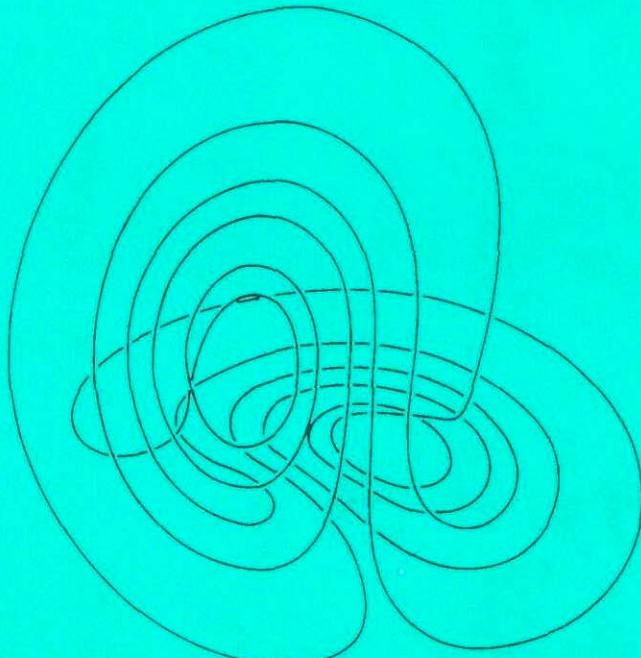


 Series on Knots and Everything — Vol. 1

Louis H. Kauffman

KNOTS AND PHYSICS

Third Edition



World Scientific

KNOTS AND PHYSICS

KNOTS

AND

PHYSICS

Third Edition

Louis H. Kauffman

Department of Mathematics
Statistics and Computer Science
University of Illinois at Chicago



World Scientific

Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

First edition 1991

Second edition 1993

Library of Congress Cataloging-in-Publication Data

Kauffman, Louis, H., 1945-

Knots and physics / Louis H. Kauffman.

p. cm. -- (Series on knots and everything ; vol. 1)

Includes bibliographical references and index.

ISBN 9810203438. -- ISBN 9810203446 (pbk.)

1. Knot polynomials. 2. Mathematical physics. I. Title.

II. Series.

QC20.7.K56K38 1991

514'224--dc20

91-737

CIP

KNOTS AND PHYSICS (Third Edition)

Copyright © 2001 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-02-4111-9

981-02-4112-7 (pbk)

Printed in Singapore.

To Diane

Preface to the First Edition

This book has its origins in two short courses given by the author in Bologna and Torino, Italy during the Fall of 1985. At that time, connections between statistical physics and the Jones polynomial were just beginning to appear, and it seemed to be a good idea to write a book of lecture notes entitled **Knots and Physics**.

The subject of knot polynomials was opening up, with the Jones polynomial as the first link polynomial able to distinguish knots from their mirror images. *We were looking at the tip of an iceberg!* The field has grown by leaps and bounds with remarkable contributions from mathematicians and physicists – a wonderful interdisciplinary interplay.

In writing this book I wanted to preserve the flavor of those old Bologna/Torino notes, and I wanted to provide a pathway into the more recent events. After a good deal of exploration, I decided, in 1989, to design a book divided into two parts. The first part would be combinatorial, elementary, devoted to the bracket polynomial as state model, partition function, vacuum-vacuum amplitude, Yang-Baxter model. The bracket also provides an entry point into the subject of quantum groups, and it is the beginning of a significant generalization of the Penrose spin-networks (see Part II, section 13⁰.) Part II is an exposition of a set of related topics, and provides room for recent developments. In its first incarnation, Part II held material on the Potts model and on spin-networks.

Part I grew to include expositions of Yang-Baxter models for the Homfly and Kauffman polynomials – as discovered by Jones and Turaev, and a treatment of the Alexander polynomial based on work of Francois Jaeger, Hubert Saleur and the author. By using Yang-Baxter models, we obtain an induction-free introduction to the existence of the Jones polynomial and its generalizations. Later, Part I grew some more and picked up a chapter on the 3-manifold invariants of Reshetikhin and Turaev as reformulated by Raymond Lickorish. The Lickorish model is completely elementary, using nothing but the bracket, trickery with link diagrams, and the tangle diagrammatic interpretation of the Temperley-Lieb algebra. These 3-manifold invariants were foretold by Edward Witten in his landmark paper on quantum field theory and the Jones polynomial. Part I ends with an introduction to Witten's functional integral formalism, and shows how the knot polynomials

arise in that context. An appendix computes the Yang-Baxter solutions for spin-preserving R -matrices in dimension two. From this place the Jones and Alexander polynomials appear as twins!

Part II begins with Bayman's theory of hitches – how to prove that your horse can not get away if you tie him with a well constructed clove hitch. Then follows a discussion of the experiments available with a rubber band. In sections 3^0 and 4^0 we discuss attempts to burrow beneath the usual Reidemeister moves for link diagrams. There are undiscovered realms beneath our feet. Then comes a discussion of the Penrose chromatic recursion (for colorations of planar three-valent graphs) and its generalizations. This provides an elementary entrance into spin networks, coloring and recoupling theory. Sections 7^0 and 8^0 on coloring and the Potts model are taken directly from the course in Torino. They show how the bracket model encompasses the dichromatic polynomial, the Jones polynomial and the Potts model. Section 9^0 is a notebook on spin, quantum mechanics, and special relativity. Sections 10^0 and 11^0 play with the quaternions, Cayley numbers and the Dirac string trick. Section 11^0 gives instructions for building your very own topological/mechanical quaternion demonstrator with nothing but scissors, paper and tape. Sections 12^0 and 13^0 discuss spin networks and q -deformed spin networks. The end of section 13^0 outlines work of the author and Sostenes Lins, constructing 3-manifold invariants of Turaev-Viro type via q -spin networks. Part II ends with three essays: Strings, DNA, Lorenz Attractor. These parting essays contain numerous suggestions for further play.

Much is left unsaid. I would particularly like to have had the space to discuss Louis Crane's approach to defining Witten's invariants using conformal field theory, and Lee Smolin's approach to quantum gravity – where the states of the theory are functionals on knots and links. This will have to wait for the next time!

It gives me great pleasure to thank the following people for good conversation and intellectual sustenance over the years of this project.

Corrado Agnes	Steve Bryson
Jack Armel	Scott Carter
Randy Baadhio	Naomi Caspe
Gary Berkowitz	John Conway
Joan Birman	Paolo Cotta-Ramusino
Joe Birman	Nick Cozzarelli
Herbert Brun	Louis Crane
William Browder	Anne Dale

Tom Etter
Massimo Ferri
David Finkelstein
Michael Fisher
James Flagg
George Francis
Peter Freund
Larry Glasser
Seth Goldberg
Fred Goodman
Arthur Greenspoon
Bernard Grossman
Victor Guggenheim
Ivan Handler
John Hart
Mark Hennings
Kim Hix
Tom Imbo
Francois Jaeger
Herbert Jehle
Vaughan Jones
Alice Kauffman
Akio Kawauchi
Milton Kerker
Doug Kipping
Robion Kirby
Peter Landweber
Ruth Lawrence
Vladimir Lepetic
Anatoly Libgober
Raymond Lickorish
Sostenes Lins
Roy Lisker
Mary Lupa
Maurizio Martellini
John Mathias
Ken Millett
John Milnor
Jose Montesinos
Hugh Morton
Kunio Murasugi
Jeanette Nelson
Frank Nijhoff
Pierre Noyes
Kathy O'Hara
Eddie Oshins
Gordon Pask
Annetta Pedretti
Mary-Minn Peet
Roger Penrose
Mario Rasetti
Nicolai Reshetikhin
Dennis Roseman
Hubert Saleur
Dan Sandin
Jon Simon
Isadore Singer
Milton Singer
Diane Slaviero
Lee Smolin
David Solzman
George Spencer-Brown
Sylvia Spengler
Joe Staley
Stan Tenen
Tom Tieffenbacher
Volodja Turaev
Sasha Turbiner

David Urman	Randall Weiss
Jean-Michel Vappereau	Steve Winker
Rasa Varanka	Edward Witten
Francisco Varela	John Wood
Oleg Viro	Nancy Wood
Heinz von Foerster	David Yetter
Miki Wadati	

(As with all lists, this one is necessarily incomplete. Networks of communication and support span the globe.) I also thank the Institute des Hautes Études Scientifiques of Bures-Sur-Yvette, France for their hospitality during the academic year 1988-1989, when this book was begun.

Finally, I wish to thank Ms. Shirley Roper, Word Processing Supervisor in the Department of Mathematics, Statistics, and Computer Science at the University of Illinois at Chicago, for a great typesetting job and for endurance above and beyond the call of duty.

This project was partially supported by NSF Grant Number DMS-8822602 and the Program for Mathematics and Molecular Biology, University of California at Berkeley, Berkeley, California.

Chicago, Illinois, January 1991

Preface to the Second Edition

This second edition of Knots and Physics contains corrections to misprints that appeared in the first edition, plus an appendix that contains a discussion of graph invariants and Vassiliev invariants followed by a reprinting of four papers by the author. This discussion and the included papers constitute an update and extension of the material in the original edition of the book.

Chicago, Illinois
September 1993

Preface to the Third Edition

This third edition of Knots and Physics contains corrections to misprints in the second edition plus a new article in the appendix entitled “Knot Theory and Functional Integration”. This article provides an introduction to the relationships between Vassiliev invariants, Knotsevich integrals and Witten’s approach to link invariants via quantum field theory.

Chicago, Illinois and
Galais, France
November 15, 2000

Table of Contents

Preface to the First Edition	vii
Preface to the Second Edition	xi
Preface to the Third Edition	xiii

Part I. A Short Course of Knots and Physics

1. Physical Knots	4
2. Diagrams and Moves	8
3. States and the Bracket Polynomial	25
4. Alternating Links and Checkerboard Surfaces	39
5. The Jones Polynomial and its Generalizations	49
6. An Oriented State Model for $V_K(t)$	74
7. Braids and the Jones Polynomial	85
8. Abstract Tensors and the Yang-Baxter Equation	104
9. Formal Feynman Diagrams, Bracket as a Vacuum-Vacuum Expectation and the Quantum Group $SL(2)q$	117
10. The Form of the Universal R -matrix	148
11. Yang-Baxter Models for Specializations of the Homfly Polynomial ..	161
12. The Alexander Polynomial	174
13. Knot-Crystals – Classical Knot Theory in a Modern Guise	186
14. The Kauffman Polynomial	215
15. Oriented Models and Piecewise Linear Models	235
16. Three Manifold Invariants from the Jones Polynomial	250
17. Integral Heuristics and Witten’s Invariants	285
18. Appendix – Solutions to the Yang-Baxter Equation	316

Part II. Knots and Physics — Miscellany

1. Theory of Hitches	323
2. The Rubber Band and Twisted Tube	329
3. On a Crossing	332
4. Slide Equivalence	336
5. Unoriented Diagrams and Linking Numbers	339
6. The Penrose Chromatic Recursion	346
7. The Chromatic Polynomial	353
8. The Potts Model and the Dichromatic Polynomial	364

9. Preliminaries for Quantum Mechanics, Spin Networks and Angular Momentum	381
10. Quaternions, Cayley Numbers and the Belt Trick	403
11. The Quaternion Demonstrator	427
12. The Penrose Theory of Spin Networks.....	443
13. <i>Q</i> -Spin Networks and the Magic Weave.....	459
14. Knots and Strings – Knotted Strings	475
15. DNA and Quantum Field Theory	488
16. Knots in Dynamical Systems – The Lorenz Attractor	501
Coda	511
References	513
Index	531

Appendix

Introduction	541
Gauss Codes, Quantum Groups and Ribbon Hopf Algebras	551
Spin Networks, Topology and Discrete Physics	597
Link Polynomials and a Graphical Calculus	638
<i>(with P. Vogel)</i>	
Knots, Tangles, and Electrical Networks	684
<i>(with J. R. Goldman)</i>	
Knot Theory and Functional Integration	724

PART I

PART I. A SHORT COURSE OF KNOTS AND PHYSICS.

This first part of the book is a short course on knots and physics. It is a rapid penetration into key ideas and examples. The second half of the book consists in a series of excursions into special topics (such as the Potts Model, map coloring, spin-networks, topology of 3-manifolds, and special examples). These topics can be read by themselves, but they are informed by Part I.

Acts of abstraction, changes of mathematical mood, shifts of viewpoint are the rule rather than the exception in this short course. The course is a rapid guided ascent straight up the side of a cliff! Later, we shall use this perspective - both for the planning of more leisurely walks, and for the plotting of more complex ascents.

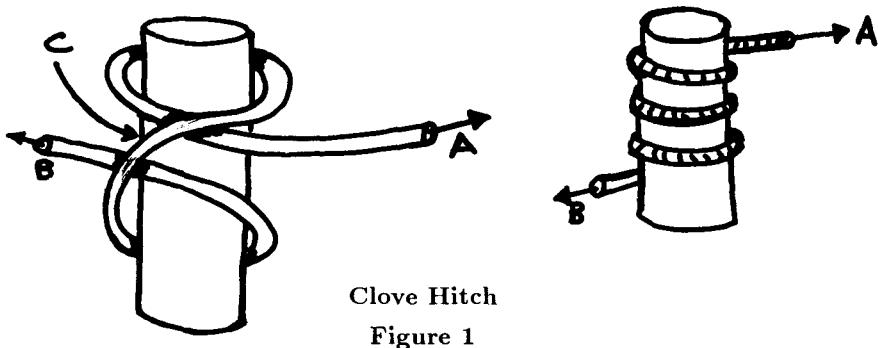
Here is the plan of the short course. First we discuss the diagrammatic model for knot theory. Then we construct the bracket polynomial [LK4], and obtain from it the original Jones polynomial. We then back-track to the diagrams and uncover abstract tensors, the Yang-Baxter equation and the concept of the quantum group. With state and tensor models for the bracket in hand, we then introduce other (generalized) link polynomials. Then comes a sketch of the Witten invariants, first via combinatorial models, then using functional integrals and gauge field theory. The ideas and techniques spiral outward toward field theory and three-dimensional topology. Statistical mechanics models occur in the middle, and continue to weave in and out throughout the play. Underneath all this structure of topology and physical ideas there sounds a theme of combinatorics - graphs, matroids, coloring problems, chromatic polynomials, combinatorial structures, enumerations. And finally, throughout there is a deep underlying movement in the relation between space and sign, geometry and symbol, the source common to mathematics, physics, language and all.

And so we begin. It is customary to begin physics texts either with mathematical background, or with the results of experiments. But this is a book of knots. Could we then begin with the physics of knots? Frictional properties? The design of the clove-hitch and the bow-line? Here is the palpable and practical

physics of the knots.

1^o. Physical Knots.

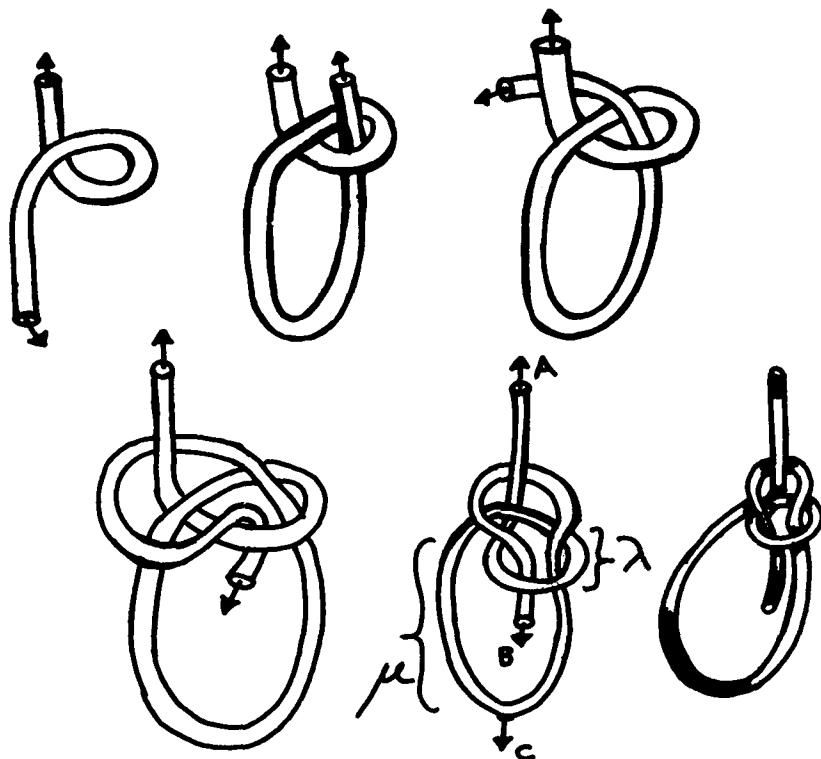
The **clove-hitch** is used to secure a line to a post or tree-trunk. The illustration in Figure 1 shows how it is made.



A little experimentation with the clove hitch shows that it holds very well indeed, with tension at *A* causing the line at *C* to grab *B* - and keep the assemblage from slipping.

Some physical mathematics can be put in back of these remarks - a project that we defer to Part II. At this stage, I wish to encourage experimentation. For example, just wind the rope around the post (as also in Figure 1). You will discover easily that the amount of tension needed at *A* to make the rope slide on the post increases exponentially with the number of windings. (The rope is held lightly at *B* with a fixed tension.) With the introduction of weaving - as in the clove hitch - interlocking tensions and frictions can produce an excellent bind.

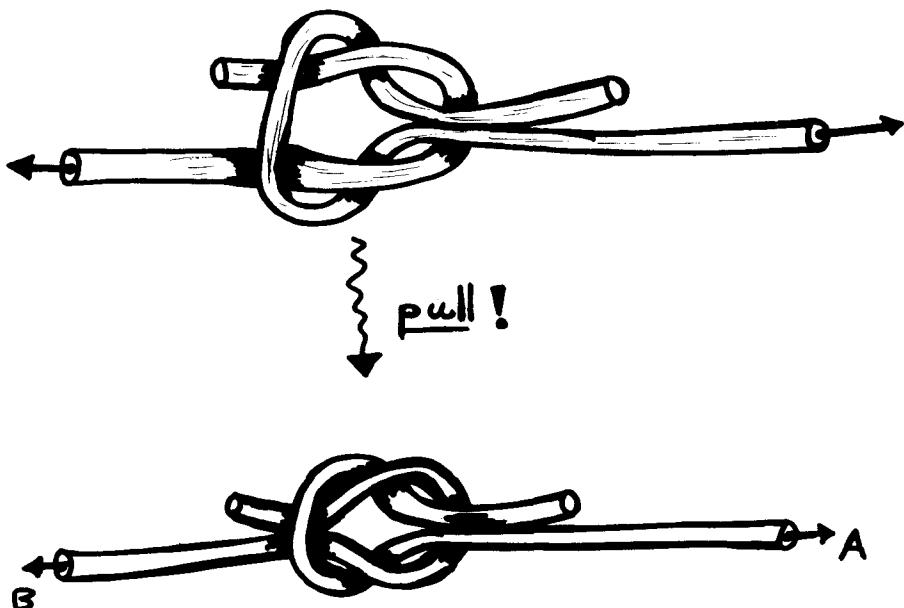
In Figure 2 we find a sequence of drawings that show the tying of the **bowline**.



The Bowline

Figure 2

In the bowline we have an excellent knot for securing a line. The concentrated part of the knot allows the loop μ to be of arbitrary size. This loop is secure under tension at A . Pulling at A causes the loop λ to shrink and grab tightly the two lines passing through it. This secures μ . Relax tension at A and the knot is easily undone, making loop adjustment easy.



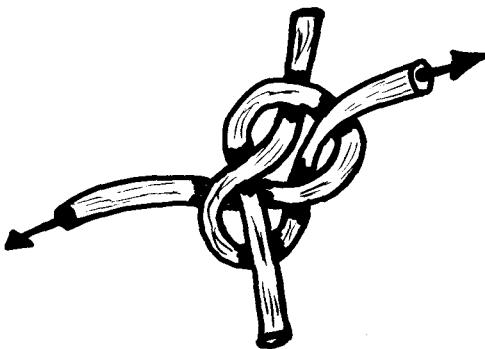
The Square Knot

Figure 3

In Figure 3 we find the **square knot**, excellent for splicing two lengths of rope. Tension at *A* and *B* creates an effect of mutual constriction - keeping the knot from slipping.

And finally in this rogue's gallery of knots and hitches we have the **granny** (Figure 4), a relative of the square knot, but quite treacherous. The granny will not hold. No mutual interlock here. A pull facilitates a flow of rope through the knot's pattern until it is undone.

It is important to come to some practical understanding of how these knots work. The facts that the square knot holds, and that the granny does not hold are best observed with actual rope models. It is a tremendous challenge to give a good mathematical analysis of these phenomena. Tension, friction and topology conspire to give the configuration a form - and within that form the events of slippage or interlock can occur.



The Granny Knot

Figure 4

I raise these direct physical questions about knotting, not because we shall answer them, but rather as an indication of the difficulty in which we stand. A book on knots and physics cannot ignore the physicality of knots of rope in space. Yet all of our successes will be more abstract, more linguistic, patterned and poised between the internal logic of patterns and their external realizations. And yet these excursions will strike deeply into topological and physical questions for knots and other realms as well.

There are other direct physical questions for knotting. We shall return to them in the course of the conversation.

2^o. Diagrams and Moves.

The transition from matters of practical knotwork to the associated topological problems is facilitated by deciding that a **knot** shall correspond to a **single closed loop of rope** (A **link** is composed of a number of such loops.). Thus the square knot is depicted as the closed loop shown in Figure 5. This figure also shows the loop-form for the granny knot.

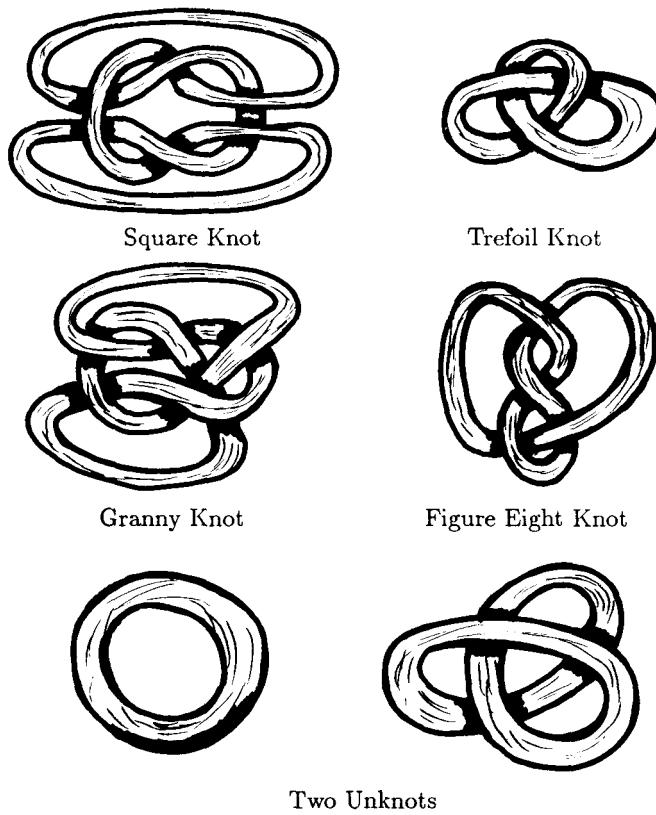


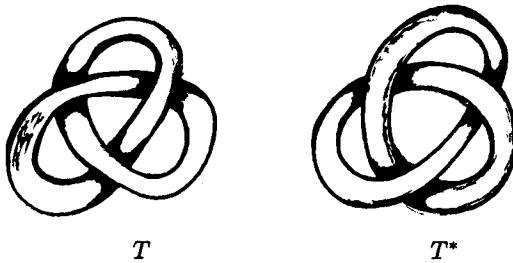
Figure 5

The advantage for topological exploration of the loop-form is obvious. One can make a knot, splice the ends, and then proceed to experiment with various deformations of it - without fear of losing the pattern as it slips off the end of the rope.

On the other hand, some of the original motivation for using the knot may be lost as we go to the closed loop form. The square knot is used to secure two lengths of rope. This is nowhere apparent in the loop form.

For topology, the mathematical advantages of the closed form are overwhelming. We have a uniform definition of a knot or link and correspondingly, a definition of **unknottedness**. A standard ring, as shown in Figure 5 is the canonical unknot. Any knot that can be continuously deformed to this ring (without tearing the rope) is said to be unknotted.

Note that we have not yet abstracted a mathematical definition of knot and link (although this is now relatively easy). Rather, I wish to emphasize the advantages of the closed loop form in doing experimental topological work. For example, one can form both the trefoil T , as shown in Figures 5 and 6, and its mirror image T^* (as shown in Figure 6). The trefoil T **cannot** be continuously deformed into its mirror image T^* . It is a remarkably subtle matter to prove this fact. One should try to actually create the deformation with a model - to appreciate this problem.



Trefoil and its Mirror Image

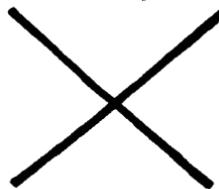
Figure 6

A knot is said to be **chiral** if it is topologically distinct from its mirror image, and **achiral** if it can be deformed into its mirror image. Remarkably, the figure eight

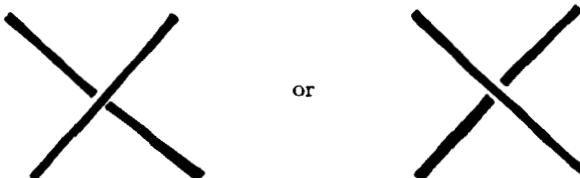
knot (see Figure 5) can be deformed into its mirror image (Exercise: Create this deformation.). Thus the figure eight knot is achiral, while the trefoil is chiral.

Yet, before jumping ahead to the chirality of the trefoil, we need to have a mathematical model in which facts such as chirality or knottedness can be proved or disproved.

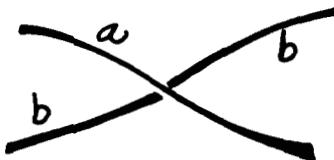
One way to create such a model is through the use of **diagrams** and **moves**. A **diagram** is an abstract and schematized picture of a knot or link composed of curves in the plane that cross transversely in 4-fold vertices:



Each vertex is equipped with extra structure in the form of a deleted segment:



The deletion indicates the **undercrossing line**. Thus in

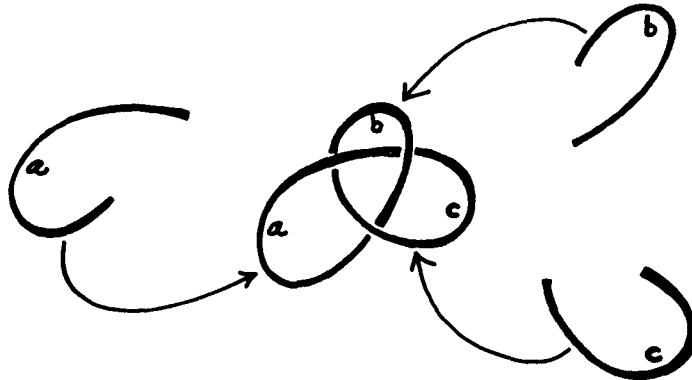


the line labelled *a* **crosses over** the line labelled *b*. The diagram for the trefoil *T* (of Figure 6) is therefore:



T

Because we have deleted small segments at each crossing, the knot or link diagram can be viewed as an interrelated collection of **arcs**. Thus for the trefoil T , I have labelled below the arcs a, b, c on the diagram:



Each arc begins at an undercrossing, and continues, without break, until it reaches another undercrossing.

The diagram is a compact **weaving pattern** from which one can make an unambiguous rope model. It has the appearance of a projection or shadow of the knot in the plane. I shall, however, take **shadow** as a technical term that indicates the knot/link diagram without the breaks, hence without under or over-crossing information. Thus in Figure 7, we have shown the trefoil, the Hopf link, the Borromean rings and their shadows. The shadows are plane graphs with 4-valent vertices. Here I mean a graph in the sense of graph theory (technically a multi-graph). A graph G consists of two sets: a set of **vertices** $V(G)$ and a set of edges $E(G)$. To each edge is associated a set of one or two vertices (the **endpoints** of that edge). Graphs are commonly realized by letting the edges be curve segments, and the endpoints of these edges are the vertices.

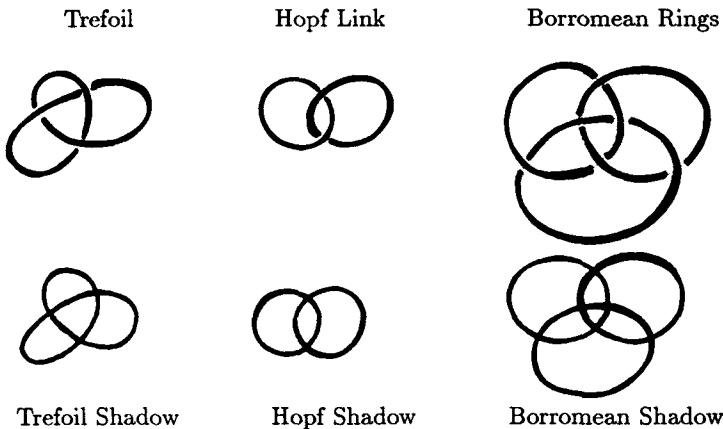
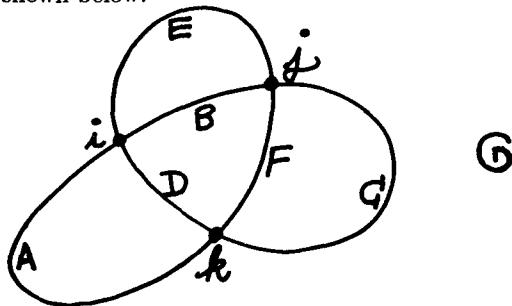


Figure 7

In the trefoil shadow, the graph G that it delineates can be described via the labelled diagram shown below:



$$E(G) = \{A, B, C, D, E, F\}$$

$$V(G) = \{i, j, k\}$$

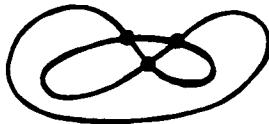
$$bA = \{i, k\} \quad bE = \{i, j\}$$

$$bB = \{i, j\} \quad bF = \{j, k\}$$

$$bC = \{j, k\}$$

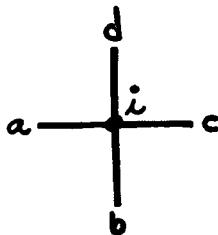
$$bD = \{i, k\}.$$

Here bX denotes the set of end-points of the edge X . The algebraic data shown above is still insufficient to specify the shadow, since we require the graph to be realized as a planar embedding. For example, the diagram

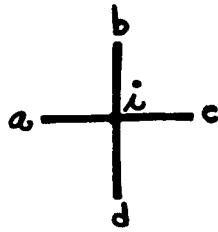


has the same abstract graph as the trefoil shadow.

The matter of encoding the information in the shadow, and in the diagram is a fascinating topic. For the present I shall let the diagrams (presented in the plane) speak for themselves. It is, however, worth mentioning, that the planar embedding of a vertex

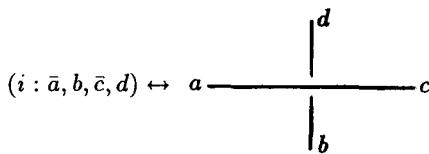


can be specified by giving the (counter-clockwise) order of the edges incident to that vertex. Thus $(i : a, b, c, d)$ can denote the vertex as drawn above, while $(i : a, d, c, b)$ indicates the vertex



Note that both $(i : a, b, c, d)$ and $(i : a, d, c, b)$ have the same cross-lines ($a \rightarrow c$ and $b \rightarrow d$). One is obtained from the other by picking up the pattern and putting it back in the plane after turning it over. An augmentation of this code can indicate the over and undercrossing states. Thus we can write $(i : \bar{a}, b, \bar{c}, d)$ to indicate that

the segment ac crosses over bd .

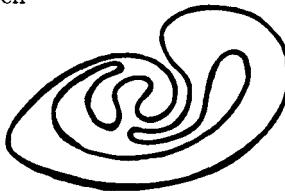


But this is enough said about graph-encodement.

Let the diagram take precedence. The diagram is to be regarded as a notational device at the same level as the letters and other symbols used in mathematical writing. Now I ask that you view these diagrams with the eyes of a topologist. Those topological eyes see no difference between



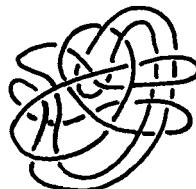
and



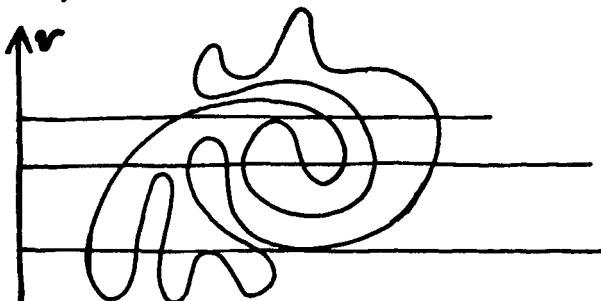
or between



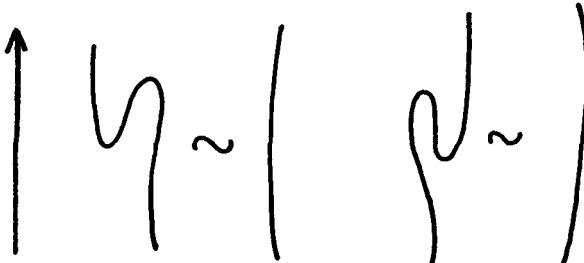
and



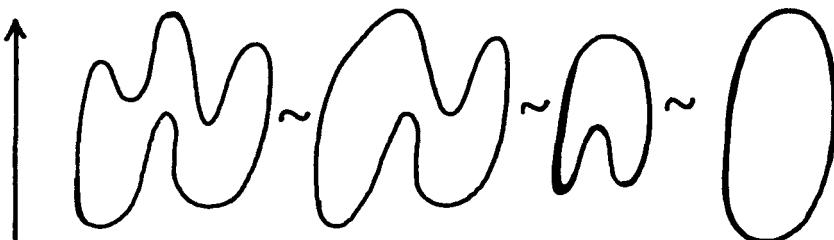
You will find that for a smoothly drawn simple closed curve in the plane (no self-intersections) there is a direction so that lines perpendicular to this direction (\vec{v}) intersect the curve either transversally (linearly independent tangent vectors) or tangentially.



With respect to this special direction the moves



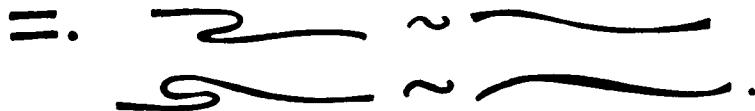
can be used repeatedly to simplify the curve until it has only one (local) maximum and one (local) minimum:



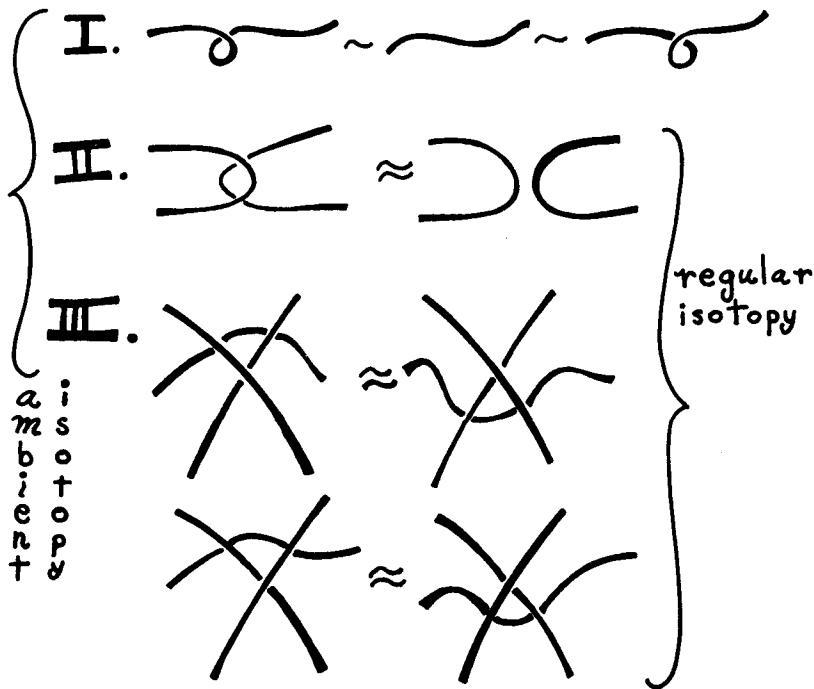
Later, this way of thinking will be very useful to us. We have just sketched a proof of the Jordan Curve Theorem [ST] for smooth (differentiable, non-zero tangent vector) curves in the plane. The Jordan Curve Theorem states that such a curve divides the plane into two disjoint regions, each homeomorphic to a disk. By moving the curve to a standard form, we accomplish these homeomorphisms.

Notationally the Jordan curve theorem is a fact about the plane upon which we write. It is the fundamental underlying fact that makes the diagrammatics of knots and links correspond to their mathematics. This is a remarkable situation – a fundamental theorem of mathematics is the underpinning of a notation for that same mathematics.

In any case, I shall refer to the basic topological deformations of a plane curve as **Move Zero**:



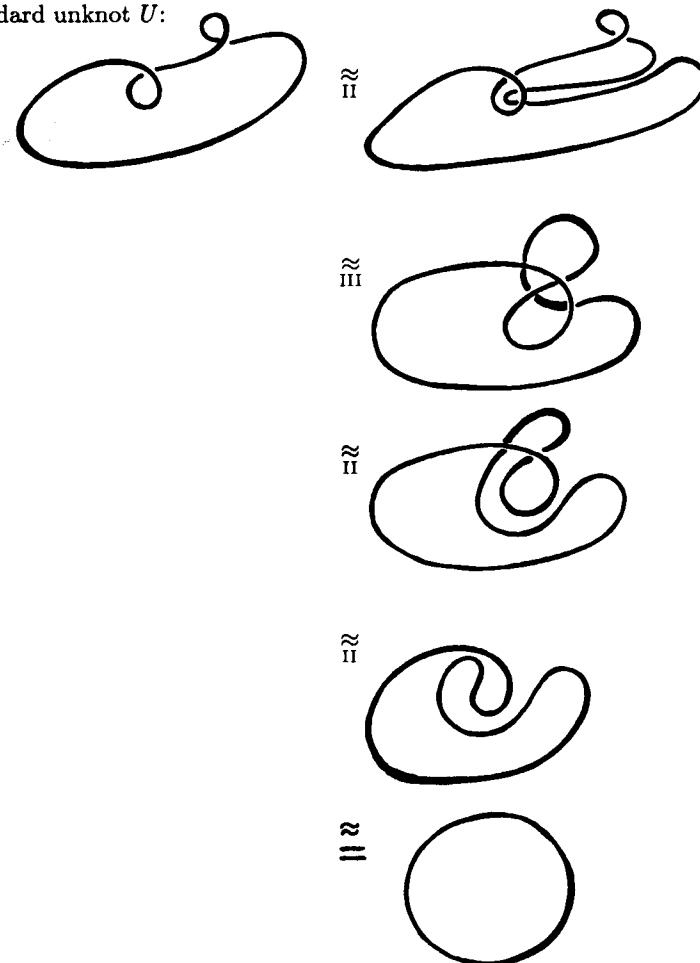
We then have the **Reidemeister Moves** for knot and link diagrams - as shown in Figure 8.



The Reidemeister Moves

Figure 8

The three moves on diagrams (I, II, III) change the graphical structure of the diagram while leaving the topological type of the embedding of the corresponding knot or link the same. That is, we understand that each move is to be performed locally on a diagram without changing that part of the diagram not depicted in the move. For example, here is a sequence of moves from a diagram K to the standard unknot U :



In this sequence I have only used the moves II and III (and the “zero move”). The equivalence relation generated by II and III is called **regular isotopy**. Because

it is possible to cancel curls () of opposite type - as shown above - I single out the regular isotopy relation.

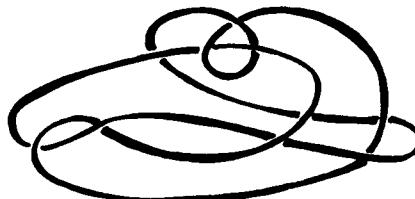
The equivalence relation on diagrams that is generated by all four Reidemeister moves is called **ambient isotopy**. In fact, Reidemeister [REI] showed that two knots or links in three-dimensional space can be deformed continuously one into the other (the usual notion of ambient isotopy) if and only if any diagram (obtained by projection to a plane) of one link can be transformed into a diagram for the other link via a sequence of Reidemeister moves (). Thus these moves capture the full topological scenario for links in three-space.

For a modern proof of Reidemeister's Theorem, see [BZ].

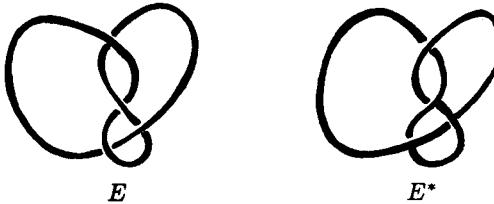
In order to become familiar with the **diagrammatic formal system** of these link diagrams together with the moves, it is very helpful to make drawings and exercises. Here are a few for starters:

Exercises.

1. Give a sequence of Reidemeister moves (ambient isotopy) from the knot below to the unknot.

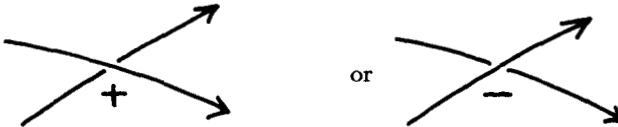


2. Give a sequence of Reidemeister moves from the figure eight knot E to its mirror image E^* .



3. A knot or link is said to be **oriented** if each arc in its diagram is assigned a

direction so that at each crossing the orientations appear either as



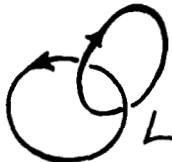
Note that each of these oriented crossings has been labelled with a sign of plus (+) or minus (-). Call this the **sign of the (oriented) crossing**.

Let $L = \{\alpha, \beta\}$ be a link of two components α and β . Define the **linking number** $\ell k(L) = \ell k(\alpha, \beta)$ by the formula

$$\ell k(\alpha, \beta) = \frac{1}{2} \sum_{p \in \alpha \cap \beta} \epsilon(p)$$

where $\alpha \cap \beta$ denotes the set of crossings of α with β (no self-crossings) and $\epsilon(p)$ denotes the sign of the crossing.

Thus



$$\ell k(L) = \frac{1}{2} (1 + 1) = 1$$

and



$$\ell k(L') = -1.$$

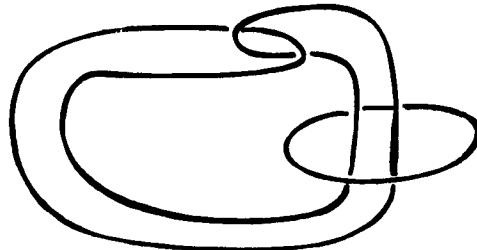
Prove: If L_1 and L_2 are oriented two-component link diagrams and if L_1 is ambient isotopic to L_2 , then $\ell k(L_1) = \ell k(L_2)$. (Check the oriented picture for oriented moves.)

4. Let K be any oriented link diagram. Let the **writhe of K** (or **twist number of K**) be defined by the formula $w(K) = \sum_{p \in C(K)} \epsilon(p)$ where $C(K)$ denotes the

set of crossings in the diagram K . Thus $w(\text{trefoil}) = +3$.

Show that regularly isotopic links have the same writhe.

5. Check that the link W below has zero linking number - no matter how you orient its components.



6.



The Borromean rings (shown here and in Figure 7)

have the property that they are linked, but the removal of any component leaves two unlinked rings. Create a link of 4 components that is linked, and so that the removal of any component leaves three unlinked rings. Generalize to n components.

7. What is wrong with the following argument? The trefoil T



has no Reidemeister moves except ones that make the diagram more complex such as



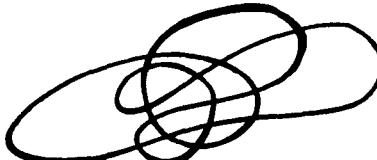
Thus the diagram T is in a minimal form, and therefore T is truly knotted (not equivalent to the unknot).

