k(\sigma^4) + ...)(1 + \langle H_k(\sigma^4)\rangle_k + ...) \\
&= \langle H_{2D}(J_k, N_k)\rangle_k + \langle H_{2D}(J_k, N_k)\rangle_k \langle H_k(\sigma^4)\rangle_k - \langle H_{2D}(J_k, N_k) H_k(\sigma^4)\rangle_k + ... \\
&= E_{approx}(J_k, N_k) + E_{approx}(J_k, N_k)\langle H_k(\sigma^4)\rangle_k - \langle H_{2D}(J_k, N_k) H_k(\sigma^4)\rangle_k + ... \\
&= E_{approx}(J_k, N_k) \\
&- \langle (H_{2D}(J_k, N_k) - \langle H_{2D}(J_k, N_k)\rangle_k)(H_k(\sigma^4) - \langle H_k(\sigma^4)\rangle_k)\rangle_k + ...
\end{aligned} \tag{8.13}$$

Plugging in the spin operators for the 2D model, and defining M_k to be the quartic coupling at the kth decimation step, we get,

$$\begin{aligned}
\langle H_{2D}(J_k, N_k)\rangle &= E_{approx}(J_k, N_k) + J_k M_k \langle \sum_{n.n.} \sigma^2 \rangle_k \langle \sum_{pl} \sigma^4 \rangle_k - J_k M_k \langle \sum_{n.n.} \sigma^2 \sum_{pl} \sigma^4 \rangle_k + ... \\
&= E_{approx}(J_k, N_k) + J_k M_k \langle (\sum_{n.n.} \sigma^2 - \langle \sum_{n.n.} \sigma^2 \rangle_k)(\sum_{pl} \sigma^4 - \langle \sum_{pl} \sigma^4 \rangle_k)\rangle_k,
\end{aligned} \tag{8.14}$$

where we calculate M_k assuming only the quadratic coupling matters in calculating it, and both $M_k < 0$ and $E_{approx} < 0$, while $J_k > 0$. We use $\sum_{pl}$ to indicate summing over all the plaquettes of four-spin interactions, and $\sum_{n.n.}$ indicates summing over nearest neighbors. Note that the last term is a connected Green's function, which could also be seen in the last line of (8.13). Although M_k is approximate, recall that the problem of a negative KL divergence occurs even with one decimation step, and we know M_1 exactly.

Near the critical point, $M_k \approx -0.056 \approx -J_k/10$. So we learn that (1) the order $M_k^2 \approx J_k^2/100$ terms can safely be neglected relative to the quadratic coupling, and (2) the terms of order $J_k M_k \approx -J_k^2/10$. Just above the critical point, the KL divergence is small: less than -0.056 (approximately -0.0052). The

biggest the KL divergence gets is around -0.02 (during which $M_k \approx -0.08$ and J_k is still roughly 10 times M_k in magnitude) so the terms are of the right order to make the KL divergence non-negative. Also, M_k gets drastically bigger with decimation step and with initial coupling constant, getting to 1/3 of J_k in magnitude. So although the correlation functions will get smaller with decimation step (they are larger at first, so not much smaller than 1), it should be compensated by the growth of M_k.

Hence, we can expect the magnitude of everything to be the correct size to make the KL divergence non-negative in the appropriate regions (the KL divergence becomes flatly zero at a large enough coupling ≈ 2.5 because at large coupling, $E_{approx} = -\log Z_{approx}$).

Furthermore, because we know that the only irrelevant operator is quartic at one decimation step, we can have some idea of the size of its contribution to the KL divergence from the connected Green's function. It is possible that other irrelevant operators make a significant contribution for greater decimation steps, and the quartic coupling changes with decimation step, so we cannot find the size of the contribution of any particular irrelevant operators at higher decimation steps. Because the quartic coupling changes with decimation step, we also cannot determine an ordering of which irrelevant operators make a greater or smaller contribution.

8.2.1 2D Decimation: Rescaling

As a further confirmation that the negative values of the KL divergence come from higher order terms that are neglected in taking the expectation value of the decimated Hamiltonian, we will calculate the KL divergence for the 2D Ising model using decimation plus rescaling. This requires us to drop quartic and higher interactions in the Hamiltonian (and therefore partition function too), but we can evaluate the expectation value of the decimated Hamiltonian exactly. We shall see that the KL divergence is non-negative in this case.

For decimation plus rescaling, the rescaled Hamiltonian is (rescaling performed after dropping the quartic and higher interactions and combining the nearest and next-nearest neighbor interactions as usual),

$$H_k = -J_k \sum_{n.n.} \sigma_i \sigma_j, \tag{8.15}$$

where the nearest neighbor sum goes over N sites. That is, it is the exact same Hamiltonian as the N-site 2D Ising model with a different coupling constant. We now evaluate the KL divergence as (the partition

function drops quartic and higher interactions so as to normalize H_k in (8.15)),

$$D_{KL}(p||q) = \log Z_k - \log Z + (-\langle H \rangle + \langle H_k \rangle)\,, \tag{8.16}$$

where we have,

$$\langle H_k \rangle = \frac{J_k}{J} \langle H \rangle. \tag{8.17}$$

This equation (8.17) follows because we can just pull the coupling constant J out of the integral and evaluate the correlation function and then substitute in the J_k coupling constant from the decimation procedure. This is just the same as multiplying the expectation value of the original Hamiltonian by J_k/J. Notice also in (8.16) that we have no factors of 2^k because there is no need to multiply copies of the decimated distribution together: the rescaled Hamiltonian has the same spin support as the original Hamiltonian.

The factors of 2^k do not make a difference anyway: $\log Z_k$ and E_k are proportional to the number of sites, so the 2^k cancels with the $N/2^k$ in front of these terms, whenever we do this decimation procedure (which makes the number of sites $N/2^k$).

We then substitute in the values of the free energy and internal energy as before ($\log Z_k$ must $= \log Z_{approx}$ because it is just normalizing a square 2D Ising model with a coupling J_k) to get,

$$\begin{aligned} D_{KL}(p||q) &= \log Z_k - \log Z + \left(\frac{J_k}{J} - 1\right) \langle H \rangle \\ &= \log Z_{approx}(N, J_k) - \log Z_{approx}(N, J) + \left(\frac{J_k}{J} - 1\right) E_{approx}(N, J) \end{aligned} \tag{8.18}$$

The results are shown in the below Figures 8.13-8.14. Notice the similarities to the earlier results when using decimation plus forming a joint distribution for the values below the critical value.

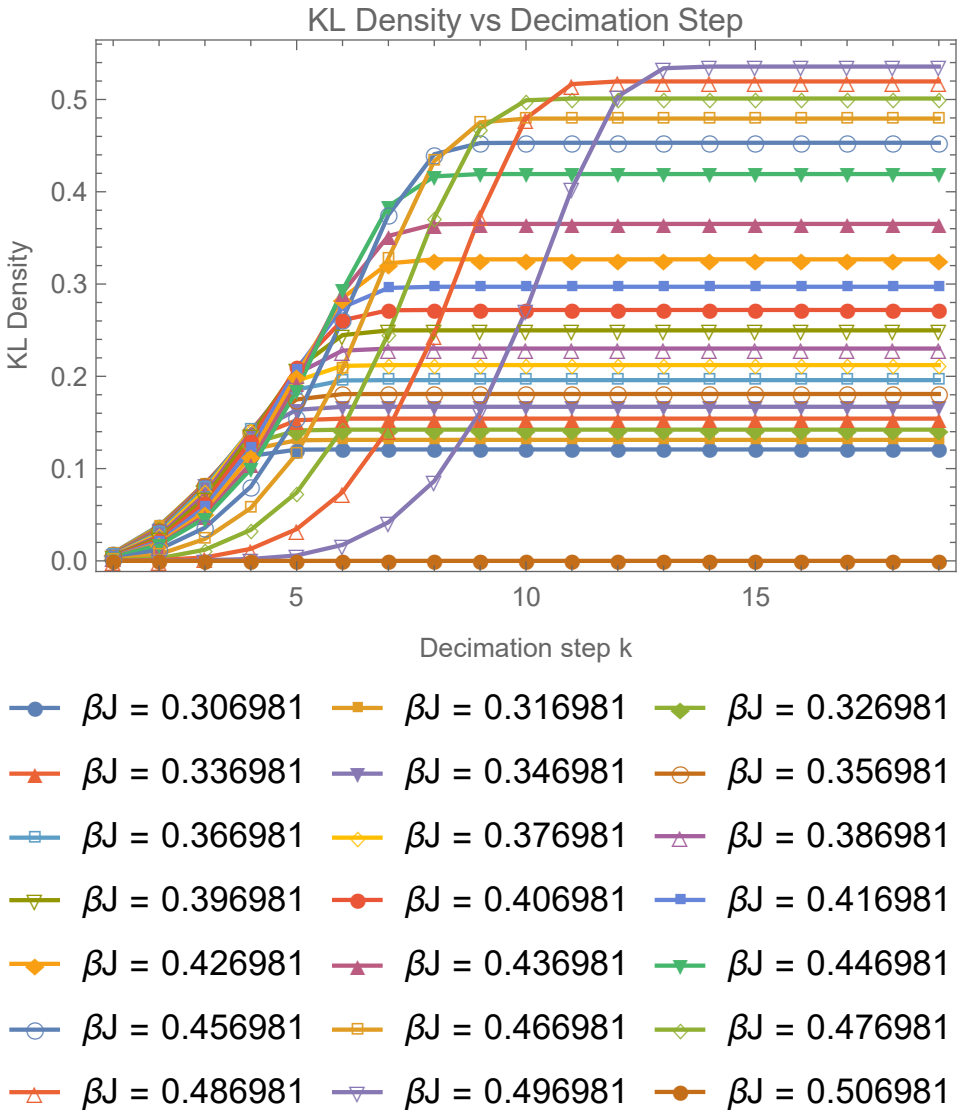

Figure 8.13: The KL density for the 2D Ising model as a function of decimation step for fixed values of βJ below the critical value of ≈ 0.50698. The distribution q is formed from decimation plus rescaling. There are similarities to the results when using decimation plus forming a joint distribution for values below the critical value.

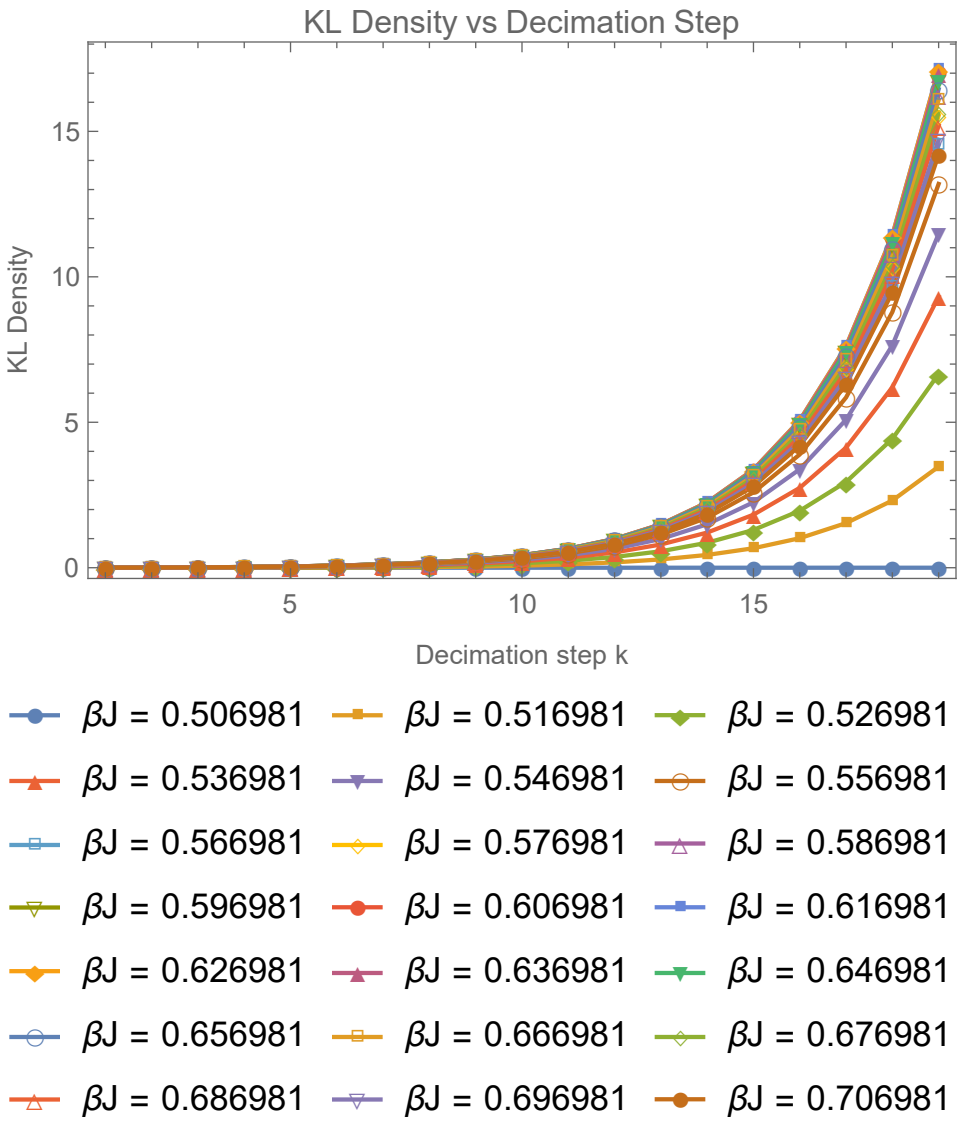

Figure 8.14: The KL density for the 2D Ising model as a function of decimation step for fixed values of βJ above the critical value of ≈ 0.50698. The distribution q is formed from decimation plus rescaling. These are very different results from when using decimation plus forming a joint distribution: notably, the curves are non-negative in the present case.

