Knots, trees, and fields: common ground between physics and mathematics

Thesis by
Ingmar Akira Saberi

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy in Physics

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2017
Defended June 16, 2016

To the memory of John F. Orsborn

ACKNOWLEDGEMENTS

This thesis represents the culmination of a five-year sojourn at Caltech. During that time, and throughout my life, I have been incomparably fortunate to have the love, kindness, generosity, and support of many wonderful people. Nothing would have been possible without you all. As much as I hope I will mention each of you here, I am quite confident I will fail; this reflects no lack of gratitude on my part, but only an (unfortunately characteristic) lack of due diligence.

First and foremost, I owe a profound debt to my advisor, Sergei Gukov. I learned a tremendous amount from him, and his steadfast support throughout my time in graduate school allowed me to learn and accomplish as much as I have. Sergei, I am enormously lucky and grateful to have had the chance to be your student.

In addition, I would like to thank from the bottom of my heart the other people I have had the pleasure of collaborating with on the work that makes up this thesis: Matt Heydeman, Hyungrok Kim, Matilde Marcolli, Satoshi Nawata, Bogdan Stoica, Marko Stošić, and Piotr Sułkowski. Each of you was a valued teacher and coworker, and I quite simply could not have done it without you. Many, many thanks.

Matilde deserves particular mention for her service on my committee. I am profoundly grateful to her, and to the committee's other members, Maria Spiropulu and John Schwarz, for support, friendship, and advice, and for steadfastly standing in my corner. Also, Maria, special thanks to you and also to Sarah Reisman: it has been a pleasure to have undeservedly been a part of so many of both your respective groups' activities.

As I was disappointed to learn in graduate school, science, like any other form of art, requires patronage to pay the bills. I owe deepest thanks and appreciation to the United States Department of Energy, which supported the work represented here through the grants DE-FG03-92-ER40701FG-02 and DE-SC-0011632, and likewise the National Science Foundation (through the grant PHY-0757647). By extension, I am indebted to the citizens and taxpayers of the United States of America, for their collective commitment to basic science and mathematics as national pursuits; I thank them also for maintaining—under the auspices of the National Park Service, United States Forest Service,

and Bureau of Land Management—its public lands and wilderness preserves in perpetuity (and, as such, for my own and everyone else's sanity). I must also gratefully acknowledge the Walter Burke Institute for Theoretical Physics at Caltech, of which I have been a part since its inception a few years ago, for its stimulating research environment; thanks are due to all its faculty, and in particular to D. Politzer for support and for many invaluable stories. Also, no mention of the Burke Institute would be complete without an appreciative nod to Carol Silberstein, without whose tireless efforts everything would have come apart at the seams long ago.

During the course of my Ph.D. work, I have been lucky to be able to travel to numerous meetings and learn an enormous amount from researchers outside my immediate sphere. This would not have been possible without the generous and largely undeserved hospitality of a number of institutions and conferences: the University of Notre Dame (with special gratitude to Stephan Stolz); C.R.M. at the Université de Montréal; I.C.M.A.T. Madrid; the Institute for Advanced Study; Princeton University; P.I.M.S. at the University of British Columbia; the Simons Center for Geometry and Physics; the University of California at Berkeley; I.M.P.A. in Rio de Janeiro (and especially Julio Andrade); the American Institute of Mathematics, which supported the collaboration leading to [109] with a SQuaRE grant; the Partnership Mathematics and Physics at Universität Heidelberg; the University of Bristol (thanks in particular to Jon Keating, for friendship and support at a difficult time); C.U. Boulder and the Gone Fishing meeting; and the Berkeley Institute for Theoretical Physics.

Moving onward: I have learned more than I can say from the peers, colleagues, and friends around Caltech who made it such a wonderful place to spend the past five years. For conversations and camaraderie around the fourth floor, I thank D. Aasen, J. Arnold, M. Beverland, K. Boddy, N. Dawra, R. Elliot, S. He, E. Herrmann, N. Hunter-Jones, M. Koloğlu, P. Kravchuk, B. Michel, C.-H. Shen, P. Solis, R. Thorngren, D. Pei, I. Yaakov, and all the other denizens. Special thanks to A. Turzillo for being willing to go backpacking at the drop of a hat. For providing a much-needed counterbalance to work, I am grateful to D. Anderson, B. Babin and V. Heckman (who introduced me to the wonders of the 24-hour spa), D. Bickerstaff, D. Chao, B. Daniels, M. Hanewich-Hollatz, M. Hesse, K. Hodge, R. Katti, S. Martis, L. Rios and I. Finneran, A. Swaminathan, and (last but by no means least) K. Zemlianova.

I owe special mention to my long-time office mate (and more recent roommate), T. McKinney, from whom I learned an enormous amount about physics and even more about the Abrahamic concept of land ownership, rituals of performative alterity among the Hopi, and a panoply of other topics. Thanks to you for putting up with me so frequently, as well as to all other aficionados of card-playing and pretentious movies: A. Arjun, B. Pang, K. Chałupka (mój przyjaciel), L. Sun, and the new arrival (mazel tov!).

Thanks to LKV for being an old friend and familiar face in L.A., to N. Kadunce and J. Hammond for becoming (albeit by a slim margin) my first new friends here and for five years of stalwart companionship and good food, to A. Jackson for getting me back into the garden, to *mi primo* Ernie for five years of lunches, to YGB for dependability and brotherhood (don't be as much of a stranger next time!), to NAD for (co?)-originating "Handles; Messiah," and to bandmates and travel companions C. Peña and C. Fernandez (congratulations to you both!).

To the close friends, companions, and roommates I have been lucky to count as family by any other name: thank you for everything over the last five years. JMGD and KLJ, good luck to you both in Chicago; the Midwest just got a little more young and restless. EAQ, EGG (couldn't resist), and Sofia Celeste, I aspire to have a family as beautiful as yours.

Onward. Love and gratitude to the happy couple, JCL and YDN—eagerly looking forward to celebrating with you on Labor Day; to LPM, for the countless ways in which you have enriched and continue to enrich my life—congratulations on hitting the big time; to BDW for being a human being of the very first water and for literally thousands of dollars' worth of accommodation, for which I promise I will someday find a way to repay you; to BEK for being a first-rate mathematician, for commiseration, and for continuing to costar in "Separated at Birth;" to ORK for going to buy stamps with me on that fateful day in 2007; and finally to DCG, for bringing great joy into my life these past months.

It remains only to thank my family: my mother and father, Lulu, Dmitri, and Mirabella. But words are scarcely adequate to that task.

ABSTRACT

One main theme of this thesis is a connection between mathematical physics (in particular, the three-dimensional topological quantum field theory known as Chern-Simons theory) and three-dimensional topology. This connection arises because the partition function of Chern-Simons theory provides an invariant of three-manifolds, and the Wilson-loop observables in the theory define invariants of knots. In the first chapter, we review this connection, as well as more recent work that studies the classical limit of quantum Chern-Simons theory, leading to relations to another knot invariant known as the A-polynomial. (Roughly speaking, this invariant can be thought of as the moduli space of flat $\mathrm{SL}(2,\mathbb{C})$ connections on the knot complement.) In fact, the connection can be deepened: through an embedding into string theory, categorifications of polynomial knot invariants can be understood as spaces of BPS states.

We go on to study these homological knot invariants, and interpret spectral sequences that relate them to one another in terms of perturbations of supersymmetric theories. Our point is more general than the application to knots; in general, when one perturbs any modulus of a supersymmetric theory and breaks a symmetry, one should expect a spectral sequence to relate the BPS states of the unperturbed and perturbed theories. We consider several diverse instances of this general lesson. In another chapter, we consider connections between supersymmetric quantum mechanics and the de Rham version of homotopy theory developed by Sullivan; this leads to a new interpretation of Sullivan's minimal models, and of Massey products as vacuum states which are entangled between different degrees of freedom in these models.

We then turn to consider a discrete model of holography: a Gaussian lattice model defined on an infinite tree of uniform valence. Despite being discrete, the matching of bulk isometries and boundary conformal symmetries takes place as usual; the relevant group is $\mathrm{PGL}(2,\mathbb{Q}_p)$, and all of the formulas developed for holography in the context of scalar fields on fixed backgrounds have natural analogues in this setting. The key observation underlying this generalization is that the geometry underlying $\mathrm{AdS}_3/\mathrm{CFT}_2$ can be understood algebraically, and the base field can therefore be changed while maintaining much of the structure. Finally, we give some analysis of A-polynomials under change of base (to finite fields), bringing things full circle.

PUBLISHED CONTENT AND CONTRIBUTIONS

Much of this thesis is adapted from articles that originally appeared in other forms. All are available as arXiv preprints; where appropriate, the DOI for the published version is included as a hyperlink from the journal citation. (Should the reader wish, hovering the mouse over the link will reveal the explicit DOI.)

[3] (adapted for chapter 1) is a review article, based on summer-school lectures given by my advisor; I was responsible for exposition and for preparing the manuscript from rough lecture notes. The remaining chapters are original research articles: [5] appears as chapter 2 and [4] as chapter 4, while chapter 3 is adapted from those sections of [1] from which I was most intimately involved. Chapter 5 consists of the draft of an unpublished article detailing another research project I undertook in graduate school. All authors contributed equally to [4] and [5]; I was largely responsible for the content of [2].

[1] S. Gukov, S. Nawata, I. Saberi, M. Stošić, and P. Sułkowski. "Sequencing BPS spectra." JHEP **03**, 004. (2016). arXiv:1512.07883 [hep-th].

[2] S. Gukov and I. Saberi. "Motivic quantum field theory and the number theory of A-polynomial curves." To appear.

[3] S. Gukov and I. Saberi. "Lectures on knot homology and quantum curves." In *Topology and field theories*. Edited by S. Stolz. Contemporary Mathematics 613. (American Mathematical Society. 2014). arXiv:1211.6075 [hep-th].

[4] M. Heydeman, M. Marcolli, I. Saberi, and B. Stoica. "Tensor networks, p-adic fields, and algebraic curves: arithmetic and the AdS_3/CFT_2 correspondence." (2016). arXiv:1605.07639 [hep-th].

[5] H. Kim and I. Saberi. "Real homotopy theory and supersymmetric quantum mechanics." (2015). arXiv:1511.00978 [hep-th].

CONTENTS

LIST OF FIGURES

I will take an egg out of the robin's nest in the orchard,
I will take a branch of gooseberries from the old bush in the garden,
and go and preach to the world;
You shall see I will not meet a single heretic or scorner,
You shall see how I stump clergymen, and confound them,
You shall see me showing a scarlet tomato, and a white pebble
from the beach.

—*Walt Whitman*

INTRODUCTION

This thesis consists of work that was originally published during my graduate-school career in the form of a series of distinct papers [109, 111, 119, 133], as well as a small amount of material from work that has yet to appear [110]. As such, the various chapters range across several different and seemingly distant topics; nonetheless, I feel that certain key themes underlie and unify all of it. Chief among these themes is the idea of bridging the gap between the disciplines now termed theoretical physics and pure mathematics.

Historically speaking, no distinction was generally made between these two disciplines until the twentieth century. As the pace of scientific discovery and concomitant specialization accelerated—and, in particular, as the project of quantum field theory came to bear fruit, requiring physicists to accept techniques such as renormalization whose basis in rigorous mathematics was obscure at best—the gulf between the two grew to appear nearly insurmountable. In 1972, Freeman Dyson quipped in his Gibbs memorial lecture that, "[as] a working physicist, I am acutely aware of the fact that the marriage between mathematics and physics, which was so enormously fruitful in past centuries, has recently ended in divorce" [73]. Dyson's classic lecture goes on to describe various instances in the history of science when the division between these two fields impoverished both. In many cases, a subject of interest to one group would have been advanced considerably by an influx of the ideas or the expertise of the other; the advances in question were uncovered years—and sometimes many years—later than they might have been. In closing, he recommended a selection of problems from quantum field theory and quantum gravity to the attention of pure mathematicians.

Dyson was prophetic. The renewal of the dialogue between these two disciplines began in earnest not long afterwards. While there are many stones at which one might point to mark the initial moment, C. N. Yang's 1977 paper [223], which pointed out the intimate connection between the physics of gauge theories and the theory of fiber bundles studied in mathematics, seems as good as any (at least from the viewpoint of someone who was not yet born at the time).

At any rate, the *rapprochement* was well begun by the early eighties, when

Donaldson's landmark 1983 paper [70] used ideas from gauge theory to construct a powerful new invariant of smooth structures on four-manifolds. The key idea of his work is simple enough to describe: One studies the smooth space X by first producing a certain auxiliary space, $\mathscr{M}(X)$, which is the moduli space of solutions to certain partial differential equations (the anti-self-dual instanton equations) on X. The topology of this moduli space depends sensitively on the smoothness structure with respect to which the PDEs are defined; relatively crude topological invariants of $\mathscr{M}(X)$, such as its homology, then become sophisticated and subtle invariants of X as a smooth space. (Other papers from around the same time [5, 27] also evidence the renewed mutual interest among members of both camps.)

Gauge theory soon made other related appearances in mathematics. Several years later, using similar PDEs, Andreas Floer [80–82] defined the homology theories for three-manifolds and Lagrangian intersections that bear his name. It was noticed that the generalizations of Donaldson's invariants to a four-manifold X with boundary $\partial X = Y$ were naturally valued in the Floer homology groups of Y. In 1988, Edward Witten was able to interpret these constructions from a physical perspective, showing that Donaldson's invariants were naturally computed by a particular truncation of a supersymmetric Yang-Mills theory. Almost simultaneously, Sir Michael Atiyah [7, 8] gave axioms for the general notion of a *topological quantum field theory*.

Topological field theory, or TQFT, is a fitting synecdoche for the interdisciplinary field that has continued to grow out of these early papers. TQFTs are of interest to physicists as toy models that possess many of the essential features of quantum field theory—or as truncated sectors of field theories of physical interest—while remaining simple enough to be mathematically rigorous and exactly solvable. Mathematicians usually think of TQFT as an organizing principle for existing invariants of manifolds, or a source of new invariants. As evidenced by Donaldson's theory, attempting to classify geometric spaces using the data of how physical theories behave on those spaces has proved enormously fruitful.

From either perspective, the definition of a topological field theory makes intuitive sense, once one has chewed it over a bit. The geometrical setting in which such a theory is defined is a *cobordism category*, in which the objects are compact oriented $(d-1)$-manifolds Y, and the morphisms from Y_0 to Y_1

are given by d-manifolds X with boundary $\bar{Y}_0 \sqcup Y_1$ (where the bar denotes orientation reversal). Put succinctly, the morphisms are oriented cobordisms. One is to imagine that, as in a usual physical theory, the state of the system is specified on a codimension-one surface playing the role of a constant-time slice, and the cobordism plays the role of a generalized time-evolution during which topology change is possible.

As is usual in quantum mechanics, the data of the theory consists of kinematic data (a description of its possible states) together with dynamical data (how those states evolve in time). The collection of possible states of the system on a space Y_i forms a Hilbert space, which we will denote $\mathscr{H}_i$. Time evolution provides a mapping of states, so that we should be able to associate to each cobordism X a mapping $\phi_X : \mathscr{H}_0 \to \mathscr{H}_1$. There are additional requirements and possible variants of these axioms, but the essential structure is hopefully clear.

In more mathematical language, one says that a TQFT is a *symmetric monoidal functor* from a geometric category (the cobordism category $\mathbf{Cob}(d)$ described in the previous paragraphs) to the linear category **Vect**. "Symmetric monoidal" just means that the functor should preserve tensor products: it carries the disjoint union of manifolds and cobordisms between them to the usual tensor product of Hilbert spaces and linear maps. In the special case where the d-manifold X has empty boundary, one gets a morphism between empty $(d-1)$-manifolds; this amounts to a linear map between copies of $\mathbb{C}$, which is a numerical invariant.

TQFTs fit into the general theme in mathematics of studying complicated, abstract, or geometric objects, such as groups, by considering their linear representations. Suppose that one were to consider an ordinary group G, viewed as a category with one object and only invertible morphisms. One could then ask for a functor from this category to **Vect**. This would amount to a choice of vector space V, to be associated to the object, and a homomorphism from G (which is the automorphisms of the unique object in the source category) to the linear automorphism group of V. In other words, one recovers the usual notion of a linear representation on an arbitrary vector space! It is easy to see from this description that a TQFT generalizes the notion of linear representation to a rather different and more complicated source category.

Physicists have often been able to provide new understanding of the invariants

defined by topological field theories, in the form of surprising and unexpected conjectures. This is possible because the physical understanding of such an invariant is rarely as an isolated construction. Rather, the TQFT describes one piece of, or calculation in, a much larger and more baroque structure: a quantum field theory. If one believes that the quantum field theory makes sense, and can see how a mathematical object fits in as a piece of it, then one has many new angles from which to look at that object, and many new relations to expect between it and other pieces of the QFT.

One interesting example of an advance that was made using this kind of reasoning took place in 1994, when Seiberg and Witten [183, 184] analyzed the infrared behavior of the $\mathcal{N} = 2$ supersymmetric SU(2) Yang-Mills theory that Witten had previously related to Donaldson's invariants. Using standard physical techniques, they argued that the theory should undergo a phase transition to a symmetry-breaking phase at low energy scales, in which the nonabelian interactions of the SU(2) theory should be replaced by an effective description in terms of abelian U(1) gauge theory. One can think of this transition as analogous to electroweak symmetry breaking in the Standard Model of particle physics. However, Witten had previously argued that the twist of this theory that produced Donaldson's invariants was independent of scale; it followed that the invariants should be able to be computed using the simpler effective physics in the infrared [214]. For mathematicians, the end result was a new set of seemingly unrelated partial differential equations (the Seiberg–Witten equations), which were far easier to deal with than Donaldson's equations, but which could be used in similar fashion to compute invariants carrying nearly identical information.

Another exemplar of this pattern happens to lie close to my own work. In 1984, Jones [123] defined a new and powerful polynomial invariant of knots; this was generalized to a family of "quantum group invariants," associated to the data of a semisimple Lie algebra $\mathfrak{g}$ and a representation R of $\mathfrak{g}$. Witten [212] gave the first physical construction of these quantum group invariants, identifying them as the expectation values of Wilson loop operators carrying the representation R in Chern-Simons theory with gauge group G, considered as functions of the level. (The level parameter k is the coupling constant of Chern-Simons theory, which must be an integer for the theory to be gauge-invariant.) In the case $G = \mathrm{SU}(2)$, Gukov [107] identified the knot invariant corresponding

to classical Chern-Simons theory, which is called the A-polynomial. The zero locus of this polynomial is the moduli space of solutions to the equations of motion in $\mathrm{SL}(2,\mathbb{C})$ Chern-Simons theory. He further observed that quantization should associate to this space a recurrence relation (the analogue of the Schrödinger equation), whose solution is the set of "colored" Jones polynomials associated to arbitrary representations of $G = \mathrm{SU}(2)$. This prediction (now verified in special cases and checked in many examples) is known as the AJ-conjecture.

In Chapter 1, taken from the work [111] by Gukov and myself, we give a pedagogical review of these developments, discussing the connections between knot theory and knot invariants, topological field theory, and quantization. We also treat newer developments in the subject, related to *categorified* quantum group invariants—*i.e.*, knot homologies. From the perspective of physics, knot homologies can be understood as spaces of BPS states in certain theories that can be geometrically engineered from M-theory, taking the data of the knot as an input. (This interpretation is reviewed later in the thesis, in §3.2.)

Chapter 2, written with Hyungrok Kim [133], explores the theme of using the states or operators of a physical theory defined on a geometric space as an algebraic model of that space. We focus on the simplest possible example, supersymmetric quantum mechanics, pointing out that the approach to homotopy theory over a field pioneered by Sullivan [190] is closely connected to physics, and giving a new interpretation of Massey products in topology as vacuum states in certain quantum-mechanical models (Sullivan's minimal models) that are entangled between different degrees of freedom.

I used some of the subtleties about supersymmetric quantum mechanics learned from that project in the course of my next work [109], with Gukov, Nawata, Stošić, and Sułkowski. The sections of that paper with which I was most involved are reproduced in Chapter 3; due to its prohibitive length, the reader interested in the remainder of the work is referred to the original paper. We were interested in coming to a physical understanding of *deformation spectral sequences* between different knot homology theories; in the end, we formulated the general proposal that spectral sequences describe the effects of perturbations of theories on their BPS spectra, in particular those that break symmetries and partially lift BPS degeneracies. While this pertains in particular to theories whose BPS spectra reproduce knot homologies, the point is much

broader and has nothing to do with knots *per se*.

In Chapter 4, based on [119] and written with Heydeman, Marcolli, and Stoica, the same theme is expressed in a much different context. We were interested in the question of how best to construct models of holography in which spacetime is discrete. This has been of interest recently in light of speculation that tensor-network models, which reproduce certain features of conformal field theory such as the entanglement entropy of the ground state, can be understood as discrete analogues of hyperbolic bulk geometry. Our models appeal to a different angle of generalization, inspired by previous work of Manin and Marcolli: we formulate $\mathrm{AdS}_3/\mathrm{CFT}_2$ as algebraically as possible, and then consider changing the underlying field in the construction of the spacetime to be a p-adic field rather than $\mathbb{C}$. The resulting bulk geometry is a p-adic maximally symmetric space: the Bruhat–Tits tree of $\mathrm{SL}(2, \mathbb{Q}_p)$. While this space is naturally discrete, it does not break any symmetries; one can still make sense of both bulk isometries and boundary conformal transformations, and identify the two as is required for the normal holographic dictionary to work.

We demonstrate that, at least for massive free scalar fields in a fixed background geometry, the holographic dictionary can be developed exactly as in the ordinary case. Certain features of AdS/CFT, such as the correspondence between the holographic direction and the boundary RG scale, are present even more sharply in our models than in the ordinary case. We also give arguments that logarithmic scaling of the entanglement entropy should hold; given this, an analogue of the Ryu-Takayanagi formula follows trivially from the geometry of the tree. Based on this, we argue that the p-adic geometry (rather than the archimedean geometry) is the natural setting in which to make contact with tensor network models, and we offer some examples of constructions through which this might be done.

Finally, in Chapter 5, I discuss the first project I worked on in graduate school, which connects A-polynomials with a similar kind of arithmetic generalization of physics to those proposed in Chapter 4. This work remains unpublished, because it is largely experimental in nature and what results there are are highly speculative; nonetheless, I offer some thought on directions for future work along these lines.

Chapter 1

QUANTIZATION AND CATEGORIFICATION IN TOPOLOGICAL QUANTUM FIELD THEORY

[1]S. Gukov and I. Saberi. "Lectures on knot homology and quantum curves." In *Topology and field theories*. Edited by S. Stolz. Contemporary Mathematics 613. (American Mathematical Society. 2014). arXiv:1211.6075 [hep-th].

Foreword

The lectures in this chapter will by no means serve as a complete introduction to the two topics of *quantization* and *categorification*. These words represent not so much single ideas as broad tools, programs, or themes in physics and mathematics; both are areas of active research, and are still not fully understood. One could easily give a full one-year course on each topic separately.

Rather, the goal of these lectures is to serve as an appetizer: to give a glimpse of the ideas behind quantization and categorification, by focusing on very concrete examples and giving a working knowledge of how these ideas are manifested in simple cases. It is our hope that the resulting discussion will remain accessible and clear while shedding at least some light on these complex ideas, and that the interest of the reader will be piqued.

Imagine the category of finite-dimensional vector spaces and linear maps. To each object in this category is naturally associated a number, the dimension of that vector space. Replacing some collection of vector spaces with a collection of numbers in this way can be thought of as a *decategorification*: by remembering only the dimension of each space, we keep some information, but lose all knowledge about (for instance) morphisms between spaces. In this sense, decategorification forgets about geometry.

Categorification can be thought of as the opposite procedure. Given some piece of information (an invariant of a topological space, for instance), one asks whether it arises in some natural way as a "decategorification": a piece of data extracted out of a more geometrical or categorical invariant, which may carry more information and thus be a finer and more powerful tool. An answer

in the affirmative to this question is a categorification of that invariant.

Perhaps the most familiar example of categorification at work is the reinterpretation of the Euler characteristic as the alternating sum of ranks of homology groups,

$$\chi(M) = \sum_{k\geq 0} (-1)^k \operatorname{rank} H_k(M)\,. \tag{1.1}$$

In light of this formula, the homology of a manifold M can be seen as a categorification of its Euler characteristic: a more sophisticated and richly structured bearer of information, from which the Euler characteristic can be distilled in some natural way. Moreover, homology theories are a far more powerful tool than the Euler characteristic alone for the study and classification of manifolds and topological spaces. This shows that categorification can be of practical interest: by trying to categorify invariants, we can hope to construct *stronger* invariants.

While the idea of categorification is rooted in pure mathematics, it finds a natural home in the realm of topological quantum field theory (TQFT), as will be discussed in §1.4. For this, however, we first need to understand what "quantum" means by explaining the quantization program, which originated squarely within physics. Its basic problem is the study of the transition between classical and quantum mechanics. The classical and quantum pictures of a physical system make use of entirely different and seemingly unconnected mathematical formalisms. In classical mechanics, the space of possible states of the system is a symplectic manifold, and observable quantities are smooth functions on this manifold. The quantum mechanical state space, on the other hand, is described by a Hilbert space $\mathscr{H}$, and observables are elements of a noncommutative algebra of operators acting on $\mathscr{H}$. Quantization of a system is the construction of the quantum picture of that system from a classical description, as is done in a standard quantum mechanics course for systems such as the harmonic oscillator and the hydrogen atom. Therefore, in some sense, quantization allows one to interpret quantum mechanics as "modern symplectic geometry." We will give a more full introduction to this idea in §1.3.

One main application of the ideas of quantization and categorification is to representation theory, where categorification, or "geometrization," leads naturally to the study of geometric representation theory [47]. Another area of mathematics where these programs bear much fruit is low-dimensional topol-

ogy, which indeed is often called "quantum" topology. This is the arena in which we will study the implications of quantization and categorification, primarily for the reason that it allows for many concrete and explicit examples and computations. Specifically, almost all of our discussion will take place in the context of knot theory. The reader should not, however, be deceived into thinking of our aims as those of knot theorists! We do not discuss quantization and categorification for the sake of their applications to knot theory; rather, we discuss knot theory because it provides a window through which we can try and understand quantization and categorification.

1.1 Why knot homology?

A *knot* is a smooth embedding of a circle S^1 as a submanifold of S^3:

$$k : S^1 \hookrightarrow S^3, \quad K \hat{=} \operatorname{im} k. \tag{1.2}$$

See *e.g.* Figs. 1.2 and 1.3 for some simple examples. Likewise, a *link* is defined as an embedding of several copies of S^1. Two knots are equivalent if the two embeddings k and k' can be smoothly deformed into one another through a family of embeddings, i.e., without self-intersections at any time. One should think of moving the knot around in the ambient space without breaking the string of which it is made.

In studying a knot, one usually depicts it using a planar *knot diagram:* this should be thought of as a projection of the knot from $\mathbb{R}^3 = S^3 \setminus \{\text{pt.}\}$, in which it lives, to some plane $\mathbb{R}^2 \subset \mathbb{R}^3$. Thus, a knot diagram is the image of an immersion of S^1 in $\mathbb{R}^2$, having only double points as singularities, and with extra data indicating which strand passes over and which under at each crossing. Examples of knot diagrams can be seen in Figs. 1.2, 1.3, and 1.4.

It should be clear that there is no unique diagram representing a given knot. We could obtain very different-looking pictures, depending on the exact embedding in $\mathbb{R}^3$ and on the choice of plane to which we project. Two knot diagrams should of course be seen as equivalent if they depict equivalent knots, but this equivalence could be nontrivial and difficult to see. The situation is made a little more tractable by a theorem of Reidemeister, which states that two knot diagrams are equivalent if and only if they can be transformed into one another by a sequence of three simple transformations. These basic transformations are the *Reidemeister moves*, which are depicted in Fig. 1.1, and

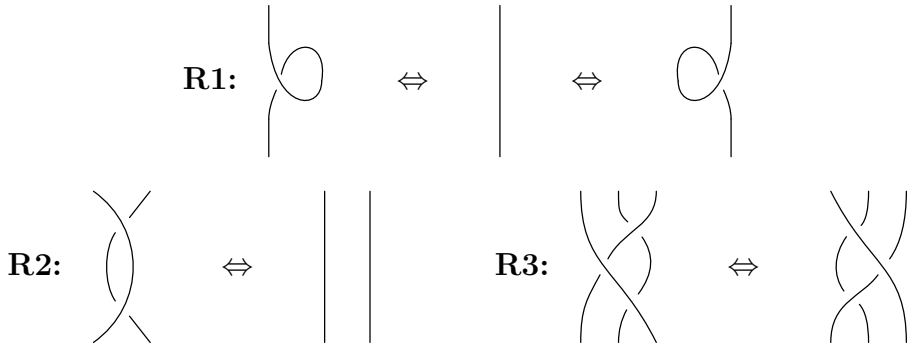


Figure 1.1: The three Reidemeister moves, which generate all equivalences between knot diagrams.

show replacements that can be made in any portion of a knot diagram to give an equivalent diagram.

Finding a sequence of Reidemeister moves that changes one given knot diagram into another, or showing that no such sequence exists, is still an *ad hoc* and usually intractable problem. As such, in attempting to classify knots, more clever methods are important. One of the most basic tools in this trade is a *knot invariant:* some mathematical object that can be associated to a knot, that is always identical for equivalent knots. In this way, one can definitively say that two knots are distinct if they possess different invariants. The converse, however, is not true; certain invariants may fail to distinguish between knots that are in fact different. Therefore, the arsenal of a knot theorist should contain a good supply of different invariants. Moreover, one would like invariants to be as "powerful" as possible; this just means that they should capture nontrivial information about the knot. Obviously, assigning the number 0 to every knot gives an invariant, albeit an extremely poor one!

Since one usually confronts a knot in the form of one of its representative knot diagrams, it is often desirable to have an invariant that can be efficiently computed from a knot diagram. Showing that some such quantity associated to a diagram is actually an invariant of knots requires demonstrating that it takes the same values on all equivalent diagrams representing the same knot. Reidemeister's theorem makes this easy to check: to show that we have defined a knot invariant, we need only check its invariance under the three moves in Fig. 1.1.*

*Nonetheless, since a knot is intrinsically an object of three-dimensional topology that

Given the goal of constructing knot invariants, it may be possible to do so most easily by including some extra structure to be used in the construction. That is, one can imagine starting not simply with a knot, but with a knot "decorated" with additional information: for instance, a choice of a Lie algebra $\mathfrak{g} = \mathrm{Lie}(G)$ and a representation R of $\mathfrak{g}$. It turns out that this additional input data from representation theory does in fact allow one to construct various invariants (numbers, vector spaces, and so on), collectively referred to as *quantum group invariants*. A large part of these lectures will consist, in essence, of a highly unorthodox introduction to these quantum group invariants.

The unorthodoxy of our approach is illustrated by the fact that we fail completely to address a natural question: what on earth do (for instance) the quantum $\mathfrak{sl}(N)$ invariants have to do with $\mathfrak{sl}(N)$? Representation theory is almost entirely absent from our discussion; we opt instead to look at an alternative description of the invariants, using a concrete combinatorial definition in terms of so-called *skein relations*. A more full and traditional introduction to the subject would include much more group theory, and show the construction of the quantum group invariants in a way that makes the role of the additional input data $\mathfrak{g}$ and R apparent [175, 212]. That construction involves assigning a so-called "quantum R-matrix" to each crossing in a knot diagram in some manner, and then taking a trace around the knot in the direction of its orientation. The connection to representation theory is made manifest; the resulting invariants, however, are the same.

Example 1. Suppose that we take an oriented knot together with the Lie algebra $\mathfrak{g} = \mathfrak{sl}(N)$ and its fundamental N-dimensional representation. With this special choice of extra data, one constructs the quantum $\mathfrak{sl}(N)$ invariant, denoted $P_N(K;q)$. Although it makes the connection to representation theory totally obscure, one can compute $P_N(K;q)$ directly from the knot diagram using the following skein relation:

$$q^N P_N(\,\diagup\!\!\diagdown\,) - q^{-N} P_N(\,\diagdown\!\!\diagup\,) = (q - q^{-1}) P_N(\,\bigcirc\,\bigcirc\,). \tag{1.3}$$

(Note that we will sometimes write $P_N(K)$ for the polynomial P_N associated to the knot or link K, suppressing the variable q; no confusion should arise.)

can be imagined without any use of diagrams, it might be hoped that one could give an obviously three-dimensional construction of various invariants that does not require a choice of a two-dimensional projection. As we discuss later in these notes, Witten's physical interpretation of the Jones polynomial in [212] does exactly this.

For now, one can think of q as a formal variable. The subdiagrams shown in (1.3) should be thought of as depicting a neighborhood of one particular crossing in a planar diagram of an oriented knot; to apply the relation, one replaces the chosen crossing with each of the three shown partial diagrams, leaving the rest of the diagram unchanged.

To apply this linear relation, one also needs to fix a normalization, which can be done by specifying P_N for the unknot. Here, unfortunately, several natural choices exist. For now, we will choose

$$P_N(\bigcirc) = \frac{q^N - q^{-N}}{q - q^{-1}} = \underbrace{q^{-(N-1)} + q^{-(N-3)} + \cdots + q^{N-1}}_{N \text{ terms}}. \tag{1.4}$$

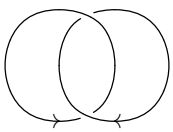
This choice gives the so-called *unnormalized* $\mathfrak{sl}(N)$ polynomial. Notice that, given any choice of $P_N(\bigcirc)$ with integer coefficients, the form of the skein relation implies that $P_N(q) \in \mathbb{Z}[q, q^{-1}]$ for every knot.

Notice further that, with the normalization (1.4), we have

$$P_N(\bigcirc) \xrightarrow[q \to 1]{} N, \tag{1.5}$$

which is the dimension of the representation R with which we decorated the knot, the fundamental of $\mathfrak{sl}(N)$. We remark that this leads to a natural generalization of the notion of dimension, the so-called *quantum dimension* $\dim_q(R)$ of a representation R, which arises from the quantum group invariant constructed from R evaluated on the unknot.

Equipped with the above rules, let us now try to compute $P_N(q)$ for some simple links. Consider the *Hopf link*, consisting of two interlocked circles:



Applying the skein relation to the upper of the two crossings, we obtain:

$$q^N P_N\Big[\underset{\text{Hopf link}}{\quad}\Big] - q^{-N} P_N\Big[\underset{\text{two unknots}}{\quad}\Big] = (q - q^{-1}) P_N\Big[\underset{\text{one unknot}}{\quad}\Big]. \tag{1.6}$$

This illustrates a general feature of the skein relation, which occurs for knots as well as links: in applying the relation to break down any knot diagram into simpler diagrams, one will generally need to evaluate P_N for links rather than

just for knots, since application of the relation (1.3) may produce links with more than one component. This means that the normalization (1.4) is not quite sufficient; we will need to specify P_N on k unlinked copies of the unknot, for $k \geq 1$.

As such, the last of our combinatorial rules for computing $P_N(q)$ concerns its behavior under disjoint union:

$$P_N(\bigcirc \sqcup K) = P_N(\bigcirc) \cdot P_N(K), \tag{1.7}$$

where K is any knot or link. Here, the disjoint union should be such that K and the additional unknot are not linked with one another.

Caution: The discerning reader will notice that our final rule (1.7) is not linear, while the others are, and so is not respected under rescaling of $P_N(q)$. Therefore, if a different choice of normalization is made, it will *not* remain true that $P_N(k \text{ unknots}) = [P_N(\bigcirc)]^k$. The nice behavior (1.7) is particular to our choice of normalization (1.4). This can be expressed by saying that, in making a different normalization, one must remember to normalize only one copy of the unknot.

To complete the calculation we began above, let's specialize to the case $N = 2$. Then we have

$$P_2(\bigcirc) = q^{-1} + q \qquad \Longrightarrow \; P_2(\text{Hopf link}) = (q^{-1} + q)^2 = q^{-2} + 2 + q^2. \tag{1.8}$$

Applying the skein relation (1.6) then gives

$$\begin{aligned} q^2 P_2(\text{Hopf link}) &= q^{-2}(q^{-2} + 2 + q^2) + (q - q^{-1})(q + q^{-1}) \\ &= q^{-4} + q^{-2} + 1 + q^2, \end{aligned} \tag{1.9}$$

so that

$$P_2(\text{Hopf link}) = q^{-6} + q^{-4} + q^{-2} + 1. \tag{1.10}$$

We are now ready to compute the $\mathfrak{sl}(N)$ invariant for any link.

From the form of the rules that define this invariant, it is apparent that dependence on the parameter N enters the knot polynomial only by way of the combination of variables q^N. As such, we can define the new variable $a \hat{=} q^N$, in terms of which our defining relations become

$$a P_{a,q}(\nearrow\!\!\searrow) - a^{-1} P_{a,q}(\nearrow\!\!\searrow) = (q - q^{-1}) P_{a,q}(\curvearrowleft\,\curvearrowright), \tag{1.11}$$

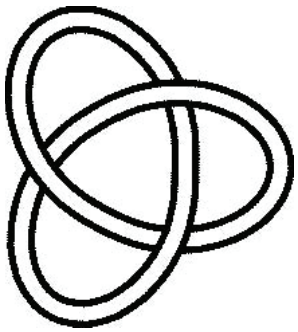


Figure 1.2: The trefoil knot 3_1. (Image from [177].)

$$P_{a,q}(\bigcirc) = \frac{a - a^{-1}}{q - q^{-1}}. \tag{1.12}$$

Together with the disjoint union property, these rules associate to each oriented link K a new invariant $P_{a,q}(K)$ in the variables a and q, called the (unnormalized) *HOMFLY-PT polynomial* of the link [88]. This is something of a misnomer, since with the normalization (1.12) the HOMFLY-PT invariant will in general be a rational expression rather than a polynomial. We have traded the two variables q, N for q and a.

For various special choices of the variables a and q, the HOMFLY-PT polynomial reduces to other familiar polynomial knot invariants:

- $a = q^N$, of course, returns the quantum $\mathfrak{sl}(N)$ invariant $P_N(q)$.
- With the particular choice $a = q^2$ ($N = 2$), the HOMFLY-PT polynomial becomes the classical *Jones polynomial* $J(L; q) \equiv P_2(q)$,

  $$J(K; q) = P_{a=q^2,q}(K). \tag{1.13}$$

  Discovered in 1984 [123], the Jones polynomial is one of the best-known polynomial knot invariants, and can be regarded as the "father" of quantum group invariants; it is associated to the Lie algebra $\mathfrak{sl}(2)$ and its fundamental two-dimensional representation.

- $a = 1$ returns the *Alexander polynomial* $\Delta(K; q)$, another classical knot invariant. This shows that the HOMFLY-PT polynomial generalizes the $\mathfrak{sl}(N)$ invariant, in some way: the evaluation $a = 1$ makes sense, even though taking $N = 0$ is somewhat obscure from the standpoint of representation theory.

Now, the attentive reader will point out a problem: if we try and compute the Alexander polynomial, we immediately run into the problem that (1.12) requires $P_{1,q}(\bigcirc) = 0$. The invariant thus appears to be zero for every link! However, this does not mean that the Alexander polynomial is trivial. Remember that, since the skein relations are linear, we have the freedom to rescale invariants by any multiplicative constant. We have simply made a choice that corresponds, for the particular value $a = 1$, to multiplying everything by zero.

This motivates the introduction of another convention: the so-called *normalized* HOMFLY-PT polynomial is defined by performing a rescaling such that

$$P_{a,q}(\bigcirc) = 1. \tag{1.14}$$

This choice is natural on topological grounds, since it associates 1 to the unknot independent of how the additional input data, or "decoration," is chosen. (By contrast, the *unnormalized* HOMFLY-PT polynomial assigns the value 1 to the empty knot diagram.) Taking $a = 1$ in the *normalized* HOMFLY-PT polynomial returns a nontrivial invariant, the Alexander polynomial.

Exercise 1. Compute the normalized and unnormalized HOMFLY-PT polynomials for the trefoil knot $K = 3_1$ (Fig. 1.2). Note that one of these will actually turn out to be polynomial!

Having done this, specialize to the case $a = q^2$ to obtain the normalized and unnormalized Jones polynomials for the trefoil. Then specialize to the case $a = q$. Something nice should occur! Identify what happens and explain why this is the case.

Solution. Applying the skein relation for the HOMFLY-PT polynomial to one crossing of the trefoil knot gives

$$aP_{a,q}(3_1) - a^{-1}P_{a,q}(\bigcirc) = (q - q^{-1})P_{a,q}(\ \bigcirc\!\!\bigcirc\ ).$$

Then, applying the relation again to the Hopf link (as in the above example) gives

$$aP_{a,q}(\ \bigcirc\!\!\bigcirc\ ) - a^{-1}P_{a,q}(\ \bigcirc\!\!\bigcirc\ ) = (q - q^{-1})P_{a,q}(\bigcirc).$$

Therefore, for the unnormalized HOMFLY-PT polynomial,

$$P(3_1) = a^{-2}P(\bigcirc) + a^{-2}(q - q^{-1})\left[a^{-1}P(\bigcirc)^2 + (q - q^{-1})P(\bigcirc)\right].$$

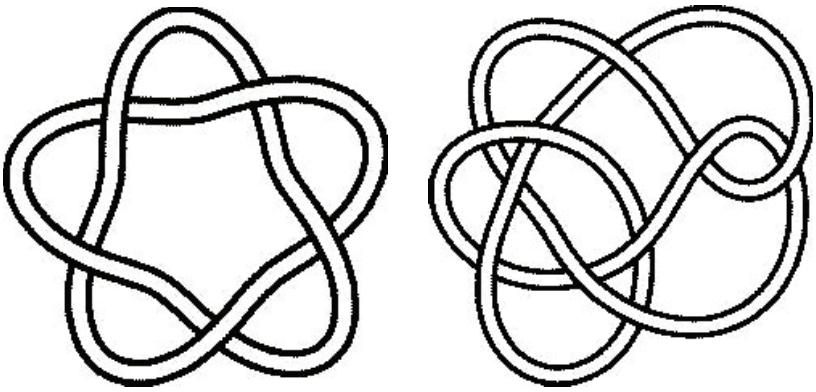


Figure 1.3: The knots 5_1 and 10_{132}. (Images from [177].)

which becomes

$$P(3_1) = \frac{a - a^{-1}}{q - q^{-1}} \left[a^{-2}q^2 + a^{-2}q^{-2} - a^{-4}\right].$$

The normalized HOMFLY-PT polynomial is simply the quantity in brackets. Specializing to $a = q^2$ gives the unnormalized Jones polynomial:

$$P_2(3_1) = \frac{q^2 - q^{-2}}{q - q^{-1}} \left[q^{-2} + q^{-6} - q^{-8}\right]. \tag{1.15}$$

Again, the normalized Jones polynomial is the factor in square brackets. Finally, we specialize to $a = q$, obtaining $P = 1$ in both the normalized and unnormalized cases! This is connected to the fact that $a = q$ corresponds to constructing the $\mathfrak{sl}(1)$ invariant, which must be vacuous since the Lie algebra is trivial. □

Remark 1. The study of this subject is made more difficult by the preponderance of various conventions in the literature. In particular, there is no agreement at all about standard usage with regard to the variables for polynomial invariants. Given ample forewarning, this should not cause too much confusion, but the reader must always be aware of the problem. In particular, it is extremely common for papers to differ from our conventions by the replacement

$$a \mapsto a^{1/2},\ q \mapsto q^{1/2}, \tag{1.16}$$

halving all powers that occur in knot polynomials. Some authors also make the change

$$a \mapsto a^{-1},\ q \mapsto q^{-1}, \tag{1.17}$$

and some make both.

We have by now seen a rich supply of knot polynomials, which can be straightforwardly computed by hand for simple enough diagrams, and are easy to write down and compare. One might then ask about the value of attempting to categorify at all. Given such simple and powerful invariants, why would one bother trying to replace them with much more complicated ones?

The simple answer is that the HOMFLY-PT polynomial and its relatives, while powerful, are not fully adequate for the job of classifying all knots up to ambient isotopy. Consider the two knot diagrams shown in Fig. 1.3, which represent the knots 5_1 and 10_{132} in the Rolfsen classification. While the knots are not equivalent, they have identical Alexander and Jones polynomials! In fact, we have

$$P_{a,q}(5_1) = P_{a,q}(10_{132})\,. \tag{1.18}$$

and, therefore, all specializations—including all $\mathfrak{sl}(N)$ invariants—will be identical for these two knots. Thus, even the HOMFLY-PT polynomial is not a perfect invariant and fails to distinguish between these two knots. This motivates us to search for a finer invariant. Categorification, as we shall see, provides one. Specifically, even though the Jones, Alexander, and HOMFLY-PT polynomials fail to distinguish the knots 5_1 and 10_{132} of our example, their respective categorifications do (cf. Fig. 1.9).

Before we step into the categorification era, let us make one more desperate attempt to gain power through polynomial knot invariants. To this end, let us introduce not one, but a whole sequence of knot polynomials $J_n(K;q) \in \mathbb{Z}[q,q^{-1}]$ called the *colored Jones polynomials*. For each non-negative integer n, the n-colored Jones polynomial of a knot K is the quantum group invariant associated to the decoration $\mathfrak{g} = \mathfrak{sl}(2)$ with its n-dimensional representation V_n. $J_2(K;q)$ is just the ordinary Jones polynomial. In *Chern-Simons theory* with gauge group $G = \mathrm{SU}(2)$, we can think of $J_n(K;q)$ as the expectation value of a Wilson loop operator on K, colored by the n-dimensional representation of $\mathrm{SU}(2)$ [212].

Moreover, the colored Jones polynomial obeys the following relations, known as *cabling formulas*, which follow directly from the rules of Chern-Simons

TQFT:

$$\begin{aligned} J_{\bigoplus_i R_i}(K;q) &= \sum_i J_{R_i}(K;q), \\ J_R(K^n;q) &= J_{R^{\otimes n}}(K;q). \end{aligned} \tag{1.19}$$

Here K^n is the n-cabling of the knot K, obtained by taking the path of K and tracing it with a "cable" of n strands. These equations allow us to compute the n-colored Jones polynomial, given a way to compute the ordinary Jones polynomial and a little knowledge of representation theory. For instance, any knot K has $J_1(K;q) = 1$ and $J_2(K;q) = J(K;q)$, the ordinary Jones polynomial. Furthermore,

$$\begin{aligned} \mathbf{2}\otimes\mathbf{2} = \mathbf{1}\oplus\mathbf{3} &\implies J_3(K;q) = J(K^2;q) - 1, \\ \mathbf{2}\otimes\mathbf{2}\otimes\mathbf{2} = \mathbf{2}\oplus\mathbf{2}\oplus\mathbf{4} &\implies J_4(K;q) = J(K^3;q) - 2J(K;q), \end{aligned} \tag{1.20}$$

and so forth. We can switch to representations of lower dimension at the cost of considering more complicated links; however, the computability of the ordinary Jones polynomial means that this is still a good strategy for calculating colored Jones polynomials.

Example 2. Using the above formulae, it is easy to find n-colored Jones polynomial of the trefoil knot $K = 3_1$ for the first few values of n:

$$\begin{aligned} J_1 &= 1, \\ J_2 &= q + q^3 - q^4, \\ J_3 &= q^2 + q^5 - q^7 + q^8 - q^9 - q^{10} + q^{11}, \\ &\vdots \end{aligned} \tag{1.21}$$

where, for balance (and to keep the reader alert), we used the conventions which differ from (1.15) by the transformations (1.16) and (1.17).

Much like the ordinary Jones polynomial is a particular specialization (1.13) of the HOMFLY-PT polynomial, its colored version $J_n(K;q)$ can be obtained by the same specialization from the so-called *colored HOMFLY-PT polynomial* $P_n(K;a,q)$,

$$J_n(K;q) = P_n(K;a=q^2,q). \tag{1.22}$$

labeled by an integer n. More generally, the colored HOMFLY-PT polynomials $P^\lambda(K;a,q)$ are labeled by Young diagrams or 2d partitions λ. In these lectures, we shall consider only Young diagrams that consist of a single row (or a single

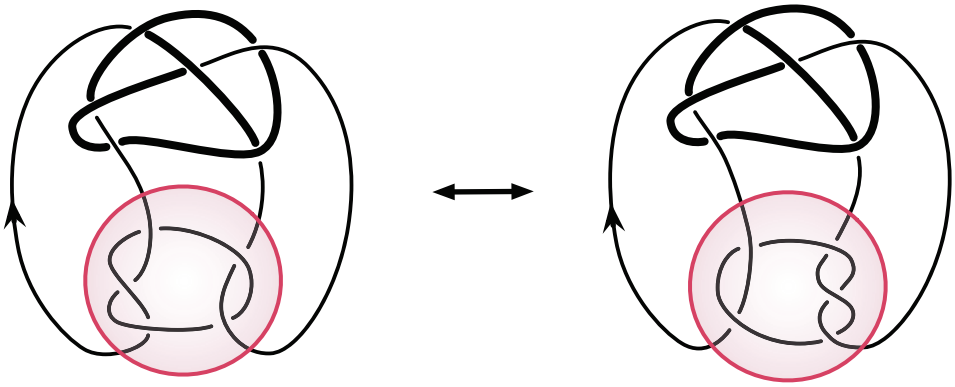


Figure 1.4: Mutant knots. (Images from [64].)

column) and by Schur-Weyl duality correspond to totally symmetric (resp. totally anti-symmetric) representations. Thus, what we call $P_n(K;a,q)$ is the HOMFLY-PT polynomial of K colored by $\lambda = S^{n-1}$.

Even though $P_n(K;a,q)$ provide us with an infinite sequence of two-variable polynomial knot invariants, which can tell apart e.g. the two knots in (1.18), they are still not powerful enough to distinguish simple pairs of knots and links called *mutants*. The operation of *mutation* involves drawing a disc on a knot diagram such that two incoming and two outgoing strands pass its boundary, and then rotating the portion of the knot inside the disc by 180 degrees. The Kinoshita-Terasaka and Conway knots shown in Fig. 1.4 are a famous pair of knots that are mutants of one another, but are nonetheless distinct; they can be distinguished by homological knot invariants, but not by any of the polynomial invariants we have discussed so far!

Theorem 1. The colored Jones polynomial, the colored HOMFLY-PT polynomial, and the Alexander polynomial cannot distinguish mutant knots [156], while their respective categorifications can [106, 170, 207].

1.2 The classical A-polynomial

In this section, we take a step back from quantum group invariants to discuss another classical invariant of knots: the so-called *A-polynomial*. Our introduction will be rather brief, intended to familiarize the reader with the general idea behind this invariant and catalogue some of its properties, rather than attempt a complete construction. For more information, we refer to the pioneering paper of Cooper et. al. [54], in which the A-polynomial was first

defined.

For a knot K, let $N(K) \subset S^3$ be an open tubular neighborhood of K. Then the *knot complement* is defined to be

$$M \doteq S^3 \setminus N(K). \tag{1.23}$$

By construction, M is a 3-manifold with torus boundary, and our goal here is to explain that to every such manifold one can associate a planar algebraic curve

$$\mathscr{C} = \{(x, y) \in \mathbb{C}^2 : A(x, y) = 0\}, \tag{1.24}$$

defined as follows. The classical invariant of M is its fundamental group, $\pi_1(M)$, which in the case of knot complements is called the *knot group*. It contains a lot of useful information about M and can distinguish knots much better than any of the polynomial invariants we saw in §1.1.

Example 3. Consider the trefoil knot $K = 3_1$. Its knot group is the simplest example of a braid group:

$$\pi_1(M) = \langle a, b : aba = bab \rangle. \tag{1.25}$$

Although the knot group is a very good invariant, it is not easy to deal with due to its non-abelian nature. To make life easier, while hopefully not giving up too much power, one can imagine considering representations of the knot group rather than the group itself. Thus, one can consider representations of $\pi_1(M)$ into a simple non-abelian group, such as the group of 2×2 complex matrices,

$$\rho : \pi_1(M) \to \mathrm{SL}(2, \mathbb{C}). \tag{1.26}$$

Associated to this construction is a polynomial invariant $A(x, y)$, whose zero locus (1.24) parameterizes in some sense the "space" of all such representations. Indeed, as we noted earlier, M is a 3-manifold with torus boundary,

$$\partial M = \partial N(K) \cong T^2. \tag{1.27}$$

Therefore, the fundamental group of ∂M is

$$\pi_1(\partial M) = \pi_1(T^2) = \mathbb{Z} \times \mathbb{Z}. \tag{1.28}$$

The generators of $\pi_1(\partial M)$ are the two basic cycles, which we will denote by m and ℓ (standing for *meridian* and *longitude*, respectively—see Fig. 1.5). m

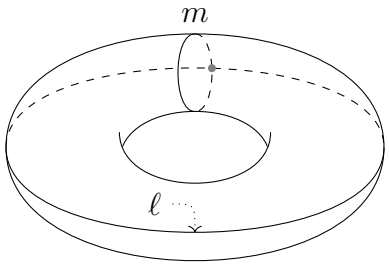


Figure 1.5: The torus $T^2 = \partial N(K)$ for $K =$ unknot, with cycles m and ℓ.

is the cycle that is contractible when considered as a loop in $N(K)$, and ℓ is the non-contractible cycle that follows the knot in $N(K)$. Of course, any representation $\pi_1(M) \to \mathrm{SL}(2,\mathbb{C})$ restricts to a representation of $\pi_1(T^2 = \partial M)$; this gives a natural map of representations of $\pi_1(M)$ into the space of representations of $\pi_1(\partial M)$.

These cycles are represented in $\mathrm{SL}(2,\mathbb{C})$ by 2×2 complex matrices $\rho(m)$ and $\rho(\ell)$ with determinant 1. Since the fundamental group of the torus is just $\mathbb{Z} \times \mathbb{Z}$, the matrices $\rho(m)$ and $\rho(\ell)$ commute, and can therefore be simultaneously brought to Jordan normal form by some change of basis, i.e., conjugacy by an element of $\mathrm{SL}(2,\mathbb{C})$:

$$\rho(m) = \begin{pmatrix} x & \star \\ 0 & x^{-1} \end{pmatrix}, \quad \rho(\ell) = \begin{pmatrix} y & \star \\ 0 & y^{-1} \end{pmatrix}. \tag{1.29}$$

Therefore, we have a map that assigns two complex numbers to each representation of the knot group:

$$\begin{aligned} \mathrm{Hom}(\pi_1(M), \mathrm{SL}(2,\mathbb{C}))/\mathrm{conj.} &\to \mathbb{C}^\star \times \mathbb{C}^\star, \\ \rho &\mapsto (x, y), \end{aligned} \tag{1.30}$$

where x and y are the eigenvalues of $\rho(m)$ and $\rho(\ell)$, respectively. The image of this map is the *representation variety* $\mathscr{C} \subset \mathbb{C}^\star \times \mathbb{C}^\star$, whose defining polynomial is the A-polynomial of K. Note, this definition of the A-polynomial does not fix the overall numerical coefficient, which is usually chosen in such a way that $A(x, y)$ has integer coefficients (we return to this property below). For the same reason, the A-polynomial is only defined up to multiplication by arbitrary powers of x and y. Let us illustrate the idea of this construction with some specific examples.

Example 4. Let $K \subset S^3$ be the unknot. Then $N(K)$ and M are both homeomorphic to the solid torus $S^1 \times D^2$. Notice that m is contractible as a loop in $N(K)$ and ℓ is not, while the opposite is true in M: ℓ is contractible and m is not. Since ℓ is contractible in M, $\rho(\ell)$ must be the identity, and therefore we have $y = 1$ for all $(x, y) \in \mathscr{C}$. There is no restriction on x, so that

$$\mathscr{C}(\bigcirc) = \{(x, y) \in \mathbb{C}^\star \times \mathbb{C}^\star : y = 1\}; \tag{1.31}$$

the A-polynomial of the unknot is therefore

$$A(x, y) = y - 1. \tag{1.32}$$

Example 5. Let $K \subset S^3$ be the trefoil knot 3_1. Then, as mentioned in (1.25), the knot group is given by

$$\pi_1(M) = \langle a, b : aba = bab \rangle, \tag{1.33}$$

where the meridian and longitude cycles can be identified as follows:

$$\begin{cases} m = a, \\ \ell = ba^2ba^{-4}. \end{cases} \tag{1.34}$$

Let us see what information we can get about the A-polynomial just by considering abelian representations of $\pi_1(M)$, i.e. representations such that $\rho(a)$ and $\rho(b)$ commute. For such representations, the defining relations reduce to $a^2b = ab^2$ and therefore imply $a = b$. (Here, in a slight abuse of notation, we are simply writing a to refer to $\rho(a)$ and so forth.) Eq. (1.34) then implies that $\ell = 1$ and $m = a$, so that $y = 1$ and x is unrestricted exactly as in Example 4. It follows that the A-polynomial contains $(y - 1)$ as a factor.

This example illustrates a more general phenomenon. Whenever M is a knot complement in S^3, it is true that the abelianization

$$\pi_1(M)^{\mathrm{ab}} = H_1(M) \cong \mathbb{Z}. \tag{1.35}$$

Therefore, the A-polynomial always contains $y - 1$ as a factor,

$$A(x, y) = (y - 1)(\cdots), \tag{1.36}$$

where the first piece carries information about abelian representations, and any additional factors that occur arise from the non-abelian representations.

In the particular case $K = 3_1$, a similar analysis of non-abelian representations of (1.25) into $\mathrm{SL}(2, \mathbb{C})$ yields

$$A(x, y) = (y - 1)(y + x^6). \tag{1.37}$$

To summarize, the algebraic curve $\mathscr{C}$ is (the closure of) the image of the representation variety of M in the representation variety $\mathbb{C}^\star \times \mathbb{C}^\star$ of its boundary torus ∂M. This image is always an affine algebraic variety of complex dimension 1, whose defining equation is precisely the A-polynomial [54].

This construction defines the A-polynomial as an invariant associated to any knot. However, extension to links requires extra care, since in that case $\partial N(L) \ncong T^2$. Rather, the boundary of the link complement consists of several components, each of which is separately homeomorphic to a torus. Therefore, there will be more than two fundamental cycles to consider, and the analogous construction will generally produce a higher-dimensional character variety rather than a plane algebraic curve. One important consequence of this is that the A-polynomial cannot be computed by any known set of skein relations; as was made clear in Exercise 1, computations with skein relations require one to consider general links rather than just knots.

To conclude this brief introduction to the A-polynomial, we will list without proof several of its interesting properties:

- For any hyperbolic knot K,

$$A_K(x, y) \neq y - 1. \tag{1.38}$$

That is, the A-polynomial carries nontrivial information about non-abelian representations of the knot group.

- Whenever K is a knot in a homology sphere, $A_K(x, y)$ contains only even powers of the variable x. Since in these lectures we shall only consider examples of this kind, we simplify expressions a bit by replacing x^2 with x. For instance, in these conventions the A-polynomial (1.37) of the trefoil knot looks like

$$A(x, y) = (y - 1)(y + x^3). \tag{1.39}$$

- The A-polynomial is *reciprocal*: that is,

$$A(x, y) \sim A(x^{-1}, y^{-1}), \tag{1.40}$$

where the equivalence is up to multiplication by powers of x and y. Such multiplications are irrelevant, because they don't change the zero locus of the A-polynomial in $\mathbb{C}^\star \times \mathbb{C}^\star$. This property can be also expressed by saying that the curve $\mathscr{C}$ lies in $(\mathbb{C}^\star \times \mathbb{C}^\star)/\mathbb{Z}_2$, where $\mathbb{Z}_2$ acts by $(x, y) \mapsto (x^{-1}, y^{-1})$ and can be interpreted as the Weyl group of $\mathrm{SL}(2, \mathbb{C})$.

- $A(x, y)$ is invariant under orientation reversal of the knot, but *not* under reversal of orientation in the ambient space. Therefore, it can distinguish mirror knots (knots related by the parity operation), such as the left- and right-handed versions of the trefoil. To be precise, if K' is the mirror of K, then

$$A_K(x, y) = 0 \iff A_{K'}(x^{-1}, y) = 0. \tag{1.41}$$

- After multiplication by a constant, the A-polynomial can always be taken to have integer coefficients. It is then natural to ask: are these integers counting something, and if so, what? The integrality of the coefficients of $A(x, y)$ is a first hint of the deep connections with number theory. For instance, the following two properties, based on the *Newton polygon* of $A(x, y)$, illustrate this connection further.

- The A-polynomial is *tempered*: that is, the faces of the Newton polygon of $A(x, y)$ define cyclotomic polynomials in one variable. Examine, for example, the A-polynomial of the figure-8 knot:

$$A(x, y) = (y - 1)\left(y^2 - (x^{-2} - x^{-1} - 2 - x + x^2)y + y^2\right). \tag{1.42}$$

- Furthermore, the slopes of the sides of the Newton polygon of $A(x, y)$ are boundary slopes of incompressible surfaces[†] in M.

While all of the above properties are interesting, and deserve to be explored much more fully, our next goal is to review the connection to physics [107], which explains known facts about the A-polynomial and leads to many new ones:

[†]A proper embedding of a connected orientable surface $F \to M$ is called incompressible if the induced map $\pi_1(F) \to \pi_1(M)$ is injective. Its boundary slope is defined as follows. An incompressible surface $(F, \partial F)$ gives rise to a collection of parallel simple closed loops in ∂M. Choose one such loop and write its homology class as $\ell^p m^q$. Then, the boundary slope of $(F, \partial F)$ is defined to be the rational number p/q.

- The A-polynomial curve (1.24), though constructed as an algebraic curve, is most properly viewed as an object of symplectic geometry: specifically, a holomorphic Lagrangian submanifold.

- Its quantization with the symplectic form

$$\omega = \frac{dy}{y} \wedge \frac{dx}{x} \tag{1.43}$$

leads to interesting wavefunctions.

- The curve $\mathscr{C}$ has all the necessary attributes to be an analogue of the Seiberg-Witten curve for knots and 3-manifolds [67, 89].

As an appetizer and a simple example of what the physical interpretation of the A-polynomial has to offer, here we describe a curious property of the A-polynomial curve (1.24) that follows from this physical interpretation. For any closed cycle in the algebraic curve $\mathscr{C}$, the integral of the Liouville one-form (see (1.49) below) associated to the symplectic form (1.43) should be quantized [107]. Schematically,[‡]

$$\oint_\Gamma \log y \frac{dx}{x} \in 2\pi^2 \cdot \mathbb{Q}. \tag{1.44}$$

This condition has an elegant interpretation in terms of algebraic K-theory and the Bloch group of $\bar{\mathbb{Q}}$. Moreover, it was conjectured in [114] that every curve of the form (1.24) — not necessarily describing the moduli of flat connections — is quantizable if and only if $\{x, y\} \in K_2(\mathbb{C}(\mathscr{C}))$ is a torsion class. This generalization will be useful to us later, when we consider a refinement of the A-polynomial that has to do with categorification and homological knot invariants.

To see how stringent the condition (1.44) is, let us compare, for instance, the A-polynomial of the figure-eight knot (1.42):

$$A(x, y) = 1 - (x^{-4} - x^{-2} - 2 - x^2 + x^4)y + y^2 \tag{1.45}$$

[‡]To be more precise, all periods of the "real" and "imaginary" part of the Liouville one-form θ must obey

$$\oint_\Gamma \Big(\log|x| d(\arg y) - \log|y| d(\arg x) \Big) \;=\; 0\,,$$
$$\frac{1}{4\pi^2} \oint_\Gamma \Big(\log|x| d\log|y| + (\arg y) d(\arg x) \Big) \;\in\; \mathbb{Q}\,.$$

and a similar polynomial

$$B(x,y) = 1 - (x^{-6} - x^{-2} - 2 - x^2 + x^6)y + y^2 \,. \tag{1.46}$$

(Here the irreducible factor $(y-1)$, corresponding to abelian representations, has been suppressed in both cases.) The second polynomial has all of the required symmetries of the A-polynomial, and is obtained from the A-polynomial of the figure-eight knot by a hardly noticeable modification. But $B(x,y)$ cannot occur as the A-polynomial of any knot since it violates the condition (1.44).

1.3 Quantization

Our next goal is to explain, following [107], how the physical interpretation of the A-polynomial in Chern-Simons theory can be used to provide a bridge between *quantum group invariants* of knots and *algebraic curves* that we discussed in §§1.1 and 1.2, respectively. In particular, we shall see how quantization of Chern-Simons theory naturally leads to a *quantization* of the classical curve (1.24),

$$A(x,y) \quad \rightsquigarrow \quad \hat{A}(\hat{x},\hat{y};q)\,, \tag{1.47}$$

i.e. a q-difference operator $\hat{A}(\hat{x},\hat{y};q)$ with many interesting properties. While this will require a crash course on basic tools of Quantum Mechanics, the payoff will be enormous and will lead to many generalizations and ramifications of the intriguing relations between quantum group invariants of knots, on the one hand, and algebraic curves, on the other. Thus, one such generalization will be the subject of §1.4, where we will discuss categorification and formulate a similar bridge between algebraic curves and *knot homologies*, finally explaining the title of these lecture notes.

We begin our discussion of the quantization problem with a lightning review of some mathematical aspects of classical mechanics. Part of our exposition here follows the earlier lecture notes [66] that we recommend as a complementary introduction to the subject. When it comes to Chern-Simons theory, besides the seminal paper [212], mathematically oriented readers may also want to consult the excellent books [9, 135].

As we discussed briefly in the introduction, the description of a system in classical mechanics is most naturally formulated in the language of symplectic geometry. In the classical world, the state of a system at a particular instant

in time is completely specified by giving $2N$ pieces of data: the values of the coordinates x_i and their conjugate momenta p_i, where $1 \leq i \leq N$. The $2N$-dimensional space parameterized by the x_i and p_i is the *phase space* $\mathscr{M}$ of the system. (For many typical systems, the space of possible configurations of the system is some manifold X, on which the x_i are coordinates, and the phase space is the cotangent bundle $\mathscr{M} = T^*X$.) Notice that, independent of the number N of generalized coordinates needed to specify the configuration of a system, the associated phase space is always of even dimension. In fact, phase space is always naturally equipped with the structure of a symplectic manifold, with a canonical symplectic form given by

$$\omega = dp \wedge dx. \tag{1.48}$$

(When the phase space is a cotangent bundle, (1.48) is just the canonical symplectic structure on any cotangent bundle, expressed in coordinates.) Recall that a symplectic form on a manifold is a closed, nondegenerate two-form, and that nondegeneracy immediately implies that any symplectic manifold must be of even dimension.

Since ω is closed, it locally admits a primitive form, the so-called *Liouville one-form*

$$\theta = p\, dx. \tag{1.49}$$

It should be apparent that $\omega = d\theta$, so that θ is indeed a primitive.

Let us now explore these ideas more concretely in the context of a simple example. As a model system, consider the one-dimensional simple harmonic oscillator. The configuration space of this system is just $\mathbb{R}$ (with coordinate x), and the Hamiltonian is given by

$$H = \frac{1}{2}p^2 + \frac{1}{2}x^2. \tag{1.50}$$

Since $dH/dt = 0$, the energy is a conserved quantity, and since $N = 1$, this one conserved quantity serves to completely specify the classical trajectories of the system. They are curves in phase space of the form

$$\mathscr{C} : \frac{1}{2}(x^2 + p^2) - \mathscr{E} = 0, \tag{1.51}$$

for $\mathscr{E} \in \mathbb{R}_+$; these are concentric circles about the origin, with radius determined by the energy. Fig. 1.6 shows the potential of this system, together

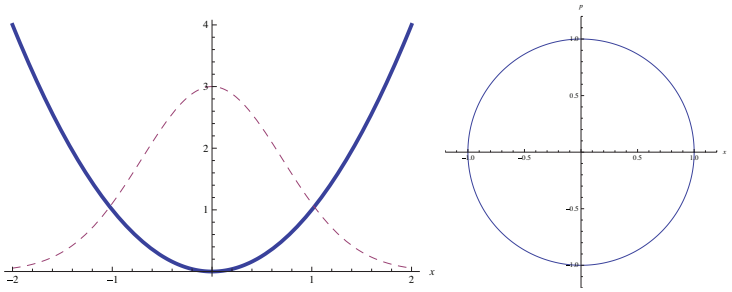


Figure 1.6: On the left, the potential and lowest-energy wavefunction for the simple harmonic oscillator. On the right, the phase space of this system, with a typical classical trajectory.

with a typical trajectory in the phase space. The dashed line represents the lowest-energy wavefunction of the system, to which we will come in a moment.

Now, recall that a *Lagrangian submanifold* $\mathscr{C} \subset (\mathscr{M}, \omega)$ is a submanifold such that $\omega|_{\mathscr{C}} = 0$, having the maximal possible dimension, i.e., $\dim \mathscr{C} = \frac{1}{2} \dim \mathscr{M}$. (If $\mathscr{C}$ has dimension larger than half the dimension of $\mathscr{M}$, the symplectic form cannot be identically zero when restricted to $\mathscr{C}$, since it is nondegenerate on $\mathscr{M}$.) It should be clear that, in the above example, the classical trajectories (1.51) are Lagrangian submanifolds of the phase space.

Moreover, since in this example the degree of the symplectic form ω is equal to the dimension of the phase space, ω is a volume form—in fact, the standard volume form on $\mathbb{R}^2$. We can therefore compute the area encompassed by a trajectory of energy $\mathscr{E}$ by integrating ω over the region $x^2 + p^2 < 2\mathscr{E}$, obtaining

$$2\pi\mathscr{E} = \int_D dp \wedge dx, \tag{1.52}$$

where D is the disc enclosed by the trajectory $\mathscr{C}$. Therefore, classically, the energy of a trajectory is proportional to the area in phase space it encompasses.

How do these considerations relate to quantization of the system? It is well known that the energy levels of the simple harmonic oscillator are given by

$$\mathscr{E} = \frac{1}{2\pi} \int_D dp \wedge dx = \hbar \left(n + \frac{1}{2} \right) \tag{1.53}$$

when the system is quantized. Thus, we expect that, in quantizing a system, the number of quantum states contained in some region of phase space will be directly proportional to its area. Moreover, we interpret $\hbar$, which has the same units as area in phase space, as the amount of classical phase space per

quantum state. Schematically,

$$\#\text{ states} \sim \text{area}/\hbar. \tag{1.54}$$

This relation has a long history in quantum physics; it is none other than the Bohr-Sommerfeld quantization condition.

Moreover, since ω admits a primitive, we can use the Stokes theorem to write

$$\mathscr{E} = \frac{1}{2\pi}\int_D \omega = \frac{1}{2\pi}\oint_{\mathscr{C}} \theta, \tag{1.55}$$

since $\mathscr{C} = \partial D$ and $d\theta = \omega$.

We have discussed counting quantum states; what about actually constructing them? In quantum mechanics, we expect the state to be a vector in a Hilbert space, which can be represented as a square-integrable wavefunction $Z(x)$. It turns out that, in the limit where $\hbar$ is small, the wavefunction can be constructed to lowest order in a manner that bears a striking resemblance to (1.55):

$$\begin{aligned} Z(x) &\xrightarrow[\hbar\to 0]{} \exp\left[\frac{i}{\hbar}\int_0^x \theta + \cdots\right] \\ &= \exp\left[\frac{i}{\hbar}\int_0^x \sqrt{2\mathscr{E} - x^2}\, dx + \cdots\right] \end{aligned} \tag{1.56}$$

Evaluating the wavefunction in this manner for the lowest-energy state of our system ($\mathscr{E} = \hbar/2$) yields

$$Z(x) \approx \exp\left[-\frac{1}{2\hbar}x^2 + \cdots\right]. \tag{1.57}$$

Indeed, $\exp(-x^2/2\hbar)$ is the exact expression for the $n = 0$ wavefunction.

We are slowly making progress towards understanding the quantization of our model system. The next step is to understand the transition between the classical and quantum notions of an *observable*. In the classical world, the observables x and p are coordinates in phase space—in other words, functions *on* the phase space:

$$x : \mathscr{M} \to \mathbb{R},\ (x,p) \mapsto x, \tag{1.58}$$

and so forth. General observables are functions of x and p, i.e., general elements of $C^\infty(\mathscr{M}, \mathbb{R})$.

In the quantum world, as is well known, x and p should be replaced by operators $\hat{x}$ and $\hat{p}$, obeying the canonical commutation relation

$$[\hat{p}, \hat{x}] = -i\hbar. \tag{1.59}$$

These operators now live in some noncommutative algebra, which is equipped with an action on the Hilbert space of states. In the position representation, for instance,

$$\hat{x}f(x) = xf(x), \quad \hat{p}f(x) = -i\hbar\frac{d}{dx}f(x), \tag{1.60}$$

where $f \in L^2(\mathbb{R})$. The constraint equation (1.51) that defines a classical trajectory is then replaced by the operator equation

$$\left[\frac{1}{2}(\hat{x}^2 + \hat{p}^2) - \mathscr{E}\right] Z(x) = 0, \tag{1.61}$$

which is just the familiar Schrödinger eigenvalue equation $\hat{H}Z = \mathscr{E}Z$. Now, unlike in the classical case, the solutions of (1.61) in the position representation will only be square-integrable (and therefore physically acceptable) for certain values of $\mathscr{E}$. These are precisely the familiar eigenvalues or allowed energy levels

$$\mathscr{E} = \hbar\left(n + \frac{1}{2}\right), \tag{1.62}$$

where $n = 0, 1, 2, \ldots$. Taking the lowest energy level ($n = 0$) as an example, the exact solution is $Z(x) = \exp(-x^2/2\hbar)$, just as we claimed above. The reader can easily verify this directly.

All of this discussion should be taken as illustrating our above claim that quantum mechanics should properly be understood as a "modern symplectic geometry," in which classical constraints are promoted to operator relations. We have constructed the following correspondence or dictionary between the elements of the classical and quantum descriptions of a system:

	Classical	*Quantum*
state space	symplectic manifold $(\mathscr{M}, \omega)$	Hilbert space $\mathscr{H}$
states	Lagrangian submanifolds $\mathscr{C} \subset \mathscr{M}$	vectors (wave functions) $Z \in \mathscr{H}$
observables	algebra of functions $f \in C^\infty(M)$	algebra of operators $\hat{f}$, acting on $\mathscr{H}$
constraints	$f_i = 0$	$\hat{f}_i Z = 0$

We now have a benchmark for what a successful quantization should accomplish: for a given classical system, it should construct the quantum counterpart for each element in the classical description, as summarized above. Of course,

we would also like the correspondence principle to hold: that is, the quantum description should dovetail nicely with the classical one in some way when we take $\hbar \to 0$.

The correspondence between the classical and quantum descriptions is not quite as cut-and-dried as we have made it appear, and there are a few points that deserve further mention. Firstly, it should be apparent from our discussion of the harmonic oscillator that not every Lagrangian submanifold will have a quantum state associated to it; in particular, only a particular subset of these (obeying the Bohr-Sommerfeld quantization condition, or equivalently, corresponding to eigenvalues of the operator $\hat{H}$) will allow us to construct a square-integrable wavefunction $Z(x)$. There can be further constraints on *quantizable* Lagrangian submanifolds [116].

Secondly, let us briefly clarify why quantum state vectors correspond to Lagrangian submanifolds of the classical phase space and not to classical 1-dimensional trajectories, as one might naively think. (In our example of the harmonic oscillator we have $N = 1$ and, as a result, both Lagrangian submanifolds and classical trajectories are one-dimensional.) The basic reason why Lagrangian submanifolds, rather than dimension-1 trajectories, are the correct objects to consider in attempting a quantization is the following. In quantum mechanics, we specify a state by giving the results of measurements of observables performed on that state. For this kind of information to be meaningful, the state must be a simultaneous eigenstate of all observables whose values we specify, which is only possible if all such observables mutually commute. As such, to describe the state space in quantum mechanics, we choose a "complete set of commuting observables" that gives a decomposition of $\mathscr{H}$ into one-dimensional eigenspaces of these operators. For time-independent Hamiltonians, one of these operators will always be $\hat{H}$.

However, at leading order in $\hbar$, the commutator of two quantum observables must be proportional to the Poisson bracket of the corresponding classical observables. Therefore, if $\hat{H}, \hat{f}_i$ form a complete set of commuting quantum-mechanical observables, we must have

$$\{H, f_i\}_{\text{P.B.}} = 0, \tag{1.63}$$

where $\{\cdot\,,\cdot\}_{\text{P.B.}}$ is the Poisson bracket. But we know that the classical time-

evolution of the quantity f_i is determined by the equation

$$\frac{df_i}{dt} + \{H, f_i\}_{\text{P.B.}} = 0. \tag{1.64}$$

As such, the quantum-mechanical observables used in specifying the state must correspond to classically conserved quantities: *constants of the motion*. And it is well-known that the maximal possible number of classically conserved quantities is $N = \frac{1}{2}\dim \mathscr{M}$, corresponding to a completely integrable system; this follows from the nondegeneracy of the symplectic form on the classical phase space. For $N > 1$, then, specifying all of the constants of the motion does not completely pin down the classical trajectory; it specifies an N-dimensional submanifold $\mathscr{C} \subset \mathscr{M}$. However, it does give all the information it is possible for one to have about the quantum state. This is why Lagrangian submanifolds are the classical objects to which one attempts to associate quantum states.

We should also remark that it is still generically true that wavefunctions will be given to lowest order by

$$Z(x) = \exp\left[\frac{i}{\hbar}\int_{x_0}^{x} \theta + \cdots\right]. \tag{1.65}$$

This form fits all of the local requirements for $Z(x)$, although it may or may not produce a globally square-integrable wavefunction.

Finally, the quantum-mechanical algebra of operators is a non-commutative deformation or *q-deformation* of the algebra of functions $C^\infty(\mathscr{M})$, where the deformation is parameterized by

$$q \mathrel{\hat{=}} e^{\hbar}. \tag{1.66}$$

In the classical limit, $q \to 1$.

How are these general ideas about quantization implemented in the context of topological quantum field theories? To illustrate the connection, we will consider a specific example of a TQFT: the Chern-Simons gauge theory.

As in any gauge theory, the starting point of this theory is the choice of a gauge group G and the action functional, which in the present case is the Chern-Simons functional:

$$\frac{1}{\hbar}\int_M \mathrm{Tr}(A \wedge dA + \frac{2}{3} A \wedge A \wedge A). \tag{1.67}$$

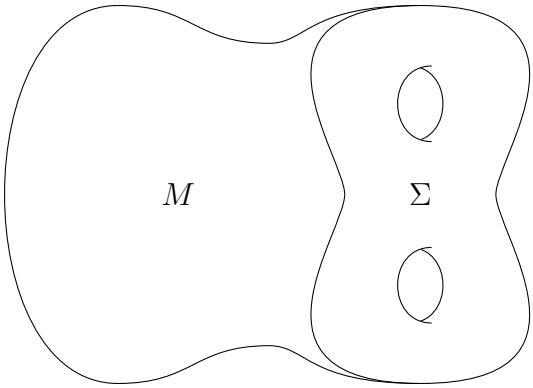


Figure 1.7: The setup for Chern-Simons theory: an oriented 3-manifold M with boundary a 2-manifold Σ.

Here M is a 3-manifold, and the gauge field A is a connection on a principal G-bundle $E \to M$. The action functional (1.67) can be interpreted roughly as a Morse function on the space of gauge fields. We search for critical points of this functional by solving the equation of motion, which is the PDE

$$dA + A \wedge A = 0. \tag{1.68}$$

This equation says that A is a flat connection. How is this gauge theory formulation related to the picture of a TQFT as a functor, in the axiomatic language of Atiyah and Segal [9]?

The answer to this question is summarized in the below table, and illustrates the way in which quantization plays a role. The action functional (1.67) defines a classical gauge theory; the classical phase space of this theory is the *moduli space of flat connections* $\mathscr{M} = \mathscr{M}_{\text{flat}}(G, \Sigma)$, where $\Sigma = \partial M$.

Now, let $\mathscr{M}_{\text{flat}}(G, M)$ be the moduli space of flat connections on M. There is a natural mapping

$$\mathscr{M}_{\text{flat}}(G, M) \hookrightarrow \mathscr{M}_{\text{flat}}(G, \Sigma) \tag{1.69}$$

induced by restriction to $\Sigma = \partial M$. The image of this map is the subspace of $\mathscr{M}$ consisting of flat connections on Σ that can be extended to M; this is a Lagrangian submanifold $\mathscr{C} \subset \mathscr{M}$.

We are now equipped with precisely the classical data referred to in our earlier discussion of the quantization problem. If we now quantize the classical Chern-Simons theory, the classical phase space $\mathscr{M}$ and the Lagrangian submanifold $\mathscr{C} \subset \mathscr{M}$ will be respectively replaced with a Hilbert space and a

state vector in that Hilbert space. But these are precisely the objects that we expect a TQFT functor to associate to Σ and M!