8. What is lacking in the following argument? The trefoil T has writhe $w(T) = +3$ (independent of how T is oriented), while its mirror image T^* has writhe $w(T^*) = -3$. Therefore T is not ambient isotopic to T^* .
9. A diagram is said to be **alternating** if the pattern of over and undercrossings

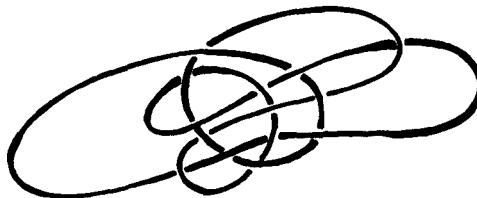
alternates as one traverses a component.



Take any knot or link shadow



Attempt to draw an alternating diagram that overlies this shadow.



Will this always work? When are alternating knot diagrams knotted? (Investigate empirically.) Does every knot have an alternating diagram? (Answer: No. The least example has eight crossings.)

Discussion.

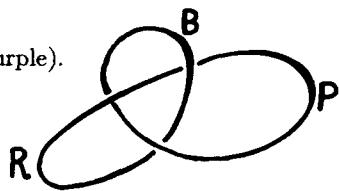
Exercises 1. → 4. are integral to the rest of the work. I shall assume that the reader has done these exercises. I shall comment later on 5. and 6. The fallacy in 7. derives from the fact that it is possible to have a diagram that represents an unknot admitting no simplifying moves.

We shall discuss 8. and 9. in due course.

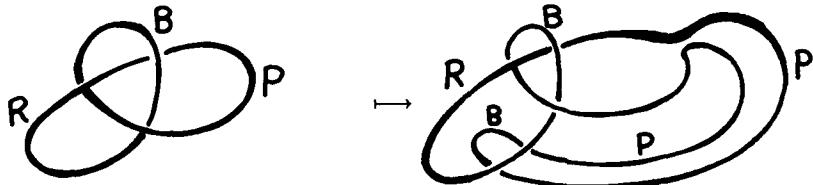
The Trefoil is Knotted.

I conclude this section with a description of how to prove that the trefoil is knotted. Here is a trefoil diagram with its arcs colored (labelled) in three distinct

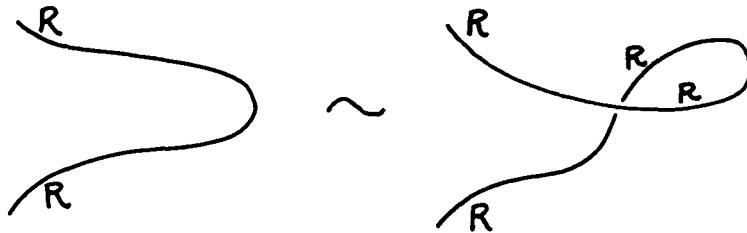
colors (R -red, B -blue, P -purple).



I claim that, with an appropriate notion of coloring, this property of being three-colored can be preserved under the Reidemeister moves. For example:



However, note that under a type I move we may be forced to retain only one color at a vertex:

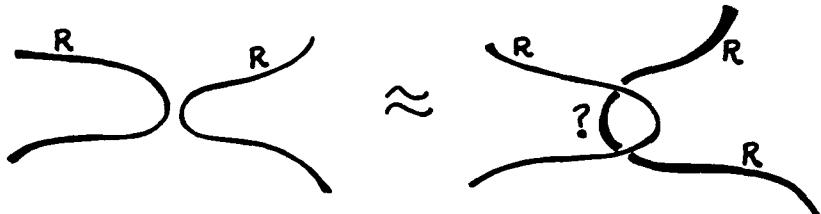


Thus I shall say that a knot diagram K is **three-colored** if each arc in K is assigned one of the three colors (R, B, P), all three colors occur on the diagram and each crossing carries either **three colors or one color**. (Two-colored crossings are not allowed.)

It follows at once from these coloring rules, that we can never transform a singly colored diagram to a three-colored diagram by local changes

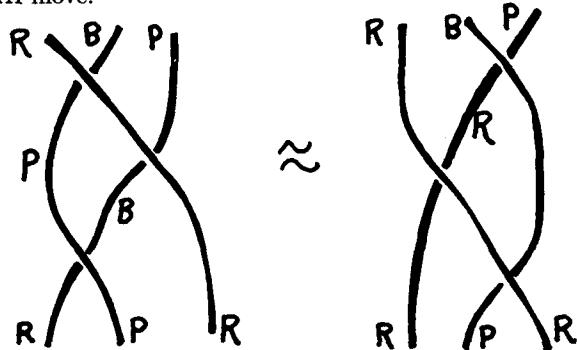


(obeying the three or one rule). For example,



The arc in question must be colored R since two R 's already occur at each neighboring vertex.

Each Reidemeister move affords an opportunity for a local color change - by coloring new arcs created by the move, or deleting some colored arcs. Consider the type III move:



Note that in preserving the (non-local) inputs (R, B, P) and output (R, P, R) we were forced to introduce a single-color vertex in the resultant of the triangle move. Check (exercise) that all color inputs do re-configure under the type III move - and that three-coloration is preserved. Finally, you must worry about the following scenario:

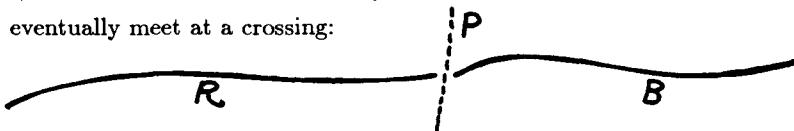


Upon performing a simplifying type II move, a colored arc is lost! Could we also

lose three-coloration as in the link below?



Well, knots do not allow this difficulty. In a knot the two arcs labelled *R* and *B* must eventually meet at a crossing:



The transversal arc at this crossing will be colored *P* in a three-colored diagram. Thus, while purple could be lost from the local arc, it will necessarily occur elsewhere in the diagram.

This completes the sketch of the proof that the trefoil is knotted. If *T* were ambient isotopic to the unknot, then the above observations plus the sequence of Reidemeister moves from *T* to unknot would yield a proof that $3 = 1$ and hence a contradiction. //

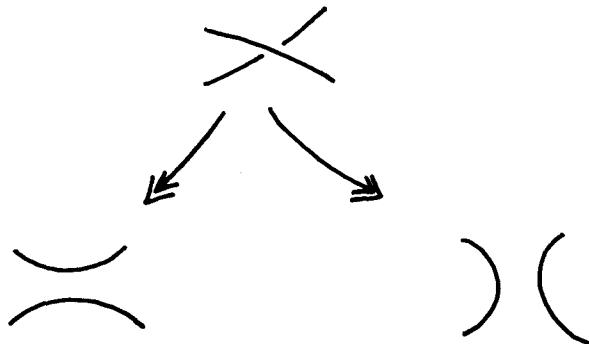
A coloration could be called a **state** of the knot diagram in analogy to the energetic states of a physical system. In this case the system admits topological deformations, and in the case of three-coloring, we have seen that there is a way to preserve the state structure as the system is deformed. Invariant properties of states then become topological invariants of the knot or link.

It is also possible to obtain topological invariants by considering all possible states (in some interpretation of that term-state) of a given diagram. Invariants emerge by summing (averaging) or integrating over the set of states for one diagram. We shall take up this viewpoint in the next section.

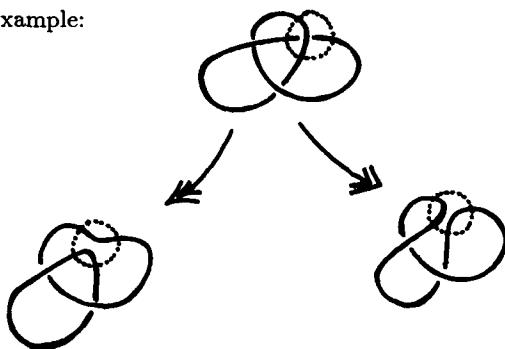
These two points of view - topological evolution of states versus integration over the space of states for a given system - appear to be quite complementary in studying the topology of knots and links. Keep watching this theme as we go along. The evolution of states is most closely related to the fundamental group of the knot and allied generalizations (section 13⁰). The integration over states is the fundamental theme of this book.

3⁰. States and the Bracket Polynomial.

Consider a crossing in an unoriented link diagram:  Two associated diagrams can be obtained by **splicing the crossing**:



For example:



One can repeat this process, and obtain a whole family of diagrams whose one ancestor is the original link diagram. (Figure 9).

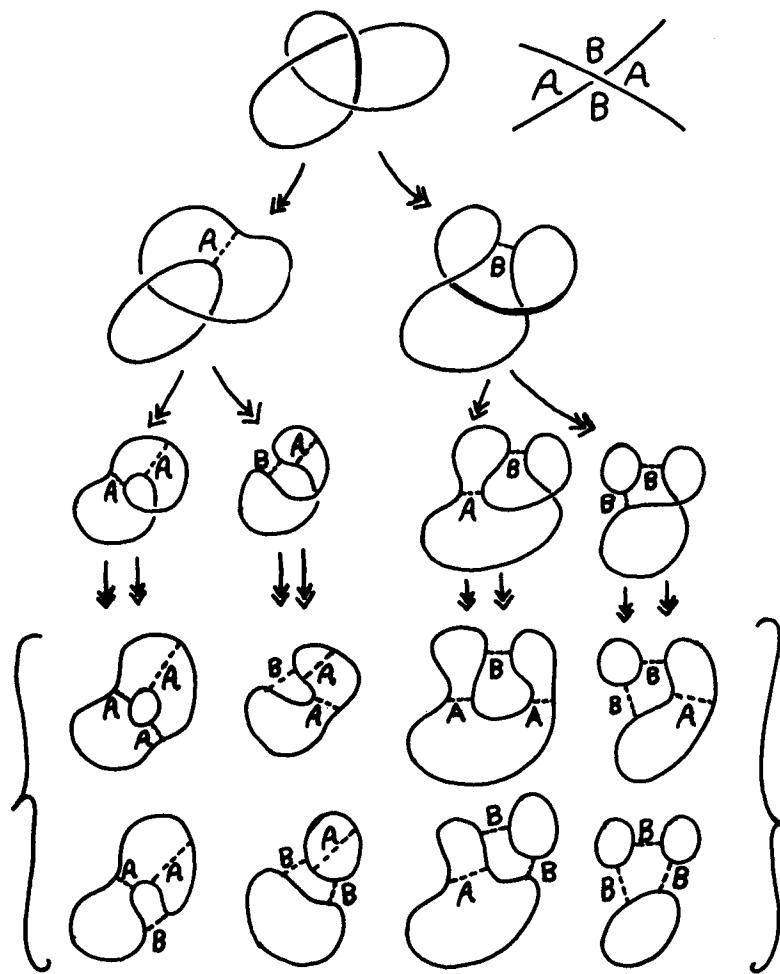
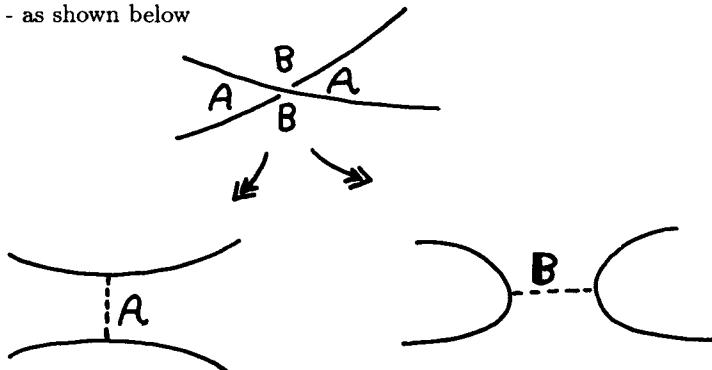


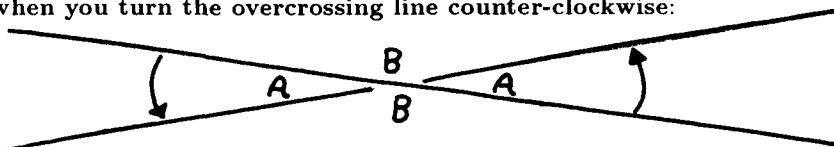
Figure 9

In Figure 9 I have indicated this splitting process with an indication of the type of split - as shown below

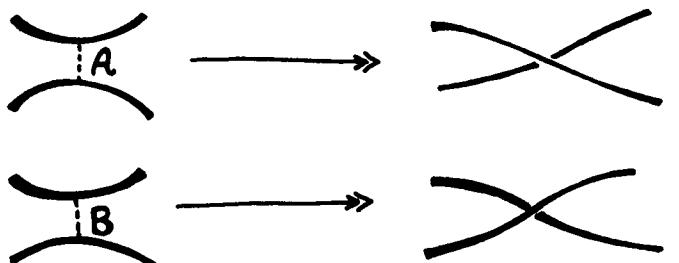


That is, a given split is said to be of **type A** or **type B** according to the convention that an **A-split** joins the regions labelled **A** at the crossing. The regions labelled **A** are those that appear on the **left** to an observer walking toward the crossing along the undercrossing segments. The **B-regions** appear on the right for this observer.

Another way to specify the **A-regions** is that the **A-regions are swept out when you turn the overcrossing line counter-clockwise**:

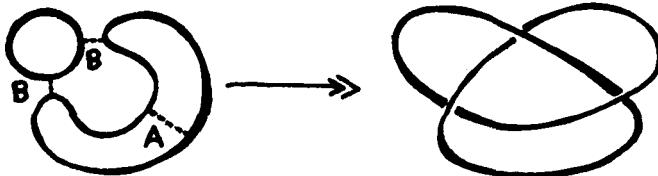


In any case, a split site labelled **A** or **B** can be **reconstructed** to form its ancestral crossing:



Therefore, by keeping track of these **A's** and **B's** we can reconstruct the ancestor link from any of its descendants. The most primitive descendants are collections

of Jordan curves in the plane. Here all crossings have been spliced. In Figure 9 we see eight such descendants. A sample reconstruction is:



Call these final descendants of the given knot or link K the **states** of K . Each state (with its labelling) can be used to reconstruct K . We shall construct invariants of knots and links by averaging over these states.

The specific form of the averaging is as follows: Let σ be a state of K . Let $\langle K | \sigma \rangle$ denote the product (commutative labels) of the labels attached to σ . Thus

$$\langle \text{trefoil} \mid \text{trefoil} \rangle = A^3$$

[Note that we deduce the labels from the structure of the state in relation to K .] Let $\|\sigma\|$ denote one less than the number of loops in σ . Thus

$$\left\| \text{figure-eight} \right\| = 2 - 1 = 1.$$

Definition 3.1. We define the **bracket polynomial** by the formula

$$\langle K \rangle = \langle K \rangle (A, B, d) = \sum_{\sigma} \langle K | \sigma \rangle d^{\|\sigma\|}$$

where A , B , and d are commuting algebraic variables. Here σ runs over all the states of K .

Remark. The bracket state summation is an analog of a partition function [BA1] in discrete statistical mechanics. In fact, for appropriate choices of A , B , d the bracket can be used to express the partition function for the Potts model. See Part II sections 7⁰ and 8⁰ for this connection.

Example. From Figure 9 we see that the bracket polynomial for the trefoil diagram is given by the formula:

$$\begin{aligned}\langle K \rangle &= A^3 d^{2-1} + A^2 B d^{1-1} + A^2 B d^{1-1} + AB^2 d^{2-1} + A^2 B d^{1-1} + AB^2 d^{2-1} \\ &\quad AB^2 d^{2-1} + B^3 d^{3-1} \\ \langle K \rangle &= A^3 d^1 + 3A^2 B d^0 + 3AB^2 d^1 + B^3 d^2.\end{aligned}$$

This bracket polynomial is not a topological invariant as it stands. We investigate how it behaves under the Reidemeister moves - and determine conditions on A , B and d for it to become an invariant.

Proposition 3.2.

$$\langle \text{X} \rangle = A \langle \text{--} \rangle + B \langle \text{O O} \rangle$$

Remark. The meaning of this statement rests in regarding each small diagram as part of a larger diagram, so that the three larger diagrams are identical except at the three local sites indicated by the small diagrams. Thus a special case of Proposition 3.2 is

$$\langle \text{OO} \rangle = A \langle \text{O} \text{O} \rangle + B \langle \text{O} \text{O} \rangle$$

The labels A and B label A and B - splits, respectively.

Proof. Since a given crossing can be split in two ways, it follows that the states of a diagram K are in one-to-one correspondence with the union of the states of K' and K'' where K' and K'' are obtained from K by performing A and B splits at a given crossing in K . It then follows at once from the definition of $\langle K \rangle$ that $\langle K \rangle = A\langle K' \rangle + B\langle K'' \rangle$. This completes the proof of the proposition. //

Remark. The above proof actually applies to a more general bracket of the form

$$\langle K \rangle = \sum_{\sigma} \langle K | \sigma \rangle \langle \sigma \rangle$$

where $\langle \sigma \rangle$ is any well-defined state evaluation. Here we have used $\langle \sigma \rangle = d^{\|\sigma\|}$ as above. We shall see momentarily that this form of state-evaluation is demanded by the topology of the plane.

Remark. Proposition 3.2 can be used to compute the bracket. For example,

$$\begin{aligned}
\langle \text{OO} \rangle &= A \langle \text{OO} \rangle + B \langle \text{OO} \rangle \\
&= A \left\{ A \langle \text{OO} \rangle + B \langle \text{OO} \rangle \right\} + \\
&\quad B \left\{ A \langle \text{OO} \rangle + B \langle \text{OO} \rangle \right\} \\
&= A^2 d^{2-1} + AB d^{1-1} + BA d^{1-1} + B^2 d^{2-1} \\
&= A^2 d^1 + 2AB d^0 + B^2 d^1.
\end{aligned}$$

Proposition 3.3.

$$\begin{aligned}
(a) \quad \langle \text{OC} \rangle &= AB \langle \text{CC} \rangle + AB \langle \text{OC} \rangle \\
&\quad + (A^2 + B^2) \langle \text{--} \rangle \\
(b) \quad \langle \text{O} \text{--} \rangle &= (Ad + B) \langle \text{--} \rangle \\
\langle \text{--} \sigma \rangle &= (A + Bd) \langle \text{--} \rangle
\end{aligned}$$

Proof. (a)

$$\begin{aligned}
 \langle \text{Diagram} \rangle &= A \langle \text{Diagram} \rangle + B \langle \text{Diagram} \rangle \\
 &= A \left\{ A \langle \text{Diagram} \rangle + B \langle \text{Diagram} \rangle \right\} + \\
 &\quad B \left\{ A \langle \text{Diagram} \rangle + B \langle \text{Diagram} \rangle \right\} \\
 &= AB \langle \text{Diagram} \rangle + AB \langle \text{Diagram} \rangle \\
 &\quad + (A^2 + B^2) \langle \text{Diagram} \rangle.
 \end{aligned}$$

Part (b) is left for the reader. Note that $\langle \text{Diagram} \rangle = d \langle \text{Diagram} \rangle$ and, in general, $\langle OK \rangle = d \langle K \rangle$ where OK denotes any addition of a disjoint circle to the diagram K . Thus

$$\langle \text{Diagram} \rangle = d \langle \text{Diagram} \rangle. \quad //$$

Corollary 3.4. If $B = A^{-1}$ and $d = -A^2 - A^{-2}$, then

$$(a) \quad \langle \text{Diagram} \rangle = \langle \text{Diagram} \rangle$$

$$(b) \quad \langle \text{Diagram} \rangle = (-A^3) \langle \text{Diagram} \rangle$$

$$\langle \text{Diagram} \rangle = (-A^{-3}) \langle \text{Diagram} \rangle.$$

Proof. The first part follows at once from 3.3. The second part follows from 3.3 and the calculation $Ad + B = A(-A^2 - A^{-2}) + A^{-1} = -A^3$. //

Remark. The formula 3.3 (a) shows that just on the basis of the assumptions

$$\langle \cancel{\infty} \rangle = A \langle \overbrace{\infty} \rangle + B \langle \circlearrowleft \circlearrowright \rangle$$

and

$$\langle \cancel{\circlearrowleft} \rangle = \langle \circlearrowleft \circlearrowright \rangle,$$

we need that $AB = 1$ and

$$AB \langle \cancel{\circlearrowleft} \rangle = (-A^2 - B^2) \langle \overbrace{\infty} \rangle$$

whence

$$\langle \cancel{\circlearrowleft} \rangle = -\left(\frac{A}{B} + \frac{B}{A}\right) \langle \overbrace{\infty} \rangle$$

or

$$\langle \cancel{\circlearrowleft} \rangle = d \langle \overbrace{\infty} \rangle.$$

This shows that the rule we started with for evaluating all simple closed curves is necessary for type II invariance. Thus from the topological viewpoint, the bracket unfolds completely from the recursive relation

$$\langle \cancel{\infty} \rangle = A \langle \overbrace{\infty} \rangle + B \langle \circlearrowleft \circlearrowright \rangle.$$

We have shown that $\langle K \rangle$ can be adjusted to be invariant under II. In fact,

Proposition 3.5. Suppose $B = A^{-1}$ and $d = -A^2 - A^{-2}$ so that

$$\langle \cancel{\circlearrowleft} \rangle = \langle \circlearrowleft \circlearrowright \rangle. \text{ Then}$$

$$\langle \cancel{\cancel{\infty}} \rangle = \langle \cancel{\cancel{\infty}} \rangle.$$

Proof.

$$\langle \cancel{\cancel{\infty}} \rangle = A \langle \cancel{\circlearrowleft} \cancel{\circlearrowright} \rangle + B \langle \cancel{\circlearrowleft} \cancel{\circlearrowright} \rangle$$

$$= A \langle \overbrace{\infty} \rangle + B \langle \cancel{\circlearrowleft} \cancel{\circlearrowright} \rangle$$

$$= A \langle \cancel{\circlearrowleft} \cancel{\circlearrowright} \rangle + B \langle \cancel{\circlearrowleft} \cancel{\circlearrowright} \rangle$$

$$= \langle \cancel{\cancel{\infty}} \rangle.$$

//

The bracket with $B = A^{-1}$, $d = -A^2 - A^{-2}$ is invariant under the moves II and III. If we desire an invariant of ambient isotopy (I, II and III), this is obtained by a normalization:

Definition 3.6. Let K be an oriented link diagram. Define the **writhe of K** , $w(K)$, by the equation $w(K) = \sum_p \epsilon(p)$ where p runs over all crossings in K , and $\epsilon(p)$ is the sign of the crossing:



(Compare with exercises 3 and 4). Note that $w(K)$ is an invariant of regular isotopy (II, III) and that

$$\begin{cases} w(\text{---}) = 1 + w(\text{---}) \\ w(\text{---}) = 1 + w(\text{---}) \end{cases}$$

$$\begin{cases} w(\text{---}) = -1 + w(\text{---}) \\ w(\text{---}) = -1 + w(\text{---}) \end{cases}$$

Thus, we can define a **normalized bracket**, \mathcal{L}_K , for oriented links K by the formula

$$\mathcal{L}_K = (-A^3)^{-w(K)} \langle K \rangle$$

and we have

Proposition 3.7. The normalized bracket polynomial \mathcal{L}_K is an invariant of ambient isotopy.

Proof. Since $w(K)$ is a regular isotopy invariant, and $\langle K \rangle$ is also a regular isotopy invariant, it follows at once that \mathcal{L}_K is a regular isotopy invariant. Thus we need only check that \mathcal{L}_K is invariant under type I moves. This follows at once. For

example

$$\begin{aligned}
 \mathcal{L}_{\text{double}} &= (-A^3)^{-w} (\text{double}) \langle \text{double} \rangle \\
 &= (-A^3)^{-[1+w]} (\text{single})] (-A^3) \langle \text{single} \rangle \\
 &= (-A^3)^{-w} (\text{single}) \langle \text{single} \rangle \\
 &= \mathcal{L}_{\text{single}} //
 \end{aligned}$$

Finally, we have the particularly important behavior of $\langle K \rangle$ and \mathcal{L}_K under mirror images:

[Note: Unless otherwise specified, we assume that $B = A^{-1}$ and $d = -A^2 - A^{-2}$ in the bracket.]

Proposition 3.8. Let K^* denote the mirror image of the (oriented) link K that is obtained by switching all the crossings of K . Then $\langle K^* \rangle(A) = \langle K \rangle(A^{-1})$ and

$$\mathcal{L}_{K^*}(A) = \mathcal{L}_K(A^{-1}).$$

Proof. Reversing all crossings exchanges the roles of A and A^{-1} in the definition of $\langle K \rangle$ and \mathcal{L}_K . QED //

Remark. In section 5⁰ (Theorem 5.2) we show that \mathcal{L}_K is the original Jones polynomial after a change of variable.

Examples. It follows from 3.8 that if we calculate \mathcal{L}_K and find that $\mathcal{L}_K(A) \neq \mathcal{L}_K(A^{-1})$, then K is not ambient isotopic to K^* . Thus \mathcal{L}_K has the potential to detect chirality. In fact, this is the case with the trefoil knot as we shall see.

$$\begin{aligned}
 \text{(i)} \quad \langle \text{trefoil} \rangle &= A \langle \text{double} \rangle + A^{-1} \langle \text{double} \rangle \\
 &= A(-A^3) + A^{-1}(-A^{-3}) \\
 \langle L \rangle &= -A^4 - A^{-4}.
 \end{aligned}$$

$$\text{(ii)} \quad \langle \text{Trefoil} \rangle = A \langle \text{Trefoil} \rangle + A^{-1} \langle \text{Trefoil} \rangle \\ = A(-A^4 - A^{-4}) + A^{-1}(-A^{-3})^2$$

$$\langle T \rangle = -A^5 - A^{-3} + A^{-7}$$

$w(T) = 3$ (independent of the choice of orientation

since T is a knot)

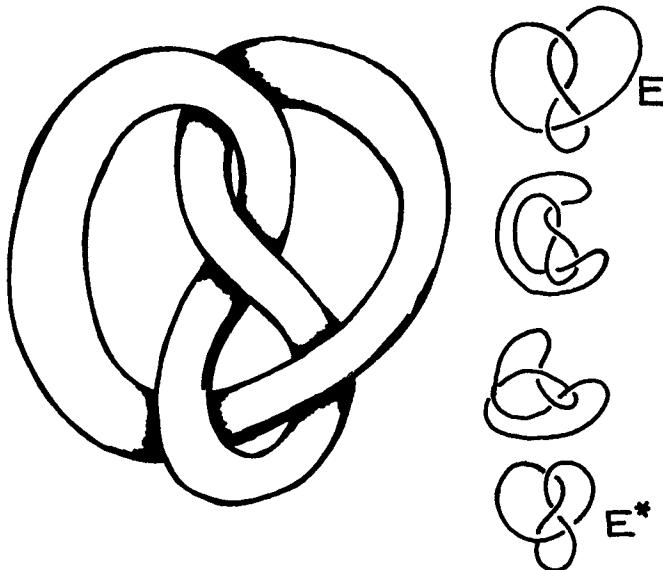
$$\therefore \mathcal{L}_T = (-A^3)^{-3} \langle T \rangle \\ = -A^{-9}(-A^5 - A^{-3} + A^{-7})$$

$$\therefore \mathcal{L}_T = A^{-4} + A^{-12} - A^{-16}$$

$$\therefore \mathcal{L}_{T^*} = A^4 + A^{12} - A^{16}.$$

Since $\mathcal{L}_{T^*} \neq \mathcal{L}_T$, we conclude that the trefoil is not ambient isotopic to its mirror image. Incidentally, we have also shown that the trefoil is knotted and that the link L is linked.

(iii) (The Figure Eight Knot)



$$\begin{aligned}
 \langle E \rangle &= A \left\langle \text{Diagram } 1 \right\rangle + A^{-1} \left\langle \text{Diagram } 2 \right\rangle \\
 &= A(-A^3) \left\langle \text{Diagram } 3 \right\rangle + A^{-1} \left\langle \text{Diagram } 4 \right\rangle \\
 &= -A^4[-A^4 - A^{-4}] + A^{-1}[-A^5 - A^{-3} + A^{-7}] \\
 &= A^8 + 1 - A^4 - A^{-4} + A^{-8} \\
 \langle E \rangle &= A^8 - A^4 + 1 - A^{-4} + A^{-8}.
 \end{aligned}$$

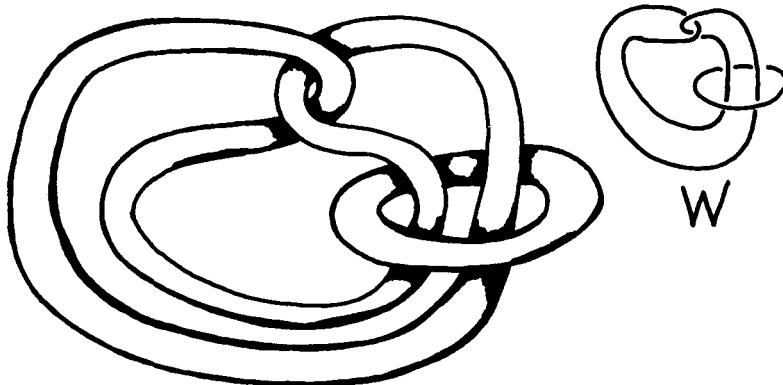
Since $w(E) = 0$, we see that

$$\mathcal{L}_E = \langle E \rangle \text{ and } \mathcal{L}_{E^*} = \mathcal{L}_E.$$

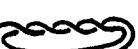
In fact, E is ambient isotopic to its mirror image.

(iv) W (The Whitehead Link).

The Whitehead link has linking number zero however one orients it. A bracket calculation shows that it is linked:



$$\begin{aligned}
\langle W \rangle &= A \left\langle \text{(knot diagram)} \right\rangle + A^{-1} \left\langle \text{(knot diagram)} \right\rangle \\
&= A \left\langle \text{(knot diagram)} \right\rangle + (A^{-1})(-A^{-3}) \left\langle \text{(knot diagram)} \right\rangle \\
&= A(-A^3) \left\langle \text{(knot diagram)} \right\rangle - A^{-4} \left\langle \text{(knot diagram)} \right\rangle \\
&= (-A^4) \left\langle \text{(knot diagram)} \right\rangle - A^{-4} \langle E^* \rangle \\
&= (-A^4)(-A^{-3}) \langle T^* \rangle - A^{-4} \langle E^* \rangle \\
\\
&= (+A^1)[-A^{-5} - A^3 + A^7] - A^{-4}[A^8 - A^4 + 1 - A^{-4} + A^{-8}] \\
\\
&= -A^{-4} - A^4 + A^8 - A^4 + 1 - A^{-4} + A^{-8} - A^{-12} \\
\\
\langle W \rangle &= A^8 + A^{-8} - 2A^4 - 2A^{-4} + 1 - A^{-12}
\end{aligned}$$

(v)     ... 

Here K_n is a **torus link of type $(2, n)$** . (I will explain the terminology below.)

$$\begin{aligned}
\left\langle \text{(knot diagram)} \right\rangle &= A \left\langle \text{(knot diagram)} \right\rangle \\
&\quad + A^{-1} \left\langle \text{(knot diagram)} \right\rangle \\
\Rightarrow \langle K_n \rangle &= A \langle K_{n-1} \rangle + A^{-1} (-A^{-3})^{n-1}.
\end{aligned}$$

Thus $\boxed{\langle K_n \rangle = A\langle K_{n-1} \rangle + (-1)^{n-1}A^{-3n+2}}$

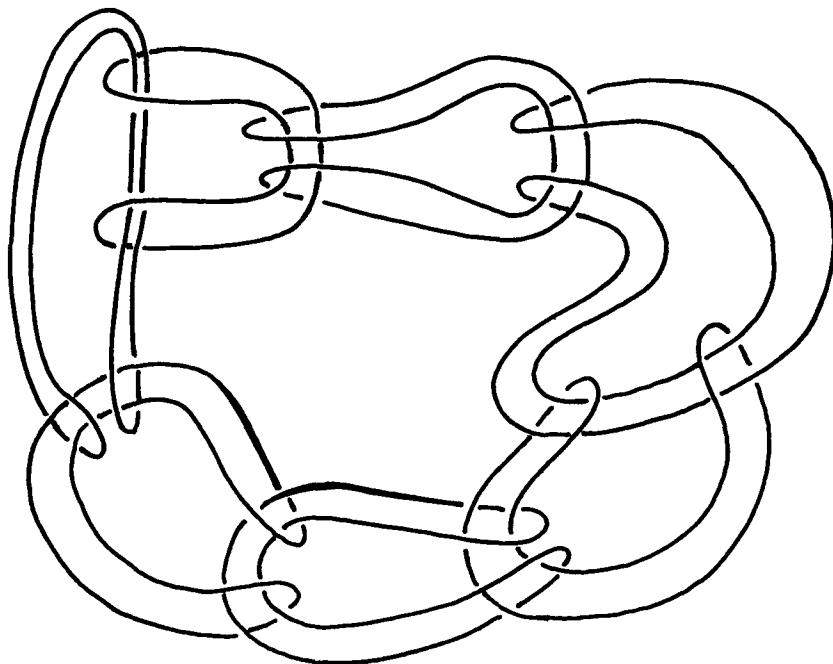
$$\langle K_1 \rangle = -A^3$$

$$\langle K_2 \rangle = A(-A^3) + (-1)^1 A^{-3 \cdot 2 + 2} = -A^4 - A^{-4}$$

$$\langle K_3 \rangle = A(-A^4 - A^{-4}) + (-1)^2 A^{-7} = -A^5 - A^{-3} + A^{-7}$$

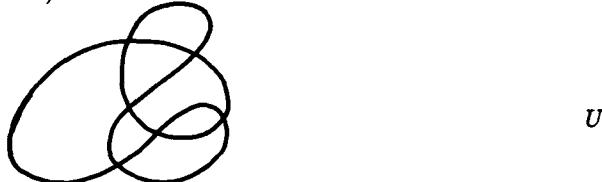
$$\langle K_4 \rangle = A(-A^5 - A^{-3} + A^{-7}) + (-1)^3 A^{-10} = -A^6 - A^{-2} + A^{-6} - A^{-10}.$$

Use this procedure to show that no torus knot of type $(2, n)$ ($n > 1$) is ambient isotopic to its mirror image.

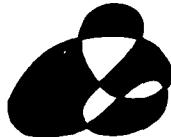


4⁰. Alternating Links and Checkerboard Surfaces.

This is a long example. To begin with, take any link shadow. (No over or undercrossings indicated.)

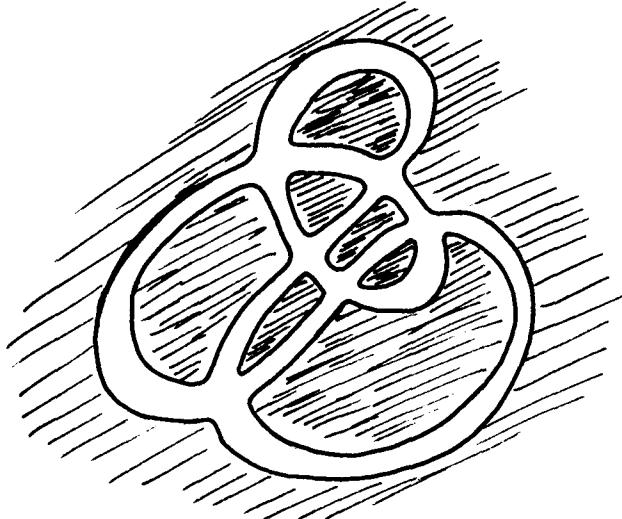


Shade the regions with two colors (black and white) so that the outer region is shaded white, and so that regions sharing an edge receive opposite colors.

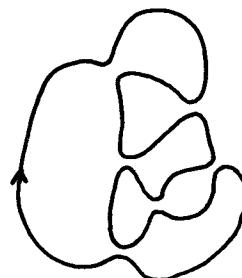
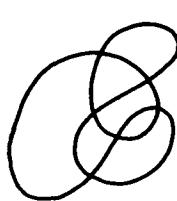
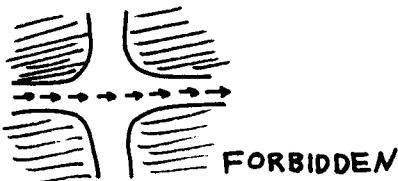


You find that this shading is possible in any example that you try. Prove that it always works!

While you are working on that one, here is another puzzle. Thicken the shadow U so that it looks like a network of roadways with street-crossings:

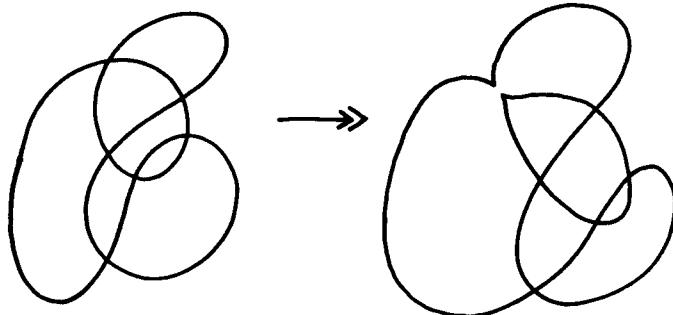


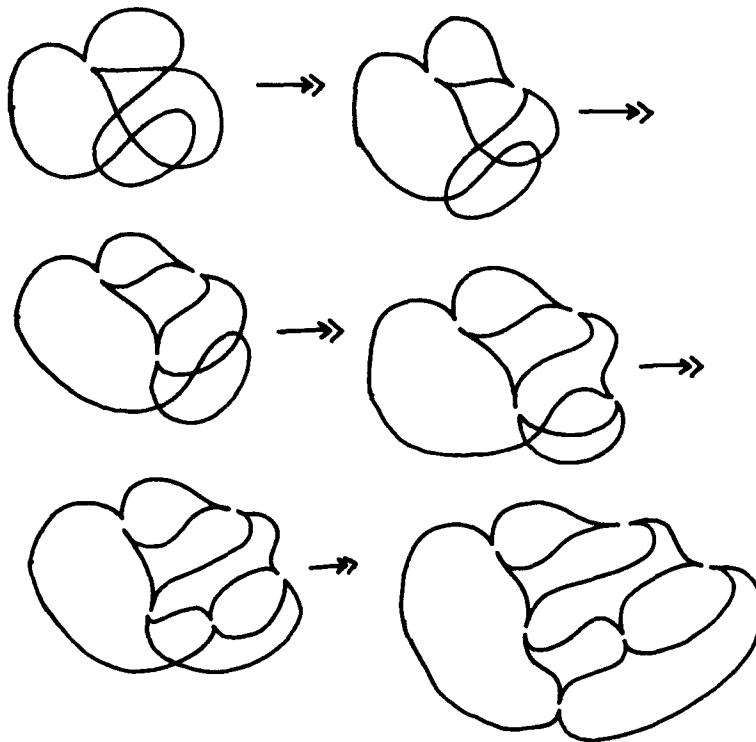
Now, find a route that traverses every street in this map **once**, and so that it always turns at each cross-street.



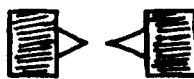
You see that the roadway problem can be rephrased as: **split each crossing of the diagram so that the resulting state has a single component**.

Now it is easy to prove that there exists such a splitting. Just start splitting the diagram making sure that you maintain connectivity at each step. For example





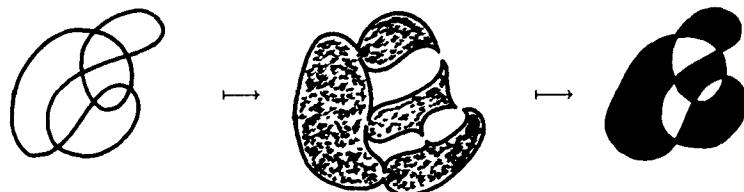
It is always possible to maintain connectivity. For suppose that $\rightarrow \leftarrow$ is disconnected. Then $\rightarrow \leftarrow$ must have the form



and hence is connected. Here, I am implicitly using the **Jordan Curve Theorem** - that a simple closed curve divides the plane into two connected pieces.

Knowing that the **roadway problem** can be solved, we in fact know that

every link diagram can be two-colored. The picture is as follows:



Split the diagram to obtain a simple closed curve. Shade the inside of the curve. Then shade the corresponding regions of the original link diagram!

Knowing that link diagrams can be two-colored, we are in a position to prove the fundamental.

Theorem 4.1. Let U be any link shadow. Then there is a choice of over/under structure for the crossings of U forming a diagram K so that K is **alternating**. (A diagram is said to be alternating if one alternates from over to under to over when travelling along the arcs of the diagram.)

Proof. Shade the diagram U in two colors and set each crossing so that it has the form

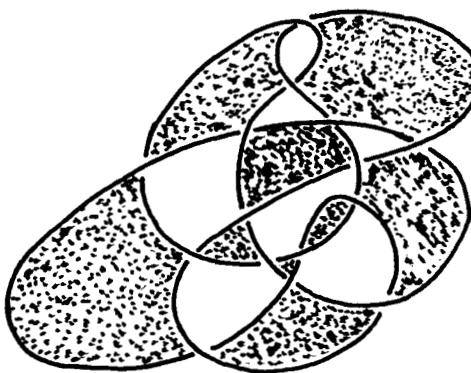


that is - so that the A -regions at this crossing are shaded. The picture below should convince you that K (as set above) is alternating:



This completes the proof. //

Example.



Now we come to the center of this section. Consider the bracket polynomial, $\langle K \rangle$, for an alternating link diagram K . If we shade K as in the proof above, so that every pair of A -regions is shaded, then the state S obtained by splicing each shading



will contribute

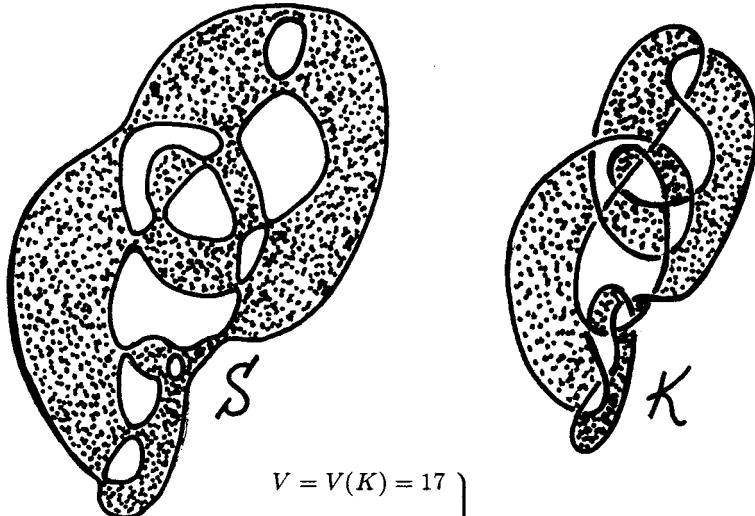
$$A^{\#(\text{crossings})}(-A^2 - A^{-2})^{\ell(S)-1}$$

where $\ell(S)$ is the number of loops in this state. Let $V(K)$ denote the number of crossings in the diagram K . Thus the **highest power term contributed by S is**

$$(-1)^{\ell(S)} A^{V(K)+2\ell(S)-2}.$$

I claim that **this is the highest degree term in $\langle K \rangle$ and that it occurs with exactly this coefficient $(-1)^{\ell(S)}$.**

In order to see this assertion, take a good look at the state S :



$$\left. \begin{array}{l} V = V(K) = 17 \\ W = W(K) = 10 \\ B = B(K) = 9 \\ R = R(K) = 19 \end{array} \right\}$$

By construction we see that

$$\|S\| = \ell(S) - 1 = W - 1$$

where $W = W(K)$ is the number of white (unshaded) regions in the two-coloring of the diagram. Thus we are asserting that the maximum degree of the bracket is given by the formula

$$\max \deg \langle K \rangle = V(K) + 2W(K) - 2$$

where $V(K)$ denotes the number of crossings in K , and $W(K)$ denotes the number of white regions in the checkerboard shading of K .