As we see, the KL divergence is non-negative, as it should be, thus confirming that the negative values from before must be coming from approximating the expectation value of the decimated Hamiltonian.

8.2.2 2D Decimation: Mean Field Theory

Because the value of the KL divergence was negative for some values of couplings, we need to make a better approximation. We will accomplish this by taking the quartic interaction into account, using mean field theory to evaluate $\langle H_k \rangle$ and $\ln Z_k$. Because thinking in terms of a constant energy shift is equivalent to thinking in terms of a changing partition function, we will work in the picture that does not have a constant energy shift but lets the partition function change with decimation step. This will make our calculations easier.

For the decimated Hamiltonian, we get ($\beta = 1$, σ^k the spins remaining at decimation step k after some chosen decimation path),

$$\begin{aligned} \langle H_k \rangle &\equiv \frac{1}{Z} \int [d\sigma] H_k \exp(-H) \\ &= \frac{1}{Z} \int [d\sigma^k] H_k \exp(-H_k), \end{aligned} \tag{8.19}$$

where,

$$H_k = -\left(J_{1,k} \sum_{n.n} \sigma_i^k \sigma_j^k + J_{2,k} \sum_{n.n.n} \sigma_i^k \sigma_j^k + J_{3,k} \sum_{plaq.} \sigma_i^k \sigma_j^k \sigma_k^k \sigma_l^k \right). \tag{8.20}$$

Mean field theory then gives,

$$\begin{aligned} H_k &= -J_{1,k} zm \sum_i \sigma_i^k - J_{2,k} zm \sum_i \sigma_i^k - J_{3,k} zm^3 \sum_i \sigma_i^k \\ &= -(J_{1,k} + J_{2,k}) zm \sum_i \sigma_i^k - J_{3,k} zm^3 \sum_i \sigma_i^k \\ &\equiv -J_k zm \sum_i \sigma_i^k - J_{3,k} zm^3 \sum_i \sigma_i^k \end{aligned} \tag{8.21}$$

where $z = 2D = 4$ ($D =$ dimension of the lattice) is the number of neighboring sites of the spin σ_i^k and we define $J_k \equiv J_{1,k} + J_{2,k}$ to simplify notation. Notice that the magnetization m is the magnetization of the system at decimation step k. Furthermore, we have made an assumption that terms of higher order than the quartic can be neglected, so that each time we decimate, we get a Hamiltonian of the above form. Also, we have assumed that any corrections to the recursion relations can be neglected and that the

quadratic term dominates, so that we still have,

$$\begin{aligned} J_k &= \frac{1}{4}\log\left(\cosh 4J_{k-1}\right) + \frac{1}{8}\log\left(\cosh 4J_{k-1}\right) \\ &= \frac{3}{8}\log(\cosh(4\beta J_{k-1})), \\ J_{3,k} &= \frac{1}{8}\log\left(\cosh 4J_{k-1}\right) - \frac{1}{2}\log\left(\cosh 2J_{k-1}\right). \end{aligned} \tag{8.22}$$

We can then calculate expectation values in this theory as,

$$\begin{aligned} \langle H_k \rangle &= \frac{1}{Z}\int [d\sigma^k] H_k \exp(-H_k) \\ &= -\frac{1}{Z}\int [d\sigma^k] J_k z m \sum_i \sigma_i^k \exp(-H_k) \\ &- \frac{1}{Z}\int [d\sigma^k] J_{3,k} z m^3 \sum_i \sigma_i^k \exp(-H_k) \\ &= -Nz\left(J_k m^2 + J_{3,k} m^4\right), \end{aligned} \tag{8.23}$$

where $N = n/2^k$ is the number of spin sites in our decimated theory with $n =$ total number of spins in the undecimated theory. We now calculate the magnetization in the decimated theory. Notice that, relative to the mean field theory of a 2D Ising model, the effective magnetic field just gets an additional term from m^3. Hence, the magnetization in our theory is the same as that in the usual 2D Ising model but with this extra term in the effective magnetic field. The magnetization then is,

$$m = \tanh(z\left(J_k m + J_{3,k} m^3\right)). \tag{8.24}$$

We can then proceed to find solutions to this equation graphically in the usual manner. The critical temperature can be found by expanding near $m = 0$,

$$\begin{aligned} m &= (z\left(J_k m + J_{3,k} m^3\right) - \frac{(zJ_k)^3}{3} m^3 \\ &\rightarrow m = 0, m = \pm \frac{\sqrt{3}\sqrt{-1 + zJ}}{\sqrt{z^3 (J_k)^3 - 3zJ_{3,k}}}. \end{aligned} \tag{8.25}$$

We see then that we need $zJ_k > 1$ and $z^3(J_k)^3 > 3zJ_{3,k}$ in order for a real non-zero solution to exist. In our case, $J_{3,k} \leq 0$ for all values, while we are working with a positive J_k, so the second condition is

automatically satisfied. So we must have $m = 0$ for whatever values of $\beta J_k, \beta J_{3,k}$ that do not satisfy the conditions. In particular, the first condition shows that $m = 0$ for small enough values of βJ_k.

For very large values of βJ_k, the $\tanh$ goes to 1, so $m = 1$ in this limit.

The expression for $\log Z_k$ is found similarly. We have,

$$\begin{aligned} Z_k &= \int [d\sigma^k] \exp(-H_k) \\ &= \prod_i^N \int [d\sigma^k] \exp(z \left(J_k m + J_{3,k} m^3\right) \sigma_i^k) \\ &= \left(2\cosh(z(J_k m + J_{3,k} m^3))\right)^N \\ &\to \log Z_k = N \log \left(2\cosh(z(J_k m + J_{3,k} m^3))\right), \end{aligned} \tag{8.26}$$

where since this is the decimated theory, as with finding the expectation value of H_k, we have $N = n/2^k$ and m is the magnetization in the decimated theory.

So in Mathematica, for small J_k, we set $m = 0$; for large J_k we set $m = 1$, and for in between we numerically solve the equation with the $\tanh$. We can then plug this value of magnetization into our expressions for $\langle H_k \rangle$ and $\log Z_k$ and calculate the KL divergence. This will be reserved for future work.

8.3 (1 + epsilon) Dimensions

We now consider an Ising model in $1 + \varepsilon$ dimensions. To do this, we take a 2D asymmetric Ising model with constant couplings in the x-direction much stronger than the constant couplings in the y-direction, $J_x >> J_y$. We use perturbation theory to calculate the KL divergence at each decimation step to lowest order, i.e., quadratic in J_y. Because J_y is much smaller than J_x, the probability distribution q can be treated as a stack of 1D Ising models with perturbations that connect the models in the y-direction,

$$H_q = H_p + \delta H = H_{1D} + \delta H = \sum_{i,j} J\sigma_{i,j}\sigma_{i+1,j} + \delta \sum_{i,j} \sigma_{i,j}\sigma_{i,j+1}, \tag{8.27}$$

where we have relabeled J_x as J and J_y as δ to emphasize that the term is being treated as a perturbation. More generally, suppose we have a probability distribution p and a distribution q that is the distribution p plus a sum of operators O that are perturbations $\Theta(x) = \lambda^i(x) O_i(x)$, where λ is taken to be small. The

partition function Z_q is,

$$\begin{aligned} Z_q &= \int \mathrm{D}\mu \; e^{-\beta H_p - \beta\Theta(x)} \\ &= \int \mathrm{D}\mu \; e^{-\beta H_p} \sum_{m=0}^{\infty} \frac{(-\beta\Theta(x))^m}{(m)!} \\ &\approx \int \mathrm{D}\mu \; \frac{e^{-\beta H_p}}{Z_p} Z_p \left(1 - \beta\Theta(x) + \frac{\beta^2}{2}\Theta(x)\Theta(y)\right) \\ &= Z_p \left(1 - \beta\langle\Theta(x)\rangle_p + \frac{\beta^2}{2}\langle\Theta(x)\Theta(y)\rangle_p\right), \end{aligned} \tag{8.28}$$

where the partition function has been expanded to second order in the final line. The KL divergence is then,

$$\begin{aligned} D(p||q) &= \int \mathrm{D}\mu \; \frac{e^{-\beta H_p}}{Z_p} \left(\log\left(\frac{Z_q}{Z_p}\right) + \beta\Theta(x)\right) \\ &= \log\left(1 - \beta\langle\Theta(x)\rangle_p + \frac{\beta^2}{2}\langle\Theta(x)\Theta(y)\rangle_p - ...\right) + \beta\langle\Theta(x)\rangle_p \\ &\approx \frac{\beta^2}{2}\left(\langle\Theta(x)\Theta(y)\rangle_p - \langle\Theta(x)\rangle_p\langle\Theta(y)\rangle_p\right), \end{aligned} \tag{8.29}$$

where the partition function and logarithm has been expanded to second order. In the case of the $(1+\varepsilon)$-D model we are considering, we have $\Theta(x) = \delta\sum_{i,j}\sigma_{i,j}\sigma_{i,j+1}$, so to lowest order, the KL divergence is,

$$D(p||q) = \frac{(\beta\delta)^2}{2}\sum_{i,j}\sum_{i',j'}\left(\langle\sigma_{i,j}\sigma_{i,j+1}\sigma_{i',j'}\sigma_{i',j'+1}\rangle_p - \langle\sigma_{i,j}\sigma_{i,j+1}\rangle_p\langle\sigma_{i',j'}\sigma_{i',j'+1}\rangle_p\right), \tag{8.30}$$

where an exact expression for the KL divergence cannot be found because of uncertainty in rewriting the expanded exponential $\sum_m(\sum_{i,j}\sigma_{i,j}\sigma_{i,j+1})^m/m!$ in terms of something tractable.

As will be shown, because theory p is just a stack of 1D Ising models, the correlation functions can be re-written in terms of correlation functions for a 1D Ising model. First note that in general $\langle\sigma_{x,y}\sigma_{x',y'}\rangle_p = 0$ when $y \neq y'$. This is because the spins in different y-columns are uncorrelated in theory p. Theory p is just a sum of independent 1D Ising models; there is no coupling between spins at different y sites. So we are limited to correlation functions with the same y.

Next, note that the partition function for theory p is just a product of 1D partition functions because the Hamiltonian is a sum of independent 1D Ising models,

$$\begin{aligned} Z_p &= \int \mathrm{D}\sigma \; e^{-\beta J \sum_{i,j} \sigma_{i,j}\sigma_{i+1,j}} \\ &= \int \prod_j \mathrm{D}\sigma_j \; \prod_j e^{-\beta J \sum_i \sigma_{i,j}\sigma_{i+1,j}} \\ &= \prod_j \int \mathrm{D}\sigma_j \; e^{-\beta H_{1D,j}} = \prod_j Z_{1D,j}. \end{aligned} \tag{8.31}$$

We then have for the two-point function,

$$\begin{aligned} \langle \sigma_{x,y}\sigma_{x',y}\rangle_p &= \int \mathrm{D}\sigma \; \sigma_{x,y}\sigma_{x',y}\frac{e^{-\beta H_p}}{Z_p} \\ &= \frac{\prod_j \int \mathrm{D}\sigma_{j\neq y} \; e^{-\beta H_{1D,j}} \int \mathrm{D}\sigma_y \; \sigma_{x,y}\sigma_{x',y} e^{-\beta H_{1D,y}}}{\prod_{j\neq y} Z_{1D,j} Z_{1D,y}} \\ &= \int \mathrm{D}\sigma_y \; \sigma_{x,y}\sigma_{x',y}\frac{e^{-\beta H_{1D,y}}}{Z_{1D,y}} \\ &= \langle \sigma_x \sigma_{x'}\rangle_{1D}, \end{aligned} \tag{8.32}$$

and the 4-point function is similarly,

$$\begin{aligned} \langle \sigma_{x,y}\sigma_{x',y}\sigma_{x,y+1}\sigma_{x',y+1}\rangle_p &= \int \mathrm{D}\sigma \; \sigma_{x,y}\sigma_{x',y}\sigma_{x,y+1}\sigma_{x',y+1}\frac{e^{-\beta H_p}}{Z_p} \\ &= \frac{\int \mathrm{D}\sigma_{y+1} \; \sigma_{x,y+1}\sigma_{x',y+1} e^{-\beta H_{1D,y+1}} \int \mathrm{D}\sigma_y \; \sigma_{x,y}\sigma_{x',y} e^{-\beta H_{1D,y}}}{Z_{1D,y+1} Z_{1D,y}} \\ &= \langle \sigma_x \sigma_{x'}\rangle_{1D} \langle \sigma_x \sigma_{x'}\rangle_{1D}. \end{aligned} \tag{8.33}$$

The KL divergence to lowest order is then,

$$\begin{aligned} D(p||q) &= \frac{(\beta\delta)^2}{2} \sum_{i,j} \sum_{i',j} \left(\langle \sigma_i \sigma_{i'}\rangle_{1D}\right)^2 \\ &= \frac{(\beta\delta)^2}{2} N_y \sum_{i} \sum_{i'} \left(\langle \sigma_i \sigma_{i'}\rangle_{1D}\right)^2 \\ &= \frac{(\beta\delta)^2}{2} N_y \sum_{i} \sum_{i'} \left(e^{-2|i-i'|/\xi} e^{-2/\chi(i,i')}\right) \\ \frac{1}{\chi(i,i')} &= -\log \frac{\lambda_{p1}^{N_x} + \lambda_{p2}^{N_x} \frac{\lambda_{p1}}{\lambda_{p2}}^{2i-2i'}}{\lambda_{p1}^{N_x} + \lambda_{p2}^{N_x}}, \end{aligned} \tag{8.34}$$

where $N_{y,x}$ is the number of sites in the y(x)-direction, and for periodic boundary conditions, the correlation function was calculated in (7.31) and $1/\xi = \log[\lambda_{p1}/\lambda_{p2}] = \log[\coth(\beta J)]$. Although N_x and N_y are distinguished for generality, we only consider 2D Ising models with $N_x = N_y$. In the limit of large N_x, only the exponential factor with the ξ survives, and the correlation function does not depend on the boundary conditions, which makes the KL divergence become,

$$\begin{aligned} D(p||q)|_{N_x>>1} &= \frac{(\beta\delta)^2}{2} N_y \sum_i \sum_{i'} \left(e^{-2|i-i'|/\xi} \right) \\ &= \frac{N_y(\beta\delta)^2}{2} \frac{1-e^{-2N_x/\xi}}{1-e^{-2/\xi}}, \end{aligned} \tag{8.35}$$

where a shift of indices has occurred to turn the double sum on the exponential into a geometric series ("center of mass" coordinates), and again, $N_x = N_y$ for our models. In using this equation for the KL divergence with decimation, the equation will only be valid while a large number of sites remains, so not as many decimation steps can be achieved accurately.