To sum up, our situation is as follows:

Geometry	*Classical CS*	*Quantum CS*
2-manifold Σ	symplectic manifold $\mathscr{M} = \mathscr{M}_{\text{flat}}(G, \Sigma)$	vector space $\mathscr{H}_\Sigma$
3-manifold M $(\partial M = \Sigma)$	Lagrangian submanifold: connections extendible to M	vector $Z(M) \in \mathscr{H}_\Sigma$

To move from the first column to the second, we define the classical Chern-Simons theory. Moving from the second column to the third consists of a quantization of this theory. The usual picture of a TQFT as a functor is the composition of these two: it moves directly from the first to the third column, ignoring the second.

Let us discuss the phase space of classical Chern-Simons theory a little further. It is known that all flat connections on Riemann surfaces are described by their holonomies; that is, the moduli space consists of maps

$$\mathscr{M} = \text{Hom}(\pi_1(\Sigma) \to G)/\text{conjugation}. \tag{1.70}$$

As emphasized in the work of Atiyah and Bott [6], this space comes equipped with a natural symplectic form,

$$\omega = \frac{1}{4\pi^2} \int_\Sigma \text{Tr}\, \delta A \wedge \delta A, \tag{1.71}$$

where δ denotes the exterior derivative on $\mathscr{M}$, so that δA is a 1-form on Σ as well as on $\mathscr{M}$. The Lagrangian submanifold we are considering is then given by

$$\mathscr{C} = \text{Hom}(\pi_1(M) \to G)/\text{conjugation}, \tag{1.72}$$

and the inclusion map is induced by the natural map $\pi_1(\Sigma) \to \pi_1(M)$. This Lagrangian submanifold can be defined by classical constraint equations of the form

$$A_i = 0. \tag{1.73}$$

Quantization will then replace these with quantum constraints; that is, operator relations

$$\hat{A}_i Z = 0, \tag{1.74}$$

just as the classical constraint (1.51) was replaced by the operator equation (1.61) in our previous example.

Exercise 2. Verify that $\mathscr{C}$ is indeed Lagrangian with respect to the symplectic form (1.71). That is, show that

$$\omega|_{\mathscr{C}\subset\mathscr{M}} = 0. \tag{1.75}$$

Exercise 3. Let g be the genus of Σ. Show that, for $g > 1$,

$$\dim \mathscr{M} = (2g-2)\dim G. \tag{1.76}$$

Solution. Consider the case where G is a simple group. The fundamental group $\pi_1(\Sigma)$ is generated by $2g$ elements A_i and B_i, $1 \le i \le g$, subject to the one relation

$$A_1 B_1 A_1^{-1} B_1^{-1} \cdots A_g B_g A_g^{-1} B_g^{-1} = 1. \tag{1.77}$$

After applying an element of $\mathrm{Hom}(\pi_1(\Sigma) \to G)$, the generators are mapped to group-valued matrices, and so we need $2g\cdot\dim G$ parameters to specify them all independently. However, there are constraints: the matrices must obey (1.77), one matrix equation which eliminates $\dim G$ degrees of freedom. Taking the quotient by conjugacy eliminates another $\dim G$ degrees of freedom, leaving

$$\dim \mathscr{M} = (2g-2)\dim G, \tag{1.78}$$

as we expected. □

Let us now specialize this general discussion and consider the theory with gauge group $G = \mathrm{SL}(2,\mathbb{C})$ on a 3-manifold that is a knot complement, $M = S^3 \setminus N(K)$. Then, of course, $\partial M = \Sigma \cong T^2$. It follows immediately that $\pi_1(\Sigma) = \mathbb{Z}\times\mathbb{Z}$, so that

$$\begin{aligned} \mathscr{M} &= \mathrm{Hom}(\mathbb{Z}\times\mathbb{Z} \to \mathrm{SL}(2,\mathbb{C}))/\text{conjugacy} \\ &= (\mathbb{C}^\star \times \mathbb{C}^\star)/\mathbb{Z}_2. \end{aligned} \tag{1.79}$$

This is exactly the space we considered in §1.2 in our discussion of the A-polynomial: it is the representation variety of the boundary torus of M! Moreover, the Lagrangian submanifold is in this case given by

$$\begin{aligned} \mathscr{C} &= \mathrm{Hom}(\pi_1(M) \to \mathrm{SL}(2,\mathbb{C}))/\text{conjugacy} \\ &= \{(x,y) \in (\mathbb{C}^\star \times \mathbb{C}^\star)/\mathbb{Z}_2 : A(x,y) = 0\}, \end{aligned} \tag{1.80}$$

where $A(x, y)$ is a familiar polynomial in x and y, interpreted now as a classical observable giving the classical constraint relation that defines the submanifold $\mathscr{C} \subset \mathscr{M}$.

The appearance of the A-polynomial in this context clarifies two mysterious statements that were made in the previous section. Firstly, it makes apparent in what sense the zero locus of the A-polynomial is a natural object in symplectic geometry. Secondly, we can now make sense of the statement that one can "quantize" the A-polynomial. Having interpreted it as a classical constraint equation defining a Lagrangian submanifold of a classical phase space, it becomes obvious that quantization replaces the A-polynomial by an operator in a quantum constraint equation of the form (1.74).

What happens when we try to quantize the A-polynomial? The natural symplectic form (1.71) on the classical phase space takes the simple form [107]:

$$\omega = \frac{dy}{y} \wedge \frac{dx}{x} = d\ln y \wedge d\ln x. \tag{1.81}$$

The canonical commutation relation is therefore

$$\left[\widehat{\ln y}, \widehat{\ln x}\right] = \hbar, \tag{1.82}$$

which can be rewritten in the form

$$\hat{y}\hat{x} = q\hat{x}\hat{y}. \tag{1.83}$$

with $q = e^{\hbar}$. Given this relation, what form do the operators $\hat{x}$ and $\hat{y}$ take in the position representation? Of course, we must have $\hat{x}f(x) = xf(x)$. Then the commutation relation becomes

$$q\hat{x}(\hat{y}f(x)) = \hat{y}(\hat{x}f(x)), \tag{1.84}$$

and implies that $\hat{y}$ should act as a *shift operator* $\hat{y}f(x) = f(qx)$. The reason for this name is the following. Notice, that the symplectic form (1.81) has the canonical form in logarithms of x and y, rather than x and y themselves. Therefore, it is natural to introduce the logarithmic variable n by the relation $x = q^n$. Then, in terms of n the action of the operators $\hat{x}$ and $\hat{y}$ looks like

$$\hat{x}f(n) = q^n f(n), \quad \hat{y}f(n) = f(n+1). \tag{1.85}$$

The quantization of the polynomial $A(x, y) = \sum_k a_k(x)y^k$ will then be an operator of the form

$$\hat{A}(\hat{x}, \hat{y}; q) = \sum_k a_k(\hat{x}; q)\hat{y}^k. \tag{1.86}$$

In general, quantization is a rather delicate and mysterious procedure [220] (see [116] for a recent discussion). However, for algebraic curves defined by classical constraint equations of the form $A(x, y) = 0$, recent progress in mathematical physics [3, 28, 61, 75, 152] has led to a systematic way of constructing the coefficients $a_k(\hat{x}; q)$ of the quantum operator (1.86) entirely from the data of the classical A-polynomial [114] (see also [25]):

$$A(x, y) \quad \rightsquigarrow \quad \hat{A}(\hat{x}, \hat{y}; q). \tag{1.87}$$

In addition, in some cases the curve $A(x, y) = 0$ comes from extra data that may be very helpful in constructing its quantum analog. For instance, the construction [165] of the ordinary A-polynomial based on the triangulation data of a 3-manifold M admits a beautiful non-commutative lift [62]. However, since in what follows we need to apply the procedure (1.87) to arbitrary curves for which the extra data is not always available, we shall mainly focus on the so-called *topological recursion* approach that involves complex analysis and noncommutative algebra on $\mathscr{C}$.

In complex analysis, one of the basic ingredients associated to the curve $\mathscr{C}$: $A(x, y) = 0$ is the so-called *Bergman kernel*. It becomes the first brick in the foundation of the construction (1.87) based on the topological recursion, which after a few more systematic and completely rigorous steps builds the q-difference operator as a power series in $\hbar$:

$$A(x, y) \quad \rightsquigarrow \quad \hat{A}(\hat{x}, \hat{y}; q) = A(\hat{x}, \hat{y}) + \hbar A_1(\hat{x}, \hat{y}) + \cdots . \tag{1.88}$$

Even though we omit the intermediate steps due to constraints of space, the reader should simply be aware that a well-defined, systematic procedure exists. The existence and uniqueness of this procedure are well-motivated based on physical considerations; in fact, these form one of the basic premises of quantum mechanics.

By looking at (1.88) it would seem that we would therefore have to compute terms to arbitrarily high order in this series to write down the operator $\hat{A}$. However, in practice, this is not the case; we usually need to compute only one or sometimes two terms in the series to know $\hat{A}$ exactly! The trick is as follows: if we know, *a priori*, that the operator we construct can be written as a rational function of $q = e^{\hbar}$, then the higher order terms in the expansion in $\hbar$ must resum nicely into an expression of this form. We also have information about

the classical limit ($q \to 1$) of this expression. Armed with this information, it is usually pretty straightforward to construct the quantization of $A(x,y)$ in closed form.

For example, if we know both the classical term and the first quantum correction $\hbar A_1(\hat{x},\hat{y})$ in the expansion (1.88), there is a good chance we can reconstruct the quantum operator

$$\hat{A}(\hat{x},\hat{y};q) \;=\; \sum_{m,n} a_{m,n}\, q^{c_{m,n}}\, \hat{x}^m\, \hat{y}^n \tag{1.89}$$

simply from the data $\{a_{m,n}\}$ of the original polynomial $A(x,y) = \sum a_{m,n} x^m y^n$ and from the exponents $\{c_{m,n}\}$ determined by $\hbar A_1(\hat{x},\hat{y})$. This trick becomes especially useful for curves that come from knots and 3-manifolds. Indeed, in such examples the leading quantum correction is determined by the "classical" knot invariant $\Delta(q)$ called the twisted Alexander polynomial. Therefore, a simple mnemonic rule to remember what goes into the construction of the operator $\hat{A}(\hat{x},\hat{y};q)$ in such situations can be schematically expressed as [114]:

$$\boxed{\text{“ } A(x,y) + \Delta(q) \;\Rightarrow\; \hat{A}(\hat{x},\hat{y};q) \text{ ”}} \tag{1.90}$$

Concretely, the exponents $c_{m,n}$ in (1.89) can be determined by requiring that the relation

$$2\sum_{m,n} a_{m,n}\, c_{m,n}\, x^m y^n \;=\; \frac{\partial A}{\partial \ln x}\left(\frac{\partial A}{\partial \ln y}\right)^{-1} \frac{\partial^2 A}{(\partial \ln y)^2} + x\frac{\partial \Delta(x)}{\partial x}\frac{\partial A}{\partial \ln y} \tag{1.91}$$

holds for all values of x and y (along with $A(x,y)=0$).

Example 6. Consider once more the trefoil knot $K = 3_1$, which has A-polynomial $A(x,y) = (y^{-1}-1)(y+x^3)$ and where, following our earlier agreement, we replaced x^2 by x to simplify the expressions, cf. (1.39). Notice, that $A(x,y)$ in this example is a degree-2 polynomial in y. Quantization (1.88) then gives an operator which is also of degree 2 in $\hat{y}$

$$\hat{A}(\hat{x},\hat{y};q) = \alpha\hat{y}^{-1} + \beta + \gamma\hat{y}, \tag{1.92}$$

where

$$\begin{cases} \alpha = \frac{x^2(x-q)}{x^2-q}; \\ \beta = q\left(1 + x^{-1} - x + \frac{q-x}{x^2-q} - \frac{x-1}{x^2q-1}\right); \\ \gamma = \frac{q-x^{-1}}{1-qx^2}. \end{cases} \tag{1.93}$$

In the representation (1.85), our quantized constraint (1.74) then gives an operator relation that takes the form of a recurrence in the variable n:

$$\hat{A}Z = 0 \implies \alpha(q^n;q)Z_{n-1} + \beta(q^n;q)Z_n + \gamma(q^n;q)Z_{n+1} = 0, \tag{1.94}$$

where we recall that n was defined so that $x = q^n$.

Exercise 4. Solve this recurrence with the initial conditions

$$Z_n = 0 \text{ for } n \leq 0; \quad Z_1 = 1. \tag{1.95}$$

That is, find the first few terms of the sequence $Z_n(q)$ for $n = 2, 3, \ldots$

Solution. Straightforward computation gives

$$\begin{aligned} Z_2(q) &= -\beta(q;q)/\gamma(q;q) \\ &= -\frac{1-q^3}{q-q^{-1}} \cdot q \left(1 + q^{-1} - q - \frac{q-1}{q^3-1}\right) \\ &= -\frac{(1-q^3)(1+q-q^2) + q(q-1)}{q-q^{-1}} \\ &= \frac{-1+q^3+q^4-q^5}{q-q^{-1}} \\ &= q + q^3 - q^4, \end{aligned} \tag{1.96}$$

as well as

$$\begin{aligned} Z_3(q) &= -(\alpha(q^2;q) + \beta(q^2;q)Z_2(q))/\gamma(q^2;q) \\ &= q^2 + q^5 - q^7 + q^8 - q^9 - q^{10} + q^{11}, \end{aligned} \tag{1.97}$$

after a little manipulation. Notice that the Z_n all turn out to be polynomials! □

Now, we come to one of the punch lines of these lectures. The reader who has completed Exercise 1 and followed through the derivation of (1.21) may have noticed a startling coincidence: Z_n produced by our our recurrence relation (1.94) is none other than the n-colored Jones polynomial; that is, the quantum group invariant of the knot decorated with extra data consisting of the Lie algebra $\mathfrak{g} = \mathfrak{sl}(2)$ and its n-dimensional representation $R = V_n$.

This is no coincidence, of course. As we reviewed in §1.1, the n-colored Jones polynomial is simply the partition function of Chern-Simons TQFT with gauge group $G = SU(2)$. On the other hand, in this section we explained that

the classical A-polynomial and its quantum, non-commutative version have a natural home in Chern-Simons TQFT with complex gauge group $G_{\mathbb{C}} = \mathrm{SL}(2,\mathbb{C})$. In particular, we saw how the usual rules of quantum mechanics replace the classical constraint (1.80) with an operator relation (1.74),

$$\mathscr{C} : A(x,y) = 0 \quad \rightsquigarrow \quad \hat{A}(\hat{x},\hat{y};q) Z_{\mathrm{CS}}(M) = 0\,, \tag{1.98}$$

where $Z_{\mathrm{CS}}(M)$ is the state vector associated by quantization to the Lagrangian submanifold $\mathscr{C}$ (or, equivalently, associated by the Chern-Simons TQFT functor to the 3-manifold M). Since $G_{\mathbb{C}} = \mathrm{SL}(2,\mathbb{C})$ is a complexification of $G = SU(2)$, the partition functions in these two theories are closely related [68, 217]. In particular, it was argued in [107] that *both* SU(2) and $\mathrm{SL}(2,\mathbb{C})$ partition functions must satisfy the quantum constraint equation (1.98). In the n-representation (1.85) it takes the form of a *recurrence relation*

$$A(x,y) = \sum_k a_k(x) y^k \quad \rightsquigarrow \quad \sum_k a_k(q^n;q) J_{n+k}(K;q) = 0\,, \tag{1.99}$$

which is precisely our q-difference equation (1.94) in the above example, where K was taken to be the trefoil knot. More generally, the equation (1.99) is a q-difference equation, describing the behavior with respect to n, or "color dependence," of the n-colored Jones polynomial that is computed by Wilson loop operators in the SU(2) Chern-Simons theory.

The relation between the quantization of the A-polynomial and the quantum group invariants (1.99) that follows from Chern-Simons theory is the statement of the *quantum volume conjecture* [107] (see [66] for a review of earlier developments that led to it). This conjecture was independently proposed in [94] around the same time, and is also known as the *AJ-conjecture*. It provides a bridge between two seemingly distant areas of knot theory, the classical A-polynomial and the study of quantum group invariants. Before the discovery of this connection, the separate communities of knot theorists working on these two different types of invariants had very little contact with one another.

Do two knots having the same A-polynomial always have all the same n-colored Jones polynomials? Based on the above connection, we would expect an affirmative answer, given that the quantization procedure for the A-polynomial is essentially unique. This has been checked for knots up to large number of crossings, although there is as yet no formal proof. If it is true, then a single

algebraic curve constructed without any reference to quantum groups encodes all the information about the whole tower of n-colored Jones polynomials:

$$A(x,y) \quad \rightsquigarrow \quad \hat{A}(\hat{x},\hat{y};q) \quad \rightsquigarrow \quad J_n(K;q)\,. \tag{1.100}$$

Nonetheless, even if all the n-colored Jones polynomials together carry no more information than the A-polynomial, their relation to quantum groups still makes them interesting objects of study in their own right. (It is also worth noting that the study of the colored Jones polynomial predates the discovery of the A-polynomial.)

Once we explained how to go, via quantization, from the classical A-polynomial to quantum group invariants (1.100) it is natural to ask whether there is a simple way to go back. The *generalized volume conjecture* [107] proposes an affirmative answer to this question and is also based on the fact that the analytic continuation of SU(2) is $\mathrm{SL}(2,\mathbb{C})$. It states that, in the classical limit $q \to 1$ accompanied by the "large color" limit $n \to \infty$, the n-colored Jones polynomial, as a Wilson line in SU(2) Chern-Simons theory [212], exhibits the exponential behavior

$$J_n(K;q=e^{\hbar}) \underset{\substack{n\to\infty\\ \hbar\to 0}}{\sim} \exp\left(\frac{1}{\hbar}S_0(x)+\cdots\right), \tag{1.101}$$

where the limits are taken with $q^n = x$ held fixed. Here $S_0(x)$ is the classical action of $\mathrm{SL}(2,\mathbb{C})$ Chern-Simons theory, which is

$$S_0(x) = \int \log y \frac{dx}{x} \tag{1.102}$$

evaluated on a path within the curve $\mathscr{C} : A(x,y) = 0$. Here, by an abuse of notation, the variable x stands in for a point on the Riemann surface; S_0 is actually a function on $\mathscr{C}$, and the integral in (1.102) is taken along a path in $\mathscr{C}$ from some fixed base point to the point at which S_0 is evaluated. Moreover, (1.102) is only well defined if the integrality condition (1.44) holds! The change ΔS_0 that comes from composing the path used in our evaluation with an arbitrary closed cycle must be valued in $2\pi\mathbb{Z}$, so that the quantity e^{iS_0} is well-defined and independent of path; the integrality condition ensures that this is so.

To summarize, the generalized volume conjecture gives us two important ways of thinking about the A-polynomial: firstly, as a characteristic variety encoding

information about $\mathrm{SL}(2,\mathbb{C})$ flat connections, and secondly, as a *limit shape* in the limit of large color.

We have now begun to see how the seemingly disparate topics we have been discussing are connected to one another. Roughly speaking, there are four major themes in these lectures: quantum group invariants, the A-polynomial, quantization, and categorification. We have now seen how quantization relates the A-polynomial and quantum group invariants, providing a bridge between seemingly unrelated knot polynomials. In what remains, we will return to ideas of categorification, hoping to give at least a glimpse of how knot polynomials arise from deeper and more powerful homological invariants.

1.4 Categorification

Categorification is a powerful and flexible idea; it can mean different things in different contexts, and a given mathematical construction may admit more than one categorification depending on how one chooses to look at its structure. In the context of topological quantum field theories, however, categorification is manifested in a very natural way. The categorification of a 3-dimensional TQFT should be a 4-dimensional TQFT, from which the 3D theory is recovered by dimensional reduction, see e.g. [56, 108]. That is,

$$\text{3D TQFT} \underset{\text{dimensional reduction}}{\overset{\text{categorification}}{\rightleftarrows}} \text{4D TQFT}$$

We can tabulate the information that each of these TQFTs should associate to geometrical objects in the below table:

Geometry	*3D TQFT*	*4D TQFT*
3-manifold M, knot $K \subset M$	number $Z(M)$, polynomial invariant $P(K)$	vector space $\mathscr{H}_K$
2-manifold Σ	vector space $\mathscr{H}_\Sigma$	category $\mathbf{Cat}_\Sigma$

Thus, to a geometrical object of given dimension, a categorified TQFT associates objects of one higher categorical level than its decategorified counterpart. (The categorical level of the object associated by a TQFT to something in geometry corresponds to its *codimension*, so that a 4D TQFT assigns numerical invariants to 4-manifolds. Famous examples of these are given by Donaldson theory [211] and Seiberg-Witten theory [214].)

$i=3$	·	·	·	·	·	·	·	·	$\mathbb{Z}$
2	·	·	·	·	$\mathbb{Z}$	·	$\mathbb{Z}/2$	·	·
1	·	·	·	·	·	·	·	·	·
0	$\mathbb{Z}$	·	$\mathbb{Z}$	·	·	·	·	·	·
	1	2	3	4	5	6	7	8	$9=j$

Figure 1.8: The Khovanov homology $H_{i,j}(K=3_1)$ of the trefoil knot.

In 2000, Mikhail Khovanov [128] succeeded in constructing a categorification of the Jones polynomial. Like the Jones polynomial, it is associated to the extra data $\mathfrak{g}=\mathfrak{sl}(2)$ and its fundamental representation $R=V_2$. To give the barest outline, his construction associates a chain complex to a diagram of a link K. The homology of this chain complex can be shown to be invariant under the Reidemeister moves, and therefore to be an invariant of K. *Khovanov homology* $H_{i,j}(K)$ is doubly graded, and the Jones polynomial is its graded Euler characteristic, cf. (1.1),

$$J(q)=\sum_{i,j}(-1)^i q^j \dim H_{i,j}(K)\,. \tag{1.103}$$

Sometimes it is convenient to encode information about the Khovanov homology in its Poincaré polynomial:

$$\mathrm{Kh}(q,t)=P_{\mathfrak{sl}(2),V_2}(q,t)=\sum_{i,j} t^i q^j \dim H_{i,j}(K). \tag{1.104}$$

The Jones polynomial is then recovered by making the evaluation at $t=-1$. As an example, the Poincaré polynomial of the trefoil knot is

$$\mathrm{Kh}(q,t;K=3_1)=q+q^3t^2+q^4t^3\,. \tag{1.105}$$

It is easy to see that the evaluation at $t=-1$ indeed returns the normalized Jones polynomial of the trefoil knot (1.21) that we saw in §1.1. By definition, this version of the homology is called *reduced*. Its close cousin, the *unreduced knot homology* categorifies the *unnormalized* polynomial invariant. Thus, for the unnormalized Jones polynomial (1.15) of $K=3_1$ the corresponding categorification is given by the unreduced Khovanov homology shown in Fig. 1.8.

Much like the Khovanov homology of a knot is a categorification of its Jones polynomial or quantum $\mathfrak{sl}(2)$ invariant, there exist generalizations [53, 85, 205, 221, 224] of the Khovanov homology categorifying the n-colored Jones

polynomials for all n:

$$J_n(K;q) = P_n(K;q,t)|_{t=-1} = \sum_{i,j} q^i t^j \dim H_{i,j}^{\mathfrak{sl}(2),V_n}(K)\Big|_{t=-1} . \tag{1.106}$$

The n-colored $\mathfrak{sl}(2)$ knot homologies satisfy recursion relations, just like their decategorified versions, and exhibit beautiful asymptotic behavior in the limit of large n. Both of these behaviors are controlled by a *refined algebraic curve*, which is an analogue of the A-polynomial [90]:

$$\mathscr{C}^{\rm ref}: \ A^{\rm ref}(x,y;t) = 0\,. \tag{1.107}$$

This curve is a t-deformation of (the image of) the representation variety of a knot complement M in the classical phase space of the Chern-Simons theory, which is the moduli space $\mathscr{M}_{\rm flat}(\mathrm{SL}(2,\mathbb{C}),\Sigma)$ of flat connections. Here $\Sigma = \partial M$. Much like the representation variety (1.80) of M, its t-deformation (1.107) is a holomorphic Lagrangian submanifold with respect to the symplectic form (1.81).

Example 7. In §1.2 we derived the A-polynomial of the trefoil knot (1.39). Then, in §1.3 we discussed its quantization, or *non-commutative* q-deformation. In both cases, the result is a quadratic polynomial in y. Similarly, the *commutative* t-deformation of the A-polynomial for the trefoil knot is a quadratic polynomial in y,

$$A^{\rm ref}(x,y;t) = y^2 - \frac{1 - xt^2 + x^3t^5 + x^4t^6 + 2x^2t^2(t+1)}{1+xt^3}y + \frac{(x-1)x^3t^4}{1+xt^3}, \tag{1.108}$$

which reduces to the ordinary A-polynomial (1.39) in the limit $t = -1$.

As in §1.3, quantization of $\mathscr{M}_{\rm flat}(\mathrm{SL}(2,\mathbb{C}),\Sigma)$ with its natural symplectic form promotes x and y to operators obeying the commutation relation

$$\hat{y}\hat{x} = q\hat{x}\hat{y} \tag{1.109}$$

and turns the planar algebraic curve (1.107) into a q-difference recursion relation, cf. (1.99),

$$\boxed{\hat{A}^{\rm ref} P_\star(K;q,t) \simeq 0\,,} \tag{1.110}$$

where $\hat{x}P_n = q^n P_n$ and $\hat{y}P_n = P_{n+1}$. This recursion relation, called the *homological volume conjecture* in [90], provides a natural categorification of the

generalized volume conjecture that was the subject of §1.3. Unlike the generalized volume conjecture, its homological version (1.110) is based on a much more sophisticated physics that involves a physical interpretation of knot homologies in terms of refined BPS invariants [108, 112] and dynamics of supersymmetric gauge theories [64, 67, 89, 218]. The details of this physical framework go way beyond the scope of these lectures and we simply refer the interested reader to the original papers.

There also exists a homology theory categorifying the HOMFLY-PT polynomial [131, 132]. As should be obvious, this theory must be triply graded; the HOMFLY-PT polynomial is recovered by taking the graded Euler characteristic, cf. (1.103),

$$P_{a,q}(K) = \sum_{ijk} (-1)^i q^j a^k \dim \mathscr{H}_{ijk}(K)\,. \tag{1.111}$$

Just as we did for *Khovanov homology*, we can construct the Poincaré polynomial associated to the *HOMFLY homology*, which will encode information about the dimensions of its groups at each level:

$$\mathscr{P}(a,q,t) = \sum_{ijk} t^i q^j a^k \dim \mathscr{H}_{ijk}(K)\,. \tag{1.112}$$

Then decategorification corresponds once more to evaluation at the value $t = -1$. It turns out that even the HOMFLY homology is not a complete invariant of knots; nonetheless, these homological invariants are strictly finer and stronger than their decategorified counterparts. For instance, HOMFLY homology can distinguish between the knots 5_1 and 10_{132}, discussed earlier, that have identical Jones, Alexander, and HOMFLY-PT polynomials (1.18).

We should remark also that n-colored generalizations of HOMFLY homology can be constructed, and that the color dependence can be encoded in an algebraic curve, just as the zero locus of the A-polynomial encodes the information about color dependence of the n-colored Jones polynomial. We will return to this point and discuss the corresponding algebraic curve in much more detail in the final section of these lectures. Meanwhile, in the rest of this section we mostly focus on the ordinary, *uncolored* HOMFLY homology aiming to explain its structure and how to compute it in practice.

As we shall see, the structure of the homological knot invariants turns out to be so rich and so powerful that, once we learn enough about it, we will be able

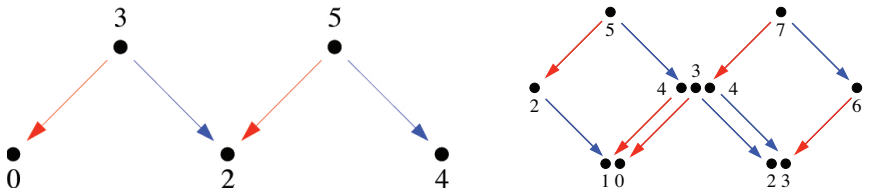


Figure 1.9: The HOMFLY homology for knots 5_1 and 10_{132}. Each dot represents a generator, with its vertical and horizontal position indicating a- and q-degree respectively. The labels represent t-degree. The diagram can be thought of as similar to a Newton diagram for the Poincaré polynomial (1.112). (Images from [72].)

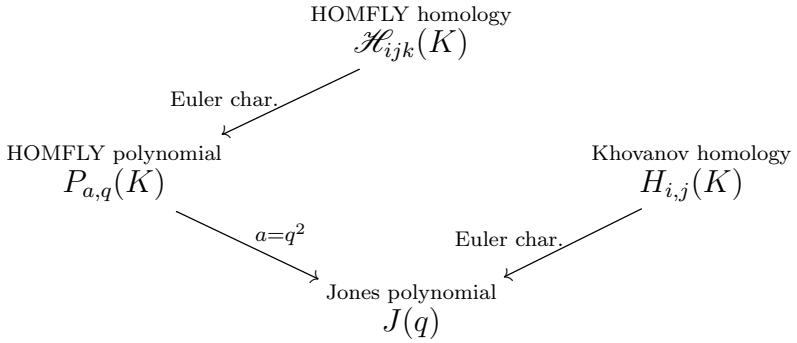
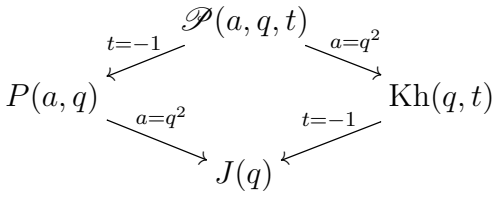


Figure 1.10: A summary of relations between homological and polynomial invariants.

to compute, say, the Khovanov homology and the HOMFLY homology of the trefoil knot solely from the data of its Jones polynomial. In other words, in a moment we will learn powerful techniques that will allow us to reproduce (1.105) without even learning the definition of Khovanov homology. And, much of this structure is present—in fact, in a richer form!—in the *colored HOMFLY homology* as well [113].

Let us start by summarizing the familiar relations (1.13), (1.103), (1.111) between homological and polynomial invariants diagramatically, as shown in Fig. 1.10. We would like to be able to fill in the missing fourth arrow, i.e., to have a way of recovering Khovanov homology directly from the HOMFLY homology. This, however, is rather delicate for a number of reasons. First, the specialization $a = q^2$ does not make sense in the context of the homology theories. At best one could try to complete the diagram by working with the

Poincaré polynomials associated to these theories:



(1.113)

As we explain shortly, even this is too naive due to a simple, yet conceptual reason. Nevertheless, for a moment let us ignore this issue and proceed as if (1.113) were actually correct.

Example 8. Let us see if we can use the information in (1.113) to reconstruct $\mathscr{P}(a,q,t)$ for the trefoil knot. We know already that

$$\begin{cases} P(a,q) = aq^{-1} + aq - a^2, \\ \mathrm{Kh}(q,t) = q + q^3t^2 + q^4t^3. \end{cases} \tag{1.114}$$

We can attempt to guess $\mathscr{P}(a,q,t)$ just by comparing terms; this gives

$$\mathscr{P}(a,q,t) = aq^{-1} + aqt^2 + a^2t^3. \tag{1.115}$$

This naive guess turns out to be correct! Using only information from the HOMFLY-PT polynomial and Khovanov homology (both of which are easily computable), we have obtained information about the triply-graded HOMFLY homology theory, which encodes information about the $\mathfrak{sl}(N)$ homological invariants for all N.

In fact, one can even get to (1.115) without knowing the Khovanov homology! Our task is to assign a t-degree to each term in the HOMFLY-PT polynomial. We can do this using the following trick: From Exercise 1, the reader should know that evaluating $P(a,q)$ at $a=q$ yields a monomial (exactly which monomial depends on a simple knot invariant and a choice of normalization). This turns out to be true for any knot: the HOMFLY-PT polynomial will always become trivial, i.e., monomial, when evaluated at $a=q$. Therefore, to ensure the needed cancellation when the specialization $a=q$ is made, the normalized HOMFLY-PT polynomial for any knot must have the following schematic form:

$$P_{a,q} = 1 + (1 - a^{-1}q)Q(a,q), \tag{1.116}$$

where Q is some polynomial factor. The basic reason for this is that taking $a = q$ corresponds to asking about the $\mathfrak{sl}(1)$ polynomial invariant, which must always be trivial. A similar simplification happens in the case $a = q^{-1}$.

What about the $\mathfrak{sl}(1)$ *homological* invariant? Since $\mathscr{P}(a, q, t)$ has only positive coefficients, $\mathscr{P}(q, q, t)$ can't be trivial—it must reduce to a monomial only because of cancellations that occur for $t = -1$. But we would not expect to be able to construct any nontrivial invariants with $\mathfrak{sl}(1)$, homological or otherwise. This is a clue that something more sophisticated must be happening in the way that one extracts Khovanov homology (generally, $\mathfrak{sl}(N)$ homology) from the HOMFLY homology.

The reason, to which we alluded earlier, is that when polynomial knot invariants are categorified one correspondingly needs to upgrade the specialization $a = q^N$ of §1.1 to the homological level. In other words, trying to use the specialization $a = q^N$ as we did in diagram (1.113) is too naive and the suitable operation should also be from the world of homological algebra.

It turns out that the correct homological lift of the specialization $a = q^N$ involves a conceptually new ingredient, which has no analog at the (decategorified) polynomial level: a family of differentials $\{d_N\}$ on the HOMFLY homology, indexed by $N \in \mathbb{Z}$. These differentials endow HOMFLY homology with a structure that is much richer than what can be seen at the polynomial level and that is responsible for our claim that (1.115) can be derived even without the knowledge of the Khovanov homology. By viewing the triply-graded homology as a complex and taking its homology with respect to this differential, one recovers the doubly-graded Khovanov homology. Specifically, in the grading conventions of [113], the differentials have degree

$$\begin{aligned} d_{N>0} &:(-1, N, -1), \\ d_{N\leq 0} &:(-1, N, -3) \end{aligned} \tag{1.117}$$

with respect to (a, q, t) grading. The homology of $\mathscr{H}_\star$, viewed as a complex with differential d_N, returns the doubly-graded $\mathfrak{sl}(|N|)$ homology theory [131] or the *knot Floer homology* [169, 174] in the special case $N = 0$, see [72] for details. In particular, its homology with respect to the differentials d_1 and d_{-1} must be trivial.

For instance, in considering the reduction of HOMFLY homology to the $\mathfrak{sl}(1)$ homological invariant, almost all of the terms in the triply-graded HOMFLY

homology will be killed by the differential d_1, leaving behind a "trivial" one-dimensional space,

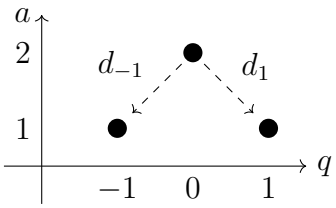
$$\dim\left(\mathscr{H}_\star, d_1\right) = 1\,. \tag{1.118}$$

Because the differential d_1 has definite grading (1.117), the Poincaré polynomial of HOMFLY homology therefore must be of the following general form

$$\mathscr{P}(a,q,t) = 1 + (1 + a^{-1}qt^{-1})Q_+(a,q,t)\,, \tag{1.119}$$

where the first term represents a contribution of the (trivial) $\mathfrak{sl}(1)$ knot homology, and $Q_+(a,q,t)$ is some polynomial with positive coefficients. Note, that the Poincaré polynomial (1.119) necessarily has all of its coefficients non-negative. Similar structure follows from the existence of another *canceling differential* d_{-1} that also kills all but one generators of the HOMFLY homology. The physical interpretation of the differentials $\{d_N\}$ can be found in [113].

Now, just from the little we learned about the differentials d_1 and d_{-1}, we can reconstruct the HOMFLY homology of the trefoil knot. First, we can get information about the a- and q-degrees of nontrivial HOMFLY homology groups just from the HOMFLY-PT polynomial. For the trefoil knot, these are depicted below:



It is clear that each of the differentials $d_{\pm 1}$ can only act nontrivially in one place. From the condition that they give rise to trivial homology, each must be surjective; this determines the relative t-degree of each group. Taking the point with (a,q)-degree $(1,-1)$ to have $t = 0$, it immediately follows that the degrees of the other groups with respect to (a,q,t) degree are $(2,0,3)$ and $(1,1,2)$. We have now managed to extract this information without even computing Khovanov homology; the results of Exercise 1 and the above trick are all we need.

1.5 Epilogue: super-A-polynomial

In this section, we give a somewhat deeper discussion of the connection between physics, homological knot invariants, and the quantization of the A-polynomial, constructing one final bridge between the ideas of quantization and categorification. This final section of the lectures can be seen as an addendum; based on recent progress [2, 89, 91, 160] it summarizes material that was covered in a talk given at the conference following the summer school, and so is somewhat more technical.

In these lectures, we saw several deformations of the classical A-polynomial $A(x,y)$ introduced in §1.2. In §1.3 we saw how quantization of $\mathrm{SL}(2,\mathbb{C})$ Chern-Simons theory leads to a *non-commutative* q-deformation (1.47). Then, in §1.4, we saw how more sophisticated physics based on refined BPS invariants leads to a categorification of the generalized volume conjecture and a *commutative* t-deformation (1.107).

These turn out to be special cases of a more general three-parameter "super-deformation" of the A-polynomial introduced in [91]. Two out of these three deformations are *commutative* and will be parametrized by a and t, while the third *non-commutative* deformation is produced essentially by the quantization procedure (1.88) of §1.3:

$$A^{\mathrm{super}}(x,y;a,t) \quad \rightsquigarrow \quad \hat{A}^{\mathrm{super}}(\hat{x},\hat{y};a,q,t)\,. \tag{1.120}$$

What is the meaning of this *super-A-polynomial*?

The best way to answer this question is to consider an example. In fact, let us repeat the analogs of Example 6 and Exercise 4:

Example 9. For our favorite example, the trefoil knot $K = 3_1$, we know from our earlier discussion that the classical A-polynomial $A(x,y) = (y-1)(y+x^3)$ is quadratic in y, and so are its t-deformation (1.108) and q-deformation (1.92). The same is true of the super-A-polynomial of $K = 3_1$,

$$A^{\mathrm{super}}(x,y;a,t) = y^2 - \frac{a\,(1-t^2x+2t^2(1+at)x^2+at^5x^3+a^2t^6x^4)}{1+at^3x}y + \frac{a^2t^4(x-1)x^3}{1+at^3x}, \tag{1.121}$$

which clearly reduces to (1.108) upon setting $a = 1$ and to the ordinary A-polynomial (1.39) upon further specialization to $t = -1$. Moreover, the

quantization procedure of §1.3 turns super-A-polynomial (9) into a q-difference operator, which can be interpreted as a recurrence relation, similar to (1.94),

$$\hat{A}^{\text{super}}(\hat{x},\hat{y};a,q,t) = \alpha + \beta\hat{y} + \gamma\hat{y}^2$$
$$\Longrightarrow \quad \alpha\mathscr{P}_n + \beta\mathscr{P}_{n+1} + \gamma\mathscr{P}_{n+2} = 0\,. \quad (1.122)$$

Here, the coefficients α, β, and γ are certain rational functions of the variables a, q, $x \equiv q^n$, and t, whose explicit form can be found in [91].

Exercise 5. As in Exercise 4, solve the recurrence (1.122) with the initial conditions

$$\mathscr{P}_n = 0 \text{ for } n \le 0; \quad \mathscr{P}_1 = 1. \quad (1.123)$$

That is, find the first few terms of the sequence $\mathscr{P}_n(q)$ for $n = 2, 3, \ldots$

Solution. Straightforward computation gives:

n	$\mathscr{P}_n(a,q,t)$
1	1
2	$aq^{-1} + aqt^2 + a^2t^3$
3	$a^2q^{-2} + a^2q(1+q)t^2 + a^3(1+q)t^3 + a^2q^4t^4 + a^3q^3(1+q)t^5 + a^4q^3t^6$
4	$a^3q^{-3} + a^3q(1+q+q^2)t^2 + a^4(1+q+q^2)t^3 + a^3q^5(1+q+q^2)t^4 +$ $+ a^4q^4(1+q)(1+q+q^2)t^5 + a^3q^4(a^2+a^2q+a^2q^2+q^5)t^6 +$ $+ a^4q^8(1+q+q^2)t^7 + a^5q^8(1+q+q^2)t^8 + a^6q^9t^9$

How should we interpret these polynomial invariants? The answer can be guessed from a couple of clues in the above table: firstly, all $\mathscr{P}_n(a,q,t)$ involve only positive integer coefficients. Secondly, we have seen $\mathscr{P}_2(a,q,t)$ before; it is the Poincaré polynomial (1.115) of the triply-graded HOMFLY homology of the trefoil knot! □

These considerations lead one to guess, correctly, that $\mathscr{P}_n(a,q,t)$ is the Poincaré polynomial of the n-colored generalization of the HOMFLY homology:

$$\mathscr{P}_n(a,q,t) = \sum_{ijk} t^i q^j a^k \dim \mathscr{H}^{(n)}_{ijk}(K)\,. \quad (1.124)$$

Naively, one might expect that making the specialization $a = q^2$ in the polynomial $\mathscr{P}_n(a,q,t)$ should return the Poincaré polynomial for the n-colored $\mathfrak{sl}(2)$ homology in (1.106), and so forth. However, in the homological world,

this specialization is a little bit more subtle. It turns out that, just as we saw earlier in §1.4, the colored homology $\mathscr{H}^{(n)}_{ijk}(K)$ comes naturally equipped with a family of differentials d_N; viewing $\mathscr{H}^{(n)}_{ijk}(K)$ as a complex and taking its homology with respect to the differential d_2 allows one to pass directly from the n-colored HOMFLY homology to the n-colored analog of the Khovanov homology.

To summarize, the super-A-polynomial encodes the "color dependence" of the *colored HOMFLY homology*, much like the ordinary A-polynomial and its t-*deformation* do for the colored Jones polynomial (1.99) and the colored $\mathfrak{sl}(2)$ homology (1.110), respectively:

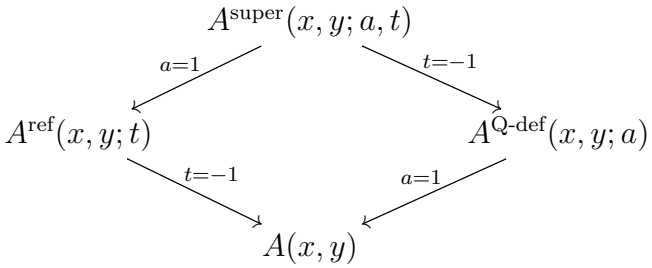
$$\boxed{\hat{A}^{\text{super}}\,\mathscr{P}_\star(a,q,t)\simeq 0.} \tag{1.125}$$

Moreover, setting $q=1$ gives the classical super-A-polynomial with two commutative parameters a and t. Its zero locus defines an algebraic curve

$$\mathscr{C}^{\text{super}}:\ A^{\text{super}}(x,y;a,t)=0, \tag{1.126}$$

which in various limits reduces to the A-polynomial curve (1.24) and its "refined" version (1.107). This curve plays the same role for colored HOMFLY homology as the ordinary A-polynomial does for the colored Jones invariants. Specifically, there is an obvious analog of the generalized volume conjecture (1.101), which states that (1.126) is the *limit shape* for the S^n-colored HOMFLY homology in the large color limit $n\to\infty$ accompanied by $q\to 1$ [91].

A simple way to remember different specializations of the two-parameter "super-deformation" of the A-polynomial is via the following diagram:



(1.127)

which should remind the reader of the diagram (1.113) expressing a similar relation between various polynomial and homological invariants discussed here. Indeed, each of the invariants in (1.113) has a n-colored analog whose

color dependence is controlled by the corresponding deformation of the A-polynomial in (1.127). In this diagram, we included yet another deformation of the A-polynomial, which can be obtained from the super-A-polynomial by setting $t = -1$. This so-called *Q-deformation* of the A-polynomial was recently studied in [2], where it was conjectured that $A^{\text{Q-def}}(x, y; a)$ agrees with the augmentation polynomial of *knot contact homology* [74, 166, 167].

As a closing remark, we should mention that the colored homological invariants have even more structure than we have so far discussed. One can also construct a family of *colored differentials*, which act by removing boxes from Young tableaux or reducing the dimension of the representation in the decoration of a link diagram [113]. For example,

$$(\mathscr{H}^{\square\square}, d_{\text{colored}}) \simeq \mathscr{H}^{\square}, \tag{1.128}$$

where $(\mathscr{H}^{\square\square}, d_{\text{colored}})$ denotes the homology of the complex with respect to the indicated differential. This can be expressed for the respective Poincaré polynomials by a relation of the form (1.119):

$$\mathscr{P}^{\square\square}(a, q, t) = a^s \mathscr{P}^{\square}(a, q^2, t) + (1 + at) Q_+(a, q, t), \tag{1.129}$$

showing the color dependence of these invariants in the form that nicely integrates with the recursion (1.125).

In general, there are many more colored differentials, which altogether form a very rich and rigid structure [113]. To fully appreciate the beauty and the power of this structure one needs to consider *homologically thick knots*. Roughly speaking, these are the knots whose homological invariants contain a lot more new information compared to their polynomial predecessors. The knot 8_{19}—equivalently, the torus knot $T^{(3,4)}$—is the first example of a homologically thick knot. Other examples of homologically thick knots and links include *mutants*.

In the case of n-colored HOMFLY homology that we discussed earlier, the colored differentials include the differentials d_N of §1.4 for special values of N in the range $-2n+3, \ldots, 1$. Note, in the uncolored theory ($n = 2$) this range contains only three differentials, $d_{\pm 1}$ and d_0, which play a very special role. Namely, the first two are canceling differentials, whereas d_0 is the differential that relates HOMFLY homology to *knot Floer homology* [72]. We emphasize that the last relation really requires the knowledge of how d_0 acts

on HOMFLY homology, which is an extra data not contained in the Poincaré polynomial $\mathscr{P}(a,q,t)$. Curiously, this extra data is automatically contained in the colored version of the HOMFLY homology, so that *knot Floer homology* can be recovered directly from $\mathscr{P}_n(a,q,t)$, even for homologically thick knots!

The reason for this is that all three special differentials d_1, d_{-1}, and d_0, have analogs in the n-colored theory. Moreover, they are part of the colored differentials d_N, with $N = -2n+3, \ldots, 1$. Specifically, in the n-colored HOMFLY homology the differentials d_1 and d_{1-n} are canceling, whereas d_{2-n} provides the relation to knot Floer homology [106, 113]. And the virtue of the colored theory is that the action of this latter differential can be deduced from the data of $\mathscr{P}_n(a,q,t)$ alone. In other words, what in the uncolored theory appears as a somewhat bizarre and irregular behavior at $N = -1, 0, +1$ becomes a natural and simple structure in the colored theory.

Chapter 2

SUPERSYMMETRIC QUANTUM MECHANICS: ALGEBRAIC MODELS OF GEOMETRIC OBJECTS

[1]H. Kim and I. Saberi. "Real homotopy theory and supersymmetric quantum mechanics." (2015). arXiv:1511.00978 [hep-th].

Introduction

Supersymmetric quantum mechanics, which describes a supersymmetric particle moving on a compact Riemannian manifold, has been studied by many authors (see, for example, [5, 209, 210]) and has deep connections to the geometry and topology of the target space M on which the particle moves. The theory is a supersymmetric sigma model in $0+1$ dimensions. Its Hilbert space is the space of complex-valued differential forms on M, equipped with the inner product that generalizes the L^2 inner product from undergraduate quantum mechanics:

$$(\alpha, \beta) = \int_M \star\bar{\alpha} \wedge \beta. \tag{2.1}$$

The bar here denotes complex conjugation. The $\mathscr{N} = 1$ supersymmetry algebra in $0+1$ dimensions [5] consists of the exterior derivative d and its adjoint $d^\dagger$ with respect to the inner product (2.1), as well as their anticommutator, the Laplacian (which is the Hamiltonian operator of the theory, and commutes with the supercharges). With respect to the grading, d carries degree one.

We will sometimes refer to this collection of operators as the de Rham algebra. Representations of this algebra are of two types: there is the standard or "long" representation, consisting of two generators of adjacent degree that are mapped to one another by the supercharges, and there are "short" one-dimensional representations that are annihilated by both d and $d^\dagger$. The "BPS bound" in this theory is simply $\Delta \geq 0$. It follows from the equation $\Delta = \{d, d^\dagger\}$, and it is easily shown that a representation is short if and only if it has zero energy. We draw a picture of the spectrum of the theory in Fig. 2.1.

The BPS states of the theory also admit a description as the cohomology of a chosen supercharge. There is only one supercharge to choose, namely d; it is

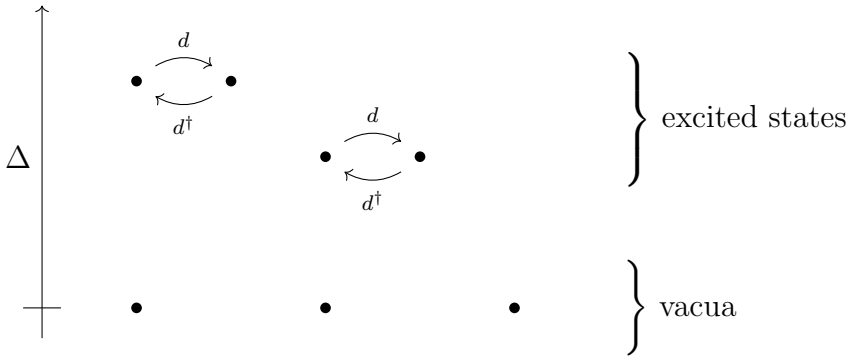


Figure 2.1: A schematic diagram of the de Rham complex of a compact manifold, drawn to emphasize the connections with supersymmetry. The vertical axis is the energy (eigenvalue of the Laplacian), and the horizontal axis is the homological degree. The compactness assumption, which ensures that the spectrum is discrete, is crucial and cannot be relaxed.

a scalar, since we are in $0+1$ dimensions, and so we can think of passing to its cohomology as a topological twist of the theory.

This story continues in many directions: for instance, one can imagine adding a superpotential term to the action. The resulting physics has a beautiful description in terms of the WKB approximation and the Morse complex of M. However, we will not continue down this path, and refer the interested reader to [210]. Instead, we wish to emphasize some points of differential topology which are less well known to most physicists.

When one studies a space M using an algebraic invariant, it is natural to ask how well that invariant distinguishes inequivalent spaces. Many different manifolds share the same Euler characteristic, while the simplicial chain complex of M has perfect information: from it we can reconstruct M up to homeomorphism by gluing together a simplicial complex. Of course, there are many different simplicial complexes corresponding to the same manifold; in this sense, despite containing perfect information, a triangulation of M is not an invariant at all.

The cohomology of M is an honest invariant, lying somewhere between these two in strength: it determines the Euler characteristic and is determined by a triangulation. But how close is it to a perfect invariant of M?

In physics, the simplicial chain complex of the target space M does not arise naturally. However, the Hilbert space associated to M (i.e. its de Rham com-

plex) plays a similar role. We are therefore led to ask: How much information about a manifold can be recovered from its de Rham complex? And how much information about the de Rham complex is contained in its cohomology—or, in physical language, how much can we learn about a supersymmetric quantum mechanics by studying its BPS spectrum?

To address these questions, we recall some structures and results of classical algebraic topology. (For more background, see [26].) Cohomology is an invariant of homotopy type: a space that is homotopy-equivalent to M will have identical cohomology. For example, all the spaces $\mathbb{R}^n$ for every n have the homotopy type of a point. However, assuming that M is compact without boundary, one can also deduce its dimension from its cohomology through Poincaré duality. One therefore has a way of distinguishing between homotopy-equivalent spaces such as M and $M \times \mathbb{R}$.

A homotopy equivalence between M and another space N induces a relationship (also called homotopy equivalence) between their respective de Rham complexes, considered as commutative differential graded algebras. Homotopy-equivalent CDGAs have identical cohomology, but the converse statement is not true. To see why, we need to consider algebraic operations defined on cohomology classes.

The de Rham complex is an algebra. Its multiplication is the wedge product of differential forms. This product descends to cohomology classes, which therefore form an algebra. It also (in more subtle fashion) induces *additional* structures: the cohomology carries a set of higher operations known as *Massey products*. These are partially defined functions and, where they are defined, may or may not vanish. Roughly speaking, the Massey products record ambiguities about how products of certain harmonic forms become exact. Another intuitive picture is that Massey products are like higher-order linking numbers: they detect or measure entangledness between sets of three or more cycles which are pairwise unlinked, like the Borromean rings (Fig. 2.2). Here we have been sloppy and identified cocycles with their Poincaré-dual homology cycles.

In the simplest case, the Massey triple product is defined as follows. Let u, v, and w be representatives of three nontrivial cohomology classes in $\mathrm{H}^\bullet(M)$, of homogeneous degree, and let $\bar{u} = (-)^{F+1}u$, where F denotes the degree operator. (This convention helps keep track of signs.) Furthermore, suppose

that the pairwise products $[u][v] = [v][w] = 0$; in other words, that $\bar{u} \wedge v = ds$ for some form s, and similarly $\bar{v} \wedge w = dt$.The Massey product of u, v, and w is then given by the following expression:

$$m(u, v, w) = [\bar{s} \wedge w + \bar{u} \wedge t]. \tag{2.2}$$

It is simple to check that the representative $\bar{s} \wedge w + \bar{u} \wedge t$ is closed. For the Massey product of cohomology classes to be well defined, one should check that the choice of representatives u, v, and w does not affect the end result. In fact, the end result is not exactly invariant, but it becomes well-defined as an element of a particular quotient of the cohomology.

If one can make a consistent choice of representatives for cohomology classes of M such that all higher Massey products simultaneously vanish, one says that M is *formal*. More transparently, a space is formal if its de Rham complex is homotopy-equivalent to its cohomology ring, viewed as a CDGA with zero differential. Passing to cohomology thus loses information about the homotopy type of a manifold exactly when that manifold fails to be formal. This is an answer to our second question above.

As to the first question, the seminal work of Sullivan [190] and Barge [14] shows that, up to "finite ambiguities," the diffeomorphism type of a simply connected smooth compact manifold is determined by the rational homotopy type of its de Rham complex.* Moreover, given a choice of Riemannian metric on the manifold, this homotopy type can be presented canonically by a finitely generated "minimal model" CDGA over $\mathbb{Q}$. We give a more detailed exposition of Sullivan's methods and results later in the paper. For now, the reader should bear in mind that the de Rham complex is, in some appropriate sense, "almost" a perfect algebraic model like a triangulation.

One goal of this paper is to understand how these results relate to physics. To make contact with Sullivan's work, we need a physical context in which the algebraic structure of the de Rham complex, or at least of the cohomology, arises. Supersymmetric quantum mechanics, by itself, does not provide such a context: the wedge product differential forms is just a variant of multiplying two wavefunctions together, which is not *a priori* meaningful in quantum mechanics. Algebraic structures do occur in physics, but they are more naturally

*To be precise, we here mean the $\mathbb{Q}$-polynomial variant of the de Rham complex, as defined by Sullivan.

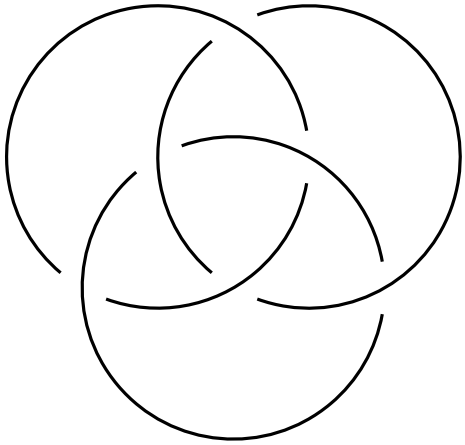


Figure 2.2: The Borromean rings.

connected to operators. As such, one needs a physical setting in which there is a state-operator correspondence.

Two-dimensional conformal field theory is the simplest such setting, and two-dimensional superconformal sigma models are well studied. In order for a 2d sigma model to admit a topological twist, it must have enhanced supersymmetry; this occurs precisely when the target space is Kähler.

Kähler manifolds also play a special role in work of Sullivan and collaborators on rational homotopy theory; the paper [58] proves that every compact Kähler manifold is formal. Moreover, this fact follows from a simple identity relating the differential operators ∂ and $\bar{\partial}$ on the de Rham complex, called the dd^{c}-lemma. The same identity is responsible for supersymmetry enhancement in Kähler sigma models; a version of the dd^{c}-lemma can be proved for quantum mechanical systems obtained from dimensional reduction from field theories with enhanced supersymmetry. We will expand on this point later.

There are strong constraints on the topology of Kähler manifolds. Formality is one of these, but much more can be said. To what extent enhanced supersymmetry imposes analogous constraints on two-dimensional $\mathscr{N} = (2,2)$ SCFTs has been studied by many authors, notably [79, 140]. A second aim of this paper is to revisit and clarify the analogies between Kähler geometry and $\mathscr{N} = (2,2)$ SCFT. For us, the key ingredient is always supersymmetric quantum mechanics: we study the $0+1$-dimensional theories obtained by dimensional reduction from supersymmetric field theories in $d \geq 2$. In theories

that are sigma models, this supersymmetric quantum mechanics is precisely the de Rham complex of the target space, or (as in the case of the B-twist of the two-dimensional sigma model) some variant thereof. Furthermore, extra structures arise when a d-dimensional theory arises as the dimensional reduction of a d'-dimensional theory, $d' > d$; this was considered by [79]. For us, $d = 1, 2, 4, 6$; as the dimension grows, the number of supercharges (and also the number of bosonic symmetries) increases accordingly. A table outlining these hierarchies can be found later in the paper (Fig. 2.7).

To reiterate: we review the well-known hierarchies of increasingly rich structures that appear in the following four contexts:

- target-space geometry that is generic, Kähler, and hyper-Kähler;[†]
- supersymmetric quantum mechanics, with one, two, and four supercharges;
- two-dimensional superconformal field theories, with $\mathscr{N} = (1,1)$, $(2,2)$, and $(4,4)$;
- minimally supersymmetric field theories in dimensions two, four, and six.

We explore similarities and differences between the structures that emerge in each case. As the structure becomes richer, the differences between the various categories become fewer and fewer.

We hope that this paper will also serve a pedagogical purpose, being of some use to those who wish to learn the well-studied subjects we review. In studying the literature on supersymmetric sigma models and Kähler geometry, certain unifying themes became apparent to us, which we felt were not adequately spelled out in existing references. We hope that at least some of this thematic unity comes across in our treatment.

Throughout, we are motivated by the following series of increasingly speculative ideas and questions: The de Rham complex is an algebraic model of a space, which naturally arises in physics as the Hilbert space of supersymmetric quantum mechanics. Physics also offers examples of other kinds of algebraic

[†]Here "Calabi-Yau" should appear strictly in between "Kähler" and "hyper-Kähler."

models for spaces: whenever a sigma model (in any dimension and with any amount of supersymmetry) can be defined on a target space M, its spaces of states or operators constitute an algebraic model of M. This algebraic model should always reduce to some variant of the de Rham complex upon dimensional reduction.

In the case of the two-dimensional $\mathscr{N} = (2, 2)$ sigma models defined when M is a Calabi-Yau threefold, this algebraic model does not retain perfect information about M or even about its cohomology. However, this is far from a failure of these ideas; the ambiguity in recovering M from this algebraic model leads precisely to the phenomenon of mirror symmetry [186]!

Given an algebraic model of some kind, it is natural to ask for sufficient conditions that ensure that it is the model of some space. One can get some feel for this by enumerating all possible identifiable structures that are present on algebraic models of honest spaces; these are then necessary conditions to be the model of a space. In physics, one often speaks about this problem in the language of "geometric" and "non-geometric" theories.

The Sullivan–Barge theorem says that, with appropriate conditions, the list of conditions and structures that are *necessary* for a CDGA to algebraically model an honest space are also *sufficient*. As such, a motivating problem for the line of work we have pursued would be to understand the algebraic structures corresponding to (some class of) physical quantum field theories, to develop the homotopy theory of such structures,[‡] and then formulate sufficient conditions for such a theory to be describable as a sigma model. If one could do this rigorously and give a statement, analogous to Sullivan–Barge, allowing one to reconstruct the target space from its algebraic model, one would almost surely have a clear understanding of the phenomenon of mirror symmetry. While we are of course far from doing any of this in this modest paper (and others [227] have previously and more expertly drawn connections between mirror symmetry and rational homotopy theory), it is our hope that the ideas we sketch here will prove useful for others to think about and eventually bear fruit of this kind.

[‡]A "homotopy between two physical theories" should be understood as a one-parameter family of theories interpolating between the two; that is to say, a path from one to the other in the appropriate moduli space.

2.1 Supersymmetry algebras and their representations in one space-time dimension

The extended supersymmetry algebra in $0+1$ dimensions [5] is as follows:

$$\{Q_i, Q_j^\dagger\} = 2\delta_{ij}H, \quad \{Q_i, Q_j\} = 0. \tag{2.3}$$

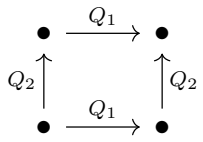
From these basic commutation relations, it follows that all supercharges commute with H (the Hamiltonian of the theory). The indices range from 1 to $\mathscr{N}$ for $\mathscr{N}$-extended supersymmetry; although there is no restriction on $\mathscr{N}$ in principle, in cases relevant to either physics or geometry, $\mathscr{N}$ is usually a small power of two.

This algebra is usually supplanted with either an operator F, defining fermion number or homological degree and taken to have integer eigenvalues, or $(-)^F$, which defines fermion number modulo two. In either case the supercharges Q_i should carry one unit of fermion number. In the former case, a representation of the algebra with $\mathscr{N} = 1$ becomes a chain complex of Hilbert spaces; in the latter, it is a $\mathbb{Z}/2\mathbb{Z}$-graded chain complex. The reader will no doubt have noticed that the $\mathscr{N} = 1$ algebra is just what we called the de Rham algebra in the introduction.

We have already mentioned the familiar classification of irreducible representations of this algebra. To remind the reader, they can be labeled with a single nonnegative number, the eigenvalue of H (as well as an integer labeling the degree). When this number (the energy) is positive, the representation is "long" and consists of two generators of adjacent degree, which are mapped to one another by the supercharges; when it is zero, the representation is one-dimensional ("short"). The key picture to keep in mind is Fig. 2.1. The reader will have no trouble drawing the analogous picture of a $\mathbb{Z}/2\mathbb{Z}$-graded complex.

Thanks to formal Hodge theory, this classification persists for representations of the extended supersymmetry algebra. The bosonic operators consist of H together with whatever degree operators $F_1, \ldots, F_{\mathscr{N}}$ are defined. If we insist that the F_i commute, representations will be joint eigenspaces for all of these, and can be labeled by their energy together with their multidegree. Short representations will still be one-dimensional; long representations will now have dimension $2^{\mathscr{N}}$, and they will consist of generators at each corner of an $\mathscr{N}$-dimensional cube in the degree space. The cohomology with respect to any supercharge Q_i is the same; it counts the short representations, and cohomol-

ogy classes therefore carry a well-defined multidegree. A long representation of $\mathscr{N} = 2$ is shown in the (F_1, F_2)-plane below:



(2.4)

Let us focus on the case $\mathscr{N} = 2$, which will be important in what follows. A crucial observation is that any representation of this algebra actually admits a $\mathbb{C}\mathrm{P}^1$ family of *different* actions (if we relax some requirements about the existence of degree operators). Define the parameterized supercharge

$$Q_t = t_1 Q_1 + t_2 Q_2, \tag{2.5}$$

where the parameter $t \in \mathbb{C}\mathrm{P}^1$, and we have chosen a representative such that $|t_1|^2 + |t_2|^2 = 1$. Then it is an easy calculation to show that

$$\{Q_t, Q_t^\dagger\} = 2\left(|t_1|^2 + |t_2|^2\right) H = 2H.$$

Further, let $\tilde{t} = [-\bar{t}_2 : \bar{t}_1]$. Then $Q_{\tilde{t}}$ is another supercharge, and

$$\{Q_t, Q_{\tilde{t}}\} = \{Q_t, Q_{\tilde{t}}{}^\dagger\} = 0.$$

These two supercharges therefore define another $\mathscr{N} = 2$ algebra. Further, they each carry total fermion number $F_1 + F_2 = 1$. However, they are no longer eigenstates of $F_1 - F_2$; conjugation with $F_1 - F_2$ acts on the $\mathbb{C}\mathrm{P}^1$ parameter space by the rule

$$[t_1 : t_2] \mapsto [t_1 : -t_2],$$

and the two fixed points correspond to our two original supercharges. Q_t cohomology is the same for all values of t; since it describes the zero-energy spectrum of the same Hamiltonian, there is no way it can change. (For physicist readers, we are just describing the consequences of SU(2) R-symmetry.)

There are two variants of this algebra that emerge naturally in geometry and physics, which we will now describe. These are in some sense intermediate between $\mathscr{N} = 1$ and $\mathscr{N} = 2$ supersymmetry. The first is what we will call the $\mathscr{N} = 1^2$ algebra: it consists of two mutually commuting copies of the de Rham algebra, and so is equivalent to an $\mathscr{N} = 2$ algebra in which the condition that the two supercharges square to the *same* Hamiltonian has been relaxed. In equations,

$$\{Q_i, Q_j^\dagger\} = 2\delta_{ij} H_i, \quad \{Q_i, Q_j\} = 0. \tag{2.6}$$

Using the Jacobi identity, it is easy to prove that H_1 and H_2 must commute. This algebra acts, for example, on the Ramond sector of any $\mathscr{N} = 2$ superconformal field theory in two dimensions. The relevant closed subalgebra of the Ramond algebra is:

$$\begin{aligned} (G_0^{\pm})^2 &= 0 \\ (\bar{G}_0^{\pm})^2 &= 0 \\ \{G_0^{\pm}, \bar{G}_0^{\pm}\} &= 0 \\ \{G_0^{+}, G_0^{-}\} &= L_0 - c/24 \\ \{\bar{G}_0^{+}, \bar{G}_0^{-}\} &= \bar{L}_0 - c/24. \end{aligned} \tag{2.7}$$

Clearly, representations are now labeled by *two* energies, E_1 and E_2. The irreducible representations are of four types, according to whether or not each E_i is zero; their dimensions are four, two, two, and one.

The construction we gave above still goes through and produces a $\mathbb{C}\mathrm{P}^1$ family of $\mathscr{N} = 1$ algebras with supercharges Q_t. However, the $\mathbb{C}\mathrm{P}^1$ family of Hamiltonians is now nontrivial:

$$\{Q_t, Q_t^{\dagger}\} = 2H_t = 2\left(|t_1|^2 H_1 + |t_2|^2 H_2\right). \tag{2.8}$$

The zero-energy spectrum of H_t is the same for almost all t, consisting of states for which E_1 and E_2 are both zero. (These are the genuinely short one-dimensional representations.) However, there are two special points in the $\mathbb{C}\mathrm{P}^1$ moduli space that exhibit *enhanced* vacuum degeneracy: the points $[0:1]$ and $[1:0]$, corresponding to our two original supercharges. The cohomology of Q_t thus jumps in total rank at these points.

One should note that these jumps do not occur unless there are states for which one Laplacian is zero but the other is not. The condition that the spaces of vacuum states for the two Laplacians agree is strictly weaker than the condition that the two operators are identical. Nonetheless, by formal Hodge theory, it is enough to ensure that Q_t-cohomology is the same for all t, and therefore that only square and singlet representations occur. As such, it is sufficient to establish the dd^{c}-lemma.

In quantum mechanics, one should expect enhanced degeneracy to be an avatar of enhanced symmetry. The obvious question is: what symmetry is enhanced at the points of our $\mathbb{C}\mathrm{P}^1$ moduli space that exhibit extra BPS states?

The answer is that precisely at these points $\mathscr{N} = 1$ supersymmetry is enhanced to $\mathscr{N} = 1^2$. It is easy to check that, when we try to define a second supercharge $Q_{\tilde{t}}$ by our prescription above, something goes wrong: namely,

$$\{Q_t, Q_{\tilde{t}}^\dagger\} = t_1 t_2 \left(H_2 - H_1\right) \neq 0.$$

As such, when t is generic, no supercharges other than Q_t and its adjoint can be found that act in a compatible way.

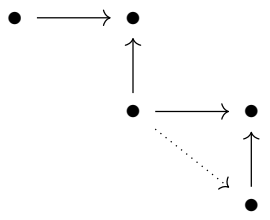
This leads us to the second interesting variant, which we will refer to as $\mathscr{N} = 1.5$. The notation is meant to convey not only intermediacy between $\mathscr{N} = 1$ and $\mathscr{N} = 2$, but also that something is "not whole": the commutation relations in this case no longer define a closed algebra. In this case, the following commutation relations are imposed:

$$\{Q_i, Q_i^\dagger\} = 2H_i, \quad \{Q_i, Q_j\} = 0. \tag{2.9}$$

However, we do not require that $S = \{Q_1, Q_2^\dagger\}$ must vanish. It is simply a new bosonic operator that can be defined, about which nothing can *a priori* be said.

On a generic complex manifold, this is the algebra that holds between the derivatives ∂ and $\bar{\partial}$. Moreover, in a generic double complex, this is the algebra satisfied by the two differentials. We will return to this point when reviewing results from complex geometry in the next section.

New, exotic irreducible representations of this algebra are possible. In addition to the standard long and short representations of $\mathscr{N} = 2$ ("squares" and "dots"), one can also have "staircases" of the sort depicted below:[§]



(2.10)

As in (2.4), the plane is the (F_1, F_2)-plane, and the arrows represent the action of the supercharges. We will refer to the total dimension of such a representation as the "length" of the staircase, which is five for the staircase pictured

[§]We are grateful to David Speyer for helpful comments at http://mathoverflow.net/questions/86947/ that pointed these facts out to us.

above; any length can occur. The dashed line indicates the operator S defined above, whose role is to go up and down stairs.

Staircases of odd length, like the one pictured in (2.10), contribute one generator to both Q_1-cohomology and Q_2-cohomology, as well as to the "total" cohomology of the supercharge $Q_1 + Q_2$. However, as can easily be seen from the picture, the generator is in a different $(F_1 - F_2)$-degree. We can no longer label cohomology classes with their "axial" quantum numbers.

Furthermore, staircases of even length (depending on their orientation) contribute *two* generators to one of the Q_i-cohomologies, but *none* to the other. Drawing the appropriate picture will make this clear. These ideas should be familiar to anyone who is familiar with the spectral sequences associated to a double complex: they begin at Q_1- or Q_2-cohomology, converge to $(Q_1 + Q_2)$-cohomology, and the differentials on the k-th page cancel pairs of generators that lie at opposite ends of a staircase of length $2k$.

As always, the two anticommuting supercharges allow one to define a $\mathbb{C}\mathrm{P}^1$ family of de Rham algebras. However, unlike in the $\mathscr{N} = 1^2$ case, the Hamiltonians corresponding to different points in this moduli space can never be simultaneously diagonalized. The spectral sequence is a formal way of describing how different representations appear and disappear in the zero-energy spectrum of this parameterized family of Hamiltonians. Further remarks on the physics of spectral sequences and supersymmetric quantum mechanics will appear in [109].

2.2 Review of Kähler, Calabi-Yau, and hyper-Kähler geometry

We now give a brief review of some classical facts about Kähler geometry, in a way that is tailored to our purposes. For more details, the reader is referred to the excellent exposition in [58], or to the book [121]. We will return to the formality result of [58] when reviewing Sullivan's theory of minimal models in §2.3. Calabi-Yau and hyper-Kähler geometry provide levels of additional structure; we address these in order of decreasing generality. Readers should keep in mind that, in the context of two-dimensional superconformal sigma models, Calabi-Yau structure is required to define an $\mathscr{N} = (2,2)$ SCFT. A hyper-Kähler target is necessary and sufficient for $\mathscr{N} = (4,4)$ superconformal symmetry.

Let M be a smooth manifold of real dimension $2n$. An almost-complex struc-

ture on M is a vector-bundle morphism $J\colon TM \to TM$ such that $J^2 = -1$. On the space of complex-valued differential forms, J will induce a decomposition

$$\Omega^n(M, \mathbb{C}) = \bigoplus_{p+q=n} \Omega^{p,q}(M), \tag{2.11}$$

which comes from the decomposition of the complexified tangent bundle into the $\pm i$ eigenspaces of J. This decomposition allows us to write the exterior derivative operator d as a sum of terms of different degree:

$$d = \sum_{r+s=1} d^{r,s}. \tag{2.12}$$

The almost-complex structure is said to be integrable when M is a complex manifold: it admits an atlas of complex-valued coordinates with holomorphic transition maps, such that multiplication by i agrees with the almost-complex structure J. This occurs precisely when only terms of degree $(1,0)$ and $(0,1)$ are present in the decomposition of d. (An equivalent condition is that $d^{0,1}$ square to zero.) In this case, we use the symbols ∂ and $\bar{\partial}$ for the operators $d^{1,0}$ and $d^{0,1}$; it is easy to check that these operators separately square to zero and anticommute with one another. A complex manifold is a space locally modeled on $\mathbb{C}^n$; one is able to define which functions are holomorphic.

The above shows immediately that the de Rham complex of a complex manifold admits an action of $\mathscr{N} = 1.5$ supersymmetry. The de Rham algebra (or $\mathscr{N} = 1$ supersymmetry algebra for quantum mechanics) acts in three meaningfully different ways:[¶] each of ∂, $\bar{\partial}$, and $d = \partial + \bar{\partial}$ may be thought of as a nilpotent supercharge.

However, on a garden-variety complex manifold, supersymmetry is *not* enhanced to $\mathscr{N} = 2$. While ∂ anticommutes with $\bar{\partial}$, it may not commute with its adjoint: $\{\partial, \bar{\partial}^\dagger\} \neq 0$. When this is the case (as we discussed in a general setting above) the respective Laplacians will not agree and will fail to commute with one another. Precisely when the manifold is Kähler, its de Rham complex provides an $\mathscr{N} = 2$ supersymmetric quantum mechanics. We will return to this point after giving the formal definition of Kähler structure.

Suppose that M is a complex manifold. A symplectic structure on M is a choice of closed, non-degenerate two-form $\omega \in \Omega^2(M, \mathbb{R})$. One says that

[¶]As we emphasized in the previous section, a $\mathbb{C}\mathrm{P}^1$ family of de Rham algebras can be defined. The three actions we mention here are three representative points in this moduli space.

the symplectic structure ω is *compatible* with the complex structure if the composite tensor

$$g(a, b) = \omega(a, Jb) \tag{2.13}$$

is a Riemannian metric on M. Symmetry of the metric tensor then implies that the form ω has (p, q)-degree $(1, 1)$. A complex manifold admitting such a compatible triple of structures is called *Kähler*.

None of the conditions in the definition of a Kähler manifold can be relaxed; counterexamples exist in all cases. For example, there are complex manifolds that admit symplectic forms, but nonetheless are not Kähler (the compatibility condition cannot be satisfied). Despite these rigid requirements, though, many manifolds are naturally Kähler. All of the complex projective spaces $\mathbb{C}\mathrm{P}^n$ are, when equipped with the Fubini–Study metric. Furthermore, a complex submanifold of a Kähler manifold is again Kähler; as such, so are all smooth projective algebraic varieties over $\mathbb{C}$.

Just as the integrability condition on an almost-complex structure may be phrased in terms of differential operators as the identity $\bar{\partial}^2 = 0$, so the conditions for M to be Kähler can be expressed by a set of *Kähler identities* between operators in its de Rham complex.

Define an operator $\Lambda^\dagger$ by the relation

$$\Lambda^\dagger \colon \alpha \mapsto \omega \wedge \alpha,$$

where ω is the Kähler form, and let $\Lambda = \star\Lambda^\dagger\star$ be its adjoint. This pair of operators are called *Lefschetz operators;* they have pure degree $(1, 1)$ and $(-1, -1)$ respectively. Their commutator $D = [\Lambda^\dagger, \Lambda]$ is a degree operator: it acts diagonally according to the rule

$$D|_{\Omega^{p,q}} = (p + q - n).$$

To check that this is true, one can simply do a linear algebra calculation corresponding to one of the fibers of the bundle. For a Hermitian vector space that is one-dimensional over $\mathbb{C}$, the result is easy to see; the exterior algebra $\Lambda^* V$ is concentrated in degrees 0, 1, and 2, so that $[\Lambda^\dagger, \Lambda]$ must be zero in the middle dimension. The general statement is then obtained by induction on the complex dimension of V; for details, the reader is referred to [121].

It follows that these three operators together form a basis for the Lie algebra $\mathrm{SL}(2,\mathbb{R})$: they satisfy the commutation relations

$$[D,\Lambda^\dagger] = 2\Lambda^\dagger, \quad [D,\Lambda] = -2\Lambda, \quad [\Lambda^\dagger,\Lambda] = D.$$

(Since the Kähler form is real, these operators actually act on the real cohomology and not merely on cohomology with complex coefficients; however, the latter is the relevant case for quantum mechanics.) The Kähler identities imply that Kähler manifolds admit a unique choice of Laplacian, that their de Rham complexes are examples of $\mathscr{N} = 2$ supersymmetric quantum mechanics, and that the $\mathrm{SL}(2,\mathbb{R})$ action defined above commutes with the Laplacian and hence descends to the cohomology. The essential identities, from which all others can be derived, are

$$\begin{aligned} [\Lambda,\bar\partial] = -i\partial^\dagger, \quad [\Lambda,\partial] = i\bar\partial^\dagger, \\ [\Lambda,\partial^\dagger] = [\Lambda,\bar\partial^\dagger] = 0. \end{aligned} \tag{2.14}$$

To establish these, we refer to [58], who point out that Kähler metrics osculate to second order to the flat metric on $\mathbb{C}^n$. This implies that any identity which is at most first-order in derivatives of the metric can be established by checking it on $\mathbb{C}^n$. The calculation in the case of (2.14) is simple.

Given these identities, it is straightforward to show that supersymmetry is enhanced:

$$\begin{aligned} \{\partial,\bar\partial^\dagger\} &= i\{\partial,[\partial,\Lambda]\} \\ &= i\left(\partial\partial\Lambda - \partial\Lambda\partial + \partial\Lambda\partial - \Lambda\partial\partial\right) \\ &= 0 \quad (\text{since } \partial^2 = 0). \end{aligned} \tag{2.15}$$

Therefore, two mutually commuting copies of the $\mathscr{N} = 1$ algebra act. It remains to show that their respective Laplacians agree. This is again a quick calculation using (2.14), the Jacobi identity, and $\{\partial,\bar\partial\} = 0$:

$$\begin{aligned} \{\partial,\partial^\dagger\} &= i\{\partial,[\Lambda,\bar\partial]\} \\ &= i\left([\Lambda,\{\partial,\bar\partial\}] + \{\bar\partial,[\partial,\Lambda]\}\right) \\ &= \{\bar\partial, i[\partial,\Lambda]\} \\ &= \{\bar\partial,\bar\partial^\dagger\}. \end{aligned} \tag{2.16}$$

As a consequence of (2.16), there is a unique harmonic representative of each cohomology class, with pure (p,q)-degree. Recall that the representations of

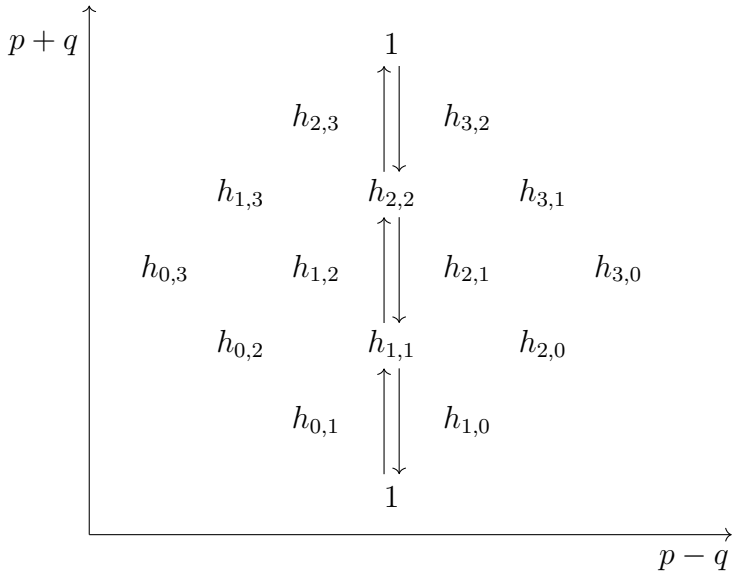


Figure 2.3: The Hodge diamond for a generic connected Kähler threefold. The raising and lowering operators, indicated by vertical arrows for $p-q=0$, are the Lefschetz operators. Many of the indicated Hodge numbers are not independent, being related by discrete symmetries.