To see the truth of this formula, consider any other state S' of K . Then S' can be obtained from S by switching some of the splices in S []. I am assuming that the diagram K is reduced. This means that K is not in the form [] with a 2-strand bridge between pieces that contain

crossings. It follows from this assumption, and the construction of S , that if S' is obtained from S by one switch then $\|S'\| = \|S\| - 1$. Hence, if S contributes $A^{V(K)}(-A^2 - A^{-2})^{\|S\|}$, then S' contributes $A^{V(K)-2}(-A^2 - A^{-2})^{\|S\|-1}$ and so the largest degree contribution from S' is 4 less than the largest degree contribution from S . It is then easy to see that no state obtained by further switching S' can ever get back up to the maximal degree for S . This simple argument proves our assertion about the maximal degree.

By the same token, the minimal degree is given by the formula

$$\min \deg\langle K \rangle = -V(K) - 2B(K) + 2$$

where $B(K)$ is the number of black regions in the shading. Formally, we have proved.

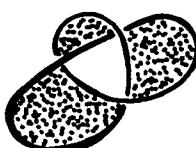
Proposition 4.2. Let K be a reduced alternating link diagram, shaded in (white and black) checkerboard form with the unbounded region shaded white. Then the maximal and minimal degrees of $\langle K \rangle$ are given by the formulas

$$\max \deg\langle K \rangle = V + 2W - 2$$

$$\min \deg\langle K \rangle = -V - 2B + 2$$

where V is the number of crossings in the diagram K , W is the number of white regions, B is the number of black regions.

For example, we have seen that the right-handed trefoil knot T has bracket $\langle T \rangle = -A^5 - A^{-3} + A^{-7}$ and in the shading we have



$$V = 3$$

$$W = 2$$

$$B = 3$$

so that

$$V + 2W - 2 = 3 + 4 - 2 = 5$$

$$-V - 2B + 2 = -3 - 6 + 2 = -7$$

and these are indeed the maximal and minimal degrees of $\langle K \rangle$.

Definition 4.3. The **span** of an unoriented knot diagram K is the difference between the highest and lowest degrees of its bracket. Thus

$$\text{span}(K) = \max \deg\langle K \rangle - \min \deg\langle K \rangle.$$

Since $\mathcal{L}_K = (-A^3)^{-w(K)}\langle K \rangle$ is an ambient isotopy invariant, we know that **span(K) is an ambient isotopy invariant.**

Now let K be a reduced alternating diagram. We know that

$$\max \deg\langle K \rangle = V + 2W - 2$$

$$\min \deg\langle K \rangle = -V - 2B + 2.$$

Hence $\text{span}(K) = 2V + 2(W + B) - 4$.

However, $W + B = R$, the total number of regions in the diagram - and it is easy to see that $R = V + 2$. Hence

$$\text{span}(K) = 2V + 2(V + 2) - 4$$

$$\text{span}(K) = 4V.$$

Theorem 4.4. ([LK4], [MUR1]). Let K be a reduced alternating diagram. Then the number of crossings $V(K)$ in K is an ambient isotopy invariant of K .

This is an extraordinary application of the bracket (hence of the Jones polynomial). The topological invariance of the number of crossings was conjectured since the tabulations of Tait, Kirkman and Little in the late 1800's. The bracket is remarkably adapted to the proof.

Note that in the process we have (not surprisingly) shown that the reduced alternating diagrams with crossings represent non-trivial knots and links. See [LICK1] for a generalization of Theorem 4.4 to so-called adequate links.

Exercise 4.5. Using the methods of this section, obtain the best result that you can about the chirality of alternating knots and links. (Hint: If K^* is the mirror image of K , then $\mathcal{L}_{K^*}(A) = \mathcal{L}_K(A^{-1})$.)

Note that $\mathcal{L}_K(A) = (-A^3)^{-w(K)}\langle K \rangle$. Hence

$$\max \deg \mathcal{L}_K = -3w(K) + \max \deg\langle K \rangle$$

$$\min \deg \mathcal{L}_K = -3w(K) + \min \deg\langle K \rangle.$$

If $\mathcal{L}_K(A) = \mathcal{L}_K(A^{-1})$, then $\max \deg \mathcal{L}_K = -\min \deg \mathcal{L}_K$, whence

$$6w(K) = +\max \deg \langle K \rangle + \min \deg \langle K \rangle.$$

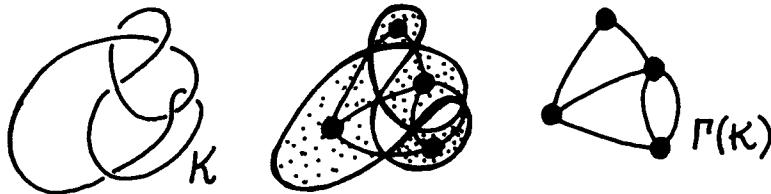
Thus if K is reduced and alternating, we have, by Proposition 4.2, that

$$6w(K) = 2(W - B)$$

whence $3w(K) = W - B$.

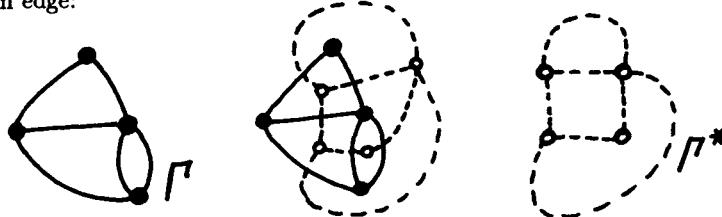
For example, if K is the trefoil  then $B = 3$, $W = 2$, $w(K) = 3$. Since $W - B = -1 \neq 9$ we conclude that K is not equivalent to its mirror image. We have shown that if K is reduced and alternating, then K achiral implies that $3w(K) = W - B$. One can do better than this result, but it is a good easy start. (Thistlethwaite [TH4] and Murasugi [MUR2] have shown that for K reduced and alternating, $w(K)$ is an ambient isotopy invariant.) Note also that in the event that $w(K) = 0$ and K is achiral, we have shown that $B = W$. See [LICK1] and [TH3] for generalizations of this exercise to adequate links.

Example 4.6. Define the graph $\Gamma(K)$ of a link diagram K as follows: Shade the diagram as a white/black checkerboard with the outer region shaded white. Choose a vertex for each black region and connect two vertices by an edge of $\Gamma(K)$ whenever the corresponding regions of K meet at a crossing.

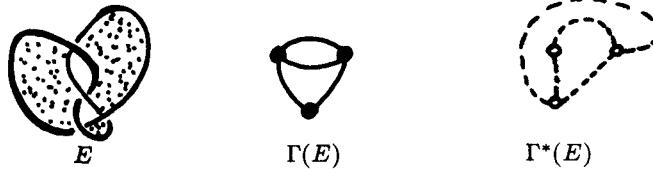


The corresponding graph $\Gamma^*(K)$ that is constructed from the white regions is the planar dual of $\Gamma(K)$. That is, $\Gamma^*(K)$ is obtained from Γ by assigning a vertex of Γ^* to each region of Γ , and an edge of Γ^* whenever two regions of Γ have a

common edge:

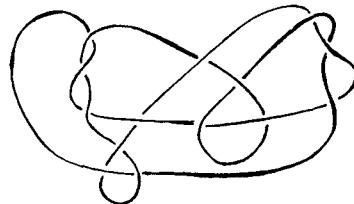


In the case of the figure eight knot E we see that $\Gamma(E)$ and $\Gamma^*(E)$ are isomorphic graphs:



The fact that these graphs are isomorphic is related to the fact that the knot E is ambient isotopic to its mirror image. In fact, if $\Gamma(K)$ and $\Gamma^*(K)$ are isomorphic graphs on the two-sphere S^2 (that is on the plane with a point at infinity added), then the associated **alternating** knots K and K^* (its mirror image) are ambient isotopic. (Exercise.) One possible converse to this statement is the **Mirror Conjecture**: If K is an alternating knot, and K is ambient isotopic to its mirror image, then there exists an alternating diagram K for K such that $\Gamma(K)$ and $\Gamma^*(K)$ are isomorphic graphs on the two-sphere S^2 .

Remark (1993). This conjecture has been shown to be false by Oliver Dasbach and Stefan Hougardy [Dasbach and Hougardy. (preprint 1993)]. Their counterexample is the 14 crossing alternating achiral knot shown below.



5⁰. The Jones Polynomial and its Generalizations.

The original 1-variable Jones polynomial was discovered [JO2] via a new representation of the Artin braid group. We shall see this in section 6⁰.

It is easy to show how the bracket polynomial gives rise to the Jones polynomial via the following definition.

Definition 5.1. The 1-variable Jones polynomial, $V_K(t)$, is a Laurent polynomial in the variable \sqrt{t} (i.e. finitely many positive and negative powers of \sqrt{t}) assigned to an oriented link K . The polynomial satisfies the properties:

- (i) If K is ambient isotopic to K' , then $V_K(t) = V_{K'}(t)$.
- (ii) $V_{\text{circle}} = 1$
- (iii) $t^{-1}V_{\text{overpass}} - tV_{\text{underpass}} = \left(\sqrt{t} - \frac{1}{\sqrt{t}}\right)V_{\text{link}}$

In the last formula, the three small diagrams stand for three larger link diagrams that differ only as indicated by the smaller diagrams.

For example, $V_{\text{G}} = V_{\text{G}} = V_{\text{G}} = 1$ by (i) and (ii). Therefore,

$$t^{-1} - t = \left(\sqrt{t} - \frac{1}{\sqrt{t}}\right)V_{\text{G}} \text{ by (iii).}$$

Let

$$\delta = \frac{t^{-1} - t}{\sqrt{t} - 1/\sqrt{t}} = \frac{(t^{-1} - t)(\sqrt{t} + 1/\sqrt{t})}{t - t^{-1}}$$

$$\therefore \delta = -(\sqrt{t} + 1/\sqrt{t}).$$

This definition gives sufficient information to compute the Jones polynomial, recursively on link diagrams. (We will discuss this shortly.) The definition is not obviously well-defined, nor is it obvious that such an invariant exists. Jones proved (i), (ii), (iii) as theorems about his invariant. However,

Theorem 5.2. Let $\mathcal{L}_K(A) = (-A^3)^{-w(K)} \langle K \rangle$ as in section 4⁰. Then

$$\mathcal{L}_K(t^{-1/4}) = V_K(t).$$

Thus the normalized bracket yields the 1-variable Jones polynomial.

Remark. In this context, I take this theorem to mean that if we let $V_K(t) = \mathcal{L}_K(t^{-1/4})$, then $V_K(t)$ satisfies the properties (i), (ii) and (iii) of Definition 5.1. Thus this theorem proves the existence (and well-definedness) of the 1-variable Jones polynomial. For the reader already familiar with $V_K(t)$ as satisfying (i), (ii), and (iii) the theorem draws the connection of this polynomial with the bracket.

Proof of 5.2. Keeping in mind that $B = A^{-1}$, we have the formulas for the bracket

$$\langle \overleftarrow{\times} \rangle = A \langle \overleftarrow{\equiv} \rangle + B \langle \circlearrowleft \circlearrowright \rangle$$

$$\langle \overrightarrow{\times} \rangle = B \langle \overleftarrow{\equiv} \rangle + A \langle \circlearrowleft \circlearrowright \rangle$$

Hence

$$B^{-1} \langle \overleftarrow{\times} \rangle - A^{-1} \langle \overrightarrow{\times} \rangle = \left(\frac{A}{B} - \frac{B}{A} \right) \langle \overleftarrow{\equiv} \rangle.$$

Thus

$$A \langle \overleftarrow{\times\!\times} \rangle - A^{-1} \langle \overrightarrow{\times\!\times} \rangle = (A^2 - A^{-2}) \langle \overrightarrow{\equiv} \rangle.$$

Let $w = w(\overrightarrow{\equiv})$ so that $w(\overrightarrow{\times\!\times}) = w + 1$ and $w(\overleftarrow{\times\!\times}) = w - 1$. Let $\alpha = -A^3$. Then

$$A \langle \overrightarrow{\times\!\times} \rangle \alpha^{-w} - A^{-1} \langle \overrightarrow{\times\!\times} \rangle \alpha^{-w} = (A^2 - A^{-2}) \langle \overrightarrow{\equiv} \rangle \alpha^{-w}.$$

Hence

$$A\alpha \langle \overrightarrow{\times\!\times} \rangle \alpha^{-(w+1)} - A^{-1}\alpha^{-1} \langle \overrightarrow{\times\!\times} \rangle \alpha^{-(w-1)} = (A^2 - A^{-2}) \langle \overrightarrow{\equiv} \rangle \alpha^{-w}$$

$$\begin{aligned} A\alpha \mathcal{L}_{\overrightarrow{\times\!\times}} &- A^{-1}\alpha^{-1} \mathcal{L}_{\overrightarrow{\times\!\times}} = (A^2 - A^{-2}) \mathcal{L}_{\overrightarrow{\equiv}} \\ -A^4 \mathcal{L}_{\overrightarrow{\times\!\times}} &+ A^{-4} \mathcal{L}_{\overrightarrow{\times\!\times}} = (A^2 - A^{-2}) \mathcal{L}_{\overrightarrow{\equiv}}. \end{aligned}$$

Letting $A = t^{-1/4}$, we conclude that

$$t^{-1} \mathcal{L}_{\overrightarrow{\times\!\times}} - t \mathcal{L}_{\overrightarrow{\times\!\times}} = \left(\sqrt{t} - \frac{1}{\sqrt{t}} \right) \mathcal{L}_{\overrightarrow{\equiv}}.$$

This proves property (iii) of Proposition 5.1. Properties (i) and (ii) follow directly from the corresponding facts about \mathcal{L}_K . This completes the proof. //

One effect of this approach to the Jones polynomial is that we get an immediate and simple proof of the **reversing property**.

Proposition 5.3. Let K and K' be two oriented links, so that K' is obtained by reversing the orientation of a component $K_1 \subset K$. Let $\lambda = \ell k(K_1, K - K_1)$ denote the total linking number of K_1 with the remaining components of K . (That is, λ is the sum of the linking number of K_1 with the remaining components of K .) Then

$$V_{K'}(t) = t^{-3\lambda} V_K(t).$$

Proof. It is easy to see that the writhes of K and K' are related by the formula: $w(K') = w(K) - 4\lambda$ where $\lambda = \ell k(K_1, K - K_1)$ as in the statement of the proposition. Thus

$$\begin{aligned} \mathcal{L}_{K'}(A) &= (-A^3)^{-w(K')} \langle K' \rangle \\ &= (-A^3)^{-w(K')} \langle K \rangle \\ &= (-A^3)^{-w(K)+4\lambda} \langle K \rangle \\ \therefore \mathcal{L}_{K'}(A) &= (-A^3)^{4\lambda} \mathcal{L}_K(A). \end{aligned}$$

Thus, by 5.2, $V_{K'}(t) = \mathcal{L}_{K'}(t^{-1/4}) = t^{-3\lambda} \mathcal{L}_K(t^{-1/4}) = t^{-3\lambda} V_K(t)$. This completes the proof. //

The reversing result for the Jones polynomial is a bit surprising (see [MO1] and [LM2]) if the polynomial is viewed from the vantage of the relation

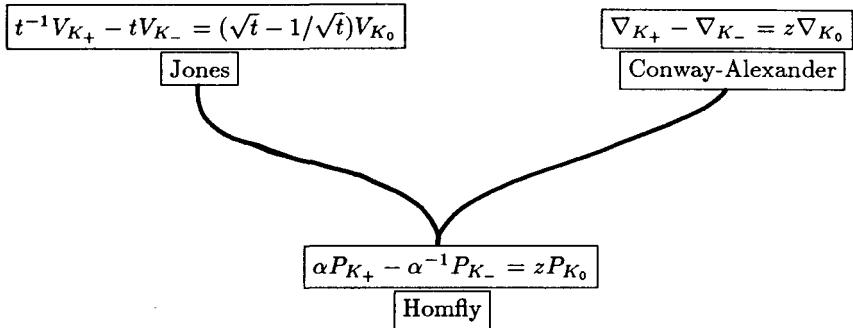
$$t^{-1} V_{\overleftarrow{\times}} - t V_{\overrightarrow{\times}} = \left(\sqrt{t} - \frac{1}{\sqrt{t}}\right) V_{\overleftarrow{\times}}.$$

This relation is structurally similar to the defining relations for the **Alexander-Conway polynomial** $\nabla_K(z) \in \mathbf{Z}[z]$ (polynomials in z with integer coefficients):

$$\left\{ \begin{array}{l} \text{(i)} \quad \nabla_K(z) = \nabla_{K'}(z) \text{ if the oriented links } K \\ \text{and } K' \text{ are ambient isotopic} \\ \\ \text{(ii)} \quad \nabla_{\text{circle}} = 1 \\ \\ \text{(iii)} \quad \nabla_{\overleftarrow{\times}} - \nabla_{\overrightarrow{\times}} = z \nabla_{\overleftarrow{\times}} \end{array} \right\}$$

These properties are John H. Conway's [CON] reformulation (and generalization) of the original Alexander polynomial [A]. We discuss the Alexander polynomial at some length and from various points of view in sections 12⁰ and 13⁰. It was Conway who perceived that the three properties ((i), (ii), (iii)) above characterize this polynomial. Moreover, $\nabla_K(z)$ is very sensitive to changes in orientation. It does not simply multiply by a power of a linking number.

Upon juxtaposing the exchange identities for the Jones and Conway-Alexander polynomials, one is led to ask for a common generalization. This generalization exists, and is called the Homfly polynomial after its many discoverers ([F], [PT]):



In the diagram above, I have indicated the basic exchange relation for the **oriented 2-variable polynomial** $P_K(\alpha, z)$. For $\alpha = t^{-1}$, $z = \sqrt{t} - 1/\sqrt{t}$, it specializes to the Jones polynomial. For $\alpha = 1$, it specializes to the Conway-Alexander polynomial. The original proofs of the existence of the Homfly polynomial were by induction on knot diagrams. We shall deduce its existence via generalizations of the bracket state model. Another approach to the Homfly polynomial is given by Jones in [JO3], that generalizes his original methods for $V_K(t)$.

Along with this oriented 2-variable polynomial there is also a **semi-oriented 2-variable polynomial** $F_K(\alpha, z)$ [LK8] (the Kauffman polynomial) that generalizes the bracket and original Jones polynomial. This polynomial is a normalization of a polynomial, L_K , defined for unoriented links and satisfying the properties:

- (i) If K is regularly isotopic to K' , then $L_K(\alpha, z) = L_{K'}(\alpha, z)$.
- (ii) $L \text{ } \bigcirc = 1$

$$(iii) L \begin{array}{c} \diagup \\ \diagdown \end{array} + L \begin{array}{c} \diagdown \\ \diagup \end{array} = z(L \begin{array}{c} \diagup \\ \diagup \end{array} + L \begin{array}{c} \diagdown \\ \diagdown \end{array})$$

$$(iv) L \begin{array}{c} \diagup \\ \diagdown \end{array} = \alpha L$$

$$L \begin{array}{c} \diagdown \\ \diagup \end{array} = \alpha^{-1} L$$

See Figure 8 and the discussion in sections 2⁰ and 3⁰ for more information about regular isotopy. Just as with the bracket, the L -polynomial is multiplicative under a type I Reidemeister move. The F -polynomial is defined by the formula

$$F_K(\alpha, z) = \alpha^{-w(K)} L_K(\alpha, z)$$

where $w(K)$ is the writhe of the oriented link K .

Again, we take up the existence of this invariant in section 14⁰. Given its existence, we can see easily that the bracket is a special case of L and the Jones polynomial is a special case of F . More precisely,

Proposition 5.4. $\langle K \rangle(A) = L_K(-A^3, A + A^{-1})$ and

$$V_K(t) = F_K(-t^{-3/4}, t^{-1/4} + t^{1/4}).$$

Proof. We have the bracket identities

$$\left\langle \begin{array}{c} \diagup \\ \diagdown \end{array} \right\rangle = A \left\langle \begin{array}{c} \diagup \\ \diagup \end{array} \right\rangle + A^{-1} \left\langle \begin{array}{c} \diagdown \\ \diagdown \end{array} \right\rangle$$

$$\left\langle \begin{array}{c} \diagup \\ \diagup \end{array} \right\rangle = A^{-1} \left\langle \begin{array}{c} \diagup \\ \diagdown \end{array} \right\rangle + A \left\langle \begin{array}{c} \diagdown \\ \diagdown \end{array} \right\rangle.$$

$$\text{Hence } \left\langle \begin{array}{c} \diagup \\ \diagdown \end{array} \right\rangle + \left\langle \begin{array}{c} \diagup \\ \diagup \end{array} \right\rangle = (A + A^{-1}) \left[\left\langle \begin{array}{c} \diagup \\ \diagup \end{array} \right\rangle + \left\langle \begin{array}{c} \diagdown \\ \diagdown \end{array} \right\rangle \right].$$

Therefore $\langle K \rangle(A) = L_K(-A^3, A + A^{-1})$. Since $V_K(t) = L_K(t^{-1/4})$ and

$$\mathcal{L}_K(A) = (-A^3)^{-w(K)} \langle K \rangle(A)$$

we have that

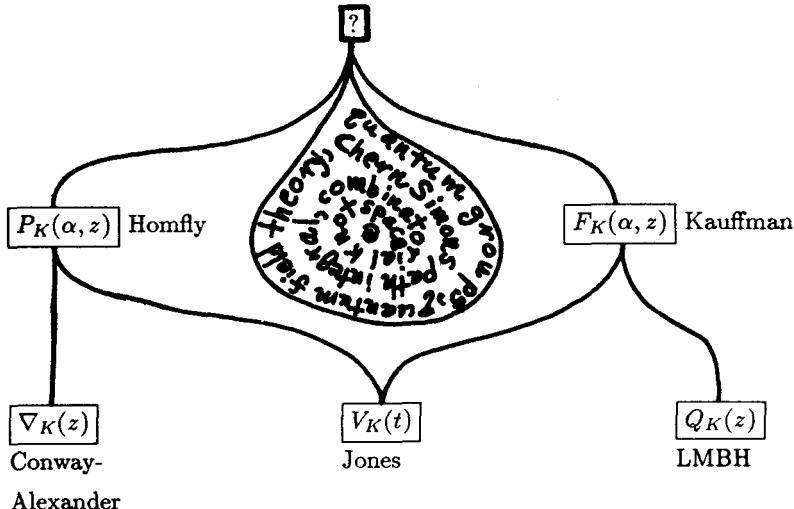
$$V_K(t) = (-t^{-3/4})^{-w(K)} \langle K \rangle(t^{-1/4}).$$

Therefore $V_K(t) = (-t^{-3/4})^{-w(K)} L_K(-t^{-3/4}, t^{-1/4} + t^{1/4})$. Hence

$$V_K(t) = F_K(-t^{-3/4}, t^{-1/4} + t^{1/4}).$$

This completes the proof. //

Thus we can form the chart



where the question mark denotes an unknown unification of the oriented and semi-oriented 2-variable polynomials. Here the invariant $Q_K(z)$ is the unoriented polynomial invariant of ambient isotopy discovered by Lickorish, Millett, Brandt and Ho ([BR], [HO]). It is, in this context, obtained by setting $\alpha = 1$ in the L -polynomial.

Both of these 2-variable polynomials - P_K and F_K are good at distinguishing mirror images. F_K seems a bit better at this game. (See [LK8].)

In this description of knot polynomial generalizations of the Jones polynomial I have avoided specific calculations and examples. However, it is worth mentioning that the oriented invariant $P_K(\alpha, z)$ can also be regarded as the normalization of a regular isotopy invariant. In this way we define the **regular isotopy homfly polynomial**, $H_K(\alpha, z)$, by the properties

- (i) If the oriented links K and K' are regularly isotopic, then

$$H_K(\alpha, z) = H_{K'}(\alpha, z).$$

$$(ii) H_{\text{circle}} = 1$$

$$(iii) H_{\overrightarrow{\nearrow} \nearrow} - H_{\overleftarrow{\nearrow} \nearrow} = z H_{\overleftarrow{\nearrow} \nearrow}$$

$$(iv) H \text{ } \begin{array}{c} \nearrow \\ \searrow \end{array} = \alpha H \rightarrow$$

$$H \text{ } \begin{array}{c} \searrow \\ \nearrow \end{array} = \alpha^{-1} H \leftarrow.$$

Again, believing in the existence of this invariant, we have

Proposition 5.5. $P_K(\alpha, z) = \alpha^{-w(K)} H_K(\alpha, z)$.

Proof. Let $W_K(\alpha, z) = \alpha^{-w(K)} H_K(\alpha, z)$. Then W_K is an invariant of ambient isotopy, and $W_{\text{circle}} = 1$. Thus it remains to check the exchange identity:

$$H \text{ } \begin{array}{c} \nearrow \\ \searrow \end{array} - H \text{ } \begin{array}{c} \searrow \\ \nearrow \end{array} = z H \rightarrow$$

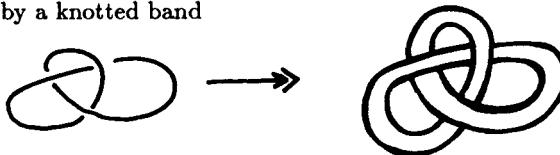
$$\Rightarrow \alpha^{+1} \alpha^{-(w+1)} H \text{ } \begin{array}{c} \nearrow \\ \searrow \end{array} - \alpha^{-1} \alpha^{-(w-1)} H \text{ } \begin{array}{c} \searrow \\ \nearrow \end{array} = z \alpha^{-w} H \rightarrow$$

where $w = w(\rightarrow)$. Hence

$$\alpha W \text{ } \begin{array}{c} \nearrow \\ \searrow \end{array} - \alpha^{-1} W \text{ } \begin{array}{c} \searrow \\ \nearrow \end{array} = z W \rightarrow .$$

This completes the proof. //

There is a conceptual advantage in working with the regular isotopy Homfly polynomial. For one thing, we see that the second variable originates in measuring the writhing of the diagrams involved in a given calculation. One can think of this as a measure of actual topological twist (rather than diagrammatic curl) if we replace the link components by embedded bands. Thus a trefoil knot can be replaced by a knotted band



Then the rule $H \text{ } \begin{array}{c} \nearrow \\ \searrow \end{array} = \alpha H \rightarrow$ becomes the rule:

$$H \text{ } \begin{array}{c} \nearrow \\ \searrow \end{array} = H \text{ } \begin{array}{c} \nearrow \\ \searrow \end{array} = \alpha H \rightarrow$$

That is, the polynomial H may be interpreted as an **ambient isotopy invariant of embedded, oriented band-links**. (No Möbius bands here. Compare [HE1] and [HE2].) The exchange identity remains essentially the same:

$$H \begin{array}{c} \text{X} \\ \diagup \quad \diagdown \\ \text{X} \end{array} - H \begin{array}{c} \text{X} \\ \diagdown \quad \diagup \\ \text{X} \end{array} = zH \begin{array}{c} \text{Y} \\ \curvearrowleft \quad \curvearrowright \\ \text{Y} \end{array}$$

With this interpretation the variables z and α acquire separate meanings. The z -variable measures the splicing and shifting operations. The α -variable measures twisting in the bands. This point of view can then be reformulated to work with **framed links** - links with an associated normal vector field.

To underline this discussion, let's compute the specific examples of Hopf link and trefoil knot:

$$(a) H \begin{array}{c} \text{Q} \\ \text{O} \end{array} = \alpha^{-1} H \begin{array}{c} \text{Q} \\ \text{O} \end{array}; H \begin{array}{c} \text{O} \\ \text{Q} \end{array} = \alpha^{-1} H \begin{array}{c} \text{O} \\ \text{Q} \end{array} = \alpha^{-1}.$$

$$(b) H \begin{array}{c} \text{G} \\ \text{G} \end{array} - H \begin{array}{c} \text{G} \\ \text{G} \end{array} = zH \begin{array}{c} \text{G} \\ \text{G} \end{array}$$

$$\alpha - \alpha^{-1} = zH \begin{array}{c} \text{G} \\ \text{G} \end{array}$$

$$\text{Let } \delta = (\alpha - \alpha^{-1})/z.$$

The same reasoning shows that

$$H_{\bullet K} = \delta H_K \text{ for any } K.$$

$$(c) H \begin{array}{c} \text{G} \\ \text{G} \end{array} - H \begin{array}{c} \text{G} \\ \text{G} \end{array} = zH \begin{array}{c} \text{G} \\ \text{G} \end{array}, \quad \text{G}_L$$

$$H_L - \delta = z\alpha$$

$$H_L = z^{-1}(\alpha - \alpha^{-1}) + z\alpha$$

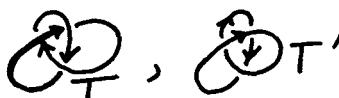
$$P_L = \alpha^{-2} H_L = z^{-1}(\alpha^{-1} - \alpha^{-3}) + z\alpha^{-1}.$$

$$(d) H_T - H_{T'} = zH_L$$

$$H_T - \alpha = z[z^{-1}(\alpha - \alpha^{-1}) + z\alpha]$$

$$H_T = (2\alpha - \alpha^{-1}) + z^2\alpha$$

$$P_T = \alpha^{-3} H_T = 2\alpha^{-2} - \alpha^{-4} + z^2\alpha^{-2}.$$



Since $P_{K^*}(\alpha, z) = P_K(\alpha^{-1}, -z)$ when K^* is the mirror image of K , we see the chirality of T reflected in this calculation of P_T .

In these calculations we have used a tree-like decomposition into simpler knots and links. Thus in the case of the trefoil, the full calculation can be hung on the tree in Figure 10.

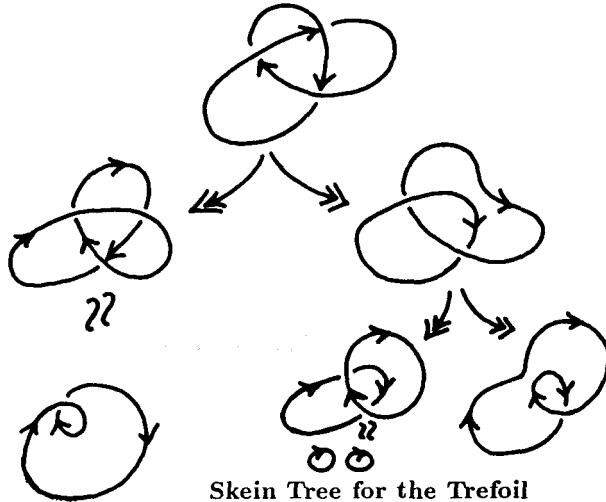
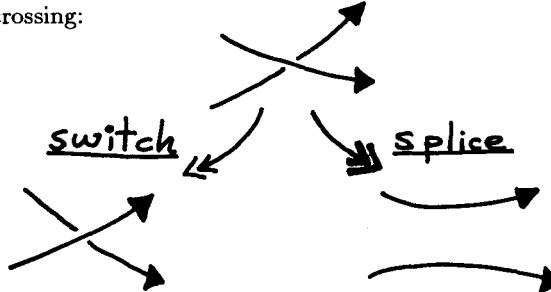


Figure 10

Each branching of this skein-tree takes the form of splicing or switching a given crossing:

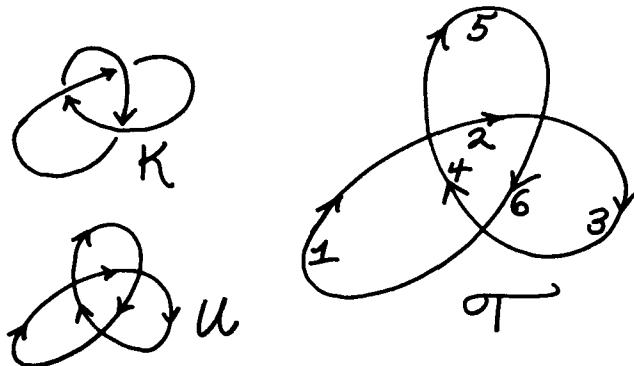


We can make this tree-generation process completely automatic by the following algorithm ([JA1], [LK13]):

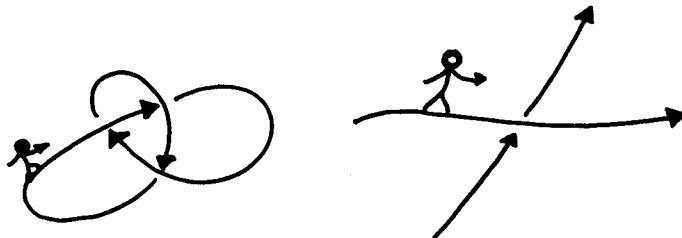
Skein-Template Algorithm.

- Let K be a given oriented link and let U be the universe underlying K . (U is the diagram for K without any indicated over or undercrossings). **Label each**

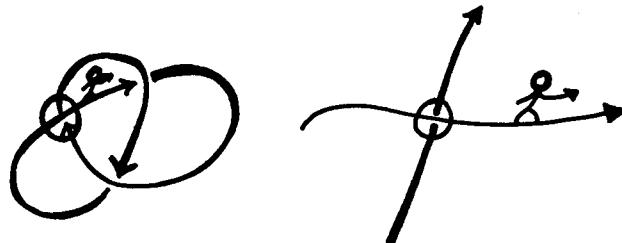
edge of U with a distinct positive integer. Call this labelled U the template T for the algorithm.



- (ii) Choose the least (unused) label on the template T . Begin walking along the diagram K (or along any diagram in the tree derived from K) in the direction of the orientation and from the center of this least labelled edge.

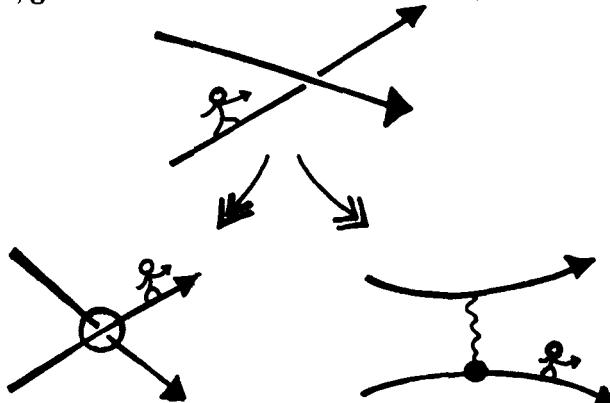


Continue walking across an overpass, and decorate it (see below).

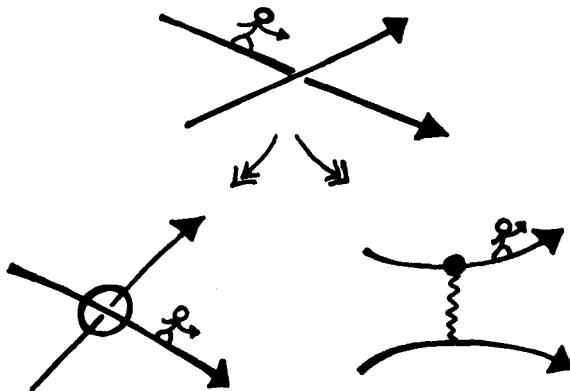


Stop walking at an underpass. If you have stopped walking at an

underpass, generate two branches of the tree by switch and splice.



Decorate the switched and spliced crossings as indicated above and also below:



These decorations are designed to indicate that the state of the site has been decided. Note that the dot → for a spliced crossing is placed along the arc of passage for the walker.

In the switched diagram a walker would have **passed over** as indicated on the decorated crossing (using the given template). In the spliced diagram the walker can **continue walking** along the segment with the dot.

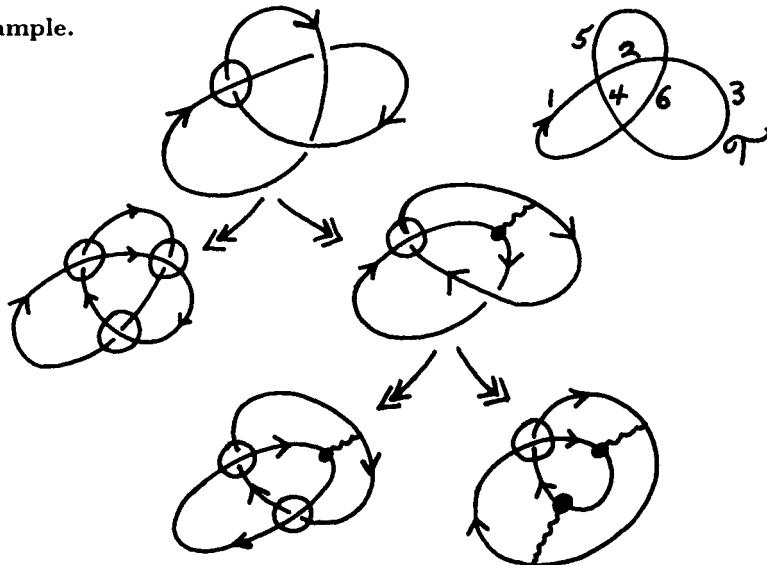
Also decorate any overpass that you walk along in the process of the algorithm.

(iii) If you have completed (ii) for a given diagram, go to the next level in the tree and apply the algorithm (ii) again – using the same template. [Since switched

and spliced descendants of K have the same edges (with sometimes doubled vertices), the same template can be used.] It is not necessary to start all over again at any of these branchings. Assume the path began from the generation node for your particular branch.

If there is no next level in the tree, then every vertex in each of the final nodes is decorated. This is the completion of the **Skein-Template Algorithm**.

Example.



Regard the **output** of the Skein-Template Algorithm as the set

$$\text{STA}(K, T) = \left\{ D \mid \begin{array}{l} D \text{ is a decorated final node (diagram)} \\ \text{in the tree generated by the Skein-} \\ \text{Template Algorithm with template } T. \end{array} \right\}$$

Thus we have

$$\text{STA}\left(\text{ (knot diagram) , (knot diagram) } \right) = \{ \text{ (knot diagram) , (knot diagram) , (knot diagram) } \}.$$

Note that (at least in this example) every element $D \in \text{STA}(K, T)$ is an unknot or unlink. This is always the case - it is a slight generalization of the fact that you will always draw an unknot if you follow the rule: **first crossing is an overcrossing**, while traversing a given one-component universe.



The tree generated by the skein-template algorithm is sufficient to calculate $H_K(\alpha, z)$ via the exchange identity and the basic facts about H_K .

In fact, with our decorations of the elements $D \in \text{STA}(K, T)$ we can write a formula:

$$H_K(\alpha, z) = \sum_{D \in \text{STA}(K, T)} \langle K | D \rangle \delta^{\|D\|}$$

where $\delta = (\alpha - \alpha^{-1})/z$, $\|D\|$ denotes one less than the number of link components in D , and $\langle K | D \rangle$ is a **product of vertex weights** depending upon the crossing types in K and the corresponding decorations in D . The local rules for these vertex weights are as follows:

$$\left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \mid \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle = z$$

$$\left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \mid \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle = -z$$

$$\left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \mid \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle = 0$$

$$\left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \mid \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle = 0$$

Positive crossings only allow first passage on the lower leg, while negative crossings only allow first passage on the upper leg. These rules correspond to our generation process. The D 's generated by the algorithm will yield no zeros.

A different and simpler rule applies to the encircled vertices:

$$\left\langle \begin{array}{c} \text{X} \\ \text{X} \end{array} \text{ or } \begin{array}{c} \text{X} \\ \text{X} \end{array} \mid \begin{array}{c} \text{X} \\ \text{X} \end{array} \right\rangle = \alpha$$

$$\left\langle \begin{array}{c} \text{X} \\ \text{X} \end{array} \text{ or } \begin{array}{c} \text{X} \\ \text{X} \end{array} \mid \begin{array}{c} \text{X} \\ \text{X} \end{array} \right\rangle = \alpha^{-1}$$

The product of these weights will give the writhe of D . Thus, we have the formula

$$\langle K|D \rangle = [(-1)^{t_-(D)} z^{t(D)} \alpha^{w(D)}]$$

where $t_-(D)$ denotes the number of splices of negative crossings (of K) in D and $t(D)$ denotes the number of spliced crossings in D .

Here it is assumed that D is obtained from $\text{STA}(K, T)$ so that there are no zero-weighted crossings.

I leave it as an exercise for the reader to check that the formula

$$H_K = \sum_{D \in \text{STA}(K, T)} \langle K|D \rangle \delta^{\parallel D \parallel}$$

is indeed correct (on the basis of the axioms for H_K). This formula shows that the tree-generation process also generates a state-model for H_K that is very similar in its form to the bracket. The real difference between this **skein model for H_K** and the state model for the bracket is that it is not at all obvious how to use the skein model as a logical or conceptual foundation for H_K . It is really a **computational expression**. The bracket model, on the other hand, gives us a direct entry into the inner logic of the Jones polynomial. I do not yet know how to build a model of this kind for $H_K(\alpha, z)$ as a whole.

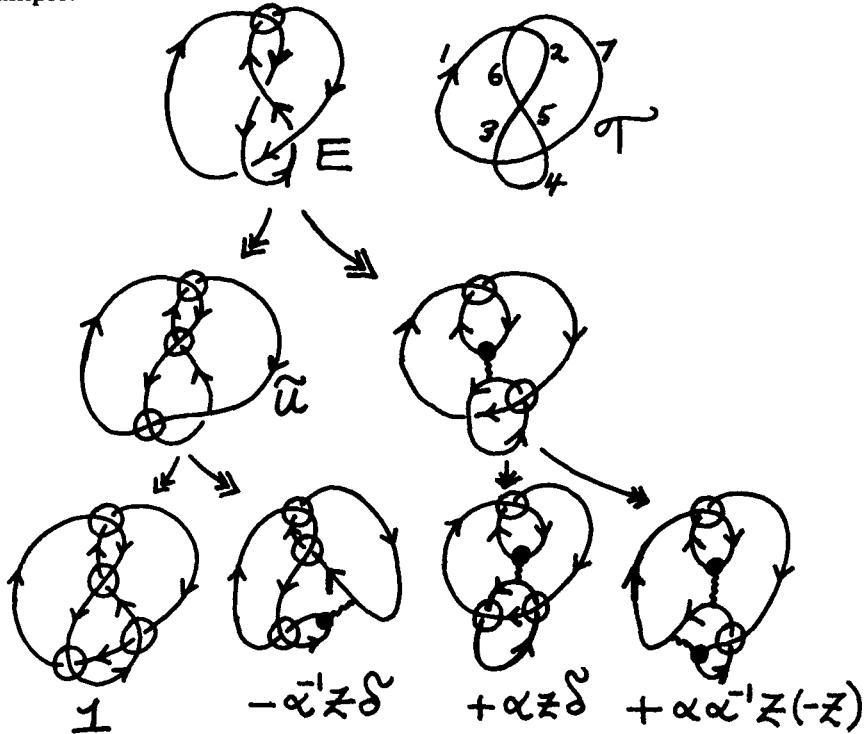
Remark. The first skein model was given by Francois Jaeger [JA1], using a matrix inversion technique. In [LK13] I showed how to interpret and generalize this model as a direct consequence of skein calculation.