The last thing needed for calculation the KL divergence at each decimation step is the recursion relation for the coupling constants. These are just the recursion relations for decimating an asymmetric square 2D Ising model with $J_y = \delta$ taken to be smaller than $J_x = J$. They are,

$$\begin{aligned} \beta J'_{1,x} = \beta J'_{1,y} = \beta J'_1 &= \frac{1}{4} \log \frac{\cosh(2\beta(J_x+J_y))}{\cosh(2\beta(J_x-J_y))}, \\ \beta J'_{2,x} &= \frac{1}{8} \log \left(\left(\frac{\cosh(\beta J_x)}{\cosh(\beta J_y)} \right)^2 \cosh(2\beta(J_x+J_y)) \cosh(2\beta(J_x-J_y)) \right), \\ \beta J'_{2,y} &= \frac{1}{8} \log \left(\left(\frac{\cosh(\beta J_y)}{\cosh(\beta J_x)} \right)^2 \cosh(2\beta(J_x+J_y)) \cosh(2\beta(J_x-J_y)) \right), \\ \beta J'_x = \beta J'_1 + \beta J'_{2,x} &= \frac{1}{8} \log \left(\left(\frac{(\cosh(\beta J_x)}{\cosh(\beta J_y)} \right)^2 \cosh^3(2\beta(J_x+J_y)) \cosh^{-1}(2\beta(J_x-J_y)) \right), \\ \beta J'_y = \beta J'_1 + \beta J'_{2,y} &= \frac{1}{8} \log \left(\left(\frac{(\cosh(\beta J_y)}{\cosh(\beta J_x)} \right)^2 \cosh^3(2\beta(J_x+J_y)) \cosh^{-1}(2\beta(J_x-J_y)) \right), \end{aligned} \tag{8.36}$$

where the same trick is being used as with the symmetric case: the four-point interaction is ignored (so no recursion relation for βJ_3 has been given) and the remaining coupling constants are added together to keep the Hamiltonian in the same form. Notice that the recursion relation reduces to the symmetric case (8.9) when $J_x = J_y$, as expected.

CHAPTER 9
Channel Capacity Example: Ising Model

We now study how to calculate the channel capacity by re-interpreting and generalizing our decimation procedure as follows. Start with an Ising model in the UV. Perform a blocking transformation (which reduces to decimation in the case the block has two spins). This results in a new Ising model. We can repeat this procedure until we arrive at an Ising model in the IR. The geometry of the setup is then a number of Ising models on B-ary trees multiplied together with a UV Ising model on one boundary and an IR Ising model on the other boundary.

Let us label the tree in the following manner. Let a tree have B branches at a node. Take a vector index **a**, where each element takes values $a_i = 0, ..., B-1$. We take **a** to be of length L, where L is the level of the tree. Our vector at level L is $(a_L, a_{L-1}, ..., a_0)$. To get the indices at level $L-1$, we leave off the first entry in the vector, giving $(a_{L-1}, ..., a_0)$. Our spins then connect as $\sigma_{a_L,a_{L-1},...,a_0}\sigma_{a_{L-1},...,a_0}$ across the levels. The total number of nodes on the tree with L levels is then $N = B^L + B^{L-1} + ... + 1$

9.1 1D Ising Model on Tree

For a 1D Ising model on a tree, the Hamiltonian looks like,

$$\begin{aligned} -H_{tree} &= J_L \sum_{a_L,a_{L-1},...,a_0} \sigma_{a_L,a_{L-1},...,a_0}\sigma_{a_{L-1},...,a_0} \\ &\quad + J_{L-1} \sum_{a_L,a_{L-1},...,a_0} \sigma_{a_{L-1},...,a_0}\sigma_{a_{L-2},...,a_0} + ... \\ &= -(H_L + H_{L-1} + ...). \end{aligned} \tag{9.1}$$

We will tak all the couplings J_i to be the same $J_i = J_{tree}$.

Taking H_{UV}, H_{IR} and $p(H_{UV})$, $p(H_{IR})$ to be the Hamiltonians and distributions for the 1D Ising models on the UV and IR boundaries of the tree, our total distribution is,

$$
\begin{aligned}
p(\sigma_{UV}, \sigma_{IR}) &= p(H_{UV})p(H_{IR}) \prod_{i,N_{IR}} p(H_{tree,i}) \\
&\equiv p(H_{UV})p(H_{IR})P(H_{tree}),
\end{aligned}
\tag{9.2}
$$

where N_{IR} is the number of sites of the 1D Ising model on the IR boundary. We can write the total distribution as a product of a conditional and marginal. To calculate channel capacity, we then have the following,

$$
\begin{aligned}
p(\sigma_{UV}, \sigma_{IR}) &= p(\sigma_{UV})P(\sigma_{IR}|\sigma_{UV}) \\
&= p(H_{UV} + h\sigma_{UV})P(\sigma_{IR}|\sigma_{UV}),
\end{aligned}
\tag{9.3}
$$

where we identify $p(\sigma_{UV}) = p(H_{UV} + h\sigma_{UV})$. If we integrate out all the IR modes to produce the UV distribution, we are left with the UV 1D Ising model and a bunch of spins attached to a magnetic field h. Our marginal distribution must then be the 1D UV Ising model plus a magnetic field, where the magnetic field is calculated by integrating out all the other modes of the tree and IR boundary. Likewise, $p(\sigma_{IR}) = p(H_{IR} + h\sigma_{IR})$.

Now, we already know what the full distribution is in terms of $p(H_{UV})$, $p(H_{IR})$, and $P(H_{tree})$, so we can find the conditional distribution as,

$$
\begin{aligned}
p(\sigma_{UV}, \sigma_{IR}) &= p(H_{UV})p(h\sigma_{UV})\frac{p(H_{tree})}{p(h\sigma_{UV})p(h\sigma_{IR})}p(H_{IR})p(h\sigma_{IR}) \\
\rightarrow P(\sigma_{IR}|\sigma_{UV}) &= p(H_{IR})p(h\sigma_{IR})\frac{p(H_{tree})}{p(h\sigma_{UV})p(h\sigma_{IR})} \\
&= p(H_{IR})\frac{p(H_{tree})}{p(h\sigma_{UV})}
\end{aligned}
\tag{9.4}
$$

The mutual information is then,

$$
\begin{aligned}
I(UV;IR) &= \int \mathrm{D}\sigma\ p(\sigma_{UV}, \sigma_{IR}) \log \frac{P(\sigma_{IR}|\sigma_{UV})}{p(H_{IR})p(h\sigma_{IR})} \\
&= \int \mathrm{D}\sigma\ p(\sigma_{UV}, \sigma_{IR}) \log \frac{p(H_{tree})}{p(h\sigma_{UV})p(h\sigma_{IR})} \\
&= \log \frac{Z_{hUV}Z_{hIR}}{Z_{tree}} \\
&\quad - \int \mathrm{D}\sigma\ p(H_{tree})p(H_{UV})p(H_{IR})(H_{tree} - h_{UV}\sum \sigma_{UV} - h_{IR}\sum \sigma_{IR}),
\end{aligned}
\tag{9.5}
$$

where,

$$\begin{aligned} Z_{hUV} &= \frac{Z_{UV}(h_{UV})}{Z_{UV}}, \\ Z_{hUV} &= \frac{Z_{IR}(h_{IR})}{Z_{IR}}, \end{aligned} \tag{9.6}$$

where $Z_{UV}(h_{UV})$ is the partition function for the UV 1D Ising model with the magnetic field h_{UV}, etc.

The problem now reduces to calculating correlation functions. First, for a 1D Ising model on a tree, we can calculate correlation functions using a transfer matrix method. We build the transfer matrix as follows.

For a B-ary tree (so a block size of B), find the transfer matrix $T(B)$ for B leaves going to a single node. This matrix $T(B)$ will be of size 2^B x 2. For a full tree with B leaves on each node (including the root), we can then find the transfer matrix at each level of the tree. The transfer matrix $T_{L;B}$ at level L (transitioning from level L to $L-1$; T_0 is just the identity) is a tensor product of the transfer matrices at each block,

$$\begin{aligned} T_{L;B} \equiv T(L \to L-1; B) &= \otimes_{N_{UV}/B} T(B) \\ &= \otimes_{B^{L-1}} T, \end{aligned} \tag{9.7}$$

where we have taken the number of UV sites $N_{UV} = B^L$ to keep the number of sites divisible by B as we go through the tree. Note that T_L is of size 2^{B^L} x $2^{B^{L-1}}$ To get the transfer matrix across more than one level of a tree, just multiply the transfer matrices together,

$$\begin{aligned} T(L \to k; B) &= T_{L;B} T_{L-1;B} ... T_{k+1;B} \\ &= \prod_{i=L}^{k+1} T_{i;B}, \end{aligned} \tag{9.8}$$

where $k < L$ is the final level of the tree to which we want to transfer (so the index i decreases in increments of 1), and we could have extended the product to k since T_0 is the identity in our convention.

We can now calculate the partition function of a full B-ary tree with a total of $\mathcal{L}$ levels and free boundary conditions by having spin vectors on both sides of the total transfer matrix (they will have length twice the number of sites at each boundary) with equal weight to spin up and spin down. For an initial state $\langle \psi_{UV}|$ and final state $|\psi_{IR}\rangle$, the partition function is,

$$Z_{tree}(\mathcal{L}, B) = \sum_{|\psi_{UV}\rangle, |\psi_{IR}\rangle} \langle \psi_{UV} | T(L \to 0; B) | \psi_{IR} \rangle. \tag{9.9}$$

Because we are interested in the partition function for a full tree, the final state is that of a single spin site: the root node. The initial state represents $B^{\mathcal{L}}$ independent spins, so our state vectors (for an all spin-up configuration) are,

$$\begin{aligned} |\psi_{UV}\rangle &= \otimes_{N_{UV}=B^{\mathcal{L}}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix} \\ |\psi_{IR}\rangle &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \end{aligned} \tag{9.10}$$

where we take the tensor product of $B^{\mathcal{L}}$ spin states. Now, we must sum over all possible configurations, e.g., a single spin down and the rest spin up represented by a 1 in the bottom of one of the vectors and 1s on top for the rest. To easily perform the sum over all configurations, notice that $\langle \psi_{UV}|$ is a $2^{B^{\mathcal{L}}}$ x 1 row vector having a single entry 1 and the rest are 0s. Each of the unique $2^{B^{\mathcal{L}}}$ configurations corresponds to a unique position in the row vector for the 1. Hence, if we summed all $2^{B^{\mathcal{L}}}$ vectors together, we would get a $2^{B^{\mathcal{L}}}$ x 1 row vector of ones. Similarly, if we summed together all the column vectors representing different configurations for the $|\psi_{IR}\rangle$, we would get a 2 x 1 column vector of 1s.

Hence, when we perform the sum over all configurations, we get,

$$\begin{aligned} Z_{tree}(\mathcal{L}, B) &= \sum_{|\psi_{UV}\rangle} \left(\langle \psi_{UV}| T(\mathcal{L} \to 0; B) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right. \\ &\quad \left. + \langle \psi_{UV}| T(\mathcal{L} \to 0) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \\ &= \sum_{|\psi_{UV}\rangle} \langle \psi_{UV}| T(\mathcal{L} \to 0; B) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \left(\begin{pmatrix} 1 & \dots & 0 \end{pmatrix} + ... + \begin{pmatrix} 0 & \dots & 1 \end{pmatrix} \right) T(\mathcal{L} \to 0; B) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & \dots & 1 \end{pmatrix} T(\mathcal{L} \to 0; B) \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \end{aligned} \tag{9.11}$$

That is, we have confirmed an intuition that the initial and final state vectors are vectors of 1s, representing that each state has equal probability of occurring. We could easily have derived the initial state vector by taking tensor products of B^L state vectors $(1, 1)$ (i.e, an equal probability of spin up or spin down for each spin site)

We could similarly calculate the partition function of a tree that is not full, which would be a product of two full trees. We could do this either by multiplying the partition functions of the two trees together (since without the root node, the two parts of the tree are independent of each other), or we could think of the two trees as a single incomplete tree and repeat the same calculation as above except change the final state vector into a tensor product of N_F column vectors $(1, 1)$, where N_F is the number of spin sites at the final state (e.g., for a full 2-level binary tree without the root node, there would be 2 spin sites at the final state).