$\mathscr{N}=2$ supersymmetry are of two types: zero-energy, one-dimensional "short" representations, and positive-energy, four-dimensional representations, which form "squares" in degrees (p,q), $(p+1,q)$, $(p,q+1)$, and $(p+1,q+1)$. The Betti numbers of the manifold can therefore be refined into Hodge numbers,

$$b_n = \sum_{p+q=n} h_{p,q}, \tag{2.17}$$

which count the short multiplets. (We emphasize once again that this is not true on a generic complex manifold! Eq. (2.17) is a consequence of the degeneration of the Hodge-to-de-Rham spectral sequence, which is in turn a consequence of the enhanced supersymmetry guaranteed by the Kähler identities.)

One usually arranges the Hodge numbers of M in a diamond, according to their bidegrees, as shown in Fig. 2.3.

The Kähler identities (2.14) further imply that the Lefschetz operators commute with the Laplacian; checking this is easy, and proceeds using the Jacobi identity. The action of $\mathrm{SL}(2,\mathbb{R})$ they provide therefore maps harmonic forms to harmonic forms, and so descends to the cohomology.

As such, a nice way to think of the Hodge diamond of a Kähler manifold is as a weight diagram for a representation of the rank-two Lie algebra $\mathrm{SU}(2) \times \mathrm{U}(1)$. The algebra acts via the operators $\Lambda^\dagger$, Λ, $D = P + Q - n$, and $P - Q$. The Hodge numbers are then arranged by their weights with respect to the Cartan of this algebra. This picture will generalize nicely to the hyper-Kähler case.

The Lefschetz operators have extra consequences for the topology of M, in addition to those that follow from enhanced supersymmetry. Since they commute with $P-Q$, it follows immediately that each vertical slice of the Hodge diamond (when $(p-q)$-degree is displayed horizontally) is a weight diagram for $\mathrm{SL}(2,\mathbb{R})$. As a consequence, the odd- and even-degree Betti numbers of a Kähler manifold are separately monotonically nondecreasing toward the middle degree. This is the content of the hard Lefschetz theorem, which is straightforward to understand from the standpoint of representation theory.

While $\mathscr{N} = (2,2)$ superconformal symmetry in two dimensions is enough to imply things like bigrading and Poincaré duality—we return to these points later—it is not sufficient to establish properties like hard Lefschetz. Indeed, $\mathscr{N} = (2,2)$ theories may not even have operators with the appropriate quantum numbers to correspond to a Kähler class—superconformal minimal models are examples of this.

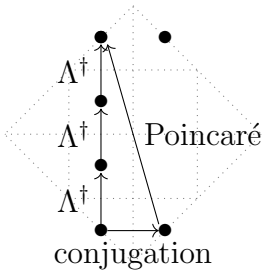
Since the volume form of a compact Kähler manifold has bidegree (n,n), Poincaré duality defines a pairing between $\mathrm{H}^{p,q}$ and $\mathrm{H}^{n-p,n-q}$, implying immediately that the corresponding Hodge numbers are equal. Poincaré duality acts on the Hodge diamond (Fig. 2.3) by reflection through the center point, or equivalently by simultaneously flipping the sign of $P - Q$ and D.

An additional symmetry of the Hodge diamond comes from considering the action of complex conjugation on the de Rham complex. This takes forms of degree (p,q) to forms of degree (q,p), while preserving total degree. It therefore acts on the Hodge diamond by reflecting left and right, flipping the sign of $P - Q$ while fixing D.

From the representation-theory standpoint, we are considering unitary representations of $\mathrm{SU}(2) \times \mathrm{U}(1)$. Every representation of $\mathrm{SU}(2)$ is symmetric with respect to the Weyl group $\mathbb{Z}/2\mathbb{Z}$, which acts on the Hodge diamond by reflections about the horizontal axis. Furthermore, for unitary representations of $\mathrm{U}(1)$, we also insist on charge-conjugation symmetry, which reflects the Hodge

diamond about the vertical axis. We therefore recover the same $(\mathbb{Z}/2\mathbb{Z})^2$ symmetry that is generated by Poincaré duality and complex conjugation.

Lastly, the so-called Hodge-Riemann bilinear relation is a compatibility condition between the Lefschetz action and these discrete symmetries: it states that, for highest-weight states, the Poincaré dual of the complex conjugate is the same (up to a complex scalar phase) as the state obtained by applying the SU(2) raising operator multiple times. In pictures, the following should commute up to a scalar:



The equation that expresses this is

$$\star\,(\Lambda^\dagger)^j\psi = \frac{(-1)^{k(k+1)/2}j!}{(n-k-j)!}(\Lambda^\dagger)^{n-k-j}\bar{\psi}, \tag{2.18}$$

where ψ is a k-form such that $\Lambda\psi = 0$. See [121] for more details, as well as a proof of the Kähler identities that starts from this formula.

A Kähler manifold M is *Calabi-Yau* if it has a nowhere-vanishing holomorphic—and therefore harmonic—$(n, 0)$-form. (An equivalent condition is that the canonical bundle be trivial.) Therefore, a Calabi-Yau manifold will have $h_{n,0} = h_{0,n} = 1$.

There are strong constraints on the fundamental groups of Kähler and Calabi-Yau manifolds as well. Just to substantiate this, we recall that any Calabi-Yau M has a finite cover of the form $T \times N$, where n is a simply connected Calabi-Yau and T is a complex torus. This fact is called *Bogomolov decomposition* [24]. In this paper, however, we will always restrict ourselves to the simply connected case.

The next level of possible structure is provided by *hyper-Kähler manifolds*, which are Riemannian manifolds equipped with three distinct complex structures. We will call these I, J, and K. They should satisfy the multiplication table of the quaternions: for instance, $IJ = -JI = K$. Further, when

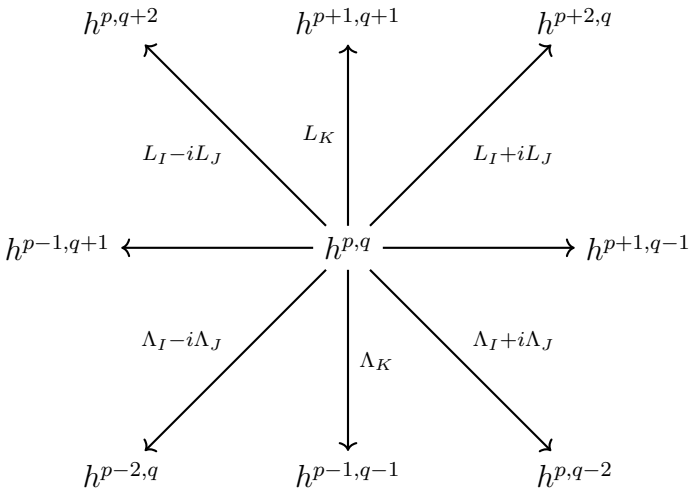


Figure 2.4: The action of the B_2 root system on the Hodge diamond of a hyper-Kähler manifold.

equipped with any of these complex structures, the manifold should be Kähler. The tangent spaces to a hyper-Kähler manifold are thus quaternionic vector spaces; while this is necessary, it is not sufficient. (Again, it is possible for compatibility conditions between the complex structures and the metric to fail; so-called quaternionic Kähler manifolds are examples of this [57].)

Hyper-Kähler manifolds are highly structured, and few compact examples are known: the hyper-Kähler structure imposes strong constraints on the topology. We review some of these constraints briefly here.

The three Kähler classes corresponding to the three complex structures define operators on the de Rham complex which close into the action of a Lie algebra, analogous to the Lefschetz SL(2) action in the Kähler case. The algebra that applies in this case is SO(5); it was first constructed by Verbitsky [197]. For the reader's convenience, we explain his construction here. We use the notation

$$L_I = \Lambda_I^\dagger \colon \alpha \mapsto \omega_I \wedge \alpha$$

(and its analogues) for the Lefschetz-type operators. (One should be careful here; Λ_I is the adjoint of L_I with respect to complex structure I. However, if one works with respect to a fixed complex structure, not all pairs of Lefschetz operators will be adjoints; this is why we prefer the notation L for raising operators in the sequel.)

Following Verbitsky, let's define

$$M_{IJ} = [L_I, \Lambda_J],$$

where the indices range over pairs of complex structures. Clearly,

$$M_{II} = M_{JJ} = M_{KK} = D.$$

Moreover, $[L_I, L_J] = [\Lambda_I, \Lambda_J] = 0$ for any pair of indices I and J, since two-forms commute. The M operators carry homological degree zero, and therefore commute with D.

One identity requires a nontrivial argument: $M_{IJ} = -M_{JI}$. Just as in the Kähler case, commutation relations like this one are really statements about linear algebra, and can be proved by an explicit calculation in the case of a one-dimensional quaternionic vector space.

Once this is established, the remaining commutation relations can be fixed by some quick calculations with the Jacobi identity:

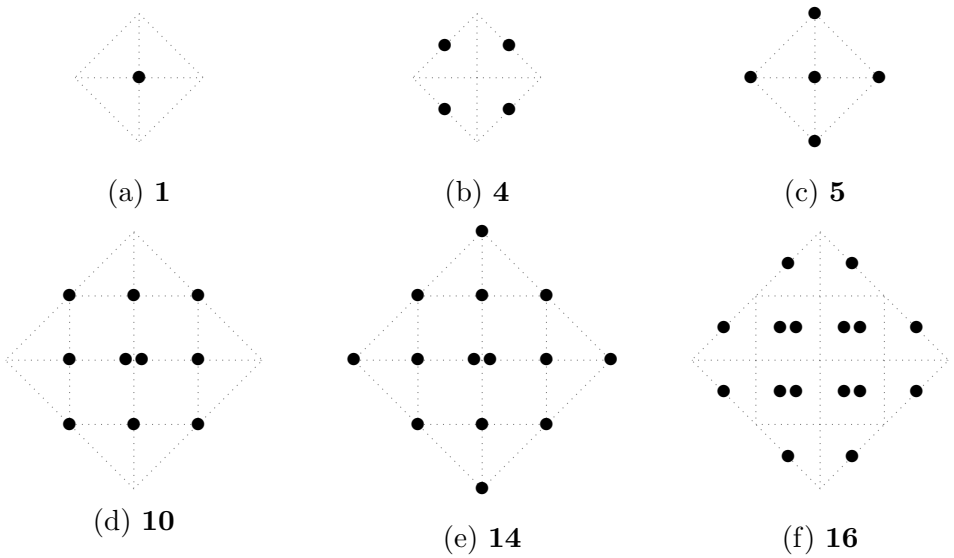
$$\begin{gathered} [M_{IJ}, \Lambda_I] = -2\Lambda_J, \quad [M_{IJ}, L_J] = 2L_I; \\ [M_{IJ}, \Lambda_K] = [M_{IJ}, L_K = 0]; \\ [M_{IJ}, M_{JK}] = 2M_{IK}. \end{gathered} \tag{2.19}$$

These operators therefore close into a ten-dimensional Lie algebra of rank two. Let's choose to fix the complex structure K on the manifold, and the corresponding Cartan subalgebra spanned by D and $-iM_{IJ}$. It is then straightforward to show that the weights of the adjoint representation form the root system B_2, corresponding to the algebra SO(5). Moreover, the bigrading defined by weights for this choice of Cartan on the de Rham complex coincides with the Hodge bigrading by homological degree and $(P - Q)$, defined with respect to complex structure K. That is,

$$-iM_{IJ} = (P - Q).$$

(See Fig. 2.4.)

The picture of the Hodge diamond as a weight diagram for a Lie algebra representation furnished by the cohomology therefore generalizes beautifully to the hyper-Kähler case. The Lie algebra in question is now SO(5), rather than SU(2) × U(1), and the restrictions imposed on the Hodge numbers by



(a) **1** (b) **4** (c) **5**

(d) **10** (e) **14** (f) **16**

Figure 2.5: SO(5) irreps that can fit inside an 8-dimensional Hodge diamond. The first three can also fit inside a 4-dimensional Hodge diamond. The **16** representation cannot occur for simply connected eight-manifolds.

representation theory are accordingly more severe. For example, only three irreducible representations of SO(5) can fit inside the Hodge diamond of a hyper-Kähler four-manifold, and only six inside that of an eight-manifold. (See Fig. 2.5.) The restrictions coming from the SO(5) action are best thought of as inequalities, rather than equalities, between Hodge numbers: they are the analogue of the monotonicity properties guaranteed for Betti numbers of Kähler manifolds by the hard Lefschetz theorem.

The discrete symmetries of the Hodge diamond are also enhanced in the hyper-Kähler case. The Weyl group of B_2 is $(\mathbb{Z}/2\mathbb{Z})^3$; since this group has order eight, Hodge numbers are repeated up to eight times. The new identity is

$$h^{p,q} = h^{p,2n-q} \quad (n = \dim_{\mathbb{C}} M).$$

Moreover, the de Rham complex on a hyper-Kähler manifold has differential operators generating an action of the $\mathscr{N} = 4$ supersymmetry algebra. Recall that, on a Kähler manifold, the real operators were the exterior derivative d and its partner $d^c = -JdJ$. On a hyper-Kähler manifold, we can write the four operators

$$d_1 = d, \quad d_I = -IdI, \quad d_J = -JdJ, \quad d_K = -KdK.$$

The commutation relations for these operators and their adjoints follow trivially from those for d and d^c, since for any pair of them one can be obtained from the other by conjugating with a specific complex structure.

To emphasize the perspective we have tried to bring out in this review: the requirements on the target space of supersymmetric sigma models, which are usually understood in terms of special holonomy, can also fruitfully be thought about in terms of dimensional reduction to supersymmetric quantum mechanics. Since dimensional reduction preserves the number of real supercharges, a minimally supersymmetric d-dimensional sigma model can be defined on a target manifold M if and only if the appropriate number of supercharges ($2^{\lfloor d/2 \rfloor}$) act on the corresponding supersymmetric quantum mechanics (furnished by the de Rham complex of M). Although this condition is obviously necessary, it is perhaps surprising that it is sufficient. Furthermore, as was first pointed out by [79], the perspective of dimensional reduction offers a natural explanation of Lefschetz-type symmetry algebras, just as it does for R-symmetries in general supersymmetric theories. In fact, one should think of Lefschetz symmetries in geometry and R-symmetries in supersymmetry as the same kind of object. We return to this in §2.4.

2.3 Homotopy theory over a field and Sullivan minimal models

In this section, we introduce readers to some rudiments of rational homotopy theory, as developed by Sullivan and collaborators, which formalizes the idea of the de Rham complex as an algebraic model of a manifold. We closely follow the exposition in [58]; other more exhaustive references are [103, 190]. We attempt to use quantum-mechanical language when possible, and offer a reinterpretation of Sullivan's minimal models in physical language, which sheds intuitive light on the origin of Massey products.

The de Rham complex is a commutative differential graded algebra (briefly, a CDGA). That is, it is a graded algebra $\Omega = \bigoplus_{i \geq 0} \Omega^i$ over a ground field k of characteristic zero, whose multiplication preserves the grading, equipped with a differential of degree one that is a derivation for the product:

$$d(x \cdot y) = dx \cdot y + (-)^{|x|} x \cdot dy. \tag{2.20}$$

Here x and y are assumed to have homogeneous degree, and $|x| = \deg x$. The product is further taken to be commutative in the graded sense familiar from

supersymmetry:

$$x \cdot y = (-)^{|x||y|} y \cdot x. \tag{2.21}$$

In words, a CDGA is a chain complex equipped with a compatible notion of multiplication.

The cohomology of a CDGA is defined as usual, and is a commutative graded algebra. It is itself a CDGA, if we understand its differential to be the zero map. We will say that Ω is *connected* if $\mathrm{H}^0(\Omega) = k$, and *simply connected* if it is connected and $\mathrm{H}^1(\Omega) = 0$.

Homomorphisms of CDGAs are algebra homomorphisms that are also morphisms of chain complexes. A CDGA homomorphism is a *quasi-isomorphism* if the map it induces on cohomology is an isomorphism. We will consider two CDGAs to be *equivalent* if there is a zigzag of quasi-isomorphisms going from one to the other.

There is a collection of ideals I_j in Ω, defined by

$$I_j = \bigoplus_{i \geq j} \Omega^i \subset \Omega. \tag{2.22}$$

The ideal of *decomposable elements* is $I_1 \cdot I_1 \subseteq I_2$. It consists of elements of Ω that can be written as products of elements of strictly lower degree.

Let M be a simply connected CDGA. We will say that M is *minimal* if (1) as an algebra, it is the free (graded)-commutative algebra generated by a finite number of elements of homogeneous degree; (2) the differential is decomposable, i.e.,

$$\operatorname{im} d \subseteq I_1 \cdot I_1.$$

From these conditions, it follows that $M^0 = k$ and $M^1 = 0$. The notion of a minimal CDGA can be generalized to the non-simply connected case, but for simplicity we will always deal with simply connected spaces and CDGAs in this paper.

Like any CDGA, a minimal CDGA M over $\mathbb{C}$ can be thought of as a supersymmetric quantum mechanics. If we choose an inner product on M (or if one occurs naturally), so that the adjoints of operators can be defined, there is an obvious action of the relevant $\mathscr{N} = 1$ supersymmetry algebra. The requirement that M be freely generated as a commutative algebra just says that it is a Fock space: each state can be built up from the (unique) vacuum state

in degree zero by the free action of a finite number of bosonic and fermionic creation operators. This is what we would expect for the Hilbert space of a free field theory, even in $0+1$ dimensions. Moreover, once states are identified with operators in this way, the algebra structure is just the obvious product of operators. Structures analogous to this one can be seen (for instance) in the description of the elliptic genus in two-dimensional $\mathscr{N} = (2,2)$ theories in terms of a plethystic exponent.

Lastly, the condition that the differential be decomposable simply insists that we include no "irrelevant" creation operators that are Q-exact and so contribute only to the excited spectrum of the theory.

A minimal CDGA is a *minimal model* of Ω if there is a quasi-isomorphism

$$f\colon M \to \Omega.$$

The utility of minimal models comes from a theorem of Sullivan, which shows that every simply connected CDGA has a minimal model that is unique up to isomorphism. Furthermore, two CDGAs are equivalent if and only if their minimal models are isomorphic. As such, minimal models cure the "ambiguity" of the de Rham complex, and furnish a true invariant of a space. It is as though one had a way of choosing a unique triangulation (and therefore a unique simplicial chain complex) representing each topological manifold.

A minimal model of a CDGA can be thought of as the unique "smallest" supersymmetric quantum mechanics that (1) reproduces the correct BPS spectrum and (2) has the Hilbert space of a free theory. One can construct a minimal model degree by degree, adding generators only as necessary to generate or kill new cohomology. We will give examples below.

Given a minimal model M, we can consider its spaces of *indecomposable elements* in each degree:

$$\pi_i \doteq M^i/(I_1 \cdot I_1)^i.$$

These vector spaces (more precisely, their duals) are the *k-de Rham homotopy groups* of M. By another remarkable theorem, if M is the minimal model of the de Rham complex of a space X (with coefficients in k), then the de Rham homotopy groups of M *are* the homotopy groups of X, up to information about torsion:

$$\pi_i(M) \cong \pi_i(X) \otimes k. \tag{2.23}$$

In fact, the minimal model contains all of the information of the homotopy type of X over the field k; if $k = \mathbb{R}$, in the words of [58], M "*is* the real homotopy type" of X. While it does not preserve the complete homotopy type of X (since there is no information about torsion), it still provides a powerful algebraic invariant.

Example 10. The cohomology of the sphere S^n is a simply connected CDGA with zero differential and zero product, with generators in degrees 0 and n only ($n > 1$). If n is odd, this CDGA is already minimal, and so is its own minimal model.

If n is even, the minimal model M must have a free bosonic generator (call it x) in degree n. The generator x must be closed in order to survive to the cohomology. Therefore x^2 is closed; however, since it does *not* survive to cohomology, it must be exact. We are therefore forced to introduce a generator y in degree $2n-1$, such that $dy = x^2$. This kills all higher powers of x ($d(yx^n) = x^{n+2}$), and therefore (since y is fermionic, so that $y^2 = 0$) we have constructed the minimal model.

It is *a priori* not obvious that these minimal models carry any information about spheres as spaces; they are minimal models of the cohomology, rather than of the de Rham complexes, of the spheres. However, it is possible to show that spheres are formal spaces, in the sense discussed in the introduction. Recallthat a minimal CDGA M is *formal* if there is a CDGA quasi-isomorphism

$$f\colon M \to \mathrm{H}^\bullet(M).$$

The algebras we constructed for spheres are therefore formal by definition; to show that spheres are formal spaces requires showing that their de Rham complexes are equivalent to their cohomology.

Once this is done, however, the minimal models we constructed above *are* the minimal models of the n-spheres, and their de Rham homotopy groups are the homotopy groups of spheres, after tensoring with $\mathbb{R}$. This shows that all the higher homotopy of odd-dimensional spheres is torsion, whereas the higher homotopy of even-dimensional spheres has rank one in degree $2n-1$ (corresponding for $n = 2$ to the Hopf fibration). Formality is thus a remarkably powerful tool.

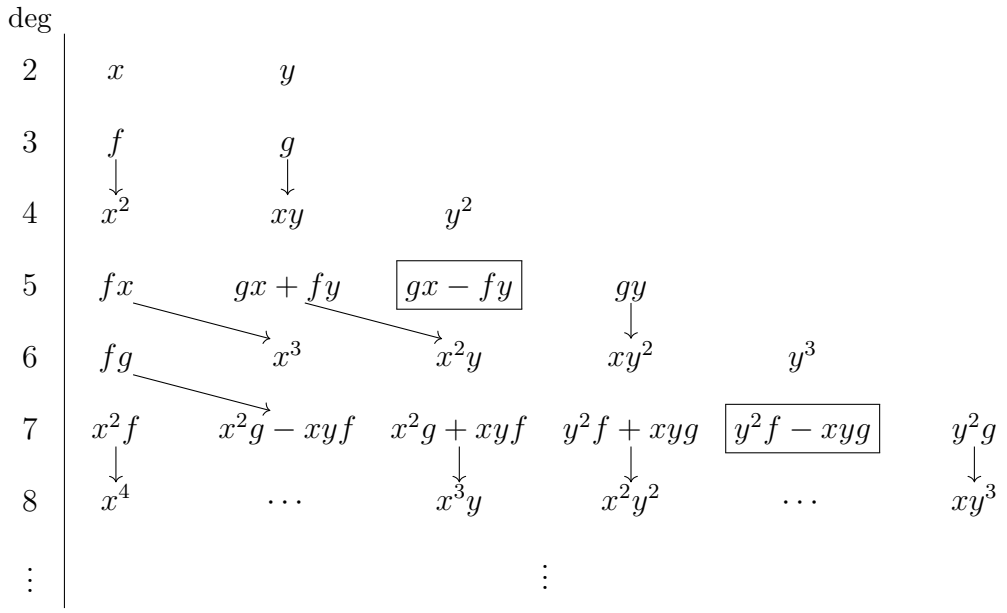


Figure 2.6: A picture of the low-degree portion of our non-formal example of a minimal CDGA. The differential is indicated by arrows, so that the long and short representations are visible. We have been sloppy about including coefficients; the indicated relationships are true up to overall scalars.

Example 11. We now discuss a minimal CDGA that fails to be formal. The example is, once again, due to [58]. Let M be freely generated by x and y in degree two and f and g in degree three. The differential will be defined by

$$dx = dy = 0 \qquad df = x^2 \qquad dg = xy. \tag{2.24}$$

This algebra cannot be formal, because it has a triple Massey product. The product xxy is zero in cohomology for two reasons (because xx and xy are both exact), and the Massey product measures the "difference" between the two. It is defined by the element

$$m(x, x, y) = fy - gx,$$

which is closed but not exact in degree five. The Massey product is ambiguous up to the ideal

$$x \cdot \mathrm{H}^3(M) + y \cdot \mathrm{H}^3(M),$$

but this vanishes since $\mathrm{H}^3(M) = 0$.

We draw a picture of this example in Fig. 2.6. To give an explanation which may be more tangible to the reader than the formal definition, Massey products

have to do with entanglement: in a minimal model of a CDGA, they are supersymmetric vacuum states that are entangled between different oscillator degrees of freedom.

To say this at more length, recall that entanglement is a property of a state in a Hilbert space, when viewed with respect to a basis that exhibits that Hilbert space as a tensor product of factors corresponding to subsystems.

In a minimal-model CDGA M, there are two natural choices of basis: one is the tensor-product basis corresponding to the different oscillator degrees of freedom that make up the system. With respect to this basis, M is freely generated as a graded polynomial algebra. The other basis is the basis consisting of energy eigenstates, which shows that M is a direct sum of irreducible representations of supersymmetry.

One of these bases is natural from the standpoint of the action of the de Rham algebra, and the other is natural from the standpoint of the multiplication. These bases may not agree; if this happens, the result is that the product of operators corresponding to zero-energy states may not be an energy eigenstate at all, but rather may be a quantum superposition of zero-energy and positive-energy states. For the same reason, a zero-energy state may not be a tensor product state of different oscillators, but may only appear as an entangled state. As always, studying the appropriate picture (Fig. 2.6) should make this clear.

In a supersymmetric quantum mechanics whose Hilbert space is a Sullivan minimal-model CDGA, the Hilbert space is that of a finite collection of bosonic and fermionic harmonic-oscillator degrees of freedom. The cohomology classes corresponding to Massey products, like the state $gx - fy$ in our example, are vacua of this system in which the oscillator subsystems exhibit nontrivial entanglement with one another. As is clear from Fig. 2.6, no tensor product state of the form

$$|f\rangle \otimes |\cdots\rangle$$

can be a vacuum state, since $|f\rangle$ is Q-exact. However, an entangled combination of states like this one *is* a new supersymmetric vacuum state.

Let us also say a few words about how one should think about the meaning of homotopy equivalence from a physical perspective. Let f and g be two homomorphisms between CDGAs A and B. The two are *homotopic* if there

exists a map $H : A \to B$ of degree -1, such that

$$f - g = d_B h + h d_A.$$

Now, an example of a homotopy equivalence between CDGAs A and B is a pair of projection and inclusion maps

$$A \xrightarrow{p} B, \quad B \xhookrightarrow{i} A,$$

such that ip is homotopic to 1_A; in other words, $ip - 1_A$ (which is projection onto the orthogonal complement of $B \subset A$) is Q-exact.

The reader may wonder why the inclusion of the cohomology into a CDGA as the set of zero-energy states does not necessarily define a homotopy equivalence between the two. The answer is that the maps i and p above are required to be homomorphisms of algebras; therefore, one cannot simply project onto any linear subspace. The kernel of p must be an ideal.

This requirement has a clear physical interpretation as well: it corresponds to our normal ideas about integrating out degrees of freedom. If the map p sends a state $|\psi\rangle$ to zero, indicating (for instance) that $|\psi\rangle$ has an energy which is above a certain cutoff value, it should also send to zero states obtained by adding additional particles to $|\psi\rangle$. In the context of CDGAs that are Fock spaces (that is to say, Sullivan algebras that are not necessarily minimal), one should only discard states corresponding to an entire harmonic-oscillator degree of freedom, when they all have nonzero energy (so that that degree of freedom is massive and can be integrated out).

It is therefore possible to think about (at least a subset of) homotopy equivalences between CDGAs, at least heuristically, as renormalization group flows. Since the RG scale is a continuous parameter of the physical theory, this is in line with the intuitive idea that a homotopy between physical theories should be a path in the moduli space.

Now let us review the main result of [58]:

Theorem 2. Compact Kähler manifolds are formal.

Proof. With the technology we have set up, the proof is almost trivial. The de Rham complex of a Kähler manifold exhibits $\mathscr{N} = 2$ supersymmetry, either with respect to supercharges ∂ and $\bar{\partial}$ or with respect to d and $d^{\mathsf{c}} \doteq i\left(\partial - \bar{\partial}\right)$.

The latter choice corresponds to $t = [1 : 1]$ in our $\mathbb{C}\mathrm{P}^1$ family. From the standpoint of complex coefficients (and therefore physics) there is no difference; however, [58] choose d and d^{c} because they are real operators, and so act on the real de Rham complex.

There is no difference between d-cohomology and d^{c}-cohomology; both count short representations of $\mathscr{N} = 2$ supersymmetry. However, passing to d^{c}-cohomology furnishes an *equivalence* of CDGAs with respect to the differential d:

$$(\Omega, d) \leftarrow (\ker d^{\mathsf{c}}, d) \rightarrow (H_{d^{\mathsf{c}}}, d). \tag{2.25}$$

It should be clear that both arrows are quasi-isomorphisms, and that the differential induced on d^{c}-cohomology by d is identically zero. The result follows immediately. □

Formality thus follows simply whenever *more than one* supercharge can be used to identify the zero-energy spectrum of the *same* Hamiltonian. This is precisely the context of $\mathscr{N} \geq 2$ supersymmetry. However, the reader should keep in mind that the absence of enhanced supersymmetry does not necessarily imply a lack of formality.

One of the initial motivations that led to this study was as follows: Not every compact manifold is formal. A large class of examples are nilmanifolds, which are obtained as quotients of nilpotent simply connected Lie groups by discrete co-compact subgroups. For instance, we can take the three-dimensional Heisenberg group consisting of matrices

$$N = \left\{ \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{R} \right\}, \tag{2.26}$$

which is homeomorphic to $\mathbb{R}^3$, and the subgroup consisting of elements for which $x, y, z \in \mathbb{Z}$. The quotient is a compact three-manifold that is not formal. One can construct simply connected manifolds with the same properties.

Since the "vanilla" flavor of supersymmetric quantum mechanics does not require any additional structure, Kähler or otherwise, we were led to ask: does physics look different on a non-formal manifold? Do the Massey products, interpreted as operations relating sets of vacuum states, have a physical significance or meaning? In light of the discussion above, the reader should

immediately see the proper response to this question. Massey products arise because of entanglement between the basis of energy eigenstates (or supersymmetry representations) and the basis that is most natural with respect to the wedge product. However, as we remarked in the introduction, the wedge product is not an obviously meaningful operation in supersymmetric quantum mechanics, and so the question does not make sense without additional structure.

It is then natural to look for other examples of physical systems for which certain sets of *operators* admit a description in terms of the cohomology of a supercharge, and therefore higher-order Massey products between (for instance) Q-cohomology classes of local operators, or supersymmetric defects with respect to the product defined by fusion [37], might occur. The results of [58] translate directly to a particular class of physical examples: $\mathscr{N} = 2$ supersymmetric sigma models in two dimensions, which require Kähler structure on the target space as a necessary and sufficient condition for supersymmetry. These sigma models admit two different well-known topological twists (the A- and B-twists), after which the TQFT Hilbert spaces are identified with the cohomology of a geometric CDGA associated to the target space. The statement that this CDGA must necessarily be formal (for the B-model, this generalization of [58] was proven by Zhou [227]) suggests that whatever physical phenomena higher cohomology operations may represent cannot occur between local operators in a two-dimensional supersymmetric sigma model—at least for the A-twist in the large volume limit where the product agrees with the wedge product. The situation is more subtle for the A-model at finite volume.

Understanding the correct notion of formality for quantum field theories is more subtle than for CDGAs or $0 + 1$-dimensional systems. The reason is that, while spaces of BPS operators, states, or defects form algebras on very general grounds [118, 140], and moreover their algebraic structure comes from natural and meaningful physical operations (such as the structure of Fock space in perturbative theories, the OPE, or defect fusion), these operations are only guaranteed to be nonsingular for BPS objects, and are highly singular in general. Formality is not a property of the cohomology ring, but rather a property of the differential algebra from which it is derived; as such, understanding the physics (and the algebraic structure) of the non-BPS part

of the theory is critical. Nonetheless, in quantum field theories that admit topological twists, one can look at the portion of the supersymmetry algebra that includes the scalar supercharge, and sometimes prove analogues of the dd^c lemma. This can be done whenever the $\mathscr{N} = 2$ algebra of quantum mechanics is a subalgebra of the higher-dimensional twisted supersymmetry algebra. We will return to this question in §2.5.

It is also interesting to notice that the above discussion suggests a way to define "homotopy groups" with complex coefficients, for $\mathscr{N} = 2$ supersymmetric quantum mechanics, or more generally for twists of theories for which a dd^c-lemma can be proven. These homotopy groups are really complex vector spaces, just like the $\mathbb{C}$-de Rham homotopy groups we discussed above. To calculate them, one should take the graded algebra of Q-cohomology classes or BPS states, view it as a CDGA with zero differential, and construct its minimal model as was described above. The graded components of the minimal-model CDGA are then the $\mathbb{C}$-de Rham homotopy groups; when the theory one started with is the $\mathscr{N} = 2$ quantum mechanics of a particle on a Kähler manifold, this procedure recovers the homotopy groups of the target space, tensored with $\mathbb{C}$. For more abstract examples of supersymmetric quantum mechanics, the meaning of these new invariants is less clear.

2.4 Physical origins of Lefschetz operators

In this section, we review work of Figueroa-O'Farrill, Köhl, and Spence [79], who showed (pursuing an idea due to Witten) that the action of Lefschetz operators in the de Rham complex of Kähler and hyper-Kähler manifolds can be understood in terms of dimensional reduction of the higher-dimensional supersymmetric sigma models that can be defined on these spaces. Our goal in doing this is to emphasize the utility of studying these theories via dimensional reduction to supersymmetric quantum mechanics, and also to point out that their argument allows analogues of the Lefschetz action to be defined in theories that are not necessarily sigma models. These Lefschetz-type symmetries are precisely the R-symmetries of the dimensionally reduced theories. We detail the consequences of these Lefschetz-type symmetries, comparing them with the consequences of two-dimensional superconformal R-symmetries.[‖] Further, we comment on some intriguing and (we believe) unexplained numerical coin-

[‖]This is motivated by the fact that two-dimensional sigma models on Calabi-Yau and hyper-Kähler manifolds are automatically superconformal.

cidences.

The central idea is simple to state. Suppose that a d-dimensional, minimally supersymmetric sigma model can be defined on a target manifold X. When $d = 4$, it is necessary and sufficient for X to be Kähler; when $d = 6$, X must be hyper-Kähler. Again, this can be seen just by counting the supercharges that act in the de Rham complex.

Ignoring questions of signature, the Lorentz group of the theory will be $\mathrm{SO}(d)$. In order to dimensionally reduce to a quantum mechanics problem, we must fix a splitting of the worldsheet coordinates into one time and $d-1$ spatial directions. This breaks the Lorentz group to $\mathrm{SO}(d-1)$, acting on the spatial directions; this symmetry should survive as a flavor symmetry in the dimensionally reduced theory.

For the four-dimensional sigma model, we therefore expect the symmetry algebra $\mathfrak{so}(3) \cong \mathfrak{su}(2)$ to act; for the six-dimensional sigma model, the relevant algebra will be $\mathfrak{so}(5)$. The result of [79] is that, in the context of Kähler and hyper-Kähler sigma models, these group actions agree precisely with the Lefschetz actions we reviewed above.

However, when arguing that these symmetry algebras must act on quantum mechanics after dimensional reduction, we did not actually use the fact that the theories in question are sigma models. The argument is quite general, and applies to the supersymmetric quantum mechanics obtained after dimensional reduction of *any* theory. As such, if we are interested in seeing how close such a supersymmetric quantum mechanics problem is to the algebraic model of an honest space, we can use the existence of a Lefschetz action without making any further assumptions.

To be concrete, consider the $\mathscr{N} = 1$ supersymmetry algebra in four dimensions:

$$\{Q_\alpha, \bar{Q}_{\dot\beta}\} = 2i\sigma^\mu_{\alpha\dot\beta} P_\mu. \tag{2.27}$$

Following [79, Table 1], make the following identification:

$$(\partial, \bar\partial^\dagger) \mapsto (Q_1, Q_2) \qquad\qquad (\partial^\dagger, \bar\partial) \mapsto (\bar{Q}_{\dot 1}, \bar{Q}_{\dot 2}).$$

Note that there is a $\mathbb{C}\mathrm{P}^1$ family of choices to be made here. We are picking a point out of the projectivization of the Weyl spinor space $\mathbb{C}^2$. This is the same as picking a direction out of three-dimensional Euclidean space.

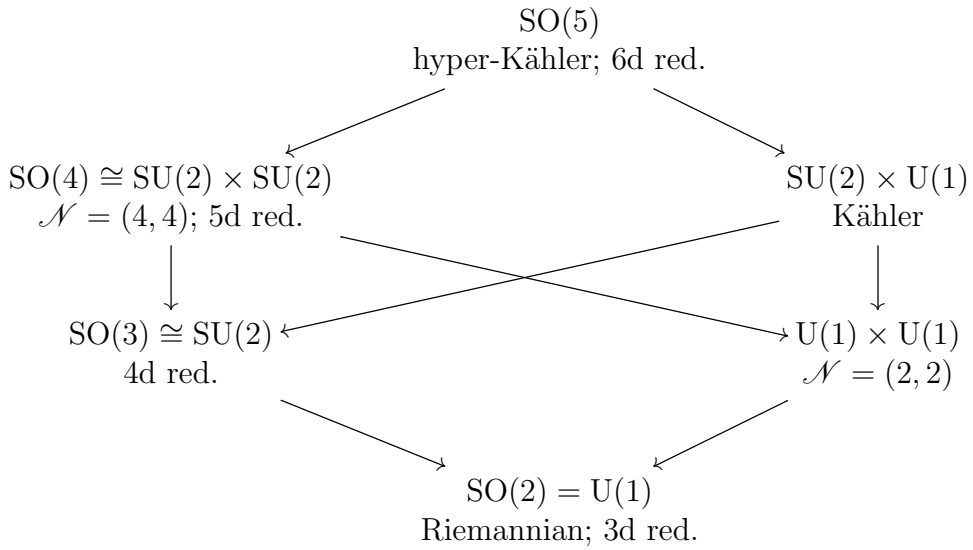


Figure 2.7: The hierarchy of group actions that occur on twists of 2d sigma models, on dimensional reductions of higher-dimensional field theories, and on the cohomology of target spaces.

It is then simple to rewrite the algebra (2.27) as follows:

$$\begin{aligned}(\Delta_\partial, \Delta_{\bar\partial}) &= (2(P_0 + P_3), 2(P_0 - P_3)),\\ \{\partial, \bar\partial\} &= \{Q_1, \bar Q_{\dot 2}\} = P_1 + iP_2,\\ \{\partial, \bar\partial^\dagger\} &= \{Q_1, Q_2\} = 0.\end{aligned} \tag{2.28}$$

Upon dimensional reduction to $0+1$ dimensions, we set the momentum operators $P_i = 0$. After doing so, this becomes the algebra of $\mathscr{N} = 2$ quantum mechanics, which we have written in the notation appropriate to Kähler manifolds.

If we do *not* dimensionally reduce to quantum mechanics, this algebra is not an $\mathscr{N} = 2$ algebra. However, using Lorentz invariance, we may without loss of generality take the momentum to lie in the P_3 direction. We then obtain the $\mathscr{N} = 1^2$ algebra, where the Laplacians are as in (2.4).

Recall from our discussion of the $\mathscr{N} = 1^2$ algebra above that the representations that spoil the dd^{c}-lemma are the two-dimensional representations that are short with respect to one supercharge and long with respect to the other. By inspection of (2.4), these are states for which $P_0 = \pm P_3$: that is to say, massless states in the four-dimensional theory.

Returning to quantum mechanics, the point is that we can naturally identify

an action of $\mathfrak{su}(2) \cong \mathfrak{so}(3)$ that is compatible with the $\mathscr{N} = 2$ supersymmetry algebra we identified above. We define the analogues of the Lefschetz operators as follows:

$$D = 2J_3 \qquad\qquad (\Lambda^\dagger, \Lambda) = J_\pm = J_1 \pm iJ_2. \tag{2.29}$$

These operators commute with the Hamiltonian P_0 by the standard Poincaré algebra:

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

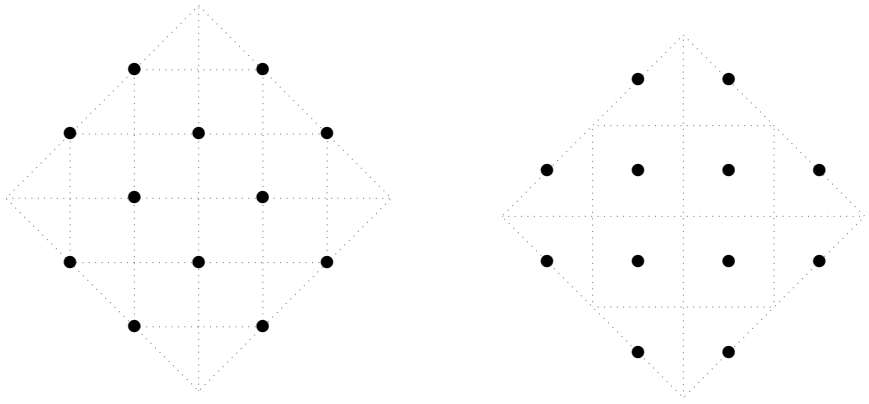
Again, although (2.29) agrees with the usual Lefschetz action in the context of sigma models, nothing in this derivation relies on having a sigma-model description of the field theory! The analogues of the crucial Kähler identities (2.14) follow from the fact that the supercharges transform as a spinor under rotations.

Given the above, it should be clear that any supersymmetric quantum mechanics arising from dimensional reduction from a field theory in d dimensions has a Lefschetz-type action of $\mathrm{SO}(d-1)$. In Fig. 2.7, we make a subgroup diagram showing these actions together with those that follow from R-symmetry in two-dimensional superconformal theories and those that follow from geometric realization as a sigma model. Each subgroup corresponds to a specific element on the partially ordered set of geometrical consequences.

This partially ordered set is not totally ordered. The consequences of $\mathscr{N} = (2,2)$ superconformal symmetry and realizability as a Kähler sigma model are incomparable, for instance. Since $\mathrm{SU}(2)$ acts vertically on the Hodge diagram for Kähler manifolds, but the $\mathrm{SU}(2) \times \mathrm{SU}(2)$ arising from $\mathscr{N} = (4,4)$ R-symmetry acts diagonally — one is not a subgroup of the other.

Furthermore, notice that the conjunction of these two requirements is strictly weaker than the consequences of realizability as a hyper-Kähler sigma model. For instance, Fig. 2.8b is consistent with $\mathscr{N} = (2,2)$ superconformal symmetry and the Kähler $\mathrm{SU}(2)$ Lefschetz action, but is inconsistent with $\mathrm{SO}(5)$ Lefschetz action (since the diagram has the wrong degeneracies in the interior of the octagon).

There is one interesting thing to note. Verbitsky constructed the $\mathrm{SO}(4,1)$ action for hyper-Kähler manifolds by using the three Kähler classes as generators. Since these three classes are independent, $b_2 \geq 3$ for any hyper-Kähler



(a) This pattern is consistent with $\mathscr{N} = (4,4)$ superconformal algebra's $\mathrm{SU}(2) \times \mathrm{SU}(2)$ action, but is incompatible with 4d Lefschetz action.

(b) This octagon is compatible with $\mathscr{N} = 4$ with 4d Lefschetz action, but is incompatible with 6d Lefschetz action.

Figure 2.8: Illustration of the differences between requirements tabulated in Fig. 2.7.

manifold. However, if b_2 is large, there are many more "candidates" for defining Lefschetz operators, and one may wonder if a larger symmetry algebra can be defined in this case.

In later work, Verbitsky [198, 199] showed that this is indeed the case. He constructed an action of $\mathrm{SO}(4, b_2 - 2)$ on the cohomology of a hyper-Kähler manifold, and it is natural to wonder (indeed, the authors of [79] already do so) whether this can be understood from a supersymmetric perspective as well.

The only compact hyper-Kähler four-manifold is the K3 surface. For hyper-Kähler manifolds of (real) dimension eight, Guan [104] has shown that the second Betti number must be one of the following:

$$b_2 \in \{3, 4, 5, 6, 7, 8, 23\},$$

although examples are only explicitly known for $b_2 = 7, 23$. We draw the Hodge diamonds of these examples [181] in Fig. 2.9.

If one were to realize Verbitsky's $\mathfrak{so}(b_2 + 2)$ actions for these target spaces in terms of Lorentz symmetry, as has been done for the standard Lefschetz actions, this would correspond to the following set of dimensions for the corresponding sigma models:

$$d \in \{6, 7, 8, 9, 10, 11, 26\}.$$

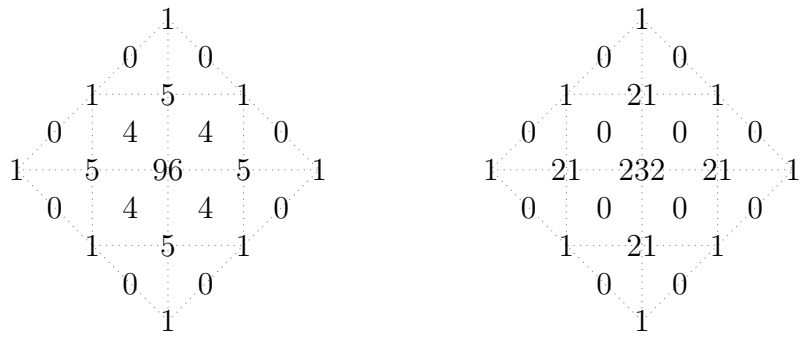


Figure 2.9: The Hodge diamonds of known examples of compact hyper-Kähler eight-manifolds with $b_2 > 3$.

The largest three—10, 11, and 26—happen to equal the spacetime dimensions of superstring theory, M-theory, and bosonic string theory. Moreover, the examples that have been constructed correspond exactly to $d = 10$ and $d = 26$. We emphasize, however, that these should be thought of as worldsheet (not target-space) dimensions!

It is worth pointing out that compactification of superstring theory on a hyper-Kähler eight-manifold leaves a two-dimensional effective theory in the target space. In some sense, it is as if a rather bizarre duality exchanges the roles of target space and worldsheet.

To the best of our knowledge, there is no straightforward explanation of these facts, since there is no such thing as the supersymmetric sigma model in dimension $d > 6$. (This can be seen in many ways: from Berger's classification of exceptional holonomy, or by noting that the algebra of complex structures required for $\mathscr{N} = 8$ would have to be that of the octonions.) In such a large spacetime dimension, the spinor representations in supersymmetry would be far bigger than the number of differential operators present on differential forms on a hyper-Kähler manifold.

These considerations suggest a scenario in which a d-dimensional theory, which is not supersymmetric, somehow acquires supersymmetry upon dimensional reduction to six (or fewer) dimensions. However, we lack any concrete proposal or meaningful evidence for this speculative scenario.

2.5 Topological twists of quantum field theories

In this section, we take inspiration from the discussion of the dd^c-lemma above and look at topological twists of field theories in higher dimensions. For some

twists, the supersymmetry algebra after twisting contains the algebra of $\mathscr{N} = 2$ quantum mechanics as a subalgebra; for others, this is not true.

4d $\mathscr{N} = 2$

The relevant part of the four-dimensional extended supersymmetry algebra is

$$\{Q^A_\alpha, \bar{Q}_{\dot\beta B}\} = 2\sigma^\mu_{\alpha\dot\beta} P_\mu \delta^A_B,$$
$$\{Q^A_\alpha, Q^B_\beta\} = \epsilon_{\alpha\beta}\epsilon^{AB} Z,$$

The adjoint relationship is, once again, $(Q^A_\alpha)^\dagger = \bar{Q}_{\dot\alpha A}$. We will only consider the case without central charges, so that the scalar supercharges obtained after twisting are honestly nilpotent. In the case when Z is nonzero, the scalar supercharge will square to Z; topological twists can still be understood in this setting, but we will not address this here. The reader is referred to [137].

The representations in which the supercharges sit are

$$Q^A_\alpha \colon (\mathbf{2}, \mathbf{1}; \mathbf{2}) \qquad\qquad \bar{Q}^B_{\dot\beta} \colon (\mathbf{1}, \mathbf{2}; \mathbf{2}).$$

The first two numbers refer to the representation of the Lorentz group $\mathrm{SO}(4) \cong \mathrm{SU}(2)_{\text{left}} \times \mathrm{SU}(2)_{\text{right}}$, and the last number is a representation of $\mathrm{SU}(2)$ R-symmetry. Up to equivalence, there is a unique way to construct a topological twist of this algebra. One chooses the homomorphism

$$\mathrm{pr}_1 \colon \mathrm{SU}(2)_{\text{left}} \times \mathrm{SU}(2)_{\text{right}} \to \mathrm{SU}(2)_R,$$

and lets the Lorentz group act on R-symmetry indices via this identification rather than trivially. The resulting representations are

$$\left(Q^1_1, \frac{1}{\sqrt{2}}(Q^2_1 + Q^1_2), Q^2_2\right) \colon (\mathbf{3}, 1)$$
$$q \doteq \frac{1}{\sqrt{2}}(Q^2_1 - Q^1_2) \colon (1, 1)$$
$$\bar{Q}^B_{\dot\beta} \colon (\mathbf{2}, \mathbf{2}).$$

The quantum numbers indicated are for the new Lorentz group $\mathrm{SU}(2)'_{\text{left}} \times \mathrm{SU}(2)_{\text{right}}$. The scalar supercharge with respect to which we twist is q; in order that q be nilpotent, we must set the central charge to zero. It is simple to

check that the "Laplace operator" corresponding to this supercharge is

$$\begin{aligned}\{q, q^\dagger\} &= \frac{1}{2}\{Q_1^2 - Q_2^1, \bar{Q}_{\dot{1}2} - \bar{Q}_{\dot{2}1}\} \\ &= \frac{1}{2}\left(2P_\mu \sigma^\mu_{1\dot{1}} + 2P_\mu \sigma^\mu_{2\dot{2}}\right) \\ &= \mathrm{tr}(P_\mu \sigma^\mu_{\alpha\beta}) \\ &= 2P_0.\end{aligned}$$

Therefore, the q-cohomology counts states on which the Hamiltonian acts by zero. This is what we should expect if the result of twisting is to be a topological theory.

There is no other scalar supercharge in the theory, so adding any more supercharges to our subalgebra will give a result that is not Lorentz invariant. Of course, this is not a problem: one merely needs the existence of an $\mathscr{N} = 2$ subalgebra to show that the dd^{c}-lemma holds, whereas a single Lorentz-invariant supercharge is enough to make the twist. It is instructive to consider the commutation relations with the supercharge

$$\tilde{q} = \frac{1}{\sqrt{2}}\left(Q_1^2 + Q_2^1\right).$$

The two supercharges commute with one another (even in the presence of central charges):

$$\begin{aligned}\{q, \tilde{q}\} &= \frac{1}{2}\{Q_1^2 - Q_2^1, Q_1^2 + Q_2^1\} \\ &= \frac{1}{2}\left(\epsilon_{12} Z^{21} - \epsilon_{21} Z^{12}\right) \\ &= 0.\end{aligned}$$

We have therefore identified a $\mathscr{N} = 1.5$ subalgebra. However, this fails to close into an $\mathscr{N} = 1^2$ algebra. It is easy to check that

$$\{q, \tilde{q}^\dagger\} = 2P_3.$$

However, the Laplacian of $\tilde{q}$ nonetheless agrees with that of q:

$$\{\tilde{q}, \tilde{q}^\dagger\} = 2P_0.$$

The reader may see a puzzle here—the Laplacians agree, so there is no way for a representation to have both zero and nonzero energy. It should follow that the one-dimensional representations are the only short representations.

The resolution to this puzzle consists of remembering that (in Lorentzian signature) $P_0 \geq |P_3|$. Therefore, if a state satisfies $P_0 = 0$, the operator P_3, which is the staircase operator in this context, must act on it by zero as well. This is sufficient to show that the only permissible staircases have length two, just as is the case for the $\mathscr{N} = 1^2$ algebra.

Indeed, the algebra we have written is equivalent to the $\mathscr{N} = 1^2$ algebra after a change of basis. Taking the supercharges to be

$$\begin{aligned} \partial = Q^2_1, \quad \bar{\partial} = -Q^1_2, \\ \partial^\dagger = \bar{Q}_{\dot{1}2}, \quad \bar{\partial}^\dagger = -\bar{Q}_{\dot{2}1}, \end{aligned}$$

it follows immediately that

$$\begin{aligned} \Delta_\partial = \{\partial, \partial^\dagger\} = 2\sigma^\mu_{11} P_\mu = P_0 + P_3 \\ \Delta_{\bar{\partial}} = \{\bar{\partial}, \bar{\partial}^\dagger\} = 2\sigma^\mu_{2\dot{2}} P_\mu = P_0 - P_3 \\ \{\partial, \bar{\partial}\} = \{\partial, \bar{\partial}^\dagger\} = 0. \end{aligned}$$

This is therefore a typical $\mathscr{N} = 1^2$ system. The new states that appear in ∂- and $\bar{\partial}$-cohomology, as compared to $q = \partial + \bar{\partial}$-cohomology, satisfy $P_3 = \pm P_0$ respectively. Put differently, *massless* states cause the failure of the dd^c-lemma. In a topological twist of a massive four-dimensional $\mathscr{N} = 2$ theory, we would expect it to hold.

3d $\mathscr{N} = 4$

Inspecting the quantum-mechanical subalgebra we were considering above, it is easy to see that it becomes an $\mathscr{N} = 2$ subalgebra upon setting $P_3 = 0$, i.e., after dimensional reduction to three dimensions. It makes sense that this should generate a new scalar supercharge: after the twist, the $\bar{Q}^B_{\dot{\beta}}$ are in a vector representation of $\mathrm{SO}(4)'$, which will become a vector and a scalar upon dimensional reduction. The new scalar is precisely $q^\dagger$; therefore, we obtain a complete Lorentz-invariant copy of the de Rham algebra, sitting inside a non-Lorentz-invariant $\mathscr{N} = 2$ algebra.

This shows that the dd^c-lemma holds for one of the two possible twists of three-dimensional $\mathscr{N} = 4$ theories. While there are two twists that are in general inequivalent [22], they are the same at the level of the supersymmetry algebra. Only the decomposition of supermultiplets differs. As such, the $\mathscr{N} = 2$ subalgebra we identified above exists in both possible twists.

4d $\mathscr{N} = 4$

This is the dimensional reduction of minimal supersymmetry in ten dimensions. The R-symmetry is $\mathrm{Spin}(6) \cong \mathrm{SU}(4)$. The representations of the supercharges before twisting are

$$Q^A_\alpha \colon (\mathbf{2}, \mathbf{1}; \mathbf{4}) \qquad\qquad \bar{Q}^B_{\dot\beta} \colon (\mathbf{1}, \mathbf{2}; \bar{\mathbf{4}}). \tag{2.30}$$

Donaldson–Witten twist This corresponds to choosing an $\mathscr{N} = 2$ subalgebra and twisting as above. The twisting homomorphism embeds $\mathrm{SU}(2)_{\text{left}}$ in the obvious way as a block diagonal inside $\mathrm{SU}(4)$. This leaves an $\mathrm{SU}(2) \times \mathrm{U}(1)$ subgroup of the R-symmetry unbroken. The R-symmetry representations of the supercharges decompose as

$$(\mathbf{2}, \mathbf{1}; \mathbf{4}) \to (\mathbf{3}, \mathbf{1}; \mathbf{1})_1 \oplus (\mathbf{1}, \mathbf{1}; \mathbf{1})_1 \oplus (\mathbf{2}, 1; \mathbf{2})_{-1} \tag{2.31}$$

$$(\mathbf{1}, \mathbf{2}; \bar{\mathbf{4}}) \to (\mathbf{2}, \mathbf{2}; \mathbf{1})_{-1} \oplus (\mathbf{1}, \mathbf{2}; \mathbf{2})_1 \tag{2.32}$$

The representations indicated at right are for $\mathrm{SU}(2)'_{\text{left}}$, $\mathrm{SU}(2)_{\text{right}}$, $\mathrm{SU}(2)_R$, and $\mathrm{U}(1)_R$. Taking $\mathrm{SU}(2)'_{\text{left}}$ to act on the first two R-symmetry indices, the scalar supercharge is

$$q = \frac{1}{\sqrt{2}} \left(Q^2_1 - Q^1_2 \right),$$

just as for the twist of the $\mathscr{N} = 2$ theory. All the calculations from above continue to be valid, so that the corresponding Laplacian is again proportional to P_0.

However, it is now easy to find another choice of supercharge that defines an $\mathscr{N} = 2$ subalgebra. For instance, one can take

$$\tilde{q} = \frac{1}{\sqrt{2}} \left(Q^4_1 - Q^3_2 \right).$$

It is simple to check the commutation relations between q, $\tilde{q}$, and their adjoints. Once again, this subalgebra is not Lorentz-invariant, but this is not important in the context of proving the dd^{c}-lemma.

Vafa–Witten twist This corresponds to embedding $\mathrm{SU}(2)_{\text{left}}$ diagonally inside the obvious $\mathrm{SO}(3) \times \mathrm{SO}(3)$ subgroup of $\mathrm{SO}(6) \cong \mathrm{SU}(4)$. The unbroken R-symmetry is $\mathrm{SU}(2)_R$, rotating the two diagonal blocks into one another. (If we write the $\mathrm{SU}(4)$ fundamental index as a pair of indices valued in $\{1, 2\}$,

$\mathrm{SU}(2)_{\text{left}}$ acts on one of these indices and the unbroken $\mathrm{SU}(2)_R$ on the other.) The representations of the supercharges decompose as follows:

$$(\mathbf{2},\mathbf{1};\mathbf{4}) \to (\mathbf{3},\mathbf{1};\mathbf{2}) \oplus (\mathbf{1},\mathbf{1};\mathbf{2}), \tag{2.33}$$

$$(\mathbf{1},\mathbf{2};\bar{\mathbf{4}}) \to (\mathbf{2},\mathbf{2};\mathbf{2}). \tag{2.34}$$

The representations indicated at right are for $\mathrm{SU}(2)'_{\text{left}} \times \mathrm{SU}(2)_{\text{right}} \times \mathrm{SU}(2)_R$. The scalar supercharges can be written explicitly as

$$q_\uparrow = \frac{1}{\sqrt{2}}\left(Q_1^2 - Q_2^1\right) \qquad\qquad q_\downarrow = \frac{1}{\sqrt{2}}\left(Q_1^4 - Q_2^3\right). \tag{2.35}$$

(The subscripts refer to the representation of $\mathrm{SU}(2)_R$.)

These supercharges are the same as the q and $\tilde{q}$ identified above, and so satisfy the same commutation relations, generating an $\mathscr{N} = 2$ subalgebra.

Marcus–Kapustin–Witten twist This corresponds to the obvious block-diagonal homomorphism

$$\mathrm{SU}(2)_{\text{left}} \times \mathrm{SU}(2)_{\text{right}} \to \mathrm{SU}(4)_R.$$

There is an unbroken $\mathrm{U}(1)_R$ symmetry, commuting with this embedding, which rotates the overall phases of the two blocks in opposite directions. Under this embedding the $\mathbf{4}$ of the R-symmetry group transforms as $(\mathbf{2},1)_1 \oplus (1,\mathbf{2})_{-1}$ with respect to the twisted Lorentz group and the unbroken $\mathrm{U}(1)_R$. The transformation of the $\bar{\mathbf{4}}$ representation is the same, but with $\mathrm{U}(1)$ charges reversed.

It follows that the supersymmetries transform after the twist as

$$\begin{aligned}(\mathbf{2},\mathbf{1};\mathbf{4}) &\to (\mathbf{1},\mathbf{1})_1 \oplus (\mathbf{3},\mathbf{1})_1 \oplus (\mathbf{2},\mathbf{2})_{-1} \\ (\mathbf{1},\mathbf{2};\bar{\mathbf{4}}) &\to (\mathbf{1},\mathbf{1})_1 \oplus (\mathbf{1},\mathbf{3})_1 \oplus (\mathbf{2},\mathbf{2})_{-1}.\end{aligned} \tag{2.36}$$

The representations indicated at right are for $\mathrm{SU}(2)'_{\text{left}} \times \mathrm{SU}(2)'_{\text{right}} \times \mathrm{U}(1)_R$. (Our convention for the sign of the $\mathrm{U}(1)_R$-charge is the opposite of that of [124], so that the scalar supercharges carry charge $+1$. This is in keeping with our use of cohomological grading conventions, for which the differential has positive degree, throughout the paper.)

Explicitly, the scalar supercharges can be written as

$$q_{\text{left}} = \frac{1}{\sqrt{2}}\left(Q_1^2 - Q_2^1\right), \quad q_{\text{right}} = \frac{1}{\sqrt{2}}\left(\bar{Q}_{\dot{1}4} - \bar{Q}_{\dot{2}3}\right). \tag{2.37}$$

The subscripts indicate the chirality of the spinor supercharge from which the scalar is derived. It is clear that these supercharges and their adjoints form the same subalgebra as for the Vafa-Witten twist, since $q_{\text{left}} = q_\uparrow$ and $q_{\text{right}} = q_\downarrow^\dagger$.

In fact, the VW and MKW twists are isomorphic for the sphere S^3 Hilbert space (equivalently, for local operators). These are just $\operatorname{tr} \phi^n$ in both cases, where ϕ is the adjoint sgaugino. The two twists are, however, very different for non-spheres, where the TQFT is *not* a subsector of the untwisted SQFT.

2d superconformal algebras

As we already mentioned above, the $\mathscr{N} = (2,2)$ superconformal algebra has a global subalgebra (for instance, in the Ramond sector) which is precisely equal to two copies of the de Rham algebra, one in the left-moving and one in the right-moving sector. As such, the algebra relevant to twists of these theories is always the $\mathscr{N} = 1^2$ algebra.

In this case, one can consider deformation problems of the sort we discussed in general, where some mixture of these four supercharges defines a nilpotent operator. This leads to the A- and B-twists, as well as to the elliptic genera of $\mathscr{N} = (2,2)$ theories at the special points where BPS degeneracies are enhanced. Further remarks on these ideas in the context of $\mathscr{N} = (2,2)$ theories will be made in [109].

All twists of $\mathscr{N} = (4,4)$ theories come from $\mathscr{N} = (2,2)$ subalgebras: there is a $\mathbb{C}\mathrm{P}^1$ family of twists [17], and this corresponds to choosing a U(1) inside SU(2) R-symmetry, which amounts to choosing $\mathscr{N} = 2$ subalgebra of $\mathscr{N} = 4$.

2.6 Superconformal symmetry and reconstructing target spaces

In this section, we briefly note a couple of the results of Sullivan and others that motivated us to begin reading about rational homotopy theory.

The theorems concern when an abstract rational homotopy type (presented, for instance, by a minimal-model CDGA) is, in fact, the rational homotopy type of an honest manifold. As such, one should think of them as stating sufficient conditions for an abstract algebraic model to be "geometrical." This is very close to the question we raised in the introduction: given (for instance) a two-dimensional $\mathscr{N} = (2,2)$ superconformal theory, which one may think of as a string theory background, what conditions are sufficient to ensure that it can be described as the sigma model on some geometrical target space?

In the context of Theorem 3, one enumerates a list of all structures that can be defined on a CDGA originating from geometry, and carefully details the properties that these structures have. This list of necessary conditions turns out to also be sufficient.

Theorem 3 (Sullivan–Barge; see [190], as well as [77], Theorem 3.2)**.** Let M be a simply connected Sullivan minimal model over $\mathbb{Q}$, whose cohomology satisfies Poincaré duality with respect to a top form in dimension d. Choose an element $p \in \oplus_i \mathrm{H}^{4i}(M)$, representing the total rational Pontryagin class.

If $4 \nmid d$, there is a compact simply connected manifold whose rational homotopy type has M as its minimal model and p as its total Pontryagin class.

If $d = 4k$, the statement remains true if the manifold is permitted to have one singular point. The singular point may be removed if, and only if:

- the intersection form is equivalent over $\mathbb{Q}$ to a quadratic form $\sum \pm x_i^2$; and
- either the signature is zero, or a choice of fundamental class can be made such that the Pontryagin numbers are integers satisfying certain congruences.

Theorem 4 ([190], Theorem 12.5)**.** The diffeomorphism type of a simply connected Kähler manifold is determined by its integral cohomology ring and total rational Pontryagin class, up to a finite list of possibilities.

One key feature in the above theorems is the choice of coefficients: integers or rational numbers, rather than complex numbers, which are most natural from the standpoint of physics. We were thus led to ask whether one could recover a natural integer lattice in the physical Hilbert space of a sigma model, corresponding to the image of the integral cohomology in cohomology with complex coefficients. In fact, another motivation led us to think about this as well. In considering the geometrical symmetries of the cohomology of a Kähler manifold, the Hodge symmetry map is *antilinear*: it identifies $\mathrm{H}^{p,q}$ with $\bar{\mathrm{H}}^{q,p}$. So far we have only discussed this map at the level of Hodge numbers. We asked whether any physical symmetry of the theory could recover this complex-antilinear operation.

In the context of physics, one has a meaningful inner product on the Hilbert space, which defines the notion of a vector of length one. However, as is familiar from undergraduate quantum mechanics, the phase of a normalized state can still in general be arbitrary.

In a unitary two-dimensional conformal field theory with Hilbert space $\mathcal{H}$, consider a state $|\phi\rangle$ with conformal weights $(h, \bar{h})$. Identifying spacetime with the complex plane, the local operator $\phi(z, \bar{z})$ corresponding to $|\phi\rangle$ via the operator-state correspondence can be Fourier-expanded as

$$\phi(z, \bar{z}) = \sum_{n \in \mathbb{Z}+\eta} \sum_{\bar{n} \in \mathbb{Z}+\bar{\eta}} \phi_{-h-n, -\bar{h}-\bar{n}} z^n \bar{z}^{\bar{n}} \tag{2.38}$$

where $\eta, \bar{\eta} \in \{0, 1/2\}$ according to specified boundary conditions. The modes' conformal weights take values in $(\mathbb{Z}+h+\eta, \mathbb{Z}+\bar{h}+\bar{\eta})$. The Hermitian adjoints of these modes have conformal weights $(\mathbb{Z}-h-\eta, \mathbb{Z}-\bar{h}-\bar{\eta})$. These two sets intersect (and coincide) precisely when $h, \bar{h} \in \frac{1}{2}\mathbb{Z}$. In that case, the following requirement may be nontrivially imposed:

$$\phi^\dagger_{n,\bar{n}} = \phi_{-n,-\bar{n}}. \tag{2.39}$$

The set of states with (half-) integer conformal weights that satisfy Eq. (2.39) form a real (but not complex) linear subspace of the entire Hilbert space. Call this subspace $\mathcal{H}_0^{\mathbb{R}}$. Then its complexification

$$\mathcal{H}_0^{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C} \equiv \mathcal{H}_0 \subset \mathcal{H}$$

is the complex linear subspace of $\mathcal{H}$ consisting of all states with (half-) integer conformal weights. By virtue of having a preferred real subspace, $\mathcal{H}_0$ is equipped with a canonical $\mathbb{C}$-antilinear automorphism, which we denote[**]

$$I \colon \mathcal{H}_0 \to \mathcal{H}_0.$$

This map fixes the conformal weights, but otherwise all other charges are conjugated. In particular, for a theory with $\mathcal{N} = (2, 2)$ superconformal symmetry, I flips the various U(1) charges:

$$I \colon (h, \bar{h}, q_{\text{left}}, q_{\text{right}}) \mapsto (h, \bar{h}, -q_{\text{left}}, -q_{\text{right}}).$$

[**]This construction is not at odds with the axiom of quantum mechanics that the (absolute) phase of a state is unobservable. The choice of operator-state map gives us the vacuum state (which maps to the identity operator), and phases of all states are measured relative to it. The phase of the vacuum (as determined by the operator-state correspondence) is arbitrary, as quantum mechanics requires.

Now, consider the parity-reversal operator Ω, which acts on the worldsheet coordinate z as

$$\Omega\colon z \mapsto \bar{z}.$$

This is a unitary operator that exchanges left- and right-moving symmetries. In particular,

$$\Omega\colon (h, \bar{h}, q_{\text{left}}, q_{\text{right}}) \mapsto (\bar{h}, h, q_{\text{right}}, q_{\text{left}}).$$

Composing the two, we obtain a $\mathbb{C}$-antilinear map

$$I \circ \Omega\colon (h, \bar{h}, q_{\text{left}}, q_{\text{right}}) \mapsto (\bar{h}, h, -q_{\text{right}}, -q_{\text{left}}).$$

It is easy to see that the above maps the (a,c)-ring to itself. In fact, in the A-model on a Calabi-Yau manifold, it is easy to see that the above acts on (p, q)-forms as

$$(p, q) \mapsto (q, p),$$

since the degree $p+q$ corresponds to the axial R-charge while $p-q$ corresponds to the vector R-charge.

Therefore, we have recovered the antilinear Hodge-symmetry map

$$\mathrm{H}^{p,q} \to \bar{\mathrm{H}}^{q,p}.$$

2.7 Conclusion

We wish to call the attention of physicists to the fact that many methods and results of rational homotopy theory are essentially quantum-mechanical in nature, and can be translated into physical language. Rational homotopy theory aims to study the geometry of manifolds as encoded in the de Rham complex; as such, supersymmetric quantum mechanics as a probe of a space is its basic ingredient, and many of the main concepts that arise (Massey products, minimal models, and so on) are quite natural and pleasant to see from the perspective of physics.

The techniques of rational homotopy theory allow one to extract the homotopy groups of a space, modulo torsion, from its de Rham complex. We observed that identical techniques can be used to assign "homotopy groups" over $\mathbb{C}$ to topologically twisted quantum field theories. If the theory in question satisfies an analogue of the dd^c-lemma, one can compute its homotopy groups just from the knowledge of the algebra of BPS states. In the simplest examples of topologically twistable sigma models, one recovers the ranks of higher homotopy

groups of the target; beyond this setting, the meaning of these new invariants is unclear.

Arising out of this, one can use supersymmetric quantum mechanics as a tool to understand how close generic classes of quantum field theories are to being geometric—that is to say, describable as sigma models. There are certain properties that immediately mark a theory as being of non-geometric origin. For instance, the series of $\mathscr{N} = 2$ superconformal minimal models are not geometric: their central charges, which would play the role of the dimension of the target space, are not integral. Similar integrality properties for R-charges (which translate in sigma models to the homological gradings) also fail.

Work of Figueroa-O'Farrill, Köhl, and Spence [79] shows that these U(1) charges sometimes fit into a larger nonabelian symmetry that acts on the quantum mechanics. When this is the case, the required integrality properties follow immediately from integrality properties of weights for representations of semisimple symmetry algebras. In the context of sigma models, [79] proved that these algebras (which are Lefschetz symmetries in that case) coincide with the algebras one expects to see from dimensional reduction of higher-dimensional sigma models; we pointed out that these symmetries arise in any quantum mechanics that is the dimensional reduction of a field theory, and compared them with the standard list of R-symmetries.

Regarding the integrality of the central charge, it is known that all representations of the two-dimensional $\mathscr{N} = (4,4)$ superconformal algebra have integer central charge. Furthermore, precisely in this case, the two $\mathrm{U}(1)_R$ charges fit into larger semisimple R-symmetries, and so also have integrality properties. This is at least circumstantial evidence that the rich structures that arise for hyper-Kähler sigma models are equivalent to the rich structure of $\mathscr{N} = (4,4)$ SCFTs, and makes it tempting to wonder whether both are equivalent to the property of coming from a six-dimensional theory. (This is obvious for $\mathscr{N} = (4,4)$ sigma models, which are the dimensional reductions of the six-dimensional sigma models defined on the same target space.)

For eight-dimensional hyper-Kähler manifolds, we observed a strange phenomenon: the Lefschetz symmetries that arise on their de Rham complexes (which in other cases arise from dimensional reduction of sigma models) correspond to the values 10 and 26 of the worldsheet dimension. The effective target-space description of a string theory compactification on such manifolds

would be two-dimensional. We have no way of understanding these numbers (and their occurrence may simply be a coincidence) but it would be interesting to search for connections between this phenomenon and the strange properties of supersymmetric quantum field theory in six dimensions. We defer this to future work.

Lastly, with an eye towards the same questions about geometric and non-geometric theories, we noted the physical relevance of results of Sullivan and others, related to the problem of reconstructing a manifold (up to diffeomorphism) from its de Rham complex. In these results, it is important that one can begin with the de Rham complex with rational or integer coefficients, rather than real or complex. We suggested that it may be possible to recover this data from $\mathscr{N} = (2, 2)$ superconformal theories, subject to certain assumptions. Understanding the appropriate analogues of theorems like Sullivan–Barge in the context of physics may lead to progress on questions such as mirror symmetry, which originate as certain ambiguities in the way that string theories encode geometric properties of the target space.

Chapter 3

PERTURBING BPS STATES: SPECTRAL SEQUENCES

[1]S. Gukov, S. Nawata, I. Saberi, M. Stošić, and P. Sułkowski. "Sequencing BPS spectra." JHEP **03**, 004. (2016). arXiv:1512.07883 [hep-th].

3.1 Spectral sequences in physics

The bulk of this paper deals with the structure of link homologies. As we explain in great detail in the next section, link homologies are realized as spaces of BPS states in several M-theory configurations, in which the geometry of certain branes is prescribed by a choice of link $L \subset S^3$ [59, 108, 112, 168, 218]. Schematically,

$$\mathscr{H}(L) \cong \mathscr{H}_{\mathrm{BPS}}.$$

Studying descriptions of this space of BPS states from various viewpoints and in various duality frames then leads to interesting predictions about the structure of link homologies.

This interpretation has led to deep insights regarding the structure and behavior of homological link invariants. In particular, the rich structural properties of link homologies include differentials of various kinds relating different homology theories, which have been formulated in terms of *spectral sequences* in mathematics [72, 98, 138, 173]. One of the main goals of this paper is to obtain a physical understanding of the spectral sequences between various link homologies. These spectral sequences imply relations between the BPS spectra of different configurations of branes in M-theory, or more generally between different physical theories. Our goal is to understand these relations.

While the notion of a spectral sequence may be unfamiliar or daunting to many physicists, we hope to demonstrate that it can be understood in simple terms, which are clearly connected to physical problems. While certain applications of spectral sequences to physics have been made before (for example, see [18, 23, 30, 69, 219]), we feel that our general understanding of the context in which a physicist should expect spectral sequences to arise is novel. Put simply, spectral sequences describe how the cohomology of a supercharge Q changes under deformations of continuous parameters. Generally, Q-cohomology does

not remain invariant under deformations of the theory, even in contexts where the deformation leaves the Witten index and other supersymmetric indices invariant. For this reason, many of our considerations apply very generally to supersymmetric theories, and should not be construed as limited to the supersymmetric systems we use to describe links.

The first sections of this paper, therefore, discuss spectral sequences and deformation problems in supersymmetric theories in very general terms, and give simple examples of how well-known results (ranging from twists of 2d $\mathscr{N} = (2,2)$ theories to specializations of 4d indices) fit into our framework. They can be read independently of the rest of the paper. Once we have developed the necessary ideas, we review the physical approach to link homologies in §3.2, and then apply them to understand the deformation spectral sequences constructed in [98, 138]. The remainder of the paper (omitted from this thesis) continues further with our study on "color" dependence of link homologies; for that material, the reader is referred to [109].

Explanations of spectral sequences can be found in many places in the mathematics literature (for instance, see [26, 125]). However, we will offer a pedestrian exposition below, in language which is hopefully both familiar to physicists and tailored to our purposes. In addition, we will present several examples of spectral sequences in the BPS spectra of physical theories. One of these examples (in the context of Landau-Ginzburg models) will be the relevant case in the context of link homology.

Generalities: deformations and spectral sequences

We are interested in the following very general question: Suppose we have a supersymmetric quantum field theory, in which a single nilpotent supercharge Q has been chosen. The supercharge could be a scalar (as in a topological twist of the theory), or more generally any element of the fermionic part of the supersymmetry algebra (for instance, in superconformal indices). We make no assumptions about the action of the Lorentz group on this supercharge.

We would then like to know what happens to Q-cohomology as we continuously vary some choice or parameter of the theory. The parameters which can be varied are naturally of two types:

1. we could change our choice of Q, while keeping the theory the same [18, 23, 30, 69];

2. we could vary some modulus of the theory, such as the superpotential or other coupling constants [219].

While the two types of perturbation are distinct, they both allow us to pose the same broad question: What happens to Q-cohomology as continuous parameters are adjusted?

In fact, the spectrum of BPS states (i.e., Q-cohomology) changes under the deformation; this is part of the reason for introducing genuinely protected quantities, such as the Witten index [208, 210], in the first place. The Witten index counts the number of vacua in supersymmetric quantum mechanics, with bosonic vacua contributing $+1$ and fermionic vacua -1:

$$\mathrm{Tr}(-1)^F e^{-\beta H} = n_B - n_F \ .$$

It is well-known that the numbers n_B and n_F are not themselves invariant. However, a long representation of supersymmetry that breaks apart into short representations will always contribute equally to n_B and n_F, so that their difference is honestly invariant under perturbations.

Moreover, while both n_B and n_F can jump, they do so in a particular fashion that can be understood on general grounds. Since the numbers n_B and n_F are degeneracies of a quantum system, one expects enhanced degeneracy to correspond to enhanced symmetry. Moreover, the degeneracies should be enhanced only on a locus of positive codimension in the parameter space, where the parameters are tuned so that symmetry-breaking effects are absent. One can argue for the same conclusion abstractly by considering a one-parameter family of differentials d_t acting in a graded vector space, and noticing that d_t-cohomology jumps only at values of t for which certain linear subspaces have non-generic intersection, i.e., when certain equations on the parameter are satisfied. In some cases, we will be able to explicitly identify the symmetry that is enhanced when Q-cohomology jumps in rank as an R-symmetry of the theory.

Let us note that this jumping behavior we are considering is to be contrasted with wall-crossing. Wall-crossing occurs at loci of real codimension one in the moduli space, known as walls of marginal stability. On these walls, the one-particle BPS spectrum is no longer separated from the multiparticle continuum, so that BPS states can decay into BPS constituents. When this occurs, even the index will jump.

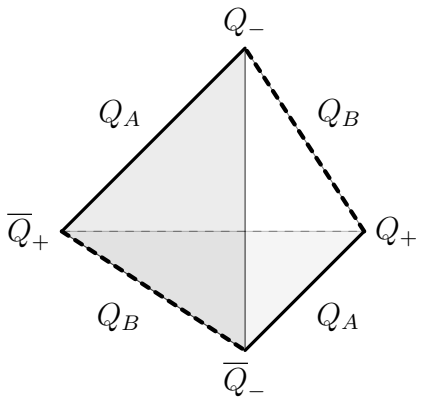


Figure 3.1: The reduced moduli space of scalar supercharges in the 2d $\mathscr{N} = (2,2)$ supersymmetry algebra (a tetrahedron in $\mathbb{R}\mathrm{P}^3$). The nilpotent locus (four of the six one-simplices) is indicated with thick lines.

We have identified two types of possible deformations to study. Let us begin with the first kind, namely, which particular supercharge is chosen. The odd part of a supersymmetry algebra is usually a vector space over $\mathbb{C}$. Since we will be interested in Q-cohomology, scaling a supercharge by a constant does not affect the result, so that the moduli space of possible choices for Q can be thought of as $\mathbb{C}\mathrm{P}^N$ (N being the number of real supercharges of the theory). Within this moduli space, *nilpotent* supercharges will lie on a certain locus, which need not (in fact, will almost never) be the whole of $\mathbb{C}\mathrm{P}^N$. Rather, it is a certain subvariety, described by the equations resulting from the condition that $Q^2 = 0$. Moreover, the nilpotent locus will not parameterize interesting choices uniquely. There are bosonic symmetries of the theory that act nontrivially on $\mathbb{C}\mathrm{P}^N$ and on the nilpotent locus inside it. To parameterize the possible choices of supercharge without redundancy, one must quotient by these symmetries as well. Let us see how this looks in some examples.

Example 1: We begin by considering the $\mathscr{N} = (2,2)$ supersymmetry algebra in 2d. That algebra has four supercharges, and is defined by the commutation relations

$$\{Q_+, \overline{Q}_+\} = \frac{1}{2}(H+P) := H_L \ , \quad \{Q_-, \overline{Q}_-\} = \frac{1}{2}(H-P) := H_R \ ,$$

with all other anticommutators equal to zero.