Example. Returning to the trefoil, we calculate $H_T(\alpha, z)$ using the skein model.

$$\text{STA}(T, T) = \left\{ \begin{array}{ccc} \text{trefoil} & , & \text{trefoil} & , & \text{trefoil} \end{array} \right\}$$

$$H_T = \alpha + z\delta + \alpha z^2 = (2\alpha - \alpha^{-1}) + z^2\alpha.$$

Example.



$$\begin{aligned}\Rightarrow H_E &= 1 - \alpha^{-1}z\delta + \alpha z\delta - z^2 \\ &= 1 - \alpha^{-1}(\alpha - \alpha^{-1}) + \alpha(\alpha - \alpha^{-1}) - z^2 \\ H_E &= (\alpha^{-2} + \alpha^2 - 1) - z^2\end{aligned}$$

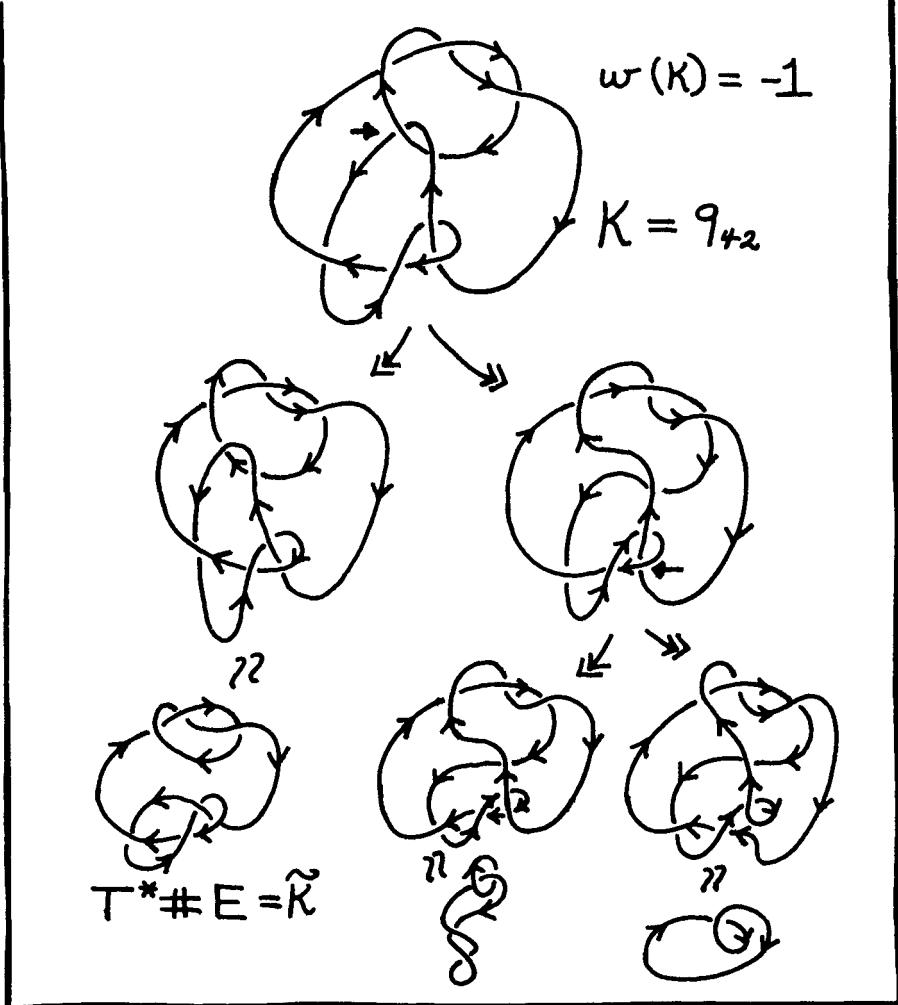
Note that for this representation of the figure eight knot E , we have $w(E) = 0$. Thus $H_E = P_E$. $H_E(\alpha, z) = H_E(\alpha^{-1}, -z)$ reflects the fact that $E \sim E^*$. That is, E is an achiral knot. Of course the fact that $P_K(\alpha, z) = P_K(\alpha^{-1}, -z)$ does not necessarily imply that K is achiral. The knot 9_{42} (next example) is the first instance of an anomaly of this sort.

This example also indicates how the Skein Template Algorithm is a typically dumb algorithm. At the first switch, we obtained the unknot \tilde{U} with $H_{\tilde{U}} = \alpha^{-2}$. The algorithm just went about its business switching and splicing \tilde{U} - this is

reflected in the calculation $1 - \alpha^{-1}z\delta = \alpha^{-2}$. A more intelligent version of the algorithm would look for unknots and unlink. This can be done in practice (Compare [EL].), but no theoretically complete method is available.

The example illustrates how, after a splice, the path continuation may lead through a previously decorated crossing. Such crossings are left undisturbed in the process - just as in the process of return in drawing a standard unlink.

Example. $K = 9_{42}$, the 42-nd knot of nine crossings in the Reidemeister tables:



Here I have created (the beginning of) a skein tree for 9_{42} by making choices and regular isotopies by hand. In this case the hand-choice reduces the calculation by an enormous amount. We see that

$$\begin{aligned}
 H_K &= [H_T^*] \left[\begin{matrix} H_E \\ H_{T^*} \end{matrix} \right] \\
 &= [H_{T^*}][H_E].
 \end{aligned}$$

In general, it is easy to see that

$$H_{A \# B} = (H_A)(H_B).$$

Thus $H_K = H_{\tilde{K}} + z \left[\alpha^2 H_{T^*} - z\alpha^{-1} \right]$, using the three diagrams in the tree.

$$H_{T^*} = L^* \text{ so that}$$

$$H_{L^*} = z^{-1}(\alpha - \alpha^{-1}) + z\alpha^{-1}$$

$$H_{T^*} = (2\alpha^{-1} - \alpha) + z^2\alpha^{-1}$$

$$H_E = (\alpha^{-2} + \alpha^2 - 1) - z^2.$$

Here I have used previous calculations in this section.

Putting this information together, we find

$$\begin{aligned}
 H_{\tilde{K}} &= H_{T^*} H_E \\
 &= ((2\alpha^{-1} - \alpha) + z^2\alpha^{-1})(\alpha^{-2} + \alpha^2 - 1 - z^2) \\
 &= 2\alpha^{-3} + 2\alpha - 2\alpha^{-1} - 2\alpha^{-1}z^2 \\
 &\quad - \alpha^{-1} - \alpha^3 + \alpha + \alpha z^2 \\
 &\quad + z^2\alpha^{-3} + z^2\alpha - z^2\alpha^{-1} - z^4\alpha^{-1}
 \end{aligned}$$

$$H_{\tilde{K}} = (2\alpha^{-3} - 3\alpha^{-1} + 3\alpha - \alpha^3) + (\alpha^{-3} - 3\alpha^{-1} + 2\alpha)z^2 - \alpha^{-1}z^4$$

$$\begin{aligned}\therefore H_K &= H_{\tilde{K}} + z\alpha^2 H_{L^*} - z^2\alpha^{-1} \\ &= H_{\tilde{K}} + z\alpha^2(z^{-1}(\alpha - \alpha^{-1}) - z\alpha^{-1}) - z^2\alpha^{-1} \\ &= H_{\tilde{K}} + \alpha^3 - \alpha - z^2\alpha - z^2\alpha^{-1} \\ &= H_{\tilde{K}} + (\alpha^3 - \alpha) + z^2(-\alpha - \alpha^{-1})\end{aligned}$$

$$H_K = (2\alpha^{-3} - 3\alpha^{-1} + 2\alpha) + (\alpha^{-3} - 4\alpha^{-1} + \alpha)z^2 - \alpha^{-1}z^4$$

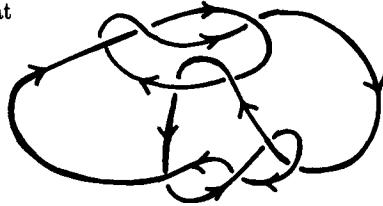
$$P_K = \alpha H_K = (2\alpha^{-2} - 3 + 2\alpha^2) + (\alpha^{-2} - 4 + \alpha^2)z^2 - z^4.$$

This calculation reveals that $P_K(\alpha, z) = P_K(\alpha^{-1}, -z)$ when K is the knot 9_{42} . Thus P_K does not detect any chirality in 9_{42} . Nevertheless, 9_{42} is chiral! (See [LK3], p. 207.)

This computation is a good example of how ingenuity can sometimes overcome the cumbersome combinatorial complexity inherent in the skein template algorithm.

Summary. This section has been an introduction to the original Jones polynomial $V_K(t)$ and its 2-variable generalizations. We have shown how the recursive form of calculation for these polynomials leads to formal state models, herein called skein models, and we have discussed sample computations for the oriented 2-variable generalization $P_K(\alpha, z)$, including a direct computation of this polynomial for the knot 9_{42} .

Exercise. (i) Show that

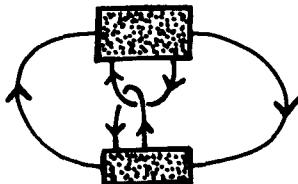


is inequivalent to its mirror image.

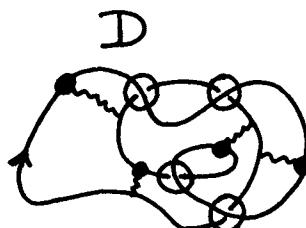
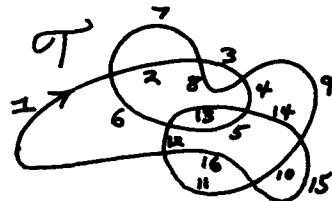
(ii) Compute H_K for K as shown below:

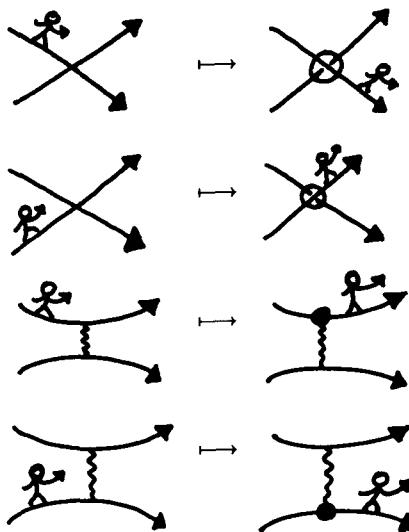


(iii) Investigate H_K when K has the form



Exercise. Let K be an oriented link diagram and U its corresponding universe. Let T be a template for U . Let \tilde{D} be any diagram obtained from U by splicing some subset of the crossings of U . Decorate \tilde{D} by using the template: That is, traverse \tilde{D} by always choosing the least possible base-point from T , and decorate each crossing as you first encounter it:





Let D denote the diagram \tilde{D} after decoration by this procedure. Let $\langle K | D \rangle$ be defined as in the Skein-Template-Algorithm. (It is now possible that $\langle K | D \rangle = 0$.)

Show. $\langle K | D \rangle \neq 0 \Leftrightarrow D \in \text{STA}(K, T)$.

This exercise shows that we can characterize the skein-model for H_K via

$$H_K = \sum_{D \in \mathcal{D}(K, T)} \langle K | D \rangle \delta^{\|D\|}$$

where $\mathcal{D}(K, T)$ is the set of all diagrams obtained as in the exercise (splice a subset of crossings of U and decorate using T).

The result of this exercise shifts the viewpoint about H_K away from the skein-tree to the structure of the states themselves. For example, in computing H_L for L the link,  , we choose a template  T , and

then consider the set $\tilde{\mathcal{D}}$ of all diagrams obtained from L by splicing or projecting crossings. These are:

$$\left\{ \text{Diagram 1}, \text{Diagram 2}, \text{Diagram 3}, \text{Diagram 4} \right\} = \tilde{\mathcal{D}}.$$

Each element of $\tilde{\mathcal{D}}$ contributes, after decoration, a term in H_K . This procedure goes as follows:

$$\begin{array}{c}
 T \quad | \quad \text{decorated} \quad L \\
 \text{decorate} \quad \longleftrightarrow \quad \delta \\
 \text{decorate} \quad \longleftrightarrow \quad 0 \\
 \text{decorate} \quad \longleftrightarrow \quad z\alpha \\
 \text{decorate} \quad \longleftrightarrow \quad 0
 \end{array}$$

$\therefore H = z\alpha + \delta = z\alpha + z^{-1}(\alpha - \alpha^{-1}).$

Example/Exercise. Finally, it is worth remarking that we can summarize our version of the skein model for H_K by an expansion formula similar to that for the bracket. Thus we write:

$$\begin{aligned}
 H \times \times &= zH \times \times + \alpha H \times \times + \alpha^{-1} H \times \times \\
 H \times \times &= -zH \times \times + \alpha H \times \times + \alpha^{-1} H \times \times
 \end{aligned}$$

to summarize the vertex weights, and

$$H_{\bullet K} = \delta H_K, \quad \delta = (\alpha - \alpha^{-1})/z$$

to summarize the loop behavior. Here it is understood that the resulting states must be compared to a given template T for admissibility. Thus, if σ is a state

(i.e. a decorated diagram without crossings that is obtained from the expansion) then $H\sigma = \epsilon \delta^{\|\sigma\|}$ ($\|\sigma\|$ = number of loops in σ minus one) where $\epsilon = 1$ if σ 's decorations follow from T and $\epsilon = 0$ if σ 's decorations do not follow from T .

It is fun to explore this a bit. For example:

$$H \text{---} \circlearrowright = zH \text{---} \circlearrowleft + \alpha H \text{---} \circlearrowright + \alpha^{-1} H \text{---} \circlearrowright$$

Because of necessity for choosing a template, some of these terms are zero. In the calculation above, we can indicate the influence of the template by showing a **base-point** (*) for the circuit calculations: Thus in



We have only the term , giving

$$H \text{---} \overset{*}{\circlearrowright} = \alpha H \text{---} \overset{*}{\circlearrowleft} = \alpha H \text{---} \overset{*}{\circlearrowright} .$$

And in we have the terms and , giving

$$H \text{---} \overset{*}{\circlearrowright} = zH \text{---} \circlearrowleft + \alpha^{-1} H \text{---} \circlearrowright *$$

$$= (z\delta + \alpha^{-1}) H \text{---} \overset{*}{\circlearrowright}$$

$$= (\alpha - \alpha^{-1} + \alpha^{-1}) H \text{---} \overset{*}{\circlearrowright}$$

$$= \alpha H \text{---} \overset{*}{\circlearrowright}$$

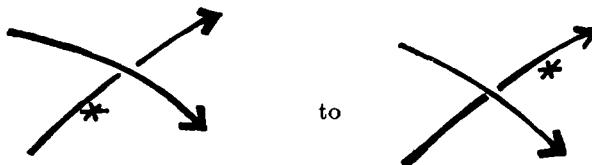
This calculation is the core of the well-definedness of the skein model.

In fact, we can sketch this proof of well-definedness. It is a little easier (due to the direct nature of the model), but of essentially the same nature as the original

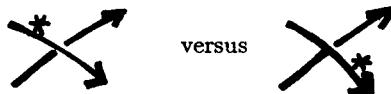
induction proofs for the Homfly polynomial [LM1]. The main point is to see that H_K is independent of the choice of template. We can first let the * (above) stand for a template label that is least among all template labels. Then going from



or from



can denote an exchange of template labels. Let's concentrate on the case



We want to show that these two choices of template will give identical calculations in the state summations. Now we have (assuming that \ and / are on the same component)

$$H \begin{array}{c} \diagup \\ \diagdown \end{array} = \alpha H \begin{array}{c} \diagdown \\ \diagup \end{array}$$

and

$$H \begin{array}{c} \diagup \\ \diagdown \end{array} = z H \begin{array}{c} \diagdown \\ \diagup \end{array} + \alpha^{-1} H \begin{array}{c} \diagup \\ \diagup \end{array}$$

Each right hand term is a sum over states with the given local configurations.

Thus it will suffice to show that

$$\alpha H \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} = z H \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} + \alpha^{-1} H \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array}$$

holds for states with respect to these templates. Each state is weighted by the contributions of its decorations. To make the notation more uniform, let $\langle \sigma \rangle$ denote the state's contribution. Thus

$$\langle \sigma \rangle = z^{t(\sigma)} (-1)^{t_-(\sigma)} \alpha^{w(\sigma)} \delta^{\parallel \sigma \parallel}$$

if the state σ is admissible, and $H_K = \sum_{\sigma} \langle \sigma \rangle$, summing over admissible states. We then must show that

$$\left\langle \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} \right\rangle = \left\langle \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} \right\rangle + \left\langle \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} \right\rangle$$

for individual states.

Remember that a state σ is, by construction, a standard unknot or unlink. Thus, if



$\begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array}$ has both edges in the **same** state component, then indicates two components and thus,

$$\left\langle \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} \right\rangle = z \delta \left\langle \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} \right\rangle$$

$$\left\langle \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} \right\rangle = \alpha^{-1} \left\langle \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} \right\rangle, (\cancel{z \delta} + \cancel{\alpha^{-1}} = \alpha)$$

$$\therefore \left\langle \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} \right\rangle + \left\langle \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} \right\rangle = \alpha \left\langle \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} \right\rangle$$

$$= \alpha \left\langle \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} \right\rangle = \left\langle \begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array} \right\rangle.$$

[$\begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array}$ denotes a neutral node that is simply carried along through all the calculations.]

If $\begin{array}{c} * \\ \diagup \quad \diagdown \\ \text{---} \end{array}$ denotes a crossing of two different components, then the same form of calculation ensues. Thus, we have indicated the proof that H_K is independent of the template labelling.

Given this well-definedness, the proof of regular isotopy invariance is easy. Just position the least label correctly:

$$\langle \text{Diagram} \rangle = \langle \text{Diagram with neutral node} \rangle = \alpha \alpha^{-1} \langle \text{Diagram with neutral node} \rangle \\ = \langle \text{Diagram} \rangle$$

Remark. I have here adopted the notation  to indicate a neutral node - i.e. a node whose vertex contribution has been already catalogued. Thus

$$\langle \text{Diagram} \rangle = \langle \text{Diagram with neutral node} \rangle \quad \text{and}$$

$$\langle \text{Diagram} \rangle = \langle \text{Diagram with neutral node} \rangle .$$

Speculation. In the skein models we may imagine particles moving on the diagram in trajectories whose initial points are dictated by the chosen template. Each state σ is a possible set of trajectories - with the specific vertex weights (hence state evaluation) determined by these starting points. Nevertheless, the entire summation H_K is independent of the choice of starting points. This independence is central for the topological meaning of the model. What does this mean physically?

In a setting involving the concept of local vertex weights determined by the global structure of trajectories (answering the question "Who arrived first?" where first refers to template order) the independence of choices avoids a multiplicity of "times". Imagine a spacetime universe with a single self-interacting trajectory moving forward and backward in time. A particle set moving on this track must be assigned a mathematical "meta-time" - but any physical calculation related to the trajectory (e.g. a vacuum-vacuum expectation) must be independent of meta-time parameters for that particle.

Later, we shall see natural interpretations for the knot polynomials as just such vacuum-vacuum expectations. In that context the skein calculation is akin to using meta-time coordinates in such a way that the dependence cancels out in the averaging.

6⁰. An Oriented State Model for $V_K(t)$.

We have already seen the relationship (Theorem 5.2) between the Jones polynomial $V_K(t)$ and the bracket polynomial:

$$V_K(t) = (-t^{3/4})^{w(K)} \langle K \rangle (t^{-1/4}).$$

In principle this gives an oriented state model for $V_K(t)$. Nevertheless, it is interesting to design such a model directly on the oriented diagrams. In unoriented diagrams we created two splits:

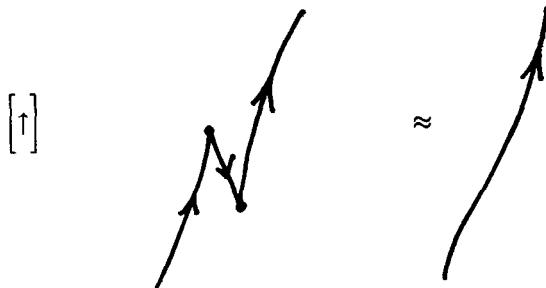


In oriented diagrams these splits become

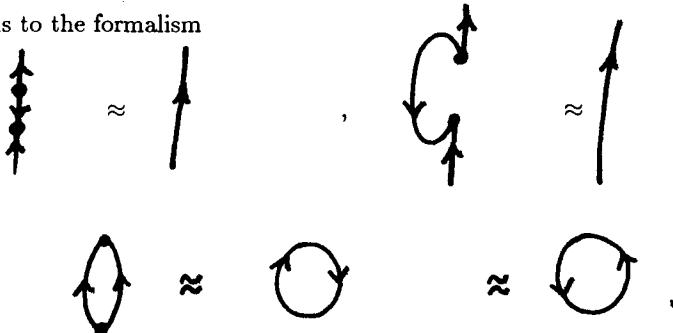


The second split acquires orientations outside the category of link diagrams.

It is useful to think of these new vertices as abstract Feynman diagrams with a local arrow of time that is coincident with the direction of the diagrammatic arrows. Then represents an interaction (a pass-by) and represent creations and annihilations. In order to have topological invariance we need to be able to cancel certain combinations of creation and annihilation. Thus

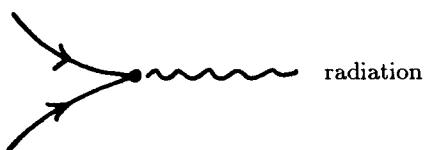


This leads to the formalism

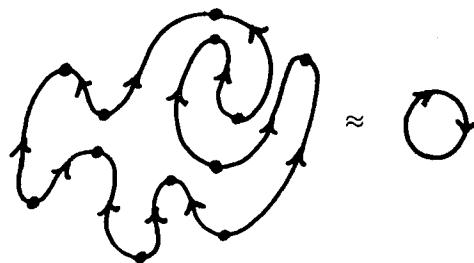


and

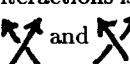
In other words, the topological diagrams know nothing (at this level of approximation) about field effects such as



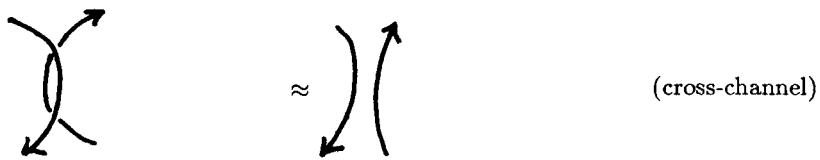
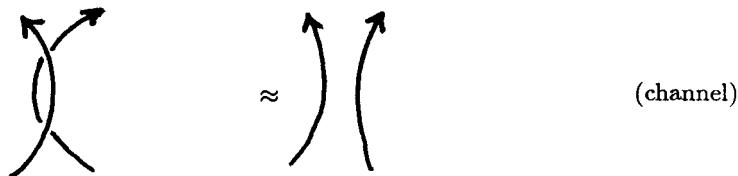
and the probability of each of the infinity of loops must be the same.



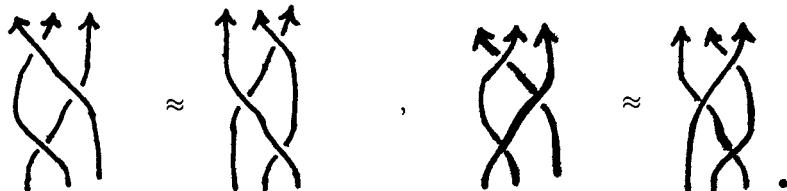
This background for diagrammatic interactions is necessary just to begin the topology. Then we want the interactions and to satisfy channel and cross-



channel unitarity:



and triangle invariance

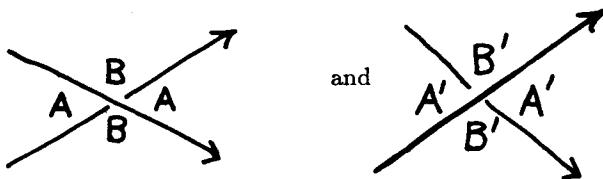


In analogy to the bracket, we posit expansions:

$$V \begin{array}{c} \nearrow \\ \searrow \end{array} = AV \begin{array}{c} \nearrow \\ \searrow \end{array} + BV \begin{array}{c} \nearrow \\ \searrow \end{array} \begin{array}{c} \nearrow \\ \searrow \end{array}$$

$$V \begin{array}{c} \nearrow \\ \searrow \end{array} = A'V \begin{array}{c} \nearrow \\ \searrow \end{array} + B'V \begin{array}{c} \nearrow \\ \searrow \end{array} \begin{array}{c} \nearrow \\ \searrow \end{array}$$

Each crossing interaction is regarded as a superposition of “pass-by” and “annihilate-create”. We can begin the model with arbitrary weights, A , B for the positive crossing, and A' and B' for the negative crossing. The mnemonics



are useful.

As in the bracket, we shall assume that an extra loop multiplies V_K by a parameter δ :

$$V_{\text{loop } K} = \delta V_K.$$

Unitarity, cross-channel unitarity and triangle invariance will determine these parameters.

1. Unitarity. V

$$= AV \begin{array}{c} \diagup \\ \diagdown \end{array} + BV \begin{array}{c} \diagdown \\ \diagup \end{array}$$

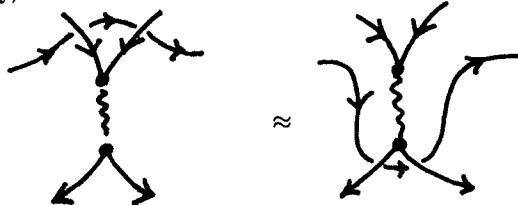
Remark. One might think that, given the tenor of the initial topological remarks about this theory, it would be required that

$$\begin{array}{c} \diagup \\ \diagdown \end{array} \approx \begin{array}{c} \diagdown \\ \diagup \end{array}$$

However, this would trivialize the whole enterprise. Thus the “field-effects” occur at the level of the interaction of creation or annihilation with the crossing interaction. As we continue, it will happen that

$$\begin{array}{c} \diagup \\ \diagdown \end{array} * \begin{array}{c} \diagdown \\ \diagup \end{array}$$

but, fortunately,



This is in keeping with the virtual character of the radiative part of this interaction. (All of these remarks are themselves virtual - to be taken as a metaphor/mathematico-physical fantasy on the diagrammatic string.)

To return to the calculation:

$$\begin{aligned}
 V &= AV + BV \\
 &= A \left\{ A'V \downarrow \downarrow \right. + B'V \left. \downarrow \downarrow \right\} \\
 &\quad + B \left\{ A'V \downarrow \downarrow \right. + B'V \left. \downarrow \downarrow \right\} \\
 V &= AA'V \downarrow \downarrow + (AB' + BB'\delta + BA')V \downarrow \downarrow
 \end{aligned}$$

Therefore, unitarity requires

$AA' = 1$
$\delta = - \left(\frac{A}{B} + \frac{A'}{B'} \right)$

2. Cross-Channel Unitarity.

$$\begin{aligned}
 V &= AV \begin{array}{c} \nearrow \\ \searrow \end{array} + BV \begin{array}{c} \swarrow \\ \nearrow \end{array} \\
 &= A \left\{ A' V \begin{array}{c} \nearrow \\ \searrow \end{array} + B' V \begin{array}{c} \swarrow \\ \nearrow \end{array} \right\} \\
 &\quad + B \left\{ A' V \begin{array}{c} \nearrow \\ \searrow \end{array} + B' V \begin{array}{c} \swarrow \\ \nearrow \end{array} \right\} \\
 &= BB' V \begin{array}{c} \nearrow \\ \searrow \end{array} + (AA'\delta + AB' + BA')V \begin{array}{c} \nearrow \\ \searrow \end{array}
 \end{aligned}$$

Therefore, cross-channel unitarity requires

$$\boxed{\begin{aligned} BB' &= 1 \\ \delta &= -\left(\frac{B'}{A'} + \frac{B}{A}\right) \end{aligned}}.$$

Thus we now have (1. and 2.)

$$\boxed{\begin{aligned} A' &= 1/A, & B' &= 1/B \\ \delta &= -\left(\frac{A}{B} + \frac{B}{A}\right) \end{aligned}}.$$

3. Annihilation and Crossing.

$$\begin{aligned}
 V \swarrow \searrow &= A'V \swarrow \searrow + B'V \swarrow \searrow \\
 &= A' \left\{ A'V \swarrow \searrow + B'V \swarrow \searrow \right\} \\
 &\quad + B' \left\{ A'V \swarrow \searrow + B'V \swarrow \searrow \right\} \\
 &= ((A')^2 + (B')^2 + \delta_{A'B'})V \swarrow \searrow + A'B'V \swarrow \searrow \\
 \therefore V \swarrow \searrow &= A'B'V \swarrow \searrow = \frac{1}{AB}V \swarrow \searrow
 \end{aligned}$$

Similarly,

$$V \swarrow \searrow = ABV \swarrow \searrow$$

and

$$V \swarrow \searrow = ABV \swarrow \searrow$$

$$V \swarrow \searrow = \frac{1}{AB}V \swarrow \searrow$$

$$\therefore V \swarrow \searrow = V \swarrow \searrow$$

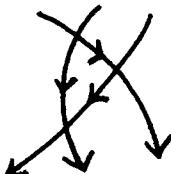
This is the annihilation-creation version of the triangle move, and it implies triangle invariance:

$$\begin{aligned}
 V \swarrow \searrow &= AV \swarrow \searrow + BV \swarrow \searrow = AV \swarrow \searrow + BV \swarrow \searrow
 \end{aligned}$$

$$\text{AV} \begin{array}{c} \diagup \\ \diagdown \end{array} + \text{BV} \begin{array}{c} \diagdown \\ \diagup \end{array} = V \begin{array}{c} \diagup \\ \diagdown \end{array}.$$

4. Oriented Triangles.

There are two types of oriented triangle moves. Those like



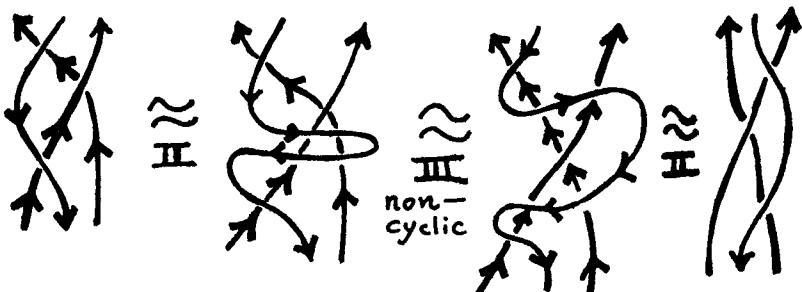
where the central triangle does **not** have a cyclic orientation, and those like



where the central triangle does have a cyclic orientation. Call the latter a **cyclic triangle move**.

Fact. In the presence of non-cyclic triangle moves and the two orientations of type II moves, cyclic triangle moves can be generated from non-cyclic triangle moves.

Proof. (Sketch) Observe the pictures below:



Exercise. Show that triangle moves involving crossings of different signs can be accomplished by triangle moves all-same-sign by using type II moves.

With the help of these remarks and the exercise we see that we have now done enough work to ensure that the model:

$$V \begin{array}{c} \nearrow \\ \searrow \end{array} = AV \begin{array}{c} \nearrow \\ \searrow \end{array} + BV \begin{array}{c} \nearrow \\ \searrow \end{array} \begin{array}{c} \nearrow \\ \searrow \end{array}$$

$$V \begin{array}{c} \nearrow \\ \searrow \end{array} = \frac{1}{A} V \begin{array}{c} \nearrow \\ \searrow \end{array} + \frac{1}{B} V \begin{array}{c} \nearrow \\ \searrow \end{array} \begin{array}{c} \nearrow \\ \searrow \end{array}$$

is a regular isotopy invariant for oriented link diagrams.

5. Type I Invariance.

$$\begin{aligned} V \begin{array}{c} \nearrow \\ \searrow \end{array} &= AV \begin{array}{c} \nearrow \\ \searrow \end{array} + BV \begin{array}{c} \nearrow \\ \searrow \end{array} \begin{array}{c} \nearrow \\ \searrow \end{array} \\ &= (A\delta + B)V \begin{array}{c} \nearrow \\ \searrow \end{array} \end{aligned}$$

$$A\delta + B = A\left(-\frac{A}{B} - \frac{B}{A}\right) + B = -A^2/B.$$

Thus

$$V \begin{array}{c} \nearrow \\ \searrow \end{array} = V \begin{array}{c} \nearrow \\ \searrow \end{array} \Leftrightarrow B = -A^2.$$

It is easy to see that this condition implies that $V \begin{array}{c} \nearrow \\ \searrow \end{array} = V \begin{array}{c} \nearrow \\ \searrow \end{array}$ as well.

Exercise. Show that this specialization contains the same topological information as the polynomial in A and B .

6. The Jones Polynomial.

We need $B = -A^2$.

Let $A = -\sqrt{t}$, $B = -t$. Then

$$V \begin{array}{c} \nearrow \\ \searrow \end{array} = -\sqrt{t} V \begin{array}{c} \nearrow \\ \searrow \end{array} - t V \begin{array}{c} \nearrow \\ \searrow \end{array} \begin{array}{c} \nearrow \\ \searrow \end{array}$$

$$V \begin{array}{c} \nearrow \\ \searrow \end{array} = -\frac{1}{\sqrt{t}} V \begin{array}{c} \nearrow \\ \searrow \end{array} - \frac{1}{t} V \begin{array}{c} \nearrow \\ \searrow \end{array} \begin{array}{c} \nearrow \\ \searrow \end{array}$$

All the work of this section assures us that this expansion yields an invariant of ambient isotopy for oriented links. Multiplying by t^{-1} and by t , and taking the difference, we find

$$\begin{aligned} t^{-1}V \begin{array}{c} \nearrow \\ \searrow \end{array} - tV \begin{array}{c} \nearrow \\ \searrow \end{array} &= \left(-\frac{\sqrt{t}}{t} + \frac{t}{\sqrt{t}} \right) V \begin{array}{c} \nearrow \\ \searrow \end{array} \\ &= \left(\sqrt{t} - \frac{1}{\sqrt{t}} \right) V \begin{array}{c} \nearrow \\ \searrow \end{array} \end{aligned}$$

Therefore $V_K(t)$ is the 1-variable Jones polynomial.

The interest of this model is that it does not depend upon writhe-normalization, and it shows how the parameters t, \sqrt{t} are intrinsic to the knot-theoretic structure of this invariant. Finally, all the issues about oriented invariants (channel and cross-channel unitarity, triangle moves, orientation conventions) will reappear later in our work with Yang-Baxter models (beginning in section 8⁰.)

It is rather intriguing to compare the art of calculating within this oriented model with the corresponding patterns of the bracket. A key lemma is

Lemma 6.1.

$$V \begin{array}{c} \nearrow \\ \searrow \end{array} = t\sqrt{t} V \begin{array}{c} \nearrow \\ \searrow \end{array}$$

$$V \begin{array}{c} \nearrow \\ \searrow \end{array} = \frac{1}{t\sqrt{t}} V \begin{array}{c} \nearrow \\ \searrow \end{array}$$

$$V \begin{array}{c} \nearrow \\ \searrow \end{array} = t\sqrt{t} V \begin{array}{c} \nearrow \\ \searrow \end{array}$$

$$V \begin{array}{c} \nearrow \\ \searrow \end{array} = \frac{1}{t\sqrt{t}} V \begin{array}{c} \nearrow \\ \searrow \end{array}$$

Proof.

$$\begin{aligned} V \begin{array}{c} \nearrow \\ \searrow \end{array} &= -\sqrt{t} V \begin{array}{c} \nearrow \\ \searrow \end{array} - tV \begin{array}{c} \nearrow \\ \searrow \end{array} \begin{array}{c} \nearrow \\ \searrow \end{array} \\ &= \left(-\sqrt{t} - t \left(-\sqrt{t} - \frac{1}{\sqrt{t}} \right) \right) V \begin{array}{c} \nearrow \\ \searrow \end{array} \\ &= t\sqrt{t} V \begin{array}{c} \nearrow \\ \searrow \end{array} \end{aligned}$$

The rest proceeds in the same way. //

Thus the model is an ambient isotopy invariant, but regular isotopy patterns of calculation live inside it!

Remark. It is sometimes convenient to use another parameterization of the Jones polynomial. Here we write

$$t\hat{V} \text{ (diagram)} - t^{-1}\hat{V} \text{ (diagram)} = \left(\sqrt{t} - \frac{1}{\sqrt{t}}\right) \hat{V} \text{ (diagram)}$$

Note that the loop variable $\hat{\delta} = (\sqrt{t} + \frac{1}{\sqrt{t}})$ for this model. The expansion is given by the formulas:

$$\hat{V} \text{ (diagram)} = \frac{1}{\sqrt{t}} \hat{V} \text{ (diagram)} - \frac{1}{t} \hat{V} \text{ (diagram with loop)}$$

$$\hat{V} \text{ (diagram)} = \sqrt{t} \hat{V} \text{ (diagram)} - t \hat{V} \text{ (diagram with loop)}$$

In the corresponding state expansion we have

$$\hat{V}_K = \sum_{\vec{\sigma}} \langle K | \vec{\sigma} \rangle \left(\sqrt{t} + \frac{1}{\sqrt{t}} \right)^{\|\vec{\sigma}\|}$$

where $\langle K | \vec{\sigma} \rangle$ is the product of the vertex weights $(\sqrt{t}, 1/\sqrt{t}, -t, -1/t)$ and $\|\vec{\sigma}\|$ is the number of loops in the oriented state. For each state, the sign is $(-1)^\pi$ where π is the parity of the number of creation-annihilation splices in the state. This version of the Jones polynomial may lead some insight into the vexing problem of cancellations in the state summation.

Unknot Problem. For a knot K , does $V_K = 1$ ($\hat{V}_K = 1$) imply that K is ambient isotopic to the unknot?

7⁰. Braids and the Jones Polynomial.

In this section I shall demonstrate that the normalized bracket

$\mathcal{L}_K(A) = (-A^3)^{-w(K)} \langle K \rangle$ is a version of the original Jones polynomial $V_K(t)$ by way of the theory of braids. The Jones polynomial has been subjected to extraordinary generalizations since it was first introduced in 1984 [JO2]. These generalizations will emerge in the course of discussion. Here we stay with the story of the original Jones polynomial and its relation with the bracket.

Jones constructed the invariant $V_K(t)$ by a route involving braid groups and von Neumann algebras. Although there is much more to say about von Neumann algebras, it is sufficient here to consider a sequence of algebras A_n ($n = 2, 3, \dots$) with multiplicative generators e_1, e_2, \dots, e_{n-1} and relations:

- 1) $e_i^2 = e_i$
- 2) $e_i e_{i \pm 1} e_i = \tau e_i$
- 3) $e_i e_j = e_j e_i \quad |i - j| > 2$

(τ is a scalar, commuting with all the other elements.) For our purposes we can let A_n be the free additive algebra on these generators viewed as a module over the ring $\mathbb{C}[\tau, \tau^{-1}]$ (\mathbb{C} denotes the complex numbers.). The scalar τ is often taken to be a complex number, but for our purposes is another algebraic variable commuting with the e_i 's. This algebra arose in the theory of classification of von Neumann algebras [JO7], and it can itself be construed as a von Neumann algebra.

In this von Neumann algebra context it is natural to study a certain tower of algebras associated with an inclusion of algebras $N \subset M$. With $M_0 = N$, $M_1 = M$ one forms $M_2 = \langle M_1, e_1 \rangle$ where $e_1 : M_1 \rightarrow M_0$ is projection to M_0 and $\langle M_1, e_1 \rangle$ denotes an algebra generated by M_1 with e_1 adjoined. Thus we have the pattern

$$M_0 \subset M_1 \subset M_2 = \langle M_1, e_1 \rangle, \quad e_1 : M_1 \rightarrow M_0, \quad e_1^2 = e_1.$$

This pattern can be iterated to form a tower

$$M_0 \subset M_1 \subset M_2 \subset M_3 \subset \dots \subset M_n \subset M_{n+1} \subset \dots$$

with $e_i : M_i \rightarrow M_{i-1}$, $e_i^2 = e_i$ and $M_{i+1} = \langle M_i, e_i \rangle$. Jones constructs such a tower of algebras with the property that $e_i e_{i \pm 1} e_i = \tau e_i$ and $e_i e_j = e_j e_i$ for $|i - j| > 1$.

Here he found that $\tau^{-1} = [M_1 : M_0]$ a generalized notion of index for these algebras. Furthermore, he defined a **trace** $\text{tr} : M_n \rightarrow \mathbf{C}$ (complex numbers) such that it satisfied the

$$\boxed{\begin{array}{l} \text{Markov Property: } \text{tr}(we_i) = \tau \text{ tr}(w) \\ \text{for } w \text{ in the algebra generated by } M_0, e_1, \dots, e_{i-1} \end{array}}.$$

A function from an algebra \mathcal{A} to the complex numbers is said to be a **trace** (tr) if it satisfies the identity

$$\text{tr}(ab) = \text{tr}(ba)$$

for $a, b \in \mathcal{A}$. Thus ordinary matrix traces are examples of trace functions.

The **tower construction** is very useful for studying the index $[M_1 : M_0]$ for special types of von Neumann algebras. While I shall not discuss von Neumann algebras in this short course, we shall construct a model of such a tower that is directly connected with the bracket polynomial. Thus the combinatorial structure of the tower construction will become apparent from the discussion that follows.

Now to return to the story of the Jones polynomial: Jones was struck by the analogy between the relations for the algebra A_n and the generating relations for the n -strand **Artin braid group** B_n . View Figure 11 for a comparison of these sets of relations

A_n	B_n
$e_i^2 = e_i$ $e_i e_{i \pm 1} e_i = \tau e_i$ $e_i e_j = e_j e_i, i - j > 1$ <hr style="border: 1px solid black; margin: 10px 0;"/> e_1, e_2, \dots, e_{n-1}	$\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}$ $\sigma_i \sigma_j = \sigma_j \sigma_i, i - j > 1$ <hr style="border: 1px solid black; margin: 10px 0;"/> $\sigma_1^{\pm 1}, \dots, \sigma_{n-1}^{\pm 1}$

Jones Algebra

Artin Braid Group

Figure 11

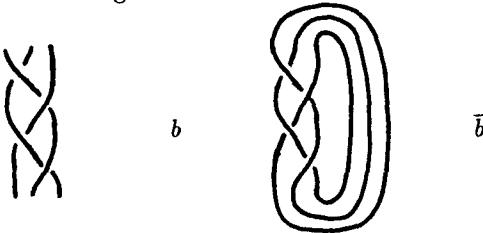
Jones constructed a representation $\rho_n : B_n \rightarrow A_n$ of the Artin Braid group to the algebra A_n . The representation has the form

$$\rho_n(\sigma_i) = ae_i + b$$

with a and b chosen appropriately. Since A_n has a trace $\text{tr} : A_n \rightarrow \mathbf{C}[t, t^{-1}]$ one can obtain a mapping $\text{tr} \circ \rho : B_n \rightarrow \mathbf{C}[t, t^{-1}]$. Upon appropriate normalization this mapping is the Jones polynomial $V_K(t)$. It is an ambient isotopy invariant for oriented links. While the polynomial $V_K(t)$ was originally defined only for braids, it follows from the theorems of Markov (see [B2]) and Alexander [ALEX1] that (due to the Markov property of the Jones trace) it is well-defined for arbitrary knots and links.

These results of Markov and Alexander are worth remarking upon here. First of all there is **Alexander's Theorem: Each link in three-dimensional space is ambient isotopic to a link in the form of a closed braid.**

Braids. A braid is formed by taking n points in a plane and attaching strands to these points so that parallel planes intersect the strands in n points. It is usually assumed that the braid begins and terminates in the same arrangement of points so that it has the diagrammatic form



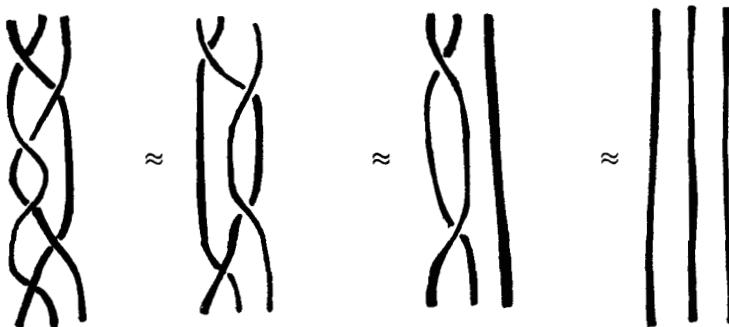
Here I have illustrated a 3-strand braid $b \in B_3$, and its closure \bar{b} . The closure \bar{b} of a braid b is obtained by connecting the initial points to the end-points by a collection of parallel strands.

It is interesting and appropriate to think of the braid as a diagram of a physical process of particles interacting or moving about in the plane. In the diagram, we take the arrow of time as moving up the page. Each plane (spatial plane) intersects the page perpendicularly in a horizontal line. Thus successive slices give a picture of the motions of the particles whose world-lines sweep out the braid.

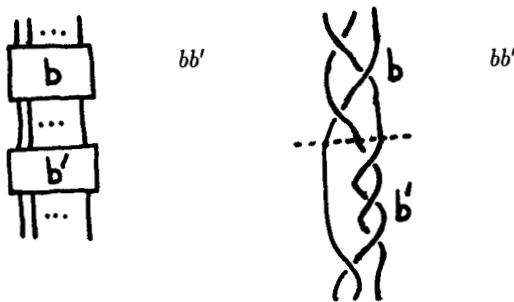
Of course, from the topological point of view one wants to regard the braid as a purely spatial weave of descending strands that are fixed at the top and the bottom of the braid. Two braids in B_n are said to be **equivalent** (and we write $b = b'$ for this equivalence) if there is an ambient isotopy from b to b' that keeps

the end-points fixed and does not move any strands outside the space between the top and bottom planes of the braids. (It is assumed that b and b' have identical input and output points.)