Having calculated the partition function for the tree, we can now calculate correlation functions for the tree. We insert an operator that is made up of a tensor product of 2 x 2 identity matrices with Pauli matrices σ_z inserted at the spin sites of interest $i = 0, ..., B^{L-1}$ at layer L. The operator is,

$$\Sigma(\{\sigma_z(i)\})_L \equiv I_2 \otimes ... \otimes \sigma_z(i) \otimes ... \otimes I_2, \tag{9.12}$$

where we are taking a tensor product of a total of B^L objects and I_N is an $N \times N$ identity matrix. The correlation functions are then (using (1_N) to denote a row vector of ones of size N and $(1_N)^T$ the column vector),

$$\langle \sigma(i)_L ... \sigma(i')'_L \rangle = \frac{1}{Z_{tree}} \times (1_{2^{N_{UV}}}) T(\mathcal{L} \to L; B) \Sigma(\{\sigma_z(i)\})_L T(L \to L'; B) \Sigma(\{\sigma_z(i')\})'_L T(L' \to 0; B) (1_{2^{N_{UV}}})^T. \tag{9.13}$$

These results can then be taken further to calculate the KL divergence as in the earlier Ising model examples. However, it is a computational problem that we leave for future work.

CHAPTER 10
Continuum Examples

We have shown how the lattice renormalization group results matches the results of the continuum renormalization group, and we have calculated the KL divergence between lattice models: Ising models. We will now make some sample calculations for some continuum theories: a scalar field ϕ^4 theory, a $T\bar{T}$ deformation of a CFT, and a Kaluza-Klein theory. The first two models are related to Ising models. The scalar ϕ^4 theory is the continuum limit of a 4d Ising model at the critical value. The 2D Ising model becomes a theory of 2D massless fermions (a CFT) in the continuum limit; adding an irrelevant deformation is then just a perturbation around the conformal fixed point. The Kaluza-Klein theory is of interest for measuring the information of an extra-dimensional and compactified UV theory that remains in an IR theory.

In this section, we sketch out and set up the calculations but do not carry them to completion because they will need to be done numerically.

10.1 Scalar field

To show how our method of calculating the KL divergence with continuum theories produces a source term in the IR theory and a UV logarithmic divergence (as we noted in Section 6.3), consider a toy example in 4 dimensions where we have free scalar fields, slow ϕ modes and fast χ modes, along with $\phi(x)\chi(x)^3$ and $\phi(x)^2\chi(x)^2$ interaction terms. Notice that this is the result of having a scalar Φ^4 theory and then splitting the field $\Phi = \phi + \chi$,

$$\begin{aligned} S &= \int \mathrm{d}^4x \ \frac{1}{2}(\partial\Phi)^2 + \frac{m_\Phi^2}{2}\Phi^2 + \frac{g}{4!}\Phi^4 \\ &= \int \mathrm{d}^4x \ \frac{1}{2}(\partial(\phi+\chi))^2 + \frac{m_\Phi^2}{2}(\phi+\chi)^2 + \frac{g}{4!}(\phi+\chi)^4 \\ &\to S_0[\phi] + S_0[\chi] + \int \mathrm{d}^4x \ g_1\phi(x)\chi(x)^3 + \frac{g_2^2}{2}\phi(x)^2\chi(x)^2, \end{aligned} \tag{10.1}$$

where S_0 denotes the free parts of the action (including any mass terms).

We want to integrate out the χ fields, where the χ fields are very heavy relative to the cut-off of the theory (so the kinetic terms are much smaller than the mass term, giving a propagator of $1/M^2$ to leading order); equivalently, the χ fields are constant or slowly varying with respect to the cut-off. Before doing this, re-write our action as,

$$\begin{aligned} S &= S_0[\phi] + S_0[\chi] + \int \mathrm{d}^4x \; g_1\phi(x)\chi(x)^3 + \frac{g_2^2}{2}\phi(x)^2\chi(x)^2 \\ &= S_0[\phi] + S_0[\chi] + \int \mathrm{d}^4x \; \phi(x)\frac{\partial L[0,\chi(x)]}{\partial\phi} + \frac{\phi(x)^2}{2}\frac{\partial^2 L[0,\chi(x)]}{\partial\phi^2}, \end{aligned} \tag{10.2}$$

so we see that the form of the action has the derivative of the interaction Lagrangian multiplying a single ϕ field. Now we integrate out the χ fields. This produces an effective action with a leading order $J\phi$ interaction, where J must have units of M^3, and a $Q\phi^2/2$ interaction, where Q must have units of M^2. The integrated two-point function is then,

$$\begin{aligned} J^2\int \mathrm{d}^4x\mathrm{d}^4y \; \langle\phi(x)\phi(y)\rangle_0 + \frac{Q^2}{4}\int \mathrm{d}^4x\mathrm{d}^4y \; \langle\phi(x)^2\phi(y)^2\rangle_0 &\sim \frac{J^2}{\Lambda_{IR}^4}\int \mathrm{d}z \; \frac{z^3}{z^2} + \frac{Q^2}{\Lambda_{IR}^4}\int \mathrm{d}z \; \frac{z^3}{z^4} \\ &= \frac{J^2}{2\Lambda_{IR}^4\Lambda_{IR}^2} + \frac{Q^2}{\Lambda_{IR}^4}\log\frac{\Lambda_{UV}}{\Lambda_{IR}} \\ &\sim (\frac{M}{\Lambda_{IR}})^6 + (\frac{M}{\Lambda_{IR}})^4\log\frac{\Lambda_{UV}}{\Lambda_{IR}}, \end{aligned} \tag{10.3}$$

which diverges with the space-time volume and logarithmically diverges in the UV (the first term is UV finite). Because of this no greater than UV $\log$ divergences, one could instead start with the IR theory and then add the $\phi\chi^3$ and $\phi^2\chi^2$ interactions as a UV completion of the IR theory.

Notice also that this result in general only holds when the scaling dimension of the field is $(D-2)/2 < \Delta < D/2$. For dimensions that are greater than $D/2$, we can add more χ fields in our UV theory that couple to the ϕ field and include the other interactions that are compatible with the symmetries of the Lagrangian.

10.2 Irrelevant deformation

In the context of CFTs, adding an irrelevant deformation to the CFT moves the QFT to a higher energy scale: it is a perturbation around a conformal fixed point. In the AdS/CFT picture, an irrelevant deformation moves the QFT into the bulk. We will consider applying our set-up to a CFT of 2D massless

fermions with a particular irrelevant deformation: the $T\bar{T}$ deformation (($T\bar{T} = \psi\partial\psi\tilde{\psi}\bar{\partial}\tilde{\psi}$)). The $T\bar{T}$ deformation and other irrelevant deformations have previously been studied by [147], [148], [149], [150]. The action is,

$$S = \int \mathrm{d}^2x \ \psi\bar{\partial}\psi + \tilde{\psi}\partial\tilde{\psi} + \mu T\bar{T}. \tag{10.4}$$

We first split into fast and slow modes ($\psi = \psi_{fast} + \psi_{slow}$). For the deformation (the rest of the action gives a free action that splits into fast and slow mode free actions, since mixed modes should vanish due to momentum conservation) this gives,

$$\psi\partial\psi\tilde{\psi}\bar{\partial}\tilde{\psi} = \psi_{fast}\partial\psi_{fast}\tilde{\psi}_{fast}\bar{\partial}\tilde{\psi}_{fast} + \psi_{slow}\partial\psi_{slow}\tilde{\psi}_{slow}\bar{\partial}\tilde{\psi}_{slow} + ..., \tag{10.5}$$

where the rest is all the remaining permutations and the μ is ignored because it simply multiplies everything (it will be restored in the final result). Taking the leading order terms gives,

$$\begin{aligned}
S[0,\psi_{fast}] &= \int \mathrm{d}^2x \ \psi_{fast}\partial\psi_{fast}\tilde{\psi}_{fast}\bar{\partial}\tilde{\psi}_{fast} + \psi_{fast}\partial\psi_{slow}\tilde{\psi}_{fast}\bar{\partial}\tilde{\psi}_{fast} \\
&\quad + \psi_{fast}\partial\psi_{fast}\tilde{\psi}_{fast}\bar{\partial}\tilde{\psi}_{slow} + \psi_{fast}\partial\psi_{slow}\tilde{\psi}_{fast}\bar{\partial}\tilde{\psi}_{slow} \\
\frac{\delta S[0,\psi_{fast}]}{\delta\psi_{slow}}\cdot\psi_{slow} &= \int \mathrm{d}^2x \ \psi_{slow}\partial\psi_{fast}\tilde{\psi}_{fast}\bar{\partial}\tilde{\psi}_{fast} \\
\frac{\delta S[0,\psi_{fast}]}{\delta\tilde{\psi}_{slow}}\cdot\tilde{\psi}_{slow} &= \int \mathrm{d}^2x \ \psi_{fast}\partial\psi_{fast}\tilde{\psi}_{slow}\bar{\partial}\tilde{\psi}_{fast} \\
\frac{\delta^2 S[0,\psi_{fast}]}{\delta\psi_{slow}^2}\cdot\frac{\psi_{slow}^2}{2} &= 0 \\
\tilde{\psi}_{slow}\cdot\frac{\delta^2 S[0,\psi_{fast}]}{\delta\tilde{\psi}_{slow}\delta\psi_{slow}}\cdot\frac{\psi_{slow}}{2} &= \int \mathrm{d}^2x \ \frac{\tilde{\psi}_{slow}\psi_{slow}}{2}\partial\psi_{fast}\bar{\partial}\tilde{\psi}_{fast} \\
&= -\psi_{slow}\cdot\frac{\delta^2 S[0,\psi_{fast}]}{\delta\psi_{slow}\delta\tilde{\psi}_{slow}}\cdot\frac{\tilde{\psi}_{slow}}{2}.
\end{aligned} \tag{10.6}$$

When adding the mixed partial derivatives, their contribution goes to zero because they add with opposite sign.

Integrating out fast modes results in,

$$\begin{aligned}
S_{eff} &= S_0[\psi_{slow},\tilde{\psi}_{slow}] + \mu\int \mathrm{d}^2x \ J_2\partial\psi_{slow} + \tilde{J}_2\bar{\partial}\tilde{\psi}_{slow} \\
&\quad + Q\partial\psi_{slow}\bar{\partial}\tilde{\psi}_{slow} + J_1\psi_{slow} + \tilde{J}_1\tilde{\psi}_{slow},
\end{aligned} \tag{10.7}$$

where the J_1 have mass dimension $M^{7/2}$, the J_2 dimension $M^{5/2}$, and the Q dimension M. Notice that J are fermionic, while the Q is an ordinary number. Also, the μ has dimension M^{-2}.

The two-point functions for the J_2 and Q terms are descendants of the two-point function for the J_1 term.

10.3 Kaluza-Klein tower of masses

As an exercise to see what occurs in theories with extra compactified dimensions, as in string theory, consider a 5D free, massless scalar field theory and place the 5th dimension on a circle. This allows the 5D field to be expressed as a free, massless 4D theory plus a tower of free, massive scalar fields,

$$\begin{aligned} S_{5D} &= \int \mathrm{d}^4x\mathrm{d}y\ \partial^M\phi\partial_M\phi \\ &= \int \mathrm{d}^4x\mathrm{d}y\ \sum_{n=-\infty}^{\infty} \partial^\mu\phi_n(x^\mu)\partial_\mu\phi_n(x^\mu)^* - \frac{n^2}{r^2}|\phi_n|^2 \\ &= 2\pi r\int \mathrm{d}^4x\ \partial^\mu\phi_0\partial_\mu\phi_0^* + ... = S_{4D} + ... \end{aligned} \tag{10.8}$$

We see that the resulting 4D theory is the action for a free massless scalar field plus the sum of actions for free massive scalar fields. As a result, producing the split between fast and slow modes will not produce any non-zero interactions. Instead, the actions exactly divide into free actions for the fast and for the slow modes. There is thus no UV/IR mixing.