The odd part of the superalgebra is isomorphic (as a vector space) to $\mathbb{C}^4$, and so the moduli space of supercharges is $\mathbb{C}\mathrm{P}^3$, corresponding to a general supercharge

$$Q = a_+ Q_+ + a_- Q_- + b_+ \overline{Q}_+ + b_- \overline{Q}_-.$$

Using the supersymmetry algebra, we find that

$$Q^2 = a_+ b_+ H_+ + a_- b_- H_-.$$

It follows that Q is nilpotent only when the equations $a_+ b_+ = a_- b_- = 0$ are satisfied. The nilpotent locus therefore consists of four distinct complex lines in $\mathbb{C}\mathrm{P}^3$. Each $\mathbb{C}\mathrm{P}^1$ intersects two of the others (but not the third) in a point; one should think of two horizontal and two vertical lines in the plane, intersecting to form a square. Concretely, the variety has four components:

$$a_+ = a_- = 0, \quad b_+ = b_- = 0; \quad a_+ = b_- = 0, \quad b_+ = a_- = 0.$$

The four intersection points correspond to the four individual supercharges $Q_\pm$ and $\overline{Q}_\pm$.

Now we can ask how the bosonic Lorentz and R-symmetries act on our picture. In this case, the action is simple to describe: it can be used to set all of the phases of the coefficients to zero. The supercharges are U(1) eigenstates for all three symmetries—Lorentz, J_L, and J_R. The three symmetries can be used to set the phases of $a_\pm$ and $b_\pm$ all equal; the remaining overall phase is eliminated upon passing to projective space. As such, after taking the quotient by these symmetries, the reduced moduli space can be identified with the set of points in $\mathbb{R}\mathrm{P}^3$ for which all coordinates are positive. Fig. 3.1 depicts a three-dimensional polyhedral region; in fact, it is a tetrahedron, whose four vertices are the four supercharges $Q_\pm$ and $\overline{Q}_\pm$. (Eight such tetrahedra would comprise the whole of $\mathbb{R}\mathrm{P}^3$.) The nilpotent locus lies along four lines which are edges of the tetrahedron, and fit together in a square: each line intersects two of the other three in a single point.

For our purposes, we will often find it more convenient to describe the unreduced nilpotent locus and remember its symmetries separately.

Example 2: As another example, let us think about the 4d $\mathcal{N} = 1$ supersymmetry algebra. Again, there are four real supercharges, denoted Q_α and $\overline{Q}_{\dot{\alpha}}$ in

usual 4d notations. The supersymmetry algebra is

$$\{Q_\alpha, \overline{Q}_{\dot\alpha}\} = 2P_{\alpha\dot\alpha} \ ,$$

with all other anticommutators zero. If we consider a generic fermionic charge,

$$Q = a_\alpha Q_\alpha + b_{\dot\alpha}\overline{Q}_{\dot\alpha} \ ,$$

it is clear that the nilpotent locus is described by the four equations

$$a_\alpha b_{\dot\alpha} = 0 \ , \quad \alpha, \dot\alpha = 1, 2 \ .$$

These equations require that either $a = 0$ or $b = 0$, resulting in two nonintersecting complex lines in $\mathbb{C}\mathrm{P}^3$. Moreover, the Lorentz group $\mathrm{SU}(2) \times \mathrm{SU}(2)$ acts transitively on each line, so that there are only two points (corresponding to a single supercharge of either chirality) in the reduced moduli space. Thus, the BPS states that contribute to the $\mathscr{N} = 1$ superconformal index [178] are the only possible Q-cohomology in $\mathscr{N} = 1$ theories. (For a related remark on the 4d chiral ring, see §3.1.)

Now, any choice of Q from the nilpotent locus allows one to define the Q-cohomology of the theory. But how does the resulting cohomology vary as Q is chosen from different points along the locus? To answer this question, let us first consider a $\mathbb{C}\mathrm{P}^1$ family of nilpotent supercharges,

$$Q_t = t_1 Q_1 + t_2 Q_2 \ ,$$

associated to two anticommuting nilpotent supercharges Q_1 and Q_2. (This is in fact the situation that applies to each component of the nilpotent locus in Example 1.) A graded Hilbert space equipped with two anticommuting differentials Q_1 and Q_2 is called a *bicomplex* or *double complex* by mathematicians, and the relationship between the cohomology of a single differential and the cohomology of a linear combination of the two has been studied in great detail. It is described by a *spectral sequence*, which provides the answer to our familiar question: How does Q_t-cohomology depend on t?

In order to give intuition about what the spectral sequence is, we will consider this example in great detail. Q_t-cohomology will indeed jump as Q_t varies, and there are essentially three different cases, depending on the anticommutation relations between the supercharges and their adjoints. We treat these from

the least to the most general. For more detail, the reader is referred to [133].

Enhanced supersymmetry: We can ask that the two supercharges, together with their adjoints, close into the $\mathscr{N} = 2$ algebra of supersymmetric quantum mechanics:

$$\{Q_i, Q_j^\dagger\} = 2\delta_{ij} H, \quad \{Q_i, Q_j\} = 0.$$

In this case, the kernel of the positive self-adjoint operator H can be canonically identified with Q_i-cohomology for either supercharge, and therefore with Q_t-cohomology for all $t \in \mathbb{C}\mathrm{P}^1$. There is no jumping behavior.

Simple (but nontrivial) jumping: To see jumping behavior, we can imagine relaxing the condition that the two supercharges square to the same Hamiltonian, while preserving the requirement that $\{Q_1, Q_2^\dagger\} = 0$:

$$\{Q_i, Q_j^\dagger\} = 2\delta_{ij} H_j, \quad \{Q_i, Q_j\} = 0. \tag{3.1}$$

It follows from these relations that $[H_1, H_2] = 0$. Note that this algebra is precisely the 2d super-Poincaré algebra we studied above in Example 1. In this case, the positive operator whose kernel corresponds to Q_t-cohomology is

$$H_t = \{Q_t, Q_t^\dagger\} = 2\left(|t_1|^2 H_1 + |t_2^2| H_2\right).$$

For generic t, only states with $H_1 = H_2 = 0$ appear in Q_t-cohomology. However, there are two special points: the poles $[0:1]$ and $[1:0]$ of $\mathbb{C}\mathrm{P}^1$. At such points, a state needs to saturate only *one* of the two BPS bounds $H_i \geq 0$ to contribute to the cohomology.

Let us recall that points of enhanced degeneracy, where Q_t-cohomology jumps, would be characterized by enhanced symmetry. This is indeed the case here: there is a $\mathrm{U}(1)_R$ symmetry of the algebra (3.1) under which Q_i has charge $(-1)^{i+1}$. This symmetry does not commute with the operator H_t defining the BPS condition, unless t is at one of the poles.

For this argument to hold up, we must assume that we started with a bigraded vector space, and that the Q_i have bidegrees $(1,0)$ and $(0,1)$ respectively. These gradings correspond to left- and right-moving R-charge in Example 1, and it is usual to assume their existence in the mathematical definition of a bicomplex as well.

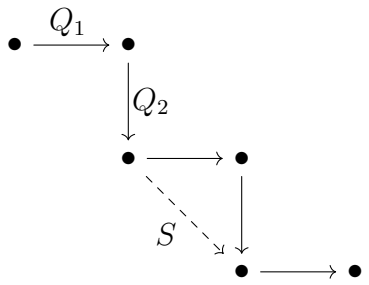


Figure 3.2: A "staircase" representation in a bicomplex

A generic bicomplex: This is the most complicated example, in which it becomes apparent that the notion of a spectral sequence can be nontrivial. We relax all conditions on the commutation relation $\{Q_i, Q_j^\dagger\}$; in particular, the commutation relations no longer *a priori* define a closed algebra. A new bosonic operator $S = \{Q_1, Q_2^\dagger\}$ appears, and there is no way to understand the action of S or express it in terms of other bosonic operators on general grounds.

In addition to the standard long representations and "shortened" representations (in which certain supercharges act by zero), there are now representations of arbitrary dimension that take the form of "staircases" in which S acts nontrivially (Fig. 3.2): The action of S is indicated by the dashed line in the above picture; Q_1 is represented by the horizontal arrows, and Q_2 by the vertical arrows. It is easy to see that the representation drawn in the picture contributes no generators to Q_1-cohomology (or in fact to Q_t-cohomology for $t \neq [0:1]$), but two generators to Q_2-cohomology. Moreover, the two generators are not canceled by the differential induced by Q_1 on Q_2-cohomology, since they do not have the appropriate bidegrees.

The spectral sequence deals with this issue. It consists of a book with infinitely many pages; each page is a bigraded complex equipped with a certain differential, and the $(k+1)$-st page is the cohomology of the k-th page with respect to its differential. One often denotes the pages by (E_k, d_k), where

$$E_k = \bigoplus E_k^{(p,q)}.$$

The E_0 page is the original bicomplex; for the spectral sequence we are considering, d_0 is Q_2, and d_1 is the differential induced by Q_1. The differential d_k has bidegree $(1-k, k)$; it cancels pairs of generators that lie at opposite ends of a staircase of length $2k$. The cohomology of a generic supercharge Q_t

(where t is away from either pole of $\mathbb{C}\mathrm{P}^1$) is the "E_∞ page;" in general, one should be careful about what it means for the spectral sequence to converge, if differentials can occur on infinitely many pages.

In physical language, the bidegree in the original complex corresponds to two commuting U(1) symmetries of the system. Let us call their generators U and V. The original Q_i are the only choices of supercharge that have well-defined quantum numbers with respect to both of these; a generic Q_t has $U+V$-degree $+1$, but breaks the $U-V$ symmetry. This means that, upon deformation, states cancel in adjacent $U+V$-degree, but can differ by *any* value of the $U-V$ quantum number (corresponding to page number in the spectral sequence). However, by studying the deformation in detail, physical understanding may allow us to gain information about which pages' differentials may be nontrivial.

To conclude our discussion of generalities, we should say a couple of words about deformation problems of the other type. So far, we have concentrated on examples on the possibility of deforming the choice of Q in a fixed theory. We could also ask about deforming moduli of the theory: for instance, adding terms to the superpotential in a 2d $\mathscr{N} = (2,2)$ theory. This will also give rise to a spectral sequence, describing the changing spectrum of BPS states; indeed, it is this type of deformation that will give rise to Lee's and Gornik's spectral sequences [98, 138] between link homologies, which we will explore in §3.2.

From elliptic genus to (twisted) chiral ring

Let us see how the rather abstract considerations of the previous section play out concretely in some simple and down-to-earth examples. We will begin by considering a 2d theory of an $\mathscr{N} = (2,2)$ chiral superfield $\Phi = \phi + \theta\psi + \theta^2 F$. The supersymmetry transformations of the chiral multiplet are as follows:

$$\begin{gathered}
[Q_\pm, \phi] = \psi_\pm, \quad [\overline{Q}_\pm, \overline{\phi}] = \overline{\psi}_\pm, \quad [Q_\pm, \overline{\phi}] = 0, \quad [\overline{Q}_\pm, \phi] = 0, \\
\{Q_\pm, \psi_\pm\} = \pm \frac{\partial \overline{W}}{\partial \overline{\phi}}, \quad \{\overline{Q}_\pm, \overline{\psi}_\pm\} = \pm \frac{\partial W}{\partial \phi}, \\
\{\overline{Q}_\pm, \psi_\pm\} = -2i\partial_\pm \phi, \quad \{Q_\pm, \overline{\psi}_\pm\} = 2i\partial_\pm \overline{\phi}.
\end{gathered} \tag{3.2}$$

In (3.1), we have included a nonzero superpotential $W(\Phi)$ for later convenience; our first example, though, will be a free chiral, for which $W = 0$.

	H_L	J_L	K	index
ϕ	0	0	1	x
$\partial_+\overline{\phi}$	0	0	-1	qx^{-1}
ψ_+	$\frac{1}{2}$	-1	1	$-q^{\frac{1}{2}}y^{-1}x$
$\overline{\psi}_+$	$\frac{1}{2}$	1	-1	$-q^{\frac{1}{2}}yx^{-1}$
∂_+	1	0	0	q

Figure 3.3: Single letters in a chiral multiplet annihilated by $\overline{Q}_-$.

It is well-known that the elliptic genus counts Q-cohomology in a 2d $\mathcal{N} = (2,2)$ theory with signs. More precisely, the elliptic genus in the RR sector counts right-moving ground states:

$$I(x;q,y) = \mathrm{Tr}_{RR}(-1)^F q^{H_L}\overline{q}^{H_R}y^{J_L}x^K,$$

where J_L is the left-moving R-charge and K is the generator of a flavor symmetry. Writing the fermion number $F = F_L + F_R$, one can see that only states for which $H_R = 0$ can contribute to the elliptic genus, and therefore that I is a holomorphic function of q. (The same argument does not apply to the left-moving quantum numbers because of the presence of the fugacity y^{J_L}.) These states can be identified with either Q_--cohomology or $\overline{Q}_-$-cohomology. We will choose to consider the cohomology of $\overline{Q}_-$ in what follows.

To see spectral sequence explicitly (and for calculational convenience), we shall calculate the elliptic genus in the NS-NS sector instead:

$$I(x;q,y) = \mathrm{Tr}_{\text{NS-NS}}(-1)^F q^{H_L}y^{J_L}x^K.$$

In superconformal theories, the two indices contain identical information, thanks to the spectral flow. However, in the NS-NS sector, the vacuum state is unique, so that one can straightforwardly count operators without worrying about subtleties due to fermion zero modes. The Ramond-sector calculation is given, for example, in [16, 215].

From the transformations (3.1), one can read off the letters annihilated by $\overline{Q}_-$ that contribute to the elliptic genus [92]. In Table 3.3, we list them with their charges under all relevant symmetries. Each such operator contributes together with all of its left-moving derivatives (which are certain conformal descendants). As a result, all modes (derivatives) from each letter contribute a factor

$$(f;q)_\infty = \prod_{k\geq 0}(1 - fq^k)$$

to the elliptic genus, where f is the product of all fugacities for the field. This factor appears in the numerator for fermions, and in the denominator for bosons.

It is thus easy to see that the elliptic genus of a free chiral is given by

$$I_\chi(q,y;x) = \frac{(q^{1/2}y^{-1}x;q)_\infty (q^{1/2}yx^{-1};q)_\infty}{(x;q)_\infty (qx^{-1};q)_\infty} \ .$$

After a shift of $y \to q^{1/2}y$ (which accounts, up to an overall normalization factor, for spectral flow), we could also write it in the form

$$I_\chi(q,y;x) = y^{-1/2}\frac{\theta_1(q,yx^{-1})}{\theta_1(q,x^{-1})}. \tag{3.3}$$

The Jacobi theta function which appears here is defined in terms of the variables $q = e^{2\pi i\tau}$ and $y = e^{2\pi iz}$ as

$$\theta_1(q,y) := -iq^{1/8}y^{1/2}\prod_{k=1}^{\infty}(1-q^k)(1-yq^k)(1-y^{-1}q^{k-1}) \ .$$

Now, let us now consider deforming the supercharge away from $\overline{Q}_-$, either in the direction of $Q_A = Q_+ + \overline{Q}_-$ or in the direction of $Q_B = \overline{Q}_+ + \overline{Q}_-$ (see Fig. 3.1). Either of these deformations gives rise to a spectral sequence from the elliptic genus to the A- or B-type chiral ring, which moreover collapses at the E_2 page. Thus, we can think of it as simply the differential induced by Q_+ or $\overline{Q}_+$ on the $\overline{Q}_-$-cohomology.

By the 2d $\mathscr{N} = (2,2)$ supersymmetry algebra relations and the Jacobi identity, one can easily see that all derivative operators are Q-exact with respect to either Q_A or Q_B. Therefore only the "single letters" can contribute to the chiral rings. More explicitly, for the twisted chiral ring (Q_A-cohomology) of a free chiral, one can read off from (3.1) that the supercharge (differential) Q_+ acts on the $\overline{Q}_-$-cohomology as

$$Q_+ : \partial_+^k\phi \to \partial_+^k\psi_+ \ , \quad Q_+ : \partial_+^k\bar{\psi}_+ \to \partial_+^{k+1}\bar{\phi} \ , \qquad \text{for} \quad k \in \mathbb{Z}_{\geq 0} \ .$$

Therefore, the twisted chiral ring is trivial, consisting only of the identity operator. On the other hand, for the chiral ring (Q_B-cohomology), the supercharge $\overline{Q}_+$ acts on the $\overline{Q}_-$-cohomology as

$$\overline{Q}_+ : \partial_+^k\psi_+ \to \partial_+^{k+1}\phi, \quad \overline{Q}_+ : \partial_+^{k+1}\overline{\phi} \to \partial_+^{k+1}\overline{\psi}_+ \ , \qquad \text{for} \quad k \in \mathbb{Z}_{\geq 0}.$$

It follows that the chiral ring is generated by the bosonic operator ϕ and the fermionic operator $\xi = \overline{\psi}_+ - \overline{\psi}_-$. It is isomorphic to the graded polynomial ring $\mathbb{C}[\phi] \otimes \Lambda^*[\xi]$ (see Fig. 3.4), and also to the unknot HOMFLY homology, as we will see in §3.2. All in all, non-trivial jumps of Q-cohomology can be seen as we move away from the vertices in Fig. 3.1 in either direction along the nilpotent locus.

It follows from this that the index (3.3) can be written in the form

$$I_\chi(x, y; q) = \frac{1 - y^{-1}x}{1 - x} + q(1 - qy)B(x, y, q) \ , \tag{3.4}$$

where the first term corresponds to the states contributing to the chiral ring and the second term contains the contributions of operators paired by $\overline{Q}_+$. Note that the charges of $\overline{Q}_+$ are $H_L = \frac{1}{2}$ and $J_L = 1$, so that the action of Q_+ corresponds to the factor $(1 - qy)$. By an identical argument, we could also write

$$I_\chi(x; y, q) = 1 + (1 - qy^{-1})A(x, y, q) \ , \tag{3.5}$$

using the spectral sequence from the elliptic genus to the (trivial) twisted chiral ring, which just consists of the differential induced from Q_+.

It is worth remarking that the fugacity parameter y in the elliptic genus records the charges of states under J_L. When we consider the spectral sequence from the elliptic genus to the chiral ring, the interpretation of y changes slightly. States in the A- or B-type chiral rings are graded by only one surviving U(1) R-symmetry: either axial or vector. The remainder is lost to the A- or B-type topological twist. Powers of y in (3.4) and (3.5) above record the appropriate remaining R-symmetry properties; however, the identification of y with a vector or axial R-symmetry may require a charge conjugation, $y \leftrightarrow y^{-1}$.

Deformation by superpotential

One of the attractive features of the elliptic genus is that it is invariant along renormalization group flow trajectories, and therefore independent of the superpotential W of a Landau-Ginzburg model, with one minor exception: W determines the allowable R-charges of the fields in the superconformal IR theory, and therefore prescribes the graded structure of the index. With this subtlety in mind, the computation for a theory with a Landau-Ginzburg description is identical to that for a free theory.

The reader may be tempted to conclude that no interesting spectral sequence can therefore appear, relating elliptic genera as the superpotential is deformed. But this is not true! While the elliptic genus is always described by the same rational function, it depends on the R-charge assignment as well as on the fugacities, and cancellations occur (or fail to occur) between the numerator and denominator, depending on the precise values of the R-charges. These cancellations reflect the disappearance of states from the $\overline{Q}_-$-cohomology.

To convince oneself, one can simply recall that the elliptic genus in the Ramond sector contains the graded dimension of the chiral ring: the latter corresponds, by spectral flow, to the set of Ramond vacua, which are counted by the elliptic genus after setting $q = 0$. However, the chiral ring depends strongly on W, and spectral sequences of type 2 are therefore easy to see. The calculation is easy to do directly in a theory with one chiral superfield. If one turns on the superpotential $W = \phi^{N+1}$, corresponding to a renormalization group flow from a free theory in the UV to the A_N superconformal minimal model in the IR, the supersymmetry transformations are perturbed. In particular, the fermionic operator $\xi = \overline{\psi}_+ + \overline{\psi}_-$ that contributes to the elliptic genus transforms according to the rule

$$\{Q_B, \overline{\psi}_+ + \overline{\psi}_-\} = 2\frac{\partial W}{\partial \phi} = 2\phi^N. \tag{3.6}$$

The new differential induced on $\overline{Q}_B$-cohomology is completely characterized by (3.6). Therefore, in the presence of the superpotential $W = \phi^{N+1}$, the chiral ring of the theory is isomorphic to $\mathbb{C}[\phi]/(\phi^N)$. In the context of link homology, turning on such a superpotential perturbs the spectrum of BPS states, as described by the d_N differential.

We can again write the elliptic genus in a form corresponding to (3.4), as the graded dimension of the chiral ring plus an additional term corresponding to operators that pair up after turning on the superpotential $W = \phi^{N+1}$:

$$\frac{1 - y^{-1}x}{1 - x} \quad \xrightarrow{W=\phi^{N+1}} \quad (1 + x + \cdots + x^{N-1}) + \frac{x^N - y^{-1}x}{1 - x}\,. \tag{3.7}$$

The different fermionic numerator in (3.7)—as contrasted with (3.4)—corresponds to the different degree assignments for the new differential. A helpful picture of the spectral sequence from the unperturbed to the perturbed chiral ring is shown in Fig. 3.4. This is exactly the d_N differential [72, 173] on the HOMFLY

homology of the unknot where the perturbed chiral ring is isomorphic to the $\mathfrak{sl}(N)$ homology of the unknot.

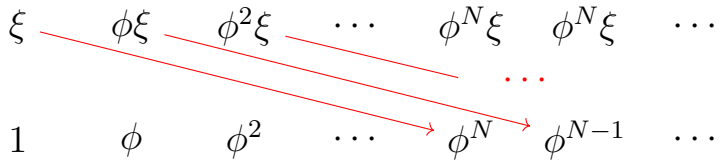


Figure 3.4: The action of the differential induced by turning on the superpotential $W = \phi^{N+1}$ on the B-type chiral ring of a single free chiral superfield.

It is interesting to notice that, after the superpotential interactions have been turned on, only bosonic states remain in the BPS spectrum (or B-model chiral ring). This leads to an important difference between perturbations corresponding to the d_N differentials and perturbations $W = \phi^{N+1} \to \phi^{N+1} + \phi^{M+1}$ that carry the $\mathfrak{sl}(N)$ theory to the $\mathfrak{sl}(M)$ theory for $M < N$. The spectral sequences induced by the latter deformations can only be trivial, since any differential must cancel bosonic and fermionic states in pairs. This means that, under this perturbation, we obtain an A_M-type Landau-Ginzburg theory together with some isolated massive vacua, which decouple in the IR limit (see Fig. 3.15). As we will see in §3.2, these factors appear in deformation spectral sequences, such as Gornik's, as trivial $\mathfrak{sl}(1)$ factors. For this reason, Gornik's spectral sequence never has any interesting differentials at all in the unknot homology.

For an example of a different type, let us look at one more interesting BPS spectral sequence, related to the 2d $\mathcal{N} = (2,2)$ superconformal gauged linear sigma model [213]. (We include this example, as well as the discussion in §3.1, mostly to convince the reader that our discussion here is not limited to the world of knots.) To that end, we write the on-shell supersymmetric transformation of a vector multiplet

$$V = \theta^-\bar{\theta}^- v_- + \theta^+\bar{\theta}^+ v_+ - \theta^-\bar{\theta}^+\sigma - \theta^+\bar{\theta}^-\bar{\sigma} + i\theta^2\bar{\theta}\bar{\lambda} + i\bar{\theta}^2\theta\lambda + \theta^2\bar{\theta}^2 D$$

as

$$\begin{aligned} [\overline{Q}_\mp, v_\pm] = \pm i\lambda_\pm, \quad & [Q_\mp, v_\pm] = \pm i\overline{\lambda}_\pm, \\ [\overline{Q}_-, \sigma] = -i\lambda_-, \quad [Q_+, \sigma] = i\overline{\lambda}_+, \quad & [Q_-, \overline{\sigma}] = -i\overline{\lambda}_-, [\overline{Q}_+, \overline{\sigma}] = i\lambda_+, \\ [\overline{Q}_+, \sigma] = 0, \quad [Q_-, \sigma] = 0, \quad & [Q_+, \overline{\sigma}] = 0, \quad [\overline{Q}_-, \overline{\sigma}] = 0, \end{aligned}$$

	H_L	J_L	index
$\partial_+\sigma$	$\frac{1}{2}$	1	$q^{3/2}y$
$\overline{\sigma}$	$\frac{1}{2}$	-1	$q^{1/2}y^{-1}$
λ_+	1	0	$-q$
$\overline{\lambda}_+$	1	0	$-q$
∂_+	1	0	q

Figure 3.5: Single letters in a vector multiplet annihilated by $\overline{Q}_-$.

$$\begin{aligned} &\{Q_+,\lambda_+\} = -2\partial_+\overline{\sigma}, \quad \{Q_-,\lambda_-\} = 2\partial_-\sigma, \quad \{\overline{Q}_\pm,\lambda_\pm\} = \partial_\sigma\widetilde{\mathscr{W}}, \\ &\{\overline{Q}_+,\overline{\lambda}_+\} = -2\partial_+\sigma, \quad \{\overline{Q}_-,\overline{\lambda}_-\} = 2\partial_-\overline{\sigma}, \quad \{Q_\pm,\overline{\lambda}_\pm\} = \partial_{\overline{\sigma}}\widetilde{\mathscr{W}}. \end{aligned} \tag{3.8}$$

The single letters that contribute to the elliptic genus are summarized in Table 3.5. Up to a shift of y, the elliptic genus of a vector multiplet can be written as

$$I_V(q,y) = \frac{(q;q)_\infty(q;q)_\infty}{(y^{-1};q)_\infty(qy;q)_\infty} = \frac{i\eta(q)^3}{\theta_1(q,y)} \ ,$$

where we use the Dedekind eta function

$$\eta(\tau) = q^{1/24}\prod_{k=1}^{\infty}(1-q^k) \ .$$

Now, let us study the simplest example of a superconformal gauged linear sigma model: the $T^*\mathbb{C}\mathrm{P}^1$ model, which is a U(1) gauge theory with three chirals Φ_1,Φ_2,Φ_3, with gauge charges $(1,1-2)$ respectively. The Higgs branch of the theory is the total space of the $\mathcal{O}(-2)$ line bundle (i.e. cotangent bundle) over $\mathbb{C}\mathrm{P}^1$. Introducing a flavor fugacity a_i for each chiral, the elliptic genus of the theory [16, 92] is

$$\begin{aligned} I_{T^*\mathbb{C}\mathrm{P}^1}(q,y;a_i) &= \frac{i\eta(q)^3}{\theta_1(q,y)}\int\frac{dz}{2\pi iz}\frac{\theta_1(q,z^{-1}a_1y)}{\theta_1(q,z^{-1}a_1)}\frac{\theta_1(q,z^{-1}a_2y)}{\theta_1(q,z^{-1}a_2)}\frac{\theta_1(q,z^2a_3y)}{\theta_1(q,z^2a_3)} \\ &= \frac{1}{2}\sum_{k,\ell=0}^{1}y^{-\ell}\frac{\theta_1(q,e^{2\pi i(k+\ell\tau)/2}a_1a_3^{1/2}y)}{\theta_1(q,e^{2\pi i(k+\ell\tau)/2}a_1a_3^{1/2})}\frac{\theta_1(q,e^{2\pi i(k+\ell\tau)/2}a_2a_3^{1/2}y)}{\theta_1(q,e^{2\pi i(k+\ell\tau)/2}a_2a_3^{1/2})} \ . \end{aligned}$$

After integrating out the gauge fugacity z, the elliptic genus counts all gauge-invariant operators comprised of letters from the three chirals and the U(1) vector multiplet.

Next, we deform the supercharge $\overline{Q}_-$ to either Q_A or Q_B. It is clear from (3.1) and (3.8) that there is no non-trivial Q_B-cohomology in the vector multiplet,

and the Q_B-cohomology consists of the gauge invariant operators that can be written in terms of the following letters:

$$\phi_1\ ,\quad \phi_2\ ,\quad \phi_3\ ,\quad \overline{\psi}_{1+}-\overline{\psi}_{1-}\ ,\quad \overline{\psi}_{2+}-\overline{\psi}_{2-}\ ,\quad \overline{\psi}_{3+}-\overline{\psi}_{3-}.$$

Although it is tedious, one can convince oneself that the other $\overline{Q}_-$-cohomology states are paired by $\overline{Q}_+$. As we have seen above, the generating function of the chiral ring can be obtained by taking the $q=0$ limit in the part of the elliptic genus that the chiral multiplets contribute. More explicitly, we can write the generating function as

$$I_{T^*\mathbb{C}\mathrm{P}^1\,\chi\text{ ring}}=y^{-1}+\frac{1}{2}\sum_{i=0,1}\frac{(1+(-1)^i a_1 a_3^{1/2}y)(1+(-1)^i a_2 a_3^{1/2}y)(1-y)}{y^3(1+(-1)^i a_1 a_3^{1/2})(1+(-1)^i a_2 a_3^{1/2})}.$$

Similarly, the supercharge Q_+ pairs all the Q_--cohomology states except the twisted chiral operator $\bar{\sigma}$. Because of the commutation relation $\{Q_+,\bar{\lambda}_+\}=\partial_{\bar{\sigma}}\widetilde{\mathscr{W}}$, the twisted chiral ring depends on the twisted superpotential $\widetilde{\mathscr{W}}$, which is known to receive quantum corrections. Actually, a one-loop computation yields the effective twisted superpotential of this model [213]:

$$\widetilde{\mathscr{W}}_{\text{eff}}(\bar{\sigma})=-t\bar{\sigma}-\sum_i q_i(\bar{\sigma}-a_i)(\log(q_i(\bar{\sigma}-a_1))-1),$$

where t is the complexified FI parameter and q_i are gauge charges. By solving the vacuum equation

$$\exp\left(\frac{\partial\widetilde{\mathscr{W}}_{\text{eff}}(\bar{\sigma})}{\partial\bar{\sigma}}\right)=1,$$

we obtain the twisted chiral ring of the $T^*\mathbb{C}\mathrm{P}^1$ model

$$\frac{(\bar{\sigma}-a_1)(\bar{\sigma}-a_2)}{4(\bar{\sigma}-a_3)^2}=e^{-t}\ .$$

Since the effective twisted superpotential of a gauged linear sigma model is dynamically generated, one can consider that the Q_A-cohomology can be obtained from the $\overline{Q}_-$-cohomology via a *dynamical BPS spectral sequence*.

Remark on 4d chiral ring

We would like to briefly take note of an important, but subtle, distinction between various constructions used in building invariants (Q-cohomology or chiral rings) of quantum field theories. Specifically, every construction we have mentioned so far—the (twisted) chiral rings in 2d and, BPS states that

contribute to the elliptic genus [16, 92, 215] and 4d $\mathscr{N} = 1$ index [178]—takes the form of a Q-cohomology.

A 4d analogue of the chiral ring has also been considered in the literature (see *e.g.* [38] and references therein). It is defined as the collection of operators annihilated by *all* supercharges of one chirality, modulo any operator that is exact with respect to *any* such supercharge. That is,

$$\chi = (\cap_\alpha \text{Ker } Q_\alpha) \,/\, (\oplus_\alpha \text{Im } Q_\alpha) \ .$$

However, the 4d chiral ring cannot be thought of as the Q-cohomology of *any* supercharge, be it any of the Q_α's or any combination.

This is easy to see in an example. Consider a chain complex with three generators, one bosonic in degree zero and two fermionic in degree one. Let the action of two supercharges be defined by

$$Q_1\phi = \psi_1 \ , \quad Q_2\phi = \psi_2 \ , \quad Q_i\psi_j = 0 \ .$$

Then both Q_i-cohomologies, as well as (Q_1+Q_2)-cohomology, are one-dimensional, with support in degree one. Moreover, the Euler characteristic of any possible Q-cohomology must equal that of the complex itself, which is -1.

However, the chiral ring is empty, and has Euler characteristic zero! This means that (despite its superficially similar construction) the chiral ring is not any kind of Q-cohomology, and its behavior cannot be expected to follow the same pattern.

Note that upon dimensional reduction to 2d, the 4d $\mathscr{N} = 1$ algebra becomes the $\mathscr{N} = (2,2)$ algebra, and the two supercharges Q_α of the same chirality descend to the two supercharges that constitute Q_B in the reduced theory. So it is tempting to suspect that the cohomology of $\sum_\alpha Q_\alpha$ would give the 4d chiral ring. But we have seen in two ways that this cannot be the case: by a direct argument in the preceding paragraph, and in §3.1 by arguing that any two linear combinations of the Q_α are related by a Lorentz transformation in 4d, and so do not give different choices of Q in the reduced nilpotent moduli space. When we dimensionally reduce to 2d, we break the Lorentz symmetry from $\text{SO}(4) = \text{SU}(2) \times \text{SU}(2)$ to $\text{SO}(2) \times \text{SO}(2)$, and the quotient by this smaller symmetry no longer identifies all choices of supercharge. This is why it is possible to describe the chiral ring by a Q-cohomology in the dimensionally reduced theory.

BPS spectral sequences in 4d $\mathscr{N}=2$ SCFTs

In this subsection, we give an example of how the generalities we have been discussing apply in a setting that is quite far removed from the world of Landau-Ginzburg models and link homologies, and therefore from the mainstream of this paper. We do so both because we find the example interesting in its own right, and in order to highlight (and convince the reader of) the breadth and generality of the framework we have been discussing. We will explain how various limits of the 4d $\mathscr{N}=2$ superconformal index [93] can be understood as arising from BPS spectral sequences. To be self-contained, we will briefly review the analysis in [93]; for more details, we refer the reader to the original paper. The indices and relationships between them that we discuss here are not novel; rather, the novelty consists in our interpretation of these results, and the way in which many facts can be organized—and easily and systematically understood—in our picture.

The fermionic part of the 4d $\mathscr{N}=2$ superconformal algebra $\mathrm{SU}(2,2|2)$ can be expressed as

$$\{Q^i_\alpha, \overline{Q}_{j\dot\alpha}\} = 2\delta^i_j P_{\alpha\dot\alpha}, \qquad \{Q^i_\alpha, Q^j_\beta\} = \{\overline{Q}_{i\dot\alpha}, \overline{Q}_{j\dot\beta}\} = 0, \tag{3.9}$$

$$\{\overline{S}^{i\dot\alpha}, S^\alpha_j\} = 2\delta^i_j K^{\dot\alpha\alpha}\ , \qquad \{S^\alpha_i, S^\beta_j\} = \{\overline{S}^{i\dot\alpha}, \overline{S}^{j\dot\beta}\} = 0\ , \tag{3.10}$$

$$\{Q^i_\alpha, \overline{S}^{j\dot\alpha}\} \quad = \quad \{S^\alpha_i, \overline{Q}_{j\dot\alpha}\} = 0, \tag{3.11}$$

as well as

$$\begin{aligned} \{Q^i_\alpha, S^\beta_j\} &= 4(\delta^i_{\ j}(M_\alpha^{\ \beta} - \tfrac{i}{2}\delta_\alpha^{\ \beta} D) - \delta_\alpha^{\ \beta} R^i_{\ j})\ , \\ \{\overline{S}^{i\dot\alpha}, \overline{Q}_{j\dot\beta}\} &= 4(\delta^i_{\ j}(\overline{M}^{\dot\alpha}_{\ \dot\beta} + \tfrac{i}{2}\delta^{\dot\alpha}_{\ \dot\beta} D) - \delta^{\dot\alpha}_{\ \dot\beta} R^i_{\ j})\ . \end{aligned} \tag{3.12}$$

The 4d $\mathscr{N}=2$ superconformal index counts the 1/8-BPS states annihilated by one supercharge and its superconformal partner (adjoint), say $\overline{Q}_{1\dot-}$ and $\overline{S}^{1\dot-}$. In other words, it counts $\overline{Q}_{1\dot-}$-cohomology, and these states saturate the bound

$$\bar\delta_{1\dot-} := \{\overline{S}^{1\dot-}, \overline{Q}_{1\dot-}\} = E - 2j_2 - 2R + r = 0\ .$$

Recalling that R-symmetry indices are raised and lowered with the antisymmetric invariant tensor ϵ^{ij}, it is easy to see from (3.9) that the three supercharges

$$Q_{1-}\ , \qquad Q_{1+}\ , \qquad \overline{Q}_{2\dot+}\ ,$$

commute with $\overline{Q}_{1\dot-}$ and $\overline{S}^{1\dot-}$, so that the $\mathscr{N}=2$ superconformal index can be defined by introducing fugacities ρ, σ, and τ for the corresponding bosonic

generators of the commutant subalgebra SU(1, 1|2):

$$\mathcal{I}(\rho,\sigma,\tau) = \mathrm{Tr}(-1)^F \rho^{\frac{1}{2}\delta_{1-}} \sigma^{\frac{1}{2}\delta_{1+}} \tau^{\frac{1}{2}\tilde{\delta}_{2\dot{+}}} e^{-\beta\,\tilde{\delta}_{1\dot{-}}} . \tag{3.13}$$

Here the δ's can be read off from (3.12):

$$\begin{aligned} \delta_{1-} &:= \{Q_{1-}, S^{1-}\} = E - 2j_1 - 2R - r \ , \\ \delta_{1+} &:= \{Q_{1+}, S^{1+}\} = E + 2j_1 - 2R - r \ , \\ \overline{\delta}_{2\dot{+}} &:= \{\overline{S}^{2\dot{+}}, \overline{Q}_{2\dot{+}}\} = E + 2j_2 + 2R + r \ . \end{aligned}$$

The usual parametrization in terms of fugacities (p, q, t) is related to (3.13) via the change of variables

$$p = \tau\sigma \,, \quad q = \tau\rho \,, \quad t = \tau^2 \,. \tag{3.14}$$

These parameters gives the (p, q) labels of the elliptic Gamma function as well as the (q, t) variables for Macdonald indices. In terms of these fugacities, the index can be expressed as

$$\mathcal{I}(p,q,t) = \mathrm{Tr}(-1)^F p^{\frac{1}{2}\delta_{1+}} q^{\frac{1}{2}\delta_{1-}} t^{R+r} e^{-\beta\,\tilde{\delta}_{1\dot{-}}} \ . \tag{3.15}$$

As we did for the elliptic genus in §3.1 and §3.1, one can list single letters for the $\mathcal{N} = 2$ vector multiplet and half-hypermultiplet that contribute to the index:

Letters	E	j_1	j_2	R	r	$\mathcal{I}(\sigma,\rho,\tau)$	$\mathcal{I}(p,q,t)$
ϕ	1	0	0	0	-1	$\sigma\rho$	pq/t
$\lambda_{1\pm}$	$\frac{3}{2}$	$\pm\frac{1}{2}$	0	$\frac{1}{2}$	$-\frac{1}{2}$	$-\sigma\tau,\ -\rho\tau$	$-p,\ -q$
$\bar{\lambda}_{1\dot{+}}$	$\frac{3}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\tau^2$	$-t$
$\bar{F}_{\dot{+}\dot{+}}$	2	0	1	0	0	$\sigma\rho\tau^2$	pq
$\partial_{-\dot{+}}\lambda_{1+} + \partial_{+\dot{+}}\lambda_{1-} = 0$	$\frac{5}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$\sigma\rho\tau^2$	pq
q	1	0	0	$\frac{1}{2}$	0	τ	$\sqrt{t}$
$\bar{\psi}_{\dot{+}}$	$\frac{3}{2}$	0	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\sigma\rho\tau$	$-pq/\sqrt{t}$
$\partial_{\pm\dot{+}}$	1	$\pm\frac{1}{2}$	$\frac{1}{2}$	0	0	$\sigma\tau,\ \rho\tau$	$p,\ q$

Figure 3.6: Contributions to the index from "single letters," with the equation of motion $\partial_{-\dot{+}}\lambda_{1+} + \partial_{+\dot{+}}\lambda_{1-} = 0$. We denote the components of the adjoint $\mathcal{N} = 2$ vector multiplet by $(\phi, \bar{\phi}, \lambda_{I,\alpha}, \bar{\lambda}_{I\,\dot{\alpha}}, F_{\alpha\beta}, \bar{F}_{\dot{\alpha}\dot{\beta}})$ and the components of the half-hypermultiplet by $(q, \bar{q}, \psi_\alpha, \bar{\psi}_{\dot{\alpha}})$. The letters $\partial_{\alpha\dot{\alpha}}$ represent the spacetime derivatives.

The authors of [93] have considered four limits of the index shown in Table 3.7. These limits count states subject to enhanced BPS shortening conditions; after

taking them, the index becomes simpler and can often be expressed in terms of certain special functions.

	Fugacities	states annihilated by
Macdonald index	$\sigma \to 0$	$Q_{1+},\ \overline{Q}_{1\dot{-}}$
Hall-Littlewood index	$\sigma \to 0,\ \rho \to 0$	$Q_{1\pm},\ \overline{Q}_{1\dot{-}}$
Schur index	$\rho = \tau$	$Q_{1+},\ \overline{Q}_{1\dot{-}}$
Coulomb-branch index	$\tau \to 0$	$\overline{Q}_{2\dot{+}},\ \overline{Q}_{1\dot{-}}$

Figure 3.7: Various limits of the $\mathscr{N} = 2$ index. The middle column lists the specializations of fugacities. The right column indicates the supercharges that annihilate states contributing to the corresponding index.

In fact, several (but not all) of these limits are related by BPS spectral sequences to the original index. To begin, there is a BPS spectral sequence from $\overline{Q}_{1\dot{-}}$-cohomology to the states annihilated by Q_{1+} and $\overline{Q}_{1\dot{-}}$ (that contribute to the Macdonald index). We can identify these states with the cohomology of the perturbed supercharge $Q_{1+} + \overline{Q}_{1\dot{-}}$.

To see this spectral sequence explicitly, note that the supersymmetry transformations of the $\mathscr{N} = 2$ vector multiplet can be written as follows:

$$\begin{aligned} \delta_\epsilon A_\mu &= i\epsilon_i \sigma_\mu \overline{\lambda}^i - i\lambda_i \sigma_\mu \overline{\epsilon}^i \ , \\ \delta_\epsilon \lambda_i &= F_{\mu\nu}\sigma^{\mu\nu}\epsilon_i + \sqrt{2} i D_\mu \phi\, \sigma^\mu \overline{\epsilon}_i + D\epsilon_i \ , \\ \delta_\epsilon \phi &= \sqrt{2}\epsilon_i \lambda^i \ . \end{aligned}$$

We also recall the transformations of the half-hypermultiplet, which are

$$\begin{aligned} \delta_\epsilon q_i &= -\sqrt{2}\epsilon_i \psi + \sqrt{2}\overline{\epsilon}_i \overline{\psi} \ , \\ \delta_\epsilon \psi &= -\sqrt{2} i D_\mu q_i \sigma^\mu \overline{\epsilon}^i - 2F^i \epsilon_i \ . \end{aligned}$$

From these supersymmetry transformations, one can see that the supercharge Q_{1+} acts on the $\overline{Q}_{1\dot{-}}$-cohomology by

$$\begin{aligned} &Q_{1+} : \partial^k_{\pm\dot{+}} \phi \to \partial^k_{\pm\dot{+}} \lambda_{1+} \ , \qquad Q_{1+} : \partial^k_{\pm\dot{+}} \overline{F}_{\dot{+}\dot{+}} \to \partial^k_{\pm\dot{+}} \partial_{+\dot{+}} \overline{\lambda}_{1\dot{+}} \ , \\ &Q_{1+} : \partial^k_{\pm\dot{+}} \overline{\psi}_{\dot{+}} \to \partial^k_{\pm\dot{+}} \partial_{+\dot{+}} q \ , \end{aligned}$$

for $k \in \mathbb{Z}_{k\geq 0}$. Thus, the single letters that contribute to the Macdonald index consist of λ_-, $\overline{\lambda}_{\dot{+}}$, q and $\partial_{-\dot{+}}$ which satisfy $\delta_{1+} = 0$. At the level of the $\mathscr{N} = 2$ index, one can write

$$\mathscr{I}_{\text{full}}(\rho, \sigma, \tau) = \mathscr{I}_{\text{Mac}}(\rho, \tau) + \sigma(1 - \rho^{-1}\tau) A(\rho, \sigma, \tau) \ .$$

Since the supercharge Q_{1+} has $\delta_{1-} = -2$ and $\delta_{2\dot{+}} = 2$, the spectral sequence due to Q_{1+} corresponds to the factor $(1 - \rho^{-1}\tau)$. This is why the Schur limit $\rho = \tau$ also leads to the same 1/4-BPS condition as the Macdonald limit $\sigma \to 0$ does.

There is also a BPS spectral sequence from $\overline{Q}_{1\dot{-}}$-cohomology to the states annihilated by $\overline{Q}_{2\dot{+}}$ and $\overline{Q}_{1\dot{-}}$ (that contribute to the Coulomb-branch index). For the two kinds of multiplet we are considering, the action of $\overline{Q}_{2\dot{+}}$ on $\overline{Q}_{1\dot{-}}$-cohomology is

$$\begin{aligned} &\overline{Q}_{2\dot{+}} : \partial^k_{\pm\dot{+}}\lambda_{1\pm} \to \partial^{k+1}_{\pm\dot{+}}\phi \ , \qquad \overline{Q}_{2\dot{+}} : \partial^k_{\pm\dot{+}}\overline{\lambda}_{1\dot{+}} \to \partial^k_{\pm\dot{+}}\overline{F}_{\dot{+}\dot{+}} \ , \\ &\overline{Q}_{2\dot{+}} : \partial^k_{\pm\dot{+}}q \to \partial^k_{\pm\dot{+}}\overline{\psi}_{\dot{+}} \ , \end{aligned}$$

for $k \in \mathbb{Z}_{k\geq 0}$. Thus, ϕ is the only single letter that contributes to the Coulomb-branch index. At the level of the index, the relation between the full $\mathscr{N} = 2$ index and the Coulomb-branch index can be written as

$$\mathscr{I}_{\text{full}}(\rho, \sigma, \tau) = \mathscr{I}_{\text{Coulomb}}(\sigma, \rho) + \tau(1 - \sigma\rho)B(\rho, \sigma, \tau) \ ,$$

where the action of $\overline{Q}_{2\dot{+}}$ amounts to the factor $(1 - \sigma\rho)$, since the supercharge $\overline{Q}_{2\dot{+}}$ has $\delta_{1-} = 2$ and $\delta_{2\dot{+}} = 2$. This implies that the Coulomb-branch index can be also obtained by taking the specialization $\sigma = \rho^{-1}$ in the full $\mathscr{N} = 2$ index.

Moreover, by checking the commutation relations (3.9) and (3.12) above, the reader may check using the general methodology we outlined in §3.1 that both of these spectral sequences must collapse at the E_2 page; that is, there are no interesting higher differentials.

In contrast with this, the states satisfying the 3/8-BPS condition and contributing to the Hall-Littlewood index cannot be thought of as the Q-cohomology of any choice of supercharge. Analogous to what we have seen for the 4d chiral ring in §3.1, the action of the Lorentz group relates the supercharges $Q_{1\pm}$, so that the cohomology of $Q_{1+} + Q_{1-} + \overline{Q}_{1\dot{-}}$ must be identical to that of just $Q_{1+} + \overline{Q}_{1\dot{-}}$. Thus, the Hall-Littlewood index is not the graded dimension of any Q-cohomology, and one cannot see a spectral sequence from $\overline{Q}_{1\dot{-}}$-cohomology (or from the Macdonald index) to the 3/8-BPS states. Indeed, it is easy to see that the supercharge Q_{1-} must act trivially on states contributing to the Macdonald index in another way: it carries charge $\delta_{1+} = -2$ for the fugacity σ—but this has already been set to zero.

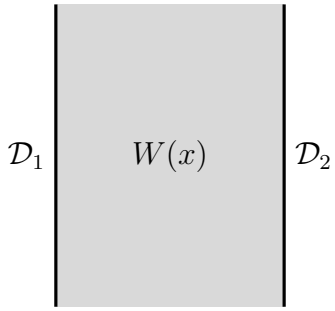


Figure 3.8: A Landau-Ginzburg model on a strip

LG models on a strip

Let us now consider yet another example of jumping behavior of BPS spectra under deformations of a theory, which will bring us back towards our goal of understanding spectral sequences between knot homologies. In contrast with those in §3.1, this example is a spectral sequence of the second type (i.e. associated to moduli of the theory, rather than to a choice of supercharge). Its novelty consists in two features: (1) we are considering an RG flow trajectory between two interacting theories, and (2) we include certain boundary conditions or defects. The context will still be that of Landau-Ginzburg theories, but we will consider open-string rather than closed-string BPS states.

Let us consider the three-parameter family of Landau-Ginzburg superpotentials

$$W(x) = (x - u_1)(x - u_2)(x - u_3) \ , \tag{3.16}$$

corresponding to the relevant perturbation of the A_2 minimal model $W = x^3$. We are interested in the spectrum of open-string states when the LG model described by $W(x)$ is placed on a strip (Fig. 3.8). Each side has a boundary condition, which may be chosen independently. Possible choices of boundary condition that preserve B-type supersymmetry in Landau-Ginzburg theories are described by matrix factorizations of the superpotential.

A matrix factorization $\mathcal{D}$ is a $\mathbb{Z}_2$-graded free module over a polynomial ring R, equipped with an odd endomorphism $Q_{\rm bd}$ that squares to a polynomial potential $W \in R$:

$$Q_{\rm bd} : \mathcal{D} \to \mathcal{D}, \quad Q_{\rm bd}^2 = W \cdot {\rm id}_{\mathcal{D}} \, .$$

The variables of R correspond to the chiral superfields of the model, and $Q_{\rm bd}$ specifies the action of the unbroken supercharge Q_B on the boundary degrees

of freedom. If the matrix factorization has rank one, so that $\mathcal{D} = R_0 \oplus R_1$, we can write Q_{bd} in components as

$$Q_{\text{bd}} = \begin{pmatrix} 0 & f \\ g & 0 \end{pmatrix} .$$

One could also write the matrix factorization as

$$R_0 \xrightarrow{f} R_1 \xrightarrow{g} R_0 .$$

Every choice of superpotential admits a trivial rank-one matrix factorization, for which the boundary supercharge is simply

$$R_0 \xrightarrow{1} R_1 \xrightarrow{W} R_0 .$$

We will often regard two matrix factorizations as equivalent if one can be obtained from the other by taking the direct sum with some number of trivial matrix factorizations.

We will denote the space of maps from $\mathcal{D}_1$ to $\mathcal{D}_2$ (considered as graded R-modules) by $\text{Mat}(\mathcal{D}_1, \mathcal{D}_2) \cong \mathcal{D}_1 \otimes \overline{\mathcal{D}}_2$; it can be thought of as a space of matrices. Furthermore, the boundary supercharge Q_{bd} for the tensor product $\mathcal{D}_1 \otimes \overline{\mathcal{D}}_2$ makes $\text{Mat}(\mathcal{D}_1, \mathcal{D}_2)$ into a $\mathbb{Z}/2\mathbb{Z}$-graded complex. The cohomology $\mathcal{H}^*(\mathcal{D}_1, \mathcal{D}_2)$ of this complex is the space of open-string BPS states between these two boundary conditions. As we will see in §3.2, the generators in $\mathcal{H}^*(\mathcal{D}_1, \mathcal{D}_2)$ are also called *defect-changing operators*.

Let us see how the spectrum of open-string BPS states jumps in the parameter space of the potential (3.16). We explicitly specify the two boundary conditions for the potential (3.16) as

$$\begin{aligned} \mathcal{D}_1 : \qquad & R_0 \xrightarrow{x-u_1} R_1 \xrightarrow{(x-u_2)(x-u_3)} R_0 \\ \mathcal{D}_2 : \qquad & R_0 \xrightarrow{x-u_2} R_1 \xrightarrow{(x-u_1)(x-u_3)} R_0 , \end{aligned}$$

where $R = \mathbb{C}[x]$. Note that these boundary conditions are unobstructed: they make sense over the whole parameter space of the superpotential.

We then calculate their tensor product

$$\mathcal{D}_1 \otimes \overline{\mathcal{D}}_2 : \quad \begin{bmatrix} R \\ \bigoplus \\ R \end{bmatrix}_0 \xrightarrow{d^0} \begin{bmatrix} R \\ \bigoplus \\ R \end{bmatrix}_1 \xrightarrow{d^1} \begin{bmatrix} R \\ \bigoplus \\ R \end{bmatrix}_0 ,$$

where the maps are defined

$$d^0 = \begin{bmatrix} x-u_1 & u_2-x \\ -(x-u_1)(x-u_3) & (x-u_2)(x-u_3) \end{bmatrix}, \quad d^1 = \begin{bmatrix} x-u_2 & x-u_2 \\ x-u_1 & x-u_1 \end{bmatrix}.$$

The cohomology of this system is

$$\begin{aligned} \mathcal{H}^0(\mathcal{D}_1,\mathcal{D}_2) &= \text{Ker } d^0/\text{Im } d^1 \cong \begin{cases} \mathbb{C}[x]/(x-u) & \text{for} \quad u_1=u_2=u \\ 1 & \text{for} \quad u_1 \neq u_2 \end{cases}, \\ \mathcal{H}^1(\mathcal{D}_1,\mathcal{D}_2) &= \text{Ker } d^1/\text{Im } d^0 \cong \begin{cases} \mathbb{C}[x]/(x-u) & \text{for} \quad u_1=u_2=u \\ 1 & \text{for} \quad u_1 \neq u_2 \end{cases}. \end{aligned}$$

Therefore, at the locus $u_1 = u_2$ (of complex codimension one) in the parameter space (u_1, u_2, u_3), we can see a jump of the BPS spectrum. Away from the locus, the BPS states disappear in pairs, one bosonic and one fermionic, just as we expected would happen.

It is straightforward to generalize this result to the perturbation of A_N minimal models. Suppose that the potential is

$$W(x) = \prod_{u_i \in S} (x-u_i) \quad \text{for } |S| = N,$$

and we have the two boundary conditions

$$\begin{aligned} Q^{(1)}_{\text{bd}} &= \begin{bmatrix} 0 & \prod_{u_i \in U_2}(x-u_i) \\ \prod_{u_i \in U_1}(x-u_i) & 0 \end{bmatrix} \\ Q^{(2)}_{\text{bd}} &= \begin{bmatrix} 0 & \prod_{u_i \in V_2}(x-u_i) \\ \prod_{u_i \in V_1}(x-u_i) & 0 \end{bmatrix}, \end{aligned}$$

where $U_1 \cup U_2 = V_1 \cup V_2 = S$. (These unions are not necessarily disjoint, since roots will occur in the set S with some multiplicity.) The BPS spectrum turns out to be

$$\mathcal{H}^0(\mathcal{D}_1,\mathcal{D}_2) \cong \mathcal{H}^1(\mathcal{D}_1,\mathcal{D}_2) \cong \frac{\mathbb{C}[x]}{\prod_{U_1 \cap U_2 \cap V_1 \cap V_2}(x-u_i)}. \tag{3.17}$$

It is thus easy to see that the jumping phenomena of this type are ubiquitous in LG model on a strip when one varies parameters of the potential, corresponding to relevant deformations of A_N-series minimal models.

There is one other important point to note. We stated that the space $\mathcal{H}(\mathcal{D}_1,\mathcal{D}_2)$ of open-string BPS states should be taken as the space of morphisms in the category of B-type boundary conditions. However, our calculation (3.17) shows

that this space can in fact empty, even when $\mathcal{D}_1$ and $\mathcal{D}_2$ are the same defect! For instance, consider the following boundary condition:

$$\mathcal{D} : R_0 \xrightarrow{(x-u)^{p+1}} R_1 \xrightarrow{(x-v)^{q+1}} R_0 \ . \tag{3.18}$$

The relevant superpotential is clearly $W = (x-u)^{p+1}(x-v)^{q+1}$. When $u = v$, $\mathcal{D}$ is a nontrivial rank-one boundary condition in the A_{p+q+1} minimal model, and

$$\dim \mathcal{H}^0(\mathcal{D}, \mathcal{D}) = \min(p+1, q+1) > 0 \ .$$

In fact, we would expect that it must be, since any category must admit at least the identity morphism from an object to itself. Notice, however, that if either boundary condition were the trivial matrix factorization ($p = -1$), the space of morphisms would be zero-dimensional. This is in keeping with our claim above that the trivial defect "decouples" from the actual category of boundary conditions: it has zero morphisms with any object, even itself.

Now let us consider deforming u and v away from one another. Our calculation (3.17) shows that as soon as this is done, $\dim \mathcal{H}^0(\mathcal{D}, \mathcal{D}) = 0$! Not even an identity morphism is present. Nevertheless, this is not a problem, and in fact should have been expected. The theory described by the perturbed superpotential no longer has a unique vacuum state. Rather, it has two vacua; one corresponding to an A_p conformal theory and the other to an A_q conformal theory. The theory splits into two superselection sectors, which do not interact. Therefore, one ought to think of the boundary condition (3.18) as representing a kind of composite of the trivial defect in the A_p theory and the trivial defect in the A_q theory. The idea that boundary conditions form a category applies to the collection of boundary conditions in a fixed theory; the idea breaks down in the presence of superselection sectors.

Moreover, models of this type actually include LG models on a cylinder, with any number of defects extended in the time direction! This will be exactly the setup that we find in the context of knot homology. For example, consider a LG model with two defects on a cylinder. If the two domains separated by the defects have different potentials in the same variable (chiral field) x, i.e. $W_1(x)$ and $W_2(x)$, this is equivalent to LG model on a strip with the potential $W_1(x) - W_2(x)$ by using the folding trick as in Fig. 3.9b. Thus, interesting spectral sequences occur in this model under deformation of the potentials, in contrast to the unknot case where the two domains have the same potential of

different variables, $W(x)$ and $W(y)$, as in Fig. 3.12. Even with more than two defects (as in Fig. 3.11), the LG model on a cylinder undergoes the spectral sequence under the deformations of superpotentials if all the potentials are of the same variable x. As we will explain later, one can fuse the defects together one by one, until only two remain.

3.2 Fivebranes and links

In this section, we study homological invariants of knots and links using the bird's-eye view of fivebrane systems [59, 108, 112, 168, 218]. The physical realizations of link homologies predict their rich structural properties, which are accomplished by action of differentials and algebras [72, 99, 100, 113, 173]. In particular, we build a new vantage point in terms of Landau-Ginzburg model in the fivebrane systems for link homologies §3.2. In this viewpoint, the considerations in §3.1 will give a physical understanding of Lee-Gornik spectral sequences [98, 138] of $\mathfrak{sl}(N)$ link homology in §3.2. Since our focus will change and in fact grow much more specific in what follows, the reader can think of this section as an introduction to the second part of the paper, which will investigate the color-dependence of link homology.

There are several good reasons why the focus of knot theory in the twenty-first century has shifted from polynomial invariants to their categorified (*i.e.* homological) versions [56]. By definition, a categorification of a (Laurent) polynomial with r variables is a $(r+1)$-graded homology theory whose graded Euler characteristic is the polynomial in question. As such, it is usually a stronger invariant of knots and links, since some information is lost in passing to the Euler characteristic. More importantly, there are many operations one can do with vector spaces that are simply impossible at the polynomial level (maps, for instance), which opens a door into the beautiful world of homological algebra.

The quantum group invariants $P^{G}_{\lambda_1,\ldots,\lambda_n}(L;q)$ computed by Chern-Simons TQFT [176, 212] provide an infinite set of polynomial link invariants that depend on the choice of gauge group G, representations λ_i of G (one for each link component), a variable q (related to the coupling constant of Chern-Simons theory), and, of course, the link L itself. All these polynomial invariants have a remarkable property: they are (Laurent) polynomials in variable q with integer coefficients. Therefore, they should be categorified by doubly graded homology

theories. Luckily, by now these homology theories have been constructed [53, 85, 131, 205, 221, 224] for any choice of G and $\vec{\lambda} = (\lambda_1, \ldots, \lambda_n)$, of which the most familiar example is probably Khovanov homology [128] corresponding to $G = \mathrm{SU}(2)$ and $\lambda_i = \square$. However, how exactly do these invariants depend on the "representation theory data" G and $\vec{\lambda} = (\lambda_1, \ldots, \lambda_n)$?

The dependence of $P^G_{\vec{\lambda}}(L; q)$ on the group G turns out to be very simple for classical groups of Cartan type A, B, C, or D. For each of these root systems, the dependence on the rank is beautifully packaged by a 2-variable polynomial invariant: the colored HOMFLY polynomial $P_{\vec{\lambda}}(L; a, q)$ for Cartan type A and the so-called colored Kauffman polynomial $F_{\vec{\lambda}}(L; a, q)$ for groups of type B, C, or D. Thus, for $G = \mathrm{SU}(N)$ one has

$$P^{\mathrm{SU}(N)}_{\lambda_1,\ldots,\lambda_n}(L; q) \;=\; P_{\lambda_1,\ldots,\lambda_n}(L; a = q^N, q)\ . \tag{3.19}$$

This means that quantum group invariants for different groups are not that different after all; there is a lot of regularity captured by the extra variable a. Since for each choice of λ_i the colored HOMFLY polynomial $P_{\lambda_1,\ldots,\lambda_n}(L; a, q)$ depends on two variables, a and q, its categorification must be a triply-graded homology theory. The existence [112] and properties [72] of this homology theory came from physics as a surprise, contrary to the expectations in math literature; see *e.g.* [129]. In the "uncolored" case, *i.e.* when all $\lambda_i = \square$, the triply-graded HOMFLY homology has been given a rigorous mathematical definition [132, 206]. Moreover, the Poincaré polynomial of the colored HOMFLY homology (often called colored *superpolynomial*)

$$\mathscr{P}_{\vec{\lambda}}(L; a, q, t) \;:=\; \sum_{i,j,k} a^i q^j t^k \dim \mathscr{H}_{\vec{\lambda}}(L)_{i,j,k} \tag{3.20}$$

is, roughly speaking, an intermediate object: much like colored HOMFLY polynomial, it is still a polynomial (with positive integer coefficients), but it captures more information and provides a window into a homological world.

However, the construction of triply-graded homology to general groups with arbitrary "color" label $\vec{\lambda} = (\lambda_1, \ldots, \lambda_n)$ as well as the computations of their Poincaré polynomials (even for simple knots) remain a major challenge. This makes any predictions especially valuable. In fact, using the perspective of fivebrane systems, many predictions and conjectures on structure of colored knot homologies have been made [72, 99, 113, 115, 161], which enables us to determine colored superpolynomials of many knots [89–91, 160, 161]. The

structure of knot homology appears as a large set of differentials, called d_N and colored differentials, which should be formulated in terms of spectral sequences [173]. As such, we first provide a physical meaning of spectral sequences in link homology based on the understanding in §3.1.

From the next section, we will study the structure of link homology as well as the "color"-dependence $\vec{\lambda} = (\lambda_1, \ldots, \lambda_n)$. Actually, there is an important difference between HOMFLY invariants of knots and links. Unlike the reduced colored HOMFLY polynomial for a knot which is a genuine Laurent polynomial, in the case of links it has a nontrivial denominator. Since the corresponding homology theory that categorifies it is consequently infinite-dimensional, colored HOMFLY homology of links loses some of properties for that of knots, which makes the analysis more difficult. However, the link homology turns out to enjoy surprising regularities in its dependence on color $\vec{\lambda} = (\lambda_1, \ldots, \lambda_n)$, that can be described in terms of *sliding properties*, *homological blocks*, and *associated varieties*. The reader is referred to the published version [109] for more on these results.

Link homology as Q-cohomology

In every extant physical approach to homological link invariants, they are realized as spaces of BPS states annihilated by a supercharge Q (modulo Q-exact states), which are also known as BPS states [67, 108, 112, 218] (see also [1] for torus knots):

$$\mathscr{H}_{\vec{\lambda}}(L) \;\cong\; \{Q\text{-cohomology}\} \;\equiv\; \mathscr{H}_{\rm BPS}\,. \tag{3.21}$$

Moreover, in every physical realization of link homologies all spatial dimensions are effectively compact, so that the system reduces to supersymmetric quantum mechanics in one non-compact "time" direction $\mathbb{R}_t$.

All physical approaches to link homologies are essentially different ways to look at the same physical system, which has two phases—"deformed conifold" and "resolved conifold"—related by the *geometric transition* [97]. The former describes doubly-graded $\mathfrak{sl}(N)$ link homologies for fixed N, whereas the latter reproduces triply-graded HOMFLY homologies:

deformed conifold	$\Leftrightarrow$	doubly-graded $\mathfrak{sl}(N)$ homology
resolved conifold	$\Leftrightarrow$	triply-graded HOMFLY homology

On the "deformed conifold" side the physical setup is:

$$\begin{array}{lccccc} \text{space-time:} & \mathbb{R}_t\times & TN_4 & \times & T^*S^3 \\ N \text{ M5-branes:} & \mathbb{R}_t\times & D & \times & S^3 \\ \text{M5}'\text{-branes:} & \mathbb{R}_t\times & D & \times & M_L \end{array} \qquad \text{(def-M5)}$$

where $TN_4 \cong \mathbb{R}^4_{\epsilon_1,\epsilon_2}$ is the Taub-NUT 4-manifold, $D \cong \mathbb{R}^2_{\epsilon_1}$ is the two-dimensional "cigar" (holomorphic Lagrangian submanifold) in the Taub-NUT space, and the Lagrangian 3-manifold $M_L \subset T^*S^3$ is the conormal bundle to the link $L \subset S^3$, such that

$$L = S^3 \cap M_L \ .$$

After the geometric transition (i.e. on the "resolved conifold" side) the corresponding system is

$$\begin{array}{lccccc} \text{space-time:} & \mathbb{R}_t & \times & TN_4 & \times & X \\ \text{M5}'\text{-branes:} & \mathbb{R}_t & \times & D & \times & M_L \end{array} \qquad \text{(res-M5)}$$

The effective 3d $\mathcal{N} = 2$ theory on $\mathbb{R}_t \times D$ is what is often called $T[M_L]$; it is a 3d $\mathcal{N} = 2$ theory labeled by a 3-manifold M_L, or equivalently by the link L [49, 55, 63–65, 67, 89–91, 139, 194, 222].

To be more precise, the BPS states in question are the so-called *open* BPS states $\mathscr{H}(L) \cong \mathscr{H}^{\text{open}}_{\text{BPS}}$, meaning that they are represented by bordered Riemann surfaces in X with boundary on M_L. In contrast, the *closed* BPS states are represented by membranes supported on closed holomorphic curves in X, with no boundary. In other words, the space of open BPS states is determined by the pair (X, M_L), whereas the space of closed BPS states depends only on X. It has been conjectured [117] that closed BPS states form an algebra $\mathcal{A}$. Similarly, in [113] it was argued that open BPS states furnish a module M for this algebra:

$$\begin{array}{rl} \text{(refined) open BPS states :} & M := \mathscr{H}^{\text{open}}_{\text{BPS}} \\ & \circlearrowleft \\ \text{(refined) closed BPS states :} & \mathcal{A} := \mathscr{H}^{\text{closed}}_{\text{BPS}} \end{array} \qquad (3.22)$$

In fact, this relation between closed and open BPS states is supposed to hold in general, regardless of the specific nature of the Calabi-Yau 3-fold X and the Lagrangian subvariety M_L. (In this general context, the notion of *refined* BPS states exists only for rigid X.) However, in application to knots, the action of $\mathcal{A}$ on triply-graded homology (3.21) is especially useful and accounts for

certain differentials that act on colored HOMFLY homology [113]. Although at present one can only identify certain elements of $\mathcal{A}$ but not the entire algebra, it is expected to be related to the *double affine Hecke algebra* [46] and the *rational Cherednik algebra* [100] for torus knots. (See also [159] for a review.)

Both systems (def-M5) and (res-M5) enjoy a $\mathrm{U}(1)_P \times \mathrm{U}(1)_F$ symmetry that acts on TN_4 and gives rise to two gradings (= conserved charges), q-grading and t-grading*:

$$TN_4 \cong \mathbb{R}^4_{\epsilon_1,\epsilon_2} , \qquad D \cong \mathbb{R}^2_{\epsilon_1} , \qquad q = e^{\epsilon_1} , \qquad t = -e^{-(\epsilon_1+\epsilon_2)} . \tag{3.23}$$

In addition, the system (res-M5) has $\mathrm{U}(1)_a$ gauge symmetry in five dimensions $\mathbb{R}_t \times TN_4$ that gives rise to the a-grading of HOMFLY homology. Indeed, compactification of eleven-dimensional M-theory on the resolved conifold X is a simple example of *geometric engineering* of SUSY gauge theories [126], that in the present case engineers a super-Maxwell theory on $\mathbb{R}_t \times TN_4$ with gauge group $\mathrm{U}(1)_a$.

The Omega-background [162] associated with the isometry $\mathrm{U}(1)_P \times \mathrm{U}(1)_F$ implements equivariant localization to the fixed point, the origin $\{0\} \in TN_4$. As a result, the effective theory on the branes in (def-M5) and (res-M5) is $(0+1)$-dimensional quantum mechanics with only one supercharge Q and its hermitian conjugate $Q^\dagger$:

$$\{Q, Q^\dagger\} \;=\; 2H.$$

There are two standard ways to describe knots and links in this setup: either via Wilson / 't Hooft line operators (that are naturally labeled by representations of $G = \mathrm{SU}(N)$) or via monodromy defects (that are labeled by G-valued holonomies). The system (def-M5), where knots and links are introduced via M5′-branes supported on $\mathbb{R}_t \times D \times M_L$, provides a realization of the monodromy defects.

In order to obtain a realization based on Wilson and 't Hooft operators one needs to replace the set of M5′-branes in (def-M5) by M2′-branes (a.k.a. mem-

*Up to exchange of ϵ_1 and ϵ_2, this identification and its relation with the refinement of topological string [122], the Nekrasov-Shatashvili limit [163], *etc.* is explained in [90, §4]. Note that the unrefined limit corresponds to $\epsilon_1 = -\epsilon_2$, *i.e.* $t = -1$.

branes):

$$\begin{array}{lccccc} \text{space-time:} & \mathbb{R}_t\times & TN_4 & \times & T^*S^3 \\ N \text{ M5-branes:} & \mathbb{R}_t\times & D & \times & S^3 \\ \text{M2}'\text{-branes:} & \mathbb{R}_t\times & \{0\} & \times & \Sigma_L \end{array} \qquad \text{(def-M2)}$$

where Σ_L is the total space of $T^*L \subset T^*S^3$ that meets the zero section, S^3, along the link L. Interestingly, these two systems are connected by the Hannay-Witten effect [59, 84]; when we push the M5′-branes to the cotangent fiber of T^*S^3, the M2′-branes are dynamically generated.

Now, if we perform the geometric transition starting with the system (def-M2), then on the "resolved conifold" side we get

$$\begin{array}{lccccc} \text{space-time:} & \mathbb{R}_t & \times & TN_4 & \times & X \\ \text{M2}'\text{-branes:} & \mathbb{R}_t & \times & \{0\} & \times & \Sigma_L \end{array} \qquad \text{(res-M2)}$$

which seems to be a new duality frame, little studied in the literature.[†]

Our goal is to use the physical systems (def-M5) and (def-M2) (resp. (res-M5) and (res-M2)) to explain the structural properties and, ideally, compute the doubly-graded $\mathfrak{sl}(N)$ homologies (resp. the triply-graded HOMFLY homologies) denoted by $\mathscr{H}_{\vec{\lambda}}^{\mathfrak{sl}(N)}(L)$ (resp. $\mathscr{H}_{\vec{\lambda}}(L)$) in (3.21).

Effective quantum mechanics

In all of the four duality frames, equivariant localization (*i.e.* the Omega-background in the directions of TN_4) effectively reduces the theory to a quantum mechanics with "time" direction $\mathbb{R}_t$ and two real supercharges, Q and $Q^\dagger$. The space of supersymmetric states in this effective quantum mechanics is the desired space $\mathscr{H}(L)$, which is invariant under isotopies of $L \subset S^3$ due to the topological twist along the S^3 directions. Because of this, in (def-M5) and (def-M2) we could, in fact, replace S^3 by *any* 3-manifold M_3 without breaking any of the symmetries or supersymmetries. In particular, one could still associate a Hilbert space $\mathscr{H}(M_3, L)$ of supersymmetric states to a link $L \subset M_3$ defined as Q-cohomology, which should be an exciting direction for future research. In this paper, we focus on the simplest case $M_3 = S^3$ merely

[†]Part of the reason this system is more subtle than its counterpart (res-M5) involving M5-branes is that it is more difficult to "lift" M2-branes off the M5-branes in (def-M2). The resulting surface Σ_L in (res-M2) has asymptotic boundary $\partial\Sigma_L = L \subset S^3 \subset \partial X$ and presumably can be constructed along the lines of [193].

for simplicity and in order to keep contact with current mathematical developments, which at this stage are limited to colored link homologies in S^3.

As explained *e.g.* in [108], closing up the time direction $\mathbb{R}_t$ into a circle corresponds to "decategorification." In other words, replacing $\mathbb{R}_t$ by S^1 in (def-M5), (res-M5), (def-M2), and (res-M2) gives physical systems effectively compact in all directions. Therefore, instead of producing the Hilbert space $\mathscr{H}(L)$ as in $0+1$ dimensional quantum mechanics, this 0-dimensional system computes the partition function, which is the graded Poincaré polynomial of $\mathscr{H}(L)$. In duality frames (res-M5) and (res-M2) it yields the colored superpolynomial (3.20), whereas in duality frames (def-M5) and (def-M2) it gives the Poincaré polynomial of the doubly-graded $\mathfrak{sl}(N)$ link homology,

$$\bar{\mathscr{P}}_{\vec{\lambda}}^{\mathfrak{sl}(N)}(L;q,t) \;:=\; \sum_{i,j} q^i t^j \dim \bar{\mathscr{H}}_{\vec{\lambda}}^{\mathfrak{sl}(N)}(L)_{i,j}\,.$$

The richness of physics does not stop with the four duality frames (def-M5), (res-M5), (def-M2), and (res-M2). In each of the duality frames, one can look at the Q-cohomology (3.21) in a number of different ways. (See [113, 159] for overviews of different vantage points.) For example, looking at it from the vantage point of the Calabi-Yau 3-fold gives a description of $\mathscr{H}(L)$ in terms of enumerative invariants. A look from the vantage point of $\mathbb{R}_t \times TN_4$ gives a description of the space $\mathscr{H}(L)$ in terms of gauge theory with a surface operator. Focusing on the physics of the surface operator leads to yet another equivalent description of the same system (and hence the space $\mathscr{H}(L)$) as the space of BPS states in the 3d $\mathscr{N}=2$ theory $T[M_L]$ on $\mathbb{R}_t \times D$. In particular, this provides an identification of the effective quantum mechanics in question with a 3d-5d coupled system in the Omega-background. Schematically,

$$\begin{array}{ccc} \text{Quantum mechanics} & \cong & \text{3d } \mathscr{N}=2 \text{ theory } T[M_L] \\ \text{on } \mathbb{R}_t & & \text{on } \mathbb{R}_t \times \mathbb{R}^2_{\epsilon_1} \end{array}$$

where we used the identification (3.23).

Link homology and fusion of defect lines in LG models

Note that in (def-M5), (res-M5), and (def-M2) the most "interesting" part of the setup (where the physical degrees of freedom that contribute to $\mathscr{H}(L)$ "live") is actually the four-dimensional factor

$$\mathbb{R}_t \times \mathbb{R}^2_{\epsilon_1} \times L\,,$$

where we used (3.23) to emphasize that $D \cong \mathbb{R}^2_{\epsilon_1}$ is effectively compact due to Omega-background. (For instance, the equivariant volume $\mathrm{Vol}(D) = \frac{1}{\epsilon_1}$ is finite for finite values of ϵ_1.) In the setup (res-M2) the physical degrees of freedom that contribute to $\mathscr{H}(L)$ live on $\mathbb{R}_t \times L$ (so that D does not even appear as part of their support). In either of these cases, what is the effective theory on $\mathbb{R}_t \times L \cong \mathbb{R}_t \times S^1$?

In the deformed conifold side with either the probe M5′-branes (def-M5) or the probe M2′-branes (def-M2), it was argued [113, 115] that the answer to this question involves Landau-Ginzburg (LG) models that appear [130–132] in Khovanov-Rozansky formulation of link homologies. Indeed, in order to understand the effective theory on $\mathbb{R}_t \times L$ away from crossings, one can choose L to be a piece of strand or an unknot that runs, say, along the great circle (the "equator")

$$S^1_\sigma \subset S^3\,, \tag{3.24}$$

parametrized by a (periodic) coordinate σ (Fig. 3.13). Then, the effective theory on $\mathbb{R}_t \times L$ describes a surface operator coupled to a 4d topological field theory on $\mathbb{R}_t \times S^3$ [48]. And, if L covers S^1_σ k times, the physics on $\mathbb{R}_t \times S^1_\sigma$ away from crossings can be described by a LG model with the superpotential

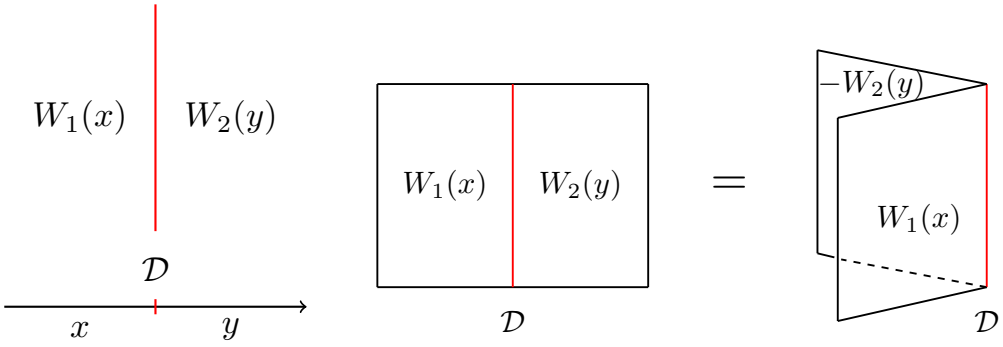
$$W = W_{\lambda_1} + \ldots + W_{\lambda_k},$$

where λ_i's are representations ("colors") carried by the strands, *cf.* Fig. 3.13. For instance, when $\lambda = \square$ is the fundamental representation one recovers the potential

$$W = x^{N+1} \tag{3.25}$$

used by Khovanov-Rozansky in construction of $\mathfrak{sl}(N)$ link homology [131]. We will use the physical realization of Khovanov-Rozansky homology [131, 132] to understand the physical origin/meaning of the differentials and spectral sequences connecting different link homology to one another.

In the construction of link homology given by Khovanov-Rozansky, a cone of two matrix factorizations is assigned to each crossing of a knot, as we shall review in greater detail in what follows. Recall from §3.1 that matrix factorizations represent B-type boundary conditions in Landau-Ginzburg models. Moreover, defects (interfaces) can also be represented by matrix factorizations [36], via the *folding trick* (Fig. 3.9b). A B-type defect $\mathcal{D}$ joints together a LG theory with chiral superfields x_i and a superpotential $W_1(x_i)$, and a different



(a) A defect between two LG models

(b) The folding trick

Figure 3.9: Topological defects in two-dimensional theories.

LG theory with superfields y_i and a superpotential $W_2(y_i)$ (Fig. 3.9a). The B-type defect is described in the language of a matrix factorization with the potential $W(x_i, y_i) = W_1(x_i) - W_2(y_i)$:

$$R_0 \xrightarrow{d_+} R_1 \xrightarrow{d_-} R_0 \ , \qquad d_+ \circ d_- = d_- \circ d_+ = W_1(x_i) - W_2(y_i) \ ,$$

where R_0 and R_1 are the polynomial ring $\mathbb{C}[x_i, y_i]$.

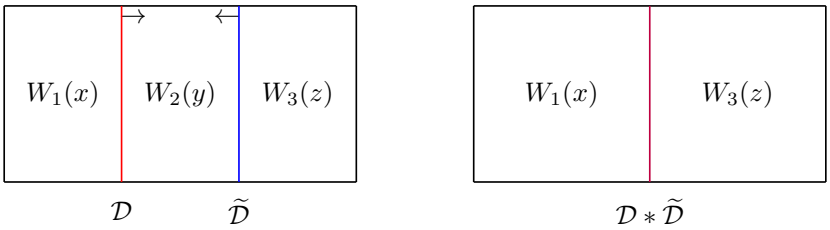


Figure 3.10: Fusion of topological defects, which is described by a tensor product of the corresponding matrix factorizations.