For example, we see the following equivalence

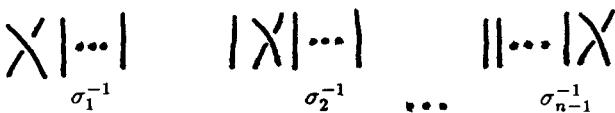
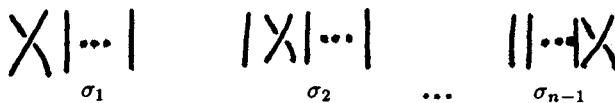


The braid consisting in n parallel descending strands is called the **identity braid** in B_n and is denoted by 1 or 1_n if need be. B_n , the collection of n -strand braids, up to equivalence, (i.e. the set of equivalence classes of n -strand braids) is a group - the **Artin Braid Group**. Two braids b, b' are multiplied by joining the output strands of b to the input strands of b' as indicated below:



Every braid can be written as a product of the **generators** $\sigma_1, \sigma_2, \dots, \sigma_{n-1}$ and their **inverses** $\sigma_1^{-1}, \sigma_2^{-1}, \dots, \sigma_{n-1}^{-1}$. These elementary braids σ_i and σ_i^{-1} are obtained by interchanging only the i -th and $(i+1)$ -th points in the row of inputs.

Thus



These generators provide a convenient way to catalog various weaving patterns.
For example



$$= \sigma_1^{-1} \sigma_2 \sigma_1^{-1} \sigma_2 \sigma_1^{-1} \sigma_2$$

A 360° twist in the strands has the appearance

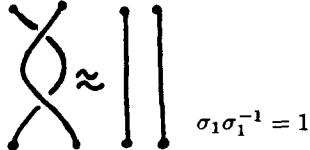


$$\begin{aligned} & \sigma_1 \sigma_2 \sigma_1 \sigma_2 \sigma_1 \sigma_2 \\ & || \\ & (\sigma_1 \sigma_2)^3. \end{aligned}$$

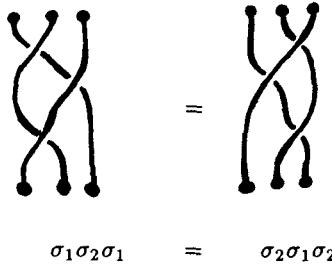
The braid group B_n is completely described by these generators and relations.
The relations are as follows:

$$\left\{ \begin{array}{ll} \sigma_i \sigma_i^{-1} = 1, & i = 1, \dots, n-1 \\ \sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}, & i = 1, \dots, n-2 \\ \sigma_i \sigma_j = \sigma_j \sigma_i, & |i-j| > 1. \end{array} \right.$$

Note that the first relation is a version of the type II move,

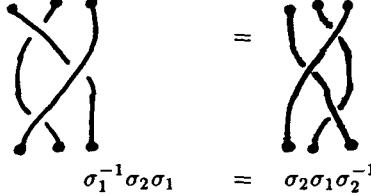


while the second relation is a type III move:



Note that since $\sigma_1 \sigma_2 \sigma_1 = \sigma_2 \sigma_1 \sigma_2$ is stated in the **group B_3** , we also know that $(\sigma_1 \sigma_2 \sigma_1)^{-1} = (\sigma_2 \sigma_1 \sigma_2)^{-1}$, whence $\sigma_1^{-1} \sigma_2^{-1} \sigma_1^{-1} = \sigma_2^{-1} \sigma_1^{-1} \sigma_2^{-1}$.

There are, however, a few other cases of the type III move. For example:



However, this is algebraically equivalent to the relation $\sigma_2 \sigma_1 \sigma_2 = \sigma_1 \sigma_2 \sigma_1$ (multiply both sides by σ_1 on the left, and σ_2 on the right). In fact, we can proceed directly as follows:

$$\begin{aligned}\sigma_1^{-1} \sigma_2 \sigma_1 &= \sigma_1^{-1} (\sigma_2 \sigma_1 \sigma_2) \sigma_2^{-1} \\ &= \sigma_1^{-1} (\sigma_1 \sigma_2 \sigma_1) \sigma_2^{-1} \\ &= (\sigma_1^{-1} \sigma_1) (\sigma_2 \sigma_1 \sigma_2^{-1}) \\ &= \sigma_2 \sigma_1 \sigma_2^{-1}.\end{aligned}$$

I emphasize this form of the equivalence because it shows that the type III move with a mixture of positive and negative crossings can be accomplished via a combination of type II moves and type III moves where all the crossings have the same sign.

Note that there is a homomorphism π of the braid group B_n onto the permutation group S_n on the set $\{1, 2, \dots, n\}$. The map $\pi : B_n \rightarrow S_n$ is defined by taking the permutation of top to bottom rows of points afforded by the braid. Thus

$$\pi \begin{bmatrix} 1 & 2 & 3 \\ \diagdown & \diagup & \diagdown \\ 1 & 2 & 3 \end{bmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

where the notation on the right indicates a permutation $\rho : \{1, 2, 3\} \rightarrow \{1, 2, 3\}$ with $\rho(1) = 3$, $\rho(2) = 1$, $\rho(3) = 2$. If $\rho = \pi(b)$ for a braid b , then $\rho(i) = j$ where j is the lower endpoint of the braid strand that begins at point i .

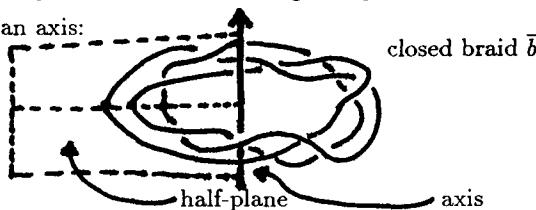
Letting $\tau_k : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ denote the transposition of k and $k+1$: $\tau_k(i) = i$ if $i \neq k, k+1$, $\tau_k(k) = k+1$, $\tau_k(k+1) = k$. We have that $\pi(\sigma_i) = \tau_i$, $i = 1, \dots, n-1$. In terms of these transpositions, S_n has the presentation

$$S_n = (\tau_1, \dots, \tau_{n-1} \mid \tau_i^2 = 1, \tau_i \tau_{i+1} \tau_i = \tau_{i+1} \tau_i \tau_{i+1}).$$

The permutation group is the quotient of the braid group B_n , obtained by setting the squares of all the generators equal to the identity.

Alexander's Theorem.

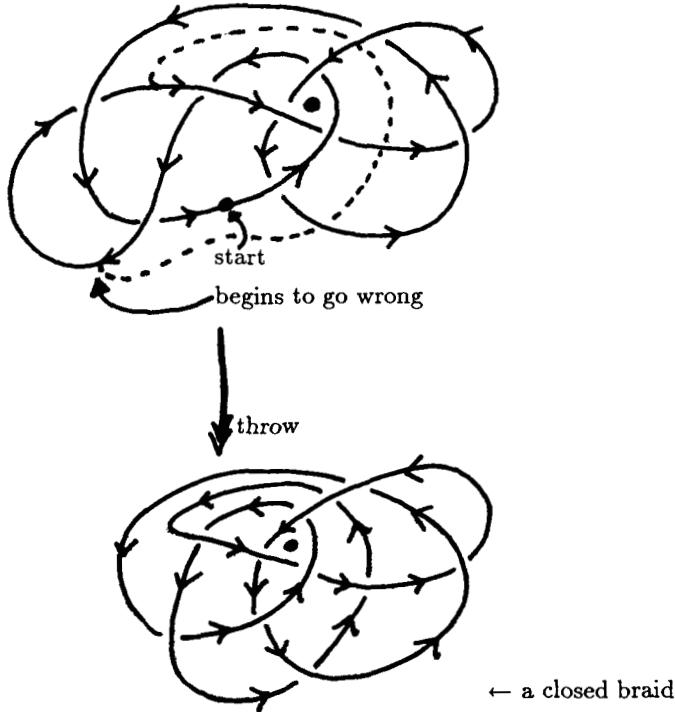
As we mentioned a few paragraphs ago, Alexander proved [ALEX1] that any knot or link could be put in the form of a closed braid (via ambient isotopy). Alexander proved this result by regarding a closed braid as a looping of the knot around an axis:



Thinking of three-dimensional space as a union of half-planes, each sharing the axis, we require that \bar{b} intersect each half-plane in the same number of points – (the number of braid-strands). As you move along the knot or link, you are circulating the axis in either a clockwise or counterclockwise orientation. Alexander's method was to choose a **proposed braid axis**. Then follow along the knot or link,

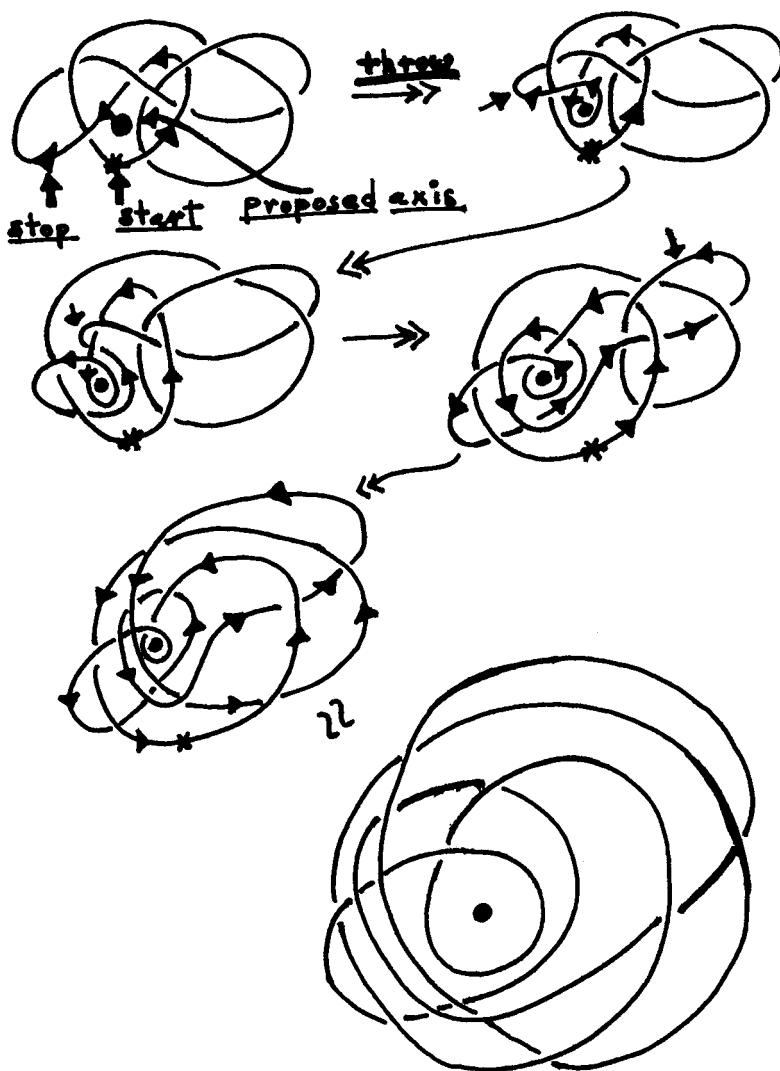
throwing the strand over the axis whenever it began to circulate incorrectly. Eventually, you have the link in braid form.

Figure 12 illustrates this process for a particular choice of axis. Note that it is clear that this process will not always produce the most efficient braid representation for a given knot or link. In the example of Figure 12 we would have fared considerably better if we had taken the axis at a different location - as shown below.



One throw over the new axis is all that is required to obtain this braid.

These examples raise the question: **How many different ways can a link be represented as a closed braid?**



Alexander's Theorem

Figure 12

There are some simple ways to modify braids so that their closures are ambient isotopic links. First there is the **Markov move**: Suppose β is a braid word in B_n (hence a word in $\sigma_1, \sigma_2, \dots, \sigma_{n-1}$ and their inverses). Then the three braids β , $\beta\sigma_n$ and $\beta\sigma_n^{-1}$ all have ambient isotopic closures. For example,

$$\beta = \text{Diagram of a braid} = \sigma_1^{-1}\sigma_2\sigma_1^{-1}\sigma_2 \in B_3$$

$$\beta\sigma_3 = \text{Diagram of a braid with three strands} \in B_4$$

$$\overline{\beta} = \text{Diagram of a closed link} \quad \overline{\beta\sigma_3} = \text{Diagram of a closed link}$$

Thus $\overline{\beta\sigma_n^{\pm 1}}$ is obtained from $\overline{\beta}$ by a type I Reidemeister move.

A somewhat more diabolical way to make a braid with the same closure is to choose any braid g in B_n and take the **conjugate braid** $g\beta g^{-1}$. When we close $g\beta g^{-1}$ to form $\overline{g\beta g^{-1}}$ the braid g and its inverse g^{-1} can cancel each other out by interacting through the closure strands. The fundamental theorem that relates the theory of knots and the theory of braids is the

Markov Theorem 7.1. Let $\beta_n \in B_n$ and $\beta'_m \in B_m$ be two braids in the braid groups B_n and B_m respectively. Then the links (closures of the braids β , β') $L = \overline{\beta_n}$ and $L' = \overline{\beta'_m}$ are ambient isotopic if and only if β'_m can be obtained from β_n by a series of

- 1) equivalences in a given braid group.
- 2) conjugation in a given braid group. (That is, replace a braid by some conjugate of that braid.)
- 3) **Markov moves:** (A Markov move replaces $\beta \in B_n$ by $\beta\sigma_n^{\pm 1} \in B_{n+1}$ or the inverse of this operation - replacing $\beta\sigma_n^{\pm 1} \in B_{n+1}$ by $\beta \in B_n$ if β has no occurrence of σ_n .)

For a proof of the Markov theorem the reader may wish to consult [B2].

The reader may enjoy pondering the question: How can Alexander's technique for converting links to braids be done in an algorithm that a computer can perform? (See [V].)

With the Markov theorem, we are in possession of the information needed to use the presentations of the braid groups B_n to extract topological information about knots and links. In particular, it is now possible to explain how the Jones polynomial works in relation to braids. For suppose that we are given a commutative ring R (polynomials or Laurent polynomials for example), and functions $J_n : B_n \rightarrow R$ from the n -strand braid group to the ring R , defined for each $n = 2, 3, 4, \dots$. Then the Markov theorem assures us that the family of functions $\{J_n\}$ can be used to construct link invariants if the following conditions are satisfied:

1. If b and b' are equivalent braid words, then $J_n(b) = J_n(b')$. (This is just another way of saying that J_n is well-defined on B_n .)
2. If $g, b \in B_n$ then $J_n(b) = J_n(gbg^{-1})$.
3. If $b \in B_n$, then there is a constant $\alpha \in R$, independent of n , such that

$$J_{n+1}(b\sigma_n) = \alpha^{+1} J_n(b)$$

$$J_{n+1}(b\sigma_n^{-1}) = \alpha^{-1} J_n(b).$$

We see that for the closed braid $\overline{b'} = \overline{b\sigma_n}$ the result of the Markov move $b \mapsto b'$ is to perform a type I move on \overline{b} . Furthermore, $\overline{b\sigma_n}$ corresponds to a type I move

of positive type, while $\overline{b\sigma_n^{-1}}$ corresponds to a type I move of negative type. It is for this reason that I have chosen the conventions for α and α^{-1} as above. Note also that, orienting a braid downwards, as in



has positive crossings corresponding to σ_i 's with positive exponents.

With these remarks in mind, let's define the **writhe of a braid**, $w(b)$, to be its **exponent sum**. That is, we let $w(b) = \sum_{\ell=1}^k a_\ell$ in any braid word

$$\sigma_{i_1}^{a_1} \sigma_{i_2}^{a_2} \dots \sigma_{i_k}^{a_k}$$

representing b . From our previous discussion of the writhe, it is clear that $w(b) = w(\bar{b})$ where \bar{b} is the oriented link obtained by closing the braid b (with downward-oriented strands). Here $w(\bar{b})$ is the writhe of the oriented link \bar{b} .

Definition 7.2. Let $\{J_n : B_n \rightarrow R\}$ be given with properties 1., 2., 3. as listed above. Call $\{J_n\}$ a **Markov trace** on $\{B_n\}$. For any link L , let $L \sim \bar{b}$, $b \in B_n$ via Alexander's theorem. Define $J(L) \in R$ via the formula

$$J(L) = \alpha^{-w(b)} J_n(b).$$

Call $J(L)$ the **link invariant for the Markov trace** $\{J_n\}$.

Proposition 7.3. Let J be the link invariant corresponding to the Markov trace $\{J_n\}$. Then J is an invariant of ambient isotopy for oriented links. That is, if $L \sim L'$ (\sim denotes ambient isotopy) then $J(L) = J(L')$.

Proof. Suppose, by Alexander's theorem, that $L \sim \bar{b}$ and $L' \sim \bar{b}'$ where $b \in B_n$ and $b' \in B_n$ are specific braids. Since L and L' are ambient isotopic, it follows that \bar{b} and \bar{b}' are also ambient isotopic. Hence \bar{b}' can be obtained from \bar{b} by a sequence of Markov moves of the type 1., 2., 3. Each such move leaves the function $\alpha^{-w(b)} J_n(b)$ ($b \in B_n$) invariant since the exponent sum is invariant under conjugation, braid moves, and it is used here to cancel the effect of the type 3. Markov move. This completes the proof. //

The Bracket for Braids.

Having discussed generalities about braids, we can now look directly at the bracket polynomial on closed braids. In the process, the structure of the Jones polynomial and its associated representations of the braid groups will naturally emerge.

In order to begin this discussion, let's define $\langle \cdot \rangle : B_n \rightarrow \mathbf{Z}[A, A^{-1}]$ via $\langle b \rangle = \langle \bar{b} \rangle$, the evaluation of the bracket on the closed braid b . In terms of the Markov trace formalism, I am letting $J_n : B_n \rightarrow \mathbf{Z}[A, A^{-1}] = R$ via $J_n(b) = \langle \bar{b} \rangle$. In fact, given what we know about the bracket from section 3⁰, it is obvious that $\{J_n\}$ is a Markov trace, with $\alpha = -A^3$.

Now consider the **states of a braid**. That is, consider the states determined by the recursion formula for the bracket:

$$\langle \text{X} \rangle = A \langle \text{ } \rangle \langle \text{ } \rangle + A^{-1} \langle \text{U} \rangle.$$

In terms of braids this becomes

$$\langle \| \cdots | \text{X} | \cdots \| \rangle = A \langle \| \cdots | \| \| \cdots \| \rangle + A^{-1} \langle \| \cdots | \text{U} | \cdots \| \rangle$$

$$\langle \sigma_i \rangle = A \langle 1_n \rangle + A^{-1} \langle U_i \rangle$$

where 1_n denotes the identity element in B_n (henceforth denoted by 1), and U_i is a **new** element written in braid input-output form, but with a cup-U cap-U

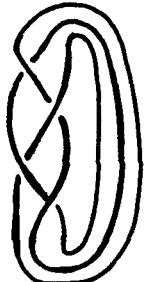
combination  at the i -th and $(i+1)$ -th strands:

$$\text{U} \| \quad , \quad \| \text{U} \quad , \quad \| \| \text{U} \quad (\text{for 4-strands})$$

$U_1 \qquad \qquad U_2 \qquad \qquad U_3$

Since a state for \bar{b} is obtained by choosing splice direction for each crossing of b , we see that **each state of \bar{b} can be written as the closure of an (input-output) product of the elements U_i** . (See section 3⁰ for a discussion of bracket states.)

For example, let $L = \bar{b}$ be the link

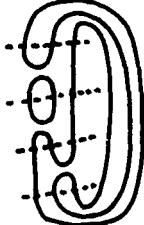


L

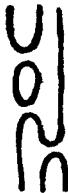


$$b = \sigma_1^{-1} \sigma_2 \sigma_1^{-1} \sigma_2$$

Then the state s of L shown below corresponds to the product $U_1^2 U_2$.



$s \leftrightarrow$



$$\overline{U_1^2 U_2} = s$$

In fact it is clear that we can use the following formalism: Write

$$\sigma_i \equiv A + A^{-1} U_i, \quad \sigma_i^{-1} \equiv A^{-1} + A U_i$$

Given a braid word b , write $b \equiv \mathcal{U}(b)$ where $\mathcal{U}(b)$ is a sum of products of the U_i 's, obtained by performing the above substitutions for each σ_i . Each product of U_i 's, when closed gives a collection of loops. Thus if U is such a product, then $\langle U \rangle = \langle \bar{U} \rangle = \delta^{\|U\|}$ where $\|U\| = \#(\text{of loops in } \bar{U}) - 1$ and $\delta = -A^2 - A^{-2}$. Finally if $\mathcal{U}(b)$ is given by

$$\mathcal{U}(b) = \sum_s \langle b | s \rangle U_s$$

where s indexes all the terms in the product, and $\langle b | s \rangle$ is the product of A 's and A^{-1} 's multiplying each U -product U_s , then

$$\begin{aligned} \langle b \rangle &= \langle \mathcal{U}(b) \rangle = \sum_s \langle b | s \rangle \langle U_s \rangle \\ \langle b \rangle &= \sum_s \langle b | s \rangle \delta^{\|s\|}. \end{aligned}$$

This is the braid-analog of the state expansion for the bracket.

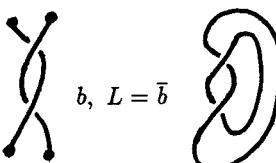
Example. $b = \sigma_1^2$.

Then

$$\mathcal{U}(b) = (A + A^{-1}U_1)(A + A^{-1}U_1)$$

$$\mathcal{U}(b) = A^2 + 2U_1 + A^{-2}U_1^2$$

$$\langle b \rangle = \langle \mathcal{U}(b) \rangle = A^2\langle 1_2 \rangle + 2\langle U_1 \rangle + A^{-2}\langle U_1^2 \rangle$$

$$\left\{ \begin{array}{l} \parallel \Rightarrow \langle 1_2 \rangle = \delta \\ \cap \Rightarrow \langle U_1 \rangle = 1 \\ \cup \Rightarrow \langle U_1^2 \rangle = \delta \end{array} \right\}$$


$$\therefore \langle L \rangle = A^2(-A^2 - A^{-2}) + 2 + A^{-2}(-A^2 - A^{-2})$$

$$= -A^4 - 1 + 2 - 1 - A^{-4}$$

$$\langle L \rangle = -A^4 - A^{-4}.$$

This is in accord with our previous calculation of the bracket for the simple link of two components.

The upshot of these observations is that in calculating the bracket for braids in B_n it is useful to have the **free additive algebra** \mathcal{A}_n with generators U_1, U_2, \dots, U_{n-1} and multiplicative relations coming from the interpretation of the U_i 's as cup-cap combinations. This algebra \mathcal{A}_n will be regarded as a module over the ring $\mathbf{Z}[A, A^{-1}]$ with $\delta = -A^2 - A^{-2} \in \mathbf{Z}[A, A^{-1}]$ the designated loop value. I shall call \mathcal{A}_n the **Temperley-Lieb Algebra** (see [BA1], [LK4]).

What are the multiplicative relations in \mathcal{A}_n ? Consider the pictures in Figure 13. They illustrate the relations:

$$[\mathcal{A}] \quad \left\{ \begin{array}{rcl} U_i U_{i \pm 1} U_i & = & U_i \\ U_i^2 & = & \delta U_i \\ U_i U_j & = & U_j U_i, \text{ if } |i - j| > 1 \end{array} \right\}.$$

$$\begin{array}{ccc}
 \text{Diagram 1} & \cong & \text{Diagram 2} \\
 U_1U_2U_1 & = & U_1
 \end{array}$$

$$\begin{array}{ccc}
 \text{Diagram 3} & \cong & \text{Diagram 4} \\
 U_1^2 & = & \delta U_1
 \end{array}$$

$$\begin{array}{ccc}
 \text{Diagram 5} & \cong & \text{Diagram 6} \\
 U_1U_3 & = & U_3U_1
 \end{array}$$

Figure 13

In fact, these are precisely the relations for The Temperley-Lieb algebra. Note that the Temperley-Lieb algebra and the Jones algebra are closely related. In fact, if we define $e_i = \delta^{-1}U_i$, then $e_i^2 = e_i$ and $e_i e_{i \pm 1} e_i = \tau e_i$ where $\tau = \delta^{-2}$. Thus, by considering the state expansion of the bracket polynomial for braids, we recover the formal structure of the original Jones polynomial.

It is convenient to view A_n in a more fundamental way: Let D_n denote the collection of all (topological) equivalence classes of diagrams obtained by connecting pairs of points in two parallel rows of n points. The arcs connecting these points must satisfy the following conditions:

- 1) All arcs are drawn in the space between the two rows of points.
- 2) No two arcs cross one another.
- 3) Two elements $a, b \in D_n$ are said to be **equivalent** if they are topologically equivalent via a planar isotopy through elements of D_n . (That is, if there is a continuous family of embeddings of arcs - giving elements $C_t \in D_n$ ($0 \leq t \leq 1$) with $C_0 = a$, $C_1 = b$ and C_t the identity map on the subset of endpoints for each t , $0 \leq t \leq 1$).

Call D_n the **diagram monoid on $2n$ points**.

Example. For $n = 3$, D_3 has the following elements:

$$\begin{array}{ccccc}
 \text{Diagram } l_3 & \quad \text{Diagram } U_1 & \quad \text{Diagram } U_2 & \quad \text{Diagram } r & \quad \text{Diagram } s
 \end{array}$$

Elements of the diagram monoid D_n are multiplied like braids - by attaching the output row of a to the input row of b - forming ab . Multiplying in this way, closed loops may appear in ab . Write $ab = \delta^k c$ where $c \in D_n$, and k is the number of closed loops in the product.

For example, in D_3

$$rs = \begin{array}{c} \text{Diagram of } rs \\ \text{a braid with two loops} \end{array} = \delta \begin{array}{c} \text{Diagram of } c \\ \text{a braid with one loop} \end{array} = \delta U_2.$$

Proposition 7.4. The elements $1, U_1, U_2, \dots, U_{n-1}$ generate D_n . If an element $x \in D_n$ is equivalent to two products, P and Q , of the elements $\{U_i\}$, then Q can be obtained from P by a series of applications of the relations $[\mathcal{A}]$.

See [LK8] for the proof of this proposition. The point of this proposition is that it lays bare the underlying combinatorial structure of the Temperley-Lieb algebra. And, for computational purposes, the multiplication table for D_n can be obtained easily with a computer program.

We can now define a mapping

$$\rho : B_n \rightarrow \mathcal{A}_n$$

by the formulas:

$$\begin{aligned} \rho(\sigma_i) &= A + A^{-1}U_i \\ \rho(\sigma_i^{-1}) &= A^{-1} + AU_i. \end{aligned}$$

We have seen that for a braid b , $\langle \bar{b} \rangle = \sum_s \langle B|s\rangle \langle \bar{U}_s \rangle$ where $\rho(b) = \sum_s \langle b|s\rangle U_s$ is the explicit form of $\rho(b)$ obtained by defining $\rho(xy) = \rho(x)\rho(y)$ on products. (s runs through all the different products in this expansion.) Here $\langle \bar{U}_s \rangle$ counts one less than the number of loops in \bar{U}_s .

Define $\text{tr} : \mathcal{A}_n \rightarrow \mathbb{Z}[A, A^{-1}]$ by $\text{tr}(U) = \langle \bar{U} \rangle$ for $U \in D_n$. Extend tr linearly to \mathcal{A}_n . This mapping - by loop counts - is a realization of Jones' trace on the von Neumann algebra A_n . We then have the formula: $\langle b \rangle = \text{tr}(\rho(b))$.

This formalism explains directly how the bracket is related to the construction of the Jones polynomial via a trace on a representation of the braid group to the Temperley-Lieb algebra.

We need to check certain things, and some comments are in order. First of all, the trace on the von Neumann algebra A_n was not originally defined diagrammatically. It was, defined in [JO7] via normal forms for elements of the Jones algebra A_n . Remarkably, this version of the trace matches the diagrammatic loop count. In the next section, we'll see how this trace can be construed as a modified matrix trace in a representation of the Temperley-Lieb algebra.

Proposition 7.5. $\rho : B_n \rightarrow A_n$, as defined above, is a representation of the Artin Braid group.

Proof. It is necessary to verify that $\rho(\sigma_i)\rho(\sigma_i^{-1}) = 1$, $\rho(\sigma_i\sigma_{i+1}\sigma_i) = \rho(\sigma_{i+1}\sigma_i\sigma_{i+1})$ and that $\rho(\sigma_i\sigma_j) = \rho(\sigma_j\sigma_i)$ when $|i - j| > 1$. We shall do these in the order - first, third, second.

First.

$$\begin{aligned} \rho(\sigma_i)\rho(\sigma_i^{-1}) &= (A + A^{-1}U_i)(A^{-1} + AU_i) \\ &= 1 + (A^{-2} + A^2)U_i + U_i^2 \\ &= 1 + (A^{-2} + A^2)U_i + \delta U_i \\ &= 1 + (A^{-2} + A^2)U_i + (-A^{-2} - A^2)U_i \\ &= 1 \end{aligned}$$

Third. Given that $|i - j| > 1$:

$$\begin{aligned} \rho(\sigma_i\sigma_j) &= \rho(\sigma_i)\rho(\sigma_j) \\ &= (A + A^{-1}U_i)(A + A^{-1}U_j) \\ &= (A + A^{-1}U_j)(A + A^{-1}U_i), [U_iU_j = U_jU_i \text{ if } |i - j| > 1] \\ &= \rho(\sigma_j\sigma_i). \end{aligned}$$

Second.

$$\begin{aligned}
 \rho(\sigma_i \sigma_{i+1} \sigma_i) &= (A + A^{-1}U_i)(A + A^{-1}U_{i+1})(A + A^{-1}U_i) \\
 &= (A^2 + U_{i+1} + U_i + A^{-2}U_iU_{i+1})(A + A^{-1}U_i) \\
 &= A^3 + AU_{i+1} + AU_i + A^{-1}U_iU_{i+1} + A^{-1}U_i^2 + AU_i + A^{-1}U_{i+1}U_i \\
 &\quad + A^{-3}U_iU_{i+1}U_i \\
 &= A^3 + AU_{i+1} + (A^{-1}\delta + 2A)U_i + A^{-1}(U_iU_{i+1} + U_{i+1}U_i) + A^{-3}U_i \\
 &= A^3 + AU_{i+1} + (A^{-1}(-A^2 - A^{-2}) + 2A + A^{-3})U_i \\
 &\quad + A^{-1}(U_iU_{i+1} + U_{i+1}U_i) \\
 &= A^3 + A(U_{i+1} + U_i) + A^{-1}(U_iU_{i+1} + U_{i+1}U_i).
 \end{aligned}$$

Since this expression is symmetric in i and $i + 1$, we conclude that

$$\rho(\sigma_i \sigma_{i+1} \sigma_i) = \rho(\sigma_{i+1} \sigma_i \sigma_{i+1}).$$

This completes the proof that $\rho : B_n \rightarrow A_n$ is a representation of the Artin Braid Group. //

8⁰. Abstract Tensors and the Yang-Baxter Equation.

In this section and throughout the rest of the book I begin a notational convention that I dub **abstract (diagrammatic) tensors**. It is really a diagrammatic version of matrix algebra where the matrices have many indices. Since tensors are traditionally such objects - endowed with specific transformation properties - the subject can be called abstract tensors.

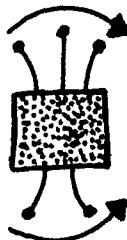
The diagrammatic aspect is useful because it enables us to hide indices. For example, a matrix $M = (M_{ij}^i)$ with entries M_{ij}^i for i and j in an index set \mathcal{I} will be a box with an upper strand for the upper index, and a lower strand for the lower index:

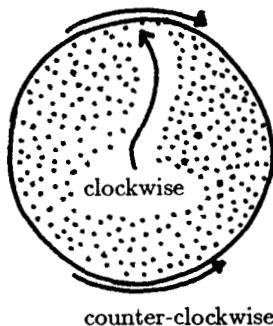
$$M = (m_j^i) \leftrightarrow \boxed{M} .$$


In general, a tensor-like object has some upper and lower indices. These become lines or strands emanating from the corresponding diagram:

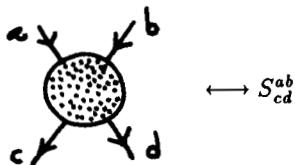
$$T_{\ell m}^{ijk} \leftrightarrow \boxed{T}$$


Usually, I will follow standard typographical conventions with these diagrams. That is, the upper indices are in correspondence with upper strands and ordered from left to right, the lower indices correspond to lower strands and are also ordered from left to right. Note that because the boundary of the diagram may be taken to be a Jordan curve in the plane, the upper left-right order proceeds clockwise on the body of the tensor, while the lower left-right order proceeds counter-clockwise.

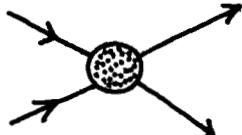




If we wish to discriminate indices in some way that is free of a given convention of direction, then the corresponding lines can be labelled. The simplest labelling scheme is to put an arrow on the line. Thus we may write



to indicate “inputs” $\{a, b\}$ and “outputs” $\{c, d\}$. Here we see one of the properties of the diagrammatic notation: Diagrammatic notation acts as a mathematical metaphor. The picture



reminds us of a process (input/output, scattering). But we are free to take this reminder or leave it according to the demands of context and interpretation.

Normally, I take the **strand**



to denote a Kronecker delta:

$$\text{Diagram showing two strands } a \text{ and } b \text{ meeting at a point.} \leftrightarrow \delta_a^b = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}$$

There are contexts in which some bending of the strand, as in $a \curvearrowright b \leftrightarrow M_{ab}$ will not be a Kronecker delta, but we shall deal with these later.

Generalized Multiplication.

Recall the definition of matrix multiplication: $(MN)_j^i = \sum_k M_k^i N_j^k$. This is often written as $M_i^k M_j^k$ where it is assumed that one sums over all occurrences of a pair of repeated upper and lower indices (Einstein summation convention). The corresponding convention in the diagrammatics is that a line that connects two index strands connotes the summation over all occurrences of an index value for that line.

$$\sum_k M_k^i N_j^k \leftrightarrow M_i^k N_j^k \leftrightarrow \boxed{M} \text{ (box)} \text{ (box)} N = \boxed{M} \text{ (box)} \text{ (box)} N$$

$\text{trace}(M) = \text{tr}(M) = \sum_i M_i^i \leftrightarrow M_i^i \leftrightarrow \boxed{M}$

Thus matrix multiplication corresponds to plugging one box into another, and the standard matrix trace corresponds to plugging a box into itself.

We can now write out various products diagrammatically:

$$M_{ij}^{ab} N_{kl}^{ij} A_c^l B_d^k \leftrightarrow \boxed{M} \text{ (box)} \text{ (box)} N \text{ (box)} \text{ (box)} A \text{ (box)} \text{ (box)} B$$

In this last example a crossover occurs - due to index reordering. Here I adopt the convention that

$$\text{Diagram showing two strands } c \text{ and } d \text{ crossing.} \leftrightarrow \delta_d^a \delta_c^b.$$

The crossed lines are independent Kronecker deltas.

Exercise. Let the index set $\mathcal{I} = \{1, 2\}$ and define

$$\epsilon_{ij} = \begin{cases} 1 & i < j \\ -1 & i > j \\ 0 & \text{otherwise} \end{cases}$$

and $\epsilon^{ij} = \epsilon_{ij}$. Let \prod be the diagram for ϵ_{ij} and Π the diagram for ϵ^{ij} . Show that

$$\prod = \boxed{\begin{array}{c} \text{II} \\ \prod \end{array}} = \boxed{\begin{array}{c} \text{II} \\ \prod \end{array}} - \times \quad \prod = \epsilon_{ij}$$

$$\left[\text{Note } \prod = \prod \prod \right]$$

You will find that this is equivalent to saying

$$\epsilon^{ab} \epsilon_{cd} = \delta_c^a \delta_d^b - \delta_d^a \delta_c^b .$$

This identity will be useful later on.

To return to the crossed lines, they are not quite innocuous. They do represent a permutation. Thus, if $\mathcal{I} = \{1, 2\}$ as in the exercise, and

$$P_{cd}^{ab} = \delta_d^a \delta_c^b = \boxed{\times},$$

then P is a matrix representing a transposition. Similarly, for three lines, we have

$$\{ \boxed{\times I}, \boxed{I \times}, \boxed{\times \times}, \boxed{\times \times}, \boxed{\times \times}, \boxed{III} \}$$

representing S_3 , the symmetric group on three letters.

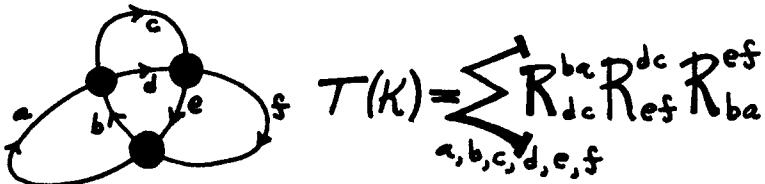
Knot Diagrams as Abstract Tensor Diagrams.

By now it must have become clear that we have intended all along to interpret a diagram of a knot or link as an abstract tensor diagram. How can this be done? Actually there is more than one way to do this, but the simplest is to use an oriented diagram and to associate two matrices to the two types of crossing. Thus,

$$\begin{array}{ccc} \times \rightarrow & \leftrightarrow & = R_{cd}^{ab} \\ \times \leftarrow & \leftrightarrow & = \overline{R}_{cd}^{ab} \end{array}$$

With this convention, any oriented link diagram K is mapped to a specific contracted (no free lines) abstract tensor $T(K)$.

Example.  $K \mapsto$  $T(K)$. Thus, if we label the lines of $T(K)$ with indices, then $T(K)$ corresponds to a formal product - or rather a sum of products, since repeated indices connote summations:



If there is a commutative ring \mathcal{R} and an index set \mathcal{I} such that $a, b, c, \dots \in \mathcal{I}$ and $R_{cd}^{ab} \in \mathcal{R}$ then we can write

$$T(K) = \sum_{a,b,\dots,f \in \mathcal{I}} R_{dc}^{ba} R_{ef}^{dc} R_{ba}^{ef},$$

and one can regard a choice of labels from \mathcal{I} for the edges of $T(R)$ as a state of K . That is, a state σ of K is a mapping $\sigma : E(K) \rightarrow \mathcal{I}$ where E denotes the edge set of K and \mathcal{I} is the given index set.

In this regard, we are seeing $T(K)$ as identical to the oriented graph that underlies the link diagram K with labelled nodes (black or white) corresponding to the crossing type in K . Such a diagram then translates directly into a formal product by associating R_{cd}^{ab} with positive crossings and \bar{R}_{cd}^{ab} with negative crossings.

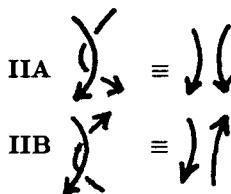
Using the concept of a state σ we can rewrite

$$T(K) = \sum_{\sigma} \langle K | \sigma \rangle$$

where σ runs over all states of K , and $\langle K | \sigma \rangle$ denotes the product of the vertex weights R_{cd}^{ab} (or \bar{R}_{cd}^{ab}) assigned to the crossings of K by the given state.

We wish to see, under what circumstances $T(K)$ will be invariant under the Reidemeister moves. In particular, the relevant question is invariance under moves of type II and type III (regular isotopy). We shall see that the model, using abstract tensors, is good for constructing representations of the Artin braid group. It requires modification to acquire regular isotopy invariance.

But let's take things one step at a time. There are two versions of the type II move:



It is necessary that the model, as a whole, be invariant under both of these moves. The move (IIA) corresponds to a very simple matrix condition on R and \bar{R} :

\equiv

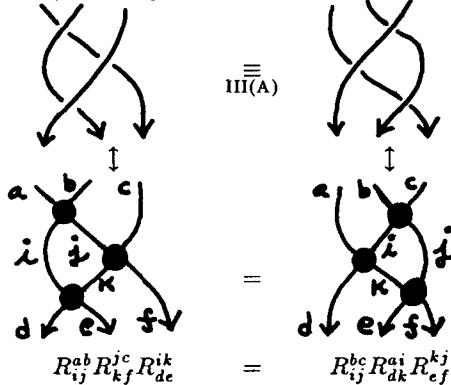
$\left. \begin{array}{c} \text{channel unitarity} \\ \bar{R}_{ij}^{ab} R_{cd}^{ij} = \delta_c^a \delta_d^b \end{array} \right\}$

That is, (IIA) will be satisfied if R and \bar{R} are inverse matrices. Call this condition on R **channel unitarity**. The direct requirements imposed by move (IIB) could be termed **cross-channel unitarity**:

\equiv

$\left. \begin{array}{c} \text{cross-channel unitarity} \\ R_{jb}^{ia} R_{ic}^{jd} = \delta_c^a \delta_b^d \end{array} \right\}$

Recall that we already discussed the concepts of channel and cross-channel unitarity in section 6⁰. We showed that in order to have type III invariance, in the presence of channel and cross-channel unitarity, it is sufficient to demand invariance under the move of type III(A) [with all crossings positive (as shown below) or all crossings negative]. See Figure 14.



The Yang-Baxter Equation Corresponding to a Move of type III(A)
Figure 14

As shown in Figure 14, there is a matrix condition that will guarantee invariance of $T(K)$ under the moves III(A, +) and III(A, -) where the + or - signs refer to the crossing types in these moves. These equations are:

$$\begin{aligned} \text{III}(A, +) : \sum_{i,j,k \in \mathcal{I}} R_{ij}^{ab} R_{kf}^{jc} R_{de}^{ik} &= \sum_{i,j,k \in \mathcal{I}} R_{ij}^{bc} R_{dk}^{ai} R_{ef}^{kj} \\ \text{III}(A, -) : \sum_{i,j,k \in \mathcal{I}} \bar{R}_{ij}^{ab} \bar{R}_{kf}^{jc} \bar{R}_{de}^{ik} &= \sum_{i,j,k \in \mathcal{I}} \bar{R}_{ij}^{bc} \bar{R}_{dk}^{ai} \bar{R}_{ef}^{kj} \end{aligned}$$

and will be referred to as the **Yang-Baxter Equation** for R (\bar{R}). In the above forms, I have written these equations with a summation to indicate that the repeated indices are taken from the given index set \mathcal{I} .

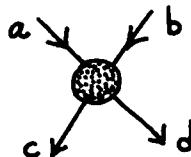
Thus we have proved the

Theorem 8.1. If the matrices R and \bar{R} satisfy

- 1) channel unitarity
- 2) cross-channel unitarity and
- 3) Yang-Baxter Equation

then $T(K)$ is a regular isotopy invariant for oriented diagrams K .

Remark. The Yang-Baxter Equation as it arises in mathematical physics involves extra parameters - sometimes called **rapidity** or **momentum**. In fact, if we interpret the vertex



as a particle interaction with incoming spins (or charges) a and b and outgoing spins (charges) c and d , then R_{cd}^{ab} can be taken to represent the scattering amplitude for this interaction. That is, it can be regarded as the probability amplitude for this particular combination of spins in and out.

Under these circumstances it is natural to also consider particle momenta and other factors.

For now it is convenient to consider only the spins. The conservation of spin (or charge) suggests the rule that $a+b = c+d$ whenever $R_{cd}^{ab} \neq 0$. This turns out to be a good starting place for the construction of solutions to the YBE (Yang-Baxter Equation).

Example. Let the index set $\mathcal{I} = \{1, 2, 3, \dots, n\}$ and let $R_{cd}^{ab} = A\delta_c^a\delta_d^b + A^{-1}\delta^{ab}\delta_{cd}$, where $n = -A^2 - A^{-2}$. That is, A is chosen to satisfy the equation

$$nA^2 + A^4 + 1 = 0.$$

Here δ_b^a and δ^{ab} are Kronecker deltas:

$$\delta_b^a = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b, \end{cases}$$

$$\delta^{ab} = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b. \end{cases}$$

Here we are transcribing this solution from the bracket model (section 3⁰) for the Jones polynomial!