We must then include a mixing term for the 5D theory. It is not clear how to do this with an odd interaction, so take self-interactions that are even for the interaction Lagrangian, which in particular, will effectively create a mass term for the 5D theory,

$$S_{5D} = \int \mathrm{d}^4x\mathrm{d}y\ \partial^M\phi\partial_M\phi + S_{int}[\phi]. \tag{10.9}$$

This then results in the same 4D theory as above, but now there are even interaction actions associated with each scalar field mode. Performing the slow and fast split then leads to the following 4D action,

$$\begin{aligned} S[\phi_{slow},\phi_{fast}] = S_0^0[\phi_{fast}] + S_0^0[\phi_{slow}] + S_{int}^0[\phi_{slow},\phi_{fast}] \\ + \sum_{i=1} S_0^i[\phi_{fast}] + S_0^i[\phi_{slow}] + S_{int}^i[\phi_{slow},\phi_{fast}], \end{aligned} \tag{10.10}$$

where the 0 term is massless but the remainder have increasingly heavy masses. This is the same result as the previous scalar field example ($\phi\chi^3$), except now there are sums of fields (or equivalently, products of the previous result) and the zero mode action now has a mass (due to a ϕ^2 interaction term). Because of this, there is no reason to keep the zero mode action outside of the summation, producing,

$$\begin{aligned} S[\phi_{slow}, \phi_{fast}] &= \sum_{i=0} S_0^i[\phi_{fast}] + S_0^i[\phi_{slow}] + S_{int}^i[\phi_{slow}, \phi_{fast}], \\ S_{int} &= \sum_{i=0} S_{int}^i[\phi_{slow}, \phi_{fast}]. \end{aligned} \tag{10.11}$$

Expanding around zero and integrating out fast modes as before for the scalar field theory then produces,

$$S_{eff} = \sum_{i=0} S_0^i[\phi_{slow}] + 2\pi r J^i \cdot \phi_{slow} + 2\pi r Q^i \cdot |\phi_{slow}|^2. \tag{10.12}$$

The scale derivative of the KL divergence is then a sum of the previous scale derivative of the KL divergence.

We have thereby shown in this section a sketch of how the calculations should proceed for a number of continuum examples of interest, some of them being of interest because they are continuum versions of Ising models. Further work could carry these calculations through to completion to track the KL divergence density as we did with the Ising models.

CHAPTER 11
Conclusion

We have studied the use and interpretation of the KL divergence in quantifying RG flows in terms of information. We have interpreted the RG transformation as a noisy communication channel: there is loss of information as the irrelevant and marginally irrelevant operators of QFTs are suppressed and higher energy modes are integrated or blocked out. The KL divergence can then be used to calculate various information theoretic quantities, such as the mutual information and the channel capacity. We have also found some connection between our problem and that of the information bottleneck, which also makes use of a KL divergence: the RG transformation compresses the information about the UV variables via a blocking transformation or integrating out and then transmits the compressed information to the IR variables via rescaling. The RG flow can also be understood in terms of the AdS/CFT correspondence, as we have explained.

We then used the KL divergence to measure the proximity between QFTs. This proximity has then been used to measure the UV/IR mixing in a low energy theory that has flown down the RG from a higher energy theory. The general difficulties in using the KL divergence are a UV divergence from a contact term and an IR divergence from the space-time volume, suggesting that the physical quantity of interest is the KL divergence per volume. A difficulty in our particular case of measuring UV/IR mixing is avoiding integrating out IR modes and avoiding contact terms in the continuum case.

To circumvent these difficulties, we first studied lattice theories. Lattice theories provide a natural UV cutoff and a position space renormalization procedure via decimation that thus avoids notions of integrating out IR modes. We also found that there were two notions of decimation: decimation plus rescaling and forming a joint distribution of decimated theories. As expected, we found that both corresponded to the continuum notion of Wilsonian renormalization. By making use of the decimation procedure, we found that the KL divergence is negative in the case of the 2D Ising model when using decimation plus forming a joint distribution: signaling the absence of the irrelevant operators and thereby putting a bound on how important the irrelevant operators are to the IR theory. In the case of a positive

KL divergence in the other cases, we are more directly measuring the amount of information remaining from the UV operators in the IR theory.

We then studied continuum theories. By expanding the UV theories about a low energy point, we found a way to measure UV/IR mixing without integrating out IR modes. The channel capacity concept was found to be another promising way to avoid integrating out IR modes by re-interpreting the UV distribution as encoding the fluctuations of couplings (i.e., sources) in the IR theory. Our results for the KL divergence of continuum theories were scheme dependent, so we found a UV completion to remove the contact term. We also found the lowest order beta function of the KL divergence when used to measure UV-IR mixing.

Some questions for further study remain. We have shown a way to calculate the KL divergence for an 1D Ising model on a tree, but we did not perform the calculation because of computational difficulties: What does the KL divergence for this model produce? We have also found a correspondence between our treatment and that of the information bottleneck but could not set up the correspondence exactly: How exactly do we assign the distributions and calculate the information bottleneck for our situation? It seems that we know $p(x|, \tilde{x})$ already (the kernel of the RG transformation) and need to invert the equation for the bottleneck to find $p(y|\tilde{x})$, although it is not clear how knowing this distribution is helpful. Perhaps a variational method will be required to invert the equation, seeing how Tishby uses a variational method to find $p(x|, \tilde{x})$. There is also a question about how lattice gauge theories would work in our setup, although we have suggested a way by thinking of the gauge fields as living on the links of the lattice. This question should be explored more and example the KL divergence of lattice gauge theories calculated.

We did not calculate the channel capacity for simple models, although we set up the calculation for finding the mutual information for the 1D Ising model on a tree. What would a channel capacity calculation produce? Is there a way to show that in the continuum case a CFT optimizes the information rate of the channel and thereby produces the channel capacity? The AdS/CFT connection to the RG procedure and to our information theoretic interpretation is also interesting but complicated: Is there some simple calculation that could be done under an AdS/CFT setup that will show how the KL divergence changes with coupling and scale?

There was another simple model we could have calculated: the hierarchical Ising model, which makes the RG decimation procedure to be exact (we had to make an approximation in 2d in this thesis).

Also, numerical methods could be used to calculate more realistic examples and might be necessary for calculating the 1D Ising model on a tree example.

APPENDIX A: MATHEMATICA CODE FOR 1D ISING MODEL FIGURES

We list the code used for generating the figures for the 1D Ising model when using decimation plus forming a joint distribution, Figures 8.1-8.4.

```
calc1DDensity[βfirst_, pow_, density_: 1] := Module[{βlast = βfirst, n = 2^pow, β, J, Num, β
    temp = βfirst, i, d = 0, rawLogZ, rawE0, Z, lnZ0, E0, lnZtemp, Etemp, lambdaplus,
    lambdaminus, dkldensity },
  (*
  By default, this calculates the relative entropy density by first calculating the
      relative entropy and then dividing by a volume factor. The relative entropy is
      computed between a 1D Ising Model with no magnetic field and its decimated version.
      The decimated theory is a joint distribution formed from the other decimated
      versions of the original theory.

  The relative entropy for a given β (the β is absorbed into the J) is calculated at each
      decimation step from 1 to pow, where pow is the number of decimation steps to carry
      out. Of course, the β changes with each decimation step, which is determined by a
      recursion relation.

  If one inputs '0' as the third optional argument, the relative entropy will be computed
      instead of the relative entropy density.
  *)


  (* The partition function and internal energy = average energy of the Hamiltonian for
      the 1D Ising Model with no magnetic field. Found via the transfer matrix method. *)
  lambdaplus = (Exp[β*J] + Exp[-β*J]);
```

```
lambdaminus = (Exp[β*J] − Exp[−β*J]);
rawLogZ = (Log[lambdaplus^Num + lambdaminus^Num]);
rawE0 = −((D[rawLogZ, β]) /. β −> 1) /. J −> β;
rawLogZ = rawLogZ /. J −> 1;


(* Keeps the computation from failing when the coupling is small or large by using large
     and small coupling expansions. *)
If[βfirst < 10^(−15),
 E0 = (Series[rawE0, {β, 0, 4}] //
     Normal) /. {β −> βfirst, Num −> n},
 If[(βfirst > 1000),
  E0 = −n*βfirst,
  E0 = rawE0 /. {β −> βfirst, Num −> n}
  ]
 ];

(* Force the energy to be a numerical value. *)
E0 = N[E0, 50];

(* Keeps the computation from failing when the coupling is small or large by using large
     and small coupling expansions. *)
If[βfirst < 10^(−2),
 lnZ0 = (Series[Log[Z], {β, 0, 4}] //
     Normal) /. {β −> βfirst, Num −> n},
 If[(βfirst > 1000),
  lnZ0 = Log[2] + n*βfirst,
  lnZ0 = rawLogZ /. {β −> βfirst, Num −> n}
  ]
```

```
];

lnZ0 = N[lnZ0, 50];

(* Create an array to store values of the relative entropy. *)
dkldensity = Array[0 &, pow];

(* This computes the relative entropy at each step and stores the value. *)
For[i = 1, i <= pow, i++,


 (* The recursion relation for the new coupling constant. *)
 If[βlast < 10^(−2),
   βtemp = βtemp^2 − (2/3)*βtemp^4,
   If[(βlast > 1000),
    βtemp = N[(βtemp − (1/2)*Log[2]), 50],
   βtemp = N[(1/2)*Log[Cosh[2*βtemp]], 50]
   ]
  ];


 If[βtemp < 10^−100,
  βtemp = 0,
  Continue
  ];


 (* Calculate energy and LogZ at each step. *)
```

```
(* Keeps the computation from failing when the coupling is small or large by using
     large and small coupling expansions. *)
If[βtemp < 10^(−15),
 Etemp = (Series[rawE0, {β, 0, 4}] //
     Normal) /. {β −> βtemp},
 If[(βtemp > 1000),
  Etemp = −(n/2^i)*βtemp,
  Etemp = rawE0 /. {β −> βtemp}
  ]
 ];


(* Forces the Etemp to a numerical value *)
Etemp = N[Etemp /. Num −> (n/2^i), 50];


(* Keeps the computation from failing when the coupling is small or large by using
     large and small coupling expansions. *)
If[βtemp < 10^(−2),
 lnZtemp = (Series[rawLogZ, {β, 0, 4}] //
     Normal) /. {β −> βtemp},
 If[(βtemp > 1000),
  lnZtemp = Log[2] + (n/2^i)*βtemp,
  lnZtemp = rawLogZ /. {β −> βtemp}
  ]
 ];


(* Forces the lnZtemp to a numerical value *)
```

```
lnZtemp = N[lnZtemp /. Num -> (n/2^i), 50];


(** The following block keeps from dividing by the volume if density selected is 0, i.e
    ., calculates relative entropy, not relative entropy density, if density selected
    is 0.

The relative entropy is computed using an equation I derived.
**)


If[ density != 0,
  dkldensity [[ i ]] =
   N[(((2^i)*lnZtemp - lnZ0) + (-E0 + (2^i)*Etemp))/(n), 50],
  dkldensity [[ i ]] =
   N[(((2^i)*lnZtemp - lnZ0) + (-E0 + (2^i)*Etemp)), 50]
  ];

βlast = βtemp

(* End of For Loop *)
 ];

(* Return *)
 dkldensity
 ];


betafunc1D[βmin_, βmax_, stepSize_: 1, k_] := Module[{dkl, diffs },
```

```
(*
Calculates the Beta Function of the KL divergence = (KL Density_(k+1) − KL Density_(k))
    /2^k for the 1D Ising model, where k is the decimation step. The KL Density is
    calculated by making a table of values for different k and β, and then taking the
    difference along each list that corresponds to a different β. The step size is an
    optional argument so that β doesn't have to only increase by 1.
*)


(* Table of values for the KL divergence of the 1D Ising Model. *)


dkl = Table[
  calc1DDensity[j, k], {j, βmin, βmax, stepSize}];

(* Prepare array to hold the beta function of the KL divergence. *)
 diffs = Array[0 &, Length[dkl]];


(* Computes the beta function of the KL divergence. *)
For[i = 1, i <= Length[dkl], i++,

 kl = dkl[[i]];
 diffs [[i]] =
  Table[N[(kl[[m + 1]] − kl[[m]])/2^m, 50], {m, 1, k − 1}]

 ];

(* Return *)
```

```
    diffs
    ];


(* Plot the KL density. *)
ListLinePlot [Table[calc1DDensity[j, 19], {j, 0.2, 1.7, 0.1}], PlotMarkers -> Automatic,
    PlotLabel -> "KL␣Density␣vs␣Decimation␣Step", Frame -> True, FrameLabel -> {"
    Decimation␣step␣k", "KL␣Density"}, PlotLegends -> LineLegend[Table[Row[{"βJ = ", j}], {j,
    0.2, 1.7, 0.1}], LegendMarkers -> Automatic], PlotMarkers -> Automatic, PlotRange -> All]

(* Plot the beta function *)
ListLinePlot [betafunc1D[0.2, 1.7, 0.1, 19], PlotMarkers -> Automatic, PlotLabel -> "Beta␣
    Function␣of␣KL␣Density␣vs␣Decimation␣Step", Frame -> True, FrameLabel -> {"
    Decimation␣step␣k", "Beta␣Function"}, PlotLegends -> LineLegend[Table[Row[{"βJ = ", j}],
    {j, 0.2, 1.7, 0.1}], LegendMarkers -> Automatic], PlotMarkers -> Automatic, PlotRange ->
    All]
```