Since these defects are topological, one can move and distort them in general. In particular, two defects $\mathcal{D}$ and $\widetilde{\mathcal{D}}$ can be put on top of each other, creating a new defect $\mathcal{D} * \widetilde{\mathcal{D}}$, which is usually called a *fusion* of the defects (Fig. 3.10). The fusion amounts to taking the tensor product of the matrix factorizations $\mathcal{D}$ and $\widetilde{\mathcal{D}}$ defined by

$$\mathcal{D} \otimes \widetilde{\mathcal{D}} : \quad \begin{bmatrix} R_0 \otimes \widetilde{R}_0 \\ \bigoplus \\ R_1 \otimes \widetilde{R}_1 \end{bmatrix}_0 \xrightarrow{d^0} \begin{bmatrix} R_1 \otimes \widetilde{R}_0 \\ \bigoplus \\ R_0 \otimes \widetilde{R}_1 \end{bmatrix}_1 \xrightarrow{d^1} \begin{bmatrix} R_0 \otimes \widetilde{R}_0 \\ \bigoplus \\ R_1 \otimes \widetilde{R}_1 \end{bmatrix}_0 , \tag{3.26}$$

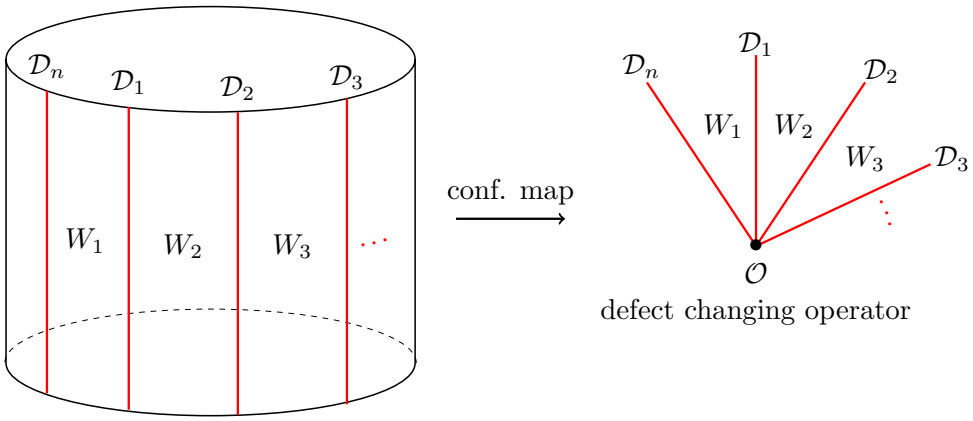


Figure 3.11: Landau-Ginzburg theory with defects on a cylinder. Via the state-operator correspondence, the BPS states of the LG model can be identified with the defect-changing operators.

where the differentials are defined by

$$d^0 = \begin{bmatrix} d_+ \otimes \mathrm{id} & \mathrm{id} \otimes \widetilde{d}_+ \\ -\mathrm{id} \otimes \widetilde{d}_- & d_- \otimes \mathrm{id} \end{bmatrix}, \quad d^1 = \begin{bmatrix} d_- \otimes \mathrm{id} & -\mathrm{id} \otimes \widetilde{d}_+ \\ \mathrm{id} \otimes \widetilde{d}_- & d_+ \otimes \mathrm{id} \end{bmatrix},$$

which satisfies $d^0 \circ d^1 = d^1 \circ d^0 = W_1(x) - W_3(z)$. In fact, the resulting matrix factorization is generally of infinite rank. Physically speaking, the chiral fields y squeezed in between the two defects are promoted to new degrees of freedom on the new defects when the two defects coincide. However, if both $\mathcal{D}$ and $\widetilde{\mathcal{D}}$ are of finite rank, the matrix factorization $(\mathcal{D} \otimes \widetilde{\mathcal{D}})_{\mathbb{C}[x,z]}$ can be reduced to finite rank by throwing away an infinite number of trivial matrix factorizations. After this reduction, we denote the resulting matrix factorization by

$$\mathcal{D} * \widetilde{\mathcal{D}} := (\mathcal{D} \otimes \widetilde{\mathcal{D}})^{\mathrm{red}}_{\mathbb{C}[x,z]} .$$

If there are n defects with the superpotentials $W_1, \ldots, W_n$ and $W_{n+1} = W_1$, one can think that the LG model with the defects is placed on a cylinder. The tensor product $\mathcal{C} = \otimes_i \mathcal{D}_i$ is a tensor product of the matrix factorizations $\mathcal{D}_i$ of the potentials $W_i - W_{i+1}$; since $d^2 = \sum_{i=1}^n W_i - W_{i+1} = 0$, it is a complex. The homology $\mathcal{H}^*(\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_n)$ of this complex is the BPS spectrum of the theory. By the state-operator correspondence, the BPS spectrum can be identified with the spectrum of operators at the junctions of the defects [36] (Fig. 3.11).

The LG model with defects on a cylinder is very akin to the Khovanov-Rozansky construction. For example, the unknot can be represented as a clo-

sure of braid on one strand, *i.e.* as two semi-circles glued together (Fig. 3.12.a). The first semi-circle is represented by a matrix factorization $\mathcal{D}_1$ of $W(x)-W(y)$

$$R_0 \xrightarrow{x-y} R_1 \xrightarrow{p} R_0 \ , \tag{3.27}$$

and the second semi-circle corresponds to a matrix factorization $\mathcal{D}_2$ of $W(y)-W(x)$

$$R_0 \xrightarrow{y-x} R_1 \xrightarrow{p} R_0 \ , \tag{3.28}$$

where $R=\mathbb{C}[x,y]$ and

$$p=p(x,y)=\frac{W(x)-W(y)}{x-y} \ .$$

The homology of the unknot is given as the homology of the tensor product (3.26) of the matrix factorizations (3.27) and (3.28):

$$\begin{bmatrix} R \\ \oplus \\ R \end{bmatrix}_0 \xrightarrow{d^0} \begin{bmatrix} R \\ \oplus \\ R \end{bmatrix}_1 \xrightarrow{d^1} \begin{bmatrix} R \\ \oplus \\ R \end{bmatrix}_0 , \qquad d^0=\begin{bmatrix} x-y & y-x \\ -p & p \end{bmatrix}, \quad d^1=\begin{bmatrix} p & x-y \\ p & x-y \end{bmatrix}.$$

It is easy to see that $d^0\circ d^1=d^1\circ d^0=0$. Consequently, a simple computation yields

$$H^0=\mathrm{Ker}\, d^0/\mathrm{Im}\, d^1\cong \mathbb{C}[x]/W'(x)\ , \quad H^1=\mathrm{Ker}\, d^1/\mathrm{Im}\, d^0\cong \begin{cases} 0 & \text{for } p\neq 0 \\ \mathbb{C}[x] & \text{for } p=0 \end{cases} .$$

With the potential $W(x)=x^{N+1}$ as in (3.25), we obtain the $\mathfrak{sl}(N)$ knot homology of the unknot

$$\overline{\mathscr{H}}^{\mathfrak{sl}(N)}_{\square}(\bigcirc)\cong \mathbb{C}[x]/(x^N). \tag{3.29}$$

The q-grading of the chiral field x is two, so that the Poincaré polynomial is

$$\overline{\mathscr{P}}^{\mathfrak{sl}(N)}_{\square}(\bigcirc)=1+q^2+\cdots+q^{2N-2} \ .$$

When $W=0$, we have the HOMFLY homology of the unknot. In this case, both H^0 and H^1 are nontrivial and isomorphic to $\mathbb{C}[x]$, so the homology is supported in two different homological degrees. In particular, the Poincaré polynomial is

$$\overline{\mathscr{P}}_{\square}(\bigcirc)=\frac{1}{1-q^2}+\frac{a^2t}{1-q^2} \ ,$$

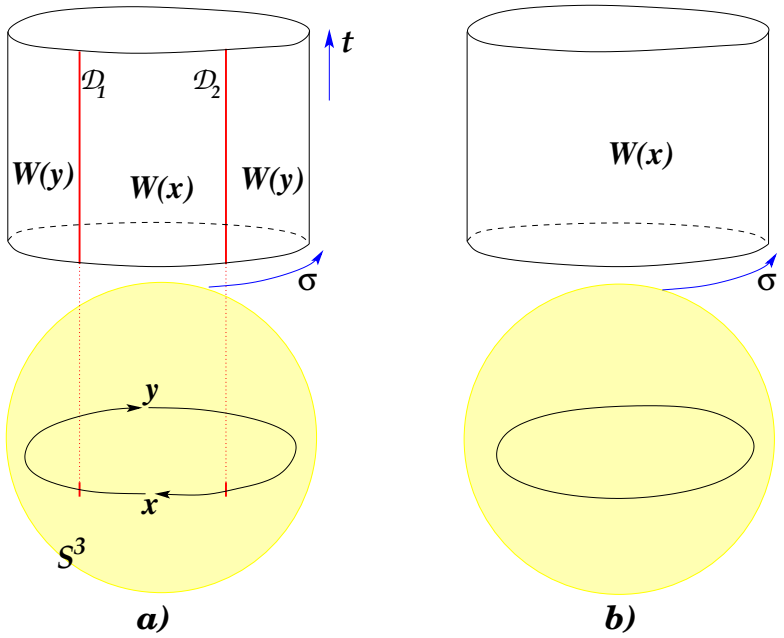


Figure 3.12: The Landau-Ginzburg model for the unknot. Since $\mathcal{D}_2$ is the inverse of $\mathcal{D}_1$, they can annihilate each other leaving the Landau-Ginzburg model without any defect lines.

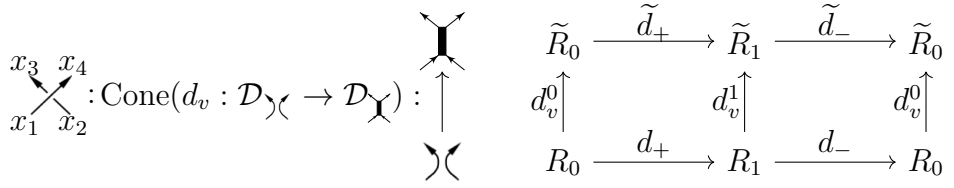
where the first term corresponds to homological degree zero whereas the second term corresponds to homological degree one as well as the a-grading is shifted by two. On the resolved side (res-M5), the free $\mathcal{N} = 2$ chiral superfield x comes from M2-branes ending on M5-branes.

In fact, in the above discussion $\mathcal{D}_2 = \bar{\mathcal{D}}_1$ is the inverse of the defect $\mathcal{D}_1$, so that $\mathcal{D}_1$ and $\mathcal{D}_2$ can annihilate each other leaving behind a LG theory with superpotential W on a cylinder without any defect lines. (Fig. 3.12) Hence, we have $\mathcal{H}(\mathcal{D}_1, \mathcal{D}_2) \cong \mathrm{Jac}(W)$, the Jacobi ring of W (or space of closed-string BPS states). The effective theory on the most "interesting" part of the unknot theory (which is the product $\mathbb{R}_t \times S^1_\sigma$ of the time and the equator (3.24)) in the deformed conifold side, (def-M5) or (def-M2), is described by a LG model.

Having identified B-type defects between LG models as matrix factorizations, one can make use of the defects for physical realization of the Khovanov-Rozansky construction of link homology on $\mathbb{R}_t \times S^1_\sigma$.

Recall, that in the Khovanov-Rozansky formulation [131, 132, 173] one first constructs a cube of resolutions by replacing each crossing either with)(or with ><. Each resulting resolution is a planar Murakami-Ohtsuki-Yamada

graph [158]. Then, one associates a matrix factorization to each resolved crossing, together with a map from one $\mathcal{D}_{)(}$ to the other $\mathcal{D}_{)(}$ at a positive crossing $\times$, corresponding to an edge of the cube. The part of the complex associated to one such crossing is called a *mapping cone*:

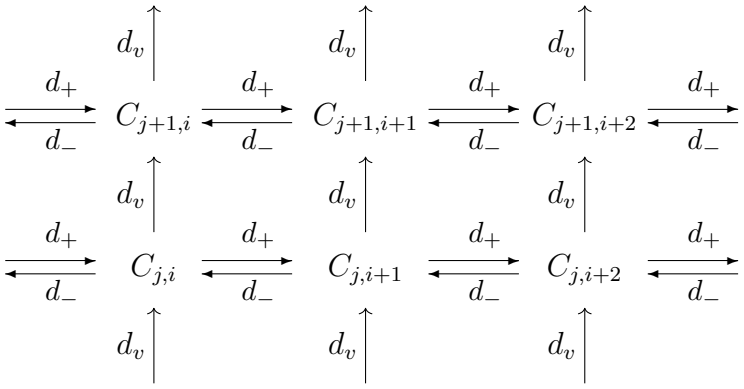


(3.30)

where all rings are isomorphic to $R = \mathbb{C}[x_1, x_2, x_3, x_4]/(x_3 + x_4 - x_1 - x_2)$ and the maps are defined by

$$\begin{aligned} d_+ &= (x_3 - x_1)(x_2 - x_3)\ , & d_- &= \frac{W(x_3) + W(x_4) - W(x_1) - W(x_2)}{(x_3 - x_1)(x_2 - x_3)}\ , \\ \widetilde{d}_+ &= (x_3 - x_1)\ , & \widetilde{d}_- &= \frac{W(x_3) + W(x_4) - W(x_1) - W(x_2)}{x_3 - x_1}\ , \\ d_v^0 &= x_2 - x_3\ , & d_v^1 &= 1\ . \end{aligned} \tag{3.31}$$

Note that the map d_v is induced in such a way that the diagram commutes. At a negative crossing $\times$, we associate the inverse of (3.30). (See [131, 132, 173] for more detail.) One then takes the tensor product of all these complexes to produce the total complex



. (3.32)

The homology of this complex turns out to be a knot invariant. When $W = 0$, it is easy to see from (3.31) that d_- is the zero map. Thus, the complex for HOMFLY homology has $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$-gradings corresponding to the (q, t, a)-grading where the q-grading is the internal grading ($q(x_i) = 2$) and (t, a)-grading is inherited from the bi-grading (j, i) of (3.32). Note that d_v is of degree $(0, 1, 0)$ and d_+ is of degree $(2, 0, 2)$. When $W \neq 0$, the complex is

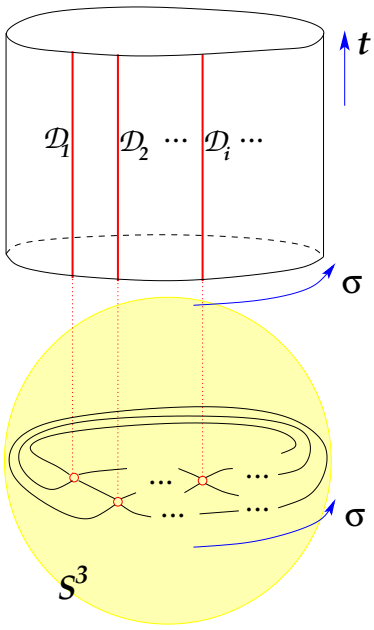


Figure 3.13: LG model with defects on $\mathbb{R}_t \times S^1_\sigma$ in the fivebrane system.

well-defined only after collapsing the $\mathbb{Z}$-grading to $\mathbb{Z}_2$-grading because of the presence of d_-. Therefore, the complex for $\mathfrak{sl}(N)$ homology loses the a-grading so that it is $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}_2$-graded where the $\mathbb{Z}_2$-grading comes from a matrix factorization. It is worth mentioning that this link homology is supported in a range of t-degrees that is at most the crossing number plus one, because a map $\times :)(\xrightarrow{d_v} \rangle\langle$, which increases t-degree by one, is assigned at each crossing.

To provide a physical realization of this formulation, using topological invariance along S^3 we represent a link L as a thin braid of k strands that runs along the equator $S^1_\sigma \subset S^3$:

$$k\text{-tuple cover}: \quad L \to S^1_\sigma,$$

as shown in Fig. 3.13. In Fig. 3.13, "$\circ$" stands for any of the basic crossings $\times$ and $\times$. Since we associate matrix factorizations (3.30) at each crossing, this means that the corresponding defect is placed at the position of the crossing in LG model on $\mathbb{R}_t \times S^1_\sigma$.

More precisely, a mapping cone (3.30) of two matrix factorizations admits a physical interpretation as a bound state of the two defects $\mathcal{D}_{)(}$ and $\mathcal{D}_{\rangle\langle}$, formed due to a fermionic *defect-changing operator* $d_v \in \mathcal{H}^1(\mathcal{D}_{)(}, \mathcal{D}_{\rangle\langle})$ [37, 204]. We

assign the bound state of the two defects at each defect line in Fig. 3.13. In fact, the bound state formation can be understood as a perturbation of the supercharge: since the mapping cone (3.30) can be expressed as

$$\mathcal{D}_{\times} = \mathrm{Cone}(d_v : \mathcal{D}_{)(} \to \mathcal{D}_{\mathrm{I}}) : \ R_1 \oplus \widetilde{R}_0 \underset{c_0}{\overset{c_1}{\rightleftarrows}} R_0 \oplus \widetilde{R}_1$$

with

$$c_1 = \begin{pmatrix} d_- & 0 \\ d_v & \widetilde{d}_+ \end{pmatrix}, \quad c_0 = \begin{pmatrix} d_+ & 0 \\ d_v & \widetilde{d}_- \end{pmatrix},$$

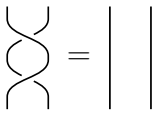
we see that the two non-interacting (direct sum) defects $\mathcal{D}_{)(}$ and $\mathcal{D}_{\mathrm{I}}$ are bound together by perturbing the boundary supercharge with the defect-changing operator d_v

$$Q_{\mathrm{bd}} = \begin{pmatrix} d_\mp & 0 \\ 0 & \widetilde{d}_\pm \end{pmatrix} \quad \longrightarrow \quad Q_{\mathrm{bd}} + \delta Q_{\mathrm{bd}} = \begin{pmatrix} d_\mp & 0 \\ d_v & \widetilde{d}_\pm \end{pmatrix} . \tag{3.33}$$

In summary, in LG theory with the potential W on $\mathbb{R}_t \times S^1_\sigma$, a bound state, either $\mathcal{D}_{\times}$ or $\mathcal{D}_{\times}$, is assigned at each crossing (Fig. 3.13). Each strand carries a $\mathscr{N} = 2$ chiral superfield x_i on a strip between the defect lines. As a result, the tensor product $\mathcal{D}_1 * \cdots * \mathcal{D}_n$ of the defects provides the complex (3.32), and therefore the BPS spectrum in this setup can be identified with the link homology

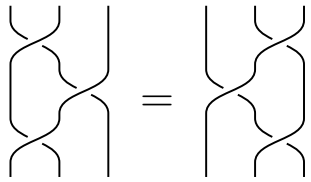
$$\mathcal{H}(\mathcal{D}_1, \cdots, \mathcal{D}_n) = \overline{\mathscr{H}}(L) \ .$$

The topological invariance of link homology tells us the fusion structure of the defects. The invariance under Reidemeister move II implies that the product of an over and under-crossing is the identity



$$\longrightarrow \quad \mathcal{D}_{\times} * \mathcal{D}_{\times} = \mathrm{Id} = \mathcal{D}_{\times} * \mathcal{D}_{\times} \ .$$

In addition, the Reidemeister move III means it obeys the braid group relation



$$\longrightarrow \quad \mathcal{D}^{(i)} * \mathcal{D}^{(i+1)} * \mathcal{D}^{(i)} = \mathcal{D}^{(i+1)} * \mathcal{D}^{(i)} * \mathcal{D}^{(i+1)}, \tag{3.34}$$

with the obvious relation $\mathcal{D}^{(i)} * \mathcal{D}^{(j)} = \mathcal{D}^{(j)} * \mathcal{D}^{(i)}$ for $|i-j| > 1$, where we denote the defect corresponding to a braid of i and $(i+1)$-strands by $\mathcal{D}^{(i)}$. In other words, $(\mathcal{D}, *)$ is an object of a braided monoidal category.

The structure of defects is similar to that of B-branes considered in Seidel-Thomas [185]. Therefore, the construction above strongly suggests that it should satisfy a spherical condition

$$\mathcal{H}^i(\mathcal{D}, \mathcal{D}) = \begin{cases} \mathbb{C}\,, & i = 0, n \\ 0\,, & \text{otherwise} \end{cases} \tag{3.35}$$

for some n. In addition, the defects should form a sequence of spherical matrix factorizations

$$\dim\ \mathcal{H}(\mathcal{D}^{(i)}, \mathcal{D}^{(j)}) = \begin{cases} 1\,, & |i-j| = 1 \\ 0 & |i-j| > 1 \end{cases}\,. \tag{3.36}$$

These conditions indeed give rise to the braid group relation (3.34) [37]. See also [108, §2.2].

Gornik's spectral sequence

Having realized the Khovanov-Rozansky formulation of link homology in the fivebrane systems, let us now investigate $\mathfrak{sl}(N)$ homology of the trefoil knot 3_1. It is known for two-bridge knots with a small number of crossings [141] that the homology $H_*(C_{j,i}, d_\pm)$ with respect to $d_\pm$ in (3.32) at each j is isomorphic to a direct sum of several copies of $\mathbb{C}[x]/(W')$, and the induced differential d_v on $H_*(C_{j,i}, d_\pm)$ is multiplication by W''. In particular, in the case of the trefoil, at $\deg t = j = 0, 2, 3$, the homology $H_*(C_{j,i}, d_\pm)$ with respect to $d_\pm$ in (3.32) is isomorphic to $\mathbb{C}[x]/(W')$, and the induced differential d_v is non-trivial between t-degrees 2 and 3. For instance, let's consider $\mathfrak{sl}(2)$ homology, for which $W = x^3$. Then, the complex can be expressed as follows:

t-degree	0	1	2	3
complex	$\mathbb{C}[x]/(x^2)$	$\xrightarrow{d_v}$ 0	$\xrightarrow{d_v}$ $\mathbb{C}[x]/(x^2)$	$\xrightarrow{d_v}$ $\mathbb{C}[x]/(x^2)$
	$\cup$		$\cup$	$\cup$
generators	$\textcircled{x}$		$\textcircled{x}$	x
	$\textcircled{1}$		1	$\textcircled{1}$

($d_v = x$: $1 \to \textcircled{1}$) (3.37)

This complex is essentially the same as the complex $[3_1]$ coming from the cube of resolutions in [15, Fig. 3]. Here, the induced differential d_v is a multiplication

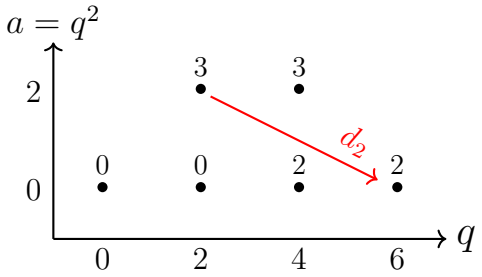


Figure 3.14: Unreduced $\mathfrak{sl}(2)$ homology of the trefoil

by $W'' = 2x$, and its non-trivial action is depicted by the red arrow above. Thus, the unreduced $\mathfrak{sl}(2)$ homology $\overline{\mathscr{H}}^{\mathfrak{sl}(2)}_{\square}(3_1) \cong H_*(H_*(C_{j,i}, d_\pm), d_v)$ is four-dimensional (the red circles above), and with appropriate shifts of q-degrees, its Poincaré polynomial can be written as

$$\overline{\mathscr{P}}^{\mathfrak{sl}(2)}_{\square}(3_1) = 1 + q^2 + q^4t^2 + q^8t^3 \ .$$

The induced differential d_v can be understood in the following way. It was proposed in [72] that the HOMFLY homology is endowed with a set of differentials $\{d_N\}_{N\in\mathbb{Z}}$ where the homology with respect to d_N is isomorphic to the $\mathfrak{sl}(|N|)$ homology. To obtain the unreduced $\mathfrak{sl}(2)$ homology $\overline{\mathscr{H}}^{\mathfrak{sl}(N)}_{\square}(K)$ from the reduced HOMFLY homology $\mathscr{H}_{\square}(K)$, first one takes the tensor product $\mathscr{H}_{\square}(K) \otimes \mathbb{C}[x]/(x^N)$ with the $\mathfrak{sl}(N)$ unknot homology, and then computes the homology with respect to d_N whose (a, q, t)-degree is $(-2, 2N, -1)$. For instance, the reduced HOMFLY homology of the trefoil is three-dimensional and its Poincaré polynomial (up to factor a^2q^{-2}) is

$$\mathscr{P}_{\square}(3_1) = 1 + q^4t^2 + a^2q^2t^3 \ .$$

To get the unreduced $\mathfrak{sl}(2)$ homology, one takes the d_2 differential after multiplying the $\mathfrak{sl}(2)$ unknot factor $(1+q^2)$ as illustrated in Fig. 3.14. In this example, it is easy to see that $\mathscr{H}_{\square}(3_1) \otimes \mathbb{C}[x]/(x^2)$ is isomorphic to $H_*(C_{j,i}(3_1), d_\pm)$, and the action of d_2 amounts to that of the induced differential d_v while the direction of the arrow is opposite. We thus see that, in this simple example, the induced differential d_v for $\mathfrak{sl}(N)$ homology is essentially equivalent to the d_N differential.

In fact, the spectral sequences corresponding to differentials $d_{N>0}$ have been constructed explicitly [173]. In physics, one should think of these spectral sequences as triggered by the deformation of the boundary supercharge (3.33).

Let us note that since the superpotential is homogeneous, the LG model has $\mathrm{U}(1)_V \equiv \mathrm{U}(1)_P$ vector R-symmetry and the $\mathrm{U}(1)_V$ R-charge is intrinsically the q-degree. Indeed, one can see from Fig. 3.14 that the differential d_v preserves the q-degree.

Deformation spectral sequences in knot homology have been investigated by Lee [138] for $\mathfrak{sl}(2)$ and by Gornik [98] for $\mathfrak{sl}(N)$ homology. There, changing the superpotential $W = x^{N+1}$ to $W = x^{N+1} + (N+1)\beta^N x$ leads to the deformation of the boundary supercharge $d_\pm$ in (3.31). Furthermore, because the potential is no longer homogeneous, the $\mathrm{U}(1)_V$ R-symmetry is broken, so that the deformed complex $C^{\mathrm{def}}(L)$ is filtered instead of bi-graded. Then, one can summarize the results of [98] as follows (see §3.1 for notations.):

- the E_1 page of the spectral sequence associated to $C^{\mathrm{def}}(L)$ is isomorphic to the undeformed $\mathfrak{sl}(N)$ homology $\overline{\mathscr{H}}_{\square}(L)$.
- the homology of $C^{\mathrm{def}}(L)$ in the E_∞ page is isomorphic to the tensor product $[\overline{\mathscr{H}}^{\mathfrak{sl}(N)}_{\square}(\bigcirc)]^{\otimes n}$ of n-copies of the $\mathfrak{sl}(N)$ homology of the unknot where n is the number of the component of L.

Actually, the Chinese remainder theorem tells us that the Jacobi ring of the deformed potential is isomorphic to

$$\mathbb{C}[x]/(x^N - \beta^N) \cong \bigoplus_{i=0}^{N-1} \mathbb{C}[x]/(x - \beta\omega^i) \ ,$$

where ω is an N-th root of unity. Its physical interpretation is as follows. When the superpotential is perturbed by the relevant operator $x^{N+1} \to x^{N+1} + (N+1)\beta x$, the Landau-Ginzburg model $\mathrm{LG}[\mathfrak{sl}(N)]$ undergoes RG flow and decomposes into N decoupled theories $\mathrm{LG}[\mathfrak{sl}(1)] \oplus \cdots \oplus \mathrm{LG}[\mathfrak{sl}(1)]$, each with the superpotential $x - \beta\omega^i$ $(i = 0, \ldots, N-1)$.

However, in the LG model $\mathrm{LG}[\mathfrak{sl}(1)]$, there is only the trivial defect (matrix factorization)! It is therefore immediate that the defects in the unperturbed theory (which encode the choice of link L) become trivial, *i.e.* invisible, from the standpoint of the $\mathfrak{sl}(1)$ theories which remain after the perturbation. The homology of $\mathrm{LG}[\mathfrak{sl}(1)]$ is therefore just the Jacobi ring $\mathbb{C}[x]/(x - \beta\omega^i)$, which is one-dimensional. In sum, the spectrum of BPS states in the IR is N^n-dimensional for an n-component link L, just as Gornik proved.

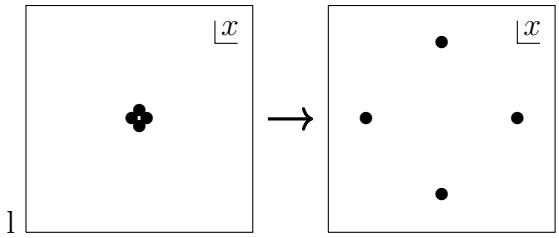


Figure 3.15: Deformation of the potential $x^5 \to x^5 + x$ breaks a degenerate vacuum into massive vacua.

Let us now see examples. In the case of the unknot, there is no non-trivial spectral sequence, *i.e.* the E_1 page is the same as the E_∞ page. Under the deformation of the superpotentials $x^{N+1} \to x^{N+1} + x$, the dimension of the Jacobi ring is invariant while critical points of the potential become non-degenerate (Fig. 3.15).

To see the non-trivial spectral sequence, we again look at the example of the $\mathfrak{sl}(2)$ homology of the trefoil. When we perturb the potential

$$W = x^3 \quad \longrightarrow \quad W = x^3 - 3x \ , \tag{3.38}$$

the complex (3.37) is deformed as shown in the following table.

t-degree	0	1	2	3
complex	$\mathbb{C}[x]/(x^2-1)$	$\xrightarrow{d_v}$ 0 $\xrightarrow{d_v}$	$\mathbb{C}[x]/(x^2-1)$	$\xrightarrow{d_v}$ $\mathbb{C}[x]/(x^2-1)$
	$\cup$		$\cup$	$\cup$
generators	$\textcircled{x}$		x	x
	$\textcircled{1}$		1	1

Since the induced differential $d_v = x$ is a multiplication by x, the generator x at $t = 2$ is mapped to the generator $x^2 \sim 1$ at $t = 3$ in the deformed ring $\mathbb{C}[x]/(x^2 - 1)$ (the red dotted arrow). This is exactly the Gornik's spectral sequence, and the homology is isomorphic to the $\mathfrak{sl}(2)$ unknot homology. Translating this into physics language, the perturbation (3.38) of the potential triggers RG flow from LG[$\mathfrak{sl}(2)$] to LG[$\mathfrak{sl}(1)$]$\oplus$LG[$\mathfrak{sl}(1)$]. In the infrared limit, the complex can be expressed as follows. In the ring $\mathbb{C}[x]/(x \pm 1)$, there is only one generator 1 and the induced differential is a multiplication by 1. Therefore,

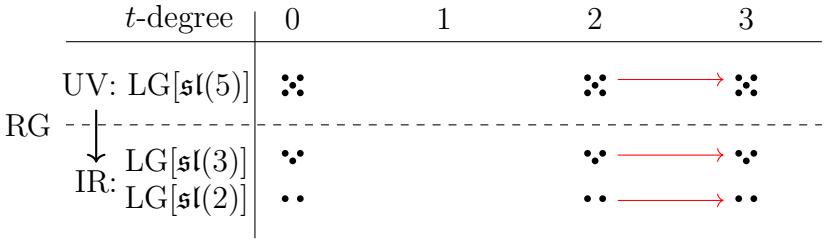
the homology of the complex is localized at $t = 0$ and it is two-dimensional.

t-degree	0	1	2	3
complex	$\mathbb{C}[x]/(x+1)$ $\oplus$ $\mathbb{C}[x]/(x-1)$	$\xrightarrow{d_v}$ 0 $\xrightarrow{d_v}$	$\mathbb{C}[x]/(x+1)$ $\oplus$ $\mathbb{C}[x]/(x-1)$	$\xrightarrow{d_v}$ $\mathbb{C}[x]/(x+1)$ $\oplus$ $\mathbb{C}[x]/(x-1)$
generators	$\cup$ ① ①		$\cup$ 1 1	$\cup$ 1 1

It is easy to generalize Gornik's spectral sequence. One can perturb the potential $W = x^{N+1}$ in such a way that the deformed potential W^{def} has the form

$$\partial_x W^{\mathrm{def}} = \prod_{i=1}^{k} (x - u_i)^{n_i} \ , \qquad \text{with} \quad N = \sum_{i=1}^{k} n_i \ .$$

Then, the same argument tells us that $\mathrm{LG}[\mathfrak{sl}(N)]$ flows to $\mathrm{LG}[\mathfrak{sl}(n_1)] \oplus \cdots \oplus \mathrm{LG}[\mathfrak{sl}(n_k)]$ under this relevant perturbation. Thus, there exists the spectral sequence associated to the deformed complex where the complex in the E_2 page is isomorphic to the $\mathfrak{sl}(N)$ homology, and the homology of the deformed complex in the E_∞ page is isomorphic to $\overline{\mathscr{H}}^{\mathfrak{sl}(n_1)}_{\square}(L) \oplus \cdots \oplus \overline{\mathscr{H}}^{\mathfrak{sl}(n_k)}_{\square}(L)$. For instance, one can schematically express this spectral sequence in the trefoil homology when one perturb the potential $x^6 \to x^6 - 12x^5/5 + 3x^4/2$ so that $\partial_x W^{\mathrm{def}} = x^3(x-1)^2$:



Chapter 4

DISCRETIZED HOLOGRAPHY: p-ADIC LOCAL FIELDS

[1]M. Heydeman, M. Marcolli, I. Saberi, and B. Stoica. "Tensor networks, p-adic fields, and algebraic curves: arithmetic and the AdS_3/CFT_2 correspondence." (2016). arXiv:1605.07639 [hep-th].

4.1 Introduction

Much attention has been paid of late to ideas that allow certain features of conformal field theory, such as long-range correlations, to be reproduced in lattice systems or other finitary models. As an example, the multiscale entanglement renormalization ansatz (or MERA), formulated by Vidal in [201], provides an algorithm to compute many-qubit quantum states whose entanglement properties are similar to those of the CFT vacuum. In Vidal's method, the states of progressively more distant qubits are entangled using successive layers of a self-similar network of finite tensors.

This tensor network can be thought of as a particular quantum circuit; other new connections between quantum information theory and holography were made in [4], which pointed out that bulk reconstruction and bulk locality in the AdS/CFT correspondence bear strong similarities to the properties of quantum error-correcting codes. This intuition was used in [172] to construct a family of "holographic" quantum codes, associated with hyperbolic tilings. In these codes, bulk qubits are thought of as the logical inputs, the boundary qubits at the periphery of the tiling constitute the encoded state, and the error-correcting properties of the code mimic features of holography such as the Ryu-Takayanagi formula [180].

Given the similarity of these networks (in which the number of tensors scales exponentially with the number of layers) to the geometry of hyperbolic space, it was natural to search for a connection with holography. In [191], Swingle proposed that MERA might be a natural discretization of AdS/CFT, in which the holographic direction (or renormalization scale) corresponds to the successive layers of the tensor network, and individual tensors are associated with "primitive cells" of the bulk geometry. However, successive work [13] identi-

fied constraints that prevent an AdS/MERA correspondence of this kind from fully reproducing all features of the bulk physics.

In this paper, we propose that discrete holographic models should be understood as approximating bulk geometry in a fundamentally different way. We are guided by considering a new and orthogonal direction in which the AdS_3/CFT_2 correspondence can be generalized. In these models, based on the p-adic numbers, discrete bulk geometry appears naturally. Despite this, essential and basic features of AdS/CFT, such as bulk isometries and boundary conformal symmetry (which are destroyed by a naive discretization), have analogues and can be fully understood in the discrete setting.

The bulk geometries relevant to the AdS_3/CFT_2 correspondence are well understood. The most well-known black hole solution is that of Bañados, Teitelboim, and Zanelli [11]; this solution was generalized to a family of higher-genus Euclidean black holes by Krasnov [136]. These solutions can be understood in general using the technique of *Schottky uniformization*, which presents a higher-genus black hole as the quotient of empty AdS_3 by a particular discrete subgroup of its isometries.

In [150], a holographic correspondence was established for these three-dimensional geometries. This correspondence expresses the conformal two point correlation function on the conformal boundary at infinity (a Riemann surface X_Γ of genus g) in terms of geodesic lengths in the bulk space (a hyperbolic handlebody $\mathcal{H}_\Gamma$ of genus g). The formula relating the boundary theory to gravity in the bulk is based on Manin's result [148] on the Arakelov Green's function.

However, we consider AdS_3/CFT_2 not merely because it is a simple setting for holography. For us, the crucial property of conformal field theory in two dimensions is its strong ties to algebraic geometry. These occur because every compact Riemann surface is a projective algebraic curve, so that many of the analytic concepts that arise in physics can (in two-dimensional contexts) be reformulated in purely algebraic terms. Once a concept can be formulated algebraically, it has many natural generalizations, obtained by changing the field of numbers one is considering. For instance, given a Riemann surface as the zero locus of a polynomial equation with rational coefficients, one can ask for the set of solutions over $\mathbb{C}$, over $\mathbb{R}$, over more exotic fields like the p-adics, or even over the integers.

The aforementioned holographic formula—and the the whole geometric setting of the correspondence, consisting of the Euclidean hyperbolic space AdS_3, its conformal boundary $\mathrm{P}^1(\mathbb{C})$, and quotients by actions of Schottky groups $\Gamma \subset \mathrm{PSL}(2,\mathbb{C})$—has a natural analogue in which the field is the p-adic numbers $\mathbb{Q}_p$. The bulk space becomes the Bruhat–Tits tree of $\mathbb{Q}_p$, which is a manifestly discrete infinite graph of uniform valence. Its conformal boundary at infinity is $\mathrm{P}^1(\mathbb{Q}_p)$, which can be thought of as the spacetime for an unusual class of CFTs. Black hole solutions are understood to be quotients of this geometry by p-adic Schottky groups $\Gamma \subset \mathrm{PGL}(2,\mathbb{Q}_p)$; these are known as *Mumford curves* in the mathematics literature. The results of Drinfeld and Manin [149] on periods of p-adic Schottky groups provide the corresponding holographic formula in this non-archimedean setting. We will give what we hope are intuitive introductions to these possibly unfamiliar concepts in the bulk of the paper.

Conformal field theory on p-adic spacetime has previously been developed, for the most part, in the context of the p-adic string theory (see, for instance, [33] and references therein), but has also been considered abstractly [153]. However, our perspective on the subject will be somewhat different: rather than using the p-adics as a worldsheet to construct real-space string amplitudes, our goal in this paper is to further develop the original holographic correspondence of [150] for the higher-genus black holes, informed by recent developments in the understanding of the AdS/CFT correspondence. We will emphasize the large extent to which algebraic structure allows familiar ideas, concepts, and arguments from ordinary $\mathrm{AdS}_3/\mathrm{CFT}_2$ can be carried over—in many cases line by line—to the p-adic setting. In addition to the holographic formulas of Manin and Marcolli, the standard semiclassical holographic analysis of scalar fields propagating without backreaction in anti-de Sitter space applies almost without alteration to the Bruhat–Tits tree. We discuss this in detail in §4.4.

In some cases, intuitions about how holography works in the archimedean case are supported even more sharply over the p-adics. For example, one normally thinks of the holographic direction as corresponding to a renormalization-group scale. Over the p-adics, as shown in §4.4, boundary modes contribute to the reconstruction of bulk functions only up to a height determined by their wavelength, and reconstruct precisely to zero above this height in the tree.

One of the most important new ideas in the AdS/CFT correspondence is the

study of entanglement entropy in boundary states and its connection (via the Ryu-Takayanagi formula) to the geometry of the bulk. We argue that, at least for the p-adic free boson CFT, an analogue of the familiar logarithmic scaling of the ground-state entanglement entropy should hold. Given such a formula, the Ryu-Takayanagi formula follows immediately from simple considerations of the geometry of the tree.

Tensor network models are often of interest because they reproduce our expectations about ground-state entanglement entropy, and in some cases (like the holographic quantum code of Pastawski *et al.* [172]) also satisfy formulas similar to Ryu-Takayanagi that relate the entanglement entropy to the size of paths or surfaces in the interior of the network. Given that our models exhibit a discrete bulk spacetime, a Ryu-Takayanagi formula, and a meaningful (and unbroken) group of bulk isometries/boundary conformal mappings, we suggest that the p-adic geometry is the natural one to consider in attempting to link tensor network models to spacetimes. We offer some ideas in this direction in §4.3.

Finally, on an even more speculative note, it is natural to wonder if the study of p-adic models of holography can be used to learn about the real case. So-called "adelic formulas" relate quantities defined over the various places (finite and infinite) of $\mathbb{Q}$; it was suggested in [146] that fundamental physics should be adelic in nature, with product formulae that relate the archimedean side of physics to a product of the contributions of all the p-adic counterparts. We briefly speculate about adelic formulas for the entanglement entropy in §4.5; one might hope that such formulas could be used to prove inequalities for entanglement entropy like those considered in [12], using ultrametric properties of the p-adics. We hope to further develop the adelic perspective, and return to these questions, in future work.

4.2 Review of necessary ideas

Basics of p-adic numbers

We begin with a lightning review of elementary properties of the p-adic fields. Our treatment here is far from complete; for a more comprehensive exposition, the reader is referred to [134], or to another of the many books that treat p-adic techniques.

When one constructs the continuum of the real numbers from the rationals, one

completes with respect to a metric: the distance between two points $x, y \in \mathbb{Q}$ is

$$d(x, y) = |x - y|_\infty, \tag{4.1}$$

where $|\cdot|_\infty$ is the usual absolute value. There are Cauchy sequences of rational numbers for which successive terms become arbitrarily close together, but the sequence does not approach any limiting rational number. The real numbers "fill in the gaps," such that every Cauchy sequence of rational numbers converges to a real limit by construction. That property is known as metric completeness.

The p-adic fields $\mathbb{Q}_p$ are completions of $\mathbb{Q}$ with respect to its other norms; these are defined by

$$\operatorname{ord}_p(x) = n \text{ when } x = p^n(a/b) \text{ with } a, b \perp p; \tag{4.2}$$

$$|x|_p = p^{-\operatorname{ord}_p(x)}. \tag{4.3}$$

By a theorem of Ostrowski, every norm on $\mathbb{Q}$ is equivalent to one of the p-adic norms or the usual (∞-adic) norm. It is common to refer to the different possible completions as the different "places" of $\mathbb{Q}$.

A number is p-adically small when it is divisible by a large power of p; one can think of the elements of $\mathbb{Q}_p$ as consisting of decimal numbers written in base p, which can extend infinitely far *left* (just as real numbers can be thought of as ordinary decimals extending infinitely far *right*). $\mathbb{Q}_p$ is uncountable and locally compact with respect to the topology defined by its metric; as usual, a basis for this topology is the set of open balls,

$$B_\epsilon(x) = \{y \in \mathbb{Q}_p : |x - y|_p < \epsilon\}. \tag{4.4}$$

The ring of integers $\mathbb{Z}_p$ of $\mathbb{Q}_p$ is also the unit ball about the origin:

$$\mathbb{Z}_p = \{x \in \mathbb{Q}_p : |x|_p \leq 1\}. \tag{4.5}$$

It can be described as the inverse limit of the system of base-p decimals with no fractional part and finite (but increasingly many) digits:

$$\mathbb{Z}_p = \varprojlim \left(\cdots \to \mathbb{Z}/p^{n+1}\mathbb{Z} \to \mathbb{Z}/p^n\mathbb{Z} \to \mathbb{Z}/p^{n-1}\mathbb{Z} \to \cdots \right). \tag{4.6}$$

$\mathbb{Z}_p$ is a discrete valuation ring; its unique maximal ideal is $\mathfrak{m} = p\mathbb{Z}_p$, and the quotient of $\mathbb{Z}_p$ by $\mathfrak{m}$ is the finite field $\mathbb{F}_p$. In general, for any finite extension of $\mathbb{Q}_p$, the quotient of its ring of integers by its maximal ideal is a finite field $\mathbb{F}_{p^n}$; we give more detail about this case in §4.6.

The Bruhat–Tits tree and its symmetries

In this section, we will describe the Bruhat–Tits Tree T_p and its symmetries. It should be thought of as a hyperbolic (though discrete) bulk space with conformal boundary $\mathrm{P}^1(\mathbb{Q}_p)$. Since these trees are a crucial part of the paper and may be unfamiliar to the reader, our treatment is informal, and aims to build intuition. Out of necessity, our discussion is also brief; for a more complete treatment, the reader may consult notes by Casselman [44] for constructions and properties related to the tree, or [225] for analysis on the tree and connections to the p-adic string.

We begin with a description of the boundary and its symmetries, which are completely analogous to the global conformal transformations of $\mathrm{P}^1(\mathbb{C})$. We then turn our attention to the bulk space T_p, focusing on its construction as a coset space and the action of $\mathrm{PGL}(2,\mathbb{Q}_p)$ on the vertices. Despite the fractal topology of the p-adic numbers, we will find (perhaps surprisingly) that many formulas from the real or complex cases are related to their p-adic counterparts by the rule $|\cdot|_\infty \to |\cdot|_p$.

Conformal group of $\mathrm{P}^1(\mathbb{Q}_p)$

The global conformal group on the boundary is $\mathrm{SL}(2,\mathbb{Q}_p)$, which consists of matrices of the form

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \text{ with } a,b,c,d \in \mathbb{Q}_p,\ ad - bc = 1. \tag{4.7}$$

This acts on points $x \in \mathrm{P}^1(\mathbb{Q}_p)$ by fractional linear transformations,

$$x \to \frac{ax+b}{cx+d}. \tag{4.8}$$

It can be checked that matrix multiplication corresponds to composition of such maps, so that the group action is well-defined. This is analogous to the $\mathrm{SL}(2,\mathbb{C})$ action on the Riemann sphere $\mathrm{P}^1(\mathbb{C})$. (We will sometimes also refer to $\mathrm{PGL}(2,\mathbb{Q}_p)$; the two differ only in minor details.)

The existence of a local conformal algebra for $\mathbb{Q}_p$, in analogy with the Virasoro symmetry in two-dimensional conformal field theory or general holomorphic mappings of $\mathrm{P}^1(\mathbb{C})$, is a subtle question. It is difficult to find definitions of a p-adic derivative or an infinitesimal transformation that are satisfactory for this purpose. In particular, since the "well-behaved" complex-valued functions

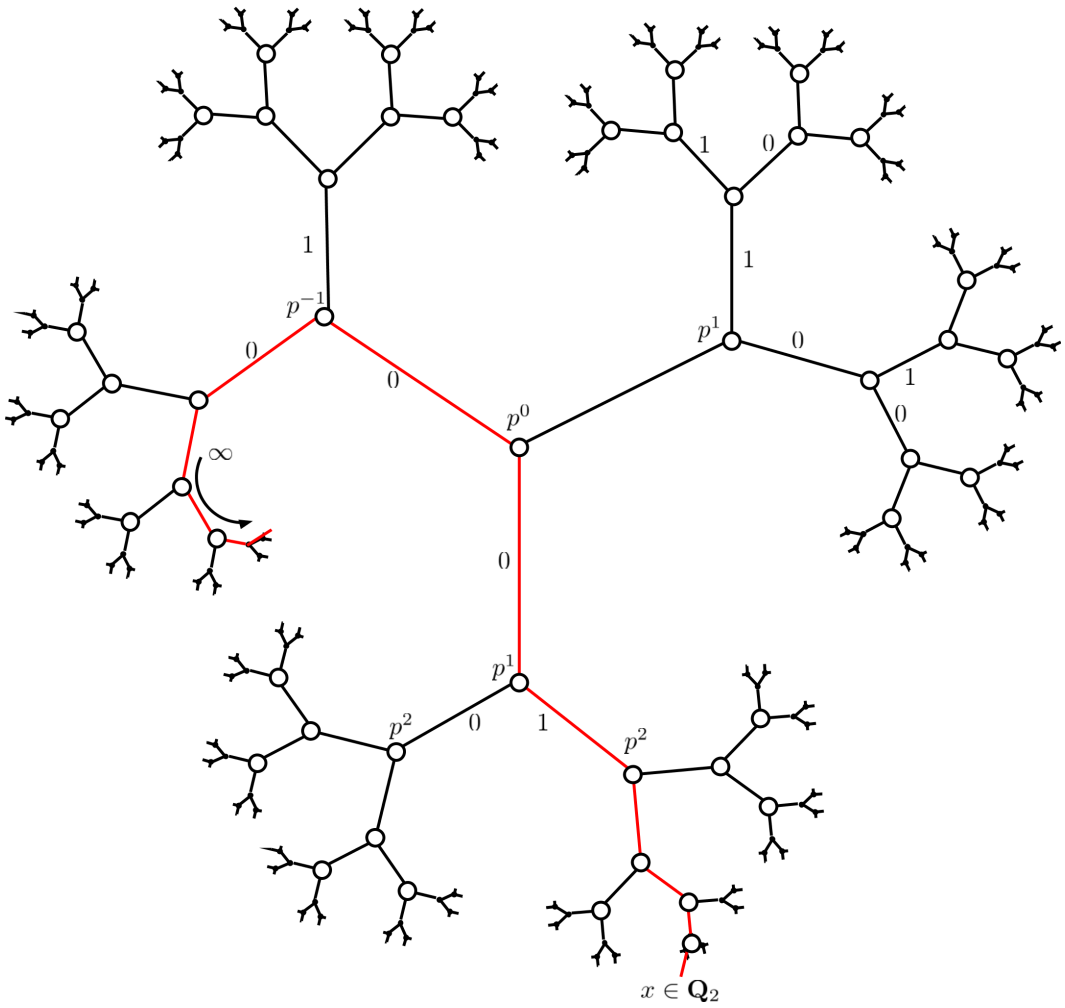


Figure 4.1: The standard representation of the Bruhat–Tits tree. The point at infinity and the center are arbitrary as the tree is homogeneous. Geodesics such as the highlighted one are infinite paths through the tree from ∞ to the boundary which uniquely specify elements of $\mathbb{Q}_p$. This path as a series specifies the digits of the decimal expansion of $x \in \mathbb{Q}_2$ in this example. At the nth vertex, we choose either 0 or 1 corresponding to the value of x_n in the p^nth term of x. Negative powers of p correspond to larger p-adic norms as we move towards the point ∞.

on $\mathbb{Q}_p$ are in some sense locally constant, there are no interesting derivations that act on the space of fields [153]. In this paper, we will concern ourselves only with global symmetries, which can still be used to constrain the properties of p-adic conformal field theories. We speculate about the possibility of enhanced conformal symmetry in §4.6.

The determinant condition implies that there are three free p-adic numbers which specify an element of $\mathrm{SL}(2, \mathbb{Q}_p)$. A convenient way to decompose a

general $\mathrm{SL}(2,\mathbb{Q}_p)$ transformation is to view it as the product of a special conformal transformation, a rotation, a dilatation, and a translation:

$$\begin{pmatrix} 1 & 0 \\ cp^{-m}a^{-1} & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} p^m & 0 \\ 0 & p^{-m} \end{pmatrix} \begin{pmatrix} 1 & bp^{-m}a^{-1} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} p^m a & b \\ c & p^{-m}a^{-1}(1+bc) \end{pmatrix}, \quad (4.9)$$

where $a, b, c \in \mathbb{Q}_p$ and $|a|_p = 1$. One can verify that the product is an arbitrary element of $\mathrm{SL}(2,\mathbb{Q}_p)$, where the determinant condition has been used to eliminate the d parameter. This represents a translation by $bp^{-m}a^{-1}$, a dilatation by p^{2m}, a rotation by a^2, and a special conformal translation by $cp^{-m}a^{-1}$. We have separated the diagonal subgroup into multiplication by elements of the unit circle, $a \in \mathbb{U}_p \subset \mathbb{Z}_p$, which do not change the p-adic norm (and thus are "rotations" in a p-adic sense), and multiplication by powers of p which scale the p-adic norm (and so correspond to dilatations). Representations of the multiplicative group of unit p-adics provide an analogue of the spin quantum number; we discuss this further in §4.4. It is worth stressing that these transformations are finite, and so we are characterizing the symmetry group rather than the algebra.

As is often the case in real conformal field theories, we can focus on the dilation subgroup. A diagonal matrix in $\mathrm{SL}(2,\mathbb{Q}_p)$ and its action on the coordinate is

$$\begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix}, \quad x \to x' = \alpha^2 x. \quad (4.10)$$

This has the effect of changing the p-adic norm by

$$|x'|_p = |\alpha|_p^2 |x|_p. \quad (4.11)$$

So if $|\alpha|_p \neq 1$, this will scale the size of coordinate. This parallels the complex case in which a dilatation changes the complex norm by $|z'| = |\alpha|^2|z|$. It will turn out to be the case that 2-point functions of spinless operators in p-adic conformal field theory will depend only on the p-adic norm of their separation. Schematically,

$$\langle \phi(x)\phi(y) \rangle \approx \frac{1}{|x-y|_p^{2\Delta}}. \quad (4.12)$$

Dilations will thus affect correlation functions of the p-adic conformal field theory exactly as in the complex case.

$\mathrm{PGL}(2,\mathbb{Q}_p)$ action on the tree T_p

We have seen that fractional linear transformations of the boundary coordinate work as in the real case. The action of the symmetry on the bulk space T_p is slightly more complicated to describe. Were we working in the archimedean theory, we would identify $\mathrm{PSL}(2,\mathbb{R})$ as the isometry group of the hyperbolic upper half space $\mathbb{H} = \mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$. Here $\mathrm{SO}(2)$ is a maximal compact subgroup. Similarly, in the context of $\mathrm{AdS}_{2+1}/\mathrm{CFT}_2$, we can think of the hyperbolic upper-half 3-space as a quotient space of the isometry group by its maximal compact subgroup: $\mathbf{H}^3 = \mathrm{SL}(2,\mathbb{C})/\mathrm{SU}(2)$.

Following this intuition, we define the *Bruhat–Tits tree* to be the quotient of the p-adic conformal group by its maximal compact subgroup:

$$T_p = \mathrm{PGL}(2,\mathbb{Q}_p)/\mathrm{PGL}(2,\mathbb{Z}_p). \tag{4.13}$$

In contrast with the archimedean examples, T_p is a discrete space: it is a homogeneous infinite tree, with vertices of valence $p+1$, whose boundary can be identified with the p-adic projective line. We expect isometries to correspond to rigid transformations of the vertices. Formally, the tree represents the incidence relations of equivalence classes of lattices in $\mathbb{Q}_p \times \mathbb{Q}_p$. As outlined in the appendix of [33], the group $\mathrm{PGL}(2,\mathbb{Q}_p)$ acts by matrix multiplication on the lattice basis vectors and takes one between equivalence classes. These transformations are translations and rotations of the points in the tree; they preserve distances, which are measured in the tree by just counting the number of edges along a given path. Since any two vertices in a connected tree are joined by exactly one path, this is well-defined; all paths are geodesics.

A standard way of representing T_p is depicted in Fig. 4.1 for the case $p = 2$. This is a regular tree with $p+1$ legs at each vertex; the exponential growth in the number of vertices with distance from a base point reflects the "hyperbolic" nature of distance in the the tree. Since paths are unique, there is a one-to-one correspondence between infinite paths in the tree starting at ∞ and elements of $\mathbb{Q}_p$. (This can be viewed like a p-adic version of stereographic projection.)

The choice of the apparent center and geodesic corresponding to infinity are arbitrary. Just as in the archimedean case, we must fix three boundary points to identify a p-adic coordinate on the projective line, corresponding to 0, 1, and ∞. Once these arbitrary choices are made, the geodesics joining them form a Y in the bulk, whose center is taken to be the centerpoint of the tree.

We can then understand the geodesic connecting ∞ to x as labeling the unique p-adic decimal expansion for $x = p^\gamma(x_0 + x_1 p + x_2 p^2 + \dots)$, where each of the x_n take values in $0, 1, \dots p-1$ corresponding to the p possible choices to make at each vertex. Each vertex of the tree is naturally marked with a copy of the finite field $\mathbb{F}_p$, identified with one "digit" of a p-adic number.

Viewing the tree as the space of p-adic decimal expansions may in some ways be more useful than the definition in terms of equivalence classes of lattices. Geometrically moving closer or further from the boundary corresponds to higher or lower precision of p-adic decimal expansions. Even with no reference to quantum mechanics or gravity, we see some hint of holography and renormalization in the tree- a spatial direction in the bulk parameterizes a scale or precision of boundary quantities. This is explored more fully in §4.4.

We now illustrate some examples of $\mathrm{PGL}(2, \mathbb{Q}_p)$ transformations on the tree. First note that the choice of the center node is arbitrary. We can take this point to be the equivalence class of unit lattices modulo scalar multiplication. One can show that this equivalence class (or the node it corresponds to) is invariant under the $\mathrm{PGL}(2, \mathbb{Z}_p)$ subgroup, so these transformations leave the center fixed and rotate the branches of the tree about this point.

More interesting is a generator such as

$$g = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \in \mathrm{PGL}(2, \mathbb{Q}_p). \tag{4.14}$$

This transformation (and others in $\mathrm{PGL}(2, \mathbb{Q}_p)$) act by translating the entire tree along a given geodesic (one can see this either from the lattice incidence relations, or from translating or shifting the p-adic decimal series expansion). This is illustrated in Fig. 4.2. We can think of these transformations as the lattice analogs of translations and dilatations of the real hyperbolic plane.

Integration measures on p-adic spaces

Just as is the case for $\mathbb{C}$, there are two natural measures on $\mathbb{Q}_p$ (or more properly, on the projective line over $\mathbb{Q}_p$); they can be understood intuitively by thinking of $\mathbb{Q}_p$ as the boundary of T_p. The first is the Haar measure $d\mu$, which exists for all locally compact topological groups. With respect to either measure, the size of the set of p-adic integers is taken to be 1:

$$\mu(\mathbb{Z}_p) = 1. \tag{4.15}$$

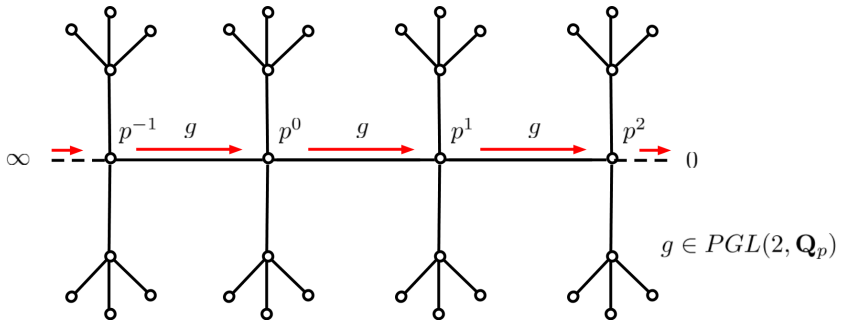


Figure 4.2: An alternative representation of the Bruhat–Tits Tree (for $p = 3$) in which we have unfolded the tree along the 0 geodesic. The action of elements of $\mathrm{PGL}(2, \mathbb{Q}_p)$ acts by translating the entire tree along different possible geodesics. In this example we translate along the 0 geodesic, which can be thought of as multiplication of each term in a p-adic decimal expansion by p. This map has two fixed points at 0 and ∞. In this "unfolded" form, a point in $\mathrm{P}^1(\mathbb{Q}_p)$ is specified by a geodesic that runs from ∞ and follows the 0 geodesic until some level in the tree where it leaves the 0 geodesic towards the boundary. The p-adic norm is simply p to the inverse power of the point where it leaves the 0 geodesic (so that leaving "sooner" leads to a larger norm, and later to a smaller norm).

The Haar measure is then fixed by multiplicativity and translation invariance; any open ball has measure equal to the p-adic norm of its radius. It is helpful to think of $\mathbb{Q}_p$ as being "flat" when considered with this measure.

The other measure, the Patterson-Sullivan measure, is the p-adic analogue of the Fubini-Study metric on $\mathrm{P}^1(\mathbb{C})$. It is most easily defined with reference to the tree, in which we fix a basepoint C (to be thought of as the unique meeting point of the geodesics joining 0, 1, and ∞ when a coordinate is chosen on the boundary). Recall that the open balls in $\mathbb{Q}_p$ correspond to the endpoints of branches of the tree below a vertex v. In the Patterson-Sullivan measure,

$$d\mu_0(B_v) = p^{-d(C,v)}. \tag{4.16}$$

The two measures are related by

$$\begin{aligned} d\mu_0(x) &= d\mu(x), \qquad |x|_p \leq 1; \\ d\mu_0(x) &= \frac{d\mu(x)}{|x|_p^2}, \qquad |x|_p > 1. \end{aligned} \tag{4.17}$$

(Later on, we will at times use the familiar notation dx to refer to the Haar measure.) The most intuitive way to picture the Patterson-Sullivan measure is to imagine the tree pointing "radially outward" from its centerpoint. It is

then easy to understand the transformation rule (4.17); it says that when all geodesics point downward from infinity and the boundary is "flat" at the lower end of the picture, points far from zero (outside $\mathbb{Z}_p$) can only be reached by geodesics that travel upward from C before turning back down towards the boundary.

Schottky uniformization of Riemann surfaces

In this section, we review Schottky uniformization, which allows one to think of a higher-genus Riemann surface as a quotient of the projective line by a particular discrete subgroup of its Möbius transformations.

A Schottky group of rank $g \geq 1$ is a discrete subgroup of $\mathrm{PSL}(2,\mathbb{C})$ which is purely loxodromic and isomorphic to a free group on g generators. The group $\mathrm{PSL}(2,\mathbb{C})$ acts on $\mathrm{P}^1(\mathbb{C})$ by fractional linear transformations,

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : z \mapsto \frac{az+b}{cz+d}.$$

The loxodromic condition means that each nontrivial element $\gamma \in \Gamma \setminus \{1\}$ has two distinct fixed points $z_\gamma^\pm$ (one attractive and one repelling) in $\mathrm{P}^1(\mathbb{C})$. The closure in $\mathrm{P}^1(\mathbb{C})$ of the set of all fixed points of elements in Γ is the limit set Λ_Γ of Γ, the set of all limit points of the action of Γ on $\mathrm{P}^1(\mathbb{C})$. In the case $g = 1$ the limit set consists of two points, which we can choose to identify with $\{0,\infty\}$, while for $g > 1$ the set Λ_Γ is a Cantor set of Hausdorff dimension $0 \leq \dim_H(\Lambda_\Gamma) < 2$. The Hausdorff dimension is also the exponent of convergence of the Poincaré series of the Schottky group: $\sum_{\gamma\in\Gamma} |\gamma'|^s$ converges for $s > \dim_H(\Lambda_\Gamma)$ [31].

It is well known that any compact smooth Riemann surface X admits a Schottky uniformization, namely $X = \Omega_\Gamma/\Gamma$, where $\Gamma \subset \mathrm{PSL}(2,\mathbb{C})$ is a Schottky group of rank equal to the genus $g = g(X)$ of the Riemann surface, and $\Omega_\Gamma = \mathrm{P}^1(\mathbb{C}) \setminus \Lambda_\Gamma$ is the *domain of discontinuity* of the action of Γ on $\mathrm{P}^1(\mathbb{C})$. There is a well known relation between Schottky and Fuchsian uniformizations of compact Riemann surfaces of genus $g \geq 2$; see [192].

A *marking* of a rank g Schottky group $\Gamma \subset \mathrm{PSL}(2,\mathbb{C})$ is a choice of a set of generators $\{\gamma_1,\ldots,\gamma_g\}$ of Γ and a set of $2g$ open connected regions D_i in $\mathrm{P}^1(\mathbb{C})$, with $C_i = \partial D_i$ the boundary Jordan curves homemorphic to S^1, with the following properties:

1. the closures of the D_i are pairwise disjoint
2. $\gamma_i(C_i) \subset C_{g+i}$
3. $\gamma_i(D_i) \subset \mathrm{P}^1(\mathbb{C}) \setminus D_{g+i}$.

The marking is *classical* if all the C_i are circles. (All Schottky groups admit a marking, but not all admit a classical marking.) A fundamental domain F_Γ for the action of the Schottky group Γ on the domain of discontinuity $\Omega_\Gamma \subset \mathrm{P}^1(\mathbb{C})$ can be constructed by taking

$$F_\Gamma = \mathrm{P}^1(\mathbb{C}) \setminus \cup_{i=1}^g (D_i \cup \bar{D}_{g+i}).$$

This satisfies $\cup_{\gamma\in\Gamma}\gamma(F_\Gamma) = \Omega_\Gamma$. In the case of genus $g = 1$, with $\Gamma = q^{\mathbb{Z}}$, for some $q \in \mathbb{C}$ with $|q| > 1$, the region F_Γ constructed in this way is an annulus A_q, with D_1 the unit disk in $\mathbb{C}$ and D_2 the disk around ∞ given by complement in $\mathrm{P}^1(\mathbb{C})$ of the disk centered at zero of radius $|q|$, so that $q^{\mathbb{Z}}A_q = \mathbb{C}^* = \mathrm{P}^1(\mathbb{C}) \setminus \{0, \infty\} = \Omega_{q^{\mathbb{Z}}}$. The resulting quotient $E_q = \mathbb{C}^*/q^{\mathbb{Z}}$ is the Tate uniformization of elliptic curves.

Hyperbolic handlebodies and higher-genus black holes

The action of $\mathrm{PSL}(2,\mathbb{C})$ by fractional linear transformations on $\mathrm{P}^1(\mathbb{C})$ extends to an action by isometries on the real 3-dimensional hyperbolic space $\mathbf{H}^3$, with $\mathrm{P}^1(\mathbb{C})$ its conformal boundary at infinity. In coordinates $(z, y) \in \mathbb{C} \times \mathbb{R}^*_+$ in $\mathbf{H}^3$, the action of $\mathrm{PSL}(2,\mathbb{C})$ by isometries of the hyperbolic metric is given by

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : (z, y) \mapsto \left(\frac{(az+b)\overline{(cz+d)} + a\bar{c}y^2}{|cz+d|^2 + |c|^2y^2}, \frac{y}{|cz+d|^2 + |c|^2y^2} \right).$$

Given a rank g Schottky group $\Gamma \subset \mathrm{PSL}(2,\mathbb{C})$, we can consider its action on the conformally compactified hyperbolic 3-space $\overline{\mathbf{H}^3} = \mathbf{H}^3 \cup \mathrm{P}^1(\mathbb{C})$. The only limit points of the action are on the limit set Λ_Γ that is contained in the conformal boundary $\mathrm{P}^1(\mathbb{C})$, and hence a domain of discontinuity for this action is given by

$$\mathbf{H}^3 \cup \Omega_\Gamma \subset \overline{\mathbf{H}^3} = \mathbf{H}^3 \cup \mathrm{P}^1(\mathbb{C}).$$

The quotient of $\mathbf{H}^3$ by this action is a 3-dimensional hyperbolic handlebody of genus g

$$\mathcal{H}_\Gamma = \mathbf{H}^3/\Gamma,$$

with conformal boundary at infinity given by the Riemann surface $X_\Gamma = \Omega_\Gamma/\Gamma$,

$$\overline{\mathcal{H}}_\Gamma = \mathcal{H}_\Gamma \cup X_\Gamma = (\mathbf{H}^3 \cup \Omega_\Gamma)/\Gamma.$$

Given a marking of a rank g Schottky group Γ (for simplicity we will assume the marking is classical), let D_i be the discs in $\mathrm{P}^1(\mathbb{C})$ of the marking, and let $\mathcal{D}_i$ denote the geodesic domes in $\mathbf{H}^3$ with boundary $C_i = \partial D_i$, namely the $\mathcal{D}_i$ are the open regions of $\overline{\mathbf{H}^3}$ with boundary $S_i \cup D_i$, where the S_i are totally geodesic surfaces in $\mathbf{H}^3$ with boundary C_i that project to D_i on the conformal boundary. Then a fundamental domain for the action of Γ on $\mathbf{H}^3 \cup \Omega_\Gamma$ is given by

$$\mathcal{F}_\Gamma = F_\Gamma \cup (\mathbf{H}^3 \setminus \cup_{i=1}^g (\mathcal{D}_i \cup \bar{\mathcal{D}}_{g+i}).$$

The boundary curves C_i for $i = 1, \dots, g$ provide a collections of A-cycles, that give half of the generators of the homology of the Riemann surface X_Γ: the generators that become trivial in the homology of the handlebody $\bar{\mathcal{H}}_\Gamma$. The union of fundamental domains $\gamma(\mathcal{F}_\Gamma)$ for $\gamma \in \Gamma$ can be visualized as in Fig. 4.4.

In the case of genus $g = 1$ with $\Gamma = q^{\mathbb{Z}}$, acting on $\mathbf{H}^3$ by

$$\begin{pmatrix} q^{1/2} & 0 \\ 0 & q^{-1/2} \end{pmatrix} (z, y) = (qz, |q|y),$$

with limit set $\{0, \infty\}$ the fundamental domain $\mathcal{F}_\Gamma$ consists of the space in the upper half space $\mathbf{H}^3$ contained in between the two spherical domes of radius 1 and $|q| > 1$. The generator q of the group acts on the geodesic with endpoints 0 and ∞ as a translation by $\log |q|$. The quotient $\mathbf{H}^3/q^{\mathbb{Z}}$ is a hyperbolic solid torus, with the Tate uniformized elliptic curve $E_q = \mathbb{C}^*/q^{\mathbb{Z}}$ as its conformal boundary at infinity, and with a unique closed geodesic of length $\log q$. It is well known (see [21], [143] and §2.3 of [150]) that the genus one handlebodies $\mathcal{H}_{q^{\mathbb{Z}}}$ are the Euclidean BTZ black holes [11], where the cases with $q \in \mathbb{C} \setminus \mathbb{R}$ correspond to spinning black holes. The geodesic length $\log |q|$ is the area of the event horizon, and hence is proportional to the black hole entropy.

The higher-genus hyperbolic handlebodies correspond to generalizations of the BTZ black hole to the higher-genus asymptotically AdS_3 black holes considered in [136] and [150].

In these more general higher-genus black hole, because of the very different nature of the limit set (a fractal Cantor set instead of two points) the structure of the black hole event horizon is significantly more complicated. In the

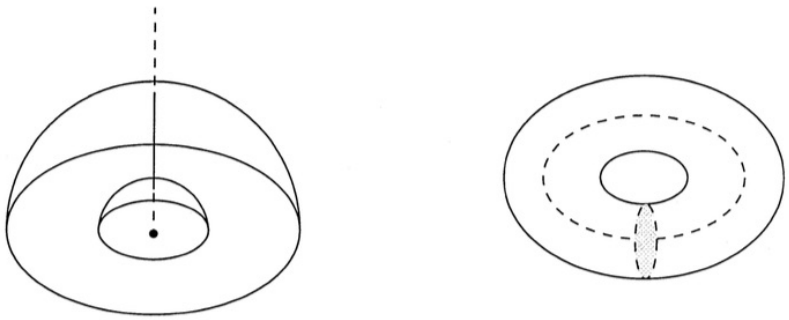


Figure 4.3: Fundamental domain and quotient for the Euclidean BTZ black hole. Compare with the p-adic BTZ geometry, shown in Fig. 4.10.

Euclidean BTZ black hole, the only infinite geodesic that remains confined into a compact region inside the hyperbolic solid torus $\mathcal{H}_{q^{\mathbb{Z}}}$ for both $t \to \pm\infty$ is the unique closed geodesic (the image in the quotient of the geodesic in $\mathbf{H}^3$ given by the vertical line with endpoints 0 and ∞. On the other hand, in the higher-genus cases, the geodesics in the hyperbolic handlebody $\mathcal{H}_\Gamma = \mathbf{H}^3/\Gamma$ can be classified as:

1. *Closed geodesics*: these are the images in the quotient $\mathcal{H}_\Gamma$ of geodesics in $\mathbf{H}^3$ with endpoints $\{z_\gamma^+, z_\gamma^-\}$, the attractive and repelling fixed points of some element $\gamma \in \Gamma$.

2. *Bounded geodesics:* these images in the quotient $\mathcal{H}_\Gamma$ of geodesics in $\mathbf{H}^3$ with endpoints on Λ_Γ. If the endpoints are not a pair of fixed points of the same element of Γ the geodesic in the quotient is not closed, but it remains forever confined within a compact region inside $\mathcal{H}_\Gamma$, the *convex core* $\mathcal{C}_\Gamma$.

3. *Unbounded geodesics:* these are images in the quotient $\mathcal{H}_\Gamma$ of geodesics in $\mathbf{H}^3$ with at least one of the two endpoints in Ω_Γ. These are geodesics in $\mathcal{H}_\Gamma$ that wander off (in at least one time direction $t \to \infty$ or $t \to -\infty$) towards the conformal boundary X_Γ at infinity and eventually leave every compact region in $\mathcal{H}_\Gamma$.

The convex core $\mathcal{C}_\Gamma \subset \mathcal{H}_\Gamma$ is the quotient by Γ of the geodesic hull in $\mathbf{H}^3$ of the limit set Λ_Γ. It is a compact region of finite hyperbolic volume in $\mathcal{H}_\Gamma$, and it is a deformation retract of $\mathcal{H}_\Gamma$. A natural replacement for the event horizon of the BTZ black hole in these higher-genus cases can be identified in terms

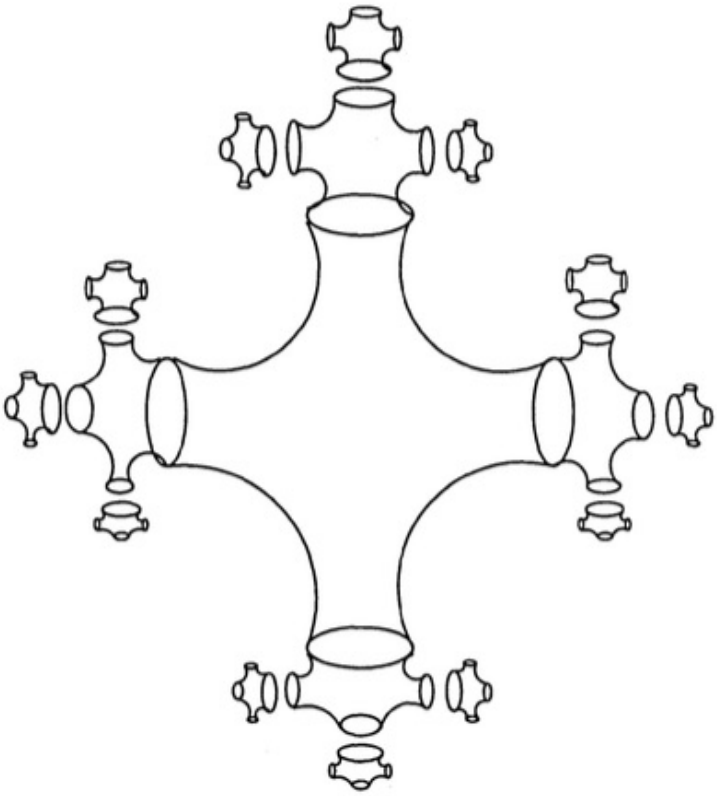


Figure 4.4: Fundamental domains for the action of Γ on $\mathbf{H}^3$.

of the convex core $\mathcal{C}_\Gamma$, where we think of $\mathcal{C}_\Gamma$ as the region from which geodesic trajectories cannot escape and must remain forever confined. The complement $\mathcal{H}_\Gamma \setminus \mathcal{C}_\Gamma$ is homeomorphic to $\partial\mathcal{C}_\Gamma \times \mathbb{R}_+$ (see [41] for a more general treatment of convex cores of Kleinian groups and ends of hyperbolic 3-manifolds). The boundary $\partial\mathcal{C}_\Gamma$ is the event horizon of the higher-genus black hole, with the black hole entropy proportional to the area of $\partial\mathcal{C}_\Gamma$.

In [147] and [148], Manin proposed to interpret the tangle of bounded geodesics inside the hyperbolic handlebody $\mathcal{H}_\Gamma$ as a model for the missing "closed fiber at infinity" in Arakelov geometry. This interpretation was based on the calculation of the Arakelov Green function [148], and the analogy with the theory of Mumford curves [157] and the computations of [149] for p-adic Schottky groups. The results of [148] and their holographic interpretation in [150], as well as the parallel theory of Mumford curves and periods of p-adic Schottky groups, will form the basis for our development of a p-adic and adelic form of the AdS_{+1}/CFT correspondence. The interpretation of the tangle of bounded

geodesics in $\mathcal{H}_\Gamma$ as "closed fiber at infinity" of Arakelov geometry was further enriched with a cohomological interpretation in [52] (see also [50], [51] for the p-adic counterpart).

Bruhat–Tits trees, p-adic Schottky groups, and Mumford curves

The theory of Schottky uniformization of Riemann surfaces as conformal boundaries of hyperbolic handlebodies has a non-archimedean parallel in the theory of Mumford curves, uniformized by p-adic Schottky groups, seen as the boundary at infinity of a quotient of a Bruhat–Tits tree.

Some basic facts regarding the geometry of the Bruhat–Tits tree T_p of $\mathbb{Q}_p$ have been recalled in §4.2. More generally, the geometry we consider here applies to any finite extension $\mathbf{k}$ of the p-adic field $\mathbb{Q}_p$. By identifying $(\mathcal{O}_{\mathbb{Q}_p}/\mathfrak{m}^r) \otimes \mathcal{O}_\mathbf{k} = \mathcal{O}_\mathbf{k}/\mathfrak{m}^{re_\mathbf{k}}$, where $e_\mathbf{k}$ is the ramification index of $\mathbf{k}$ over $\mathbb{Q}_p$, we see that the Bruhat–Tits tree $T_\mathbf{k}$ for a finite extension $\mathbf{k}$ of $\mathbb{Q}_p$ is obtained from the Bruhat–Tits tree of $\mathbb{Q}_p$ by adding $e_\mathbf{k}-1$ new vertices in each edge of $T_{\mathbb{Q}_p}$ and increasing the valence of all vertices to p^f+1, where $f = [\mathbf{k} : \mathbb{Q}_p]/e_\mathbf{k}$, the degree of the extension normalized by the ramification index.

Let $\mathcal{O}_\mathbf{k}$ denote the ring of integers of $\mathbf{k}$ and $\mathfrak{m} \subset \mathcal{O}_\mathbf{k}$ the maximal ideal, so that the residue field $\mathcal{O}_\mathbf{k}/\mathfrak{m} = \mathbb{F}_q$ is a finite field with $q = p^r$ for some $r \in \mathbb{N}$. The set of vertices $V(T_\mathbf{k})$ of the Bruhat–Tits tree $T_\mathbf{k}$ of $\mathbf{k}$ is the set of equivalence classes of free rank two $\mathcal{O}_\mathbf{k}$-modules, under the equivalence $M_1 \sim M_2$ if $M_1 = \lambda M_2$, for some $\lambda \in \mathbf{k}^*$. For a pair of such modules with $M_2 \subset M_1$, one can define a distance function $d(M_1, M_2) = |l-k|$, where $M_1/M_2 = \mathcal{O}_\mathbf{k}/\mathfrak{m}^l \oplus \mathcal{O}_\mathbf{k}/\mathfrak{m}^k$. This distance is independent of representatives in the equivalence relation. There is an edge in $E(T_\mathbf{k})$ connecting two vertices in $V(T_\mathbf{k})$ whenever the corresponding classes of modules have distance one. The resulting tree $T_\mathbf{k}$ is an infinite homogeneous tree with vertices of valence $q+1$, where $q = \#\mathcal{O}_\mathbf{k}/\mathfrak{m}$ is the cardinality of the residue field. The boundary at infinity of the Bruhat–Tits tree is identified with $\mathrm{P}^1(\mathbf{k})$. One can think of the Bruhat–Tits tree as a network, with a copy of the finite field $\mathbb{F}_q$ (or better of the projective line $\mathrm{P}^1(\mathbb{F}_q)$) associated to each vertex: this will be the guiding viewpoint in our approach to non-archimedean tensor networks.

The reader should beware that there is an unavoidable clash of notation: q is the standard notation for the modular parameter of an elliptic curve, but is also used to denote a prime power $q = p^r$ in the context of finite fields

or extensions of the p-adics. While both uses will be made in this paper, particularly in this section and in §4.6, we prefer not to deviate from standard usage; it should be apparent from context which is intended, and hopefully no confusion should arise.

There is an action of $\mathrm{PGL}(2,\mathbf{k})$ on the set of vertices $V(T_{\mathbf{k}})$ that preserves the distance, hence it acts as isometries of the tree $T_{\mathbf{k}}$. A p-adic Schottky group is a purely loxodromic finitely generated torsion free subgroup of $\mathrm{PGL}(2,\mathbf{k})$. The Schottky group Γ is isomorphic to a free group on g-generators, with g the rank of Γ.

In this p-adic setting the loxodromic condition means that every nontrivial element γ in Γ has two fixed points $z_\gamma^\pm$ on the boundary $\mathrm{P}^1(\mathbf{k})$. Equivalently, an element γ is loxodromic if the two eigenvalues have different p-adic valuation. The closure of the set of fixed points $z_\gamma^\pm$, or equivalently the set of accumulation points of the action of Γ on $T_{\mathbf{k}} \cup \mathrm{P}^1(\mathbf{k})$ is the limit set Λ_Γ of the Schottky group Γ. The complement $\mathrm{P}^1(\mathbf{k}) \setminus \Lambda_\Gamma = \Omega_\Gamma(\mathbf{k})$ is the domain of discontinuity of the action of Γ on the boundary.