Recall that the bracket is defined by the equations:

$$\langle \times \rangle = A \langle \) (\ \rangle + A^{-1} \langle \cup \cap \rangle$$

$$\langle \circ \rangle = -A^2 - A^{-2}.$$

See section 3⁰ for the other conventions for dealing with the bracket. This suggests a matrix model where

$$\begin{aligned} R_{cd}^{ab} &= \text{Diagram } \times \quad = A \left(\begin{array}{c} b \\ c \\ d \end{array} \right) \left(\begin{array}{c} b \\ d \end{array} \right) + A^{-1} \left(\begin{array}{c} b \\ c \\ d \end{array} \right) \cup \\ &= A \delta_c^a \delta_d^b + A^{-1} \delta^{ab} \delta_{cd}. \end{aligned}$$

In such a model the loop value, $\langle O \rangle$, is the trace of a Kronecker delta: $\langle O \rangle = \delta_a^a = n$ if $\mathcal{I} = \{1, 2, \dots, n\}$. Thus we require that A satisfies the equation

$$n = -A^2 - A^{-2}.$$

This completes the motivation picking the particular form of the R -matrix. Certain consistency checks are needed. For example, if without orientation we assigned

$$R_{cd}^{ab} \leftrightarrow \text{Diagram } \times \quad \text{↑ time's arrow}$$

then (viewing the interaction with time's arrow turned by 90°) we must have

$$\bar{R}_{db}^{ca} \leftrightarrow \text{Diagram } \times \quad \text{← time's arrow.}$$

Thus we need

$$\bar{R}_{db}^{ca} = R_{cd}^{ab}$$

whence

$$\begin{aligned}
\overline{R}_{cd}^{ab} &= R_{ac}^{bd} \\
&= A\delta_a^b \delta_c^d + A^{-1}\delta^{bd}\delta_{ac} \\
&= A^{-1}\delta_c^a \delta_d^b + A\delta^{ab}\delta_{cd} \\
&= A^{-1} \begin{array}{c} \nearrow b \\ \searrow d \\ c \end{array} + A \begin{array}{c} \nearrow b \\ \nwarrow d \\ c \end{array}
\end{aligned}$$

Essentially the same checks that proved the regular isotopy invariance of $\langle K \rangle$ now go over to prove that \overline{R} is the inverse of R , and that R and \overline{R} satisfy the YBE. This is a case where R and \overline{R} satisfy both channel and cross-channel unitarity.

As far as link diagrams are concerned, we could define an R -matrix via

$$R_{cd}^{ab} = A\delta_c^a \delta_d^b + B\delta^{ab}\delta_{cd}$$

with

$$\begin{array}{c} \nearrow b \\ \searrow d \\ c \end{array} \leftrightarrow A \quad \left(\begin{array}{c} \nearrow b \\ \nwarrow d \\ c \end{array} + B \begin{array}{c} \nearrow b \\ \searrow d \\ c \end{array} \right) \leftrightarrow R_{cd}^{ab}.$$

If the index set is $\mathcal{I} = \{1, 2, \dots, n\}$, then the resulting “tensor contraction” $T(K)$ will satisfy:

$$T\left(\begin{array}{c} \times \\ \diagup \\ \diagdown \end{array}\right) = AT\left(\begin{array}{c} \diagup \\ \diagdown \end{array}\right) + BT\left(\begin{array}{c} \diagup \\ \diagdown \\ \diagup \\ \diagdown \end{array}\right)$$

$$T(OK) = nT(K)$$

$$T(O) = n.$$

Thus $T(K) = \langle K \rangle$ where $\langle K \rangle$ is a generalized bracket with an integer value n for the loop.

A priori, specializing $B = A^{-1}$ and $n = -A^2 - A^{-2}$ does not guarantee that R will satisfy the YBE. But the fact that $\langle K \rangle$ is, under these conditions, invariant under the third Reidemeister move, provides strong motivation for a direct check on the R -matrix.

Checking the R -matrix.

Let

$$R_{cd}^{ab} = A \begin{pmatrix} a \\ c \end{pmatrix} \begin{pmatrix} b \\ d \end{pmatrix} + B \begin{array}{c} \text{---} \\ \diagup \quad \diagdown \\ c \quad d \end{array}$$
(notation)

where it is understood that the arcs denote Kronecker deltas with indices from an index set $\mathcal{I} = \{1, 2, \dots, n\}$. In this case, we can perform a direct expansion of the triangle configuration into eight terms:

$$\begin{aligned} \left[\begin{array}{c} \diagup \quad \diagdown \\ a \quad b \end{array} \right] &= A^3 \left[\begin{array}{|c|c|} \hline \diagup & \diagdown \\ \hline \end{array} \right] + A^2 B \left[\begin{array}{|c|c|} \hline \diagup & \diagdown \\ \hline \diagup & \diagdown \\ \hline \end{array} \right] \\ &\quad + AB^2 \left[\begin{array}{|c|c|} \hline \diagup & \diagdown \\ \hline \diagup & \diagdown \\ \hline \diagup & \diagdown \\ \hline \end{array} \right] \\ &\quad + B^3 \left[\begin{array}{|c|c|} \hline \diagup & \diagdown \\ \hline \diagup & \diagdown \\ \hline \diagup & \diagdown \\ \hline \end{array} \right] \end{aligned}$$

and

$$\begin{aligned} \left[\begin{array}{c} \diagup \quad \diagdown \\ a \quad b \end{array} \right] &= A^3 \left[\begin{array}{|c|c|} \hline \diagdown & \diagup \\ \hline \end{array} \right] + A^2 B \left[\begin{array}{|c|c|} \hline \diagdown & \diagup \\ \hline \diagdown & \diagup \\ \hline \end{array} \right] \\ &\quad + AB^2 \left[\begin{array}{|c|c|} \hline \diagdown & \diagup \\ \hline \diagdown & \diagup \\ \hline \diagdown & \diagup \\ \hline \end{array} \right] \\ &\quad + B^3 \left[\begin{array}{|c|c|} \hline \diagdown & \diagup \\ \hline \diagdown & \diagup \\ \hline \diagdown & \diagup \\ \hline \end{array} \right] \end{aligned}$$

Letting n denote the loop value, we then have an equation for the difference:

$$\begin{aligned}
 \left[\begin{array}{c} \diagup \\ \diagdown \end{array} \right] - \left[\begin{array}{c} \diagdown \\ \diagup \end{array} \right] &= A^2 B \left[\begin{array}{c|c} \cup & \cup \\ \cap & \cap \end{array} \right] - \left[\begin{array}{c|c} \cup & \cup \\ \cap & \cap \end{array} \right] \\
 &\quad + AB^2 \left[\begin{array}{c|c} \cup & \cup \\ \cap & \cap \end{array} \right] - \left[\begin{array}{c|c} \cup & \cup \\ \cap & \cap \end{array} \right] \\
 &\quad + B^3 \left[\begin{array}{c|c} \cup & \cup \\ \cap & \cap \end{array} \right] \\
 &= (A^2 B + nAB^2 + B^3) \left[\begin{array}{c|c} \cup & \cup \\ \cap & \cap \end{array} \right]
 \end{aligned}$$

Thus in order for R to satisfy the YBE it is sufficient for

$$B(A^2 + nAB + B^2) = 0.$$

Thus $B = 0$ gives a trivial solution to YBE (a multiple of the identity matrix). Otherwise, we require that $A^2 + nAB + B^2 = 0$ and so (assuming $A \neq 0 \neq B$) we have $n = -(\frac{A}{B} + \frac{B}{A})$ for the loop value.

Given this choice of loop value it is interesting to go back and examine the behavior of the bracket under a type II move:

$$\begin{aligned}
 \langle \text{D}\text{C} \rangle &= AB \langle \text{D}\text{C} \rangle + (A^2 + B^2 + ABn) \langle \text{C}\text{C} \rangle \\
 &= AB \langle \text{D}\text{C} \rangle.
 \end{aligned}$$

Thus, even if $B \neq A^{-1}$, this model for the bracket will be invariant under the type II move up to a multiplication by a power of AB . I leave it as an exercise for the reader to see that normalization to an ambient isotopy invariant yields the usual version of the Jones polynomial.

In any case, we have verified directly that

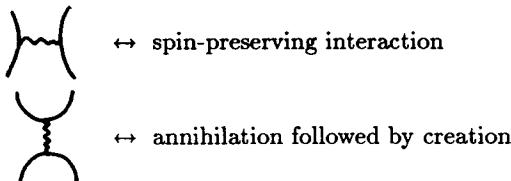
$$\begin{aligned}
 R_{cd}^{ab} &= A\delta_c^a\delta_d^b + A^{-1}\delta^{ab}\delta_{cd} \\
 n &= -A^2 - A^{-2}
 \end{aligned}$$

is a solution to the Yang-Baxter Equation.

In the course of this derivation I have used the notation

$$R = A \text{ } \begin{array}{c} \diagup \\ \diagdown \end{array} + B \text{ } \begin{array}{c} \diagdown \\ \diagup \end{array}$$

to keep track of the composition of interactions. It is nice to think of these as **abstract Feynman diagrams**, where, with a vertical arrow of time [\uparrow], we have



The next section follows this theme into a different model for the bracket and other solutions to the Yang-Baxter Equation.

9⁰. Formal Feynman Diagrams, Bracket as a Vacuum-Vacuum Expectation and the Quantum Group $SL(2)q$.

The solutions to the Yang-Baxter Equation that we constructed in the last section are directly related to the bracket, but each solution gives the bracket at special values corresponding to solutions of $n + A^2 + A^{-2} = 0$ for a given positive integer n . We end up with Yang-Baxter state models for infinitely many specializations of the bracket.

In fact there is a way to construct a solution to the Yang-Baxter Equation and a corresponding state model for the bracket - giving the whole bracket polynomial in one model. In order to do this it is conceptually very pleasant to go back to the picture of creations, annihilations and interactions - taking it a bit more seriously.

If we take this picture seriously, then a **fragment** such as a maximum or a minimum, need no longer be regarded as a Kronecker delta. The fragment

$$\uparrow \quad \text{cup}^b \leftrightarrow M^{ab},$$

with the time's arrow as indicated, connotes a **creation** of spins a and b from the vacuum.

In any case, we now allow matrices M_{ab} and M^{ab} corresponding to **caps** and **cups** respectively. From the viewpoint of time's arrow running up the page, cups are creations and caps are annihilations. The matrix values M_{ab} and M^{ab} represent (abstract) amplitudes for these processes to take place.

Along with cups and caps, we have the R -matrices:

$$\begin{array}{c} a \diagup \quad b \\ \diagdown \quad \diagup \\ c \quad d \end{array} \leftrightarrow R_{cd}^{ab}$$

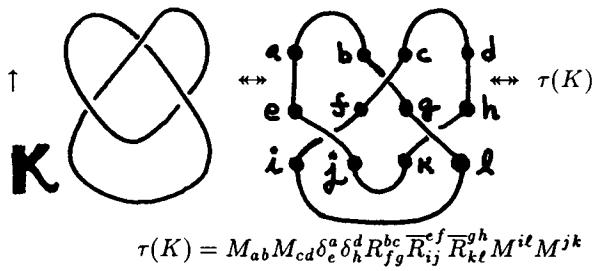
$$\begin{array}{c} a \quad b \\ \diagup \quad \diagdown \\ c \quad d \end{array} \leftrightarrow \bar{R}_{cd}^{ab}$$

corresponding to our knot-theoretic interactions. Note that the crossings corresponding to R and \bar{R} are now differentiated relative to time's arrow. Thus for R the over-crossing line goes from right to left as we go up the page.

A given link-diagram (**unoriented**) may be represented with respect to time's arrow so that it is naturally decomposed (via time as the height function) into cups, caps and interactions. There may also be a few residual Kronecker deltas (curves with no critical points vis-a-vis this height function):



Example.



In this example we have translated a trefoil diagram K into its corresponding expression $\tau(K)$ in the language of annihilation, creation and interaction. If the tensors are numerically valued (or valued in a commutative ring), then $\tau(K)$ represents a **vacuum-vacuum expectation** for the processes indicated by the diagram and this arrow of time.

The expectation is in accord with the principles of quantum mechanics. In quantum mechanics (see [FE]) the **probability amplitude for the concatenation of processes is obtained by summing the products of the amplitudes of the intermediate configurations in the process over all possible internal configurations**. Thus if we have a process $\xrightarrow[a]{P}\xrightarrow[b]{} b$ with “input” a and “output” b , and another process $\xrightarrow[c]{Q}\xrightarrow[d]{} d$, then the amplitude for $\xrightarrow[a]{P}\xrightarrow[i]{Q}\xrightarrow[b]{} b$ (given input a and output b) is the sum

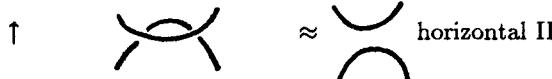
$$\sum_i P_{ai}Q_{ib} = (PQ)_{ab}$$

and corresponds exactly to matrix multiplication. For a vacuum-vacuum expectation there are no inputs or outputs. The expression $\tau(K)$ (using Einstein summation

tion convention) represents the sum over all internal configurations (spins on the lines) of the products of amplitudes for creation, annihilation and interaction.

Thus $\tau(K)$ can be considered as the basic form of vacuum-vacuum expectation in a highly simplified quantum field theory of link diagrams. We would like this to be a topological quantum field theory in the sense that $\tau(K)$ should be an invariant of regular isotopy of K . Let's look at what is required for this to happen.

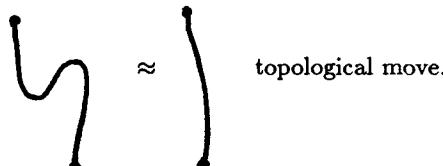
Think of the arrow of time as a specified vertical bottom-to-top direction on the page. Regular isotopies of the link diagram are generated by the Reidemeister moves of type II and type III. A type II move can occur at various angles with respect to the time-arrow. Two extremes are illustrated below:

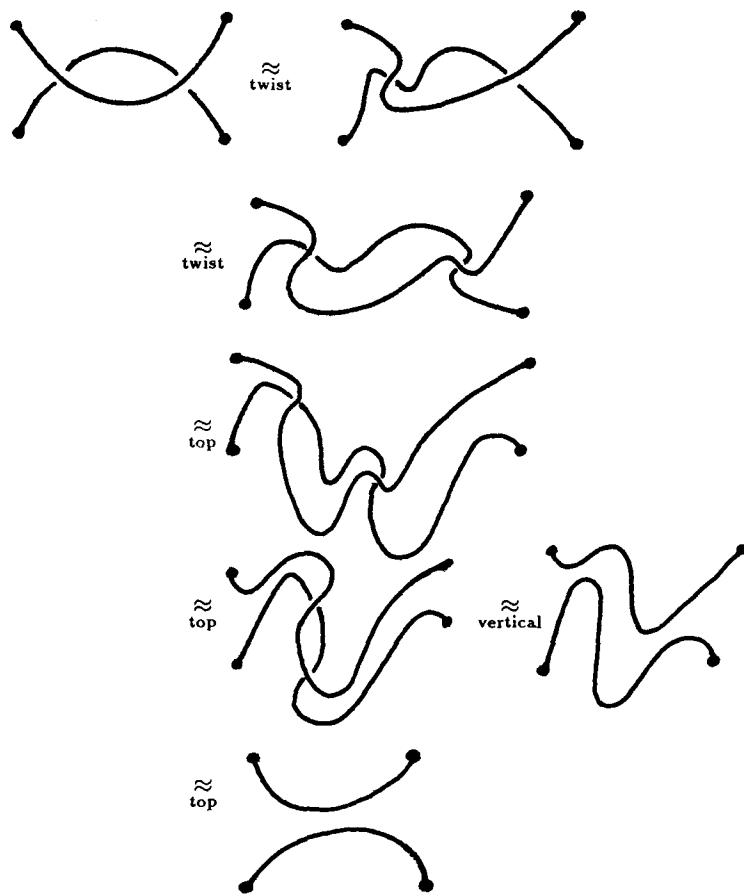


Similarly a type III move can occur at various angles, and we can twist a crossing keeping its endpoints fixed:



It is interesting to note that vertical II plus the twist generates horizontal II. See Figure 15. This generation utilizes the basic topological move of canceling pairs of critical points (maxima and minima) as shown below:



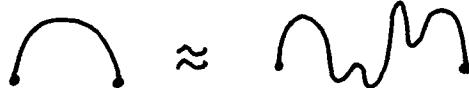


twist + vertical + top \Rightarrow horizontal

Figure 15

Note that in the deformation of Figure 15 we have kept the angles of the lines at their endpoints fixed. This allows a proper count of maxima and minima generated by the moves, and it means that these deformations can fit smoothly into larger diagrams of which the patterns depicted are a part. A simplest instance of this

angle convention is the maxima (annihilation):



If we were allowed to move the end-point angles, then a maximum could turn into a minimum: \approx



The same remarks apply to the horizontal version of the type III move, and we find that two link diagrams arranged transversal to a given time direction (height function) are regularly isotopic if and only if one can be obtained from the other by a sequence of moves of the types:

- a) topological move (canceling maxima and minima)
- b) twist
- c) vertical type II move
- d) vertical type III move with all crossings of the same type relative to time's arrow.

(Transversal means that a given level (of constant time) intersects the diagram in isolated points that are either transversal intersections, maxima or minima.)

We can immediately translate the conditions of these relativized Reidemeister moves to a set of conditions on the abstract tensors for creation, annihilation and interaction that will guarantee that $\tau(K)$ is an invariant of regular isotopy. Let's take the moves one at a time:

- a) topological move

$$\begin{array}{ccc} \text{Diagram showing two link components } a \text{ and } b \text{ interacting via a crossing.} & \approx & \Leftrightarrow M_{bi}M^{ia} = \delta_b^a \\ \text{Diagram showing the same components } a \text{ and } b \text{ with the crossing resolved.} & \approx & \Leftrightarrow M^{ai}M_{ib} = \delta_a^b \end{array}$$

The matrices M_{ab} and M^{ab} are inverses of each other.

b) the twist

$$\bar{R}_{cd}^{ab} = M_{ce} R_{gh}^{ef} M^{hb} \delta_f^a \delta_d^g$$

c) vertical type II

$$\bar{R}_{ij}^{ab} R_{cd}^{ij} = \delta_c^a \delta_d^b$$

$$\bar{R}\bar{R} = I = R\bar{R}$$

d) vertical type III

This is the Yang-Baxter Equation for R and for \bar{R} .

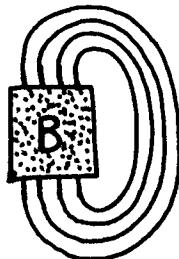
Thus we see (compare with 8.1):

Theorem 9.1'. If the interaction matrix R and its inverse \bar{R} satisfy the Yang-Baxter relation plus the interrelation with M_{ab} and M^{ab} specified by the twist and if M_{ab} and M^{ab} are inverse matrices then $\tau(K)$ will be an invariant of regular isotopy.

Braids.

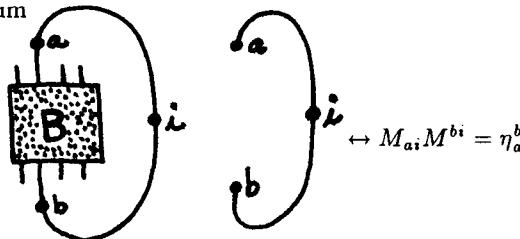
The vacuum-vacuum-expectation viewpoint reflects very beautifully on the

case of braids. For consider $\tau(K)$ when K is a closed braid:



$$K = \overline{B}.$$

If $K = \overline{B}$ where B is a braid, then each closure strand has one maximum and one minimum



The braid itself consists entirely of interactions, with no creations or annihilations. As we sweep time up from the vacuum state the minima from the closure strands give pair creations. The left hand member of each pair participates in the braid while its right hand twin goes up the trivial braid of the closure. At the top everybody pairs up and all cancel. Each braid-strand contributes a matrix of the form

$$\eta_a^b = M_{ai}M^{bi}$$

as illustrated above. If $\rho(B)$ denotes the interaction tensor (composed of R -matrices) coming from the braid, then we find that

$$\tau(K) = \text{Trace}(\eta^{\otimes n} \rho(B))$$

where there are n braid strands, and $\rho(B)$ is regarded as living in a tensor product of matrices as in section 7⁰. Note that this version of $\tau(K)$ has the same form as the one we discussed vis-a-vis the Markov Theorem (Markov trace).

Obtaining the Bracket and its R -matrix.

If we wish $\tau(K)$ to satisfy the bracket equation

$$\tau\left(\begin{array}{c} \diagup \\ \diagdown \end{array}\right) = A\tau\left(\begin{array}{c} \diagup \\ \diagup \end{array}\right) + A^{-1}\tau\left(\begin{array}{c} \diagdown \\ \diagdown \end{array}\right)$$

then a wise choice for the R -matrix is

$$\begin{array}{c} \diagup \\ \diagdown \end{array} = R_{cd}^{ab} = A \begin{array}{c} \diagup \\ \diagup \end{array} + A^{-1} \begin{array}{c} \diagdown \\ \diagdown \end{array}.$$

Hence

$$R_{cd}^{ab} = AM^{ab}M_{cd} + A^{-1}\delta_c^a\delta_d^b.$$

Suppose, for the moment, that the M -matrices give the correct loop value. That is, suppose that

$$\begin{array}{c} \bullet \\ \circ \end{array} b = \sum_{a,b} M_{ab}M^{ab} = d = -A^2 - A^{-2}.$$

Letting $U = M^{ab}M_{cd} = \begin{array}{c} \diagup \\ \diagup \end{array}$, we have $U^2 = dU$ and $R = AU + A^{-1}I$. The proof that R (with the given loop value) satisfies the YBE is then identical to the proof given for the corresponding braiding relation in Proposition 7.5.

Thus, all we need, to have a model for the bracket and a solution to the Yang-Baxter Equation is a matrix pair (M_{ab}, M^{ab}) of inverse matrices whose loop value (d above) is $-A^2 - A^{-2}$. Let us assume that $M_{ab} = M^{ab}$ so that we are looking for a matrix M with $M^2 = I$.

$$\begin{array}{c} \diagup \\ \diagup \end{array} \approx \begin{array}{c} \diagup \\ \diagup \end{array} \Leftrightarrow M^2 = I$$

Then the loop value, d , is the sum of the squares of the entries of M :

$$d = \sum_{a,b} (M_{ab})^2.$$

If we let $M = \begin{bmatrix} 0 & \sqrt{-1} \\ -\sqrt{-1} & 0 \end{bmatrix}$, then $M^2 = I$ and $d = (-1) + (-1) = -2$. This is certainly a special case of the bracket, with $A = 1$ or $A = -1$. Furthermore, it is very easy to deform this M to obtain

$$M = \begin{bmatrix} 0 & \sqrt{-1}A \\ -\sqrt{-1}A^{-1} & 0 \end{bmatrix}$$

with $M^2 = I$ and $d = -A^2 - A^{-2}$, as desired.

This choice of creation/annihilation matrix M gives us a tensor model for $\langle K \rangle$, and a solution to the Yang-Baxter Equation.

In direct matrix language, we find

$$M = \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 0 & \sqrt{-1}A \\ \hline 2 & -\sqrt{-1}A^{-1} \\ \hline \end{array}$$

$$U = M \otimes M = \begin{array}{|c|c|c|c|} \hline 11 & 12 & 21 & 22 \\ \hline 0 & 0 & 0 & 0 \\ \hline 12 & 0 & -A^2 & 1 \\ \hline 21 & 0 & 1 & -A^{-2} \\ \hline 22 & 0 & 0 & 0 \\ \hline \end{array}$$

and $R = AM \otimes M + A^{-1}I \otimes I$.

$$R = \begin{array}{|c|c|c|c|} \hline A^{-1} & 0 & 0 & 0 \\ \hline 0 & -A^3 + A^{-1} & A & 0 \\ \hline 0 & A & 0 & 0 \\ \hline 0 & 0 & 0 & A^{-1} \\ \hline \end{array}$$

There is much to say about this solution to the Yang-Baxter Equation. In order to begin this discussion, we first look at the relationship of the group $SL(2)$ and the cases $A = \pm 1$.

Spin-Networks, Binors and $SL(2)$.

The special case of the bracket with $M = \begin{bmatrix} 0 & \sqrt{-1} \\ -\sqrt{-1} & 0 \end{bmatrix}$ is of particular significance. Let's write this matrix as

$$\begin{bmatrix} 0 & \sqrt{-1} \\ -\sqrt{-1} & 0 \end{bmatrix} = \sqrt{-1} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \sqrt{-1}\epsilon$$

where ϵ_{ab} denotes the alternating symbol:

$$\epsilon_{ab} = \begin{cases} 1 & \text{if } a < b \\ -1 & \text{if } a > b \\ 0 & \text{if } a = b \end{cases}$$

$$a, b \in \{1, 2\} = \mathcal{I}.$$

Let us also use the following diagrammatic for this matrix:

$$\epsilon_{ab} = \prod_{\substack{a \\ b}} \quad \quad \epsilon^{ab} = \prod_{\substack{a \\ b}}$$

The alternating symbol is fundamental to a number of contexts.

Lemma 9.2. Let $P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a matrix of commuting (associative) scalars. Then

$$P\epsilon P^T = \text{DET}(P)\epsilon$$

where P^T denotes the transpose of P .

Proof.

$$\begin{aligned} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} b & d \\ -a & -c \end{bmatrix} \\ &= \begin{bmatrix} ab - ba & ad - bc \\ cb - da & cd - dc \end{bmatrix} \\ &= \begin{bmatrix} 0 & ad - bc \\ -ad + bc & 0 \end{bmatrix} \\ &= (ad - bc) \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \end{aligned}$$

//

Thus we have the definition.

Definition 9.3. Let \mathcal{R} be a commutative, associative ring with unit. Then $SL(2, \mathcal{R})$ is defined to be the set of 2×2 matrices P with entries in \mathcal{R} such that

$$P\epsilon P^T = \epsilon.$$

In the following, I shall write $SL(2)$ for $SL(2, \mathcal{R})$. $SL(2)$ is identified as the set of matrices P leaving the ϵ -symbol invariant.

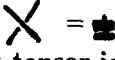
Note also how this invariance appears diagrammatically: Let $P = (P_b^a) = (P_{ab})$

$$(P\epsilon P^T) = P_i^a \epsilon^{ij} P_j^b, \quad P_b^a$$



Thus the invariances $P\epsilon P^T = \epsilon$ and $P^T\epsilon P = \epsilon$ correspond to the diagrammatic equations

In this sense, the link diagrams at $A = \pm 1$ correspond to $SL(2)$ -invariant tensors.

Also, the identity  =     for the R -matrix becomes a specific $SL(2)$ -invariant tensor identity:

$$\epsilon^{ab} \epsilon_{cd} = \delta_c^a \delta_d^b - \delta_d^a \delta_c^b$$

$$\Leftrightarrow -(\sqrt{-1} \Pi)(\sqrt{-1} \Pi) =)(-X$$

$$\Leftrightarrow - \bigcup =)(-X$$

$$\Leftrightarrow X = \bigcup +) ([A=+1].$$

In this calculus of $SL(2)$ -invariant tensor diagrams there is no distinction between an over-crossing and an under-crossing. The calculus is a topologically invariant calculation for curves immersed in the plane. All Jordan curves receive the same value of -2 . This form of a calculus for $SL(2)$ -invariant tensors is the **binor calculus** of Roger Penrose [PEN1]. His binors are precisely the case for $A = -1$ so that $\text{X} + \text{X} +) (= 0$. Here each crossing receives a minus sign and we have the invariance $\text{X} = \text{O}$ as well.

The Penrose binors are a special case of the bracket, and hence a special case of the Jones polynomial. The binors are the underpinning for the Penrose theory of spin-networks.

For now, it is worth briefly re-tracing steps that led Penrose to discover the binors. Penrose began by considering **spinors** ψ^A . Now a spinor is actually a 2-vector over the complex numbers \mathbf{C} . Thus $A \in \{1, 2\}$. $SL(2, \mathbf{C})$ acts (via matrix multiplication) on these spinors, and one wants an inner product $\psi\psi^* \in \mathbf{R}$ (real numbers) that is $SL(2, \mathbf{C})$ invariant. Let $\psi_A^* = \epsilon_{AB}\psi^B$ so that

$$\psi\psi^* = \psi^A\psi_A^* = \psi^A\epsilon_{AB}\psi^B.$$

If $(U_B^A) = U \in SL(2, \mathbf{C})$ then $\psi\psi^*$ is invariant under the action of U :

$$\begin{aligned} (U\psi)(U\psi)^* &= (U_I^A\psi^I)\epsilon_{AB}(U_J^B\psi^J) \\ &= (U_I^A U_J^B \epsilon_{AB})\psi^I\psi^J \\ &= \epsilon_{IJ}\psi^I\psi^J \\ &= \psi\psi^*. \end{aligned}$$

A calculus of diagrams that represents this inner product is suggested by the “natural” method of lowering an index:

$$\psi^A \leftrightarrow \boxed{\bullet}, \quad \psi_A^* \leftrightarrow \boxed{\circ}$$

Of course,

$$\psi\psi^* \leftrightarrow \boxed{\bullet} \curvearrowright \boxed{\circ}$$

and one is led to take

$$\cap \leftrightarrow \epsilon_{ab}$$

$$\cup \leftrightarrow \epsilon^{ab}.$$

However, the resulting diagrams are not topologically invariant:

$$\text{Diagram: } \begin{array}{c} \text{a} \\ \curvearrowleft \\ \text{b} \end{array} \leftrightarrow \epsilon_{ai} \epsilon^{ib} = -\delta_a^b \leftrightarrow - \begin{array}{c} \text{b} \\ \curvearrowright \\ \text{a} \end{array},$$

and

$$\text{Diagram: } \begin{array}{c} \text{i} \\ \curvearrowleft \\ \text{i} \\ \curvearrowright \\ \text{a} \quad \text{b} \end{array} \leftrightarrow \delta_a^i \delta_b^j \epsilon_{ji} = \epsilon_{ba} = -\epsilon_{ab}.$$

Penrose solved this difficulty by associating a minus sign to each minimum and each crossing. This changes the loop from +2 to -2 and gives the binor identity

$$\times + \text{Diagram} +) (= 0.$$

That the resulting calculus is indeed topologically invariant, we know well from our work with the bracket. Note that

$$\cup \leftrightarrow \sqrt{-1} \text{Diagram}$$

$$\cap \leftrightarrow \sqrt{-1} \text{Diagram}$$

works just as well as the Penrose convention - distributing the sign over maxima and minima.

It is amusing to do our usual topological verifications in the binor context:

$$\text{Diagram} = - \text{Diagram} - \text{Diagram}$$

$$= - \text{Diagram} - \times$$

$$= - \text{Diagram} + \text{Diagram} + \text{Diagram}$$

$$= \text{Diagram}.$$

There is much more to the theory of the binors and the spin networks. (See [PEN1], [PEN3], and sections 12⁰ and 13⁰ of Part II.)

The *R*-matrix.

We can now view the bracket calculation as stemming from the particular deformation $\tilde{\epsilon} = \begin{bmatrix} 0 & A \\ -A^{-1} & 0 \end{bmatrix}$ of the spinor epsilon (alternating symbol). We use the shorthand

$$M_{ab} = \sqrt{-1} \tilde{\epsilon}_{ab} = \sqrt{-1} \prod_a^{\text{II}} \prod_b^{\text{I}}$$

Thus

\cap	$\leftrightarrow \sqrt{-1} \prod^{\text{II}} A^{\text{II}}$
\cup	$\leftrightarrow \sqrt{-1} \prod^{\text{I}} A^{\text{I}}$

are the creation and annihilation matrices for the bracket. This symbolism, plus the Fierz identity

$$\frac{\prod}{\prod} =)(-X$$

make it easy for us to re-express the *R*-matrix:

$$R = A \cap + A^{-1})()$$

$$= A(\sqrt{-1})^2 \frac{\prod}{\prod} A^{\text{II+II}} + A^{-1})()$$

$$= - \frac{\prod}{\prod} A^{\text{II+II+1}} + A^{-1})()$$

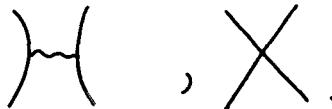
$$= [X -)() A^{\text{II+II+1}} + [)() A^{-1}$$

$$= (A^{-1} - A^{\text{II+II+1}})[)() + A^{\text{II+II+1}}[X].$$

From the viewpoint of the particle interactions, the Fierz identity lets us replace the annihilation \rightsquigarrow creation



by a combination of exchange and crossover



This actually leads to a different state model for the bracket - one that generalizes to give a series of models for infinitely many specializations of the Homfly polynomial! (Compare [JO4].)

In order to see this let's calculate R a bit further:

$$R = (A^{-1} - A^{\Pi+\Pi+1})[\text{)(}] + A^{\Pi+\Pi+1}[\text{ X }].$$

The index set is $\mathcal{I} = \{1, 2\}$ with ${}_{12}^\Pi = +1$, ${}_{21}^\Pi = -1$, ${}_{11}^\Pi = {}_{22}^\Pi = 0$. The lines and cross-lines in the abstract tensors for the diagrams above are Kronecker deltas. Thus for $[\text{)}(]_{12}$ the coefficient is

$$A^{-1} - A^{\frac{12}{\Pi+\Pi+1}} = A^{-1} - A^{2+1} = A^{-1} - A^3.$$

We find:

$$R = (A^{-1} - A^3)[\text{)}(] + A^{-1}[\text{)=}(] + A[\text{ X }].$$

Thus

$$R = A^{+1} \{(A^{-2} - A^2)[\text{)}(] + A^{-2}[\text{)=}(] + [\text{ X }]\}.$$

From this, we can abstract an R -matrix

$$R_q = (q - q^{-1})[\text{)}>[\text{ }]+q[\text{)}=[\text{ }]+[\text{ X }]$$

defined for an arbitrary ordered index set \mathcal{I} . By the same token, let

$$\bar{R}_q = (q^{-1} - q)[\text{)}<[\text{ }]+q^{-1}[\text{)}=[\text{ }]+[\text{ X }].$$

Note that

$$R_q - \bar{R}_q = (q - q^{-1})\{[\downarrow] > [\uparrow] + [\downarrow] < [\uparrow] + [\downarrow] = [\uparrow]\}$$

$$\therefore R_q - \bar{R}_q = (q - q^{-1})[\downarrow \uparrow].$$

If $R_q = [\cancel{\downarrow \uparrow}]$, $\bar{R}_q = [\cancel{\uparrow \downarrow}]$ then this is the analog of the exchange identity for the regular isotopy version of the Homfly polynomial.

As a start in this direction, we observe that

Proposition 9.4. R_q and \bar{R}_q are inverses. Each is a solution to the Yang-Baxter Equation for any ordered index set \mathcal{I} .

Remark. We have already verified 9.4 for $\mathcal{I} = \{1, 2\}$. The present form of R_q makes the more general verification easy. From the point of view of particle interactions, these solutions are quite remarkable in that they have a left-right asymmetry - R_q weights only for $\downarrow < \uparrow$ while \bar{R}_q weights for $\downarrow > \uparrow$. In the case of $\mathcal{I} = \{1, 2\}$, this **asymmetry** is not at first apparent from the expansion formula for R but we see that it arises from the specific choice of time's arrow that converts one abstract tensor to an exchange, and the other to a creation/annihilation. The loop-adjustment $d = -A^2 - A^{-2}$ assures us that the calculations will be independent of the choice of time/space split. One asymmetry (left/right) compensates for another (time/space).

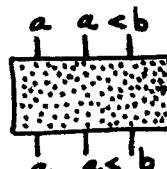
Where is the physics in all of this? The mathematics is richly motivated by physical ideas, and this part of the story will continue to spin its web. It is also possible that the concept of a topological interaction pattern, such as the knot/link diagram, may contain a clue for modeling observed processes. The link diagram (or the link in three-space) could as a **topological whole** represent a particle or process. This speculation moves in the direction of **embedded** (knotted and linked) **strings**. We have the **choice** to interpret a link diagram as an abstract network with expectation values computable from a mixture of quantum mechanical and topological notions **or** as an embedding in three-space. The three-space auxiliary to a link diagram need not be regarded as the common space of physical observation. The interpretive problem is to understand when these contexts (mathematical versus observational) of three dimensions come together.

Proof of Proposition 9.4. It is easy to verify directly that R_q and \bar{R}_q are inverses. Let's write

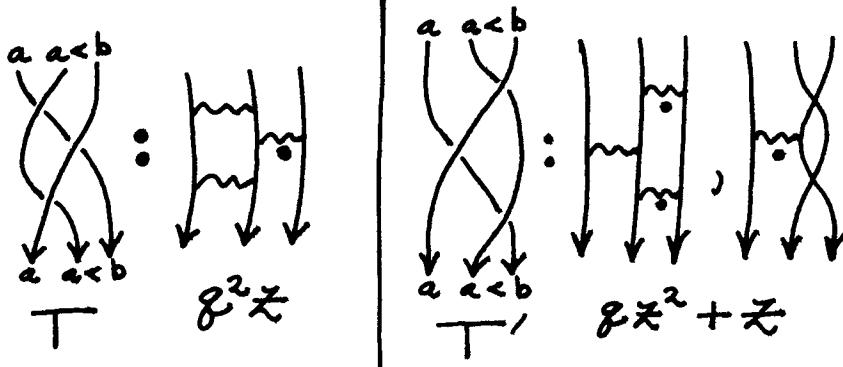
$$R = z \begin{pmatrix} & < \\ & + \\ & q \end{pmatrix} = \begin{pmatrix} & + \\ & \times \end{pmatrix}$$

and examine the conditions under which R yields a solution to YBE.

Case I.



Assume that the input and output spins are as indicated in the diagram above. Then we have:



Here I have indicated the possible interaction patterns for the two sides of the Yang-Baxter ledger. I have also adopted the following notation:

$$\begin{pmatrix} & < \\ & + \\ & q \end{pmatrix} \leftrightarrow \begin{pmatrix} & = \\ & \downarrow \end{pmatrix} \leftrightarrow \text{equal spin labels}$$

$$\begin{pmatrix} & < \\ & \circ \\ & \downarrow \end{pmatrix} \leftrightarrow \begin{pmatrix} & < \\ & \downarrow \end{pmatrix} \leftrightarrow \text{left label less than right label}$$

$$\begin{pmatrix} & > \\ & \circ \\ & \downarrow \end{pmatrix} \leftrightarrow \begin{pmatrix} & > \\ & \downarrow \end{pmatrix} \leftrightarrow \text{left label greater than right label}$$

$$\begin{pmatrix} & \times \\ & \downarrow \end{pmatrix} \leftrightarrow \begin{pmatrix} & \times \\ & \downarrow \end{pmatrix} \leftrightarrow \text{crossover with unequal labels}$$

Returning to Case I, we see that for T (the first triangle condition) there is only one admissible interaction consisting in three exchanges (no crossover) and a product of vertex weights of $q^2 z$. For T' we have two possibilities - exchanges, or a pair of canceling crossovers. Here the sum of products of vertex weights is $qz^2 + z$.

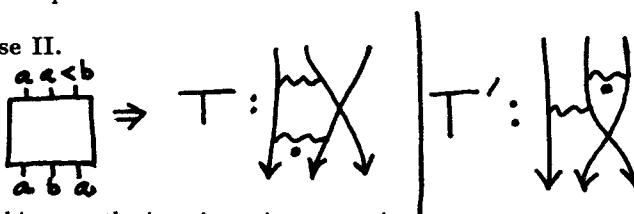
In order for the YBE to hold, it is necessary that the contributions from T and T' match. Therefore, we need $q^2 z = qz^2 + z$. Hence

$$\begin{aligned} q^2 &= qz + 1 \\ z &= q - q^{-1}. \end{aligned}$$

This completes Case I.

The other cases are all similar. The only demand on R is $z = q - q^{-1}$. Here is a sample case. We leave the other cases for the reader.

Case II.



In this case, the invariance is automatic.

Modulo the remaining cases, this completes the proof of Proposition 9.4. //

Discussion. We have seen already that for the index set $\mathcal{I} = \{-1, +1\}$ this solution to the YBE is closely related to the group $SL(2)$.

The general solution shown here for any ordered index set is related to the group $SL(n)$ and these solutions can be used to construct knot polynomial models. Before creating these models, I wish to return to our discussion of $SL(2)$ and show how an extension of its symmetry leads to the idea of a “quantum group.”

Bracket, Spin Nets and the Quantum Group for $SL(2)$.

In the discussion surrounding Lemma 9.2 we have seen solutions to the Yang-Baxter Equation and models for the Jones polynomial (a. k. a. bracket) emerge from the deformed epsilon, $\tilde{\epsilon} = \begin{bmatrix} 0 & A \\ -A^{-1} & 0 \end{bmatrix}$. Since $\epsilon = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is the fundamental defining invariant for $SL(2)$, it is natural to ask the question: **For what algebraic structure is $\tilde{\epsilon}$ the basic invariant?**

Therefore suppose that a 2×2 matrix $P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is given and that the entries of P belong to an associative but not necessarily commutative algebra. Let us demand the invariances:

$$\left\{ \begin{array}{l} P\tilde{\epsilon}P^T = \tilde{\epsilon} \\ P^T\tilde{\epsilon}P = \tilde{\epsilon} \end{array} \right\} \quad (*)$$

where it is assumed that A commutes with the entries of P . A bit of calculation will reveal the relations demanded by (*).

Proposition 9.5. With $\tilde{\epsilon} = \begin{pmatrix} 0 & A \\ -A^{-1} & 0 \end{pmatrix}$, $P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the equations

$$\left. \begin{array}{rcl} 1) & P\tilde{\epsilon}P^T &= \tilde{\epsilon} \\ 1') & P^T\tilde{\epsilon}P &= \tilde{\epsilon} \end{array} \right\} \quad (*)$$

are equivalent to the set of relations (**) shown boxed below: (with $\sqrt{q} = A$)

$ba = qab$	$dc = qcd$
$ca = qac$	$db = qbd$
$bc = cb$	
$ad - da = (q^{-1} - q)bc$	
$ad - q^{-1}bc = 1$	

(**)

Proof. $P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\tilde{\epsilon} = \begin{pmatrix} 0 & A \\ -A^{-1} & 0 \end{pmatrix}$.

$$P\tilde{\epsilon}P^T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & A \\ -A^{-1} & 0 \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

$$= \begin{pmatrix} -A^{-1}b & Aa \\ -A^{-1}d & Ac \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

$$P\tilde{\epsilon}P^T = \begin{pmatrix} -A^{-1}ba + Aab & -A^{-1}bc + Aad \\ -A^{-1}da + Acb & -A^{-1}dc + Acd \end{pmatrix}$$

$$P^T\tilde{\epsilon}P = \begin{pmatrix} -A^{-1}ca + Aac & -A^{-1}cb + Aad \\ -A^{-1}da + Abc & -A^{-1}db + Abd \end{pmatrix}$$

Therefore, the relations that follow directly from $P\tilde{\epsilon}P^T = \tilde{\epsilon}$ and $P^T\tilde{\epsilon}P = \tilde{\epsilon}$ are as follows:

$$\begin{array}{ll} A^{-1}ba = Aab & A^{-1}dc = Acd \\ A^{-1}ca = Aac & A^{-1}db = Abd \\ -A^{-1}bc + Aad = A & -A^{-1}cb + Aad = A \\ -A^{-1}da + Acb = -A^{-1} & -A^{-1}da + Abc = A^{-1}. \end{array}$$

It follows that $bc = cb$, and that an equivalent set of relations is given by:

$$\left\{ \begin{array}{ll} ba = qab & dc = qcd \\ ca = qac & db = qbd \\ bc = cb & \\ ad - da = (q^{-1} - q)bc & \\ ad - q^{-1}bc = 1 & \end{array} \right.$$

where $q = A^2$. Note that the last relation, $ad - q^{-1}bc = 1$, becomes the condition $ad - bc = \text{DET}(P) = 1$ when $q = 1$. Also, when $q = 1$ these relations tell us that the elements of the matrix P commute among themselves.

This completes the proof of Proposition 9.5. //

Remark. The non-commutativity for elements of P is essential. The only non-trivial case where these elements all commute is when $q = 1$. But one, perhaps unfortunate, consequence of non-commutativity is that if P and Q are matrices satisfying (**), then PQ does not necessarily satisfy (**) (Since $(PQ)^T \neq Q^TP^T$ in the non-commutative case.). Thus Proposition 9.5 does **not** give rise to a natural generalization of the group $SL(2)$ to a group $SL(2)_q$ leaving $\tilde{\epsilon}$ invariant. Instead, we are left with the universal associative algebra defined by the relations (**). **This algebra will be denoted by $SL(2)_q$.** It is called “the quantum group $SL(2)_q$ ” in the literature [DRIN1].