APPENDIX B: MATHEMATICA CODE FOR 2D ISING MODEL FIGURES

We list the code used for generating the figures for the 2D Ising model when using decimation plus forming a joint distribution, Figures 8.7-8.12.

```
calc2DDensity[βfirst_, pow_, density_: 1] := Module[{βlast = βfirst, n = 2^pow, βtemp = βfirst,
    β, i, d = 0, E0, lnZ0, lnZtemp, Etemp, Ktemp, dkldensity, κ, c = 0},
  (*
  By default, this calculates the relative entropy density by first calculating the
      relative entropy and then dividing by a volume factor. The relative entropy is
      computed between a 2D Ising Model with no magnetic field and its decimated version.
      The decimated theory is a joint distribution formed from the other decimated
      versions of the original theory.

  The relative entropy for a given β (the β is absorbed into the J) is calculated at each
      decimation step from 1 to pow, where pow is the number of decimation steps to carry
      out. Of course, the β changes with each decimation step, which is determined by a
      recursion relation.

  If one inputs '0' as the third optional argument, the relative entropy will be computed
      instead of the relative entropy density.
  *)


  (* The internal energy = average energy of the Hamiltonian for the 2D Ising model with
      no magnetic field. This value was found in Pathria. *)
  κ = 2*Sinh[2*βfirst]/Cosh[2*βfirst]^2;
```

```
If[βfirst < 10^−15,
 E0 = N[−n^2∗(2∗βfirst^2 + (10/3)∗βfirst^4), 50],
 If[βfirst > 10^3,
  E0 = N[−n^2∗βfirst∗2, 50],
  E0 =
   N[−n^2∗βfirst∗
     Coth[2∗βfirst]∗(1 + (2/Pi)∗
       EllipticK[κ^2]∗(2∗Tanh[2∗βfirst]^2 − 1)), 50]
  ]
 ];

(∗ The Log of the partition function. Also found in Pathria. n^2 = the total number of
    sites. ∗)
If[κ < 10^−10,
 lnZ0 =
  N[n^2∗Log[Sqrt[2]∗Cosh[2∗βfirst]] + (n^2/2)∗Log[2], 50],
 lnZ0 =
  N[n^2∗Log[Sqrt[2]∗Cosh[2∗βfirst]] + (n^2/Pi)∗
    NIntegrate[
     Log[1 + Sqrt[1 − (κ)^2∗Sin[\[Phi]]^2]], {\[Phi], 0,
      Pi/2}], 50]
 ];

(∗ Create an array to store values of the relative entropy. ∗)
dkldensity = Array[0 &, pow];
```

```
(* This computes the relative entropy at each step and stores the value. *)
For[i = 1, i <= pow, i++,


 If[βlast < 10^(−2),
  Ktemp = N[−2*βtemp^4, 50],
   If[(βlast > 1000),
    Ktemp = N[(3/8)*Log[2] − (1/2)*βtemp, 50],
   Ktemp =
    N[(1/8)*Log[Cosh[4*βtemp]] − (1/2)*
       Log[Cosh[2*βtemp]], 50]
   ]
  ];

(* The recursion relation for the new coupling constant. *)
If[βlast < 10^(−2),
 βtemp = N[3*βtemp^2 − 8*βtemp^4, 50],
   If[(βlast > 1000),
    βtemp = N[3*(4*βtemp − Log[2])/8, 50],
   βtemp = N[(3/8)*Log[Cosh[4*βtemp]], 50]
   ]
  ];




(* Calculate energy and LogZ at each step. *)
κ = 2*Sinh[2*βtemp]/Cosh[2*βtemp]^2;

If[βtemp < 10^(−15),
```

```
Etemp =
 N[−(n^2/2^i)∗(1)∗(2∗βtemp^2 + (10/3)∗βtemp^4), 50],
If[βtemp > 10^3,
 Etemp = N[−(n^2/2^i)∗βtemp∗2, 50],
 Etemp =
  N[−(n^2/2^i)∗βtemp∗
    Coth[2∗βtemp]∗(1 + (2/Pi)∗
      EllipticK[κ^2]∗(2∗Tanh[2∗βtemp]^2 − 1)), 50]
 ]
];

κ = 2∗Sinh[2∗βtemp]/Cosh[2∗βtemp]^2;

If[κ < 10^−10,
 lnZtemp =
  N[(n^2/2^i)∗Log[Sqrt[2]∗Cosh[2∗βtemp]] + ((n^2/2^i)/2)∗
    Log[2], 50],
 lnZtemp =
  N[(n^2/2^i)∗Log[Sqrt[2]∗Cosh[2∗βtemp]] + ((n^2/2^i)/Pi)∗
    NIntegrate[
     Log[1 + Sqrt[1 − (κ)^2∗Sin[\[Phi]]^2]], {\[Phi], 0,
      Pi/2}], 50]
 ];

(∗∗ The following block keeps from dividing by the volume if density selected is 0, i.e
    ., calculates relative entropy, not relative entropy density, if density selected
    is 0.

The relative entropy is computed using an equation I derived.
```

```
    **)

    If[ density  != 0,
     dkldensity [[ i ]]  = N[(((2^ i )*lnZtemp − lnZ0) + (−E0 + (2^i)*Etemp))/(n^2),  50],
     dkldensity [[ i ]]  = N[(((2^ i )*lnZtemp − lnZ0) + (−E0 + (2^i)*Etemp)),  50]
     ];
    βlast = βtemp;


    (* End of For Loop *)
    ];

   (* Return *)
   dkldensity
   ];


betafunc2D[βmin_, βmax_, stepSize_: 1, k_] :=
  Module[{dkl,  diffs },
   (*
   Calculates the Beta Function of the KL divergence = (KL Density_(k+1) − KL Density_(k))
       /2^k for the 2D Ising model, where k is the decimation step. The KL Density is
       calculated by making a table of values for different k and β, and then taking the
       difference along each list that corresponds to a different β. The step size is an
       optional argument so that β doesn't have to only increase by 1.
   *)


   (* Table of values for the KL divergence of the 2D Ising Model. *)
```

```
dkl = Table[
  calc2DDensity[j, k], {j, βmin, βmax, stepSize}];

(* Prepare array to hold the beta function of the KL divergence. *)
diffs = Array[0 &, Length[dkl]];


(* Computes the beta function of the KL divergence. *)
For[i = 1, i <= Length[dkl], i++,

 kl = dkl[[i]];
 diffs[[i]] =
  Table[N[(kl[[m + 1]] - kl[[m]])/2^m, 50], {m, 1, k - 1}]

 ];

(* Return *)
diffs
];


(* Plots. Our figures below our plotted in terms of the critical point of the 2D Ising
   model that we used in our calculations: 0.506981.... We look at points above and below
   the critical point in even steps of 0.01 or 0.1, etc., so we use points like
   0.306981... for our starting and/or ending values in the plots. *)

(* Plot the KL density *)
```

```
ListLinePlot [Table[calc2DDensity[j, 19], {j,
    0.30698101907756665365267663270784405816039132035413,
    0.50698101907756665365267663270784405816039132035413, 0.01}], PlotMarkers ->
    Automatic, PlotLabel -> "KL␣Density␣vs␣Decimation␣Step", Frame -> True, FrameLabel
    -> {"Decimation␣step␣k", "KL␣Density"}, PlotLegends -> LineLegend[Table[Row[{"βJ = ", j
    }], {j, 0.30698101907756665365267663270784405816039132035413,
    0.50698101907756665365267663270784405816039132035413, 0.01}], LegendMarkers ->
    Automatic], PlotMarkers -> Automatic, PlotRange -> All]

(* Plot the beta function *)
ListLinePlot [betafunc2D[0.50698101907756665365267663270784405816039132035413,
    0.70698101907756665365267663270784405816039132035413, 0.01, 19], PlotMarkers ->
    Automatic, PlotLabel -> "Beta␣Function␣of␣KL␣Density␣vs␣Decimation␣Step", Frame ->
    True, FrameLabel -> {"Decimation␣step␣k", "Beta␣Function"}, PlotLegends -> LineLegend[
    Table[Row[{"βJ = ", j}], {j, 0.50698101907756665365267663270784405816039132035413,
    0.70698101907756665365267663270784405816039132035413, 0.01}], LegendMarkers ->
    Automatic], PlotMarkers ->Automatic, PlotRange -> All]


(* Prepare to plot the KL density versus coupling *)
kltempplot = Table[calc2DDensity[j, 19], {j,
    0.00698101907756665365267663270784405816039132035413,
    1.30698101907756665365267663270784405816039132035413, 0.01}];

(* Plot the KL density versus coupling at different decimation steps *)
ListLinePlot [Table[kltempplot[[1 ;; All, k]], {k, 1, 8, 1}], DataRange ->
    {0.00698101907756665365267663270784405816039132035413,
    1.30698101907756665365267663270784405816039132035413}, PlotMarkers -> Automatic,
    PlotLabel -> "KL␣Density␣vs␣Coupling", Frame -> True, FrameLabel -> {"βJ", "KL␣
```

```
Density"}, PlotLegends −> LineLegend[Table[Row[{"Decimation␣step␣k␣=␣", j}], {j, 1, 8, 1}],
 LegendMarkers −> Automatic], PlotMarkers −> Automatic, PlotRange −> All]
```

REFERENCES

[1] Michael R. Douglas. The Statistics of string / M theory vacua. *JHEP*, 05:046, 2003. doi: 10.1088/1126-6708/2003/05/046.

[2] Washington Taylor and Yi-Nan Wang. The F-theory geometry with most flux vacua. *JHEP*, 12: 164, 2015. doi: 10.1007/JHEP12(2015)164.

[3] N. Tishby, F. C. Pereira, and W. Bialek. The information bottleneck method. 2000.

[4] Juan Martin Maldacena. The Large N limit of superconformal field theories and supergravity. *Int. J. Theor. Phys.*, 38:1113–1133, 1999. doi: 10.1023/A:1026654312961. [Adv. Theor. Math. Phys.2,231(1998)].

[5] S. S. Gubser, Igor R. Klebanov, and Alexander M. Polyakov. Gauge theory correlators from noncritical string theory. *Phys. Lett.*, B428:105–114, 1998. doi: 10.1016/S0370-2693(98)00377-3.

[6] Edward Witten. Anti-de Sitter space and holography. *Adv. Theor. Math. Phys.*, 2:253–291, 1998.

[7] Jan de Boer, Erik P. Verlinde, and Herman L. Verlinde. On the holographic renormalization group. *JHEP*, 08:003, 2000. doi: 10.1088/1126-6708/2000/08/003.

[8] Jacques Calmet and Xavier Calmet. Distance between physical theories based on information theory. *Mod. Phys. Lett.*, A26:319–323, 2011. doi: 10.1142/S0217732311034955.

[9] Vijay Balasubramanian, Jonathan J. Heckman, and Alexander Maloney. Relative Entropy and Proximity of Quantum Field Theories. *JHEP*, 05:104, 2015. doi: 10.1007/JHEP05(2015)104.

[10] Pasquale Calabrese and John Cardy. Entanglement entropy and conformal field theory. *J. Phys.*, A42:504005, 2009. doi: 10.1088/1751-8113/42/50/504005.

[11] H. Casini and Marina Huerta. On the RG running of the entanglement entropy of a circle. *Phys. Rev.*, D85:125016, 2012. doi: 10.1103/PhysRevD.85.125016.

[12] Robert C. Myers and Aninda Sinha. Holographic c-theorems in arbitrary dimensions. *JHEP*, 01: 125, 2011. doi: 10.1007/JHEP01(2011)125.

[13] Tatsuma Nishioka, Shinsei Ryu, and Tadashi Takayanagi. Holographic Entanglement Entropy: An Overview. *J. Phys.*, A42:504008, 2009. doi: 10.1088/1751-8113/42/50/504008.

[14] Vijay Balasubramanian, Michael B. McDermott, and Mark Van Raamsdonk. Momentum-space entanglement and renormalization in quantum field theory. *Phys. Rev.*, D86:045014, 2012. doi: 10.1103/PhysRevD.86.045014.

[15] Pasquale Calabrese and John L. Cardy. Entanglement entropy and quantum field theory. *J. Stat. Mech.*, 0406:P06002, 2004. doi: 10.1088/1742-5468/2004/06/P06002.

[16] H. Casini and M. Huerta. Entanglement entropy in free quantum field theory. *J. Phys.*, A42: 504007, 2009. doi: 10.1088/1751-8113/42/50/504007.

[17] Ronak M Soni and Sandip P. Trivedi. Aspects of Entanglement Entropy for Gauge Theories. *JHEP*, 01:136, 2016. doi: 10.1007/JHEP01(2016)136.

[18] A. B. Zamolodchikov. Irreversibility of the Flux of the Renormalization Group in a 2D Field Theory. *JETP Lett.*, 43:730–732, 1986. [Pisma Zh. Eksp. Teor. Fiz.43,565(1986)].

[19] John L. Cardy. Is There a c Theorem in Four-Dimensions? *Phys. Lett.*, B215:749–752, 1988. doi: 10.1016/0370-2693(88)90054-8.

[20] H. Osborn. Derivation of a Four-dimensional c Theorem. *Phys. Lett.*, B222:97–102, 1989. doi: 10.1016/0370-2693(89)90729-6.

[21] Zohar Komargodski and Adam Schwimmer. On Renormalization Group Flows in Four Dimensions. *JHEP*, 12:099, 2011. doi: 10.1007/JHEP12(2011)099.

[22] Daniel L. Jafferis, Igor R. Klebanov, Silviu S. Pufu, and Benjamin R. Safdi. Towards the F-Theorem: N=2 Field Theories on the Three-Sphere. *JHEP*, 06:102, 2011. doi: 10.1007/JHEP06(2011)102.

[23] Efrat Gerchkovitz, Jaume Gomis, and Zohar Komargodski. Sphere Partition Functions and the Zamolodchikov Metric. *JHEP*, 11:001, 2014. doi: 10.1007/JHEP11(2014)001.