There is a unique geodesic ℓ_γ in $T_{\mathbf{k}}$ with endpoints $\{z_\gamma^-, z_\gamma^+\}$, the *axis* of a loxodromic element γ. The subgroup $\gamma^{\mathbb{Z}}$ acts on $T_{\mathbf{k}}$ by translations along ℓ_γ. There is a smallest subtree $T_\Gamma \subset T_{\mathbf{k}}$ that contains all the axes ℓ_γ of all the nontrivial elements $\gamma \in \Gamma$. The boundary at infinity of the subtree T_Γ is the limit set Λ_Γ. T_Γ is the non-archimedean analog of the geodesic hull of the limit set of a Schottky group in $\mathbf{H}^3$.

The quotient $X_\Gamma(\mathbf{k}) = \Omega_\Gamma(\mathbf{k})/\Gamma$ is a *Mumford curve* with its p-adic Schottky uniformization, [157]. The quotient $T_{\mathbf{k}}/\Gamma$ consists of a finite graph T_Γ/Γ with infinite trees appended at the vertices of T_Γ/Γ, so that the boundary at infinity of the graph $T_{\mathbf{k}}/\Gamma$ is the Mumford curve $X_\Gamma(\mathbf{k})$. The finite graph $G_{\mathbf{k}} = T_\Gamma/\Gamma$ is the dual graph of the special fiber X_q (a curve over $\mathbb{F}_q$ which consists of a collection of $\mathrm{P}^1(\mathbb{F}_q)$ at each vertex of $G_{\mathbf{k}}$, connected along the edges). A family of finite graphs $G_{\mathbf{k},n}$, for $n \in \mathbb{N}$, is obtained by considering neighborhoods $T_{\Gamma,n}$ of T_Γ inside $T_{\mathbf{k}}$ consisting of T_Γ together with all vertices in $T_{\mathbf{k}}$ that are at a distance at most n from some vertex in T_Γ and the edges between them (these trees are preserved by the action of Γ), and taking the quotients $G_{\mathbf{k},n} = T_{\Gamma,n}/\Gamma$. The endpoints (valence one vertices) in $G_{\mathbf{k},n}$ correspond to the $\mathbb{F}_{q^n}$ points in the special fiber, $X_q(\mathbb{F}_{q^n})$, see [147]. One sees in this way, geometrically, how the $\mathbf{k}$-points in the Mumford curve $X_\Gamma(\mathbf{k})$ are obtained as limits, going along the

infinite ends of the graph $T_{\mathbf{k}}/\Gamma$, which correspond to successively considering points of X_q over field extensions $\mathbb{F}_{q^n}$. Conversely, one can view the process of going into the tree from its boundary $X_\Gamma(\mathbf{k})$ towards the graph $G_{\mathbf{k}}$ in the middle of $T_{\mathbf{k}}/\Gamma$ as applying reductions to $\mathbb{F}_{q^n}$. We will see later in the paper how this process should be thought of physically as a form of renormalization. The finite graph $G_{\mathbf{k}} = T_\Gamma/\Gamma$ is the non-archimedean analog of the convex core $\mathcal{C}_\Gamma$ of the hyperbolic handlebody $\mathcal{H}_\Gamma$, while the infinite graph $T_{\mathbf{k}}/\Gamma$ is the non-archimedean analog of $\mathcal{H}_\Gamma$ itself, with the Mumford curve $X_\Gamma(\mathbf{k})$ replacing the Riemann surface $X_\Gamma = X_\Gamma(\mathbb{C})$ as the conformal boundary at infinity of $T_{\mathbf{k}}/\Gamma$.

Geodesics in the bulk space $T_{\mathbf{k}}/\Gamma$ correspond to images in the quotient of infinite paths without backtracking in the tree $T_{\mathbf{k}}$, with endpoints at infinity on $\mathrm{P}^1(\mathbf{k})$. Again, one can subdivide these in several cases. When the endpoints are the attractive and repelling fixed points $z_\gamma^\pm$ of some element $\gamma \in \Gamma$, the path in $T_{\mathbf{k}}/\Gamma$ is a closed loop in the finite graph $G_{\mathbf{k}}$. If the endpoints are both in Λ_Γ but not the fixed points of the same group element, then the geodesic is an finite path in $G_{\mathbf{k}}$ that is not a closed loop (but which winds around several closed loops in $G_{\mathbf{k}}$ without a fixed periodicity). If at least one of the endpoints is in $\Omega_\Gamma(\mathbf{k})$ then the path in $T_{\mathbf{k}}/\Gamma$ eventually (for either $t \to +\infty$ or $t \to -\infty$) leaves the finite graph $G_{\mathbf{k}}$ and wonders off along one of the attached infinite trees towards the boundary $X_\Gamma(\mathbf{k})$ at infinity. We still refer to these cases as closed, bounded, and unbounded geodesics, as in the archimedean case. We refer the reader to [95], [145], [157], [196] for a more detailed account of the geometry of Mumford curves.

4.3 Tensor networks

Motivated by the idea that the Bruhat–Tits tree T_p is a discrete (while still geometric) analogue of anti-de Sitter space, we will use this section to consider some relations between tensor networks that have been considered in the literature and the tree. One might imagine that many such relations can be drawn, and we have made no effort to be exhaustive; indeed, a primary aim in writing this paper is to bring the p-adic geometry to the attention of researchers working on tensor-network approaches to holography, who (we hope) will find it both interesting and useful.

Throughout this section the basic Hilbert spaces in the bulk and the boundary will be those of finite dimensional qudits and the primary object of study will

be the entanglement structure. In §4.4, the finite dimensional Hilbert spaces are replaced with those of a field theory valued in $\mathbb{R}$ or $\mathbb{C}$. We will find many aspects of holography hold in this field theoretic model and provide evidence for an exact correspondence. This connection puts the tensor network models of holography on a more equal footing with dynamical models, since both are defined from the same discrete spacetime.

We will focus our attention on the networks used by Pastawski, Yoshida, Harlow, and Preskill [172], (or "HaPPY"), in their construction of holographic quantum error-correcting codes. Such codes are easy to describe and admit many variations; in the simplest case, they are associated to a regular hyperbolic tiling of the plane. We will refer to such tilings by their *Schläfli symbol;* the notation $\{s, n\}$ refers to a tiling in which n regular polygons, each with s sides, meet at every vertex. A simple calculation shows that the tiling is hyperbolic whenever

$$n > \frac{2s}{s-2}. \tag{4.18}$$

For instance, if pentagons are used, $n = 4$ is the smallest possible choice ($n = 3$ would give the dodecahedron). If the tiles are heptagons or larger, any $n \geq 3$ gives a hyperbolic tiling.

Due to constraints of space, we will not fully review the HaPPY construction here; for details, the reader is referred to the original paper. The key point is that each tile carries a *perfect tensor*, which has an even number of indices, each of which refers to a qudit Hilbert space of fixed size. Such tensors are characterized by the property that any partition of the indices into two equal sets yields a maximally entangled state; we review perfect tensors in more detail, and construct a family of them associated to finite fields, in §4.3. Due to the appearance of finite fields in the construction of the tree, we feel this is the most natural family of tensors to consider.

The gist of our argument is that the natural "geometric" setting of a HaPPY tensor network (for certain uniform tilings) is the Bruhat–Tits tree corresponding to a prime p. We are motivated in this argument not only by the algebraic similarity between the constructions of T_p and AdS_3, but also by the fact that field-theoretic models of holography can be defined on the tree which exhibit it as the natural discrete setting for the AdS/CFT correspondence. In particular, the Bruhat–Tits tree with all edges of equal length can be thought of as a discrete analog of (empty) Euclidean AdS_3, and we conjecture that it is

dual to the vacuum of the CFT living on $\mathbb{Q}_p$. It is therefore logical to guess that HaPPY tensor networks naturally encode information about the entanglement structure of the conformal field theories living at the finite places, and not about the CFT living at the Archimedean place.

The precise connection we identify is that, at least for certain choices, the tiling used in HaPPY's code (when thought of as a graph) has a spanning tree that is a Bruhat–Tits tree. In some sense, therefore, the tree represents the union of as many geodesics as can be marked on the tiling without creating closed paths in the bulk.

In HaPPY's original paper, a "greedy algorithm" (related to reconstruction of the quantum state input at the bulk or "logical" qubits of the code) is used to define a region of the bulk, the perimeter of which is then called a "geodesic." We will show in §4.3 that tree geodesics can be understood to correspond to these "greedy" geodesics for the $\{5, p+2\}$ tilings.

An analogue of the Ryu-Takayanagi formula holds for these codes, essentially because the length of the geodesic counts the number of bonds (contracted tensor indices) that cut across it, and—due to the properties of perfect tensors—each contributes a constant amount (the logarithm of the qudit dimension) to the entanglement entropy. For us, it will be crucial to note that the length of a (unique boundary anchored) tree geodesic is related to the p-adic size of the boundary region it defines. We will elaborate on this in §4.5; for now, we will simply remark on a few features of the formula that we will need in this section.

At the Archimedean place, entanglement entropy measures the entanglement between the degrees of freedom living on a spatial domain A of a QFT and those living on the complement A^c. In AdS_3, the domain A is usually taken to be an interval, or a collection of intervals. The finite place analogue of an interval is just an open ball (as defined in §4.2); the notion of "codimension" is counterintuitive in the p-adic setting! One can see that there is no topological difference between (for instance) an open subset of the unit circle, which would be an interval in the normal case, and a generic open subset.

If the Ryu-Takayanagi formula holds, the entanglement entropy between A and A^c is given by the length of the unique geodesic $\gamma\left(x_A, y_A\right)$ in the Bruhat–Tits

tree connecting boundary points x_A and y_A,

$$S_{\mathrm{EE}}(A) = \# \cdot \mathrm{length}\left(\gamma\left(x_A, y_A\right)\right). \tag{4.19}$$

Just as in the case of AdS_3, entanglement entropy is a logarithmically divergent quantity. The divergence arises because if x_A and y_A are on the boundary, the number of legs on the geodesic is infinite. To regularize this divergence, we introduce a cutoff ϵ_p such that the length of the geodesic is

$$\mathrm{length}\left(\gamma\left(x_A, y_A\right)\right) = 2\log_p \left|\frac{x_A - y_A}{\epsilon_p}\right|_p, \tag{4.20}$$

with $|\cdot|_p$ the p-adic norm. (For details about this, see §4.5.) This gives the entanglement entropy between A and its complement in $\mathbb{Q}_p$ as*

$$S_{\mathrm{EE}}(A) = \# \log_p \left|\frac{x_A - y_A}{\epsilon_p}\right|_p. \tag{4.21}$$

The proportionality constant will be left undetermined for now.

Perfect tensors and quantum error-correcting codes from finite fields

Our goal in this subsection is to recall some features of quantum error-correcting codes associated to $2n$-index perfect tensors, as used by Pastawski, Yoshida, Harlow, and Preskill [172]. We will review the three-qutrit code and associated four-index perfect tensor that they construct, and then show how this is the case $q = 3$ of a family of perfect tensors associated to powers $q = p^m$ of odd primes. While the corresponding quantum error-correcting codes are not new [102], our goal is to highlight the properties of these particular codes that make them relevant to p-adic holography. In particular, as we recalled in §4.2, each vertex of the Bruhat–Tits tree for a degree-n unramified extension k of $\mathbb{Q}_p$ is marked with a copy of the residue field $\mathbb{F}_{p^n}$. As such, finite fields appear as important ingredients both in the construction of holographic tensor networks and in the algebraic setting of the Bruhat–Tits tree. We feel that the codes discussed here are natural candidates to consider in connecting p-adic geometry to tensor network models, although of course this choice is not inevitable and any code with the right properties will define a holographic code.

*For the length of the geodesic between x_A and y_A to warrant the interpretation of entanglement entropy, it must be the case that the tensor network bonds cutting across it, when extended all the way to the boundary, connect between A and A^c. This can be done and is explained in §4.3.

The three-qutrit code

In their paper, Pastawski *et al* consider the following quantum error-correcting code, which encodes a one-qutrit logical Hilbert space in a three-qutrit physical Hilbert space:

$$\begin{aligned}|0\rangle &\mapsto |000\rangle + |111\rangle + |222\rangle \\ |1\rangle &\mapsto |012\rangle + |120\rangle + |201\rangle \\ |2\rangle &\mapsto |021\rangle + |102\rangle + |210\rangle \,.\end{aligned}$$

The encoded data is protected against erasure of any single qutrit. If we represent the state by a tensor,

$$|a\rangle \mapsto T_{abcd}\,|bcd\rangle \,,$$

then T_{abcd} is perfect in the sense of Pastawski *et al*, and defines a perfect state:

$$|\psi\rangle = T_{abcd}\,|abcd\rangle \,.$$

(Throughout, we use Einstein's summation convention.) To recall, a tensor with $2n$ indices, each representing a qudit Hilbert space of any chosen fixed size, is perfect when it satisfies any of the following equivalent conditions:

- Given any partition of the indices into two disjoint collections $A \sqcup B$, where $|A| \leq |B|$, the tensor defines an injection of Hilbert spaces $\mathscr{H}_A \hookrightarrow \mathscr{H}_B$: a linear map that is a unitary isomorphism of its domain with its image (carrying the subspace norm).
- The corresponding perfect state is maximally entangled between any tensor factors $\mathscr{H}_{A,B}$ of equal size (each consisting of n qudits). That is, after tracing out n of the $2n$ qudits, the remaining n-qudit density matrix is proportional to the identity operator.

It is straightforward to check that the above tensor T_{abcd} is an $n = 2$ perfect tensor on qutrits. To rephrase the way it is constructed so as to make its generalization to larger codes more apparent, we notice that the particular states that appear in the encoding of a basis state $|a\rangle$ are lines of slope a in $\mathbb{F}_3^2$: if the three qudits are labeled by an element x of F_3, then the states are of the form $\otimes_x |f(x)\rangle$, where $f(x) = ax + b$, and we sum over the three

possible choices of $b \in \mathbb{F}_3$. The result is a perfect tensor because a line is determined either by two of its points, or by one point and knowledge of the slope; conversely, given any two points, or any one point and one slope, exactly one corresponding line exists.

Perfect polynomial codes

We would like to generalize this to a family of perfect tensors in which the qudit Hilbert spaces are of size $q = p^m$, so that a basis can be labelled by the elements of $\mathbb{F}_q$. An obvious guess is to associate a function or collection of functions $f_a : \mathbb{F}_q \to \mathbb{F}_q$ to each logical basis state $|a\rangle$, generalizing the collection of lines $f_a(x) = ax + b$ that were used when $q = 3$. These functions should have the property that knowledge of some number of evaluations of f_a will uniquely specify a, whereas knowing any smaller number of evaluations will give no information about a whatsoever. The encoded states will then take the form $\sum_b (\otimes_x |f_a(x)\rangle)$, for some collection of x's in $\mathbb{F}_q$. Here b stands for a collection of numbers parameterizing the set of functions f_a.

The simplest choice of such a class of functions are polynomials of fixed degree d:

$$f_a(x) = ax^d + b_{d-1}x^{d-1} + \cdots + b_1 x + b_0.$$

Over the real numbers, $d+1$ points determine such a polynomial. Over finite fields, one must be a little careful—by Fermat's little theorem,

$$x^q - x = 0 \ \forall x \in \mathbb{F}_q.$$

As such, if $d \geq q$, we can't determine a polynomial uniquely by its evaluations—after all, there are at most q possible evaluations over a finite field! However, polynomials of degree $d < q$ can be recovered uniquely; in fact, every function from $\mathbb{F}_q$ to $\mathbb{F}_q$ is a polynomial function, uniquely represented by a polynomial of degree $d < q$ (there are exactly q^2 elements of each collection).

However, if we choose d too large, the resulting code will not have error-correcting properties: we will need almost all of the physical qudits to recover the logical one. We know that for codes obtained from $2n$-index perfect tensors, one logical qudit is encoded in $2n-1$ physical qudits, and is recoverable from any n of them. This is sometimes called a $[[2n-1, 1, n]]_q$ code. For polynomial codes, we must have $2n-1 \leq q$ (since there are at most q possible evaluations of the code function), and furthermore $n = d+1$. Thus, the largest

possible perfect tensor we can obtain from this class of codes has $p+1$ indices, corresponding to the $[[q, 1, (q+1)/2]]_q$ code; the polynomials used in making this code are of degree $d = (q-1)/2$. Recall that we are assuming $p \neq 2$ everywhere; $q = 3$ recovers the linear qutrit code that we discussed above.

To be concrete, when $q = 5$, the code takes the following form:

$$|a\rangle \mapsto \sum_{b_0, b_1 \in \mathbb{F}_5} |b_0, b_0 + b_1 + a, b_0 + 2b_1 + 4a, b_0 + 3b_1 + 4a, b_0 + 4b_1 + a\rangle \,.$$

The numbers that appear are just x as the coefficient of b_0 and x^2 as the coefficient of a. This encoded state already contains 25 basis states, and the perfect state $|\psi_5\rangle$ constructed from this tensor is a combination of $5^3 = 125$ basis states.

These $p+1$-index perfect tensors seem like logical candidates to use in constructing a family of quantum codes associated to Bruhat–Tits trees. In particular, they are naturally associated to the data of a finite field $\mathbb{F}_q$, which appears at each vertex of the tree; moreover, they have $q+1$ qudit indices, which agrees with the valence of the tree.

However, the exact way to combine these ingredients remains a little unclear. In particular, since paths in the tree correspond to geodesics in the p-adic hyperbolic space, it seems more natural to think of the legs of the tree as *cutting across* contractions of tensor indices, rather than representing them. We expand on this idea in the section that follows.

Bruhat–Tits trees and tensor networks

We now investigate the connection between Bruhat–Tits trees and tensor networks. The gist of this section is that, while the tree corresponds to "geometry," the tree alone cannot define a tensor-network topology in the most naive way (tensors at vertices with indices contracted along edges). This is because, in typical tensor-network models of holography, the Ryu-Takayanagi formula holds because each unit distance along a geodesic corresponds to a bond (i.e., contracted tensor index) which is "cut" by the path and contributes a fixed amount (the logarithm of the dimension of the qudit Hilbert space) to the entanglement entropy. Since the paths in the tree correspond to geodesics in the bulk, one cannot hope to connect the tree to HaPPY's holographic code without adding tensors in such a way that their indices are contracted *across* the edges of the tree.

The extra structure we need to account for the network can be as simple as grouping the vertices of the tree in some fashion, associating bulk indices to the groups, and demanding perfect tensor structure, as we now explain.

> **A basic set of rules for constructing entanglement:** Group the vertices in the tree in some way; to each grouping we associate one or more bulk vertices. If two groupings share two tree vertices, then there is a tensor network bond connecting the bulk vertices of the groupings. The resulting tensor network should be composed of perfect tensors. This constructs a tensor network mapping between the boundary and the bulk.

It is not clear what the most general rules for associating the tensor vertices to the tree vertices should be. In particular, we are not demanding planarity (the Bruhat–Tits tree has no intrinsic planar structure), so the resulting network could be quite complicated, or even pathological. In order for the nice properties of a bulk-boundary tensor network to hold additional criteria should likely be imposed. We leave the general form of these criteria for future work; in the following, we focus on one specific set of rules that works.

Bruhat–Tits spanning trees of regular HaPPY tilings

Although the most general set of rules for assigning tensors is unclear, HaPPY tensor networks of uniform tiling can easily be constructed from the minimal proposal above with the addition of a few simple rules. These extra rules introduce planarity, so that the Bruhat–Tits tree becomes the spanning tree of the graph consisting of the edges of the HaPPY tensor network tiles. For $q > 3$, we can construct a HaPPY tensor network associated to a $[[q, 1, (q+1)/2]]_q$ code by grouping the vertices of the tree into sets of q, corresponding to tessellation tiles, and adding one bulk vertex to each tile. These tiles are organized into "alleyways;" each tile consists of vertices connected by a geodesic for tiles that are the starting points of alleyways, or of vertices living on two geodesics for tiles along the alleyway (see Fig. 4.5). The edges of each tile consist of either $q-1$ or $q-2$ segments coming from the geodesics, and one or two fictitious segments (the dotted lines in Fig. 4.5) respectively, that we draw only to keep track of which tree vertices have been grouped. Furthermore, each vertex connects to exactly one dashed edge. Since the tree has valence $q+1$, the HaPPY tensor network tiling has $q+2$ tiles meeting at each vertex.

The description above works for $q > 3$. The case $q = 2$ is special and can be obtained from the $[[5, 1, 3]]_2$ code; this is the case depicted in Fig. 4.5. In fact, any size polygon could be used; the only real constraint is that the tiling be hyperbolic of the form $\{n, q + 2\}$, with q a prime power. The pathologies of low primes come from the difficulty in demanding the tiling be hyperbolic and requiring perfect tensors; for instance, the $p = 2$ case would require a 3 index perfect tensor, but all perfect tensors have an even number of indices by construction.

In this picture of tiles, the tensor network bonds can be thought of as cutting across the edges of the tiles. Indeed, because of planarity, each edge can be associated with the tensor network bond of its vertices, precisely reproducing the HaPPY construction for uniform tilings of the hyperbolic plane.[†]

An interesting feature of our construction is that it introduces a peculiar notion of distance on the boundary, in that points $x, y, \in \mathbb{Q}_p$ that are that are far apart (in terms of the norm $|x - y|_p$) can belong to the same tile, or to neighboring tiles, so they can be "close in entanglement"; this is a concrete manifestation of the dissociation between entanglement and geometry inherently present in our model.

Explicit tree-to-tessellation mapping

We now explicitly construct an identification between Bruhat–Tits trees and a HaPPY tensor network of uniform tiling. The end goal is to show that a planar graph of uniform vertex degree v admits a spanning tree of uniform vertex degree $v - 1$. Although both the degree of the tree and the size of the tile are constrained by the quantum error correcting code, for the sake of generality we will work with n-gonal tiles, $n \geq 5$, and trees of valence k, $k \geq 3$. The algorithm constructing the mapping proceeds by starting with one n-gonal tile, then builds regions of tiles moving radially outward. Each region is built counterclockwise.[‡]

The purpose of this algorithm is to build a graph of dashed and solid edges such

[†]We should always remember, however, that in our construction, unlike in [172], bulk indices and tensor network connections are fundamentally associated to groups of tree vertices, and not to the geometric elements of a tile.

[‡]There are, of course, many variations of this algorithm that work; here we only exhibit one of them. For the purposes of constructing the mapping, it does not matter which variation we use.

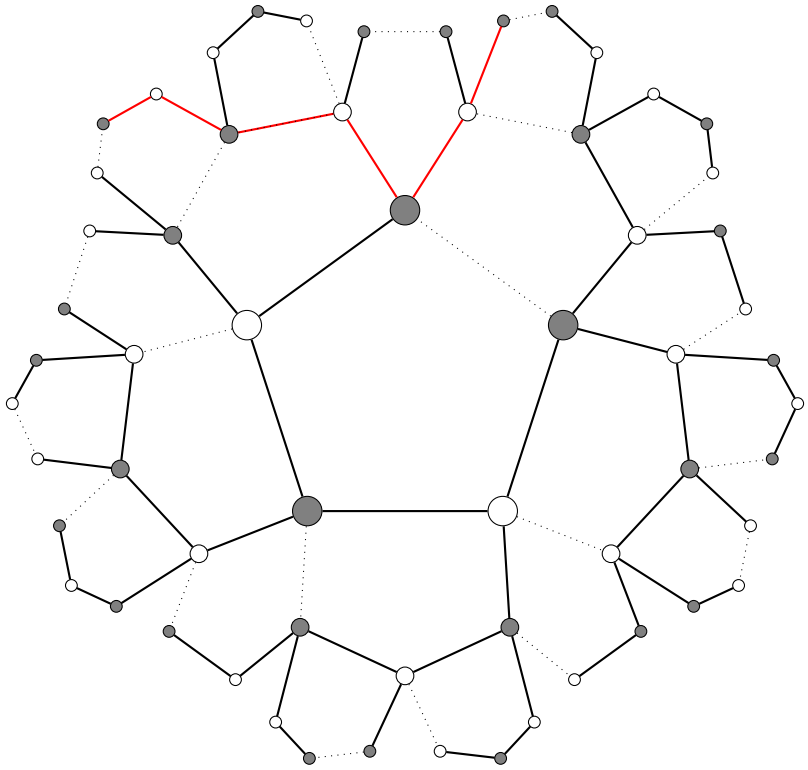


Figure 4.5: Mapping between a $p = 2$ Bruhat–Tits tree and the regular hyperbolic tiling $\{5, 4\}$. The first three regions constructed by the algorithm are shown. The red geodesic separates the causal wedge for the boundary region on which the geodesic ends from its complement in the tree. Since this is $p = 2$, the number of edges in a tile is different than our standard choice in the arbitrary q case.

that every vertex has degree $k+1$ and exactly one dashed edge connecting to it. The HaPPY tensor network tiling is given by the solid and dashed edges, and the tree is given by the solid edges, as in Fig. 4.5. The steps of the algorithm are as follows (see also Fig. 4.6 for a pictorial representation):

1. Start with an n-gonal tile with one edge dashed and $n - 1$ solid edges. This is region $r = 1$. The left vertex of the dashed line is the current vertex.

2. To construct region $r+1$, for the current vertex, add an edge e_f extending outward, then go counterclockwise around the tile being created, breaking off the edges shared with region r as soon as a vertex with degree less than $k + 1$ is encountered. Call the first new edge after breaking off e_n. If either e_f or e_n are constrained (by the condition that in the graph

we want to obtain each vertex has degree $k+1$ and precisely one dashed edge connecting to it) to be dashed, make them dashed, otherwise they are solid. If neither e_f nor e_n are dashed, make the "farthest" edge (call it e_l) dashed; otherwise, leave it as a solid edge. e_l is chosen so that its distance to the existing graph is as large as possible, and so equal on both sides if possible; if the number of new edges is even, so that this prescription is ambiguous, the choice closer to e_f is taken.

3. Move counterclockwise to the next vertex on the boundary of the current region, skipping any vertices of degree $k+1$. This new vertex becomes the current vertex.

4. Repeat the step above until we have built an edge e_f on all vertices of region r that have degree less than $k+1$. This completes region $r+1$.

5. To start on the next region, set the current vertex to be the left vertex of the dashed line on the first $r+1$ tile that we built, then go to step 2 above.

We now show why the algorithm works:

- By induction, there can be no neighboring vertices of degree greater than two on the boundary at any step, except when building a tile on the next-to-last edge of a tile from the previous region, in which case a vertex of degree 3 neighbors a vertex of some degree. This is because if the boundary has no neighboring vertices of degree $k+1$, any tile we add shares with the boundary of the previously constructed tiles at most two edges, so (since $n \geq 5$) it will have at least three free edges, adding at least two vertices of degree 2 between the vertices to which it connects.

- From the previous point, when constructing any tile, the vertices to which e_f and e_n connect cannot both have degree k before adding e_f and e_n, so either e_f or e_n can be made solid.

- If e_f and e_n are not dashed, then e_l only connects to two solid edges, so it can be made dashed.

- By the above, each new tile we add introduces exactly one dashed edge, so the graph of solid edges remains a tree at all steps.

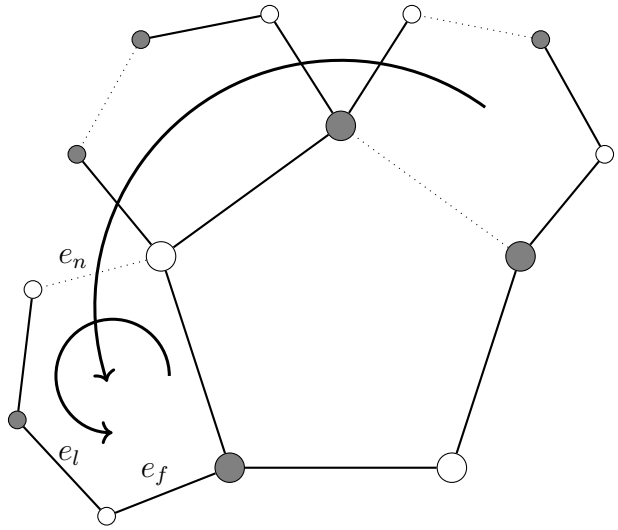


Figure 4.6: Mapping between a $p = 2$ Bruhat–Tits tree and a pentagonal HaPPY tiling, after the third tile of the second region has been built. Edge e_n is constrained to be dashed, so edges e_f and e_l are solid. The arrows represent the direction in which the algorithm constructs regions and tiles.

This completes the proof.

Bulk wedge reconstruction

In this subsection we discuss how bulk reconstruction, in the sense of [172], functions for our proposal. Although the construction in §4.3 replicates the tensor network tiling of [172], there are some differences of interpretation.

The algorithm outlined in Pastawski *et al.*'s paper prescribes that, starting with a given boundary region, one should add tiles one by one if the bulk qubits they carry can be reconstructed from the known data; i.e., if one knows (or can reconstruct) a majority of the edge qubits on the tile. When no further tiles can be added in this manner, the reconstruction is complete, and the boundary of the region is the "greedy geodesic."

In our case, for a given drawing of the tree, the first step to reconstructing a boundary open set $|x_A - y_A|_p$ is to identify a geodesic $\mathcal{G}$ that separates the causal wedge for the boundary region on which $\mathcal{G}$ ends from its complement in the tree. This is nontrivial, because due to the non-planarity of the tree, paths that end on the ball corresponding to the endpoints of the geodesic can be drawn outside the wedge. A $\mathcal{G}$ with the desired property is a geodesic

from which only one path leaves into the complement of the wedge; such an example is drawn in Fig. 4.5 in red. Given a choice of planar structure, there is a unique "outermost" separating geodesic associated to each open set. Once a separating geodesic has been identified, we can assign x_A and y_A to its endpoints, the inside of the ball to the tree inside the wedge, and the complement of the ball to the complement of the wedge in the tree.

For the tilings $\{5, n\}$, it is straightforward to see that the greedy geodesic for a boundary open ball coincides with the tree geodesic $\mathcal{G}$ that forms the "boundary" of that open ball in the chosen planar structure. The alleyways in the diagram are sequences of tiles joined along dashed lines, such that fewer than half of the edges on each tile are exposed to either side; therefore, each alleyway forms a "firebreak," which the greedy algorithm cannot jump across. If none of the dashed edges are known, none of the tiles in the alleyway can be reconstructed. Therefore, starting at a boundary open set, the greedy algorithm propagates up the alleyways whose ends lie inside the region, and fills out the region marked off by the tree geodesic. It cannot stop before the region is filled, since by construction each tile neighbors at least three tiles that are further away from the center than the tile itself is.

When the tiles are larger than pentagons, a difficulty arises when one is building tiles for which two edges touch the previous part of the picture: one may be forced (by valence requirements) to build a dashed edge at e_n or e_f, while another dashed edge exists in the previously built part, a distance of only one away. An instance where this occurs (although, of course, it causes no problems for pentagonal tiles) can be seen at the bottom right corner of Fig. 4.5. This constructs an alley that could be jumped by the greedy algorithm.

A simple fix for this problem would be to simply use a pentagon whenever this situation arises, resulting in a nonuniform tiling where the alleyways still function as firebreaks. While this will produce a valid holographic code, it is not immediate that the tiling is even regular in this case.

Another option is, if we are willing, to alter the tiling near a specifically chosen geodesic, so that the problem does not arise for that particular wedge. We explain an algorithm to construct the tiling in this case. The idea is to construct two alleyways, with the sides with one edge per tile pointing towards the wedge. This separates the plane into two regions, which for the purposes of tile building don't talk to each other. Since the rules for building a tile

from §4.3 are (almost) local, they have no information the global structure of the row being built. It is thus possible to use them to cover the two regions, moving "left" and "right" to create rows, and "up" and "down" to stack the rows. We give the explicit steps of the algorithm (see Fig. 4.7):

1. Build an alley of k-gons, by starting with a k-gon with one dashed edge (call this the root tile) and building the successive gons always on the dashed edges. For each k-gon except for the starting one, the number of solid edges on the two sides of the dashed edges should be 1 and $k-3$ respectively, with all 1's occurring on one side, and all $(k-3)$'s on the other.

2. Build a second alley of k-gons, starting on the solid edge of the root tile that neighbors the root's dashed edge on the side of 1's. For each k-gon except the starting one, the number of solid edges on the two sides of the dashed edges should be 1 and $k-3$ respectively, with all 1's facing the 1's of the first alley. The plane has now been split into two regions: wedge and complement.

3. To construct the tiling inside the wedge, start on some edge of the two central alleyways, and construct a tile using the rules from point 2 in the algorithm in §4.3. Then move to the right, and construct new tiles rotating clockwise in the construction of each tile. To do the other side, start from the initial tile and move left, constructing tiles using the rules from point 2 in the algorithm while rotating counterclockwise. This fills a "row".

4. Once the "row" is complete, move to the row "above" it and repeat.

5. To construct the tiling of the complement of the wedge, run the two bullet points above, but having the clockwise and counter-clockwise rotations swapped, and moving "below" instead of "above".

This algorithm works because locally the construction is the same as the one in the algorithm of §4.3. The individual tile building procedure does not depend on whether it is going around a finite region (as in §4.3), or along an infinite "row". And since inside the wedge more than half of each tile's neighbors are further away from the center than the tile is, the reconstruction covers the entire wedge.

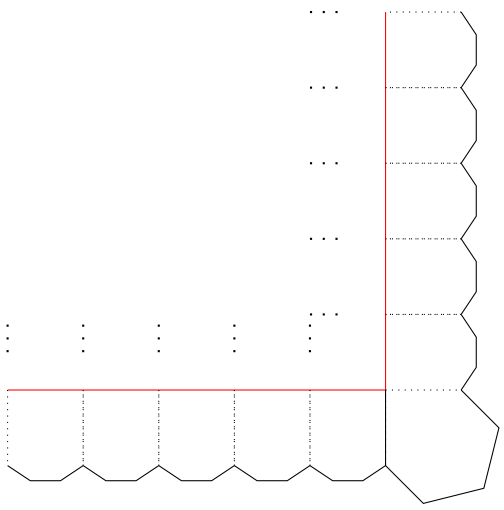


Figure 4.7: Mapping between the tree and tiling for reconstructing a chosen causal wedge. The two central alleys partition the plane into two regions: the causal wedge of the red geodesic and its complement. Tilings can be constructed to either side of the shown alley by building tiles via step 2 of the algorithm of §4.3.

While in the original HaPPY construction [172] one tensor network suffices to reconstruct the causal wedge associated to any boundary interval, in our case the tree identifies a certain collection of open sets on the boundary when pentagonal tiles are used (and, in one possible generalization to larger tiles, may even treat one boundary region as special). Even for pentagons, there may be many ways to draw the spanning tree on the same tiling. Moreover, the boundary tiles are not treated on an equal footing: they form the ends of shorter and longer alleyways. The longer the alleyway in which a boundary tile appears, the larger the first open set that includes it. One can think of these extra choices as follows: If one were to draw all possible greedy geodesics on the tiling, all edges would be marked, and there would be no information. Marking the geodesics with a subgraph is only useful when there is a unique path between pairs of boundary points, so that one knows "which way" to turn in order to recover the geodesic. This means that the marked subgraph should have no closed loops: that is, it should be a tree, and the spanning tree is (by definition) a largest possible subtree.

Entanglement bridges

By choosing a planar assignment of tensors, we have found the Bruhat-Tits tree as the spanning tree of this tiling. Although this choice was convenient, it

may be somewhat arbitrary. One might ask for more exotic non-planar way of connecting the tree with tensors. Such a non-planar structure might obscure the geometric interpretation, but it is expected that quantum gravity contains non-geometric states, so there is a sense in which at least small deviations from planarity should be physically acceptable. One simple example might be to connect two distant parts of the tree together through a non-planar tensor. If this can be done in a consistent way, one might interpret this non-planar defect a bridge of entanglement between two points of the tree. It would be interesting to investigate whether a configuration with defects or more complicated non-planar structures can be understood in terms of the ER = EPR proposal [144].

Furthermore, the rules we have identified here still generically break symmetries of the tree, since a $\mathrm{PGL}(2,\mathbb{Q}_p)$ transformation need not preserve the grouping of vertices or the planar structure. One might hope to construct a tensor network associated to the tree with more minimal auxiliary structure, so that the full symmetry group of the p-adic bulk spacetime is manifest for the network as well; however, for the reasons outlined above, it is difficult to understand how to do this while making contact with tensor network proposals existing in the current literature.

4.4 p-adic conformal field theories and holography for scalar fields

In this section, we turn from tensor networks to genuine field theories defined on p-adic spacetimes: either in the bulk of the tree T_p (or possibly a quotient by a Schottky group) or on a p-adic algebraic curve at the boundary. We will find evidence for a rich holographic structure strongly reminiscent of ordinary AdS/CFT. The conformal theory on the fractal p-adic boundary is analogous to 1+1 dimensional field theory with a p-adic global conformal group; our principle example is the p-adic free boson which permits a Lagrangian description. In the bulk, semi-classical massive scalar fields defined on the lattice model naturally couple to operators on the conformal boundary in a way that allows for precise holographic reconstruction. One can also interpret the radial direction in the tree as a renormalization scale. These observations unite discrete analogs of AdS geometry, conformal symmetry, and renormalization in a holographic way.

Generalities of p-adic CFT, free bosons, and mode expansions

While non-archimedean conformal field theory has been considered in the literature from several different perspectives [33, 45, 153], it remains much less well-studied than ordinary two-dimensional CFT. Melzer [153] defines these theories in general by the existence of an operator product algebra, where all operators in the theory are primaries with the familiar transformation law under the global conformal group $\mathrm{SL}(2,\mathbb{Q}_p)$. Descendants are absent in Melzer's formulation because there is no analogue of the derivative operators ∂ and $\bar{\partial}$ acting on complex-valued functions over $\mathbb{Q}_p$ [153], and (correspondingly) no local conformal algebra.

In this formulation, the correlation function between two primary fields $\phi_m(x)$ and $\phi_n(y)$ inserted at points x and y and having scaling dimensions Δ_n is given (after normalization) by

$$\langle \phi_m(x)\phi_n(y)\rangle = \frac{\delta_{m,n}}{|x-y|_p^{2\Delta_n}}. \tag{4.22}$$

(We will understand this formula holographically in what follows.) As in the archimedean case, as we take the points x and y to be close together (p-adically), we wish to expand the product as a sum of local field insertions: the operator product expansion. For two such primaries $\phi_m(x)$ and $\phi_n(y)$, there exists an $\epsilon > 0$ such that for $|x-y|_p < \epsilon$, the correlation function (perhaps with other primaries $\phi_{n_i}(x_i)$ inserted) is given by the expansion:

$$\langle \phi_m(x)\phi_n(y)\phi_{n_1}(x_1)\dots\phi_{n_i}(x_i)\rangle = \sum_r \tilde{C}^r_{mn}(x,y)\langle \phi_r(y)\phi_{n_1}(x_1)\dots\phi_{n_i}(x_i)\rangle, \tag{4.23}$$

where the sum runs over all primaries in the theory, and $\tilde{C}^r_{mn}(x,y)$ are real valued. This relation should hold whenever $|x-y|_p$ is smaller than the distances to the x_i's. Invariance under $\mathrm{SL}(2,\mathbb{Q}_p)$ implies

$$\tilde{C}^r_{mn}(x,y) = C^r_{mn}|x-y|_p^{\Delta_r-\Delta_m-\Delta_m} \tag{4.24}$$

with constant OPE coefficients C^r_{mn}.

Theories defined in this way enjoy a number of special properties not true of their archimedean counterparts. They are automatically unitary since they possess no descendant fields. Additionally, because $\mathbb{Q}_p$ is an *ultrametric* field, all triangles are isosceles: for $x, y, z \in \mathbb{Q}_p$, from the p-adic norm we have

$$\text{If } |x-y|_p \neq |y-z|_p \text{ , then } |x-z|_p = \max\{|x-y|_p, |y-z|_p\}. \tag{4.25}$$

This fundamental property of the p-adic numbers implies that the three- and four-point functions are *exactly* determined by the conformal weights and OPE coefficients. In the case of the four-point function, after an $\mathrm{SL}(2,\mathbb{Q}_p)$ transformation which maps three points to $0, 1$, and ∞, the only free parameter (the cross ratio of the original points) must be contained in a ball in the neighborhood of one of the other points. Since the OPE is exact in each neighborhood, one can compute the three possible cases and determine the full four-point function.

In fact, all higher-point functions are constrained by global conformal symmetry alone; by contrast, the spectrum of OPE coefficients is less constrained than in familiar CFTs. A consistent model can be constructed using the structure constants of any unital commutative algebra, subject to one simple condition. These features may be of interest in the study of conformal field theory and conformal blocks, but we do not pursue that direction here; the interested reader is referred to [153].

Let us now step back and consider the p-adic theory from the perspective of quantizing a classical field theory described by a Lagrangian. Many familiar objects from the study of quantum fields over normal (archimedean) spacetime have direct analogues in the p-adic setting. For example, one frequently makes use of the idea of a *mode expansion* of a field on flat spacetime in terms of a special class of basis functions, the plane waves:

$$\phi(x) = \int_{\mathbb{R}} dx\, e^{ikx}\tilde{\phi}(k). \tag{4.26}$$

The functions e^{ikx} are eigenfunctions of momentum, or equivalently of translations. Mathematically, we can think of these as *additive characters* of $\mathbb{R}$: they are group homomorphisms $\chi : \mathbb{R} \to \mathbb{C}$, such that $\chi(x+y) = \chi(x)\chi(y)$.

The additive characters of the fields $\mathbb{Q}_p$ are also known: they take the form [202]

$$\chi_k(x) = e^{2\pi i\{kx\}}. \tag{4.27}$$

Here $k, x \in \mathbb{Q}_p$, and the normalization factor 2π is included for convenience (in keeping with the typical math conventions for Fourier transforms). The

symbol $\{\cdot\} : \mathbb{Q}_p \to \mathbb{Q}$ denotes the *fractional part* of the p-adic number.[§] It is defined by truncating the decimal expansion to negative powers of the prime:

$$\left\{ \sum_{k=m}^{\infty} a_k p^k \right\} = \sum_{k=m}^{-1} a_k p^k, \tag{4.28}$$

where the right-hand side is interpreted as an ordinary rational number, understood to be zero when the range of the sum is empty (m is non-negative). Since a p-adic number and its fractional part differ (at least in a formal sense) by an integer, it makes sense that the complex exponential (4.27) should depend only on the fractional part of kx. (However, care should be taken: in general, it is not true for rational x that $e^{2\pi i x} = e^{2\pi i \{x\}_p}$! For instance, $0.1 = 1/10$ is a 3-adic integer.)

A wide class of scalar fields on $\mathbb{Q}_p$ can be expanded in a basis of the additive characters, just like a mode expansion in the archimedean setting:

$$\phi(x) = \int d\mu(k)\, e^{2\pi i \{kx\}}\, \tilde{\phi}(k). \tag{4.29}$$

Here $d\mu(k)$ is the Haar measure on $\mathbb{Q}_p$. The theory of the p-adic Fourier transform is developed in more detail in Appendix A.1.

Our principal example (and also by far the most well-studied instance) of a p-adic conformal field theory is the free boson: a single (real or complex) scalar field on $\mathrm{P}^1(\mathbb{Q}_p)$ or another p-adic Riemann surface, with a massless quadratic action. This theory was of interest in the context of p-adic string theory, in which the worldsheet is a p-adic space, but the target space (and hence all physically observable quantities) are ordinary. Many results were derived in that literature, including the well-known Freund–Olson–Witten tachyon scattering amplitudes [34, 86, 87].

Our interpretation of the system in question will be somewhat different, as we will emphasize the holographic nature of the interplay between field theory defined on a Riemann surface (algebraic curve) and the study of its hyperbolic filling, a quotient of the Bruhat–Tits tree. (In the p-adic string literature, it was common to view the tree as playing the role of the "interior" of the worldsheet.) Many of our results will parallel aspects of the p-adic string, but

[§] As with other notations referring to the p-adics, we will sometimes use the subscript $\{\cdot\}_p$ when it is necessary for emphasis or to make reference to a specific choice of prime.

we will view this theory as a CFT on $\mathrm{P}^1(\mathbb{Q}_p)$ without any reference to a target space.

The p-adic free boson is considered here because it permits a Lagrangian description in terms of the nonlocal Vladimirov derivative, $\partial_{(p)}$, which acts on complex- or real-valued fields of a p-adic coordinate. This derivative is defined by

$$\partial_{(p)}^n f(x) = \int_{\mathbb{Q}_p} \frac{f(x') - f(x)}{|x' - x|_p^{n+1}} dx'. \tag{4.30}$$

In the p-adic string literature, $\partial_{(p)}$ is also known as a normal derivative, for reasons that will become clear in the following sections. Intuitively, the formula is similar to Cauchy's representation of the n-th derivative of a function by a contour integral. A more detailed explanation of its properties is given in Appendix A.2. While the parameter n is often taken to be an integer, it may in principle assume any real value.

One can arrive at the following action either by "integrating out" the interior of the string worldsheet T_p as done in [225], or by hypothesis as the minimal "quadratic" action of a scalar over a p-adic coordinate [71]. The action for a single scalar is (setting the overall coupling to 1) [101, 188, 226]:

$$S_p[\phi] = -\int_{\mathbb{Q}_p} \phi(x) \partial_{(p)} \phi(x)\, dx, \tag{4.31}$$

where $\partial_{(p)}\phi(x)$ is the first Vladimirov derivative of the field ϕ. We take $\phi(x)$ to be a scalar representation of the conformal group; see [182] for a discussion of representations of $\mathrm{SL}(2, \mathbb{Q}_p)$ in general. Under an element

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

of the conformal group, where $a, b, c, d \in \mathbb{Q}_p$ and $ad - bc = 1$, quantities in the above expression transform as

$$\begin{gathered} x \to \frac{ax+b}{cx+d}, \quad x' \to \frac{ax'+b}{cx'+d}, \\ dx \to \frac{dx}{|cx+d|_p^2}, \quad dx' \to \frac{dx'}{|cx'+d|_p^2}, \\ |x'-x|_p^{-2} \to |(cx'+d)(cx+d)|_p^2 |x'-x|_p^{-2}. \end{gathered}$$

As in [153], a field $\phi_n(x)$ having conformal dimension Δ_n transforms as

$$\phi_n(x) \to \phi_n'\left(\frac{ax+b}{cx+d}\right) = |cx+d|_p^{-2\Delta_n} \phi_n(x) \tag{4.32}$$

under the p-adic conformal group. For the free boson $\phi(x)$, we claim $\Delta = 0$. With this one can see the derivative $\partial_{(p)}\phi(x)$ carries a weight $|cx+d|_p^{-2}$ and thus is a field of dimension 1. It should now be clear that the action $S_p[\phi]$ is invariant under the global conformal group.

Given the action $S_p[\phi]$, we can define the partition function in the usual way by integrating over configurations with measure $\mathcal{D}\phi$. As in the case of the p-adic string, because ϕ is a complex (and not p-adic) valued field, this integration measure is exactly the one that appears in ordinary field theory:

$$Z_p = \int \mathcal{D}\phi \, e^{-S_p[\phi]}. \tag{4.33}$$

As many authors have noted [171, 188, 226], this action and the partition function actually describe a free theory. This means the saddle point approximation to the partition function is exact, and it can be computed by Gaussian integration exactly as in the case of a real free field. Of more interest in the present discussion is the two point function. To do this we introduce sources $J(x)$ to define the generating function:

$$Z_p[J] = \int \mathcal{D}\phi \, \exp\left(-S_p[\phi] + \int_{\mathbb{Q}_p} J(x')\phi(x')dx'\right). \tag{4.34}$$

The sources for the 2-point function or propagator take the form of p-adic delta functions at the insertion points x, y are $J(x) = \delta(x'-x) + \delta(x'-y)$. Just as in the real case, we vary with respect to $\phi(x)$ and find the classical solution which extremizes the above action. This is the Green's function for the Vladimirov derivative $G(x-y)$, satisfying

$$\partial_{(p)}G(x-y) = -\delta(x-y). \tag{4.35}$$

To solve for $G(x-y) = \langle 0|\phi(x)\phi(y)|0\rangle$, we apply the p-adic Fourier transform to both sides using techniques from Appendix A.1. In Fourier space the derivative brings down one power of the momentum and the delta function becomes an additive character:

$$\widetilde{G}(k) = -\frac{\chi(ky)}{|k|_p}. \tag{4.36}$$

The 2-point function in position space can be obtained by inverse Fourier transform (with $u = x - y$):

$$G(x-y) = -\int_{\mathbb{Q}_p} \frac{\chi(k(y-x))}{|k|_p} dk \tag{4.37}$$

$$= -\int_{\mathbb{Q}_p} \frac{\chi(ku)}{|k|_p} dk. \tag{4.38}$$

This integral is divergent as $k \to 0$. We compute two similar integrals in Appendix A.2, where the apparent divergence is canceled by the numerator. Unlike in those examples, this integral really does diverge logarithmically, just as the two-point function of a dimension 0 operator in two-dimensional conformal field theory has a logarithmic divergence. Proceeding as in that case, we introduce a regulator to extract the finite part by computing

$$\lim_{\alpha\to 0} \int_{\mathbb{Q}_p} \chi(ku)|k|_p^{\alpha-1} dk. \tag{4.39}$$

This appears in the second integral computed in the appendix; in terms of the p-adic gamma function $\Gamma_p(x)$, it is

$$\lim_{\alpha\to 0} \int_{\mathbb{Q}_p} \chi(ku)|k|_p^{\alpha-1}\, dk = \lim_{\alpha\to 0} \Gamma_p(\alpha)|u|_p^{-\alpha}. \tag{4.40}$$

As $\alpha \to 0$ the gamma function has a simple pole and the norm has a log piece:

$$\lim_{\alpha\to 0} \Gamma_p(\alpha) \approx \frac{p-1}{p\ln p}\frac{1}{\alpha} \tag{4.41}$$

$$\lim_{\alpha\to 0} |u|_p^{-\alpha} \approx 1 - \alpha \ln |u|_p. \tag{4.42}$$

Finally, we restore $u = x - y$ and find that the 2-point function is given (up to normalization) by

$$\langle 0|\phi(x)\phi(y)|0\rangle \sim \ln \left|\frac{x-y}{a}\right|_p, \quad a \to 0. \tag{4.43}$$

This is exactly analogous to the correlator for the ordinary free boson in two dimensions.

The Laplacian and harmonic functions on T_p

In addition to boundary scalar fields, we will be interested in scalar fields in the "bulk," i.e., defined on the Bruhat–Tits tree. Such a field is a real- or complex-valued function on the set of vertices. We will also consider fields

that are functions on the set of edges; as we will discuss later, such functions will be analogous to higher-form fields or metric degrees of freedom in the bulk. For now we mention them for completeness and to fix some standard notation. For more information about fields in the tree, the reader can consult [225] and references given therein.

We think of the tree as the 1-skeleton of a simplicial complex, and make use of standard notations and ideas from algebraic topology. The two types of fields mentioned above are just 0- and 1-cochains; we will refer to the space of such objects as $C^*(T_p)$, where $* = 0$ or 1.

If an orientation is chosen on the edges of the tree, the boundary operator acts on its edges by $\partial e = t_e - s_e$, where s and t are the source and target maps. The corresponding coboundary operator acts on fields according to the rule

$$d : C^0(T_p) \to C^1(T_p), \quad (d\phi)(e) = \phi(t_e) - \phi(s_e). \tag{4.44}$$

The formal adjoint of this operator is

$$d^\dagger : C^1(T_p) \to C^0(T_p), \quad (d^\dagger\psi)(v) = \sum_e \pm\psi(e), \tag{4.45}$$

where the sum is over the $p+1$ edges adjacent to vertex v, with positive sign when v is the source and negative sign when it is the target of e. Whether or not $d^\dagger$ is actually an adjoint to d depends on the class of functions being considered; the L^2 inner product must be well-defined, and boundary conditions at infinity must be chosen to avoid the appearance of a boundary term.

Upon taking the anticommutator $\{d, d^\dagger\}$, we obtain an operator of degree zero, which is the proper analogue of the Laplacian. We will most often use its action on the 0-cochains, which can be represented by the formula

$$\triangle\, \phi(v) = \sum_{d(v,v')=1} \phi(v') - (p+1)\phi(v). \tag{4.46}$$

This is sometimes written using the notation $\triangle_p = t_p - (p+1)$, where t_p is the Hecke operator on the tree. The analogous formula for 1-cochains is

$$\triangle\, \psi(e) = \sum_{e'} \pm\psi(e') - 2\psi(e), \tag{4.47}$$

where the sum goes over the $2p$ edges adjacent to e at either side. Unlike for the vertices, there is a dependence on the choice of orientation here: an edge in

the sum (4.47) enters with positive sign when it points in the "same direction" as e, i.e., points out from t_e or into s_e. Edges enter with negative sign when the opposite is true. In the standard picture of the tree with ∞ at the top and all finite points of $\mathbb{Q}_p$ at the bottom, oriented vertically, we therefore have exactly one negative term in (4.47), corresponding to the unique edge above e. Notice that, for general p, the Laplacian acting on edges (unlike on vertices) will not have a zero mode; this makes sense, since the tree is a contractible space. The exception is $p = 2$, for which the standard choice of vertical orientations defines a Laplacian which annihilates constant functions of the edges. (Of course, the $p = 2$ tree is still contractible.)

We should remark on one important point: the entire analysis of this paper treats the case where the edges of the tree have uniform lengths, and argues that this is analogous to a maximally symmetric vacuum solution in ordinary gravity. It is natural to wonder what the correct analogues of the metric degrees of freedom actually are. One might speculate that allowing the edge lengths to be dynamical (breaking the $\mathrm{PGL}(2, \mathbb{Q}_p)$ symmetry) should correspond to allowing the metric to vary; after all, this would vary the lengths of paths in the tree, which are the only data that seem logically connected to the metric. By analogy with the Archimedean case, it would then make sense to assume that the edge Laplacian (4.47) will play a role in the linearized bulk equations of motion for edge lengths around a background solution. However, we will relegate further investigation of this idea to future work.

Action functional and equation of motion for scalar fields

Equipped with these ingredients, it is now straightforward to write down action functionals and equations of motion for free scalar fields. The massless quadratic action is

$$S[\phi] = \sum_e |d\phi(e)|^2 \,. \tag{4.48}$$

In what follows, we will study properties of solutions to the "wave equation" $\triangle\phi = 0$, and its massive generalization $(\triangle - m^2)\phi = 0$, on the tree. These have been considered in [225].

There is a family of basic solutions to the Laplace equation, labeled by a choice of a boundary point x and an arbitrary complex number κ. The idea is as follows: Given an arbitrary vertex v in the bulk of the tree, a unique

geodesic (indeed, a unique path) connects it to x. As such, exactly one of its $p+1$ neighbors will be closer (by one step) to x, and the other p will be farther by one step. Therefore, the function

$$\varepsilon_{\kappa,x}(v) = p^{-\kappa\, d(x,v)} \tag{4.49}$$

will be an eigenfunction of the Hecke operator, with eigenvalue $(p^\kappa + p^{1-\kappa})$.

The catch in this is that the distance $d(x,v)$ is infinite everywhere in the bulk. We need to regularize it by choosing a centerpoint C in the tree, and declaring that $d(x,C) = 0$. (This just scales the eigenfunction (4.49) by an infinite constant factor). Then $d(x,v) \to -\infty$ as $v \to x$, but we have a well-defined solution to the Laplace equation everywhere in the tree. These solutions are analogous to plane waves; the solution varies as the exponent of the (regularized) distance to a boundary point, which in the normal archimedean case is just the quantity $\mathbf{k}\cdot\mathbf{r}$.

The corresponding eigenvalue of the Laplacian is

$$\triangle\, \varepsilon_{\kappa,x} = m_\kappa^2 \varepsilon_{\kappa,x} = \left[(p^\kappa + p^{1-\kappa}) - (p+1)\right] \varepsilon_{\kappa,x}. \tag{4.50}$$

It is therefore immediate that the harmonic functions on the tree (solutions to the massless wave equation) are those with $\kappa = 0$ or 1; $\kappa = 0$ is the zero mode consisting of constant functions, whereas $\kappa = 1$ is the nontrivial zero mode. The eigenvalues (4.50) are invariant under the replacement $\kappa \to 1-\kappa$, due to the inversion symmetry of the boundary theory.

If we are considering a real scalar field, we must be able to write a basis of real solutions. Of course, when κ is real, we will always be able to do this. More generally, if $\kappa = \kappa_0 + i\gamma$, our solutions look like

$$\varepsilon \sim p^{-\kappa_0\, d} e^{-i\gamma \ln(p)\, d}, \quad p^{(1-\kappa_0)d} e^{i\gamma \ln(p)\, d}. \tag{4.51}$$

Thus, to construct a basis of real solutions, the following possibilities can occur:

- $\kappa = 1/2$. In this case, there is no restriction on γ, and the solutions look like cosines and sines of $\gamma \ln(p)\, d(x,v)$, modulated by $p^{d(x,v)/2}$.
- $\kappa > 1/2$.[¶] In this case, the amplitude parts of the two solutions are linearly independent, and so $\exp(i\gamma \ln(p)\, d)$ must be real. Since d is an integer, the choices are $\gamma \ln(p) = 0$ or $\pi \pmod{2\pi}$.

[¶]Due to the $\kappa \mapsto 1-\kappa$ symmetry, such a choice is always possible.

While it would be interesting to consider solutions with nonzero γ, we will consider only the one-parameter family of solutions with real κ in the sequel. The parameter m_κ^2 then attains its minimum value for $\kappa = 1/2$. Considering only solutions of this plane-wave form, we therefore have a bound

$$m_\kappa^2 \geq -(\sqrt{p} - 1)^2. \tag{4.52}$$

Note that we could also rewrite (4.50) in the form

$$m_\kappa^2 = -(p+1) + 2\sqrt{p}\cosh\left[\left(\kappa - \frac{1}{2}\right)\ln p\right]. \tag{4.53}$$

Bulk reconstruction and holography

It is clear from the definition that, when the real part of κ is positive, the plane wave solution (4.49) tends to zero everywhere on the boundary, except at the point x (where it tends to infinity). So we can think of it as representing the solution to the Laplace equation (taking $\kappa = 1$) in the bulk, with specified Dirichlet-type boundary conditions that look like a delta function centered at x. By linearity, we can therefore reconstruct the solution to more general Dirichlet problems by superposition: if the boundary value is to be a certain function $\phi_0(x)$ on $\partial T_p = \mathrm{P}^1(\mathbb{Q}_p)$, then the required bulk harmonic function is

$$\phi(v) = \frac{p}{p+1}\int d\mu_0(x)\,\phi_0(x)\,\varepsilon_{1,x}(v). \tag{4.54}$$

Here $d\mu_0(x)$ is the Patterson-Sullivan measure on $\mathrm{P}^1(\mathbb{Q}_p)$. The normalization factor can be fixed by taking the boundary value to be the characteristic function of any p-adic open ball in the boundary.

We can perform the analogous calculation for massive fields as well, but the sense in which $\phi(v)$ will approach $\phi_0(x)$ as $x \to v$ will be more subtle (since the equation of motion will have no constant mode). Using notation from [225], let $\delta(a \to b, c \to d)$ be the overlap (with sign) of the two indicated oriented paths in the tree, and let

$$\langle v, x\rangle = \delta(C \to v, C \to x) + \delta(v \to x, C \to v). \tag{4.55}$$

This expression makes sense for any bulk vertex v; x may be either a boundary or a bulk point. Note that $\langle z, x\rangle$ is just the negative of the "regularized distance" occurring in our previous discussion.

We would like to compute the bulk solution to the massive equation of motion obtained by integrating our primitive solution (4.49) over its boundary

argument, weighted by a boundary function. As a simple choice of boundary function, pick the characteristic function of the p-adic open ball below a vertex w in the tree:

$$\phi_w(v) = \int_{\partial B_w} d\mu_0(x)\, p^{\kappa\langle v,x\rangle}. \tag{4.56}$$

The integral is straightforward to calculate. There are two cases:

$v \notin B_w$ Here, the integrand is constant, and is just equal to $p^{\kappa\langle v,w\rangle}$. The measure of the set over which the integral is performed is $\mu_0(\partial B_w) = p^{-d(C,w)}$, so that the final result is

$$\phi_w(v) = p^{\kappa\langle v,w\rangle - d(C,w)}. \tag{4.57}$$

Note that, if v moves towards the boundary along a branch of the tree, $\langle v,w\rangle$ differs from $-d(C,v)$ by a constant, so that the solution scales as $p^{-\kappa\, d(C,z)}$.

$v \in B_w$ There are now two cases to consider: $x \in B_v$ or $x \notin B_v$. In the first scenario, the integrand is again constant; its value is $p^{\kappa\, d(C,v)}$, and the measure is $\mu_0(B_v) = p^{-d(C,v)}$.

In the second scenario, the geodesic $x \to C$ will meet the geodesic $v \to C$ at a distance h above v; by assumption, $1 \le h \le d(v,w)$. For each value of h, the integrand takes the constant value $p^{\kappa(d(C,v)-2h)}$, and the measure of the corresponding set is

$$\mu(h) = \frac{p-1}{p} p^{-d(C,v)+h}. \tag{4.58}$$

The factor $(p-1)/p$ enters because $p-1$ of the p vertices one step below the meeting vertex correspond to meeting height h (one of them is closer to v). Putting the pieces together, the result is

$$\begin{aligned}\phi_w(v) &= p^{(\kappa-1)d(C,v)} \left(1 + \frac{p-1}{p} \sum_{h=1}^{d(w,v)} \left(p^{1-2\kappa}\right)^h\right) \\ &= \left(\frac{p^{-2\kappa}-1}{p^{1-2\kappa}-1}\right) p^{(\kappa-1)d(C,v)} + \frac{p-1}{p}\left(\frac{p^{(2\kappa-1)d(C,w)}}{p^{2\kappa-1}-1}\right) p^{-\kappa\, d(C,v)}.\end{aligned} \tag{4.59}$$

The reader can check that we recover the correct answer in the massless case, $\kappa \to 1$. Furthermore, our result is a superposition of the asymptotic behavior

of the two eigenfunctions corresponding to the mass determined by our original choice of κ. To resolve the ambiguity, we will choose $\kappa > 1/2$.

At this point, we have accumulated enough understanding of scalar fields on the tree to point out how the simplest version of holography will work: namely, classical scalar fields in a non-dynamical AdS background, neglecting backreaction and metric degrees of freedom. In the archimedean case, this version of holography was neatly formulated by Witten [216] in terms of a few simple key facts. Firstly, the coupling between bulk scalar fields and boundary operators must relate the asymptotics (and hence the mass) of the bulk fields to the conformal dimension of the corresponding boundary operators; massless bulk scalars should couple to marginal operators in the boundary CFT. Secondly, the crucial fact that allows the correspondence to work is the existence of a unique solution to the generalized Dirichlet problem for the bulk equations of motion with specified boundary conditions.

Luckily, as we have now shown, all of the important features of the problem persist in the p-adic setting, and Witten's analysis can be carried over kit and caboodle to the tree. In particular, we make his holographic ansatz:

$$\left\langle \exp \int_{\mathrm{P}^1(\mathbb{Q}_p)} d\mu_0 \, \phi_0 \mathscr{O} \right\rangle_{\mathrm{CFT}} = e^{-I_{\mathrm{bulk}}[\phi]}, \tag{4.60}$$

where the bulk field ϕ is the unique classical solution extending the boundary condition ϕ_0, and $\mathscr{O}$ is a boundary operator to which the bulk field couples. In the massless case, where one literally has $\phi_0(x) = \lim_{v \to x} \phi(z)$, $\mathscr{O}$ is an exactly marginal operator in the CFT.

Given our result (4.59), it is simple to write down the correctly normalized bulk-reconstruction formula for massive fields:

$$\begin{aligned} \phi(v) &= \frac{p^{1-2\kappa} - 1}{p^{-2\kappa} - 1} \int d\mu_0(x) \phi_0(x) p^{\kappa \langle v, x \rangle}, \\ \phi(v) &\sim p^{(\kappa - 1) d(C,v)} \phi_0(x) \text{ as } v \to x. \end{aligned} \tag{4.61}$$

When the point v approaches the boundary, the exponent in the kernel becomes

$$\langle v, x \rangle = -d(C, v) + 2 \operatorname{ord}_p(x - y), \tag{4.62}$$

where y is any boundary point below v. (4.61) then becomes

$$\phi(v) = \left(\frac{p^{1-2\kappa} - 1}{p^{-2\kappa} - 1} \right) p^{-\kappa \, d(C,v)} \int d\mu(x) \frac{\phi_0(x)}{|x - y|_p^{2\kappa}}. \tag{4.63}$$

We can now understand why the Vladimirov derivative is a "normal" derivative on the boundary; it measures the rate of change in the holographic direction of the reconstructed bulk function. In particular, we have that

$$\begin{aligned}\lim_{v\to y}(\phi(v)-\phi(y))p^{\kappa d(C,v)} &= \left(\frac{p^{1-2\kappa}-1}{p^{-2\kappa}-1}\right)\int d\mu(x)\frac{\phi_0(x)-\phi_0(y)}{|x-y|_p^{2\kappa}} \\ &= \left(\frac{p^{1-2\kappa}-1}{p^{-2\kappa}-1}\right)\partial_{(p)}^{2\kappa-1}\phi(y). \end{aligned} \tag{4.64}$$

An argument precisely akin to Zabrodin's demonstration [225] that the bulk action may be written (upon integrating out the interior) as a boundary integral of the nonlocal Vladimirov action shows that we can write $I_{\text{bulk}}[\phi]$ in exactly this form. This demonstrates, exactly as in Witten's archimedean analysis, that a massive field ϕ corresponds to a boundary operator of conformal dimension κ, where $\kappa > 1/2$ is the larger of the two values that correspond to the correct bulk mass. Moreover, the boundary two-point function is proportional to $|x-y|_p^{-2\kappa}$, as expected.

Scale dependence in bulk reconstruction of boundary modes

Let us consider how the mode expansion of a boundary scalar field interacts with the reconstruction of the corresponding bulk harmonic function. We will be interested in developing the interpretation of the extra, holographic direction as a renormalization scale in our p-adic context. The idea that moving upward in the tree corresponds to destroying information or coarse-graining is already suggested by the identification of the cone above $\mathbb{Z}_p$ (or more generally any branch of the tree) with the inverse limit

$$\mathbb{Z}_p = \varprojlim \mathbb{Z}/p^n\mathbb{Z}, \tag{4.65}$$

where the set of vertices at depth n corresponds to the elements of $\mathbb{Z}/p^n\mathbb{Z}$, and the maps of the inverse system are the obvious quotient maps corresponding to the unique way to move upwards in the tree. A nice intuitive picture to keep in mind is that p-adic integers can be thought of as represented on an odometer with infinitely many $\mathbb{F}_p$-valued digits extending to the left. $\mathbb{Z}/p^n\mathbb{Z}$ is then the quotient ring obtained by forgetting all but n digits, so that there is integer overflow; the maps of the inverse system just forget successively more odometer rings. Since digits farther left are smaller in the p-adic sense, we can think of this as doing arithmetic with finite (but increasing) precision. The

parallel to the operation of coarse-graining is apparent; however, we will be able to make it more precise in what follows.

Let's consider a boundary field that is just given by an additive character (plane wave), $\phi_0(x) \sim \exp(2\pi i\{kx\}_p)$. Just as in the complex case, a plane wave in a given coordinate system won't define a solution of fixed wavelength everywhere on $\mathrm{P}^1(\mathbb{C})$; the coordinate transformation (stereographic projection) will mean that the wavelength tends to zero as one moves away from the origin, and the function will become singular at infinity. Therefore, we should instead consider a boundary function of *wavepacket* type, that looks like a plane wave, but supported only in a neighborhood of the origin.

A nice choice to make in the p-adic setting is to take the boundary function to be

$$\phi_0(x) = e^{2\pi i\{kx\}} \cdot \Theta(x, \mathbb{Z}_p), \tag{4.66}$$

where $\Theta(x, S)$ is the characteristic function of the set $S \subset \mathbb{Q}_p$. The transformation (4.17) is actually trivial inside $\mathbb{Z}_p$, so no distortion of the wavepacket occurs at all (unlike for a similar setup in $\mathbb{C}$). Of course, we ought to take $|k|_p > 1$, so that $\{kx\}$ is not constant over the whole of $\mathbb{Z}_p$.

Given this choice of boundary function, the corresponding solution to the bulk equations of motion can be reconstructed using the integral kernel (4.54):

$$\phi(v) = \frac{p}{p+1} \int_{\mathbb{Z}_p} d\mu(x)\, e^{2\pi i\{kx\}} p^{-d_C(x,v)}. \tag{4.67}$$

Recall that $d_C(x, v)$ is the distance from v to x, regularized to be zero at the centerpoint $v = C$ of the tree. We will calculate this integral when v is inside the branch of the tree above $\mathbb{Z}_p$.

Proposition 1. Let v be a vertex in the branch above $\mathbb{Z}_p$, at a depth ℓ (i.e., since $v \in \mathbb{Z}_p$, distance from the centerpoint) such that $0 \leq \ell < -\operatorname{ord}_p(k) - 2$. Then the reconstructed bulk function $\phi(v)$ is zero.

Proof. The claim relies on the simple fact that the sum of all p-th roots of unity is zero. Since v is above the red line in Fig. (4.8) (at depth equal to $-\operatorname{ord}_p(k) - 1$), both terms in the integrand are locally constant below the line, and the integral may be evaluated as a sum along the vertices at the height of the red line. Furthermore, the factor $p^{-d_C(v,x)}$ is constant for each of the p vertices on the line that descend from the same ancestor. Since the

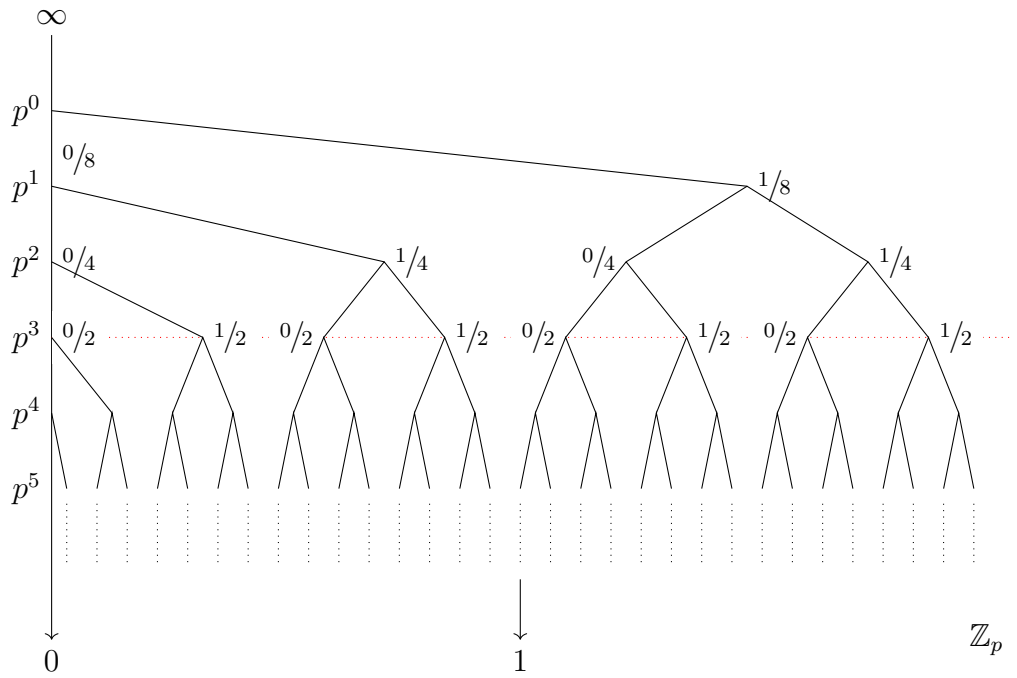


Figure 4.8: A drawing of $\mathbb{Z}_p$ ($p = 2$ for simplicity). Here $k = p^{-4}$. The marked fractions at vertices indicate contributions to $\{kx\}$, which are summed along the geodesic ending at x.

measure of each branch is equal, the integral is proportional to the sum of all p-th roots of unity, and hence to zero. Notice that this also demonstrates that the reconstructed bulk function is zero everywhere *outside* $\mathbb{Z}_p$: it is zero at the central vertex, and zero on the boundary of the open ball complementary to $\mathbb{Z}_p$. □

Even without calculating the explicit form of the bulk function for vertices below the screening height, this simple argument already allows us to make our physical point: in p-adic holography, the qualitative features of ordinary AdS/CFT persist in a setting where the bulk geometry is discrete, and in some cases are even sharpened. For instance, we have shown explicitly that modes for which $|k|_p$ is large (i.e. the short-wavelength behavior of the boundary conditions) must drop out of the reconstructed bulk field, making exactly zero contribution to it above a height in the tree precisely determined by $|k|_p$. The usual intuition that moving into the bulk along the holographic direction corresponds to integrating out UV modes is thus neatly confirmed.

The explicit form of the reconstructed bulk function at vertices below the screening height is easy to calculate, but less central to our discussion; we

leave the computation as an exercise for the reader.

The possibility of higher-spin fields

We now wish to propose an analogue of higher-spin fields that could be defined in the p-adic case. While we will motivate our proposal here, we do not investigate any properties of p-adic CFT with fields other than scalars. We will return to this question in future work.

We proceed by analogy with two-dimensional CFT, in which the conformal dimension and spin together describe a character of the multiplicative group $\mathbb{C}^\times$:

$$\phi(re^{i\theta} \cdot z) = r^\Delta e^{is\theta}\phi(z). \tag{4.68}$$

The group $\mathbb{C}^\times \simeq \mathbb{R}^\times_{>0} \times \mathrm{U}(1)$; the conformal dimension determines a character of the first factor, and the spin a character of the second, which can be thought of as scale transformations and rotations of the coordinate respectively. The existence of the logarithm function means that we can think of the multiplicative group $\mathbb{R}_+$ as isomorphic to the additive group $\mathbb{R}$.

The structure of the group of units of any local field is understood (see [164] for details). In particular, for the field $\mathbb{Q}_p$, the result is that

$$\mathbb{Q}_p^\times \simeq p^\mathbb{Z} \times \mathbb{F}_p^\times \times U^{(1)}, \tag{4.69}$$

where $U^{(1)}$ is the group of "principal units" of the form $1 + p \cdot a$, with $a \in \mathbb{Z}_p$. This decomposition just reflects the structure of the p-adic decimal expansion: since the p-adic norm is multiplicative, any number $x \neq 0$ can be written in the form

$$x = p^{\mathrm{ord}_p(x)} \left(x_0 + x_1 p + \cdots\right), \tag{4.70}$$

where $x_0 \neq 0$ (so that $x \in \mathbb{F}_p^\times \simeq C_{p-1}$) and the other x_i may be any digits chosen from $\mathbb{F}_p$. Dividing through by x_0, one gets

$$x = p^{\mathrm{ord}_p(x)} \cdot x_0 \left(1 + \sum_{i \geq 1} \tilde{x}_i p^i\right), \tag{4.71}$$

where $\tilde{x}_i = x_i/x_0$, and the factor in parentheses is a principal unit. A character of $\mathbb{Q}_p^\times$ is therefore a triple of characters, one for each factor in (4.69). The first factor, as in the normal case, corresponds to the scaling dimension of the field; the last two factors are therefore analogous to the spin. Obviously, the second

factor corresponds to a $\mathbb{Z}/(p-1)\mathbb{Z}$ phase. It is also known [164] that the set of characters of $U^{(1)}$ is countable and discrete.

In fact, we can naively understand a broader class of the characters of $\mathbb{Z}_p^\times = \mathbb{F}_p^\times \times U^{(1)}$. Recall the description of $\mathbb{Z}_p$ as the inverse limit of its finite truncations:

$$\mathbb{Z}_p = \varprojlim \mathbb{Z}/p^n\mathbb{Z}. \tag{4.72}$$

Since this is an inverse limit of rings, there are therefore projection maps between the respective multiplicative groups:

$$\mathbb{Z}_p^\times \to (\mathbb{Z}/p^n\mathbb{Z})^\times \simeq C_{p^{n-1}(p-1)}. \tag{4.73}$$

Therefore, any multiplicative character of a cyclic group $C_{p^{n-1}(p-1)}$ (i.e., any finite root of unity of order $p^{n-1}(p-1)$, for arbitrary n) will give a character of $(\mathbb{Z}/p^n\mathbb{Z})^\times$, which will in turn pull back to define a multiplicative character of the spin part of $\mathbb{Q}_p^\times$. Spin in the p-adic case is therefore both similar to and interestingly different from ordinary two-dimensional CFT.

4.5 Entanglement entropy

The entanglement entropy in quantum field theories is a notoriously difficult and subtle quantity to compute, and much effort has been expended in developing a toolbox of techniques that provide exact results. One of the first systems in which the computation became tractable was two-dimensional conformal field theory, and in particular the theory of free bosons. Since we are primarily considering the free boson in our discussion, one might hope that the same techniques can be applied in the p-adic case. While we believe that this is the case, and plan to give a full calculation of the entanglement entropy in future work, there are subtleties that arise in each technique and prevent it from being used naively. We will demonstrate these techniques, illustrate the subtle issues that arise, and justify our conjecture for the entanglement entropy in what follows.

As in the real case, we expect the entanglement entropy to have UV divergences. These are normally thought of as localized to the "boundary" of the region under consideration. Care must be used in defining what we mean by interval and boundary; the p-adic numbers have no ordering, and every element of an open set is equally (or equally not) a boundary element.

Whenever possible, we must think in terms of open sets. Over the reals, the open sets are intervals with measure or length given by the norm of the separation distance of the endpoints; as the reader will recall, p-adic open sets are perhaps best visualized using the Bruhat–Tits tree. Once a center C of the tree is picked, we can pick any other vertex v and consider the cone of points below v extending out towards the boundary, which is an open neighborhood in $\mathrm{P}^1(\mathbb{Q}_p)$. A perhaps surprising fact which follows from the definition of the p-adic norm $|x-y|_p$ $(x, y \in \mathrm{P}^1(\mathbb{Q}_p))$ is that it is related to the height of the cone required to connect x to y (see Fig. 4.9).

Following standard arguments, say of [43, 189], we can pick the boundary cone below the point v to be called the region V. The total Hilbert space on $\mathbb{Q}_p$ splits into Hilbert spaces on V and its complement, $\mathcal{H} = \mathcal{H}_V \otimes \mathcal{H}_{-V}$. The entanglement entropy defined by $S(V) = -\operatorname{Tr}(\rho_V \log \rho_V)$ and by construction satisfies $S(V) = S(-V)$. As there are an infinite number of points $x_i \in V \in \mathrm{P}^{1\,1}(\mathbb{Q}_p)$, there is an unbounded number of local degrees of freedom $\phi(x_i)$ as is typical of quantum field theory. In the continuum case this implies logarithmic divergences from modes in V entangled with those in $-V$, and we expect the same to be true in the p-adic case.

In the works of Cardy and Calabrese [39, 40], the entanglement entropy for intervals in $1+1$-dimensional conformal field theories are explicitly calculated. The p-adic field theories considered here are exactly analogous to the two-dimensional free boson; in both, the scalar $\phi(x)$ has conformal dimension zero and (as we have shown above) a logarithmically divergent propagator. We wish to understand how much of their calculation can be duplicated in the p-adic case. These authors generally follow a series of steps beginning with the replica trick, which is the observation that n powers of the reduced density matrix ρ_V can be computed by evaluating the partition function on a Riemann surface obtained by gluing n copies of the theory together along the interval V. The entanglement entropy follows from analytic continuation of these results in n, followed by the limit $n \to 1$, according to the formula

$$\operatorname{Tr}(\rho_V^n) = \frac{Z_n(V)}{Z_1^n}, \quad S_V = -\lim_{n\to 1} \frac{\partial}{\partial n} \frac{Z_n(V)}{Z_1^n}, \tag{4.74}$$

where $Z_n(V)$ is the n-sheeted partition function and Z_1 is the partition function of 1 sheet with no gluing, which is required for normalization.

In $1+1$ dimensions, the n-sheeted partition function can be viewed as a Riemann surface, and the holomorphic properties of this surface make the calculation tractable. In particular, if the interval has the boundary points x and y, the complicated world sheet topology can be mapped to the target space by defining multi-valued *twist fields* $\Phi_n(x), \Phi_n(y)$ on the plane whose boundary conditions implement the n sheeted surface. One can find that $\mathrm{Tr}(\rho_V^n)$ behaves exactly like the n^{th} power of a two point function of the twist fields, once their conformal dimension has been determined using Ward identities:

$$\mathrm{Tr}(\rho_V^n) \sim \langle 0|\Phi_n(x)\Phi_n(y)|0\rangle^n \sim \left(\frac{x-y}{\epsilon}\right)^{-\frac{c}{6}(n-\frac{1}{n})}, \tag{4.75}$$

where c is the central charge and ϵ is a normalization constant from Z_1. When $n=1$ exactly, the twist fields have scaling dimension 0 and the above correlator no longer makes sense. Instead, taking the limit as $n \to 1$, the linear term is $-n\frac{c}{3}\ln\left(\frac{x-y}{\epsilon}\right)$. Taking the derivative gives the famous universal formula for the entanglement entropy [120].