While $SL(2)_q$ is not itself a group, it does have a very interesting structure that is directly related to the Lie algebra of $SL(2)$. Furthermore, there is a natural comultiplication $\Delta : SL(2)_q \rightarrow SL(2)_q \otimes SL(2)_q$ that is itself a map of algebras. I shall first discuss the comultiplication, and then the relation with the Lie algebra.

The algebraic structure of the quantum group is intimately tied with the properties of the R -matrix and hence with associated topological invariants. The rest of this section constitutes a first pass through this region.

Bialgebra Structure and Abstract Tensors.

Let's begin with a generalization of our construction for $SL(2)_q$. Suppose that we are given a matrix $E = (E^{ij})$ (the analog of $\tilde{\epsilon}$) and a matrix $P = (P_j^i)$. The elements of E commute with the elements of P . The indices i, j are assumed to belong to a specified finite index set \mathcal{I} . Then the equation $PEP^T = E$ reads in indices as:

$$P_i^a E^{ij} P_j^b = E^{ab}$$

with summation on the repeated indices. In terms of abstract tensors we can write

$$\text{loop} \leftrightarrow P_j^i, \boxed{\text{dot}} \leftrightarrow E^{ij}$$

and

$$PEP^T = \leftrightarrow \quad \text{loop} \boxed{\text{dot}} \text{loop}$$

This last equation should be considered carefully since the order of terms in the

relations matters. We might write colloquially $\boxed{\text{dot}} = \boxed{\text{dot}}$, preserving the order of the products of elements of P . This latter form makes sense if the elements of E commute with the elements of P . Then we can say

The transposed relationship

$$P^T EP = E$$

is best diagrammed by using E with lowered indices: $E_{ij} = E^{ij}$. Then

$$P^T EP = E \leftrightarrow P_a^i E_{ij} P_b^j = E_{ab}$$

$$\leftrightarrow \boxed{\text{dot dot}} = \boxed{\text{dot}} \boxed{\text{dot}}$$

Now, given the relations $PEP^T = E$ and $P^T EP = E$, let $\mathcal{A}(E)$ denote the resulting universal algebra. Thus $\mathcal{A}(\bar{\epsilon}) = SL(2)_q$ ($\sqrt{q} = A$). Define $\Delta : \mathcal{A}(E) \rightarrow \mathcal{A}(E) \otimes \mathcal{A}(E)$ by the formula $\Delta(P_j^i) = \sum_{k \in \mathcal{I}} P_k^i \otimes P_j^k$ on the generators - extending linearly over the ground ring [The ground ring is $\mathbb{C}[A, A^{-1}]$ in the case of $SL(2)_q$].

Proposition 9.6. $\Delta : \mathcal{A}(E) \rightarrow \mathcal{A}(E) \otimes \mathcal{A}(E)$ is a map of algebras.

Proof. We must prove that $\Delta(x)\Delta(y) = \Delta(xy)$ for elements $x, y \in \mathcal{A}(E)$. It will suffice to check this on the relations $PEP^T = E$ and $P^T EP = E$. We check the first relation; the second follows by symmetry. Now $PEP^T = E$ corresponds to $P_i^a E^{ij} P_j^b = E^{ab}$. Thus we must show that $\Delta(P_i^a)E^{ij}\Delta(P_j^b) = E^{ab}\iota$ where $\iota = 1 \otimes 1$. (Note that the elements E^{ij} are scalars - commuting with everyone - and that $\Delta(1) = \iota$ by definition, since ι is the identity element in $\mathcal{A}(E) \otimes \mathcal{A}(E)$.) Since ι is the identity element in $\mathcal{A}(E) \otimes \mathcal{A}(E)$ we can regard E as a matrix on any tensor power of $\mathcal{A}(E)$ via $E^{ab} \leftrightarrow E^{ab}(1 \otimes 1 \otimes \dots \otimes 1)$. With this identification, we write

$$\Delta(P_i^a)E^{ij}\Delta(P_j^b) = E^{ab}$$

as the desired relation. Computing, we find

$$\begin{aligned} \Delta(P_i^a)E^{ij}\Delta(P_j^b) &= (P_k^a \otimes P_i^k)E^{ij}(P_\ell^b \otimes P_j^\ell) \\ &= (P_k^a P_\ell^b) \otimes (P_i^k P_j^\ell)E^{ij} \\ &= (P_k^a P_\ell^b) \otimes (P_i^k E^{ij} P_j^\ell) \\ &= (P_k^a P_\ell^b) \otimes E^{k\ell} \\ &= (P_k^a E^{k\ell} P_\ell^b) \otimes 1 \\ &= E^{ab} \otimes 1 \\ &= E^{ab}(1 \otimes 1) \\ &= E^{ab} \quad (\text{sic.}) \end{aligned}$$

This completes the proof of Proposition 9.6. //

Remark. The upshot of Proposition 9.6 is that $\Delta(\mathcal{A}(E)) \subset \mathcal{A}(E) \otimes \mathcal{A}(E)$ is also an algebra leaving the form E invariant. Furthermore, the structure of this proof

is well illustrated via the abstract tensor diagrams. We have

$$P_b^a \leftrightarrow \text{circle}, \quad E^{ab} \leftrightarrow \text{rectangle}$$

$$\Delta(P_b^a) \leftrightarrow \Delta(\text{circle}) = \text{circle} \otimes \text{circle}.$$

So that

$$\Delta(\text{circle}) \otimes \Delta(\text{rectangle}) = \text{circle} \otimes \text{circle} \otimes \text{rectangle}$$

$$= \text{circle} \otimes \text{circle} \otimes \text{rectangle}$$

$$= \text{circle} \otimes \text{circle} \otimes \text{rectangle}$$

$$= \text{circle} \otimes \text{rectangle} \otimes 1$$

$$= \text{rectangle} \otimes 1$$

$$= \text{rectangle} (1 \otimes 1)$$

$$\equiv \text{rectangle}.$$

//

The tensor diagrams give a direct view of the repeated indices (via tied lines), laying bare the structure of this calculation. In particular, we see at once that if

$$\begin{array}{c} \text{Diagram 1: } \\ \text{Two boxes connected by a curved line with two loops.} \end{array} = \begin{array}{c} \text{Diagram 2: } \\ \text{A single box with two loops attached to its top.} \end{array} \quad \text{and} \quad \begin{array}{c} \text{Diagram 3: } \\ \text{Two boxes connected by a curved line with two loops.} \end{array} = \begin{array}{c} \text{Diagram 4: } \\ \text{A single box with two loops attached to its top.} \end{array},$$

then

$$\begin{array}{c} \text{Diagram 5: } \\ \text{Two boxes connected by a curved line with two loops.} \end{array} = \begin{array}{c} \text{Diagram 6: } \\ \text{A single box with two loops attached to its top.} \end{array}.$$

In other words, if $PEP^T = E$ and $QEQT = E$ and we define

$$(P * Q)_b^a = \sum_k P_k^a \otimes Q_b^k \in \mathcal{A}_P(E) \otimes \mathcal{A}_Q(E)$$

then

$$(P * Q)E(P * Q)^T = E$$

in this extended sense of $E \equiv E(1 \otimes 1)$ in the tensor product. The tensor product formalism provides the simplest structure in which we can partially reconstruct something like a group of matrices leaving the form E invariant. (Since $P * P^{-1}$ is distinct from the identity, we do not get a group.)

To return to our specific algebra $\mathcal{A} = \mathcal{A}(\tilde{\epsilon})$ we have: $P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

$$\left\{ \begin{array}{lcl} ba = qab & & dc = qcd \\ ca = qac & & db = qbd \\ bc = cb & & \\ ad - da = (q^{-1} - q)bc & & \\ ad - q^{-1}bc = 1 & & \end{array} \right\}$$

$$\left\{ \begin{array}{lcl} \Delta(a) & = & a \otimes a + b \otimes c \\ \Delta(b) & = & a \otimes b + b \otimes d \\ \Delta(c) & = & c \otimes a + d \otimes c \\ \Delta(d) & = & c \otimes b + d \otimes d \end{array} \right\}$$

So far, we have shown that $\Delta : \mathcal{A}(\tilde{\epsilon}) \rightarrow \mathcal{A}(\tilde{\epsilon}) \otimes \mathcal{A}(\tilde{\epsilon})$ is an algebra homomorphism. In fact, Δ is coassociative. This means that the following diagram commutes ($A = \mathcal{A}(\tilde{\epsilon})$):

$$\begin{array}{ccc} A \otimes A & \xrightarrow{1 \otimes \Delta} & A \otimes A \otimes A \\ \uparrow \Delta & & \uparrow \Delta \otimes 1 \\ A & \xrightarrow{\Delta} & A \otimes A \end{array}$$

The coassociativity property is obvious in this construction, since

$$\begin{aligned}
 (1 \otimes \Delta)\Delta(P_j^i) &= (1 \otimes \Delta)(P_k^i \otimes P_j^k) \\
 &= P_k^i \otimes P_\ell^k \otimes P_j^\ell \\
 &= (\Delta \otimes 1)(P_\ell^i \otimes P_j^\ell) \\
 &= (\Delta \otimes 1)\Delta(P_j^i).
 \end{aligned}$$

An associative algebra A with unit 1 and multiplication $m : A \otimes A \rightarrow A$ is a **bialgebra** if there is a homomorphism of algebras $\Delta : A \rightarrow A \otimes A$ that satisfies co-associativity, and so that A has a **co-unit**. A **co-unit** is an algebra homomorphism $\epsilon : A \rightarrow \mathbf{C}$ (\mathbf{C} denotes the ground ring) such that the following diagrams commute:

$$\begin{array}{ccc}
 A & \xrightarrow{\Delta} & A \otimes A \\
 1 \downarrow & \epsilon \otimes 1 \downarrow & 1 \downarrow \qquad \qquad 1 \otimes \epsilon \downarrow \\
 A & \xrightarrow{1} & A \qquad A & \xrightarrow{1} & A
 \end{array}$$

In the case of our construction with $\Delta(P_j^i) = P_k^i \otimes P_j^k$, let $\epsilon(P_j^i) = \delta_j^i$ (Kronecker delta).

Then

$$\begin{aligned}
 (\epsilon \otimes 1)\Delta(P_j^i) &= \epsilon(P_k^i) \otimes P_j^k \\
 &= \delta_k^i \otimes P_j^k \\
 &= 1 \otimes P_j^i \\
 &= P_j^i.
 \end{aligned}$$

Thus we have verified that $\mathcal{A}(E)$ is a bialgebra.

A bialgebra with certain extra structure (an antipode) is called a **Hopf algebra**.

The algebra $\mathcal{A}(\tilde{\epsilon})$ is an example of a Hopf algebra. In this case the antipode is a mapping $\gamma : \mathcal{A}(\tilde{\epsilon}) \rightarrow \mathcal{A}(\tilde{\epsilon})$ defined by $\gamma(a) = d$, $\gamma(d) = a$, $\gamma(b) = -qb$, $\gamma(c) = -q^{-1}c$. Extend γ linearly on sums and define it on products so that $\gamma(xy) = \gamma(y)\gamma(x)$ (an anti-homomorphism).

Note here that

$$\begin{aligned} P' = \gamma(P) &= \begin{pmatrix} \gamma(a) & \gamma(b) \\ \gamma(c) & \gamma(d) \end{pmatrix} \\ &= \begin{pmatrix} d & -qb \\ -q^{-1}c & a \end{pmatrix} \end{aligned}$$

is actually the **inverse matrix of P** in this non-commutative setting:

$$\begin{aligned} PP' &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -qb \\ -q^{-1}c & a \end{pmatrix} \\ &= \begin{pmatrix} ad - q^{-1}bc & -qab + ba \\ cd - q^{-1}dc & -qcb + da \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

It is also true that $P'P = I$. In general, the antipode $\gamma : A \rightarrow A$ is intended to be the analog of an inverse. Its abstract definition is that the following diagram should commute:

$$\begin{array}{ccccc} A & \xrightarrow{\Delta} & A \otimes A & \xrightarrow[\substack{1 \otimes \gamma \\ \gamma \otimes 1}]{} & A \otimes A \\ \downarrow \epsilon & & & & m \downarrow \\ \mathbf{C} & \xrightarrow{\eta} & & \eta & A \end{array}$$

Here ϵ is the co-unit, and η is the unit. In the case of $\mathcal{A}(E)$ we have

$$\epsilon(P_j^i) = \delta_j^i \text{ and } \eta(\delta_j^i) = \delta_j^i.$$

Thus the condition that $\mathcal{A}(E)$ be a Hopf algebra is that

$$\begin{aligned} \gamma(P_k^i)P_j^k &= \delta_j^i \\ P_k^i\gamma(P_j^k) &= \delta_j^i. \end{aligned}$$

This is the same as saying that P and $\gamma(P)$ are inverse matrices.

Remark. The algebra $\mathcal{A}(\tilde{\epsilon})$ is a deformation of the algebra of functions (co-ordinate functions) on the group $SL(2, \mathbb{C})$. In the classical case of an arbitrary group G , let $F(G)$ denote the collection of functions $f : G \rightarrow \mathbb{C}$. Define $m : F(G) \otimes F(G) \rightarrow F(G)$ by

$$m(f \otimes g)(x) = f(x)g(x)$$

and $\Delta : F(G) \rightarrow F(G) \otimes F(G)$ by

$$\Delta(f)(x \otimes y) = f(xy).$$

Let $\gamma : F(G) \rightarrow F(G)$ be defined by $\gamma(f)(x) = f(x^{-1})$, and $\epsilon : F(G) \rightarrow \mathbf{C}$, $\epsilon(f) = f(e)$, e = identity element in G . $\eta : \mathbf{C} \rightarrow F(G)$ by $\eta(z)f(g) = zf(g)$ for all $z \in \mathbf{C}$. This gives $F(G)$ the structure of a Hopf algebra.

Now $F(SL(2, \mathbf{C}))$ (or $F(G)$ for any Lie group G) contains (by taking derivatives) the functions on the Lie algebra of $SL(2, \mathbf{C})$. Thus we should expect that $F_q(SL(2)) \stackrel{\text{def.}}{=} SL(2)_q$ should be related to **functions on a deformation of the Lie algebra of $SL(2)$** . We can draw a connection between $\mathcal{A}(\tilde{\epsilon}) = F_q(SL(2))$ and the Lie algebra of $SL(2)$.

Deforming the Lie algebra of $SL(2, \mathbf{C})$.

The Lie algebra of $SL(2, \mathbf{C})$ consists in those matrices M such that

$$e^M \in SL(2, \mathbf{C}).$$

Since $\text{DET}(e^M) = e^{\text{tr}(M)}$ where $\text{tr}(M)$ denotes the trace of the matrix M

$$\left[\text{e.g., } \text{DET} \exp \left(\begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix} \right) = \text{DET} \begin{bmatrix} e^\lambda & 0 \\ 0 & e^\mu \end{bmatrix} = e^\lambda e^\mu = e^{\lambda + \mu} \right]$$

we see that the condition $\text{DET}(e^M) = 1$ corresponds to the condition $\text{tr}(M) = 0$. Let $s\ell_2 = \{M \mid \text{tr}(M) = 0\}$. This is the **Lie algebra of $SL(2)$** . It is generated by the matrices:

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad X^+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad X^- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

$$s\ell_2 = \left\{ \begin{pmatrix} r & s \\ t & -r \end{pmatrix} = rH + sX^+ + tX^- \right\}.$$

With $[A, B] = AB - BA$ we have the fundamental relations:

$$[H, X^+] = 2X^+$$

$$[H, X^-] = -2X^-$$

$$[X^+, X^-] = H.$$

Now define a **deformation** of this Lie Algebra (and its tensor powers) and call it $\mathcal{U}_q(sl_2)$ - the quantum universal enveloping algebra of sl_2 . It is an abstract algebra generated by symbols H, X^+, X^- and the relations:

$$\begin{aligned}[H, X^+] &= 2X^+ \\ [H, X^-] &= -2X^- \\ [X^+, X^-] &= \sinh((\hbar/2)H)/\sinh(\hbar/2)\end{aligned}$$

with $q = e^\hbar$ our familiar deformation parameter. Define a co-multiplication $\Delta : \mathcal{U}_q \rightarrow \mathcal{U}_q \otimes \mathcal{U}_q$ ($\mathcal{U}_q \equiv \mathcal{U}(sl(2))_q$) by the formulas:

$$\begin{aligned}\Delta(H) &= H \otimes 1 + 1 \otimes H \\ \Delta(X^\pm) &= X^\pm \otimes e^{(\hbar/4)H} + e^{(-\hbar/4)H} \otimes X^\pm.\end{aligned}$$

One can verify directly that \mathcal{U}_q is a bialgebra (in fact a Hopf algebra). However, we can see at once that \mathcal{U}_q is really $\mathcal{A}(\tilde{\epsilon}) = SL(2)_q$ in disguise! The mask of this disguise is a process of dualization: Consider the algebra dual of \mathcal{U}_q . This is

$$\mathcal{U}_q^* = \{\rho : \mathcal{U} \rightarrow \mathbf{C}[[\hbar, \hbar^{-1}]]\}$$

(We'll work over the complex numbers.)

Note that \mathcal{U}_q^* inherits a multiplication from \mathcal{U}_q 's co-multiplication and a co-multiplication from \mathcal{U}_q 's multiplication. This occurs thus:

$$\rho, \rho' : \mathcal{U} \rightarrow \mathbf{C}[[\hbar, \hbar^{-1}]].$$

Define $\rho\rho'(\alpha) = \rho \otimes \rho'(\Delta(\alpha))$. Define $\Delta(\rho) : \mathcal{U} \otimes \mathcal{U} \rightarrow \mathbf{C}[[\hbar, \hbar^{-1}]]$ by

$$\Delta(\rho)(\alpha \otimes \beta) = \rho(\alpha\beta),$$

where $\alpha\beta$ denotes the product of \mathcal{U} .

Consider the following **representation** of \mathcal{U}_q : $\tilde{\rho} : \mathcal{U}_q \rightarrow sl_2$

$$\begin{aligned}\tilde{\rho}(H) &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ \tilde{\rho}(X^+) &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \\ \tilde{\rho}(X^-) &= \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.\end{aligned}$$

In other words, we map each of the generators of \mathcal{U}_q to its “matrix of origin” in $s\ell_2$.

It is easy to verify that $\tilde{\rho}$ is a representation of \mathcal{U}_q as algebra. Note that

$$\begin{aligned} \sinh\left(\frac{\hbar}{2}\tilde{\rho}(H)\right) &= \sinh\left(\frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right) \\ &= \frac{1}{2}\left[\begin{pmatrix} e^{\hbar/2} & 0 \\ 0 & e^{-\hbar/2} \end{pmatrix} - \begin{pmatrix} e^{-\hbar/2} & 0 \\ 0 & e^{\hbar/2} \end{pmatrix}\right] \\ &= \begin{pmatrix} \sinh(\hbar/2) & 0 \\ 0 & -\sinh(\hbar/2) \end{pmatrix}. \end{aligned}$$

Hence

$$\frac{\sinh((\hbar/2)\tilde{\rho}(H))}{\sinh(\hbar/2)} = \tilde{\rho}(H).$$

We can write $\tilde{\rho}$ as a matrix of functions on \mathcal{U}_q :

$$\tilde{\rho} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Thus

$$\begin{aligned} a(H) &= 1, & a(X^+) &= 0, & a(X^-) &= 0 \\ b(H) &= 0, & b(X^+) &= 1, & b(X^-) &= 0 \\ c(H) &= 0, & c(X^+) &= 0, & c(X^-) &= 1 \\ d(H) &= -1, & d(X^+) &= 0, & d(X^-) &= 0. \end{aligned}$$

Lemma 9.7. The (Hopf) bialgebra generated by a, b, c, d above is isomorphic to the bialgebra $\mathcal{A}(\tilde{\epsilon}) = F_q(SL(2))$. Thus $F_q(SL(2)) \simeq \mathcal{U}_q^*$ with $q = e^{\hbar/2}$.

We have already described the multiplication and comultiplication on \mathcal{U}_q^* . The counit for \mathcal{U}_q is given by $\epsilon(1) = 1$, $\epsilon(H) = \epsilon(X^\pm) = 0$. The antipode is defined via $S(H) = -H$, $S(X^\pm) = -e^{-\frac{\hbar}{4}H}X^\pm e^{\frac{\hbar}{4}H}$. In the proof of the Lemma, I will just check bialgebra structure, leaving the Hopf algebra verifications to the reader.

Proof. We will do some representative checking. Consider $ab \in \mathcal{U}_q^*$. By definition, $ab(x) = a \otimes b(\Delta(x))$ for $x \in \mathcal{U}_q$. Thus

$$\begin{aligned} ab(H) &= a \otimes b(H \otimes 1 + 1 \otimes H) \\ \Rightarrow ab(H) &= 0. \\ ab(X^+) &= a \otimes b(X^+ \otimes e^{(\hbar/4)H} + e^{-(\hbar/4)H} \otimes X^+) \\ &= 0 + a(e^{-(\hbar/4)H}) \\ &= e^{-(\hbar/4)} \\ \text{and } ba(X^+) &= e^{(\hbar/4)}. \end{aligned}$$

Thus $ba = e^{+(\hbar/2)}ab = qab$.

Now look at $\Delta(a)$: $\Delta(a)(x \otimes y) = a(xy)$. Thus $\Delta(a)(X^+ \otimes X^-) = a(X^+ X^-)$. But in general, we see that since $a(xy)$ is the 1–1 coordinate of the matrix product $\rho(x)\rho(y)$.

$$\begin{aligned} \rho(x) &= \begin{pmatrix} a(x) & b(x) \\ c(x) & d(x) \end{pmatrix} \\ \rho(x)\rho(y) &= \begin{pmatrix} a(x) & b(x) \\ c(x) & d(x) \end{pmatrix} \begin{pmatrix} a(y) & b(y) \\ c(y) & d(y) \end{pmatrix} \\ &= \begin{pmatrix} a(x)a(y) + b(x)c(y) & a(x)b(y) + b(x)d(y) \\ c(x)a(y) + d(x)c(y) & c(x)b(y) + d(x)d(y) \end{pmatrix}. \end{aligned}$$

Whence

$$\begin{aligned} \Delta(a)(x \otimes y) &= a(x)a(y) + b(x)c(y) = (a \otimes a + b \otimes c)(x \otimes y) \\ \Rightarrow \Delta(a) &= a \otimes a + b \otimes c \\ \Delta(b) &= a \otimes b + b \otimes d \\ \Delta(c) &= c \otimes a + d \otimes c \\ \Delta(d) &= c \otimes b + d \otimes d. \end{aligned}$$

The other facts about products are easy to verify, and we leave them for the reader. This completes the proof of the Lemma. //

The dual view of $SL(2)_q^*$ as

$$\mathcal{U}_q(sl_2) : \left\{ \begin{array}{l} [H, X^+] = 2X^+ \\ [H, X^-] = -2X^- \\ [X^+, X^-] = \sinh(\hbar/2 H) / \sinh(\hbar/2) \\ (\text{and } \Delta \text{ as above.}) \end{array} \right\}$$

is important because it establishes this algebra as a deformation of the Lie algebra of $SL(2)$.

Just as one can study the representation theory of $SL(2)$ (and sl_2) one can study the representation theory of these quantum groups. Furthermore, the structure of a **universal R-matrix** (a solution to the Yang-Baxter Equation) emerges from the Lie algebra formalism. This universal solution then specializes to a number of different specific solutions-depending upon the representation that we take. In the fundamental representation $(\tilde{\rho})$, we recover the R -matrix that we have obtained in this section for the bracket.

Comment. This section has lead from the bracket as vacuum-vacuum expectation through the internal symmetries of this model that give rise to the quantum group $SL(2)_q$ and its (dual) universal quantum enveloping algebra $\mathcal{U}_q(sl_2)$. This algebraic structure arises naturally from the idea of the bracket

$[\langle \mathbf{A} \rangle = A \langle \mathbf{B} \rangle + B \langle \mathbf{C} \rangle]$ and the conditions of topological invariance - coupled with elementary quantum mechanical ideas. The bare bones of a discrete topological quantum field theory and generalizations of classical symmetry are as inevitable as the construction of elementary arithmetic.

$10^0.$ The Form of the Universal R -matrix.

First of all, there is a convenient way to regard an R -matrix algebraically. Suppose that we are given a bialgebra \mathcal{U} . Let $R \in \mathcal{U} \otimes \mathcal{U}$. Then we can define three elements of $\mathcal{U} \otimes \mathcal{U} \otimes \mathcal{U}$:

$$R_{12}, R_{13}, R_{23} \in \mathcal{U} \otimes \mathcal{U} \otimes \mathcal{U},$$

by placing the tensor factors of R in the indicated factors of $\mathcal{U} \otimes \mathcal{U} \otimes \mathcal{U}$, with a 1 in the remaining factor. Thus, if

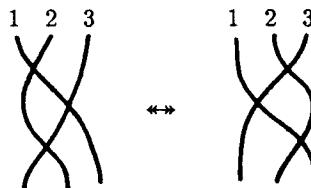
$$R = \sum_s e_s \otimes e^s \in \mathcal{U} \otimes \mathcal{U},$$

$$R_{12} = \sum_s e_s \otimes e^s \otimes 1$$

$$R_{13} = \sum_s e_s \otimes 1 \otimes e^s$$

$$R_{23} = \sum_s 1 \otimes e_s \otimes e^s.$$

The appropriate version of the Yang-Baxter Equation for this formalism is



$$R_{12}R_{13}R_{23} = R_{23}R_{13}R_{12}.$$

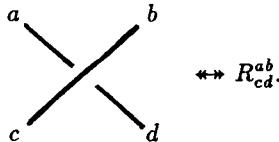
Here the vector spaces (or modules) \mathcal{U} label the lines in the diagram above the equation.

In order to relate this algebraic form of the YBE to our knot theorist's version, we must compose with a permutation. To see this, note that the knot theory uses a matrix representation of $R \in \mathcal{U} \otimes \mathcal{U}$. Let $\rho(R)$ denote such a representation. Thus $\rho(R) = (\rho_{cd}^{ab})$ where a and c are indices corresponding to the first \mathcal{U} factor and b and d are indices corresponding to the representation of the

second \mathcal{U} factor. This is the natural convention for algebra. Now define R_{cd}^{ab} by the equation

$$R_{cd}^{ab} = \rho_{dc}^{ab}$$

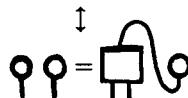
(permuting the bottom indices). We then have the diagram



Now, as desired, the crossing lines each hold the appropriate indices, and the equation $R_{12}R_{13}R_{23} = R_{23}R_{13}R_{12}$ will translate into our familiar index form of the Yang-Baxter Equation.

Now the extraordinary thing about this algebraic form of the Yang-Baxter equation is that it is possible to give general algebraic conditions that lead to solutions. To illustrate this, suppose that we know products for $\{e_s\}$ and $\{e^s\}$ in the algebra:

$$e_s e_t = m_{st}^i e_i$$



and

$$e^s e^t = \mu_i^{st} e^i \leftrightarrow \circlearrowleft \circlearrowleft = \boxed{\text{---}} \circlearrowleft \circlearrowleft .$$

Here the abstract tensors \circlearrowleft and \circlearrowright correspond to e^s and e_s , respectively. We cannot assume that \circlearrowleft and \circlearrowright commute, but the coefficient tensors $\boxed{\text{---}}$, $\boxed{\text{---}}$ commute with everybody.

Finally, assume the commutation relation

$$e_s \mu_a^{si} m_{it}^b e^t = e^t m_{ti}^b \mu_a^{is} e_s$$

Call this equation the **Yang-Baxter commutation relation**.

Theorem 10.1. Let  and  be abstract tensors satisfying a Yang-Baxter

commutation relation. Let $R = \img[alt="R tensor symbol" data-bbox="425 166 465 212"] \otimes \img[alt="R tensor symbol" data-bbox="485 166 525 212]"$ and

$$R_{12} = \img[alt="R12 tensor symbol" data-bbox="425 214 465 242"] \otimes \img[alt="R12 tensor symbol" data-bbox="485 214 525 242"] \otimes 1$$

$$R_{13} = \img[alt="R13 tensor symbol" data-bbox="425 245 465 273"] \otimes 1 \otimes \img[alt="R13 tensor symbol" data-bbox="485 245 525 273"]$$

$$R_{23} = 1 \otimes \img[alt="R23 tensor symbol" data-bbox="425 275 465 303"] \otimes \img[alt="R23 tensor symbol" data-bbox="485 275 525 303"]$$

Then $R_{12}R_{13}R_{23} = R_{23}R_{13}R_{12}$.

Proof.

$$R_{12}R_{13}R_{23} = [\img[alt="R12R13R23 tensor symbol" data-bbox="425 376 465 412"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="485 376 525 412"] \otimes 1][\img[alt="R12R13R23 tensor symbol" data-bbox="545 376 585 412"] \otimes 1 \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="605 376 645 412"]][1 \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="665 376 705 412"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="725 376 765 412"]]$$

$$= \img[alt="R12R13R23 tensor symbol" data-bbox="425 428 465 480"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="485 428 525 480"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="545 428 585 480"]$$

$$= \img[alt="R12R13R23 tensor symbol" data-bbox="425 480 465 538"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="485 480 525 538"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="545 480 585 538"]$$

$$= \img[alt="R12R13R23 tensor symbol" data-bbox="425 548 465 595"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="485 548 525 595"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="545 548 585 595"]$$

$$= \img[alt="R12R13R23 tensor symbol" data-bbox="425 605 465 658"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="485 605 525 658"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="545 605 585 658"] \quad (\text{Y-B commutation})$$

$$= \img[alt="R12R13R23 tensor symbol" data-bbox="425 660 465 718"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="485 660 525 718"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="545 660 585 718"]$$

$$= \img[alt="R12R13R23 tensor symbol" data-bbox="425 728 465 765"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="485 728 525 765"] \otimes \img[alt="R12R13R23 tensor symbol" data-bbox="545 728 585 765"]$$

$$= R_{23}R_{13}R_{12}.$$

//

The simple combinatorial pattern of this theorem controls the relationship of Lie algebras and Hopf algebras to knot theory and the YBE.

Suppose that \mathcal{U} is a bialgebra with comultiplication $\Delta : \mathcal{U} \rightarrow \mathcal{U} \otimes \mathcal{U}$. Further suppose that $\{e_s\}, \{e^s\} \subset \mathcal{U}$ with

$$e_s e_t = m_{st}^i e_i$$

$$e^s e^t = \mu_i^{st} e^i$$

and that the subalgebras A and \widehat{A} generated by $\{e_s\}$ and $\{e^s\}$ are dual in the sense that the multiplicative and comultiplicative structures in these algebras are interchanged. Thus

$$\Delta(e_t) = \mu_i^{st} e_s \otimes e^t$$

and

$$\Delta(e^i) = m_{st}^i e^s \otimes e^t.$$

Now suppose that $R = \sum_s e_s \otimes e^s$ satisfies the equation

$$R\Delta = \Delta' R \tag{*}$$

where $\Delta' = \Delta \circ \sigma$ denotes the composition of Δ with the map $\sigma : \mathcal{U} \otimes \mathcal{U} \rightarrow \mathcal{U} \otimes \mathcal{U}$ that interchanges factors.

Proposition 10.2. Under the above circumstances the equation $R\Delta = \Delta' R$ is equivalent to the Y-B commutation relation.

Proof.

$$\begin{aligned} R\Delta(e_t) &= (e_s \otimes e^s)(\mu_t^{ij} e_i \otimes e_j) \\ &= \mu_t^{ij} e_s e_i \otimes e^s e_j \\ &= \mu_t^{ij} m_{si}^\ell e_\ell \otimes e^s e_j \\ &= e_\ell \otimes e^s m_{si}^\ell \mu_t^{ij} e_j. \end{aligned}$$

$$\begin{aligned} \Delta'(e_t)R &= (\mu_t^{ij} e_j \otimes e_i)(e_s \otimes e^s) \\ &= \mu_t^{ij} e_j e_s \otimes e_i e^s \\ &= \mu_t^{ij} m_{js}^\ell e_\ell \otimes e_i e^s \\ &= e_\ell \otimes e_i \mu_t^{ij} m_{js}^\ell e^s. \end{aligned}$$

Since $\{e_\ell\}$ is a basis for A , we conclude that

$$e^s m_{si}^\ell \mu_t^{ij} e_j = e_i \mu_t^{ij} m_{js}^\ell e^s.$$

This is the Y-B commutation relation, completing the proof of the claim. //

Remark. The circumstance of $\mathcal{U} \supset A, \widehat{A}$ with these duality patterns is characteristic of the properties of quantum groups associated with Lie algebras.

In the case of $\mathcal{U}_q(sl_2)$ the algebras are the Borel subalgebras generated by H, X^+ and H, X^- . $R = \sum_s e_s \otimes e^s$ becomes a specific formula involving power series in these elements. The construction of Theorem 10.1 and its relation to the equation $R\Delta = \Delta'R$ is due to Drinfeld [DRIN1], and is called the **quantum double construction** for Hopf algebras.

While we shall not exhibit the details of the derivation of the universal R -matrix for $\mathcal{U}_q(sl(2))$, it is worth showing the end result! Here is Drinfeld's formula:

$$R = \sum_{k=0}^{\infty} \hbar^k Q_k(\hbar) \left\{ \exp \frac{\hbar}{4} [H \otimes H + k(H \otimes 1 - 1 \otimes H)] \right\} (X^+)^k \otimes (X^-)^k$$

where $Q_k(\hbar) = e^{-k\hbar/2} \prod_{r=1}^k (e^\hbar - 1)/(e^{r\hbar} - 1)$. Note that in this situation the algebras in question are infinite dimensional. Hence the sum $R = \sum_s e_s \otimes e^s$ becomes a formal power series (and the correct theory involves completions of these algebras.)

The Faddeev-Reshetikhin-Takhtajan Construction.

The equation $R\Delta = \Delta'R$ is worth contemplating on its own grounds, for it leads to a generalization of the construction of the $SL(2)$ quantum group via invariants. To see this generalization, suppose that \mathcal{U} contains an R with $R\Delta = \Delta'R$, and let $T : \mathcal{U} \rightarrow \mathcal{M}$ be a matrix representation of \mathcal{U} . If $T(u) = (t_j^i(u))$, then $t_j^i : \mathcal{U} \rightarrow \mathbb{C}$ whence $t_j^i \in \mathcal{U}^*$. [Compare with the construction in section 9⁰.] The coproduct in \mathcal{U}^* is given by

$$\begin{aligned} \Delta(t_j^i)(u \otimes v) &= t_j^i(uv) \\ &= \sum_k t_k^i(u)t_j^k(v) \end{aligned}$$

since $T(uv) = T(u)T(v)$. Therefore $\Delta(t_j^i) = t_k^i \otimes t_j^k$.

With this in mind, we can translate $R\Delta = \Delta'R$ into an equation about \mathcal{U}^* , letting

$$R_{k\ell}^{ij} = t_k^i \otimes t_\ell^j(R). \quad (\text{algebraist's convention})$$

Lemma. $R\Delta = \Delta' R$ corresponds to the equation $\boxed{RT_1T_2 = T_2T_1R}$ where

$$R = (R_{k\ell}^{ij}), \quad T = (t_j^i), \quad T_1 = T \otimes 1, \quad T_2 = 1 \otimes T.$$

Proof.

$$\begin{aligned} R\Delta(X) : \quad & t_k^i \otimes t_\ell^j(R\Delta(X)) \\ = & \sum_{m,n} t_m^i \otimes t_n^j(R)t_k^m \otimes t_\ell^n(\Delta(X)) \\ = & \sum_{m,n} R_{mn}^{ij} t_k^m t_\ell^n(X) \\ & \text{(definition of product in } \mathcal{U}^*) \end{aligned}$$

$$\begin{aligned} \Delta'(X)R : \quad & t_k^i \otimes t_\ell^j(\Delta'(X)R) \\ = & t_n^j \otimes t_m^i(\Delta(X))t_k^m \otimes t_\ell^n(R) \\ = & t_n^j t_m^i(X) R_{k\ell}^{mn}. \end{aligned}$$

Thus $R\Delta(X) = \Delta'(X)R$ implies that

$$\sum_{m,n} R_{mn}^{ij} t_k^m t_\ell^n = \sum_{m,n} t_n^j t_m^i R_{k\ell}^{mn}.$$

Letting $T = (t_j^i)$ and $T_1 = T \otimes 1$, $T_2 = 1 \otimes T$. This matrix equation reads

$$\boxed{RT_1T_2 = T_2T_1R}.$$

Faddeev, Reshetikhin and Takhtajan [FRT] take the equation $RT_1T_2 = T_2T_1R$ as a starting place for constructing a bialgebra from a solution to the Yang-Baxter Equation. We shall call this the **FRT construction**. In the FRT construction

one starts with a given solution R to YBE and associates to it the bialgebra $A(R)$ with generators $\{t_j^i\}$ and relations $T_2 T_1 R = R T_1 T_2$. The coproduct is given by $\Delta(t_j^i) = \sum_k t_k^i \otimes t_j^k$.

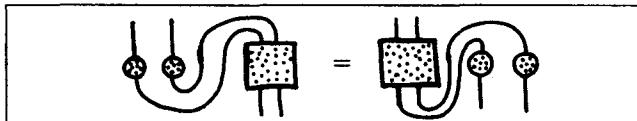
Now in order to perform explicit computations with the FRT construction, I shall assume that the R -matrix is in knot-theory format. That is, we replace R_{cd}^{ab} by R_{dc}^{ab} in the above discussion. Then it is clear that the defining equation for the FRT construction becomes

$$R T_1 T_2 = T_1 T_2 R .$$

In index form, the FRT relations become:

$$T_i^a T_j^b R_{cd}^{ij} = R_{ij}^{ab} T_c^i T_d^j .$$

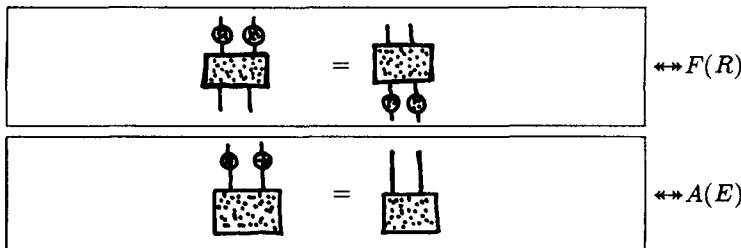
If  denotes T_j^i , and  denotes R , then these relations read:



Since we assume that the elements R_{kl}^{ij} commute with the algebra generators T_b^a , we can write the Faddeev-Reshetikhin-Takhtajan construction as:



Given an R matrix, we shall denote this algebra by $F(R)$. It is a bialgebra for exactly the same reasons as the related case of invariance (the algebra $A(E)$) discussed in section 9⁰. Diagrammatically, the comparison of $A(E)$ and $F(R)$ is quite clear:



A clue regarding the relationship of these constructions is provided by a direct look at the R -matrix for $SL(2)_q$. Recall from section 9⁰ that we have $SL(2)_q = A(E)$ where

$$E = \tilde{\epsilon} = \sqrt{-1}\epsilon_{ab}q^{\epsilon_{ab}/2}$$

and that the knot theory provided us with the associated R -matrix

$$R = \sqrt{q} \left(\text{Diagram} + \frac{1}{\sqrt{q}} \right) \left(\text{Diagram} \right)$$

Thus we see at once that any matrix \otimes satisfying $\text{Diagram} = \text{Diagram}$ will automatically satisfy the relations for $F(R)$:

The first diagram shows two strands entering a square box labeled \otimes , with two strands exiting. The second diagram shows the same setup but with a factor of $\frac{1}{\sqrt{q}}$ added. The third diagram shows the result of the multiplication, which is the original diagram plus the scaled version.

Therefore $F(R)$ has $A(E)$ as a quotient.

In fact, a direct check reveals that $F(R)$ is isomorphic to the algebra with generators a, b, c, d and relations:

$ba = qab$	$da = qad$
$db = qbd$	$ca = qac$
$bc = cb$	
$ad - da = (q^{-1} - q)bc$	

The center of $F(R)$ is generated by the element $(ad - q^{-1}bc)$, and, as we know, $A(E) \cong F(R)/(ad - q^{-1}bc)$ where the bracket denotes the ideal generated by this relation. In any case, the Faddeev-Reshetikhin-Takhtajan construction is very significant in that it shows how to directly associate a Hopf algebra or at least a bialgebra to any solution of the Yang-Baxter Equation.

The quantum double construction shows that certain bialgebras will naturally give rise to solutions to the Yang-Baxter Equation from their internal structure. The associated bialgebra (quantum group) provides a rich structure that can be utilized to understand the structure of these solutions and the structure of allied topological invariants.

Hopf Algebras and Link Invariants.

It may not yet be clear where the Hopf algebra structure is implicated in the construction of a link invariant. We have shown that the Drinfeld double construction yields formal solutions to the Yang-Baxter Equation, but this is accomplished without actually using the comultiplication if one just accepts the “Yang-Baxter commutation relation”

As for the antipode $\gamma : \mathcal{A} \rightarrow \mathcal{A}$, it doesn't seem to play any role at all. Not so! Seen rightly, the Drinfeld construction is an algebraic version of the vacuum-vacuum expectation model. In order to see this let's review the construction, and look at the form of its representation. We are given an algebra \mathcal{A} with multiplicative generators e_0, e_1, \dots, e_n and e^0, e^1, \dots, e^n . (The same formalism can be utilized to deal with formal power series if the algebras are not finite dimensional.) We can assume that $e_0 = e^0 = 1$ (a multiplicative unit) in \mathcal{A} , and we assume rules for multiplication:

$$\begin{aligned} e_i e_j &= m_{ij}^k e_k \\ e^i e^j &= \mu_k^{ij} e^k \end{aligned}$$

(summation convention operative). We further assume a comultiplication $\Delta : \mathcal{A} \rightarrow \mathcal{A} \otimes \mathcal{A}$ such that

$$\begin{aligned} \Delta(e_k) &= \mu_k^{ij} e_i \otimes e_j \\ \Delta(e^k) &= m_{ij}^k e^i \otimes e^j. \end{aligned}$$

That is, lower and upper indexed algebras are dual as bialgebras, the comultiplication of one being the multiplication of the other. Letting $\rho = e_s \otimes e^s$, we assume that ρ satisfies the Yang-Baxter commutation relation in the form $\rho\Delta = \Delta'\rho$ where Δ' is the composition of Δ with the permutation map on $\mathcal{A} \otimes \mathcal{A}$. With these assumptions it follows that ρ satisfies the algebraic form of the Yang-Baxter Equation: $\rho_{12}\rho_{13}\rho_{23} = \rho_{23}\rho_{13}\rho_{12}$, (Theorem 10.1). Finally we assume that

there is an anti-automorphism $\gamma : \mathcal{A} \rightarrow \mathcal{A}$ that is an **antipode** in the (direct) sense that

$$m((1 \otimes \gamma)\Delta(e_k)) = m((\gamma \otimes 1)\Delta(e^k)) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

(Same statement for e^k and e_k in these positions.).

Now suppose that $\text{rep} : \mathcal{A} \rightarrow M$ where M is an algebra of matrices (with commutative entries). If we let \bullet denote an element of \mathcal{A} , then $\text{rep}(\bullet) = \bullet$. That is, the representation of \bullet becomes an object with upper and lower matrix indices. For thinking about the representation, it is convenient to suppress the indices k and ℓ in e_k, e^ℓ . Thus, we can let $\bullet \rightsquigarrow \bullet$ denote $\rho = e_k \otimes e^k$ with the wavy line indicating the summation over product of elements that constitute ρ . Then we have:

$$\text{rep}(\bullet \rightsquigarrow \bullet) = \bullet \otimes \bullet$$

and we define an R -matrix via

$$R = \text{---} \otimes \text{---}$$

Note once again how the permutation of the lower lines makes R into a knot-theoretic solution to the Yang-Baxter Equation:

$$\text{---} \otimes \text{---} \approx \text{---} \otimes \text{---}$$

$$\rho_{12}\rho_{13}\rho_{23} = \rho_{23}\rho_{13}\rho_{12}.$$

For the purpose of the knot theory, we can let R correspond to a crossing of the "A" form:

$$R = \text{---} \otimes \text{---} \Leftrightarrow \text{---} \uparrow t$$

(with respect to a vertical arrow of time). Then we know that R^{-1} corresponds to the opposite crossing, and by the twist relation R^{-1} is given by the formula

$$R^{-1} = \text{Diagram with } R \text{ in a circle} = \text{Diagram with } R^{-1} \text{ in a circle}$$

where it is understood that the raising and lowering of indices on R to form R^{-1} is accomplished through the creation and annihilation matrices

$$M^{ab} \leftrightarrow \text{Diagram with } M^{ab} \text{ and } M_{ab} \leftrightarrow \text{Diagram with } M_{ab}$$

Keeping this in mind, we can deform the diagrammatic for R^{-1} so long as we keep track of the maxima and minima. Hence

$$R^{-1} = \text{Diagram with wavy lines and dots}$$

This suggests that there should be a mapping $\gamma : \mathcal{A} \rightarrow \mathcal{A}$ such that

$$\text{rep}(\gamma)(\bullet) = \text{Diagram with dot}$$

Therefore, we shall assume that the antipode $\gamma : \mathcal{A} \rightarrow \mathcal{A}$ is so represented.