[24] David Skinner. Quantum Field Theory II: The Renormalization Group. 2017.

[25] Steven Weinberg. *The quantum theory of fields. Vol. 2: Modern applications*. Cambridge University Press, 2013. ISBN 9781139632478, 9780521670548, 9780521550024.

[26] R. Machleidt and D. R. Entem. Chiral effective field theory and nuclear forces. *Phys. Rept.*, 503: 1–75, 2011. doi: 10.1016/j.physrep.2011.02.001.

[27] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995. ISBN 9780201503975, 0201503972. URL http://www.slac.stanford.edu/~mpeskin/QFT.html.

[28] W. N. Cottingham and D. A. Greenwood. *An introduction to the standard model of particle physics*. Cambridge University Press, 2007. ISBN 9780511271366, 9780521852494.

[29] Luis E. Ibanez and Angel M. Uranga. *String theory and particle physics: An introduction to string phenomenology*. Cambridge University Press, 2012. ISBN 9780521517522, 9781139227421. URL http://www.cambridge.org/de/knowledge/isbn/item6563092/?site_locale=de_DE.

[30] Pierre Binetruy. *Supersymmetry*. Oxford Graduate Texts. Oxford University Press, 2012. ISBN 9780199652730, 9780198509547.

[31] Palash B. Pal. *An Introductory Course of Particle Physics*. CRC Press, 2014. ISBN 9781482216981. doi: 10.1201/b17199.

[32] Eldad Gildener. Gauge Symmetry Hierarchies. *Phys. Rev.*, D14:1667, 1976. doi: 10.1103/PhysRevD.14.1667.

[33] Steven Weinberg. Implications of Dynamical Symmetry Breaking. *Phys. Rev.*, D13:974–996, 1976. doi: 10.1103/PhysRevD.19.1277,10.1103/PhysRevD.13.974. [Addendum: Phys. Rev.D19,1277(1979)].

[34] Leonard Susskind. Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory. *Phys. Rev.*, D20:2619–2625, 1979. doi: 10.1103/PhysRevD.20.2619.

[35] Gerard 't Hooft. Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking. *NATO Sci. Ser. B*, 59:135–157, 1980. doi: 10.1007/978-1-4684-7571-5_9.

[36] C. H. Llewellyn Smith and Graham G. Ross. The Real Gauge Hierarchy Problem. *Phys. Lett.*, 105B:38–40, 1981. doi: 10.1016/0370-2693(81)90035-6.

[37] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. Matter and Antimatter in the Universe. *New J. Phys.*, 14:095012, 2012. doi: 10.1088/1367-2630/14/9/095012.

[38] Harald Fritzsch and Peter Minkowski. Unified Interactions of Leptons and Hadrons. *Annals Phys.*, 93:193–266, 1975. doi: 10.1016/0003-4916(75)90211-0.

[39] H. Georgi and S. L. Glashow. Unity of All Elementary Particle Forces. *Phys. Rev. Lett.*, 32: 438–441, 1974. doi: 10.1103/PhysRevLett.32.438.

[40] F. Gursey, Pierre Ramond, and P. Sikivie. A Universal Gauge Theory Model Based on E6. *Phys. Lett.*, 60B:177–180, 1976. doi: 10.1016/0370-2693(76)90417-2.

[41] Kenneth Lane. Two lectures on technicolor. 2002.

[42] Savas Dimopoulos and Leonard Susskind. Mass Without Scalars. *Nucl. Phys.*, B155:237–252, 1979. doi: 10.1016/0550-3213(79)90364-X. [2,930(1979)].

[43] Estia Eichten and Kenneth D. Lane. Dynamical Breaking of Weak Interaction Symmetries. *Phys. Lett.*, 90B:125–130, 1980. doi: 10.1016/0370-2693(80)90065-9.

[44] G. R. Dvali, S. Randjbar-Daemi, and R. Tabbash. The Origin of spontaneous symmetry breaking in theories with large extra dimensions. *Phys. Rev.*, D65:064021, 2002. doi: 10.1103/PhysRevD.65.064021.

[45] N. V. Krasnikov. Ultraviolet fixed point behavior of the five-dimensional Yang-Mills theory, the gauge hierarchy problem and a possible new dimension at the TeV scale. *Phys. Lett.*, B273: 246–249, 1991. doi: 10.1016/0370-2693(91)91678-O.

[46] N. S. Manton. A New Six-Dimensional Approach to the Weinberg-Salam Model. *Nucl. Phys.*, B158:141–153, 1979. doi: 10.1016/0550-3213(79)90192-5.

[47] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. The Hierarchy problem and new dimensions at a millimeter. *Phys. Lett.*, B429:263–272, 1998. doi: 10.1016/S0370-2693(98)00466-3.

[48] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. Phenomenology, astrophysics and cosmology of theories with submillimeter dimensions and TeV scale quantum gravity. *Phys. Rev.*, D59:086004, 1999. doi: 10.1103/PhysRevD.59.086004.

[49] Gregory Gabadadze. ICTP lectures on large extra dimensions. *ICTP Lect. Notes Ser.*, 14:77–120, 2003.

[50] Nima Arkani-Hamed, Andrew G. Cohen, and Howard Georgi. Electroweak symmetry breaking from dimensional deconstruction. *Phys. Lett.*, B513:232–240, 2001. doi: 10.1016/S0370-2693(01)00741-9.

[51] Pierre Fayet. Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions. *Phys. Lett.*, 69B:489, 1977. doi: 10.1016/0370-2693(77)90852-8.

[52] Pierre Fayet. MASSIVE GLUINOS. *Phys. Lett.*, 78B:417–420, 1978. doi: 10.1016/0370-2693(78)90474-4.

[53] Pierre Fayet. Relations Between the Masses of the Superpartners of Leptons and Quarks, the Goldstino Couplings and the Neutral Currents. *Phys. Lett.*, 84B:416, 1979. doi: 10.1016/0370-2693(79)91229-2.

[54] Savas Dimopoulos and Howard Georgi. Softly Broken Supersymmetry and SU(5). *Nucl. Phys.*, B193:150–162, 1981. doi: 10.1016/0550-3213(81)90522-8.

[55] S. Dimopoulos, S. Raby, and Frank Wilczek. Supersymmetry and the Scale of Unification. *Phys. Rev.*, D24:1681–1683, 1981. doi: 10.1103/PhysRevD.24.1681.

[56] Salvatore Rappoccio. The experimental status of direct searches for exotic physics beyond the standard model at the Large Hadron Collider. *Rev. Phys.*, 4:100027, 2019. doi: 10.1016/j.revip.2018.100027.

[57] Michael Dine, Willy Fischler, and Mark Srednicki. Supersymmetric Technicolor. *Nucl. Phys.*, B189:575–593, 1981. doi: 10.1016/0550-3213(81)90582-4.

[58] Pierre Ramond. Dual Theory for Free Fermions. *Phys. Rev.*, D3:2415–2418, 1971. doi: 10.1103/PhysRevD.3.2415.

[59] Yu. A. Golfand and E. P. Likhtman. Extension of the Algebra of Poincare Group Generators and Violation of p Invariance. *JETP Lett.*, 13:323–326, 1971. [Pisma Zh. Eksp. Teor. Fiz.13,452(1971)].

[60] D. V. Volkov and V. P. Akulov. Is the Neutrino a Goldstone Particle? *Phys. Lett.*, 46B:109–110, 1973. doi: 10.1016/0370-2693(73)90490-5.

[61] J. Wess and B. Zumino. Supergauge Transformations in Four-Dimensions. *Nucl. Phys.*, B70:39–50, 1974. doi: 10.1016/0550-3213(74)90355-1. [,24(1974)].

[62] Gerard Jungman, Marc Kamionkowski, and Kim Griest. Supersymmetric dark matter. *Phys. Rept.*, 267:195–373, 1996. doi: 10.1016/0370-1573(95)00058-5.

[63] Paul Langacker. Grand Unified Theories. *eConf*, C810824:823, 1981.

[64] Graham G. Ross. UNIFIED FIELD THEORIES. *Rept. Prog. Phys.*, 44:655–718, 1981. doi: 10.1088/0034-4885/44/6/002.

[65] Graham G. Ross. *GRAND UNIFIED THEORIES*. 1985.

[66] Jogesh C. Pati and Abdus Salam. Unified Lepton-Hadron Symmetry and a Gauge Theory of the Basic Interactions. *Phys. Rev.*, D8:1240–1251, 1973. doi: 10.1103/PhysRevD.8.1240.

[67] H. Georgi, Helen R. Quinn, and Steven Weinberg. Hierarchy of Interactions in Unified Gauge Theories. *Phys. Rev. Lett.*, 33:451–454, 1974. doi: 10.1103/PhysRevLett.33.451.

[68] Luis E. Ibanez and Graham G. Ross. Low-Energy Predictions in Supersymmetric Grand Unified Theories. *Phys. Lett.*, 105B:439–442, 1981. doi: 10.1016/0370-2693(81)91200-4.

[69] Daniel Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara. Progress Toward a Theory of Supergravity. *Phys. Rev.*, D13:3214–3218, 1976. doi: 10.1103/PhysRevD.13.3214.

[70] Stanley Deser and B. Zumino. Consistent Supergravity. *Phys. Lett.*, 62B:335, 1976. doi: 10.1016/0370-2693(76)90089-7. [,335(1976)].

[71] Hans Peter Nilles. Supersymmetry, Supergravity and Particle Physics. *Phys. Rept.*, 110:1–162, 1984. doi: 10.1016/0370-1573(84)90008-5.

[72] E. Cremmer, B. Julia, Joel Scherk, S. Ferrara, L. Girardello, and P. van Nieuwenhuizen. Spontaneous Symmetry Breaking and Higgs Effect in Supergravity Without Cosmological Constant. *Nucl. Phys.*, B147:105, 1979. doi: 10.1016/0550-3213(79)90417-6.

[73] E. Cremmer, S. Ferrara, L. Girardello, and Antoine Van Proeyen. Yang-Mills Theories with Local Supersymmetry: Lagrangian, Transformation Laws and SuperHiggs Effect. *Nucl. Phys.*, B212: 413, 1983. doi: 10.1016/0550-3213(83)90679-X. [,413(1982)].

[74] E. Cremmer, B. Julia, and Joel Scherk. Supergravity Theory in Eleven-Dimensions. *Phys. Lett.*, 76B:409–412, 1978. doi: 10.1016/0370-2693(78)90894-8. [,25(1978)].

[75] J. A. Strathdee. EXTENDED POINCARE SUPERSYMMETRY. *Int. J. Mod. Phys.*, A2:273, 1987. doi: 10.1142/S0217751X87000120. [,104(1986)].

[76] E. Cremmer and B. Julia. The N=8 Supergravity Theory. 1. The Lagrangian. *Phys. Lett.*, 80B:48, 1978. doi: 10.1016/0370-2693(78)90303-9. [,78(1978)].

[77] E. Cremmer and B. Julia. The SO(8) Supergravity. *Nucl. Phys.*, B159:141–212, 1979. doi: 10.1016/0550-3213(79)90331-6.

[78] W. Nahm. Supersymmetries and their Representations. *Nucl. Phys.*, B135:149, 1978. doi: 10.1016/0550-3213(78)90218-3. [,7(1977)].

[79] Itzhak Bars. Two - time physics. In *Group theoretical methods in physics. Proceedings, 22nd International Colloquium, Group22, ICGTMP'98, Hobart, Australia, July 13-17, 1998*, pages 2–17, 1998.

[80] Mirjam Cvetič, James Halverson, Ling Lin, Muyang Liu, and Jiahua Tian. Quadrillion F-Theory Compactifications with the Exact Chiral Spectrum of the Standard Model. *Phys. Rev. Lett.*, 123 (10):101601, 2019. doi: 10.1103/PhysRevLett.123.101601.

[81] Brian R. Greene. String theory on Calabi-Yau manifolds. In *Fields, strings and duality. Proceedings, Summer School, Theoretical Advanced Study Institute in Elementary Particle Physics, TASI'96, Boulder, USA, June 2-28, 1996*, pages 543–726, 1996.

[82] P. Candelas. LECTURES ON COMPLEX MANIFOLDS. 1987.

[83] P. Candelas, Gary T. Horowitz, Andrew Strominger, and Edward Witten. Vacuum Configurations for Superstrings. *Nucl. Phys.*, B258:46–74, 1985. doi: 10.1016/0550-3213(85)90602-9.

[84] Michael B. Green, J. H. Schwarz, and Edward Witten. *SUPERSTRING THEORY. VOL. 2: LOOP AMPLITUDES, ANOMALIES AND PHENOMENOLOGY*. 1988. ISBN 9780521357531. URL http://www.cambridge.org/us/academic/subjects/physics/theoretical-physics-and-mathematical-physics/superstring-theory-volume-2.

[85] Saul Ramos-Sanchez. Towards Low Energy Physics from the Heterotic String. *Fortsch. Phys.*, 10: 907–1036, 2009. doi: 10.1002/prop.200900073.

[86] Luis E. Ibanez, Javier Mas, Hans-Peter Nilles, and Fernando Quevedo. Heterotic Strings in Symmetric and Asymmetric Orbifold Backgrounds. *Nucl. Phys.*, B301:157–196, 1988. doi: 10.1016/0550-3213(88)90166-6.