The difficulty in performing the same calculation over the p-adics consists in fixing the dimensions of the twist operators. These operators can be defined just as in the normal case; after all, all that they do is implement certain boundary conditions at branch points on the fields in a theory of n free bosons. However, the usual arguments that fix their dimension rely on the existence of a uniformizing transformation $z \mapsto z^n$ that describes the relevant n-sheeted branched cover of P^1 by P^1; the Schwarzian of this holomorphic (but not Möbius) transformation then appears as the conformal dimension. The argument using the OPE with the stress tensor is identical in content. Both cases rely on the existence of holomorphic (but not fractional-linear) transformations, and a measure—the Schwarzian or conformal anomaly—of their "failure" to be Möbius.

In the p-adic case, this is related to the question of local conformal transformations; it has been suggested [153] that no such symmetries exist. Moreover, since $\mathbb{Q}_p$ is not algebraically closed, a transformation like $z \mapsto z^n$ need not even be onto. Nevertheless, we can still define the twist operators, and we suppose that they transform as primaries with some conformal dimensions Δ_n. Their two-point function then gives the density matrix. This function is

$$\langle 0|\Phi_n^{(p)}(x)\Phi_n^{(p)}(y)|0\rangle^n \sim \left|\frac{x-y}{\epsilon}\right|_p^{-2n\Delta_n}, \tag{4.76}$$

where Δ_n are the model-dependent (and unknown) conformal dimensions. Inserting this ansatz into $-\lim_{n\to 1}\frac{\partial}{\partial n}\operatorname{Tr}(\rho_V^n)$ and taking the limit $n \to 1$, $\Delta_n \to 0$ gives:

$$\left(2n\frac{\partial\Delta_n}{\partial n}\Big|_{n=0}\right)\ln\left|\frac{x-y}{\epsilon}\right|_p. \tag{4.77}$$

While this is not a proof, it provides some evidence for the expected logarithmic scaling of the entropy. We expect that the dimensions $\Delta_n \to 0$ as $n \to 1$, since of course the twist operator on one sheet is just the identity. If we could fix the conformal dimension without using the conformal anomaly, this calculation would fix not only the logarithmic form of the answer, but also the coefficient that plays the role of the central charge. It may be possible to do this by examining the path integral with twist-operator insertions directly.

A possible way around this difficulty might be to consider a harder problem first: to think about two intervals rather than one. The genus of the Riemann surface that appears in the replica trick is $g = (n-1)(N-1)$; thus, for one interval, we are considering a branched cover of P^1 over itself, and the conformal anomaly is a necessary ingredient. However, one might hope that for two intervals, we can simply compute the partition function on a series of higher-genus Riemann surfaces (which is understood in the p-adic case), and take the limit as the genus approaches zero. Discussions of the entanglement entropy in terms of Schottky uniformization—which therefore appear tailored to our needs—have appeared in the literature [76].

Two difficulties appear in this case: the first is matching the moduli of the Riemann surface in question to the lengths of the intervals; the second is more subtle, and reflects the fact that, over the p-adics, not every branched cover of the p-adic projective line is a Mumford curve [32]. We believe that one of the strategies outlined here will succeed in producing a rigorous computation of the entanglement entropy, but we must relegate that computation to future work.

Ryu-Takayanagi formula

Let us take as given the conjecture from the previous section that the entanglement entropy of a region in the boundary CFT should be computed as the logarithm of its p-adic size. We take our interval to be the smallest p-adic open ball which contains points x and y. This interval has size $|x-y|_p$. To understand the Ryu-Takayanagi formula, it remains to compute the length of

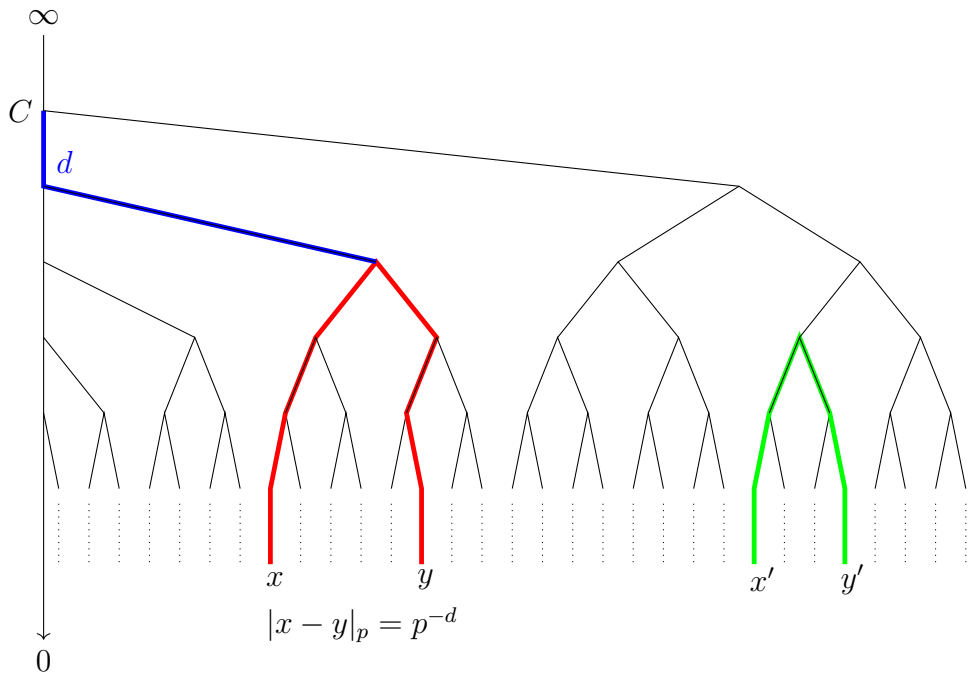


Figure 4.9: Boundary-anchored geodesics in T_p have a natural interpretation in terms of the p-adic norm. Once the arbitrary position of the center C is fixed, the norm of open sets in $\mathbb{Q}_p$ is given by p^{-d}, where d is the integer number of steps from C required before the path to the endpoints splits. In this example, $|x-y|_p$ is described by the red geodesic and the value is p^{-2}. The set corresponding to the green geodesic has a smaller norm by a factor of p because the vertex is 1 step further down the tree. As in the case of real AdS, the length of the geodesic is formally infinite because an infinite number of steps is required to reach the boundary. One may introduce a cutoff corresponding to truncation of the tree at a fixed distance, then take the limit as this cutoff goes to zero. It should be apparent that the (formally infinite) red geodesic is longer than the green one by two steps. Up to constant factors, the length of any boundary-anchored geodesic is an infinite term minus d. This explains the logarithmically divergent scaling of geodesic length with p-adic norm.

the unique geodesic connecting x to y. The tree geometry for this setup is depicted in Fig. 4.9. Since there are an infinite number of steps required to reach the boundary, the geodesic length is formally infinite, just as in the real case. We regulate this by cutting off the tree at some finite tree distance a from the center C, which can be thought of as $\mathrm{ord}_p(\epsilon)$ for some small p-adic number ϵ. We will then take this minimum number $\epsilon \to 0$ (p-adically). This limit will push the cutoff in the tree to infinite distance from C.

An $\mathrm{SL}(2,\mathbb{Q}_p)$ transformation can always be used to move the points x and y

to the $\mathbb{Z}_p$ part of the tree first to simplify the argument. Then introducing the distance cutoff a effectively truncates the decimal expansions of x and y to the first a decimal places. In the case where $|x-y|_p = 1$, the geodesic connecting the two points passes through C and has length $2a$. If $|x-y|_p < 1$, the geodesic is shorter by a factor of $2d$, where $|x-y|_p = p^{-d}$. Roughly speaking, as can be seen in Fig. 4.9, smaller boundary regions are subtended by shorter geodesics in the tree.

We see that the cutoff-dependent distance is

$$d(x,y)_a = 2a + \frac{2}{\ln p} \ln |x-y|_p. \tag{4.78}$$

We would like to take $a \to \infty$. Up to the factor of $\ln p$, we can define a to be the logarithm of a p-adic cutoff ϵ such that $a \to \infty$ as $\epsilon \to 0$. Using this definition, we find the length of a boundary-anchored geodesic to be

$$d(x,y) = \lim_{\epsilon \to 0} \frac{2}{\ln p} \ln \left| \frac{x-y}{\epsilon} \right|_p . \tag{4.79}$$

Up to the overall factor in front (which presumably depended on our choice of the length of each leg of the tree), we see the geodesic length is logarithmically divergent in interval size.

An adelic formula for entanglement?

We have argued that the general form of entanglement entropy scaling for the boundary theory is dual to a geodesic length in the bulk. At the present time we lack a p-adic notion of central charge c or theory dependent quantity which counts boundary degrees of freedom. Nevertheless, we claim the general form is

$$S_p(x-y) = c_p \ln \left| \frac{x-y}{\epsilon_p} \right|_p . \tag{4.80}$$

We now wish to speculate about the possibility of an *adelic* formula for the entanglement entropies.

In the study of p-adic numbers, there exists a surprising formula which relates the various p-adic valuations of a rational number to its real norm. This is a different form of the fundamental theorem of arithmetic, and is sometimes known as an adelic formula:

$$\prod_p^{\infty} |x|_p = 1. \tag{4.81}$$

Here $x \in \mathbb{Q}$ and the product is taken over all primes. The "prime at ∞" corresponds to the usual archimedean norm $|x|_\infty = |x|$. This equality follows by considering the unique prime factorization of x into a product of prime powers. When x contains a factor p^n, then $|x|_p = p^{-n}$. This means the infinite product over primes is well defined because only finitely many terms are not equal to 1. In fact, the product over the finite primes gives exactly the inverse of the real norm $|x|$. Therefore the product over all finite places and the infinite place is unity.

We now wish to again recall the familiar formula for the universal entanglement entropy of an interval in a 1+1 dimensional conformal field theory; written suggestively in the "prime at infinity" notation:

$$S_\infty(x-y) = \frac{c}{3} \ln \left| \frac{x-y}{\epsilon_\infty} \right|_\infty . \tag{4.82}$$

One might hope through a better understanding of p-adic conformal field theory or the holographic dual, the value of the proportionality constant or the p-adic central charge might be determined. In the (perhaps unlikely) event that the central charges of the p-adic theory agree with the real case, then using the adelic formula for the interval, we propose:

$$\begin{aligned} &\text{If } c_p = c/3 \text{ for all } p \text{ and } x-y \in \mathbb{Q}, \text{ then} \\ &\qquad \sum_p^\infty S_p(x-y) \propto \ln \left(\prod_p^\infty \left| \frac{x-y}{\epsilon_p} \right|_p \right) = 0. \end{aligned} \tag{4.83}$$

One might be suspicious about this formula; each of the entropies S_p are formally divergent. Additionally, since these quantities are entropies they are expected to be positive. Therefore care must be taken in interpreting the above.

One possible resolution is the erroneous application of the adelic formula to the cutoffs ϵ_p. In computing S_p holographically, we assumed $|\epsilon|_p \to 0$. However, if $|\epsilon_p|_p \to 0$ in one norm, it is not generally true that $|\epsilon_p|_{p'} \to 0$ for another choice p'. Therefore, we require a numerically different cutoff parameter ϵ_p for each system over p. As all these parameters are taken to 0 in their respective norms, the corresponding entanglement entropies diverge.

Understanding that the cut offs ϵ_p do not cancel on the left or right hand sides, we are left with the divergent pieces of the entropy being equal on both sides.

However, if we vary the length of $|x-y|_\infty$ in the real physical system, we see that the entropy difference associated with this interval is distributed across the S_p's such that the sum is zero. Put another way, varying the real interval length will cause some values of $|x-y|_p$ for different p to increase and others to decrease. This causes some S_p to increase and others to decrease such that the total change of entropy over all finite and infinite places is 0.

We will leave it to future work to try to derive or understand this relation further.

4.6 p-adic bulk geometry: Schottky uniformization and non-archimedean black holes

Holography for Euclidean higher-genus black holes

The first explicit form of AdS/CFT correspondence for the asymptotically AdS_3 higher-genus black holes, in the Euclidean signature, was obtained in [150], where the computation of the Arakelov Green function of [148] is shown to be a form of the holographic correspondence for these black holes, where the two-point correlation function for a field theory on the conformal boundary X_Γ is written in terms of gravity in the bulk $\mathcal{H}_\Gamma$, as a combination of lengths of geodesics.

At the heart of Manin's holographic formula lies a simple identity relating conformal geometry on $\mathrm{P}^1(\mathbb{C})$ and hyperbolic geometry on $\mathbf{H}^3$, namely the fact that the cross ratio of four points on the boundary $\mathrm{P}^1(\mathbb{C})$ can be written as the length of an arc of geodesic in the bulk $\mathbf{H}^3$. More precisely, consider the two point correlation function $g(A,B)$ on $\mathrm{P}^1(\mathbb{C})$. This is defined by considering, for a divisor $A=\sum_x m_x\, x$, the Green function of the Laplacian

$$\partial\bar\partial g_A = \pi i(\deg(A)d\mu - \delta_A),$$

with δ_A the delta current associated to the divisor, $\delta_A(\varphi)=\sum_x m_x\varphi(x)$, and $d\mu$ a positive real-analytic 2-form. The Green function g_A has the property that $g_A - m_x \log|z|$ is real analytic for z a local coordinate near x, and is normalized by $\int g_A d\mu = 0$. For two divisors A, B, with A as above and $B=\sum_y n_y\, y$ the two point function is given by $g(A,B)=\sum_y n_y g_A(y)$. For degree zero divisors it is independent of the form $d\mu$ and is a conformal invariant. If w_A is a meromorphic function on $\mathrm{P}^1(\mathbb{C})$ with $Div(w_A)=A$, and C_B is a 1-chain with

boundary B, the two point function satisfies

$$g(A,B) = Re \int_{C_B} \frac{dw_A}{w_A}.$$

For (a,b,c,d) a quadruple of points in $\mathrm{P}^1(\mathbb{C})$, the cross ratio $\langle a,b,c,d\rangle$ satisfies

$$\langle a,b,c,d\rangle = \frac{w_{(a)-(b)}(c)}{w_{(a)-(b)}(d)},$$

where $(a)-(b)$ is the degree zero divisor on $\mathrm{P}^1(\mathbb{C})$ determined by the points a,b, and the two point function is

$$g((a)-(b),(c)-(d)) = \log \frac{|w_{(a)-(b)}(c)|}{|w_{(a)-(b)}(d)|}.$$

Given two points a,b in $\mathrm{P}^1(\mathbb{C})$, let $\ell_{\{a,b\}}$ denote the unique geodesics in $\mathbf{H}^3$ with endpoints a and b. Also given a geodesic ℓ in $\mathbf{H}^3$ and a point $c \in \mathrm{P}^1(\mathbb{C})$ we write $c * \ell$ for the point of intersection between ℓ and the unique geodesic with an endpoint at c and intersecting ℓ orthogonally. We also write $\lambda(x,y)$ for the oriented distance of the geodesic arc in $\mathbf{H}^3$ connecting two given points x,y on an oriented geodesic. Then the basic holographic formula identifies the two point function with the geodesic length

$$g((a)-(b),(c)-(d)) = -\lambda(a * \ell_{\{c,d\}}, b * \ell_{\{c,d\}}).$$

One can also express the argument of the cross ratio in terms of angles between bulk geodesics (see [148], [150]). This basic formula relating the two point correlation function on the boundary to the geodesic lengths in the bulk is adapted to the higher-genus cases by a suitable procedure of averaging over the action of the group that provides an explicit construction of a basis of meromorphic differentials on the Riemann surface X_Γ in terms of cross ratios on $\mathrm{P}^1(\mathbb{C})$. A basis of holomorphic differentials on X_Γ, with

$$\int_{A_k} \omega_{\gamma_j} = 2\pi i \delta_{jk}, \qquad \int_{B_k} \omega_{\gamma_j} = \tau_{jk}$$

the period matrix, is given by

$$\omega_{\gamma_i} = \sum_{h \in S(\gamma_i)} d_z \log\langle z_h^+, z_h^-, z, z_0\rangle,$$

for $z, z_0 \in \Omega_\Gamma$, with $S(\gamma)$ the conjugacy class of γ in Γ. The series converges absolutely when $\dim_H(\Lambda_\Gamma) < 1$. Meromorphic differentials associated to a divisor $A = (a)-(b)$ are similarly obtained as averages over the group action

$$\nu_{(a)-(b)} = \sum_{\gamma \in \Gamma} d_z \log\langle a, b, \gamma z, \gamma z_0\rangle$$

and the Green function is computed as a combination $\nu_{(a)-(b)} - \sum_j X_j(a,b)\omega_{\gamma_j}$ with the coefficients $X_j(a,b)$ so that the B_k-periods vanish. Since in the resulting formula each crossed ratio term is expressible in terms of the length of an arc of geodesic in the bulk, the entire Green function is expressible in terms of gravity in the bulk space. We refer the reader to §§2.3, 2.4, and 2.5 of [150] and to [151] for a more detailed discussion and the resulting explicit formula of the holographic correspondence for arbitrary genus.

Holography on p-adic higher-genus black holes

In the special case of a genus-one curve, the relevant Schottky group is isomorphic to $q^{\mathbb{Z}}$, for some $q \in \mathbf{k}^*$, and the limit set consists of two points, which we can identify with 0 and ∞ in $\mathrm{P}^1(\mathbf{k})$. The generator of the group acts on the geodesic in $T_{\mathbf{k}}$ with endpoints 0 and ∞ as a translation by some length $n = \log|q| = v_{\mathfrak{m}}(q)$, the valuation. The finite graph $G_{\mathbf{k}}$ is then a polygon with n edges, and the graph $T_{\mathbf{k}}/\Gamma$ consists of this polygon with infinite trees attached to the vertices. The boundary at infinity of $T_{\mathbf{k}}/\Gamma$ is a Mumford curve $X_\Gamma(\mathbf{k})$ of genus one with its p-adic Tate uniformization. The graph $T_{\mathbf{k}}/\Gamma$ is the p-adic BTZ black hole, with the central polygon $G_{\mathbf{k}}$ as the event horizon. The area of the black hole and its entropy are computed by the length of the polygon (see Fig. 4.10).

The higher-genus cases are p-adic versions of the higher-genus black holes discussed above, with the finite graph $G_{\mathbf{k}}$ as event horizon, and its geodesic length proportional to the black hole entropy.

Given a set of generators $\{\gamma_1, \dots, \gamma_g\}$ of a p-adic Schottky group, let n_{γ_i} be the translation lengths that describe the action of each generator γ_i on its axis ℓ_{γ_i}. More precisely, if an element γ is conjugate in $\mathrm{PGL}(2,\mathbf{k})$ to an element of the form

$$\begin{pmatrix} q & 0 \\ 0 & 1 \end{pmatrix},$$

then the translation length is $n_\gamma = v_{\mathfrak{m}}(q) = \mathrm{ord}_{\mathbf{k}}(q)$, the order (valuation) of q. The translation lengths $\{n_{\gamma_i}\}$ are the Schottky invariants of the p-adic Schottky group Γ. It is shown in [42] that the Schottky invariants can be computed as a spectral flow.

The Drinfeld–Manin holographic formula of [149] for p-adic black holes of arbitrary genus is completely analogous to its archimedean counterpart of [148].

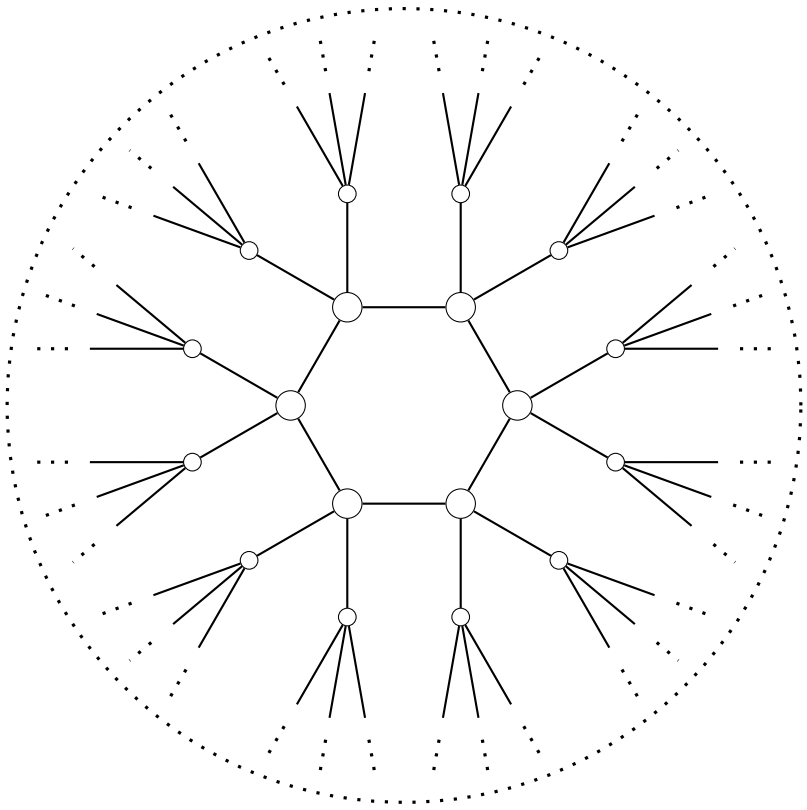


Figure 4.10: The p-adic BTZ black hole. (As pictured, $p = 3$).

There is a good notion of $\mathbf{k}$-divisor on $\mathrm{P}^1(\mathbf{k})$, as a function $\mathrm{P}^1(\bar{\mathbf{k}}) \to \mathbb{Z}$, with $z \mapsto m_z$, with the properties that $m_{z_1} = m_{z_2}$ if z_1 and z_2 are conjugate over $\mathbf{k}$; that all points z with $m_z \neq 0$ lie in the set of points of P^1 over a finite extension of $\mathbf{k}$; and that the set of points with $m_z \neq 0$ has no accumulation point. As before we write such a divisor as $A = \sum_z m_z\, z$. Given a Γ-invariant divisor A of degree zero, there exists a meromorphic function on $\Omega_\Gamma(\mathbf{k})$ with divisor A. It is given by a Weierstrass product

$$W_{A,z_0} = \prod_{\gamma\in\Gamma} \frac{w_A(\gamma z)}{w_A(\gamma z_0)},$$

where $w_A(z)$ is a $\mathbf{k}$-rational function on $\mathrm{P}^1(\mathbf{k})$ with divisor A. The convergence of this product is discussed in Proposition 1 of [149]: the non-archimedean nature of the field $\mathbf{k}$ implies that the product converges for all $z \in \Omega_\Gamma \setminus \cup_\gamma \gamma(\mathrm{supp}(A))$. The function W_{A,z_0} is a p-adic automorphic function (see [145]) with $W_{A,z_0}(\gamma z) = \mu_A(\gamma) W_{A,z_0}(z)$, with $\mu_A(\gamma) \in \mathbf{k}^*$, multiplicative in A and γ. One obtains a basis of Γ-invariant holomorphic differentials on $X_\Gamma(\mathbf{k})$ by taking

$$\omega_{\gamma_i} = d \log W_{(\gamma_i - 1)z_0, z_1},$$

where

$$W_{(\gamma-1)z_0,z_1}(z) = \prod_{h\in C(\gamma)} \frac{w_{hz_\gamma^+ - hz_\gamma^-}(z)}{w_{hz_\gamma^+ - hz_\gamma^-}(z_0)},$$

for $C(g)$ a set of representatives of $\Gamma/\gamma^{\mathbb{Z}}$.

It is shown in [149] that the order of the cross ratio on $\mathrm{P}^1(\mathbf{k})$ is given by

$$\mathrm{ord}_{\mathbf{k}} \frac{w_A(z_1)}{w_A(z_2)} = \#\{\ell_{z_1,z_2}, \ell_{a_1,a_2}\},$$

for $A = a_1 - a_2$ and $\ell_{x,y}$ the geodesic in the Bruhat–Tits tree with endpoints $x, y \in \mathrm{P}^1(\mathbf{k})$, with $\#\{\ell_{z_1,z_2}, \ell_{a_1,a_2}\}$ the number of edges in common to the two geodesics in $T_{\mathbf{k}}$. This is the basic p-adic holographic formula relating boundary two point function to gravity in the bulk.

A difference with respect to the Archimedean case is that, over $\mathbb{C}$, both the absolute value and the argument of the cross ratio have an interpretation in terms of geodesics, with the absolute value expressed in terms of lengths of geodesic arcs and the argument in terms of angles between geodesics, as recalled above. In the p-adic case, however, it is only the valuation of the two point correlation function that has an interpretation in terms of geodesic lengths in the Bruhat–Tits tree. The reason behind this discrepancy between the archimedean and non-archimedean cases lies in the fact that the Bruhat–Tits tree $T_{\mathbf{k}}$ is the correct analog of the hyperbolic handlebody $\mathbf{H}^3$ only for what concerns the part of the holographic correspondence that involves the absolute value (respectively, the p-adic valuation) of the boundary two point function. There is a more refined p-adic space, which maps surjectively to the Bruhat–Tits tree, which captures the complete structure of the p-adic automorphic forms for the action of a p-adic Schottky group Γ: Drinfeld's p-adic upper half plane, see Chapter I of [29]. Given $\mathbf{k}$ as above, let $\mathbb{C}_p$ denote the completion of the algebraic closure of $\mathbf{k}$. Drinfeld's p-adic upper half plane is defined as $\mathbf{H}_{\mathbf{k}} = \mathrm{P}^1(\mathbb{C}_p)\setminus \mathrm{P}^1(\mathbf{k})$. One can view this as an analog of the upper and lower half planes in the complex case, with $\mathbf{H}^+ \cup \mathbf{H}^- = \mathrm{P}^1(\mathbb{C}) \setminus \mathrm{P}^1(\mathbb{R})$. There is a surjection $\lambda : \mathbf{H}_{\mathbf{k}} \to T_{\mathbf{k}}$, defined in terms of the valuation, from Drinfeld's p-adic upper half plane $\mathbf{H}_{\mathbf{k}}$ to the Bruhat–Tits tree $T_{\mathbf{k}}$. For vertices $v, w \in V(T_{\mathbf{k}})$ connected by an edge $e \in E(T_{\mathbf{k}})$, the preimages $\lambda^{-1}(v)$ and $\lambda^{-1}(w)$ are open subsets of $\lambda^{-1}(e)$, as illustrated in Fig. 4.11. The map $\lambda : \mathbf{H}_{\mathbf{k}} \to T_{\mathbf{k}}$ is equivariant with respect to the natural actions of $\mathrm{PGL}(2,\mathbf{k})$ on $\mathbf{H}_{\mathbf{k}}$ and on $T_{\mathbf{k}}$. In particular, given a p-adic Schottky group $\Gamma \subset \mathrm{PGL}(2,\mathbf{k})$, we can

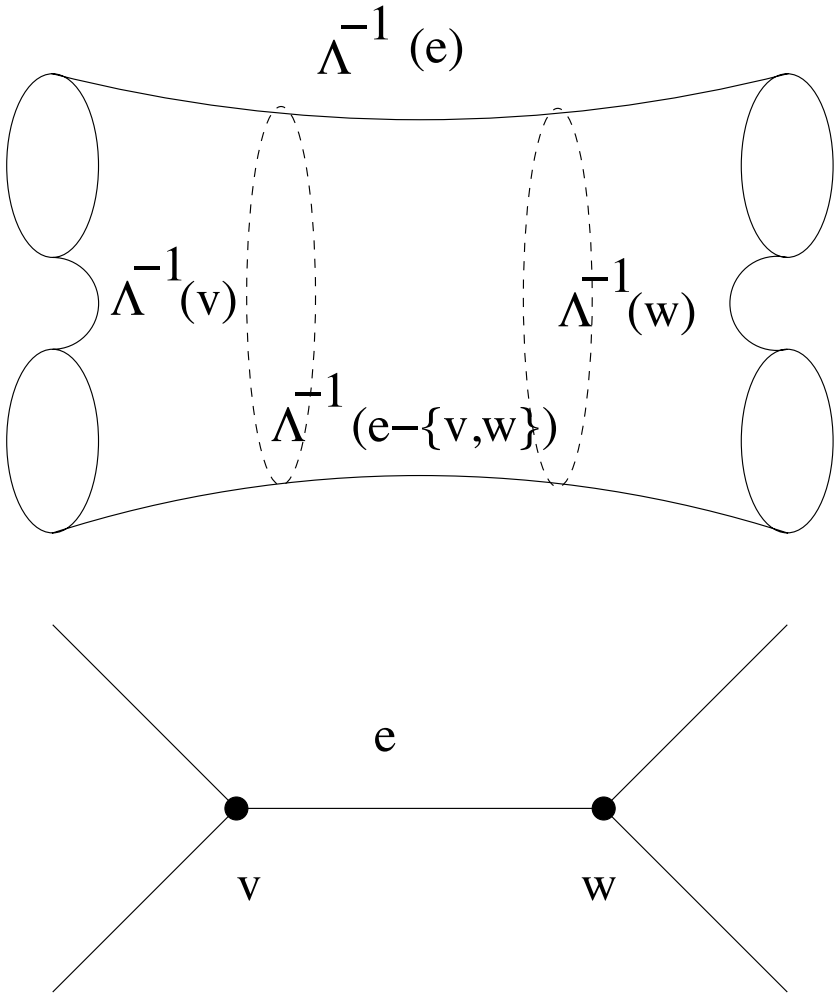


Figure 4.11: Drinfeld's p-adic upper half plane and the Bruhat–Tits tree.

consider the quotients $\tilde{\mathcal{H}}_\Gamma = \mathbf{H}_\mathbf{k}/\Gamma$ and $\mathcal{H}_\Gamma = T_\mathbf{k}/\Gamma$ and the induced projection $\lambda : \tilde{\mathcal{H}}_\Gamma \to \mathcal{H}_\Gamma$. Both quotients have conformal boundary at infinity given by the Mumford curve $X_\Gamma = \Omega_\Gamma(\mathbf{k})/\Gamma$, with $\Omega_\Gamma(\mathbf{k}) = \mathrm{P}^1(\mathbf{k}) \smallsetminus \Lambda_\Gamma$, the domain of discontinuity of the action of Γ on $\mathrm{P}^1(\mathbf{k}) = \partial\mathbf{H}_\mathbf{k} = \partial T_\mathbf{k}$. One can view the relation between $\mathbf{H}_\mathbf{k}$ and $T_\mathbf{k}$ illustrated in Fig. 4.11, and the corresponding relation between $\tilde{\mathcal{H}}_\Gamma$ and $\mathcal{H}_\Gamma$, by thinking of $\tilde{\mathcal{H}}_\Gamma$ as a "thickening" of the graph $\mathcal{H}_\Gamma$, just as in the Euclidean case one can view the union of the fundamental domains of the action of Γ on $\mathbf{H}^3$, as illustrated in Fig. 4.4, as a thickening of the Cayley graph (tree) of the Schottky group Γ, embedded in $\mathbf{H}^3$.

Thus, when considering the non-archimedean holographic correspondence and p-adic black holes of arbitrary genus, one can choose to work with either $\mathcal{H}_\Gamma$

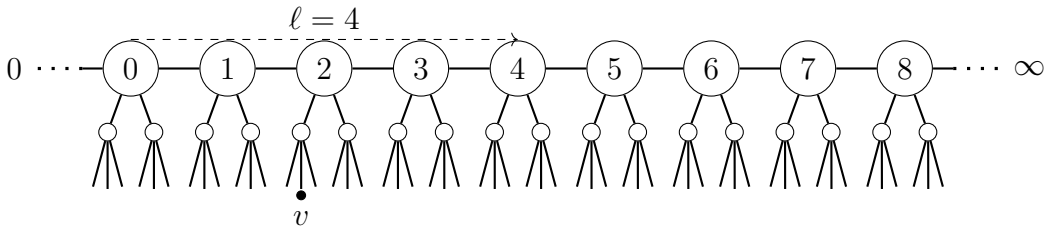


Figure 4.12: The action of a rank-one Schottky group (translation by ℓ along a fixed geodesic) on the Bruhat–Tits tree. As pictured, $n = h = 2$.

or with $\tilde{\mathcal{H}}_\Gamma$ as the bulk space, the first based on Bruhat–Tits trees and the second (more refined) based on Drinfeld's p-adic upper half spaces. In this paper we will be focusing on those aspects of the non-archimedean AdS/CFT correspondence that are captured by the Bruhat–Tits tree, while we will consider a more refined form of non-archimedean holography, based on Drinfeld's p-adic upper half planes, in forthcoming work.

Scalars on higher-genus backgrounds: sample calculation

In light of this discussion of higher-genus holography in the p-adic case, it is easy to understand how to generalize the arguments and calculations we discussed for scalar fields in §4.4 to the BTZ black hole, or to higher-genus hyperbolic handlebodies, the p-adic analogues of Krasnov's Euclidean black holes. One can simply think of the higher-genus geometry as arising from the quotient of the tree T_p (and its boundary $\mathrm{P}^1(\mathbb{Q}_p)$) by the action of a rank-g Schottky group. Any quantity that can then be made equivariant under the action of the Schottky group will then descend naturally to the higher-genus setting.

As a simple example, it is easy to construct the genus-1 analogue of our basic Green's function (4.49), using the method of images. We perform this calculation in the following paragraphs. The result makes it easy to perform the reconstruction of bulk solutions to the equations of motion in a BTZ background, with specified boundary conditions at infinity along the genus-1 conformal boundary.

Without loss of generality, we can label the distance along the geodesic which is translated by the chosen Schottky generator by integers, and imagine that the source is attached at a boundary point x connected to the vertex 0. The bulk vertex v at which we want to evaluate the Green's function will be attached

to vertex n ($0 \leq n < \ell$), at a depth h from the central geodesic. The quantity to be calculated is simply

$$\varepsilon_{\kappa,x}^{(g=1)}(v) = \sum_{g\in\mathbb{Z}} p^{\kappa\langle v,gx\rangle}, \tag{4.84}$$

where the sum ranges over the images of x under the Schottky group. We take the integrand to be normalized to 1 at the vertex where the branch containing x meets the central geodesic. The cases $n = 0$ and $n \neq 0$ are different, and we will treat them separately.

$n = 0$: In this case, the sum becomes

$$\begin{aligned}\varepsilon_{\kappa,x}(v) &= p^{\kappa\langle v,x\rangle} + 2\sum_{m>0}\left(p^{-\kappa\ell}\right)^m \\ &= p^{\kappa\langle v,x\rangle} + \frac{2p^{-\kappa h}}{p^{\kappa\ell}-1}.\end{aligned} \tag{4.85}$$

$n \neq 0$: In this case, the sum becomes

$$\begin{aligned}\varepsilon_{\kappa,x}(v) &= \sum_{m\leq 0} p^{\kappa(-n-h-|m|\ell)} + \sum_{m>0} p^{\kappa(n-h-m\ell)} \\ &= p^{-\kappa h}\left(\frac{p^{\kappa(\ell-n)}+p^{\kappa n}}{p^{\kappa\ell}-1}\right).\end{aligned} \tag{4.86}$$

In both cases, the result has the expected boundary behavior: it falls off asymptotically as $p^{-\kappa h}$ when v approaches any boundary point other than x itself.

4.7 Conclusion

In this work we have proposed an algebraically motivated way to discretize the AdS/CFT correspondence. The procedure of replacing real or complex spacetimes by $\mathbb{Q}_p$ introduces a nontrivial discrete bulk and boundary structure while still preserving many desirable features of the correspondence. The boundary conformal field theory lives on an algebraic curve in both the ordinary and non-archimedean examples; the $\mathrm{P}^1(\mathbb{Q}_p)$ theory naturally enjoys the p-adic analogue of the familiar global conformal symmetry, $\mathrm{PGL}(2,\mathbb{Q}_p)$. This same group comprises the isometries of the lattice bulk spacetime $T_p = \mathrm{PGL}(2,\mathbb{Q}_p)/\mathrm{PGL}(2,\mathbb{Z}_p)$, a maximally symmetric coset space analogous to Euclidean AdS.

In analogy with the BTZ black hole and higher-genus examples in AdS_3, higher-genus bulk spaces in the p-adic case are obtained by Schottky uniformization. One takes quotients of the geometry by p-adic Schottky groups $\Gamma \subset \mathrm{PGL}(2, \mathbb{Q}_p)$, producing Mumford curves at the boundary. These curves holographically correspond to bulk geometries consisting of discrete black holes, which appear automatically and do not need to be put in by hand.

Having found a discretization which does not break any symmetries of the problem, we then proposed one way of obtaining a holographic tensor network from a Bruhat–Tits tree. We roughly identify the tree as a space of discrete geodesics in the network. Following Pastawski *et al.*'s holographic error-correcting code, the entanglement entropy of a deleted region is reproduced by counting geodesic lengths in the bulk. This perspective puts a stronger notion of bulk geometry into tensor networks, and suggests that the p-adic systems considered here may be closer to tensor network models than their archimedean counterparts. This construction might have further applications in entangled bulk states, nongeometric bulk states, and other more exotic features of quantum gravity not present in many existing tensor network models.

After discrete bulk Hilbert spaces in tensor networks, we then turned our attention to continuous Hilbert spaces of scalar fields in the tree. In the semiclassical analysis, massless and massive scalar solutions to the lattice model couple naturally to CFT operators at the boundary, just as in the archimedean case. We identified boundary/bulk propagators in the discrete analog of empty AdS, as well as in the p-adic BTZ black hole; the method of images can be used to generalize these results to arbitrary higher-genus bulk backgrounds. We are led to believe that the semiclassical physics of the bulk "gravity" theory is dual to an exotic conformal field theory living on the fractal p-adic boundary. At the present time, little is known about these p-adic conformal field theories outside of p-adic string theory; we hope the connection to holography may draw attention to this area. Viewed as a renormalization scale, we have shown that moving up the tree corresponds to exact course graining of boundary mode expansions. The intimate relation between conformal symmetry, AdS geometry, and renormalization still holds in this entirely discrete setting.

Motivated by the tensor network models, we suggest that the entanglement entropy of regions of the field theory is computed by the unique geodesic lengths in the bulk space. While as of yet we have no formal proof in the free-

boson field theory, a number of arguments have been presented which support this conjecture. Under very specific circumstances, it might even be possible to learn certain properties of the archimedean entanglement entropies from their corresponding p-adic counterparts with the help of adelic formulas.

While we have established some essential features of p-adic holography, ranging from algebraic curves to tensor networks and from bulk/boundary propagators and renormalization scales to entanglement, much about these exotic systems remains to be understood. We propose a number of ideas to be explored in future work.

One major ingredient missing from our story is a proper description of (and quantization of) the gravitational degrees of freedom. The bulk geometries (with or without black holes in the interior) can loosely be described as p-adic discretizations of asymptotically AdS spacetimes. One way to add dynamical metric degrees of freedom without spoiling the asymptotic behavior might be to make the edge lengths of the Bruhat–Tits tree dynamical. The p-adic version of empty AdS might correspond to a solution with uniform edge lengths like the system considered here; thermal or black hole states seem to require topology change in the interior.

If we believe that the full quantum gravity Hilbert space of the interior involves fluctuating edge lengths and graph topology, one might ask if tensor network models could be adapted to this picture. More complicated tensor networks might be used to study objects such as black holes, EPR pairs, and nongeometric states. The role of planarity of the tensor network may play an important role in this story.

From the point of view of the p-adic conformal field theory, one might ask for more interesting examples than the free boson. We have already offered some speculations about higher spin fields based on representation theory of the p-adic conformal group; it would be nice to formulate these models explicitly and search for interesting gravity duals. Additionally, the models we have studied so far do not appear to have extended conformal symmetry or a central charge. These important ingredients of 1+1 dimensional CFT's might appear with the more careful inclusion of finite extensions of $\mathbb{Q}_p$. These finite extensions might also be linked to the passage to Lorentzian signature.

We finally address future work for the entanglement entropy in a p-adic holo-

graphic theory. As already mentioned, the single and multiple interval entanglement entropies will likely require a detailed replica computation. This may be possible through a more detailed study of branched covers of the p-adic plane as Mumford curves. With entanglement entropies in hand, one might ask for new and old proofs of entropy inequalities; these are expected to be simplified by the ultrametric nature of the p-adics. Finally, it remains to be seen how much can be learned about real AdS/CFT from studying these systems adelically over every prime.

Note added

As this work was being completed, we became aware of [105], which treats similar ideas from a somewhat different viewpoint, and in which some of our results in §4.4 were independently obtained.

Chapter 5

A-POLYNOMIALS OVER FINITE FIELDS

[1]S. Gukov and I. Saberi. "Motivic quantum field theory and the number theory of A-polynomial curves." To appear.

Introduction: quantization and the path integral

The entire existing formulation of quantum field theory can be based on the idea of the path integral. Since the pioneering work of Feynman [78], the most fruitful heuristic for understanding quantization has been to imagine an integral over all possible histories of a physical system, weighted by probability amplitudes $\exp(iS)$, where S is the classical action of a history. Thus, one is naturally lead to consider integrals of the type

$$\int \mathscr{D}\phi\, f(\phi), \tag{5.1}$$

where ϕ here stands for a choice of values for all physical degrees of freedom throughout the spacetime. f is a functional of the fields, and usually includes the phase weighting $\exp(iS)$ together with possible insertions of other observables of interest. Unfortunately, in generic quantum field theories of physical interest, the spaces of field configurations over spacetime are infinite-dimensional, and no way to define the measure $\mathscr{D}\phi$ or make rigorous sense of integrals of this kind is known. Lacking a way around this roadblock, (5.1) remains a heuristic rather than an algorithm for dealing with the problem of quantization.

Several examples of quantum field theories for which we are able to rigorously evaluate integrals of the form (5.1) are provided by *topological field theories*, or TQFTs. Here, the space of field configurations over which we must integrate is *a priori* infinite-dimensional, but the path integral localizes to some finite-dimensional space X, which is often the moduli space of solutions to some set of PDEs. X may be, for instance, a finite-dimensional algebraic variety. For instance, in Witten's approach to Donaldson theory, the path integral for a topological field theory obtained from a twist of $\mathscr{N} = 2$ super Yang-Mills theory in four dimensions is interpreted in terms of the moduli space of anti-self-dual instantons [211].

Topological field theories compute topological invariants of spacetimes and structures within spacetime. For instance, as first shown in [212], Chern-Simons theory computes invariants of three-manifolds, and of knots and links in three-manifolds when Wilson line observables supported along those knots are inserted into (5.1). In order to compute these invariants directly by evaluating the path integral, a measure must be chosen. It is natural to ask: what possible choices of measure could be made? Do different choices lead to different invariants, and is there a single universal or best-possible measure such that all other integrals arise from it as specializations?

If the space X is algebraic, a measure of this kind has in fact been constructed by mathematicians: the motivic integration measure $d\mu$ [127, 142, 154, 200]. Therefore, our goal in this note is to develop TQFT where the path integral is performed with the motivic measure. As one might expect, the use of this finer measure leads to theories that compute finer topological invariants. For example, in some examples of TQFTs of cohomological type, the ordinary choice of integration measure is simply the Euler density, so that the corresponding integral computes the Euler characteristic of X [155]. Using the motivic measure computes a finer invariant, akin to the Poincaré polynomial of the homology of X. The Euler characteristic can of course be recovered from this as a specialization.

The study of moduli spaces has shed new light on topological field theories from other angles as well. Recall that, in the axiomatic approach typically used by mathematicians that was first formulated by Atiyah [8], a d-dimensional TQFT is a tensor functor from the $(d-1,d)$-dimensional cobordism category to the category of complex vector spaces. Less concisely, a TQFT assigns a (finite-dimensional) Hilbert space of states to any $(d-1)$-manifold Y, which one can think of as a spacelike slice or codimension-one boundary of a spacetime on which boundary conditions are to be specified. It should assign to the disjoint union of two such manifolds the tensor product of their respective Hilbert spaces. Given a d-manifold M with boundary $\partial M = Y_0 \sqcup Y_1^\vee$, which is a morphism from Y_0 to Y_1 in the cobordism category, the TQFT should associate to it a linear map between the Hilbert spaces $\mathscr{H}_0$ and $\mathscr{H}_1$. Here $Y^\vee$ denotes Y with opposite orientation. For further details, the reader should consult the literature, e.g. [8, 20, 83] and references therein.

This is all very well, but where does one obtain an interesting example of such

Geometric object	*Physical data*	*Classical theory*	*Quantum theory*
Y (a space)	Space of possible states	Phase space: the symplectic space X of classical solutions	Hilbert space $\mathscr{H}$ of quantum states
X (a spacetime) $\partial X = Y$	Choice of a particular state	a Lagrangian submanifold $L \subset X$	a ray in $\mathscr{H}$

Figure 5.1: A schematic illustration of the way physical theories associate data to spacetimes and their boundaries.

a functor to study? In physics, a field theory (topological or not) is usually defined by writing down the collection of fields or degrees of freedom in the theory, together with the action functional. This data defines the theory, either as a classical variational problem—one asks that the action be extremized, from which condition one obtains the equations of motion or PDEs that the fields must satisfy—or as a quantum theory, in which we must either make sense of the path integral or quantize the theory by other means. As such, examples of TQFTs in physics may come with an associated classical theory, from which they are reconstructed by quantization. One can say that the functor "factors" through a classical field theory, as illustrated in Fig. 5.1.

The arrow that goes from moduli spaces (i.e., classical theories) to TQFTs is provided by quantization. Study of the quantization of classical phase spaces has been fruitful for the study of quantum invariants, leading for instance to the AJ-conjecture [107] (see also [111] for a recent review). This line of work suggests from a slightly different angle that moduli spaces hold the key to deeper understanding of TQFTs.

The motivic integration measure mentioned above can also be related to the study of the algebraic spaces X that arise as moduli spaces in TQFT over different fields, such as finite number fields or p-adic numbers. In fact, the motivic measure is intimately related to the p-adic integration measure, which can be defined by counting points on X over the rings $\mathbb{Z}/p^k\mathbb{Z}$. Motivated by this reasoning, we also report on some experimental investigation into the number-theoretic properties of various spaces X occurring in topological field theories. Specifically, we consider *A-polynomial curves*, which are invariants of knots that arise from classical $\mathrm{SL}(2,\mathbb{C})$ Chern-Simons theory [54, 107]. To

a given knot, this theory associates a planar algebraic curve that is the zero locus of a polynomial with integer coefficients. This curve is the moduli space of vacua, or the classical phase space, in the $\mathrm{SL}(2,\mathbb{C})$ theory mentioned above. We review the construction in §5.2. Since its coefficients are integers, the A-polynomial can be studied over finite fields, using techniques of arithmetic geometry.

By doing this, we immediately obtain infinitely many new positive-integer-valued knot invariants, indexed by prime powers, that are defined simply as the number of points on the A-polynomial curve over the field $\mathbb{F}_{p^k}$. Of course, these do not carry independent information, since the whole ensemble of them is computable from the A-polynomial itself. However, in light of the above, invariants of this sort may be worthy of study. One might ask, in particular, whether any of these invariants coincide with more conventional integer-valued topological invariants of knots (Seifert genus, for example). In all cases that we checked, the answer is no.

Moreover, the independent data of this collection of one integer for each prime power can be conveniently repackaged into a finite collection of only $2g-1$ integers at each prime, which are the coefficients appearing in the numerator of the local zeta function that encodes the point-counting data for X in characteristic p. Here g is the genus of the planar curve X. We identify this zeta function for the example of the figure-eight knot. In this instance, X is elliptic; by the celebrated modularity theorem [35], the global zeta function (product of all local factors) is therefore a modular form.

A somewhat more sophisticated question is whether the point-counting numbers defined above, or some relatives of theirs, are invariants of finite type. Unfortunately, this question is more subtle than it might appear (and not immediately amenable to experimental tests) since it requires definition of these invariants for links of more than one component. The A-polynomial is only a reasonably tractable planar curve for knots; for more general links, the dimension of the moduli space is larger. As such, we leave the resolution of this question to future work. The outcome of our experimental investigation consists in the identification of new regularities common to all A-polynomial curves, which occur in particular in the structure of their singularities. We offer some speculations as to the physical interpretation of these regularities.

5.1 The motivic measure

We recall the notion of a measure, as customarily defined in the theory of Lebesgue integration. A *measure space* is a triple $(X, \mathscr{F}, \mu)$, where X is a fixed set, $\mathscr{F}$ is a σ-algebra of subsets of X, and μ is a measure function on $\mathscr{F}$. μ is required to have certain properties, important among which is additivity: for any countable collection $\{E_i\}$ of sets in $\mathscr{F}$ that are pairwise disjoint from one another,

$$\mu\left(\bigcup_i E_i\right) = \sum_i \mu(E_i). \tag{5.2}$$

Here, the measure is usually required to take values in the positive real numbers. For the time being, we remain agnostic about the target space of μ, which may be any abelian group.*

We note that, if X is a topological space, there is a natural choice of σ-algebra over X: the Borel algebra generated by the closed subsets of X. Our first approximation to motivic measure can be thought of as a measure on the algebra generated by the Zariski-closed sets, where X is a variety.† That is: motivic measure should assign, to each object in the category **Var** of varieties over a field k, a "number" of some general sort (element of an abelian group), in a way that is additive: we should have

$$\mu(X) = \mu(Y) + \mu(X \setminus Y), \tag{5.3}$$

whenever Y is a Zariski-closed subset of X.

Now, there is a universal choice of the abelian group in which such a measure can be valued—that is, every other measure will factor through this "most general" possible measure. Let the *Grothendieck ring* of the category **Var** be the free abelian group on the isomorphism classes of objects in **Var**, subject to all relations of the form

$$[X] = [Y] + [X \setminus Y], \tag{5.4}$$

*In fact, we only need the set of values for our measure to admit a commutative addition operation with a unit element. Thus, we can in fact take the target space of μ to be a commutative monoid. This distinction is unimportant in practice, though, since any such monoid obeying the cancellation property $a + b = a + c \implies b = c$ can be embedded as a subset of an abelian group (the map from the monoid to its Grothendieck group is injective). It amounts to thinking of an ordinary measure as valued in the commutative monoid $\mathbb{R}_{\geq 0}$, as opposed to the abelian group $\mathbb{R}$.

†A σ-algebra should be closed under countable set operations; we consider the weaker notion of algebra of sets, where only closure under finite unions and intersections is required.

for $X, Y \in \mathrm{Ob}(\mathbf{Var})$ such that Y is Zariski-closed in X. The Grothendieck group of a category $\mathscr{C}$ is denoted $K_0(\mathscr{C})$. $K_0(\mathbf{Var})$ is a ring because $\mathbf{Var}$ is a tensor category; the product is defined by

$$[X] \cdot [X'] \hat{=} [X \times X']. \tag{5.5}$$

Now, the map

$$\mathrm{Ob}(\mathbf{Var}) \to K_0(\mathbf{Var}), \quad X \mapsto [X] \tag{5.6}$$

satisfies our additivity requirement for measures (and is universal for this requirement) by construction. Note that the class of a variety in $K_0(\mathbf{Var})$ carries strictly less information than its isomorphism class: for every locally trivial fibration $E \twoheadrightarrow B$ with typical fiber F, we have $[E] = [B] \cdot [F]$.

Thus, mapping a variety to its equivalence class in the Grothendieck ring $K_0(\mathbf{Var})$ behaves like a measure. Indeed, this "measure" is defined for *constructible subsets* of varieties, i.e., finite disjoint unions of locally closed subvarieties; we simply map $\sqcup_i Y_i$ to $\sum_i [Y_i]$. Thus it makes sense on the whole algebra of sets generated by the Zariski-closed subsets of a variety. However, we need to be able to measure a wider class of sets before we can define a satisfactory motivic integral.

Let $X \in \mathrm{Ob}(\mathbf{Var})$ be a variety over some fixed ground field (say $\mathbb{C}$). We can imagine for simplicity that X is the zero locus of a collection of polynomial equations:

$$X : f_i(x_1, \ldots, x_N) = 0, \quad 1 \leq i \leq \operatorname{codim} X.$$

The space of n-jets on X is the space of solutions to the same polynomial equations, where the variables x_j take values in $\mathbb{C}[t]/(t^{n+1})$ rather than in $\mathbb{C}$. We can think of the formal nilpotent variable t as representing an infinitesimal; then the n-jets on X are deformations of solutions of the defining equations for X in this infinitesimal parameter that solve the equations to nth order, or parameterized curves that remain on X to nth order in the parameter. In particular, the space of 1-jets on X is just the tangent bundle TX.

We denote the space of n-jets on X by $\mathscr{J}_n(X)$; it is an algebraic variety. Expanding the equations in powers of t (modulo t^{n+1}) gives a collection of $n \operatorname{codim} X$ equations for nN variables in $\mathbb{C}$. Thus, if $d = \dim X$, the space $\mathscr{J}_n(X)$ is generically nd-dimensional.

It is obvious that any n-jet on X is also an m-jet on X in a natural way for any $m \leq n$; thinking of points on $\mathscr{J}_n(X)$ as points on X with coordinates

in $\mathbb{C}[t]/(t^{n+1})$, we can just reduce modulo t^{m+1} to get a point on X with coordinates in the ring $\mathbb{C}[t]/(t^{m+1})$, i.e., an element of $\mathscr{J}_m(X)$. This defines a natural family of projection maps

$$p_m^n : \mathscr{J}_n(X) \to \mathscr{J}_m(X) \tag{5.7}$$

for all $m \leq n$. (p_n^n is the identity.)

We can now define the space of arcs on X to be the inverse limit

$$\mathscr{J}(X) = \varprojlim \mathscr{J}_n(X). \tag{5.8}$$

The points of $\mathscr{J}(X)$ are in bijection with points on X whose coordinates are formal power series in $\mathbb{C}[[t]]$. Thinking of n-jets as parameterized paths remaining in X to nth order in the parameter, we should think of arcs as representing some kind of formal parameterized trajectories in X, which are suitably analytic (representible by power series). $\mathscr{J}(X)$ is an infinite dimensional space, defined by the vanishing of countably many polynomial equations, each of which contains only a finite subset of the countably many $\mathbb{C}$-valued variables. There are natural maps $\pi_n : \mathscr{J}(X) \to \mathscr{J}_n(X)$ for all n. In general, these maps are not surjective; the space $\operatorname{im} \pi_n$ consists of those n-jets that may be lifted to arcs on X.

We say that a subset S of $\mathscr{J}(X)$ is *constructible* if there exists some m such that $S_m \mathrel{\hat{=}} \pi_m(S)$ is a constructible subset of $\mathscr{J}_m(X)$, and $S = \pi_m^{-1}(S_m)$. The constructible subsets form an algebra of sets on $\mathscr{J}(X)$. We further say that a constructible set S is *stable* if the projections π_n^{n+1} are all locally trivial fibrations. When X is nonsingular, every constructible subset is stable.

For a stable subset S, we know that

$$[\pi_n(S)] = [\mathbb{A}^1]^{(n-m)d} \cdot [\pi_m(S)] \in K_0(\mathbf{Var}), \quad \forall n \geq m. \tag{5.9}$$

Here $[\mathbb{A}^1]$ just denotes the class of the affine line in the Grothendieck ring; henceforth, we represent this element simply by L. Thus, if we allow ourselves to invert the element $L \in K_0(\mathbf{Var})$, we can extend our definition of motivic measure to all stable subsets of $\mathscr{J}(X)$ as follows: Let $\mathscr{M} \mathrel{\hat{=}} K_0(\mathbf{Var})_L$ be the ring obtained from the Grothendieck ring by inverting the class of the affine line. ($\mathscr{M}$ depends implicitly on a choice of ground field, as does the category **Var**; throughout, we choose not to indicate this dependence, and

work over $\mathbb{C}$.) Then, for stable S, we define

$$\mu(S) \hat{=} \lim_{n\to\infty} [\pi_n(S)] \cdot L^{-nd}. \tag{5.10}$$

The limit is well-defined because the sequence stabilizes for $n \geq m$. This measure generalizes the operation of mapping a variety to its class in the Grothendieck ring; if X is a nonsingular variety, then $\mu(\mathscr{J}(X)) = [X]$. (We have abused notation slightly, writing $[X]$ for its image in the localized ring $\mathscr{M}$.)

We would now like to extend μ to an additive measure on the whole algebra of constructible subsets of arc space. However, we cannot do this naively, since the limit in (5.10) may not be sensible if the sequence fails to stabilize. This will happen when the variety X has some singularities.

Example 12. Let X be the variety $\{xy = 0\} \subset \mathbb{C}^2$. Then

$$X = \mathbb{A}^1 \cup (\mathbb{A}^1 \setminus \{0\}).$$

We recall that $[\mathbb{A}^k] = L^k$, so that $\mathbb{A}^0 = 1$. Therefore,

$$[X] = 2L - 1.$$

The space $\mathscr{J}_1(X)$ consists of the algebraic set

$$(x_0, x_1, y_0, y_1) \in \mathbb{C}^4 : (x_0 + x_1 t)(y_0 + y_1 t) = 0 \pmod{t^2},$$

i.e., the zero locus of the two equations

$$x_0 y_0 = 0, \quad x_0 y_1 + x_1 y_0 = 0.$$

If x_0 is nonzero, the equations imply that $y_0 = y_1 = 0$ and x_1 is unconstrained. Thus this part of the variety looks like the locally trivial fibration $(\mathbb{A}^1 \setminus \{0\}) \times \mathbb{A}^1$. An identical argument applies if $y_0 = 0$. However, if $x_0 = y_0 = 0$, both x_1 and y_1 are unconstrained, and this part of the variety is a copy of $\mathbb{A}^2$. Therefore,

$$[\mathscr{J}_1(X)] = 2(L-1)L + L^2 = 3L^2 - 2L.$$

On the other hand, the set $\pi_1 \mathscr{J}(X)$ consists of those 1-jets that may be lifted to arcs on X. Thus, there is an additional equation that comes from order t^2:

$$x_0 y_0 = 0, \quad x_0 y_1 + x_1 y_0 = 0, \quad x_2 y_0 + x_1 y_1 + x_0 y_2 = 0, \ldots$$

If x_0 is nonzero, the equations reduce to $y_i = 0$ for all i, with x_i unconstrained. If $x_0 = y_0 = 0$, however, we additionally have the equation $x_1 y_1 = 0$; thus, this part of the variety is equivalent to X rather than to $\mathbb{A}^2$! Thus, $\pi_2 \mathscr{J}(X)$ is not all of $\mathscr{J}_2(X)$, and in fact

$$[\pi_2 \mathscr{J}(X)] = 2(L-1)L + (2L-1) = 2L^2 - 1.$$

Continuing the calculations in this manner leads us to the general results

$$[\mathscr{J}_n(X)] = (n+2)L^{n+1} - (n+1)L^n, \quad [\pi_n \mathscr{J}(X)] = 2L^{n+1} - 1. \tag{5.11}$$

These results illustrate why the projection of the arc space, rather than the full jet space, is used in our limiting procedure: there is no way to make sense of the limit of $[\mathscr{J}_n(X)]/L^n$ as $n \to \infty$ in our example. Furthermore, they illustrate that singular points on X prevent the sequence from stabilizing: we have

$$\frac{[\pi_n \mathscr{J}(X)]}{L^n} = 2L - \frac{1}{L^n},$$

which should clearly be interpreted as tending to the value $2L$. For this to work, we must pass to an appropriate completion $\widehat{\mathscr{M}}$ of $\mathscr{M}$, so that these limits are well defined. Note that the value $2L$ is the same as for the disjoint union of two affine lines; thus, the motivic measure can be thought of as naturally accomplishing some kind of resolution of singularities.

Relation to point-counting over $\mathbb{Z}/p^n\mathbb{Z}$

The motivic measure we have outlined above relies on looking at the limiting behavior of the images of successive jet spaces $\pi_n \mathscr{J}(X) \subseteq \mathscr{J}_n(X)$ in the Grothendieck ring of varieties. A similar approach is used in the theory of p-adic integration, which in fact was the original motivation for motivic integrals. We briefly outline the analogy here. Let us take for simplicity a curve defined by the vanishing of a single polynomial with integral coefficients:

$$X : f = 0.$$

It makes sense to count the number of solutions of the congruence $f \equiv 0 \pmod{n}$, for any natural number n, which can be thought of as the number of points on the variety X over the ring $\mathbb{Z}/n\mathbb{Z}$. If $n = p^{k+1}$ is a prime power, then this is the same as asking for solutions to $f = 0$ with coefficients in the ring

$$\mathbb{Z}/p^{k+1}\mathbb{Z} \cong \mathbb{F}_p[t]/(t^{k+1}),$$

where $\mathbb{F}_p$ is the finite field $\mathbb{Z}/p\mathbb{Z}$, and the formal variable t we introduced in the above equation just plays the role of the prime p. Put another way, every element $\alpha \in \mathbb{Z}/p^{k+1}\mathbb{Z}$ has a unique expansion

$$\alpha = \alpha_0 + \alpha_1 p + \cdots + \alpha_k p^k, \quad \alpha_i \in \mathbb{F}_p, \tag{5.12}$$

which gives the isomorphism above. But this shows that the points on X over $\mathbb{Z}/p^{k+1}\mathbb{Z}$ are exactly the points on $\mathscr{J}_k(X)$ over $\mathbb{F}_p$!

The proper analogue of the arc space of X, in this case, is just isomorphic to the set of points of X over the ring

$$\mathbb{F}_p[[t]] \cong \varprojlim \mathbb{Z}/p^{k+1}\mathbb{Z} = \mathbb{Z}_p,$$

the ring of p-adic integers. Just as t may be thought of as an infinitesimal quantity in the context of jet spaces, so each successive term in the expansion (5.12) is "one order smaller;" recall that, in the p-adic valuation, a number is small when it is divisible by a large power of p.

In the motivic case, we found that the measure of a variety could always be expressed in terms of powers of the symbol L, the class of the affine line. The same is true in the case of jet spaces of varieties over $\mathbb{F}_p$: here, the "measure" is just the naive count of the number of solutions, which can always be expressed as a function of the prime p. And p is just the number of points on the affine line over $\mathbb{F}_p$! Thus, we can determine the motivic measure of a variety in many cases by solving equations over $\mathbb{F}_p$, and then simply replacing p in our answer by the symbol L.

The analogue of the Grothendieck ring of varieties, in which the point-counting measure takes values, is simply $\mathbb{Z}$. Localization of the Grothendieck ring at L, in this case, corresponds to inverting the element $p \in \mathbb{Z}$, and the subsequent completion returns $\mathbb{R}$ as the final value ring for p-adic integration. We mention that a theorem of Weil relates p-adic integration on varieties to counting points on those varieties over finite fields.

5.2 Reduction of A-polynomial curves modulo p: singularities and common points

We briefly recall the definition of the A-polynomial. Let $K \subset S^3$ be a knot. Cutting out a tubular neighborhood $N(K)$ of the knot produces a three-manifold M called the knot complement, with boundary homeomorphic

to T^2. We then consider representations of the fundamental group $\pi_1(T^2)$ into $\mathrm{SL}(2,\mathbb{C})$. Since the two generators of $\pi_1(T^2)$ commute, their representation matrices can be simultaneously diagonalized, and any such representation is characterized up to conjugacy by a choice of one eigenvalue for each generator:

$$\mathrm{Hom}(\pi_1(T^2),\mathrm{SL}(2,\mathbb{C}))/\mathrm{conj.} \cong (\mathbb{C}^\star \times \mathbb{C}^\star)/\mathbb{Z}_2,$$

where $\mathbb{Z}_2$ acts by replacing each eigenvalue by its inverse and can be thought of as the Weyl group of $\mathrm{SL}(2,\mathbb{C})$. This isomorphism is not quite canonical; it depends on a choice of basis for the homology group $\mathcal{H}_1(T^2) = \pi_1(T^2)^{\mathrm{ab}}$. We can take these to be a meridian cycle—the unique generator of $\pi_1(N(K))$ whose representative lying in $\partial N(K)$ has linking number $+1$ with K—and a longitude cycle, obtained by pushing the knot laterally away from itself to the boundary of $N(K)$. The longitude is not unique, but is fixed by a choice of framing of K. We will denote the eigenvalues of the longitude and meridian cycles respectively by ℓ and m. Note furthermore that the group $\mathrm{SL}(2,\mathbb{Z})$ naturally acts on the set of all choices of basis; if we take

$$\begin{pmatrix}\gamma_\ell\\ \gamma_m\end{pmatrix} \mapsto \begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}\gamma_\ell\\ \gamma_m\end{pmatrix},$$

the A-polynomial transforms covariantly according to the rule [107]

$$A(\ell,m) \mapsto A(\ell^d m^{-b}, \ell^{-c} m^a).$$

Now, the obvious map $\pi_1(T^2) \to \pi_1(M)$ gives rise to a map of the corresponding representation spaces

$$\mathrm{Hom}(\pi_1(M),\mathrm{SL}(2,\mathbb{C})) \to \mathrm{Hom}(\pi_1(T^2),\mathrm{SL}(2,\mathbb{C})),$$

which descends to the quotient by conjugacy, and whose image after the quotient turns out to be an algebraic set in $\mathbb{C}^\star \times \mathbb{C}^\star$. One of the components of this algebraic set is a planar algebraic curve. We can choose a defining polynomial $A(m,l) \in \mathbb{Z}[m,l]$ for this curve, which is called the A-polynomial of K.

In all of our experiments with the A-polynomial, we choose to work with the natural variables m and l that are eigenvalues of our chosen meridian and longitude cycles on the boundary torus of the knot. Occasionally, we make use of the variables $(x,y) = (m^2, l)$, and we will abuse notation and use $A(m,l)$ and $A(x,y)$ to denote the A-polynomial before and after this change

of variables respectively. We should also remark that all A-polynomials are divisible by a factor $(\ell - 1)$, which arises from the abelian representations. We follow the conventions of [54], using the table of A-polynomial invariants for simple knots given in that paper and discarding this trivial component of the curve in most of our considerations.

One obvious experiment to perform is simply to plot the reduction modulo p for various A-polynomials and various values of p. The zero locus of $A(m, l)$ is a subset of $\mathbb{A}^2(\mathbb{F}_p) = (\mathbb{Z}/p\mathbb{Z})^2$, which consists of p^2 points that can be drawn as a square region of a lattice. (The projective completion $\mathbb{P}^2(\mathbb{F}_p)$ consists of $p^2 + p + 1$ points, and the curve may have points at infinity as well as points at finite distance.) One can look for two kinds of regularities in this data: regularities as p varies over primes for a fixed knot, and regularities that occur in the A-polynomial curves of all knots at some given p.

The results of these observations led to noticeable patterns only of the second kind:

Observation 5. The A-polynomial curve of any knot always has the following "universal" $\mathbb{F}_p$-rational points:

- $(m, l) = (\pm 1, -1) \in A(\mathbb{F}_p)$ for every p;
- $(m, l) = (\pm k, 1) \in A(\mathbb{F}_p)$ whenever $k^2 = -1$ in $\mathbb{F}_p$. By the quadratic reciprocity theorem, such a k exists (-1 is a quadratic residue) precisely when $p \equiv 1 \pmod 4$.

Recalling that the A-polynomial for a knot in S^3 contains only even powers of the variable m, we can summarize this result in a more compact form by saying that the points $(m^2, l) = (1, -1)$ and $(-1, 1)$ are universal points that occur on all A-polynomial curves.

A few remarks are in order. First of all, as the observant reader will no doubt have noticed, these points do not depend on the value of p for which we look at the curve; indeed, the characterization of these points in the above paragraph makes sense over any field, since 1 and -1 are always defined. Therefore, we can simply observe that the universal points are present on any such curve over any field (finite or not). In particular, these points can be seen over $\mathbb{C}$.

Secondly, the first pair of universal points are fixed points of the $\mathbb{Z}_2$ action that takes $(m, l) \mapsto (1/m, 1/l)$. However, not all fixed points of the Weyl group action on $\mathbb{C}^\star \times \mathbb{C}^\star$ are universal points; for instance, $(m, l) = (1, 1)$ is not universal (and does not occur, for instance, on the A-polynomial for the figure eight knot). Of the four fixed points, exactly two are universal. Moreover, the other pair of universal points are mapped into one another by the Weyl group action (in particular, neither is a fixed point). This is as we would expect: since the A-polynomial naturally lives in the quotient by the Weyl group, the points in a Weyl orbit of a universal point should be universal as well. However, we do remark that the second pair of universal points is special in the following sense: A-polynomial curves of knots in S^3 admit a second $\mathbb{Z}_2$ action, mapping $(m, l) \mapsto (-m, l)$. (This can be thought of as analogous to a deck transformation: since $A(m, l)$ contains only even powers of m, the change of variables $x = m^2$ gives a mapping from the A-polynomial curve to a "reduced" A-polynomial curve that is two-to-one. The bad point $m = 0$ is excluded from consideration, since our variables are valued in $\mathbb{C}^\star$.) A generic point therefore has a symmetry orbit of size four. The first pair of universal points are an orbit of size two, where the first $\mathbb{Z}_2$ (Weyl group) acts trivially, but the second exchanges the two points. The second pair of universal points are *also* an orbit of size two; the Weyl group and the deck transformation act identically and swap the two points, and the product of the two symmetries acts trivially. Therefore, if we make the change of variables $(x, y) = (m^2, l)$ which corresponds to passing to the "reduced" curve that is a quotient by the second $\mathbb{Z}_2$, both pairs of universal points project down to single points that are then fixed by the Weyl-group action on the reduced curve.

Lastly, the presence of these two points is apparent in the case of $(2, p)$ torus knots, where the A-polynomial takes the simple form $A(m, l) = 1 + lm^{2p}$. Here p is an odd prime.

We further observe the following:

Observation 6. The A-polynomial curves of all hyperbolic knots are singular. In particular, the universal points identified above are always singular points over $\mathbb{C}$.

There are easy counterexamples to this observation when the knot in question is not hyperbolic. For instance, we can simply look at the A-polynomial for

the simplest torus knot, the trefoil:

$$A(m,l) = l + m^6.$$

It is trivial to check that both pairs of universal points are present on this curve, but that the curve is nonsingular everywhere in $\mathbb{C}^\star \times \mathbb{C}^\star$.

Our discussion here should be put into context with the related discussion of singularities of A-polynomial curves in [89], §6.1. That paper uses the variables $(x,y) = (m^2, l)$, so that our (unreduced) A-polynomial curves are double covers of their (reduced) curves.

As the reader can verify by direct calculation (or by looking ahead to §5.3 where we perform the calculation in detail) the reduced A-polynomial curve for the figure eight knot has exactly two singularities at finite distance, which lie at the universal points $(x,y) = (-1,1)$ and $(1,-1)$. The discriminant of this polynomial with respect to the variable y is

$$\Delta(x) = (1-x^2)^2(1+x+x^2)(1-3x+x^2), \tag{5.13}$$

which divides the discriminant given in [89, Eq. (6.6)]. As noted there, the last of the factors in the discriminant is just the Alexander polynomial of the figure-eight knot. However, we should point out a limitation of the ability of $\Delta(x)$ to identify singularities of $A(x,y)$. A polynomial in one variable $p(y)$ has zero discriminant if and only if it is singular, i.e., there exist solutions $y = y_\star$ to the equations $p(y) = p'(y) = 0$. Taking the discriminant of $A(x,y)$ with respect to y therefore identifies all values $x_\star$ of x for which a slice at fixed x is singular: that is, the equations

$$A(x,y) = 0, \quad \partial_y A(x,y) = 0 \tag{5.14}$$

admit simultaneous solutions $y = y_\star$. This is a necessary, but not sufficient, condition for the point $(x_\star, y_\star)$ to be a singular point of the curve $A(x,y) = 0$! We must still check whether the equation

$$\partial_x A(x,y) = 0$$

is satisfied at each candidate point. If it is not, there is no singularity. A good example to keep in mind is the simple nonsingular curve $c(x,y) = y^2 - x$. Taking the discriminant with respect to y gives simply $\Delta(x) = -4x$, suggesting

that there may be a singular point with $x = 0$. However, the candidate point $(x, y) = (0, 0)$ is in fact nonsingular, since $\partial_x c(0, 0) \neq 0$.

As such, not every factor of $\Delta(x)$ in (5.13) gives a true singularity of the reduced A-polynomial curve. Indeed, only the first factor, corresponding to the values $x = \pm 1$, comes from singularities. The zeroes of the other factors of $\Delta(x)$ identify special points on $A(x, y) = 0$ that are still nonsingular. As the reader can verify directly, the conditions (5.14) for the figure-eight knot imply that $y = \pm 1$ for the candidate points. A quick check then reveals that the "candidate singularities" are

$$(x, y) = \begin{cases} \left(\frac{-1 \pm i\sqrt{3}}{2}, -1\right), & \text{for factor } 1 + x + x^2, \\ \left(\frac{3 \pm \sqrt{5}}{2}, +1\right), & \text{for factor } 1 - 3x + x^2. \end{cases} \tag{5.15}$$

None of these are singular on the curve we are considering. However, notice that the second set of candidate points, which come from zeroes of the Alexander polynomial factor of the discriminant, also lie on the zero set of $(y - 1)$, the factor corresponding to abelian representations that we excluded from our analysis of the A-polynomial. Had we included this factor (as [89] does), these points would indeed be singular, as they would lie on the intersection of two irreducible components of the curve. Since the factorization of the A-polynomial into abelian and nonabelian components does not generalize to the super-A-polynomial at other values of the deformation parameters, these should perhaps be regarded as true singular points. On the other hand, there seems to be no clear sense in which the first set of candidate points can be related to genuine singularities.

A schematic illustration of the situation is shown in Fig. 5.2, at left. The intersection of the reduced A-polynomial curve E with the plane where $x, y \in \mathbb{R}$ is shown. We also include the abelian factor $y = 1$ in a contrasting color. As the reader can see, the universal singular point $(x, y) = (-1, 1)$ is evident, and looks like a nodal self-intersection of the nonabelian curve (through which the abelian component also passes). The two other intersection points of the components are the "candidate singularities" that come from zeros of the Alexander polynomial. They are singular when both components are included, but are regular points of the nonabelian component considered alone. The other universal point at $(1, -1)$ is not visible, because the intersection of our curve with the chosen real slice there consists of a single point.

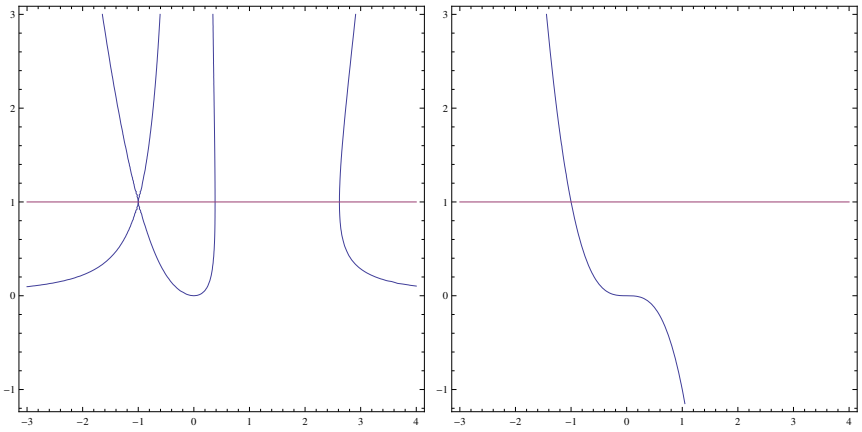


Figure 5.2: At left, a real section of the reduced A-polynomial curve for the figure eight knot. At right, the same plot for the trefoil knot.

At right, the same plot is shown for the trefoil knot. Since this is a torus knot, the universal points are not singular. However, both can be seen in the figure.

Let us say a few words about how this story generalizes to the super-A-polynomial. As is noted in [89], the super-A-polynomial for the figure-eight knot (and indeed all such polynomials that are explicitly known) possess universal singularities that live at

$$1 + at^3x^2 = 0. \tag{5.16}$$

The claim in [89] is merely that, for each choice of parameters a and t, and for each root of (5.16), there exist values for the y-coordinate that make the resulting points. We have checked this explicitly in the case of the figure-eight knot, and furthermore we find that the corresponding value of y is just $y = 1$ for all values of the parameters. Thus, these "universal singularities"[‡] can be more completely described, at least in the case of the figure-eight, by

$$1 + at^3x^2 = 0, \quad y - 1 = 0. \tag{5.17}$$

Thus, this family generalizes our second pair of universal points, $(x, y) = (-1, 1)$. The astute reader will notice that taking $a = 1$ and $t = -1$ in (5.17) indicates the presence of additional singular points at $(x, y) = (1, 1)$, which we have not mentioned. This has a simple explanation: the specialization of parameters does not precisely recover just one irreducible component of the ordinary A-polynomial. Rather, the exact relation is

$$A^{\text{super}}(x, y; a = 1, t = -1) = (x - 1)^2(y - 1)A(x, y), \tag{5.18}$$

[‡]This is the terminology used in [89].

where $A(x,y)$ is what we have termed the reduced A-polynomial. The reader can clearly see that all components of the ordinary A-polynomial occur as factors in (5.18), together with extra irreducible factors. Indeed, the presence of the factor $(x-1)^2$ means that any point with $x=1$ is singular—in particular, $(x,y)=(1,1)$ is, although it does not even occur as a zero of $A(x,y)$. This accounts for the extra singularities detected by (5.17).

We should also remark that a slight difference occurs in the universal singularities for the trefoil knot (which is not hyperbolic). Examination of the super-A-polynomial for the trefoil shows that the universal singularities live along the locus

$$1+at^3x^2=0, \quad ty+1=0. \tag{5.19}$$

The universal singularities thus have a simple (but different) description in each case. Moreover, they agree with one another when $t=-1$, i.e., when the homological deformation is made trivial. It would be interesting to check the location of universal singularities of the super-A-polynomial in more examples (in particular, for additional hyperbolic knots) to check whether the form (5.17) is always valid. It is also worth remarking that the super-A-polynomial for the unknot does not possess these universal singularities.

There is no clear way in which this universal family of singularities of the super-A-polynomial, or any of the other singularities identified in [89], is related to our first pair of universal points, $(x,y)=(1,-1)$. Moreover, the interpretation of these universal points is still unclear. However, our analysis allows us to conjecture as to the origin of the second set of universal points:

Conjecture 7. The second pair of universal points, $(m^2,l)=(-1,+1)$, and therefore the "universal singularities" of the super-A-polynomial, are related to quantizability of the A-polynomial curve.

As motivation for this conjecture, consider the following example curve that was first considered in [107]:

$$B(m,l)=1-(m^{-6}-m^{-2}-2-m^2+m^6)l+l^2. \tag{5.20}$$

This polynomial is a slight modification of the A-polynomial for the figure-eight knot, which can be written

$$A(m,l)=1-(m^{-4}-m^{-2}-2-m^2+m^4)l+l^2.$$

(Recall that we are free to multiply by powers of m and l.) Both of these curves are reciprocal, and have all the required symmetries. However, $B(m,l)$ is not quantizable, and hence does not occur as the A-polynomial of any knot. We notice that the first pair of universal points are present on $B(m,l)=0$. Indeed, they are even singular points, just as they are for $A(m,l)=0$. However, the second pair of universal points are not zeroes of $B(m,l)$, as the reader can easily check. Since the only property of the A-polynomial that $B(m,l)$ does not satisfy is the quantizability condition, it is tempting to conclude that the presence of the second pair of universal points is a signal of quantizability.

The first pair of universal points, as we mentioned above, have no clear interpretation as yet. However, they do occur on the "fake" example (5.20) considered above. It is therefore tempting to conclude that these points are not related to quantizability, but to some other, less subtle symmetry or property of the A-polynomial which (5.20) does indeed possess.

5.3 Zeta functions and modular forms

One of the most interesting and puzzling properties of the A-polynomial is that its coefficients are integral. This property allows us to view the curve $A(m,l)=0$ as a variety over the finite field $\mathbb{F}_p=\mathbb{Z}/p\mathbb{Z}$ for any prime number p, and count points on this curve, which are just solutions to the congruence $A(x,y)\equiv 0 \pmod{p}$.

To any variety X over $\mathbb{F}_p$, we can associate a *local zeta function* $Z_p(t)$. This zeta function is a generating function that encodes the numbers of points on the variety over all finite fields of characteristic p. It is defined as follows:

$$\log Z_p(t)=\sum_k N_{p^k}\frac{t^k}{k}, \quad \Longleftrightarrow \quad Z_p(t)=\exp\left(\sum_k N_{p^k}\frac{t^k}{k}\right). \tag{5.21}$$

Here $N_{p^k}\mathrel{\hat{=}}\#X(\mathbb{F}_{p^k})$, the number of points on the variety X over the finite field with p^k elements. Recall that

$$\mathbb{F}_{p^k}\mathrel{\hat{=}}\mathbb{F}_p[s]/(f), \tag{5.22}$$

where s is a formal variable, $f\in\mathbb{F}_p[s]$ is an irreducible polynomial of degree k, and $\mathbb{F}_p$ is just $\mathbb{Z}/p\mathbb{Z}$. The number of points on the variety, which for us is the algebraic curve $A(m,l)=0$ or one of its close relatives, is simply the number of solutions to the defining equation in the projective space $\mathbb{P}^2(\mathbb{F}_{p^k})$.

A well-known theorem of Weil states that the generating function (5.21) can always be resummed into a simple rational form. In particular, when the variety X is a nonsingular planar algebraic curve, the local zeta function can be written as

$$Z_p(t) = \frac{n_p(t)}{(1-t)(1-pt)}, \tag{5.23}$$

where $n_p(t) = (1-\alpha_1 t)(1-\alpha_2 t)\cdots(1-\alpha_{2g}t)$ is a polynomial of degree $2g$ (g being the genus of the curve), and the α_i are algebraic integers with various properties. This form essentially follows from the existence of a Weil cohomology theory with various good properties, including a Lefschetz trace formula. The form of the argument is as follows: Fix an algebraic closure $\bar{\mathbb{F}}_p$ of $\mathbb{F}_p$. The Frobenius endomorphism $\mathrm{Fr} : x \mapsto x^p$ is a morphism of fields in characteristic p, and we can identify

$$\mathbb{F}_{p^k} \cong \mathrm{Fix}(\mathrm{Fr}^k) \subset \bar{\mathbb{F}}_p, \quad X(\mathbb{F}_{p^k}) = \mathrm{Fix}\left(\mathrm{Fr}^k : X(\bar{\mathbb{F}}_p) \to X(\bar{\mathbb{F}}_p)\right). \tag{5.24}$$

Letting $\bar{X}$ denote $X(\bar{\mathbb{F}}_p)$, the Lefschetz fixed point formula then schematically reads

$$\#X(\mathbb{F}_{p^k}) = \#\,\mathrm{Fix}(\mathrm{Fr}^k : \bar{X} \to \bar{X}) = \sum_i (-1)^i \,\mathrm{tr}\,\mathrm{Fr}^k\,|H^i(\bar{X}). \tag{5.25}$$

We can then use this in the definition of the zeta function to see that

$$Z_p(X,t) = \prod_{i=0}^{2\dim X} \det\left(1 - t\,\mathrm{Fr}\,|H^i(\bar{X})\right)^{(-1)^{i+1}}. \tag{5.26}$$

This is the expression that reduces to the form (5.23) when X is a complex curve, so that $\dim X = 1$. The interesting piece of the zeta function (its numerator) comes from the action of the Frobenius map on $H^1(\bar{X})$.

In the case where the curve is singular, Weil's theorem does not apply, but it is a result of Aubry and Perret [10] that the zeta function still takes the form (5.23). The only change is to the form of the numerator, which is no longer as simple. However, it changes in a controlled way. Let $\nu : \tilde{X} \to X$ be a resolution of singularities of the curve X over $\mathbb{F}_p$. (That is, $\tilde{X}$ is a nonsingular curve, and ν is an isomorphism away from the singular locus of X.) Then

$$n_p(X;t) = n_p(\tilde{X};t)\cdot\xi(t), \tag{5.27}$$

where $\xi(t)$ is a product of cyclotomic factors.

As we have noted before, A-polynomial curves for hyperbolic knots are always singular algebraic varieties. As such, we will now review some notions connected to resolution of singularities that will be important in our analysis of the figure-eight knot example. In particular, we will need to consider a resolution of singularities of our curve in order to compute its Hasse-Weil zeta function, and identify the corresponding modular form.

Our chief example is the A-polynomial for the figure-eight knot, which we write in the form

$$A(m,l) = -m^4 + (1 - m^2 - 2m^4 - m^6 + m^8)l - m^4 l^2. \tag{5.28}$$

We use the notation $C : A(m,l) = 0$ for the zero locus. We will sometimes use the shorthand $p(m)$ for the coefficient of l in the above expression, viewed as an element of $\mathbb{Z}[m]$. That is, $p(m) = 1 - m^2 - 2m^4 - m^6 + m^8$. As above, by an abuse of notation, $p(x)$ will denote the same polynomial after the change of variables $x = m^2$.

To pass to a projective completion of this curve, we can view it as living in the subset

$$\mathbb{C}^\star \times \mathbb{C}^\star = \{[m : l : s] | m \neq 0, l \neq 0, s = 1\} \subset \mathbb{P}^2 \tag{5.29}$$

of two dimensional projective space. This is the embedding considered in [54]. The closure $\bar{C}$ of $C \subset \mathbb{P}^2$ is then the zero locus of the homogeneous polynomial

$$\bar{A}(m,l,s) = -m^4 s^5 + (s^8 - m^2 s^6 - 2m^4 s^4 - m^6 s^2 + m^8)l - m^4 l^2 s^3. \tag{5.30}$$

The affine chart $s = 1$ returns the ordinary A-polynomial.

This curve is singular: that is, there exist simultaneous solutions to the equations

$$\bar{A}(m,l,s) = 0, \quad \partial_m \bar{A} = \partial_l \bar{A} = \partial_s \bar{A} = 0$$

for $[m : l : s] \in \mathbb{P}^2$. By Euler's theorem on homogeneous functions, the last three of these equations imply the first, and (for instance) the equation $\partial_m \bar{A} = 0$ follows from the other three at all points with $m \neq 0$. We compute all the singular points explicitly as follows:

Case 1: $s \neq 0$**.** Work in the chart $s = 1$. Then our three equations are as follows:

$$\begin{aligned} 0 &= -m^4 + p(m)l - m^4 l^2, \\ 0 &= -4m^3(1 + l^2) + (8m^7 - 6m^5 - 8m^4 - 2m)l, \\ 0 &= p(m) - 2m^4 l. \end{aligned}$$

The first and third of these equations combine to give $m^4(l^2 - 1) = 0$. $m = 0$ cannot be singular since the third equation is not satisfied, so the only singular points in this chart occur for $l = \pm 1$. The third equation can then be factored, giving a list of possible m-values for each l, which can then be checked against the second equation. The results are the universal singular points $(m, l) = (\pm 1, -1)$ and $(\pm i, +1)$.

Case 2: $s = 0$**.** It is automatic in this case that $\partial_s A = 0$, since there is no monomial of s-degree one in $\bar{A}$. The other equations become $m^8 l = 0$, $8m^7 l = 0$, $m^8 = 0$. The unique singular point in this case is therefore $[0 : 1 : 0]$.

Since it is defined by a polynomial with integral coefficients, the curve C can naturally be studied over finite fields by reduction modulo p. However, C in the form (5.30) does not appear to be the easiest or best choice for this kind of study. It is desirable to have a nonsingular model of C. However, since C has singular points whose coordinates are not integral, methods such as the blow-up may give a new polynomial that, while having fewer singularities, is no longer integral.

Therefore, we choose instead to look at a related curve E, which is the projective closure of the zero locus of the reduced A-polynomial:

$$A(x, y) = -x^2 + (1 - x - 2x^2 - x^3 + x^4)y - x^2 y^2. \tag{5.31}$$

This curve is still singular. There are the two universal singular points at $(x, y) = (-1, 1)$ and $(1, -1)$, which are nodal of order two. In addition, the point at infinity $[0 : 1 : 0]$ is a singularity of order three. The well-known genus-degree formula,

$$g(E) = \frac{1}{2}(d - 1)(d - 2) - \sum_{p \in E} \frac{1}{2} r_p (r_p - 1), \tag{5.32}$$

then shows immediately that E is of genus one.

We note that changing variables to pass to the reduced form of the A-polynomial, although it gives a two-to-one map on $\mathbb{C}^\star \times \mathbb{C}^\star \subset \mathbb{P}^2$, does not extend to a two-to-one map from $\mathbb{P}^2$ to itself when we embed E in projective space in the same way we did for C. The map has problems at $[0:0:1]$, as well as for points at infinity. The A-polynomial curves avoid the origin, so there are no problems at finite distance. However, on both $\bar{C}$ and $\bar{E}$, there are precisely two points at infinity: $[1:0:0]$ and $[0:1:0]$. The map from $\bar{C}$ to $\bar{E}$ induced by change of variables thus fails to be two-to-one on this locus. Moreover, the map is not birational.

As is well known, every nonsingular elliptic curve can be defined by a minimal equation of the form

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6, \tag{5.33}$$

where a_i are constants. Borot and Eynard [25, §6.1.5] give the minimal equation for a nonsingular model of our elliptic curve E:

$$y^2 + xy + y = x^3 + x^2. \tag{5.34}$$

The variables x and y in (5.34) are obtained from our variables x and y in (5.31) by birational transformations.

We can study reduction modulo p of the curve (5.34) and compute its zeta-function by direct point counting. Our computational methodology is as follows. The local zeta function for a planar algebraic curve (singular or not) takes the general form (5.23). Let us use the notation $n_p(t) = \beta_0 + \beta_1 t + \cdots$ for the numerator. Ignoring questions of convergence, we can match terms order by order in t between the expressions (5.21) and (5.23) to solve for the coefficients β_i from the known numbers N_{p^k}, computed numerically for small p and k. For a fixed p, knowledge of N_{p^k} for all $k \leq n$ allows us to compute β_i for $i \leq n$.

For the nonsingular minimal model of the curve associated to the figure-eight knot, the results are tabulated in Table 5.1. The reader will observe that the polynomial $n_p(t)$ here has degree two, as it should for a nonsingular genus-one curve.

Having done this, we can proceed to compute the modular form associated to this elliptic curve. One way to state the association is as follows: To each

Prime p	*Zeta numerator* $n_p(t)$
2	$1 + t + 2t^2$
3	$1 + t$ (bad reduction)
5	$1 - t$ (bad reduction)
7	$1 + 7t^2$
11	$1 + 4t + 11t^2$
13	$1 + 2t + 13t^2$
17	$1 - 2t + 17t^2$
19	$1 - 4t + 19t^2$
23	$1 + 23t^2$
29	$1 + 2t + 29t^2$
31	$1 + 31t^2$
37	$1 + 10t + 37t^2$
41	$1 - 10t + 41t^2$
43	$1 - 4t + 43t^2$
47	$1 - 8t + 47t^2$
53	$1 + 10t + 53t^2$
59	$1 + 4t + 59t^2$

Table 5.1: Zeta function numerators for the nonsingular elliptic curve (5.34), computed directly by counting points.

elliptic curve E is associated an integer N called the *conductor*, which is roughly speaking the product of all primes in a minimal set of primes of bad reduction for the curve. This integer then defines a congruence subgroup of the modular group,

$$\Gamma_0(N) = \{\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{SL}(2,\mathbb{Z}) : c \equiv 0 \pmod{N}\}.$$

For each prime p, the point-counting data over all finite fields of characteristic p reduce to just one number: the middle coefficient of the numerator of the zeta function, which we denote $-a_p(E)$. Thus $a_p(E) = p+1-\#E(\mathbb{F}_p)$. Modularity essentially states that, after we extend this sequence multiplicatively in a well-defined way to a sequence $\{a_n\}$ defined for all natural numbers, the Fourier series

$$f = \sum_{n>0} a_n e^{2\pi i n\tau}$$

defines a *modular form* $f(\tau)$ of weight 2 on the upper half plane, for the congruence subgroup $\Gamma_0(N)$. This modular form is a cusp form. Moreover, it is a simultaneous eigenvector for all the Hecke operators that act on modular forms: in particular,

$$T_p f = a_p f,$$

where T_p is a set of Hecke operators that are defined for all prime p.

We can always identify the eigenform associated to a given curve by checking a finite number of coefficients in its Fourier expansion. This is because, for all N, the space $\mathscr{S}_2(\Gamma_0(N))$ of weight-2 cusp forms at level N is finite-dimensional. In our case, the conductor of curve (5.34) is 15, and the vector space $\mathscr{S}_2(\Gamma_0(15))$ turns out to be one-dimensional. Therefore, the modular form associated to our curve is the unique (up to normalization) weight-2 cusp form at level 15:

$$f = q - q^2 - q^3 - q^4 + q^5 + q^6 + 3q^8 + q^9 - q^{10} - 4q^{11} + \cdots . \tag{5.35}$$

The reader can easily see the correspondence between the coefficients in (5.35) at prime powers of q and the first few coefficients occurring in the numerators tabulated in Table 5.1.

5.4 Modularity and the q-Pochhammer symbol

The q-Pochhammer symbol is defined to be the product

$$(a;q)_n \mathrel{\hat{=}} \prod_{0\le k<n} (1 - aq^k). \tag{5.36}$$

When the variable q is in the open unit disc in $\mathbb{C}$, this expression makes sense for $n = \infty$. In particular, we can consider the special value $a = q$, for which

$$(q;q)_\infty = q^{-1/24}\eta(\tau) = \prod_{k>0}(1 - q^k). \tag{5.37}$$

This equation defines $\eta(\tau)$, the Dedekind eta function. (We take $\tau \in \mathbb{H}$ to be related to q by the conformal transformation $q = e^{2\pi i\tau}$.) The eta function obeys the well-known functional equation

$$\eta(-1/\tau) = \sqrt{-i\tau}\eta(\tau), \tag{5.38}$$

which, together with the obvious invariance $\eta(\tau + 1) = \eta(\tau)$, shows that η transforms with weight $1/2$ under the action of the modular group on the upper half plane (up to a twist by a phase). In particular, the function $\eta(\tau)^{24}$ is (up to a constant multiple) the discriminant function, a modular form of weight 12. We refer the reader to [60] for more details. Many other modular forms can be constructed as products of eta functions: for instance, the function

$$F(q) = q\prod_{k>0}(1 - q^k)^2(1 - q^{11k})^2 = \eta(\tau)^2\eta(11\tau)^2$$

is a modular form of weight two for the congruence subgroup $\Gamma_0(11)$ [187].

One might ask whether eta functions generate all spaces of modular forms this way. Rouse and Webb [179] consider "eta-quotients," which are functions on $\mathbb{H}$ of the form

$$f(\tau) = \prod_{d|N} \eta(d\tau)^{r_d}, \tag{5.39}$$

where r_d are integers. We have the following theorem:

Theorem 8 (M. Newman, cited in [179])**.** Suppose that the following conditions are satisfied:

- $\sum_{d|N} d r_d \equiv 0 \pmod{24}$,
- $\sum_{d|N} (N/d) r_d \equiv 0 \pmod{24}$,
- $\prod_{d|N} d^{r_d}$ is the square of a rational number.

Then the eta-quotient (5.39) defined by the r_d is a modular form for $\Gamma_0(N)$, of weight

$$k = \frac{1}{2} \sum_{d|N} r_d.$$

It is straightforward to check that there is a modular eta-quotient of weight two at level 15. Indeed, this is exactly the modular form (5.35) associated to the A-polynomial for the figure eight knot:

$$f(\tau) = \eta(\tau)\eta(3\tau)\eta(5\tau)\eta(15\tau) = q(q;q)_\infty (3q;3q)_\infty (5q;5q)_\infty (15q;15q)_\infty. \tag{5.40}$$

It would be interesting to find some relationship between (5.40) and some more "quantum" occurrence of q-Pochhammer symbols, for instance in closed-form expressions for colored Jones polynomials. However, since modularity does not generalize straightforwardly to curves of higher genus, it is difficult to see how this could come from a story that makes sense for all knots. Moreover, there does not appear to be an obvious relation with the exact expression for the (normalized) colored Jones polynomial for the figure eight,

$$J_N(4_1;q) = \sum_{j=0}^{N-1} \prod_{k=1}^{j} \left(q^{\frac{N-k}{2}} - q^{-\frac{N-k}{2}} \right) \left(q^{\frac{N+k}{2}} - q^{-\frac{N+k}{2}} \right). \tag{5.41}$$

5.5 Locally blowing up singularities

In this section, we explicitly show how the resolution of singularities on the curve $\bar{C}$ can be used to understand the extra cyclotomic factors occurring in its zeta function when compared to that for its nonsingular model $\bar{E}$. We begin by reviewing the essential idea of the *blowup* construction.

We want to come up with a general construction that will remove, or improve, the singularities that occur when the tangent line to our curve at a point is not unique. We would like to interpret these points as superpositions of *different* points on a nonsingular curve; that is, self-intersections introduced by projection. To do this, we need a way of picking apart or distinguishing between the two components that lie on top of each other at such points.

Suppose the point we want to blow up is the origin of the affine plane $\mathbb{A}^2$ (over an arbitrary field). We can consider the natural map, defined on $\mathbb{A}^2 \setminus (0,0)$, which takes a point in the plane to the line through that point and the origin, viewed as an element of $\mathbb{P}^1$. Explicitly, the map is

$$f : \mathbb{A}^2 \to \mathbb{P}^1, \quad (x, y) \mapsto [x : y],$$

which is defined while x and y are not both zero.

We can consider the graph of this mapping as a subset of $\mathbb{A}^2 \times \mathbb{P}^1$. Obviously, the graph is isomorphic to $\mathbb{A}^2 \setminus (0,0)$. Its closure is the algebraic set

$$B : xv - yu = 0,$$

where (x, y), $[u : v]$ are coordinates on $\mathbb{A}^2$ and $\mathbb{P}^1$, respectively. B contains the graph of f as the locus of points away from $(x, y) = (0, 0)$, for which projection on the first factor is an isomorphism. However, the inverse image of $(0, 0)$ under this projection is an entire copy of $\mathbb{P}^1$, which can be thought of as labeling the distinct tangent lines through the origin. When we lift a curve C passing through the origin to B, we will obtain a curve in B, together with the entire $\mathbb{P}^1$ lying above the origin (which is termed the *exceptional divisor*). The irreducible component of this algebraic set which is distinct from the exceptional divisor is the blow-up of the curve. It will intersect the exceptional divisor in a collection of points which correspond to the distinct tangent directions of the curve at the origin, so that (for instance) a nodal singular point will lift to two distinct nonsingular points on the blow-up. One can imagine that, while $f(0, 0)$ is not defined, the limit of $f(p)$ for a sequence

of points $p \in C$ tending to the origin[§] is defined, and is the point in $\mathbb{P}^1$ corresponding to the tangent line to C at $(0,0)$. By choosing an affine chart on $\mathbb{P}^1$, we can view the blowup as a lift of our curve to $\mathbb{A}^3$.

To be completely explicit, let's choose the chart on $\mathbb{P}^1$ where $u \neq 0$. Then we can think of our map f from above as the map

$$\mathbb{A}^2 \to \mathbb{A}^1, \quad (x,y) \mapsto y/x,$$

which is defined away from the locus $x = 0$. The closure of the graph of this map in $\mathbb{A}^3$ is the algebraic hypersurface

$$B_a = \{(x,y,z) \in \mathbb{A}^3 : y = xz\}.$$

The "blowdown map" is the projection

$$\pi : B_a \subset \mathbb{A}^3 \to \mathbb{A}^2, \quad (x,y,z) \mapsto (x,y).$$

This is an isomorphism away from $x = y = 0$, while it maps the entire z-axis (exceptional divisor) to the blown-up point $(0,0)$.

Now, let φ be the embedding

$$\varphi : \mathbb{A}^2 \hookrightarrow \mathbb{A}^3, \quad (x,z) \mapsto (x,xz,z).$$

This is obviously injective, and its image is contained in B_a. Indeed, $\operatorname{im}\varphi$ is exactly B_a, so that φ is an isomorphism of $\mathbb{A}^2$ with $B_a \subset \mathbb{A}^3$.

We consider the map $\psi = \pi \circ \varphi$ from the affine plane to itself. That is, we first embed the affine plane as the hypersurface $B_a \subset \mathbb{A}^3$, and then project back to $\mathbb{A}^2$ with the blowdown map. Explicitly,

$$\psi(x,z) = (x,xz).$$

ψ is certainly not an isomorphism, since its image does not contain any point of the form $(0,y)$ with $y \neq 0$. However, the restriction of ψ to the open set $U = \{(x,z) : z \neq 0\} \subset \mathbb{A}^2$ is an isomorphism of U with itself. Therefore ψ is a birational equivalence of the plane with itself. One can think of it as mapping the collection of *horizontal* lines to the collection of lines through the origin, collapsing the entire vertical axis to a point.

[§]In the singular case, these points must be chosen properly, so as to all lie in the same "direction away" from the origin.

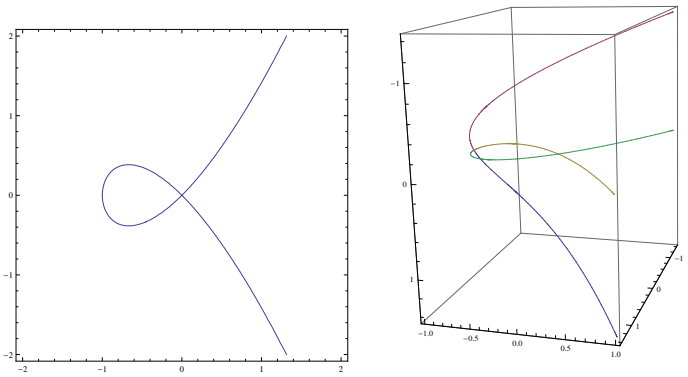


Figure 5.3: The singular algebraic curve $y^2 = x^2(x+1)$, together with its blowup in $\mathbb{A}^3$. At right, π is vertical projection and returns the original curve C, while ψ^{-1} is projection into the plane of the paper and returns the nonsingular curve C', a simple parabola.

The inverse image $\psi^{-1}(C)$ of a curve will be an algebraic set whose irreducible components are a new curve C', together with one or more[¶] copies of the exceptional divisor (z-axis). The curve C' is birationally equivalent to C, and the map $\psi : C' \to C$ gives a resolution of (sufficiently simple) singularities of C at the origin.

Example 13. Consider the curve

$$C : y^2 - x^2(x+1) = 0$$

in the affine plane. It has a nodal singularity, with two distinct tangent lines at the origin (see figure 5.3). We calculate the inverse image by making the coordinate substitution $y = xz$, obtaining the variety

$$x^2(z^2 - x - 1) = 0.$$

As we expect, two copies of the exceptional divisor $x = 0$ are present, as well as the curve

$$C' : z^2 = x + 1,$$

which is nonsingular.

We would like to use the blowup to study the relation of point counting on the singular curve E defined by (5.31) to that on its nonsingular model $\tilde{E}$,

[¶]The multiplicity of the exceptional divisor will correspond to the number of distinct tangent lines C has at the origin—strictly greater than 1 if this is a singular point.

defined by (5.34). One thing we can do to begin our analysis is simply count points on both curves separately, and compare the outcomes to check how the singularities affect these numbers. Results of this study are shown in Table 5.2. As can be seen, $\tilde{E}$ always has a nonnegative number of extra points compared to E; this makes sense if we recall that the singular points of E are self-intersections or ramification points of the normalization map $\nu : \tilde{E} \to E$, so that several points on $\tilde{E}$ may only count as one point on E. It is important to note that this intuition may not always be valid if we are working over a field that is not algebraically closed.

The number of extra points $e(p, k)$ seems to always take the form

$$e(p, k) = e_1(p) + (-1)^k. \tag{5.42}$$

That is, $e(p, k) = e_1(p) - 1$ for odd k and $e_1(p) + 1$ for even k, where $e_1(p) \geq 1$. We will see that there is a simple way of understanding this regularity.

Recalling the definition (5.21) of the zeta function, and the Taylor series expansion

$$\sum_k \frac{c^k t^k}{k} = -\log(1 - ct),$$

we can see that the zeta function numerators for E have the correct Aubry-Perret singular form. In particular, assuming the form (5.42) for the difference in numbers of points, we have

$$N_{p,k} = \tilde{N}_{p,k} - e(p, k) = \tilde{N}_{p,k} - e_1(p) - (-1)^k,$$

and therefore

$$Z_p(E; t) = Z_p(\tilde{E}; t)(1 + t)(1 - t)^{e_1(p)}. \tag{5.43}$$

This specifies the extra cyclotomic factors in the numerator (at least for the first five values of p shown in Table 5.2), which take a particularly simple form.

While one could in principle use a sequence of repeated blowups to construct an explicit normalization map to E from a nonsingular curve, we prefer to use the following trick (for which we know of no explicit reference in the literature, and which we will not justify rigorously except by observing that it works correctly). Since the blow-up is a local construction [195], we should be able to work in an affine chart, treat each singular point separately, and count points lying above the singular point on the blown-up curve to reproduce

p^k	$\#E(\mathbb{F}_{p^k})$	$\#\tilde{E}(\mathbb{F}_{p^k})$
2^1	4	4
2^2	6	8
2^3	4	4
2^4	14	16
2^5	44	44
2^6	54	56
2^7	116	116
3^1	5	5
3^2	7	9
3^3	29	29
3^4	79	81
3^5	245	245
5^1	5	5
5^2	23	25
5^3	125	125
5^4	623	625
5^5	3125	3125
7^1	7	8
7^2	61	64
7^3	343	344
7^4	2301	2304
11^1	15	16
11^2	125	128
11^3	1263	1264
11^4	14845	14848

Table 5.2: Direct comparison of point-counting data for the original A-polynomial in reduced variables and the minimal model of Borot-Eynard.

the above results on "extra" points. (This basically amounts to counting the distinct tangent directions through each singular point.)

Let's move to the affine chart $y = 1$ for our curve E, since all singular points are at finite distance in this chart. The resulting polynomial is

$$-x^2z^3 + (z^4 - xz^3 - 2x^2z^2 - x^3z + x^4) - x^2z.$$

The singular points in this chart are $(x, z) = (0, 0)$, $(-1, 1)$, and $(-1, -1)$. Let w denote the new coordinate introduced in the blow-up. If we blow up the origin, two points sit above: $w = 0$ and $w = \infty$ (recall that, properly speaking, w is a coordinate on $\mathbb{P}^1$). Blowing up $(-1, 1)$ gives us the equation $5 + 10w + 4w^2 = 0$ for points lying above the singular point, and similarly blowing up $(-1, -1)$ leaves us with the equation $3 - 6w + 4w^2 = 0$. Our

hope is that studying these equations modulo p will allow us to reproduce the overcounting data from above.

Over $\mathbb{F}_3$, the first of our quadratic polynomials becomes $w^2 + w - 1$, which is irreducible and has no points over $\mathbb{F}_3$. The second becomes $w^2 = 0$, which has the unique solution $w = 0$. Together with the two points lying over $(0, 0)$, this suggests that the counts on E and $\tilde{E}$ should be equal.

Over $\mathbb{F}_5$, the first polynomial becomes $-w^2 = 0$, which has just one point. The second becomes $3 - w - w^2 = 0$, which has no solutions in $\mathbb{F}_5$. Again, we expect equality between point-counting on E and $\tilde{E}$, which is correct.

Over $\mathbb{F}_7$, the first polynomial becomes $5 + 3w + 4w^2$, which has no roots in $\mathbb{F}_7$. The second polynomial becomes $3 + w + 4w^2$, which has the two distinct roots $w = 2$ and $w = 3$. Together with the two points over the origin, we expect one extra point on $\tilde{E}$, which is indeed the case.

Over $\mathbb{F}_{11}$, the polynomials are $5 - w + 4w^2$ and $3 + 5w + 4w^2$. The first has the two roots $w = 6$ and $w = 8$. The second has no roots in $\mathbb{F}_{11}$. Once again, we expect a total of one extra point, just as we should.

The periodicity-two behavior in k we noted above should be related to the fact that the equations here are quadratic, and so two extra points appear when a quadratic that is irreducible over $\mathbb{F}_p$ splits in a quadratic extension field. Recalling that there is a unique finite field (up to isomorphism) of each prime power cardinality, it is evident that $\mathbb{F}_{p^2}$ is a splitting field for any irreducible quadratic over $\mathbb{F}_p$.

If, indeed, we can always track the "extra" points of $\tilde{E}$ locally by looking for roots of our blow-up polynomials, we then have a good explanation for the restricted form of the number of extra points $e(p, k)$ in its dependence on p and k. Indeed, we expect that

$$e(p, 1) \in \{-1, 0, 1, 2, 3\},$$

since we are replacing our three singular points with two points (lying over the origin) together with two more points for each of our two quadratics that splits in $\mathbb{F}_p$ (or one point for a quadratic that reduces to a perfect square). The discriminants of our polynomials are 20 and -12 respectively, so we do not expect the last case unless $p = 2, 3, 5$. Thus, for larger p, we anticipate $e(p, 1) \in \{-1, 1, 3\}$. Moreover, if neither polynomial reduces to a perfect

square, then we anticipate $e(p,k) = 3$ for all even k, since both polynomials should split in this case. Thus there would seem to be three possibilities for the extra data $e(p,k)$, and we can discriminate between these cases by just checking $e(p,1)$. $p = 7$ and $p = 11$, as we have seen, have $e(p,1) = 1$. Are the other cases ever realized?

The answer is yes. As can be easily checked, neither of our blowup polynomials has a root over $\mathbb{F}_{17}$. Thus we expect $e(17,1) = -1$, and this is in fact the case.

Moreover, working over $\mathbb{F}_{19}$, both of our blowup polynomials split (they have roots $\{1,6\}$ and $\{14,16\}$, respectively), and so we expect to find $e(19,1) = 3$ (in fact, $e(19,k) = 3$ for all k). Direct calculation verifies this at $k = 1$ and 2.

How do each of these three cases appear in the zeta function numerator? We can write the extra data in the form

$$e(p,k) = \begin{cases} 1^k + (-1)^k + (-1)^k, & e(p,1) = -1, \\ 1^k + 1^k + (-1)^k, & e(p,1) = 1, \\ 1^k + 1^k + 1^k, & e(p,1) = 3. \end{cases} \tag{5.44}$$

Here each term can be thought of as representing one of our singular points. The singularity contributes 1^k if its blowup polynomial splits in $\mathbb{F}_p$, so that two points always lie above the singular point on the resolution. It contributes $(-1)^k$ if its blowup polynomial is irreducible in $\mathbb{F}_p$, so that no points lie above the singular point for odd k. The origin always contributes 1^k in this case, since there are always two rational points that lie above it. (For the special values of p where a blowup polynomial has a repeated root in $\mathbb{F}_p$, that singularity contributes 0^k.)

Therefore, for all primes $p \geq 7$, the singular zeta function should take the form

$$Z_p(E;t) = Z_p(\tilde{E};t)(1+t)^a(1-t)^{3-a}, \tag{5.45}$$

where $a \in \{0,1,2\}$ is the number of blowup polynomials that are irreducible over $\mathbb{F}_p$. Thus, $e(p,1) = 3-2a$. If this reasoning is correct, the only cyclotomic polynomials that occur are for first and second roots of unity because the singularities that need to be resolved are nodal ($r = 2$), and therefore have blowup polynomials which are quadratic. Other cyclotomic factors should correspond to singularities of higher degree.

Knot	Genus of $A(m,l)$	Product of bad primes
4_1	1	$3 \cdot 5$
5_2	2	$7 \cdot 23$
6_1	3	$3 \cdot 257$
6_2	4	$11 \cdot 1777$
6_3	7	$3 \cdot 5 \cdot 11 \cdot 13 \cdot 31$
7_2	4	$11 \cdot 4409$
7_3	5	$13 \cdot 78301$
7_4 (1)	1	$3 \cdot 5$
7_4 (2)	2	$5 \cdot 59$
7_5	9	$11 \cdot 17 \cdot 443 \cdot 10589$
7_6	12	$19 \cdot 139 \cdot 1091 \cdot 649787$
7_7	72	?
8_1	5	$13 \cdot 92051$
8_{20}	5	$2 \cdot 3 \cdot 5 \cdot 733$

Table 5.3: Primes of bad reduction for the A-polynomial curves listed by Cooper et al. [54].

5.6 Conductors of A-polynomial curves

We close with another puzzling empirical observation about the arithmetic structure of A-polynomials. The table below shows a list of primes of bad reduction for these curves. The product of these primes will correspond to the radical of the conductor. For a curve of genus g, the conductor N is bounded from below by 11^g (for instance, 11 is the smallest integer occurring as the conductor of an elliptic curve).

For these curves, it seems that the conductor is always relatively small. Indeed, for some knots (such as 8_{20}) the product of all bad primes is actually *less* than 11^g. There is no contradiction here, since primes may occur in the conductor with multiplicity more than one.

AFTERWORD

In the sea, Biscayne, there prinks
The young emerald, evening star,
Good light for drunkards, poets, widows,
And ladies soon to be married.

By this light the salty fishes
Arch in the sea like tree-branches,
Going in many directions
Up and down.

This light conducts
The thoughts of drunkards, the feelings
Of widows and trembling ladies,
The movements of fishes.

How pleasant an existence it is
That this emerald charms philosophers,
Until they become thoughtlessly willing
To bathe their hearts in later moonlight,

Knowing that they can bring back thought
In the night that is still to be silent,
Reflecting this thing and that,
Before they sleep!

It is better that, as scholars,
They should think hard in the dark cuffs
Of voluminous cloaks,
And shave their heads and bodies.

It might well be that their mistress
Is no gaunt fugitive phantom.
She might, after all, be a wanton,
Abundantly beautiful, eager,

Fecund,
From whose being by starlight, on sea-coast,
The innermost good of their seeking
Might come in the simplest of speech.

It is a good light, then, for those
That know the ultimate Plato,
Tranquillizing with this jewel
The torments of confusion.

—Wallace Stevens, "Homunculus et la Belle Étoile"

BIBLIOGRAPHY

[1]M. Aganagic and S. Shakirov. "Knot homology and refined Chern-Simons index." Commun. Math. Phys. **333**, 187–228. (2015). arXiv:1105.5117 [hep-th].

[2]M. Aganagic and C. Vafa. "Large N duality, mirror symmetry, and a Q-deformed A-polynomial for knots." (2012). arXiv:1204.4709 [hep-th].

[3]A. S. Alexandrov, A. Mironov, and A. Morozov. "M-theory of matrix models." Teor. Mat. Fiz. **150**, 179–192. (2007). arXiv:hep-th/0605171.

[4]A. Almheiri, X. Dong, and D. Harlow. "Bulk locality and quantum error correction in AdS/CFT." Journal of High Energy Physics **2015**, 1–34. (2015).

[5]L. Álvarez-Gaumé. "Supersymmetry and the Atiyah–Singer index theorem." Communications in Mathematical Physics **90**, 161–173. (1983).

[6]M. Atiyah and R. Bott. "The Yang-Mills equations over Riemann surfaces." Phil. Trans. R. Soc. London A **308**, 523–615. (1982).

[7]M. Atiyah. "New invariants of 3- and 4-dimensional manifolds." Proc. Symp. Pure Math. **48**, 285–299. (1988).

[8]M. Atiyah. "Topological quantum field theories." Publications Mathématiques de l'Institut des Hautes Études Scientifiques **68**, 175–186. (1988).

[9]M. Atiyah. *The geometry and physics of knots*. (Cambridge University Press. 1990).

[10]Y. Aubry and M. Perret. "A Weil theorem for singular curves." In *Arithmetic, geometry and coding theory*. Edited by R. Pellikaan, M. Perret, and S. G. Vladut. (de Gruyter. Berlin–New York. 1996).

[11]M. Bañados, C. Teitelboim, and J. Zanelli. "Black hole in three-dimensional spacetime." Physical Review Letters **69**, 1849. (1992).

[12]N. Bao, S. Nezami, H. Ooguri, B. Stoica, J. Sully, and M. Walter. "The holographic entropy cone." (2015). arXiv:1505.07839 [hep-th].

[13]N. Bao et al. "Consistency conditions for an AdS multiscale entanglement renormalization ansatz correspondence." Physical Review D **91**, 125036. (2015).

[14]J. Barge. "Structures différentiables sur les types d'homotopie rationnelle simplement connexes." Annales Scientifiques de l'École Normale Supérieure. 4th ser. **9**, 469–501. (1976).

[15]D. Bar-Natan. "On Khovanov's categorification of the Jones polynomial." Algebraic & Geometric Topology **2.1**, 337–370. (2002). arXiv:math/0201043 [math.GT].

[16]F. Benini, R. Eager, K. Hori, and Y. Tachikawa. "Elliptic genera of two-dimensional $\mathcal{N} = 2$ gauge theories with rank-one gauge groups." Lett. Math. Phys. **104**, 465–493. (2014). arXiv:1305.0533 [hep-th].

[17]N. Berkovits and C. Vafa. "$\mathcal{N} = 4$ topological strings." Nuclear Physics B **433**, 123–180. (1995). arXiv:hep-th/9407190 [hep-th].

[18]M. Bertolini, I. V. Melnikov, and M. R. Plesser. "Hybrid conformal field theories." JHEP **05**, 043. (2014). arXiv:1307.7063 [hep-th].

[19]A. Kh. Bikulov and A. P. Zubarev. "On one real basis for $L^2(\mathbb{Q}_p)$." (2015). arXiv:1504.03624.

[20]D. Birmingham, M. Blau, M. Rakowski, and G. Thompson. "Topological field theory." Phys. Rept. **209**, 129–340. (1991).

[21]D. Birmingham, C. Kennedy, S. Sen, and A. Wilkins. "Geometrical finiteness, holography, and the Bañados-Teitelboim-Zanelli black hole." Physical Review Letters **82**, 4164. (1999).

[22]M. Blau and G. Thompson. "Aspects of $\mathcal{N}_T \geq 2$ topological gauge theories and D-branes." Nuclear Physics B **492**, 545–590. (1997). arXiv:hep-th/9612143 [hep-th].

[23]J. de Boer and T. Tjin. "Quantization and representation theory of finite $\mathcal{W}$-algebras." Commun. Math. Phys. **158**, 485–516. (1993). arXiv:hep-th/9211109 [hep-th].

[24]F. A. Bogomolov. "On the decomposition of Kähler manifolds with trivial canonical class." Mathematics of the USSR-Sbornik **22**, 580–583. (1974).

[25]G. Borot and B. Eynard. "All-order asymptotics of hyperbolic knot invariants from non-perturbative topological recursion of A-polynomials." (2012). arXiv:1205.2261 [math-ph].

[26]R. Bott and L. Tu. *Differential Forms in Algebraic Topology*. Graduate Texts in Mathematics 82. (Springer Verlag, New York. 1982).

[27]R. Bott. "On some recent interactions between mathematics and physics." Canad. Math. Bull. **28**, 129–164. (1985).

[28]V. Bouchard, A. Klemm, M. Mariño, and S. Pasquetti. "Remodeling the B-model." Commun. Math. Phys. **287**, 117–178. (2009). arXiv:0709.1453 [hep-th].

[29]J.-F. Boutot and H. Carayol. "Uniformisation p-adique des courbes de Shimura: les théorèmes de Čerednik et de Drinfeld." Astérisque **196–97**, 45–158. (1991).

[30]P. Bouwknegt, J. G. McCarthy, and K. Pilch. "BRST analysis of physical states for 2D gravity coupled to $c \leq 1$ matter." Commun. Math. Phys. **145**, 541–560. (1992).

[31]R. Bowen. "Hausdorff dimension of quasi-circles." Publications Mathématiques de l'IHÉS **50**, 11–25. (1979).

[32]P. E. Bradley. "Cyclic coverings of the p-adic projective line by Mumford curves." manuscripta mathematica **124**, 77–95. (2007).

[33]L. Brekke and P. G. O. Freund. "p-adic numbers in physics." Physics Reports **233**, 1–66. (1993).

[34]L. Brekke, P. G. O. Freund, M. Olson, and E. Witten. "Non-archimedean string dynamics." Nuclear Physics B **302**, 365–402. (1988).

[35]C. Breuil, B. Conrad, F. Diamond, and R. Taylor. "On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises." Journal of the American Mathematical Society, 843–939. (2001).

[36]I. Brunner and D. Roggenkamp. "B-type defects in Landau-Ginzburg models." JHEP **0708**, 093. (2007). arXiv:0707.0922 [hep-th].

[37]I. Brunner, D. Roggenkamp, and S. Rossi. "Defect perturbations in Landau-Ginzburg models." JHEP **1003**, 015. (2010). arXiv:0909.0696 [hep-th].

[38]F. Cachazo, M. R. Douglas, N. Seiberg, and E. Witten. "Chiral rings and anomalies in supersymmetric gauge theory." JHEP **12**, 071. (2002). arXiv:hep-th/0211170 [hep-th].

[39]P. Calabrese and J. L. Cardy. "Entanglement entropy and quantum field theory." J. Stat. Mech. **0406**, P06002. (2004). arXiv:hep-th/0405152 [hep-th].

[40]P. Calabrese and J. L. Cardy. "Entanglement entropy and conformal field theory." J. Phys. **A42**, 504005. (2009). arXiv:0905.4013 [stat-mech].

[41]R. D. Canary. "Ends of hyperbolic 3-manifolds." Journal of the American Mathematical Society **6**, 1–35. (1993).

[42]A. Carey, M. Marcolli, and A. Rennie. "Modular index invariants of Mumford curves." (2009). arXiv:0905.3157.

[43]H. Casini and M. Huerta. "Universal terms for the entanglement entropy in $2+1$ dimensions." Nuclear Physics B **764**, 183–201. (2007). arXiv:hep-th/0606256 [hep-th].

[44]B. Casselman. "The Bruhat–Tits tree of SL(2)." available from personal website. 2016.

[45]L. O. Chekhov, A. D. Mironov, and A. V. Zabrodin. "Multiloop calculations in p-adic string theory and Bruhat–Tits trees." Commun. Math. Phys. **125**, 675. (1989).

[46]I. Cherednik. "Jones polynomials of torus knots via DAHA." International Mathematics Research Notices **2013**, 5366–5425. (2013). arXiv:1111.6195 [math.QA].

[47]N. Chriss and V. Ginzburg. *Representation theory and complex geometry.* (Birkhauser. 1997).

[48]S. Chun, S. Gukov, and D. Roggenkamp. "Junctions of surface operators and categorification of quantum groups." (2015). arXiv:1507.06318 [hep-th].

[49]H.-J. Chung, T. Dimofte, S. Gukov, and P. Sułkowski. "3d-3d correspondence revisited." (2014). arXiv:1405.3663 [hep-th].

[50]C. Consani and M. Marcolli. "Spectral triples from Mumford curves." International Mathematics Research Notices **2003**, 1945–1972. (2003).

[51]C. Consani and M. Marcolli. "New perspectives in Arakelov geometry." In Number theory. CRM Proceedings & Lecture Notes 36. (2004), pp. 81–102.

[52]C. Consani and M. Marcolli. "Noncommutative geometry, dynamics, and ∞-adic Arakelov geometry." Selecta Mathematica **10**, 167–251. (2004).

[53]B. Cooper and V. Krushkal. "Categorification of the Jones-Wenzl projectors." Quantum Topology **3**, 139–180. (2012). arXiv:1005.5117 [math.GT].

[54]D. Cooper, M. Culler, H. Gillet, D. D. Long, and P. B. Shalen. "Plane curves associated to character varieties of 3-manifolds." Inventiones Mathematicae **118**, 47–84. (1994).

[55]C. Cordova and D. L. Jafferis. "Complex Chern-Simons from M5-branes on the squashed three-sphere." (2013). arXiv:1305.2891 [hep-th].

[56]L. Crane and I. Frenkel. "Four-dimensional topological quantum field theory, Hopf categories, and the canonical bases." J. Math. Phys. **35**, 5136. (1994). arXiv:hep-th/9405183 [hep-th].

[57]A. S. Dancer. "Hyper-Kähler manifolds." In *Essays on Einstein manifolds*. Surveys in Differential Geometry 6. (International Press. 1998), pp. 15–38.

[58]P. Deligne, P. Griffiths, J. Morgan, and D. Sullivan. "Real homotopy theory of Kähler manifolds." Inventiones Mathematicae **29**, 245–274. (1975).

[59]D. E. Diaconescu, V. Shende, and C. Vafa. "Large N duality, lagrangian cycles, and algebraic knots." Commun. Math. Phys. **319**, 813–863. (2013). arXiv:1111.6533 [hep-th].

[60]F. Diamond and J. Shurman. *A first course in modular forms.* Graduate Texts in Mathematics 228. (Springer-Verlag. New York. 2005).

[61]R. Dijkgraaf, H. Fuji, and M. Manabe. "The volume conjecture, perturbative knot invariants, and recursion relations for topological strings." Nuclear Physics B **849**, 166–211. (2011). arXiv:1010.4542 [hep-th].

[62]T. Dimofte. "Quantum Riemann surfaces in Chern-Simons theory." Advances in Theoretical and Mathematical Physics **17**, 479–599. (2013). arXiv:1102.4847 [hep-th].

[63]T. Dimofte, M. Gabella, and A. B. Goncharov. "K-decompositions and 3d gauge theories." (2013). arXiv:1301.0192 [hep-th].

[64]T. Dimofte, D. Gaiotto, and S. Gukov. "3-Manifolds and 3d Indices." Advances in Theoretical and Mathematical Physics **17**, 975–1076. (2013). arXiv:1112.5179 [hep-th].

[65]T. Dimofte, D. Gaiotto, and S. Gukov. "Gauge theories labelled by three-manifolds." Commun. Math. Phys. **325**, 367–419. (2014). arXiv:1108.4389 [hep-th].

[66]T. Dimofte and S. Gukov. "Quantum field theory and the volume conjecture." Contemp. Math. **541**, 41–67. (2011). arXiv:1003.4808 [math.GT].

[67]T. Dimofte, S. Gukov, and L. Hollands. "Vortex counting and Lagrangian 3-manifolds." Lett. Math. Phys. **98**, 225–287. (2010). arXiv:1006.0977v1 [hep-th].

[68]T. Dimofte, S. Gukov, J. Lenells, and D. Zagier. "Exact results for perturbative Chern-Simons theory with complex gauge group." Comm. Num. Thy. and Phys. **3**, 363–443. (2009). arXiv:0903.2472v1 [hep-th].

[69]J. A. Dixon. "Calculation of BRS cohomology with spectral sequences." Commun. Math. Phys. **139**, 495–526. (1991).

[70]S. K. Donaldson. "An application of gauge theory to four-dimensional topology." Journal of Differential Geometry **18**, 279–315. (1983).

[71]B. Dragovich, A. Yu. Khrennikov, S. V. Kozyrev, and I. V. Volovich. "On p-adic mathematical physics." Anal. Appl. **1**, 1–17. (2009). arXiv:0904.4205 [math-ph].

[72]N. M. Dunfield, S. Gukov, and J. Rasmussen. "The superpolynomial for knot homologies." Experimental Mathematics **15**, 129–159. (2006). arXiv:math/0505662 [math.GT].

[73]F. J. Dyson. "Missed opportunities." Bull. Amer. Math. Soc. **78**, 635–639. (1972).

[74]T. Ekholm, J. Etnyre, L. Ng, and M. Sullivan. "Filtrations on the knot contact homology of transverse knots," Math. Ann. **355**, 1561–1591. (2013). arXiv:1010.0450 [math.SG].

[75]B. Eynard and N. Orantin. "Invariants of algebraic curves and topological expansion." (2007). arXiv:math-ph/0702045.

[76]T. Faulkner. "The entanglement Rényi entropies of disjoint intervals in AdS/CFT." (2013). arXiv:1303.7221 [hep-th].

[77]Y. Félix, J. Oprea, and D. Tanré. *Algebraic models in geometry*. Oxford Graduate Texts in Mathematics 17. (Oxford University Press. 2008).

[78]R. P. Feynman. "Space-time approach to non-relativistic quantum mechanics." Rev. Mod. Phys. **20**, 367–387. (1948).

[79]J. M. Figueroa-O'Farrill, C. Köhl, and B. J. Spence. "Supersymmetry and the cohomology of (hyper) Kähler manifolds." Nuclear Physics B **503**, 614–626. (1997). arXiv:hep-th/9705161 [hep-th].

[80]A. Floer. "An instanton invariant for 3-manifolds." Communications in Mathematical Physics **118**, 215–240. (1988).

[81]A. Floer. "Morse theory for Lagrangian intersections." Journal of Differential Geometry **28**, 513–547. (1988).

[82]A. Floer. "Witten's complex and infinite dimensional Morse theory." Journal of Differential Geometry **30**, 207–221. (1989).

[83]D. Freed. *1992 lectures on TQFT*. (1992) http://www.ma.utexas.edu/users/dafr.

[84]E. Frenkel, S. Gukov, and J. Teschner. "Surface operators and separation of variables." JHEP **01**, 179. (2016). arXiv:1506.07508 [hep-th].

[85]I. Frenkel, C. Stroppel, and J. Sussan. "Categorifying fractional Euler characteristics, Jones-Wenzl projectors and 3j-symbols." Quantum Topology **3**, 181–253. (2012). arXiv:1007.4680 [math.GT].

[86]P. G. O. Freund and M. Olson. "Non-archimedean strings." Physics Letters B **199**, 186–190. (1987).

[87]P. G. O. Freund and E. Witten. "Adelic string amplitudes." Physics Letters B **199**, 191–194. (1987).

[88]P. Freyd et al. "A new polynomial invariant of knots and links." Bull. Amer. Math. Soc. **12**, 239–246. (1985).

[89]H. Fuji, S. Gukov, M. Stošić, and P. Sułkowski. "3d analogs of Argyres-Douglas theories and knot homologies." JHEP **1301**, 175. (2013). arXiv:1209.1416 [hep-th].

[90]H. Fuji, S. Gukov, and P. Sułkowski. "Volume conjecture: refined and categorified (with an appendix by Hidetoshi Awata)." Advances in Theoretical and Mathematical Physics **16**, 1669–1777. (2012). arXiv:1203.2182 [hep-th].

[91]H. Fuji, S. Gukov, and P. Sułkowski. "Super-A-polynomial for knots and BPS states." Nuclear Physics B **867**, 506–546. (2013). arXiv:1205.1515 [hep-th].

[92]A. Gadde and S. Gukov. "2d index and surface operators." JHEP **03**, 080. (2014). arXiv:1305.0266 [hep-th].

[93]A. Gadde, L. Rastelli, S. S. Razamat, and W. Yan. "Gauge theories and Macdonald polynomials." Commun. Math. Phys. **319**, 147–193. (2013). arXiv:1110.3740 [hep-th].

[94]S. Garoufalidis. "On the characteristic and deformation varieties of a knot." Geometry and Topology Monographs **7**, 291–304. (2004). arXiv:0306230v4 [math].

[95]L. Gerritzen and M. van der Put. "Schottky groups and Mumford curves." Bull. Amer. Math. Soc. **8**, 375–377. (1983).

[96]D. Ghoshal and T. Kawano. "Towards p-adic string in constant B-field." Nucl. Phys. **B710**, 577–598. (2005). arXiv:hep-th/0409311 [hep-th].

[97]R. Gopakumar and C. Vafa. "On the gauge theory/geometry correspondence." Adv. Theor. Math. Phys. **3**, 1415–1443. (1999). arXiv:hep-th/9811131.

[98]B. Gornik. "Note on Khovanov link cohomology." (2004). arXiv:math/0402266.

[99]E. Gorsky, S. Gukov, and M. Stošić. "Quadruply-graded colored homology of knots." (2013). arXiv:1304.3481 [math.QA].

[100]E. Gorsky, A. Oblomkov, J. Rasmussen, and V. Shende. "Torus knots and the rational DAHA." Duke Math. J. **163**, 2709–2794. (2014). arXiv:1207.4523.

[101]P. Grange. "Deformation of p-adic string amplitudes in a magnetic field." Phys. Lett. **B616**, 135–140. (2005). arXiv:hep-th/0409305 [hep-th].

[102]M. Grassl, T. Beth, and M. Rötteler. "On optimal quantum codes." International Journal of Quantum Information **2**, 55–64. (2004).

[103]P. A. Griffiths and J. Morgan. *Rational homotopy theory and differential forms*. 2nd ed. Progress in Mathematics 16. (Birkhäuser. 2013).

[104]D. Guan. "On the Betti numbers of irreducible compact hyperkähler manifolds of complex dimension four." Math. Res. Lett. **8**, 663–669. (2001).

[105]S. S. Gubser, J. Knaute, S. Parikh, A. Samberg, and P. Witaszczyk. "p-adic AdS/CFT." (2016). arXiv:1605.01061 [hep-th].

[106]S. Gukov. "The power of refinement." to appear.

[107]S. Gukov. "Three-dimensional quantum gravity, Chern-Simons theory, and the A-polynomial." Commun. Math. Phys. **255**, 577–627. (2005). arXiv:hep-th/0306165 [hep-th].

[108]S. Gukov. "Gauge theory and knot homologies." Fortschritte der Physik **55**, 473–490. (2007). arXiv:0706.2369 [hep-th].

[109]S. Gukov, S. Nawata, I. Saberi, M. Stošić, and P. Sułkowski. "Sequencing BPS spectra." JHEP **03**, 004. (2016). arXiv:1512.07883 [hep-th].

[110]S. Gukov and I. Saberi. "Motivic quantum field theory and the number theory of A-polynomial curves." To appear.

[111]S. Gukov and I. Saberi. "Lectures on knot homology and quantum curves." In *Topology and field theories*. Edited by S. Stolz. Contemporary Mathematics 613. (American Mathematical Society. 2014). arXiv:1211.6075 [hep-th].

[112]S. Gukov, A. Schwarz, and C. Vafa. "Khovanov-Rozansky homology and topological strings." Lett. Math. Phys. **74**, 53–74. (2005). arXiv:0412243v3 [hep-th].

[113]S. Gukov and M. Stošić. "Homological algebra of knots and BPS states." Geometry & Topology Monographs **18**, 309–367. (2012). arXiv:1112.0030 [hep-th].

[114]S. Gukov and P. Sułkowski. "A-polynomial, B-model, and quantization." JHEP **1202**, 070. (2012). arXiv:1108.0002 [hep-th].

[115]S. Gukov and J. Walcher. "Matrix factorizations and Kauffman homology." (2005). arXiv:hep-th/0512298 [hep-th].

[116]S. Gukov and E. Witten. "Branes and quantization." Adv. Theor. Math. Phys. **13**. (2009). arXiv:0809.0305 [hep-th].

[117]J. A. Harvey and G. W. Moore. "Algebras, BPS states, and strings." Nuclear Physics B **463**, 315–368. (1996). arXiv:hep-th/9510182 [hep-th].

[118]J. A. Harvey and G. W. Moore. "On the algebras of BPS states." Communications in Mathematical Physics **197**, 489–519. (1998). arXiv:hep-th/9609017 [hep-th].

[119]M. Heydeman, M. Marcolli, I. Saberi, and B. Stoica. "Tensor networks, p-adic fields, and algebraic curves: arithmetic and the AdS_3/CFT_2 correspondence." (2016). arXiv:1605.07639 [hep-th].

[120]C. Holzhey, F. Larsen, and F. Wilczek. "Geometric and renormalized entropy in conformal field theory." Nuclear Physics B **424**, 443–467. (1994).

[121]D. Huybrechts. *Complex geometry: an introduction*. Universitext. (Springer-Verlag. 2005).

[122]A. Iqbal, C. Kozcaz, and C. Vafa. "The refined topological vertex." JHEP **0910**, 069. (2009). arXiv:hep-th/0701156 [hep-th].

[123]V. F. R. Jones. "A polynomial invariant for knots via von Neumann algebras." Bull. Amer. Math. Soc. **12**, 103–111. (1985).

[124]A. Kapustin and E. Witten. "Electric–magnetic duality and the geometric Langlands program." Communications in Number Theory and Physics **1**, 1–236. (2007). arXiv:hep-th/0604151 [hep-th].

[125]G. Kato. *The heart of cohomology*. (Springer-Verlag. 2006).

[126]S. H. Katz, A. Klemm, and C. Vafa. "Geometric engineering of quantum field theories." Nuclear Physics B **497**, 173–195. (1997). arXiv:hep-th/9609239 [hep-th].

[127]D. Kazhdan. *Lecture notes in motivic integration.* http://www.ma.huji.ac.il/~kazhdan/.

[128]M. Khovanov. "A categorification of the Jones polynomial." Duke Mathematical Journal **3**, 359–426. (2000). arXiv:math/9908171 [math.QA].

[129]M. Khovanov. "$\mathfrak{sl}(3)$ link homology." Algebr. Geom. Topol. **4**, 1045–1081. (2004). arXiv:math/0304375 [math.QA].

[130]M. Khovanov and L. Rozansky. "Virtual crossings, convolutions, and a categorification of the $\mathfrak{so}(2N)$ Kauffman polynomial." (2007). arXiv:math/0701333 [math.QA].

[131]M. Khovanov and L. Rozansky. "Matrix factorizations and link homology." Fundamenta Mathematicae **199**, 1–91. (2008). arXiv:math/0401268 [math.QA].

[132]M. Khovanov and L. Rozansky. "Matrix factorizations and link homology II." Geom. Topol. **12**, 1387–1425. (2008). arXiv:math/0505056 [math.QA].

[133]H. Kim and I. Saberi. "Real homotopy theory and supersymmetric quantum mechanics." (2015). arXiv:1511.00978 [hep-th].

[134]N. Koblitz. *p-adic numbers, p-adic analysis, and zeta-functions.* Graduate Texts in Mathematics 58. (Springer-Verlag. New York. 2012).

[135]T. Kohno. *Conformal field theory and topology.* Iwanami Series in Modern Mathematics. (American Mathematical Society. 1998).

[136]K. Krasnov. "Holography and Riemann surfaces." Adv. Theor. Math. Phys. **4**, 929–979. (2000). arXiv:hep-th/0005106 [hep-th].

[137]J. M. F. Labastida and M. Mariño. "Twisted $\mathcal{N} = 2$ supersymmetry with central charge and equivariant cohomology." Communications in Mathematical Physics **185**, 37–71. (1997). arXiv:hep-th/9603169 [hep-th].

[138]E. S. Lee. "An endomorphism of the Khovanov invariant." Adv. Math. **197.2**, 554–586. (2005). arXiv:math/0210213 [math.GT].

[139]S. Lee and M. Yamazaki. "3d Chern-Simons theory from M5-branes." JHEP **1312**, 035. (2013). arXiv:1305.2429 [hep-th].

[140]W. Lerche, C. Vafa, and N. P. Warner. "Chiral rings in $\mathcal{N} = 2$ superconformal theories." Nuclear Physics B **324**, 427–474. (1989).

[141]L. Lewark and A. Lobb. "New Quantum Obstructions to Sliceness." (2015). arXiv:1501.07138 [math.GT].

[142]F. Loeser. *Arizona winter school lecture notes on p-adic and motivic integration.* (2003) http://swc.math.arizona.edu/oldaws/03Notes.html.

[143]J. M. Maldacena and A. Strominger. "AdS_3 black holes and a stringy exclusion principle." JHEP **12**, 005. (1998). arXiv:hep-th/9804085 [hep-th].

[144]J. Maldacena and L. Susskind. "Cool horizons for entangled black holes." Fortsch. Phys. **61**, 781–811. (2013). arXiv:1306.0533 [hep-th].

[145]Yu. I. Manin. "p-adic automorphic functions." Journal of Mathematical Sciences **5**, 279–333. (1976).

[146]Yu. I. Manin. "Reflections on arithmetical physics." In *Conformal invariance and string theory*. (Academic Press. 1987).

[147]Yu. I. Manin. "Closed fibers at infinity in Arakelov's geometry." preprint PAM-479, Center for Pure and Applied Mathematics, University of California, Berkeley. 1989.

[148]Yu. I. Manin. "Three-dimensional hyperbolic geometry as ∞-adic Arakelov geometry." Inventiones mathematicae **104**, 223–243. (1991).

[149]Yu. I. Manin and V. Drinfeld. "Periods of p-adic Schottky groups." Journal für die reine und angewandte Mathematik **262**, 239–247. (1973).

[150]Yu. I. Manin and M. Marcolli. "Holography principle and arithmetic of algebraic curves." Advances in Theoretical and Mathematical Physics **5**, 617–650. (2002). arXiv:hep-th/0201036.

[151]M. Marcolli. *Arithmetic noncommutative geometry*. University Lecture Series 36. (American Mathematical Society. Providence, Rhode Island. 2005).

[152]M. Mariño. "Open string amplitudes and large order behavior in topological string theory." JHEP **0803**, 060. (2008). arXiv:hep-th/0612127.

[153]E. Melzer. "Non-archimedean conformal field theories." Int. J. Mod. Phys. **A4**, 4877. (1989).

[154]J. S. Milne. *Motives: Grothendieck's dream*. (2012) http://www.jmilne.org/math/.

[155]D. Montano and J. Sonnenschein. "The topology of moduli space and quantum field theory." Nuclear Physics B **324**, 348–370. (1989).

[156]H. R. Morton and P. R. Cromwell. "Distinguishing mutants by knot polynomials." J. Knot Theory Ramif. **5**, 225–238. (1996).

[157]D. Mumford. "An analytic construction of degenerating curves over complete local rings." Compositio Mathematica **24**, 129–174. (1972).

[158]H. Murakami, T. Ohtsuki, and S. Yamada. "HOMFLY polynomial via an invariant of colored planar graphs." Enseign. Math. **44**, 325–360. (1998).

[159]S. Nawata and A. Oblomkov. "Lectures on knot homology." (2015). arXiv:1510.01795 [math-ph].

[160]S. Nawata, P. Ramadevi, X. Sun, and Zodinmawia. "Super-A-polynomials for twist knots." JHEP **1211**, 157. (2012). arXiv:1209.1409 [hep-th].

[161]S. Nawata, P. Ramadevi, and Zodinmawia. "Colored Kauffman homology and super-A-polynomials." JHEP **1401**, 126. (2014). arXiv:1310.2240 [hep-th].

[162]N. A. Nekrasov. "Seiberg-Witten prepotential from instanton counting." Advances in Theoretical and Mathematical Physics **7**, 831–864. (2003). arXiv:hep-th/0206161 [hep-th].

[163]N. A. Nekrasov and S. L. Shatashvili. "Quantization of integrable systems and four-dimensional gauge theories." (2009). arXiv:0908.4052 [hep-th].

[164]J. Neukirch. *Algebraic number theory*. Grundlehren der mathematischen Wissenschaften 322. (Springer-Verlag. 2013).

[165]W. D. Neumann and D. Zagier. "Volumes of hyperbolic three-manifolds." Topology **24**, 307–332. (1985).

[166]L. Ng. "Framed knot contact homology." Duke Math. J. **141**, 365–406. (2008). arXiv:math/0407071 [math.GT].

[167]L. Ng. "Combinatorial knot contact homology and transverse knots." Adv. Math. **227**, 2189–2219. (2011). arXiv:1010.0451 [math.SG].

[168]H. Ooguri and C. Vafa. "Knot invariants and topological strings." Nuclear Physics B **577**, 419–438. (2000). arXiv:hep-th/9912123 [hep-th].

[169]P. Ozsváth and Z. Szabó. "Holomorphic disks and knot invariants." Advances in Mathematics **186**, 58–116. (2004).

[170]P. Ozsváth and Z. Szabó. "Knot Floer homology, genus bounds, and mutation." Topol. Appl. **141**, 59–85. (2004). arXiv:math/0303225.

[171]G. Parisi. "On p-adic functional integrals." Mod. Phys. Lett. **A3**, 639–643. (1988).

[172]F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill. "Holographic quantum error-correcting codes: toy models for the bulk/boundary correspondence." JHEP **2015**, 1–55. (2015). arXiv:1503.06237 [hep-th].

[173]J. A. Rasmussen. "Some differentials on Khovanov-Rozansky homology." (2006). arXiv:math/0607544 [math.GT].

[174]J. A. Rasmussen. "Floer homology and knot complements." (2003). arXiv:math/0306378 [math.GT].

[175]N. Yu. Reshetikhin and V. Turaev. "Invariants of 3-manifolds via link polynomials and quantum groups." Invent. Math. **103**, 547–597. (1991).

[176]N. Yu. Reshetikhin and V. G. Turaev. "Ribbon graphs and their invaraints derived from quantum groups." Comm. Math. Phys. **127**, 1–26. (1990).

[177]D. Rolfsen. *Knots and links*. (AMS Chelsea Publishing. 1976).

[178]C. Römelsberger. "Counting chiral primaries in $\mathcal{N} = 1$, $d = 4$ superconformal field theories." Nuclear Physics B **747**, 329–353. (2006). arXiv:hep-th/0510060.

[179]J. Rouse and J. J. Webb. "On spaces of modular forms spanned by eta-quotients." (2013). arXiv:1311.1460.

[180]S. Ryu and T. Takayanagi. "Holographic derivation of entanglement entropy from AdS/CFT." Phys. Rev. Lett. **96**, 181602. (2006). arXiv:hep-th/0603001 [hep-th].

[181]S. M. Salamon. "On the cohomology of Kähler and hyper-Kähler manifolds." Topology **35**, 137–155. (1996).

[182]P. J. Sally Jr. "An introduction to p-adic fields, harmonic analysis, and the representation theory of SL(2)." Lett. Math. Phys. **46**, 1–47. (1998).

[183]N. Seiberg and E. Witten. "Electric-magnetic duality, monopole condensation, and confinement in $\mathcal{N} = 2$ supersymmetric Yang–Mills theory." Nuclear Physics B **426**, 19–52. (1994).

[184]N. Seiberg and E. Witten. "Monopoles, duality and chiral symmetry breaking in $\mathcal{N} = 2$ supersymmetric QCD." Nuclear Physics B **431**, 484–550. (1994).

[185]P. Seidel and R. P. Thomas. "Braid group actions on derived categories of coherent sheaves." Duke Mathematical Journal **108**, 37–108. (2001). arXiv:math/0001043 [math-ag].

[186]S. L. Shatashvili and C. Vafa. "Superstrings and manifolds of exceptional holonomy." Selecta Mathematica **1**, 347–381. (1995).

[187]J. H. Silverman and J. Tate. *Rational points on elliptic curves*. (Springer. 1992).

[188]B. L. Spokoiny. "Quantum geometry of nonarchimedean particles and strings." Phys. Lett. **B208**, 401–406. (1988).

[189]M. Srednicki. "Entropy and area." Phys. Rev. Lett. **71**, 666–669. (1993). arXiv:hep-th/9303048 [hep-th].

[190]D. Sullivan. "Infinitesimal computations in topology." Publications Mathématiques de l'Institute des Hautes Études Scientifiques **47**, 269–331. (1977).

[191]B. Swingle. "Entanglement renormalization and holography." Phys. Rev. **D86**, 065007. (2012). arXiv:0905.1317 [cond-mat.str-el].

[192]L. Takhtajan and P. Zograf. "On uniformization of Riemann surfaces and the Weil–Petersson metric on Teichmüller and Schottky spaces." Math. USSR Sbornik **60**, 297–313. (1988).

[193]C. H. Taubes. "Lagrangians for the Gopakumar–Vafa conjecture." Geometry & Topology Monographs **8**, 73–95. (2006). arXiv:math/0201219.

[194]Y. Terashima and M. Yamazaki. "SL(2, $\mathbb{R}$) Chern-Simons, Liouville, and gauge theory on duality walls." JHEP **1108**, 135. (2011). arXiv:1103.5748.

[195]R. Vakil. *Foundations of algebraic geometry*. http://math.stanford.edu/~vakil/216blog/.

[196]M. van der Put. "Discrete groups, Mumford curves, and theta functions." In Annales de la faculté des sciences de Toulouse: Mathématiques. Vol. 1. 3. (1992), pp. 399–438.

[197]M. S. Verbitsky. "The action of a Lie algebra $\mathfrak{so}(5)$ on the cohomology of a hyperkähler manifold." Functional Analysis and Its Applications **24**, 229–230. (1990).

[198]M. S. Verbitsky. *Cohomology of compact hyperkähler manifolds*. 1995. arXiv:9501001 [math.AG].

[199]M. S. Verbitsky. "Cohomology of compact hyperkähler manifolds and its applications." Geometric and Functional Analysis **6**, 601–611. (1996). arXiv:9511009 [math.AG].

[200]W. Veys. "Arc spaces, motivic integration, and stringy invariants." Adv. Stud. Pure Math **43**, 529–572. (2004). arXiv:math/0401374.

[201]G. Vidal. "Class of quantum many-body states that can be efficiently simulated." Phys. Rev. Lett. **101**, 110501. (2008).

[202]V. S. Vladimirov, I. V. Volovich, and E. I. Zelenov, eds. *p-adic analysis and mathematical physics*. Series on Soviet and East European Mathematics 1. (World Scientific. 1994).

[203]V. S. Vladimirov. "Tables of integrals of complex-valued functions of p-adic arguments." Sovremennye Problemy Matematiki **2**, 3–88. (2003).

[204]J. Walcher. "Stability of Landau-Ginzburg branes." Journal of Mathematical Physics **46**, 082305. (2005). arXiv:hep-th/0412274 [hep-th].

[205]B. Webster. "Knot invariants and higher representation theory II: the categorification of quantum knot invariants." (2010). arXiv:1005.4559 [math.GT].

[206]B. Webster and G. Williamson. "A geometric construction of colored HOMFLYPT homology." (2009). arXiv:0905.0486 [math.GT].

[207]S. Wehrli. "Khovanov homology and Conway mutation." (2003). arXiv:math/0301312 [math.GT].

[208]E. Witten. "Dynamical breaking of supersymmetry." Nuclear Physics B **188**, 513–554. (1981).

[209]E. Witten. "Constraints on supersymmetry breaking." Nuclear Physics B **202**, 253–316. (1982).

[210]E. Witten. "Supersymmetry and Morse theory." Journal of Differential Geometry **17**, 661–692. (1982).

[211]E. Witten. "Topological quantum field theory." Communications in Mathematical Physics **117**, 353–386. (1988).

[212]E. Witten. "Quantum field theory and the Jones polynomial." Comm. Math. Phys. **121**, 351–399. (1989).

[213]E. Witten. "Phases of $\mathscr{N} = 2$ theories in two dimensions." Nuclear Physics B **403**, 159–222. (1993). arXiv:hep-th/9301042 [hep-th].

[214]E. Witten. "Monopoles and four-manifolds." Math. Res. Lett. **1**, 769–796. (1994). arXiv:hep-th/9411102 [hep-th].

[215]E. Witten. "On the Landau-Ginzburg description of $\mathscr{N} = 2$ minimal models." International Journal of Modern Physics A **9**, 4783–4800. (1994). arXiv:hep-th/9304026 [hep-th].

[216]E. Witten. "Anti-de Sitter space and holography." Adv. Theor. Math. Phys. **2**, 253–291. (1998). arXiv:hep-th/9802150 [hep-th].

[217]E. Witten. "Analytic continuation of Chern-Simons theory." In *Chern-Simons gauge theory: 20 years after*. Edited by J. E. Andersen. Studies in Advanced Mathematics. (American Mathematical Society/International Press. 2011), pp. 347–446. arXiv:1001.2933 [hep-th].

[218]E. Witten. "Fivebranes and knots." (2011). arXiv:1101.3216 [hep-th].

[219]K. Wong. "Spectral sequences and vacua in $\mathscr{N} = 2$ gauged linear quantum mechanics with potentials." (2015). arXiv:1511.05159 [hep-th].

[220]N. Woodhouse. *Geometric quantization*. (Oxford Univ. Press. 1997).

[221]H. Wu. "A colored $\mathfrak{sl}(N)$ homology for links in S^3." (2009). arXiv:0907.0695 [math.GT].

[222]J. Yagi. "3d TQFT from 6d SCFT." JHEP **1308**, 017. (2013). arXiv:1305.0291 [hep-th].

[223]C. N. Yang. "Magnetic monopoles, fiber bundles, and gauge fields." Annals of the New York Academy of Sciences **294**, 86–97. (1977).

[224]Y. Yonezawa. "Quantum $(\mathfrak{sl}_n, \wedge V_n)$ link invariant and matrix factorizations." Nagoya Mathematical Journal **204**, 69–123. (2011). arXiv:0906.0220 [math.GT].

[225]A. V. Zabrodin. "Non-archimedean strings and Bruhat–Tits trees." Commun. Math. Phys. **123**, 463. (1989).

[226]R. Zhang. "Lagrangian formulation of open and closed p-adic strings." Phys. Lett. **B209**, 229–232. (1988).

[227]J. Zhou. "Rational homotopy types of mirror manifolds." (1999). arXiv:math/9910027 [math.DG].

Appendix A

SUPPLEMENTAL BACKGROUND ON p-ADIC ANALYSIS

A.1 p-adic integration

Here we review some aspects of p-adic integration, including basic properties and examples, the Fourier transform, and the p-adic gamma function Γ_p. A more comprehensive review is found in [33]. For formal proofs, as well as extensive integration tables, the reader may consult [203].

As already discussed, the unique additive Haar measure dx on $\mathbb{Q}_p$ is normalized so that

$$\int_{\mathbb{Z}_p} dx = 1. \tag{A.1}$$

To find the volume of the set B^r, which consists of $x \in \{\mathbb{Q}_p, |x|_p \leq p^r\}$, we may scale the measure and reduce this to the integral above on $\mathbb{Z}_p$ as:

$$\int_{B^r} dx = p^r \int_{\mathbb{Z}_p} dx = p^r. \tag{A.2}$$

As $r \to \infty$, the volume diverges as in the real case. Compactifying the point at infinity amounts to switching from the Haar measure to the Patterson-Sullivan measure $d\mu_0(x)$; these measures agree on $\mathbb{Z}_p$ and differ in the complement by $d\mu_0(x) = dx/|x|_p^2$.

With this measure the volume is computed with a change of variables:

$$\int_{\mathbb{Q}_p} d\mu_0(x) = \int_{\mathbb{Z}_p} dx + \int_{\mathbb{Q}_p - \mathbb{Z}_p} |x|_p^{-2} dx \tag{A.3}$$

$$= 1 + \frac{1}{p} \int_{\mathbb{Z}_p} du, \quad u = \frac{1}{px}, \quad du = \frac{p\, dx}{|x|_p^2} \tag{A.4}$$

$$= \frac{p+1}{p}. \tag{A.5}$$

A large class of elementary integrals may be evaluated using these methods; see the above references for complete details.

We now turn our attention to the p-adic Fourier transform of a function $f(x) : \mathbb{Q}_p \to \mathbb{C}$. As discussed in section 4.4, this involves integrating the function against the additive character $\chi(x) = e^{2\pi i\{kx\}}$ over all $\mathbb{Q}_p$. This generates a

new complex valued function in terms of the p-adic momentum $k \in \mathbb{Q}_p$:

$$\widetilde{f}(k) = \int_{\mathbb{Q}_p} \chi(kx) f(x) dx, \tag{A.6}$$

$$f(x) = \int_{\mathbb{Q}_p} \chi(-kx) \widetilde{f}(k) dk. \tag{A.7}$$

The analogy with the real Fourier transform should be clear. In practice evaluating this kind of integral often requires one to divide $\mathbb{Q}_p$ into spheres consisting of points with $|x|_p = p^n$ and performing the integral on each sphere. This can be seen in the example:

$$\int_{B^r} \chi(kx) dx = \begin{cases} p^r, & |k|_p \leq p^{-r} \\ 0, & \text{otherwise.} \end{cases} \tag{A.8}$$

As in the real case, one may find tables with numerous p-adic Fourier transforms of elementary functions in the literature.

The final integral expression is that of the Gelfand-Graev-Tate Γ function:

$$\Gamma_p(\alpha) = \int_{\mathbb{Q}_p} \chi(x) |x|_p^{\alpha-1} dx = \frac{1 - p^{s-1}}{1 - p^{-s}}. \tag{A.9}$$

This function has some similar properties to the ordinary gamma function. It is fairly ubiquitous in certain p-adic integral calculations, and we refer the reader to literature on p-adic string theory for details.

A.2 p-adic differentiation

As already discussed, complex fields living on the boundary $\mathrm{P}^1(\mathbb{Q}_p)$ are maps

$$f(x) : \mathrm{P}^1(\mathbb{Q}_p) \to \mathbb{C}. \tag{A.10}$$

In the archimedean case of 2d conformal field theory, we have $f(z, \bar{z}) : \mathrm{P}^1(\mathbb{C}) \to \mathbb{C}$ and it makes sense to define holomorphic and antiholomorphic derivatives $\frac{\partial f}{\partial z}$ and $\frac{\partial f}{\partial \bar{z}}$. In the p-adic case, analogous differentiation expressions no longer make sense, as we would be dividing a complex number by a p-adic number and such an operation is not defined.

The only notion of derivative we may use is the *Vladimirov derivative* [71, 203], and it can be thought of as a nonlocal pseudo-differential operator. Roughly speaking, this operation is the p-adic analog of Cauchy's Differentiation Formula in which the derivative of a function at a point is expressed as a weighted

integral of the function over a curve. It is also know as a *normal derivative* [225] in the context of the p-adic string where it is interpreted as the derivative of the embedding coordinates X^μ normal to the boundary of the worldsheet. Becaues this operator is defined on $\mathbb{Q}_p$ without any reference to an embedding or worldsheet, we opt to refer to it as a Vladimirov derivative. The n^{th} Vladimirov derivative is defined by

$$\partial^n_{(p)} f(x) = \int_{\mathbb{Q}_p} \frac{f(x') - f(x)}{|x' - x|_p^{n+1}} dx'. \tag{A.11}$$

Some authors may choose a different normalization in front of this integral; usually in the form of p-adic gamma functions. At first sight the expression above may not resemble any familiar notions of differentiation. We may see this as a good notion for derivative in two ways; in the case of the p-adic string this expression is the limit of the normal derivative on T_p as we go to the boundary, as shown in [225]. We may also compute the Vladimirov derivative of some special p-adic functions and compare with the real case. This is done in the following section.

Examples

We wish to first compute the derivative of the additive character, $\chi(kx)$. This function is the p-adic analog of a plane wave with momentum k, so we expect it to be an eigenfunction of the derivative with eigenvalue related to k.

$$\partial^n_{(p)} \chi(kx) = \int_{\mathbb{Q}_p} \frac{\chi(kx') - \chi(kx)}{|x' - x|_p^{n+1}} dx'. \tag{A.12}$$

Using the properties of the additive Haar measure of $\mathbb{Q}_p$, we can shift the integration measure,

$$y = k(x' - x), \quad dy = |k|_p dx' \tag{A.13}$$

and simplify the integral

$$\int_{\mathbb{Q}_p} \frac{\chi(kx') - \chi(kx)}{|x' - x|_p^{n+1}} dx' \tag{A.14}$$

$$= |k|_p^n \int_{\mathbb{Q}_p} \frac{\chi(y + kx) - \chi(kx)}{|y|_p^{n+1}} dy \tag{A.15}$$

$$= |k|_p^n \chi(kx) \int_{\mathbb{Q}_p} \frac{\chi(y) - 1}{|y|_p^{n+1}} dy, \tag{A.16}$$

where we used the additive property of the character to extract the x dependence. The integral appears to diverge at $y \sim 0$; this divergence is actually canceled by the numerator and can be seen by introducing an infrared cutoff. Regularization of this integral is discussed in [19] and [96]. The result is

$$\int_{\mathbb{Q}_p} \frac{\chi(y)-1}{|y|_p^{n+1}} dy = \frac{1-p^{-n-1}}{1-p^n} = \Gamma_p(-n), \tag{A.17}$$

where we have used the definition of the p-adic gamma function in Eq. (A.9). So the end result is

$$\partial_{(p)}^n \chi(kx) = \Gamma_p(-n)|k|_p^n \chi(kx). \tag{A.18}$$

Up to the factor of the gamma function (which could be absorbed into the normalization of the derivative,) we see the additive character $\chi(kx)$ is an eigenfunction of the Vladimirov derivative with the eigenvalue given by the p-adic norm of its "momentum."

Another example we may wish to compute is the n^{th} derivative of $|x|_p^s$ for some $s \in \mathbb{C}$. This may be most easily be computed by Fourier transform and serves as an example of an alternative representation of the Vladimirov derivative:

$$\partial_{(p)}^n |x|_p^s = \int \chi(-kx)|k|_p^n \widetilde{|x|^s}_p dk, \tag{A.19}$$

where $\widetilde{|x|^s}_p$ is the p-adic Fourier transform of $|x|_p^s$, given in [203, 226]:

$$\widetilde{|x|^s}_p = \int \chi(kx)|x|_p^s dx = \Gamma_p(s+1)|k|_p^{-s-1} \tag{A.20}$$

everywhere it is defined. Applying this formula twice to the derivative we wish to compute, we arrive at

$$\partial_{(p)}^n |x|_p^s = \Gamma_p(s+1)\Gamma_p(n-s)|x|_p^{s-n}, \tag{A.21}$$

which should resemble the ordinary n^{th} derivative of a polynomial function.

INDEX