[87] Lance J. Dixon, Jeffrey A. Harvey, C. Vafa, and Edward Witten. Strings on Orbifolds. *Nucl. Phys.*, B261:678–686, 1985. doi: 10.1016/0550-3213(85)90593-0. [,678(1985)].

[88] Lance J. Dixon, Jeffrey A. Harvey, C. Vafa, and Edward Witten. Strings on Orbifolds. 2. *Nucl. Phys.*, B274:285–314, 1986. doi: 10.1016/0550-3213(86)90287-7.

[89] Edward Witten. String theory dynamics in various dimensions. *Nucl. Phys.*, B443:85–126, 1995. doi: 10.1016/0550-3213(95)00158-O. [,333(1995)].

[90] Timo Weigand. F-theory. *PoS*, TASI2017:016, 2018.

[91] Cumrun Vafa. Evidence for F theory. *Nucl. Phys.*, B469:403–418, 1996. doi: 10.1016/0550-3213(96)00172-1.

[92] David R. Morrison and Cumrun Vafa. Compactifications of F theory on Calabi-Yau threefolds. 2. *Nucl. Phys.*, B476:437–469, 1996. doi: 10.1016/0550-3213(96)00369-0.

[93] David R. Morrison and Cumrun Vafa. Compactifications of F theory on Calabi-Yau threefolds. 1. *Nucl. Phys.*, B473:74–92, 1996. doi: 10.1016/0550-3213(96)00242-8.

[94] John H. Schwarz. An SL(2,Z) multiplet of type IIB superstrings. *Phys. Lett.*, B360:13–18, 1995. doi: 10.1016/0370-2693(95)01405-5,10.1016/0370-2693(95)01138-G. [Erratum: Phys. Lett.B364,252(1995)].

[95] Paul S. Aspinwall. Some relationships between dualities in string theory. *Nucl. Phys. Proc. Suppl.*, 46:30–38, 1996. doi: 10.1016/0920-5632(96)00004-7.

[96] Michael Dine, V. Kaplunovsky, Michelangelo L. Mangano, C. Nappi, and N. Seiberg. Superstring Model Building. *Nucl. Phys.*, B259:549–571, 1985. doi: 10.1016/0550-3213(85)90001-X.

[97] David J. Gross, Jeffrey A. Harvey, Emil J. Martinec, and Ryan Rohm. Heterotic String Theory. 1. The Free Heterotic String. *Nucl. Phys.*, B256:253, 1985. doi: 10.1016/0550-3213(85)90394-3.

[98] Luis E. Ibanez, F. Marchesano, and R. Rabadan. Getting just the standard model at intersecting branes. *JHEP*, 11:002, 2001. doi: 10.1088/1126-6708/2001/11/002.

[99] G. Aldazabal, S. Franco, Luis E. Ibanez, R. Rabadan, and A. M. Uranga. D = 4 chiral string compactifications from intersecting branes. *J. Math. Phys.*, 42:3103–3126, 2001. doi: 10.1063/1.1376157.

[100] G. Aldazabal, S. Franco, Luis E. Ibanez, R. Rabadan, and A. M. Uranga. Intersecting brane worlds. *JHEP*, 02:047, 2001. doi: 10.1088/1126-6708/2001/02/047.

[101] Ralph Blumenhagen, Lars Goerlich, Boris Kors, and Dieter Lust. Noncommutative compactifications of type I strings on tori with magnetic background flux. *JHEP*, 10:006, 2000. doi: 10.1088/1126-6708/2000/10/006.

[102] Mirjam Cvetic, Gary Shiu, and Angel M. Uranga. Chiral four-dimensional N=1 supersymmetric type 2A orientifolds from intersecting D6 branes. *Nucl. Phys.*, B615:3–32, 2001. doi: 10.1016/S0550-3213(01)00427-8.

[103] Mirjam Cvetic, Gary Shiu, and Angel M. Uranga. Three family supersymmetric standard - like models from intersecting brane worlds. *Phys. Rev. Lett.*, 87:201801, 2001. doi: 10.1103/PhysRevLett.87.201801.

[104] Bobby Samir Acharya and Edward Witten. Chiral fermions from manifolds of G(2) holonomy. 2001.

[105] Michael Atiyah and Edward Witten. M theory dynamics on a manifold of G(2) holonomy. *Adv. Theor. Math. Phys.*, 6:1–106, 2003. doi: 10.4310/ATMP.2002.v6.n1.a1.

[106] M. Bershadsky, C. Vafa, and V. Sadov. D-branes and topological field theories. *Nucl. Phys.*, B463: 420–434, 1996. doi: 10.1016/0550-3213(96)00026-0.

[107] Sheldon H. Katz and Cumrun Vafa. Matter from geometry. *Nucl. Phys.*, B497:146–154, 1997. doi: 10.1016/S0550-3213(97)00280-0.

[108] Mirjam Cvetic, Gary Shiu, and Angel M. Uranga. Chiral type II orientifold constructions as M theory on G(2) holonomy spaces. In *Supersymmetry and unification of fundamental interactions. Proceedings, 9th International Conference, SUSY'01, Dubna, Russia, June 11-17, 2001*, pages 317–326, 2001. doi: 10.1142/9789812778192_0057.

[109] Th. Kaluza. Zum Unitätsproblem der Physik. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, 1921:966–972, 1921. doi: 10.1142/S0218271818700017. [Int. J. Mod. Phys.D27, no.14,1870001(2018)].

[110] Oskar Klein. Quantum Theory and Five-Dimensional Theory of Relativity. (In German and English). *Z. Phys.*, 37:895–906, 1926. doi: 10.1007/BF01397481. [,76(1926)].

[111] R. Slansky. Group Theory for Unified Model Building. *Phys. Rept.*, 79:1–128, 1981. doi: 10.1016/0370-1573(81)90092-2.

[112] Ling Lin and Timo Weigand. Towards the Standard Model in F-theory. *Fortsch. Phys.*, 63(2): 55–104, 2015. doi: 10.1002/prop.201400072.

[113] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999. doi: 10.1103/PhysRevLett.83.3370.

[114] Lisa Randall and Raman Sundrum. An Alternative to compactification. *Phys. Rev. Lett.*, 83: 4690–4693, 1999. doi: 10.1103/PhysRevLett.83.4690.

[115] Hao Yu, Zi-Chao Lin, and Yu-Xiao Liu. Gravitational waves and extra dimensions: a short review. *Commun. Theor. Phys.*, 71(8):991–1006, 2019. doi: 10.1088/0253-6102/71/8/991.

[116] Kris Pardo, Maya Fishbach, Daniel E. Holz, and David N. Spergel. Limits on the number of spacetime dimensions from GW170817. *JCAP*, 1807(07):048, 2018. doi: 10.1088/1475-7516/2018/07/048.

[117] Luca Visinelli, Nadia Bolis, and Sunny Vagnozzi. Brane-world extra dimensions in light of GW170817. *Phys. Rev.*, D97(6):064039, 2018. doi: 10.1103/PhysRevD.97.064039.

[118] M. Aaboud et al. Search for heavy ZZ resonances in the $\ell^+\ell^-\ell^+\ell^-$ and $\ell^+\ell^-\nu\bar{\nu}$ final states using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J.*, C78(4):293, 2018. doi: 10.1140/epjc/s10052-018-5686-3.

[119] Jan Kretzschmar. Searches for extra dimensions with the ATLAS and CMS detectors. *Nucl. Part. Phys. Proc.*, 273-275:541–545, 2016. doi: 10.1016/j.nuclphysbps.2015.09.080.

[120] Georges Aad et al. Search for Quantum Black Hole Production in High-Invariant-Mass Lepton+Jet Final States Using pp Collisions at $\sqrt{s} = 8$ TeV and the ATLAS Detector. *Phys. Rev. Lett.*, 112 (9):091804, 2014. doi: 10.1103/PhysRevLett.112.091804.

[121] Walter D. Goldberger, Yasunori Nomura, and David Tucker-Smith. Warped supersymmetric grand unification. *Phys. Rev.*, D67:075021, 2003. doi: 10.1103/PhysRevD.67.075021.

[122] V. A. Rubakov and M. E. Shaposhnikov. Do We Live Inside a Domain Wall? *Phys. Lett.*, 125B: 136–138, 1983. doi: 10.1016/0370-2693(83)91253-4.

[123] Ignatios Antoniadis, Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. New dimensions at a millimeter to a Fermi and superstrings at a TeV. *Phys. Lett.*, B436:257–263, 1998. doi: 10.1016/S0370-2693(98)00860-0.

[124] L. P. Kadanoff. Scaling laws for Ising models near T(c). *Physics Physique Fizika*, 2:263–272, 1966. doi: 10.1103/PhysicsPhysiqueFizika.2.263.

[125] Tom Kennedy. Introduction to Mathematical Physics. 2008.

[126] Kenneth G. Wilson. The Renormalization Group: Critical Phenomena and the Kondo Problem. *Rev. Mod. Phys.*, 47:773, 1975. doi: 10.1103/RevModPhys.47.773.

[127] Joseph Polchinski. Renormalization and Effective Lagrangians. *Nucl. Phys.*, B231:269–295, 1984. doi: 10.1016/0550-3213(84)90287-6.

[128] Oliver J. Rosten. Fundamentals of the Exact Renormalization Group. *Phys. Rept.*, 511:177–272, 2012. doi: 10.1016/j.physrep.2011.12.003.

[129] C. Bagnuls and C. Bervillier. Exact renormalization group equations. An Introductory review. *Phys. Rept.*, 348:91, 2001. doi: 10.1016/S0370-1573(00)00137-X.

[130] Jan de Boer. The Holographic renormalization group. *Fortsch. Phys.*, 49:339–358, 2001. doi: 10.1002/1521-3978(200105)49:4/6<339::AID-PROP339>3.0.CO;2-A.

[131] Kostas Skenderis. Lecture notes on holographic renormalization. *Class. Quant. Grav.*, 19: 5849–5876, 2002. doi: 10.1088/0264-9381/19/22/306.

[132] Leonard Susskind and Edward Witten. The Holographic bound in anti-de Sitter space. 1998.

[133] Juan Martin Maldacena. Large N field theories, string theory and gravity. In *Topics in mathematical physics, general relativity and cosmology. Proceedings, International Conference, Mexico City, Mexico, September 17-20, 2002*, pages 301–315, 2002.

[134] Claude Elwood Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27: 379–423, 1948. [Bell Syst. Tech. J.27,623(1948)].

[135] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951. doi: doi:10.1214/aoms/1177729694.

[136] S. Kullback. *Information Theory and Statistics*. John Wiley & Sons, 1959. ISBN 0844656259.

[137] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2006. ISBN 0471241954.

[138] Vijay Balasubramanian. Statistical inference, occam's razor and statistical mechanics on the space of probability distributions. 1996.

[139] S.-I. Amari and H. Nagaoka. *Methods of Information Geometry*, volume vol. 191 of *Translations of Mathematical Monographs*. American Mathematical Society and Oxford University Press, Providence, RI, 2000.

[140] V. Vedral. The role of relative entropy in quantum information theory. *Rev. Mod. Phys.*, 74: 197–234, 2002. doi: 10.1103/RevModPhys.74.197.

[141] V. Giovannetti, S. Montangero, and Rosario Fazio. Quantum multiscale entanglement renormalization ansatz channels. *Phys. Rev. Lett.*, 101:180503, Oct 2008. doi: 10.1103/PhysRevLett.101.180503. URL https://link.aps.org/doi/10.1103/PhysRevLett.101.180503.

[142] G. Vidal. Entanglement Renormalization. *Phys. Rev. Lett.*, 99(22):220405, 2007. doi: 10.1103/PhysRevLett.99.220405.

[143] G. Vidal. Class of Quantum Many-Body States That Can Be Efficiently Simulated. *Phys. Rev. Lett.*, 101:110501, 2008. doi: 10.1103/PhysRevLett.101.110501.

[144] P. Di Francesco, P. Mathieu, and D. Senechal. *Conformal Field Theory*. Graduate Texts in Contemporary Physics. Springer-Verlag, New York, 1997. ISBN 9780387947853, 9781461274759. doi: 10.1007/978-1-4612-2256-9. URL http://www-spires.fnal.gov/spires/find/books/www?cl=QC174.52.C66D5::1997.

[145] H. A. Kramers and G. H. Wannier. Statistics of the two-dimensional ferromagnet. part i. *Phys. Rev.*, 60:252–262, Aug 1941. doi: 10.1103/PhysRev.60.252. URL https://link.aps.org/doi/10.1103/PhysRev.60.252.

[146] R. K. Pathria. *Statistical Mechanics*. Butterworth-Heinemann, 1996. ISBN 9780080541716.

[147] Lauren McGough, Márk Mezei, and Herman Verlinde. Moving the CFT into the bulk with $T\bar{T}$. 2016.

[148] Amit Giveon, Nissan Itzhaki, and David Kutasov. $\mathrm{T\overline{T}}$ and LST. *JHEP*, 07:122, 2017. doi: 10.1007/JHEP07(2017)122.

[149] F. A. Smirnov and A. B. Zamolodchikov. On space of integrable quantum field theories. *Nucl. Phys.*, B915:363–383, 2017. doi: 10.1016/j.nuclphysb.2016.12.014.

[150] Allan Adams, Nima Arkani-Hamed, Sergei Dubovsky, Alberto Nicolis, and Riccardo Rattazzi. Causality, analyticity and an IR obstruction to UV completion. *JHEP*, 10:014, 2006. doi: 10.1088/1126-6708/2006/10/014.