Then we see that the equation $RR^{-1} = I$ becomes:

$$\text{Diagram with } R \text{ and } R^{-1} = \text{Diagram with dot} = \text{Diagram with dot}$$

In other words, we must have $\sum_s (e_s \otimes e^s)(\gamma(e_t) \otimes e^t) = 1 \otimes 1$ in the algebra.

With γ the antipode of \mathcal{A} , this equation is true:

$$\begin{aligned} & (e_s \otimes e^s)(\gamma(e_t) \otimes e^t) \\ &= e_s \gamma(e_t) \otimes e^s e^t \\ &= e_s \gamma(e_t) \otimes \mu_t^{st} e^t \\ &= \mu_t^{st} e_s \gamma(e_t) \otimes e^t \\ &= m(1 \otimes \gamma)(\mu_t^{st} e_s \otimes e_t) \otimes e^t \\ &= m((1 \otimes \gamma)\Delta(e_t)) \otimes e^t \quad (\text{by assumption about comultiplication}) \\ &= 1 \otimes 1 \quad (\text{by definition of antipode}). \end{aligned}$$

Thus we have shown:

Theorem 10.3. If \mathcal{A} is a Hopf algebra satisfying the assumptions of the Drinfeld construction, and if \mathcal{A} is represented so that the antipode is given by matrices M_{ab} , M^{ab} (inverse to one another) so that

$$\text{rep}(\gamma(\bullet)) = \begin{array}{c} \bullet \\ \curvearrowleft \\ \curvearrowright \end{array}$$

Then the matrices R , R^{-1} , M_{ab} , M^{ab} (specified above) satisfy the conditions for the construction of a regular isotopy invariant of links via the vacuum-vacuum expectation model detailed in this section.

The formalism of the Hopf algebra provides an algebraic model for the category of abstract tensors that provide regular isotopy invariants. In this sense the Drinfeld construction creates a universal link invariant (compare [LK26] and [LAW1]).

A Short Guide to the Definition of Quasi-Triangular Hopf Algebras.

Drinfeld [DRIN1] generalized the formalism of the double construction to the notion of **quasi-triangular Hopf algebra**. A Hopf algebra \mathcal{A} is said to be quasi-triangular if it contains an element R satisfying the following identities:

- (1) $R\Delta = \Delta'R$
- (2) $R_{13}R_{23} = (\Delta \otimes 1)(R)$
- (3) $R_{12}R_{13} = (1 \otimes \Delta)(R)$. We have already discussed the motivation behind (1).

To understand (2) and (3) let's work in the double formalism; then

$$\begin{aligned} R_{13}R_{23} &= e_s \otimes e_t \otimes e^s e^t \\ &= e_s \otimes e_t \otimes \mu_i^{st} e^i \\ &= \mu_i^{st} e_s \otimes e_t \otimes e^i \\ &= \Delta(e_i) \otimes e^i \\ \therefore R_{13}R_{23} &= (\Delta \otimes 1)(R). \end{aligned}$$

Similarly, $R_{12}R_{13} = (1 \otimes \Delta)(R)$.

The notion of quasi-triangularity encapsulates the duality structure of the double construction indirectly and axiomatically. We obtain index-free proofs of all the standard properties. For example, if \mathcal{A} is quasi-triangular, then $(\Delta' \otimes 1)(R) =$

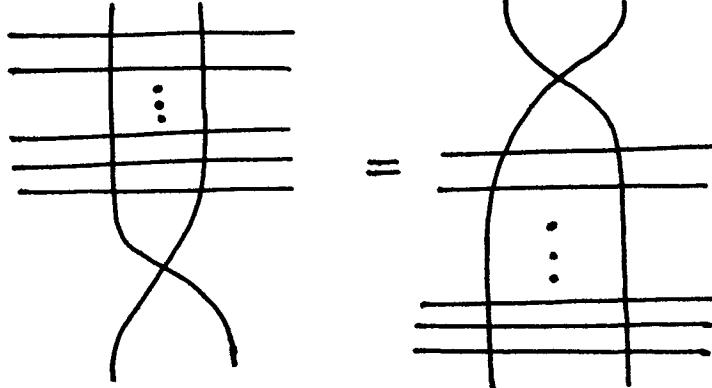
$R_{23}R_{13}$ since this formula is obtained by switching two factors in $(\Delta \otimes 1)R = R_{13}R_{23}$ (interchange 1 and 2). From this we calculate

$$\begin{aligned} R_{12}R_{13}R_{23} &= R_{12}(\Delta \otimes 1)(R) \\ &= (\Delta' \otimes 1)(R)R_{12} \quad (R\Delta = \Delta'R) \\ &= R_{23}R_{13}R_{12}. \end{aligned}$$

Hence R satisfies the Yang-Baxter Equation. It is also the case that $R^{-1} = (1 \otimes \gamma)R$ when \mathcal{A} is quasi-triangular.

Statistical Mechanics and the FRT Construction.

The Yang-Baxter Equation originated in a statistical mechanics problem that demanded that an R -matrix associated with a 4-valent vertex commute with the row-to-row transfer matrix for the lattice [BA1]. Diagrammatically, this has the form



In these terms, we can regard this equation as a representation of the equation $T_1T_2R = RT_1T_2$ with $T = \begin{smallmatrix} & & \\ \text{---} & \text{---} & \\ & & \end{smallmatrix}$ and T_1T_2 denoting matrix multiplication as shown in the diagram. In particular, this means that, for a knot-theoretic R solving the Yang-Baxter Equation, $(t_b^a)_j^i = R_{bj}^{ai}$ gives a solution to the FRT equations $t_i^a t_j^b R_{cd}^{ij} = R_{ij}^{ab} t_c^i t_d^j$. Thus, for R a solution to the Yang-Baxter Equation, the FRT construction comes equipped with a matrix representation that is constructed from submatrices of the R -matrix itself.

11⁰. Yang-Baxter Models for Specializations of the Homfly Polynomial.

We have seen in section 9⁰ that a vacuum-vacuum expectation model for the bracket leads directly to a solution to the YBE and that this solution generalizes to the R -matrices

$$R = (q - q^{-1}) \begin{array}{c} \text{---} \\ \diagup \quad \diagdown \\ a \quad b \end{array} + q \begin{array}{c} \text{---} \\ \diagup \quad \diagdown \\ a \quad b \end{array} \times \begin{array}{c} \text{---} \\ \diagup \quad \diagdown \\ a \quad b \end{array}$$

$$\bar{R} = (q^{-1} - q) \begin{array}{c} \text{---} \\ \diagup \quad \diagdown \\ a \quad b \end{array} + q^{-1} \begin{array}{c} \text{---} \\ \diagup \quad \diagdown \\ a \quad b \end{array} \times \begin{array}{c} \text{---} \\ \diagup \quad \diagdown \\ a \quad b \end{array}$$

The edges of these small diagrams are labelled with “spins” from an arbitrary index set \mathcal{I} and

$$\begin{array}{c} a \\ \diagup \quad \diagdown \\ \text{---} \\ \diagup \quad \diagdown \\ b \end{array} \leftrightarrow a < b, \quad \begin{array}{c} a \\ \diagup \quad \diagdown \\ \text{---} \\ \diagup \quad \diagdown \\ b \end{array} \leftrightarrow a > b$$

$$\begin{array}{c} a \\ \diagup \quad \diagdown \\ \text{---} \\ \diagup \quad \diagdown \\ b \end{array} \leftrightarrow a = b$$

$$\begin{array}{c} a \quad b \\ \diagup \quad \diagdown \\ \text{---} \\ \diagup \quad \diagdown \\ b \end{array} \leftrightarrow a \neq b$$

so that (e.g.) $\begin{array}{c} a \\ \diagup \quad \diagdown \\ c \quad d \end{array} = [a < b]\delta_c^a\delta_d^b$ where $[a < b] = \begin{cases} 1 & \text{if } a < b \\ 0 & \text{if } a \not< b \end{cases}$

(In general, $[P] = \begin{cases} 1 & \text{if } P \text{ is true} \\ 0 & \text{if } P \text{ is false} \end{cases}$). Thus

$$R_{cd}^{ab} = (q - q^{-1})[a < b]\delta_c^a\delta_d^b + q[a = b]\delta_c^a\delta_d^b + [a \neq b]\delta_d^a\delta_c^b.$$

The structure of these diagrammatic forms fits directly into corresponding state models for knot polynomials - as we shall see.

The models that we shall consider in this section are defined for oriented knot and link diagrams. I shall continue to use the bracket formalism:

$$\langle K \rangle = \sum_{\sigma} \langle K | \sigma \rangle \delta^{\parallel \sigma \parallel}$$

where the states σ , products of vertex weights $\langle K | \sigma \rangle$, and state norm $\parallel \sigma \parallel$ will be defined below. We shall then find that $\langle K \rangle$ can be adjusted to give invariants of

regular isotopy for an infinite class of specializations of the Homfly polynomial. This gives a good class of models, enough to establish the existence of the full 2-variable Homfly polynomial. Recall that we have discussed this polynomial and a combinatorial (not Yang-Baxter) model in section 5⁰.

How should this state model be defined? Since the R -matrix has a nice expansion, it makes sense to posit:

$$\begin{aligned}\langle \text{X} \rangle &= (q - q^{-1}) \langle \text{Y} \rangle + q \langle \text{Z} \rangle + \langle \text{A} \rangle \\ \langle \text{A} \rangle &= (q^{-1} - q) \langle \text{Y} \rangle + q^{-1} \langle \text{Z} \rangle + \langle \text{X} \rangle\end{aligned}$$

for the state model. This is equivalent to saying that the R -matrix determines the local vertex weights.

Note that, we then have

$$\langle \text{X} \rangle - \langle \text{A} \rangle = (q - q^{-1}) [\langle \text{Y} \rangle + \langle \text{Z} \rangle + \langle \text{A} \rangle]$$

and, since given two spins a, b we have either $a < b$, $a > b$ or $a = b$, we have

$$\langle \text{A} \rangle = \langle \text{Y} \rangle + \langle \text{Z} \rangle + \langle \text{A} \rangle.$$

Therefore,

$$\langle \text{X} \rangle - \langle \text{A} \rangle = (q - q^{-1}) \langle \text{A} \rangle.$$

Hence the model satisfies the exchange identity for the regular isotopy version of the Homfly polynomial. Since the edges are labelled with spins from the index set \mathcal{I} , we see that the **states** σ of this model are obtained by

1. Replace each crossing by either a decorated splice:



or by a graphical crossing:



2. Label the resulting diagram σ with spins from \mathcal{I} so that each loop in σ has

constant spin, and so that the spins obey the rules:

$$\text{a) } \begin{array}{c} a \\ \searrow \\ \curvearrowleft \end{array} \begin{array}{c} b \\ \swarrow \end{array} \Rightarrow a < b , \quad \text{b) } \begin{array}{c} a \\ \searrow \\ \curvearrowright \end{array} \begin{array}{c} b \\ \swarrow \end{array} \Rightarrow a = b$$

$$\begin{array}{c} a \\ \searrow \\ \curvearrowleft \end{array} \begin{array}{c} b \\ \swarrow \end{array} \Rightarrow a > b$$

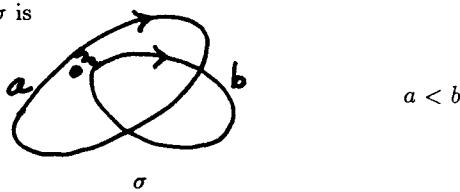
$$\begin{array}{c} a \\ \searrow \\ \cancel{\curvearrowright} \end{array} \begin{array}{c} b \\ \swarrow \end{array} \Rightarrow a \neq b.$$

[A diagram may have no labelling - in such a case it does not contribute to $\langle K \rangle$.]

Example. If K is a trefoil diagram



then a typical state σ is



Note that the loops in a state have no self-crossings (since such a crossing would force a label to be unequal to itself).

In conformity with the notion that $\langle K \rangle$ is a vacuum-vacuum expectation (generalizing from the case of the unoriented bracket of section 3⁰), we see that each individual loop of constant spin a must receive a value that depends only on a and the rotational direction of the loop. We choose this evaluation by the formula

$$\delta^{\|\sigma\|} = \langle \sigma \rangle$$

where δ is a variable whose relation with q is as-yet to be elucidated, and $\|\sigma\|$ is the sum of the spins of the loops in σ multiplied by ± 1 according to the rotational sense of the loop. That is,

$$\|\sigma\| = \sum_{\ell \in \text{components}(\sigma)} \text{rot}(\ell) \cdot \text{label}(\ell)$$

where $\text{label}(\ell)$ is the spin assigned to the loop ℓ and $\text{rot}(\ell) = \pm 1$ with the convention:

$$\text{rot}\left(\begin{array}{c} \circlearrowleft \\ \circlearrowright \end{array}\right) = +1$$

$$\text{rot}\left(\begin{array}{c} \circlearrowright \\ \circlearrowleft \end{array}\right) = -1.$$

Thus for



we have $\|\sigma\| = -a - b$ and $\langle \sigma \rangle = \delta^{-a-b}$.

Remark. In terms of the literal picture of a vacuum-vacuum expectation we need to specify creation and annihilation matrices

$$\begin{array}{cccc} \curvearrowleft & \curvearrowright & \curvearrowleft & \curvearrowright \\ \overrightarrow{M}_{ab} & \overrightarrow{M}_{ab} & \overrightarrow{M}^{ab} & \overrightarrow{M}^{ab} \end{array}$$

so that

$$\langle \begin{array}{c} \circlearrowleft \\ \circlearrowright \end{array} \rangle = \sum_{a \in \mathcal{I}} q^{-a}$$

and

$$\langle \begin{array}{c} \circlearrowright \\ \circlearrowleft \end{array} \rangle = \sum_{a \in \mathcal{I}} q^a$$

since a given loop has states with arbitrary labels from \mathcal{I} .

There is one case where these matrices appear naturally. Suppose that $a \in \mathcal{I} \Leftrightarrow -a \in \mathcal{I}$. (Positive spin is theoretically accompanied by corresponding negative spin.) Then interpret

$$a \curvearrowleft b = q^{a/2} \Delta_{a,b}$$

as a deformation of a reversed Kronecker delta:

$$\Delta_{a,b} = \delta_{a,-b} = \begin{cases} 1 & a = -b \\ 0 & \text{otherwise.} \end{cases}$$

Here we use the vertical direction (of time for the vacuum-vacuum process) so that a constant labelling of the arc \cap is interpreted as

$$\uparrow +a \quad \text{and} \quad \downarrow -a$$

depending on the arrow's ascent or descent.

(A particle travelling backwards in time reverses its spin and charge.) The matrices become:

$$\begin{aligned} \text{Diagram 1: } & a \uparrow \text{ and } b \uparrow \quad = q^{a/2} \Delta_{a,b} \\ \text{Diagram 2: } & a \downarrow \text{ and } b \uparrow \quad = q^{-a/2} \Delta_{a,b} \\ \text{Diagram 3: } & a \uparrow \text{ and } b \downarrow \quad = q^{-a/2} \Delta^{a,b} \\ \text{Diagram 4: } & a \downarrow \text{ and } b \downarrow \quad = q^{a/2} \Delta^{a,b}. \end{aligned}$$

For the symmetrical index set ($a \in \mathcal{I} \Leftrightarrow -a \in \mathcal{I}$) these matrices produce the given expectations for the trivial loops. The matrices simply state probabilities for the creation or annihilation of particles/antiparticles with opposite spin or the continued trajectory of a given particle "turning around in time". The whole point about a topological theory of this type is that the vacuum-vacuum expectations are independent of time and of the direction of time's arrow.

Returning now to the model itself we have $\langle K \rangle = \sum_{\sigma} \langle K | \sigma \rangle \langle \sigma \rangle$ where the σ can be regarded as diagrammatic "state-holders" (ready to be labelled with spins) or as labelled spin-states. A state-holder such as



is the diagrammatic form of a particular process whose expectation involves the summation over all spin assignments that satisfy the process. Note that we are free

to visualize particles moving around the loops, or a sequence of creations, interactions and annihilations. The latter involves a choice for the arrow of time. The former involves a choice of **meta-time** for the processes running in these trajectories. The topological theory maintains independence of meta-time conventions just as it is aloof from time's arrow.

In order for this theory to really be topological, we must examine the behavior of $\langle K \rangle$ under Reidemeister moves. We shall adjust δ to make $\langle K \rangle$ an invariant of regular isotopy.

For arbitrary q , δ and index set \mathcal{I} , it is easy to see that $\langle K \rangle$ is an invariant of moves II(A) and III(A) (non-cyclic with all crossings of some type). II(A) is easy since R and \bar{R} are inverses and since the state evaluations $\delta^{\parallel\sigma\parallel}$ do not involve crossing types. Nevertheless, the technique of checking II(A) by state-expansion is worth illustrating. Therefore

Lemma 11.1. $\left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle = \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle$ for any index set \mathcal{I} and variables q and δ .

Proof.

$$\begin{aligned} \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle &= z \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle + q \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle + \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle \\ (z = q - q^{-1}) \\ \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle &= z \left\{ -z \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle + q^{-1} \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle + \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle \right\} \\ &\quad + q \left\{ -z \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle + q^{-1} \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle + \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle \right\} \\ &\quad + \left\{ -z \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle + q^{-1} \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle + \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle \right\}. \end{aligned}$$

Now certain combinations are incompatible. Thus $\left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle = 0 = \left\langle \begin{array}{c} \text{Diagram} \\ \downarrow \end{array} \right\rangle$.

Therefore

$$\begin{aligned}\langle \text{Diagram} \rangle &= z \left\{ 0 + 0 + \langle \text{Diagram} \rangle \right\} \\ &\quad + q \left\{ 0 + q^{-1} \langle \text{Diagram} \rangle + 0 \right\} \\ &\quad + \left\{ -z \langle \text{Diagram} \rangle + \langle \text{Diagram} \rangle \right\}.\end{aligned}$$

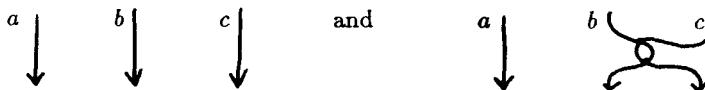
Now note that $\langle \text{Diagram} \rangle = \langle \downarrow \downarrow \rangle - \langle \downarrow \text{loop} \downarrow \rangle$ and that $\langle \text{Diagram} \rangle = \langle \downarrow \downarrow \rangle$. Thus $\langle \text{Diagram} \rangle = \langle \downarrow \downarrow \rangle$ and this proves the lemma. //

Lemma 11.2. The state model $\langle K \rangle$, as described above, is invariant under the oriented move III(A) for any ordered set \mathcal{I} and any choice of the variables q and δ .

Proof. Note that the model has the form

$$\langle K \rangle = \sum_{\sigma} \langle K | \sigma \rangle \delta^{\|\sigma\|}$$

where σ runs over all spin-labellings of the diagram K , $\langle K | \sigma \rangle$ is the product of vertex weights corresponding to the matrices R and \bar{R} and $\|\sigma\|$ is the loop count as explained above. Since R and \bar{R} satisfy the YBE, it suffices to check that for given inputs and outputs to a triangle (for the III - move) $\|\sigma\|$ is the same for all contributing internal states. A comparison with the verification in Proposition 9.4 shows that this is indeed the case. For example



have the same norm since corresponding labels and rotation numbers are the same. This completes the proof. //

The next lemma is the key to bracket behavior under the reverse type II move.

Lemma 11.3. $\langle K \rangle$ is invariant under the reversed type II move, and hence is an invariant of regular isotopy, if and only if the following conditions are satisfied:

$$(i) q^{-1} \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{one arrow clockwise, one counter-clockwise} \end{array} \right\rangle = q \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows clockwise} \end{array} \right\rangle$$

$$(ii) q^{-1} \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows counter-clockwise} \end{array} \right\rangle = q \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{one arrow clockwise, one counter-clockwise} \end{array} \right\rangle$$

Call these the **cycle conditions** for the model $\langle K \rangle$.

Remark. Note that we have reverted to the direct notation - indicating by inequality signs the spin relationships between neighboring lines.

Proof. The proof is given for one type of orientation for the type IIB move. The other case is left for the reader. Let $z = q - q^{-1}$. Expanding the form of the II move, we find:

$$\begin{aligned} \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{one arrow clockwise, one counter-clockwise} \end{array} \right\rangle &= z \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows clockwise} \end{array} \right\rangle + q \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{one arrow clockwise, one counter-clockwise} \end{array} \right\rangle + \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows counter-clockwise} \end{array} \right\rangle \\ &= z \left\{ -z \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows clockwise} \end{array} \right\rangle + \bar{q}^{-1} \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows counter-clockwise} \end{array} \right\rangle + \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{one arrow clockwise, one counter-clockwise} \end{array} \right\rangle \right\} \\ &\quad + \bar{q} \left\{ -z \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows clockwise} \end{array} \right\rangle + \bar{q}^{-1} \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows counter-clockwise} \end{array} \right\rangle + \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{one arrow clockwise, one counter-clockwise} \end{array} \right\rangle \right\} \\ &\quad + \left\{ -z \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows clockwise} \end{array} \right\rangle + \bar{q}^{-1} \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows counter-clockwise} \end{array} \right\rangle + \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{one arrow clockwise, one counter-clockwise} \end{array} \right\rangle \right\} \\ &= -z^2 \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows clockwise} \end{array} \right\rangle + z \bar{q}^{-1} \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows counter-clockwise} \end{array} \right\rangle - z \bar{q} \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows clockwise} \end{array} \right\rangle + \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{one arrow clockwise, one counter-clockwise} \end{array} \right\rangle + \left\langle \begin{array}{c} \text{circle with two arrows} \\ \text{both arrows counter-clockwise} \end{array} \right\rangle. \end{aligned}$$

In order to simplify this expression, note that

$$\langle \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \rangle = \langle \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \rangle + \langle \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \rangle$$

since crossed lines have unequal indices.

Also, it is easy to check that

$$\langle \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \rangle = \langle \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \rangle.$$

[This is an identity about rotation numbers:

$$(i) \quad \langle \begin{array}{c} Q_a \\ Q_a \end{array} \rangle = \delta^{-2a}, \quad \langle \begin{array}{c} \text{circle} \\ \text{wavy line} \end{array} \rangle = \delta^{-2a}$$

$$(ii) \quad \langle \begin{array}{c} \text{circle} \\ \text{wavy line} \end{array} \rangle = \delta^{-a}, \quad \langle \begin{array}{c} Q_a \\ Q_a \\ Q_a \end{array} \rangle = \delta^{a-a+a} = \delta^{-a}. \quad]$$

Therefore,

$$\begin{aligned} \langle \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \rangle &= -z^2 \langle \begin{array}{c} \text{circle} \\ \text{wavy line} \end{array} \rangle \\ &\quad + z\bar{z}^{-1} \langle \begin{array}{c} \text{circle} \\ \text{wavy line} \end{array} \rangle - z\bar{z} \langle \begin{array}{c} \text{circle} \\ \text{wavy line} \end{array} \rangle \\ &\quad + \langle \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \rangle. \end{aligned}$$

Thus, a necessary condition for invariance under the move IIB is that the sum of the first three terms on the right hand side of this equation should be zero.

Dividing by z , we have:

$$\begin{aligned}\phi &= -z \left\langle \begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} \right\rangle + \bar{z}^{-1} \left\langle \begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array} \right\rangle + \bar{z}^1 \left\langle \begin{array}{c} \text{Diagram 5} \\ \text{Diagram 6} \end{array} \right\rangle \\ &= -z \left\{ \begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} \right\} + \left\{ \begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array} \right\} + \bar{z}^1 \left\{ \begin{array}{c} \text{Diagram 5} \\ \text{Diagram 6} \end{array} \right\}\end{aligned}$$

Hence

$$\phi = -z \left\langle \begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} \right\rangle + \bar{z}^{-1} \left\langle \begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array} \right\rangle.$$

This proves the first cycle condition. The second follows from the need for the invariance

$$\left\langle \begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} \right\rangle = \left\langle \begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array} \right\rangle$$

with all essential details the same. This completes the proof of the Lemma. //

At this point it is possible to see that the index sets

$$\mathcal{I}_n = \{-n, -n+2, -n+4, \dots, n-2, n\}$$

($n = 1, 2, 3, \dots$) will each yield invariants of regular isotopy for $\langle K \rangle$. More precisely:

Proposition 11.4. Let $\mathcal{I}_n = \{-n, -n+2, \dots, n-2, n\}$ ($n \in \{1, 2, 3, \dots\}$) and let $\delta = q$, $z = q - q^{-1}$ so that $\langle K \rangle = \sum_{\sigma} \langle K | \sigma \rangle q^{\|\sigma\|}$. Then $\langle K \rangle$ is an invariant of

regular isotopy, and it has the following behavior under type I moves:

$$\begin{aligned}\left\langle \overrightarrow{\text{O}} \right\rangle &= q^{n+1} \left\langle \overrightarrow{\text{C}} \right\rangle \\ \left\langle \overleftarrow{\text{O}} \right\rangle &= q^{-n-1} \left\langle \overrightarrow{\text{C}} \right\rangle.\end{aligned}$$

Proof. To check regular isotopy, it is sufficient to show that the cycle conditions (Lemma 11.3) are satisfied. Here is the calculation for the first condition.

$$\begin{aligned}q^{-1} \left\langle \overrightarrow{\text{O}} \right\rangle - q \left\langle \overrightarrow{\text{O}} \right\rangle \\ = \sum_{a>c} \left[q^{-1} \sum_{a>b \geq c} q^{-b} - q \sum_{a \geq b > c} q^{-b} \right] \left\langle \overrightarrow{\text{C}} \right\rangle.\end{aligned}$$

With the choice of index set as given above, it is easy to see that each coefficient $\left[q^{-1} \sum_{a>b \geq c} q^{-b} - q \sum_{a \geq b > c} q^{-b} \right]$ vanishes identically. The second cycle condition is satisfied in exactly the same manner.

Direct calculation now shows the behavior under the type I moves. Since we shall generalize this calculation in the next proposition, I omit it here. //

Remark. With index set \mathcal{I}_n , and $\delta = q$ we can define

$$P_K^{(n)}(q) = (q^{n+1})^{-w(K)} \langle K \rangle / \langle 0 \rangle$$

where $w(K)$ denotes the writhe of K . Then $P_k^{(n)}$ satisfies the conditions

1. K ambient isotopic to $K' \Rightarrow P_K^{(n)}(q) = P_{K'}^{(n)}(q)$.

2. $P_O^{(n)} = 1$

$$q^{n+1} P_K^{(n)} - q^{-n-1} P_K^{(n)} = (q - q^{-1}) P_C^{(n)}.$$

Thus we see that $P_K^{(n)}$ is a one-variable specialization of the Homfly polynomial (Compare [JO4]). The entire collection of specializations $\{P_K^{(n)} \mid n = 1, 2, 3, \dots\}$ is sufficient to establish the existence of the Homfly polynomial itself (as a 2-variable polynomial invariant of knots and links). Although this route to the 2-variable oriented polynomial is a bit indirect, it is a very good way to understand the nature of the invariant.

Note that in this formulation the original Jones polynomial appears for the index set $\mathcal{I}_1 = \{-1, +1\}$ and $q = \sqrt{i}$. This is an oriented state model for the Jones polynomial based on the Yang-Baxter solution

$$R = (q - q^{-1}) \begin{pmatrix} & \\ \downarrow & \end{pmatrix} \begin{pmatrix} & \\ + & q \end{pmatrix} = \begin{pmatrix} & \\ + & \end{pmatrix} \begin{pmatrix} & \\ \cancel{\times} & \end{pmatrix}$$

We have already remarked (9⁰) on the relationship between this R (for a two-element index set) and the Yang-Baxter solution that arises naturally from the bracket model. It is a nice exercise to see that these models translate into one another via the Fierz identity, as explained in section 10⁰, for the R -matrices.

A further mystery about this sequence of models is the way they avoid the classical Alexander-Conway polynomial. We may regard the Conway polynomial as an ambient isotopy invariant ∇_K , with $\nabla_O = 1$ and

$$\nabla_{\cancel{\times}} - \nabla_{\times} = z \nabla_{\cancel{\cup}}.$$

Thus ∇_K seems to require $n = -1$, a zero dimensional spin set! In fact ∇_K has no direct interpretation in this series of models, but there is a Yang-Baxter model for it via a different R -matrix. We shall take up this matter in the next section.

To conclude this section, I will show that $\langle K \rangle$ really needed an equally spaced spin set in the form $\mathcal{I} = \{-a, -a + \Delta, -a + 2\Delta, \dots, a - \Delta, a\}$ in order to satisfy $\langle \cancel{\sigma}^{\rightarrow} \rangle = \alpha \langle \sigma^{\rightarrow} \rangle$ for some α . Thus the demand for multiplicativity under the type I move focuses these models into the specific sequence of regular isotopy invariants that we have just elucidated.

Proposition 11.5. Let $\langle K \rangle$ denote the state model of this section, with independent variables z, δ and arbitrary ordered index set $\mathcal{I} = \{N_0 < N_1 < \dots < N_m\}$. Suppose that $\langle K \rangle$ is simple in the sense that there exists an algebraic element α such that

$$\begin{aligned} \langle \cancel{\sigma}^{\rightarrow} \rangle &= \alpha \langle \sigma^{\rightarrow} \rangle \quad \text{and} \\ \langle \cancel{\sigma}^{\leftarrow} \rangle &= \alpha^{-1} \langle \sigma^{\leftarrow} \rangle. \end{aligned}$$

Then all the gaps $N_{k+1} - N_k$ are equal and $N_0 = -N_m$.

Proof.

$$\begin{aligned}
 \langle \text{Diagram} \rangle &= z \left\langle \text{Diagram} \right\rangle + q \left\langle \text{Diagram} \right\rangle \\
 &= z \sum_{a>b} \delta^{-b} \left\langle \text{Diagram} \right\rangle + q \sum_a \delta^{-a} \left\langle \text{Diagram} \right\rangle \\
 &= \sum_a \left(\sum_{a>b} z \delta^{-b} + q \delta^{-a} \right) \left\langle \text{Diagram} \right\rangle.
 \end{aligned}$$

In order for the model to be simple, we need that all of these coefficients are equal:

$$\begin{aligned}
 \alpha &= q \delta^{-N_0} \\
 &= (q - q^{-1}) \delta^{-N_0} + q \delta^{-N_1} \\
 &= \dots \\
 &= (q - q^{-1}) \sum_{i=0}^{m-1} \delta^{-N_i} + q \delta^{-N_m}.
 \end{aligned}$$

Thus

$$\begin{aligned}
 q \delta^{-N_0} &= q \delta^{-N_0} - q^{-1} \delta^{-N_0} + q \delta^{-N_1} \\
 \Rightarrow q^{-1} \delta^{-N_0} &= q \delta^{-N_1} \\
 \Rightarrow q^2 &= \delta^{N_1 - N_0}.
 \end{aligned}$$

Similarly, $q^2 = \delta^{N_{k+1} - N_k}$, $k = 0, \dots, m-1$. This completes the proof. //

12⁰. The Alexander Polynomial.

The main purpose of this section is to show that a Yang-Baxter model essentially similar to the one discussed in section 11⁰ can produce the Alexander-Conway polynomial.

Along with this construction a number of interesting relationships emerge - both about solutions to the Yang-Baxter Equation and about the classical viewpoints for the Alexander polynomial.

The model discussed in this section is the author's re-working [LK23] of a model for the Alexander polynomial, discovered by Francois Jaeger [JA4] (in ice-model language). Other people (Wadati [AW3], H. C. Lee [LEE1]) have seen the relevance of the R -matrix we use below to the Alexander polynomial. In [LK18] and [JA8] it is shown how this approach is related to free-fermion models in statistical mechanics. (See also the exercise at the end of section 13⁰.) Also, see [MUK2] for a related treatment of the multi-variable Alexander polynomial. The R -matrix itself first appeared in [KUS], [PS].

The first state model for a knot polynomial was the author's FKT model [LK2] for the Alexander-Conway polynomial. The FKT-model gives a (non Yang-Baxter) state model for the (multi-variable) Alexander-Conway polynomial. Its structure goes back to Alexander's original definition of the polynomial. It came as something of a surprise to realize that there was a simple Yang-Baxter model for the Alexander polynomial, and that this model was also related to the classical topology. This connection will be drawn in section 13⁰.

I shall begin by giving the model and stating the corresponding Yang-Baxter solution. We shall then backtrack across a number of the related topics.

The Yang-Baxter Model.

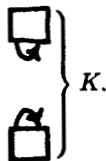
Recall that the Conway version of the one-variable Alexander polynomial is determined by the properties

$$(i) \quad \nabla_K = \nabla_{K'} \text{ whenever } K \text{ is ambient isotopic to } K'.$$

$$(ii) \quad \nabla_O = 1$$

$$(iii) \quad \nabla_{\text{}} \begin{array}{c} \nearrow \\ \searrow \end{array} - \nabla_{\text{}} \begin{array}{c} \searrow \\ \nearrow \end{array} = z \nabla_{\text{}} \begin{array}{c} \leftrightarrow \\ \leftrightarrow \end{array}$$

One important consequence of these axioms is that ∇_K vanishes on split links K . For suppose K is split. Then it can be represented diagrammatically as



(Take this as a definition of **split**.) Then we have

$$\nabla_A - \nabla_B = z\nabla_K$$

where

$$A = \left\{ \begin{array}{c} \text{square with arrow} \\ \diagdown \quad \diagup \\ \text{square with arrow} \end{array} \right\} \quad \text{and} \quad \left\{ \begin{array}{c} \text{square with arrow} \\ \diagup \quad \diagdown \\ \text{square with arrow} \end{array} \right\} = B.$$

Since A and B are ambient isotopic (via a 2π twist), we conclude that $z\nabla_K = 0$, whence $\nabla_K = 0$.

This vanishing property seems to raise a problem for state models of the form

$$\langle K \rangle = \sum_{\sigma} \langle K | \sigma \rangle \delta^{\parallel \sigma \parallel}$$

since these models have the property that $\langle \bullet K \rangle = \langle \bullet \rangle \langle K \rangle$ and we would need therefore that $\langle \bullet \rangle = 0$ for this equation to hold - sending the whole model to zero!

In principle, there is a way out (and the idea works for Alexander - as we shall see). We go back to the idea of knot theory as the study of knots (links) on a string with end-points:



One is allowed ambient or regular isotopy **keeping the endpoints fixed** and so that no movement is allowed past the endpoints. More precisely, the diagrams all live in strip $\mathbf{R} \times I$ for a finite interval I . A link L is represented as a single-input,

single-output tangle with one end attached to $\mathbf{R} \times 0$, and the other end attached to $\mathbf{R} \times 1$. Only the end-points of L touch $(\mathbf{R} \times 0) \cup (\mathbf{R} \times 1)$. All Reidemeister moves are performed inside $\mathbf{R} \times I$ (i.e. in the interior of $\mathbf{R} \times I$).

This two-strand tangle theory of knots and links is equivalent to the usual theory (exercise) - and I shall use it to create this model. For we can now imagine a model where

$$\langle \text{---} \curvearrowright \text{---} \rangle = 1$$

but

$$\langle \text{---} \circlearrowleft \text{---} \rangle = 0 \text{ and } \langle \text{---} \circlearrowright \text{---} \rangle = 0.$$

The split link property is possible in this context.

[An **n -strand tangle** is a link diagram with n free ends; usually the free ends are configured on the outside of a **tangle-box** (a rectangle in the plane) with the rest of the diagram confined to the interior of the box.]

To produce the model explicitly, we need a Yang-Baxter solution. Deus ex machina, here it is:

Let $\mathcal{I} = \{-1, +1\}$ be the index set. Let

$$R = (q - q^{-1}) \begin{array}{c} \text{---} \\ \text{---} \end{array} \quad \bar{R} = (q^{-1} - q) \begin{array}{c} \text{---} \\ \text{---} \end{array}$$

Here we use the same shorthand as in section 11⁰. Thus

$$\begin{array}{c} \text{---} \\ \text{---} \end{array} + \begin{array}{c} \text{---} \\ \text{---} \end{array} = \begin{cases} 1 & \text{if } a = c = +1 \text{ and} \\ & b = d = -1 \\ 0 & \text{otherwise.} \end{cases}$$

Note that

$$\begin{aligned} R - \bar{R} &= (q - q^{-1}) \left[\begin{array}{c} \text{---} \\ \text{---} \end{array} + \begin{array}{c} \text{---} \\ \text{---} \end{array} \right] \left[\begin{array}{c} \text{---} \\ \text{---} \end{array} + \begin{array}{c} \text{---} \\ \text{---} \end{array} \right] \\ &= (q - q^{-1}) \left[\begin{array}{c} \text{---} \\ \text{---} \end{array} \right]. \end{aligned}$$

Thus this tensor is a candidate for a model satisfying the identity

$$\nabla \begin{array}{c} \nearrow \\ \times \\ \searrow \end{array} - \nabla \begin{array}{c} \nearrow \\ \times \\ \searrow \end{array} = z \nabla \begin{array}{c} \downarrow \\ \downarrow \end{array}$$

if we associate R to positive crossings, and \bar{R} to negative crossings. That R and \bar{R} are indeed solutions to the Yang-Baxter Equation, and inverses of each other can be checked directly. [See also the Appendix where we derive this solution in the course of finding all spin preserving solutions to YBE for $\mathcal{I} = \{-1, +1\}$].

Therefore, suppose that we do set up a model

$$\langle K \rangle = \sum_{\sigma} \langle K | \sigma \rangle \delta^{\|\sigma\|}$$

using this R -matrix and $\mathcal{I} = \{-1, +1\}$ with the norms $\|\sigma\|$ calculated by labelling a rotation exactly as in section 11⁰. We want $\langle \curvearrowleft \curvearrowright \rangle = \langle \curvearrowright \curvearrowleft \rangle$, but a first pass is to try for simplicity: $\langle \curvearrowleft \curvearrowright \rangle = \alpha \langle \curvearrowright \curvearrowleft \rangle$ and see if that restriction determines the structure of the model. Therefore, we must calculate:

$$\begin{aligned} \langle \curvearrowleft \curvearrowright \rangle &= (q - q^{-1}) \left\langle \begin{array}{c} \curvearrowleft \\ \circlearrowleft \\ \curvearrowright \end{array} \right\rangle + q \left\langle \begin{array}{c} \curvearrowleft \\ \circlearrowright \\ \curvearrowright \end{array} \right\rangle - q^{-1} \left\langle \begin{array}{c} \curvearrowleft \\ \circlearrowleft \\ \curvearrowleft \end{array} \right\rangle \\ &= (q - q^{-1}) \delta^{-1} \left\langle \begin{array}{c} \curvearrowleft \\ \curvearrowleft \\ \curvearrowright \end{array} \right\rangle + q \delta^{-1} \left\langle \begin{array}{c} \curvearrowleft \\ \circlearrowright \\ \curvearrowright \end{array} \right\rangle - q^{-1} \delta \left\langle \begin{array}{c} \curvearrowleft \\ \circlearrowleft \\ \curvearrowleft \end{array} \right\rangle \\ &= [(q - q^{-1}) \delta^{-1} - q^{-1} \delta] \left\langle \begin{array}{c} \curvearrowleft \\ \curvearrowleft \\ \curvearrowright \end{array} \right\rangle + q \delta^{-1} \left\langle \begin{array}{c} \curvearrowleft \\ \circlearrowright \\ \curvearrowright \end{array} \right\rangle. \end{aligned}$$

Thus, we demand that

$$(q - q^{-1}) \delta^{-1} - q^{-1} \delta = q \delta^{-1}.$$

Whence

$$q^{-1}(\delta^{-1} + \delta) = 0.$$

Therefore $\delta^{-1} + \delta = 0$, and we shall take $\delta = i$ where $i^2 = -1$.

It is now easy to verify that the model is multiplicative for all choices of curl. In fact, it also does not depend upon the crossing type:

Lemma 12.1. With $\delta = i$ ($i^2 = -1$) and $\langle K \rangle$ as described above, we have

$$(i) \quad \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right\rangle = -iq \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right\rangle$$

$$(ii) \quad \left\langle \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right\rangle = -iq \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right\rangle$$

$$(iii) \quad \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right\rangle = iq^{-1} \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right\rangle$$

$$(iv) \quad \left\langle \begin{array}{c} \curvearrowleft \\ \curvearrowleft \end{array} \right\rangle = iq^{-1} \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right\rangle$$

Proof. We verify (iii) and leave the rest for the reader.

$$\begin{aligned} \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right\rangle &= (q^{-1} - q) \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right\rangle + q^{-1} \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right\rangle - q \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right\rangle \\ &= ((q^{-1} - q)i - qi^{-1}) \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right\rangle + q^{-1}i \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right\rangle \\ &= q^{-1}i \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right\rangle. \end{aligned} \quad //$$

The lemma implies that if

$$\nabla_K = (iq^{-1})^{-\text{rot}(k)} \langle K \rangle,$$

then $\nabla_{\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array}} = \nabla_{\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array}}$ for a curl. Certainly $\nabla_{\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array}} - \nabla_{\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array}} = (q - q^{-1})\nabla_{\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array}}$. It remains to verify that ∇_K is invariant under the type IIB move. Once we verify this, ∇_K becomes an invariant of regular isotopy and hence ambient isotopy - via the curl invariance.

Note that we shall assign rotation number zero to the bare string:

$$\text{rot}(\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array}) = 0.$$

Also, we assume that $\left\langle \begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right\rangle + \left\langle \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right\rangle = \left\langle \begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right\rangle = 1$. The model now has the form

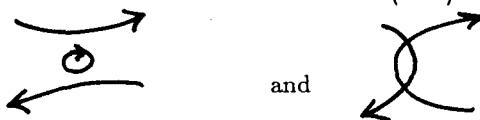
$$\langle K \rangle = \sum_{\sigma} \langle K | \sigma \rangle i^{\|\sigma\|}$$

where $i^2 = -1$, $\langle K | \sigma \rangle$ is the product of vertex weights determined by the R -matrix, and $\|\sigma\|$ is the rotational norm for index set $\mathcal{I} = \{-1, +1\}$.

Proposition 12.2. With $\langle K \rangle$ defined as above,

$$\left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle = \left\langle \begin{array}{c} \nearrow \\ \swarrow \end{array} \right\rangle.$$

Proof. Consider the terms in the expansion of $\left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle$. The contributing forms are



Hence

$$\begin{aligned} \left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle &= qq^{-1} \left\langle \begin{array}{c} \nearrow \\ \searrow \\ + \\ Q \\ + \end{array} \right\rangle + (-q^{-1})(-q) \left\langle \begin{array}{c} \nearrow \\ \searrow \\ - \\ Q \\ - \end{array} \right\rangle \\ &\quad + zq^{-1} \left\langle \begin{array}{c} \nearrow \\ \searrow \\ - \\ Q \\ + \end{array} \right\rangle + (-q^{-1})(-z) \left\langle \begin{array}{c} \nearrow \\ \searrow \\ - \\ Q \\ + \end{array} \right\rangle \\ &\quad + \left\langle \begin{array}{c} \nearrow \\ \searrow \\ - \\ Q \\ + \end{array} \right\rangle + \left\langle \begin{array}{c} \nearrow \\ \searrow \\ + \\ Q \\ - \end{array} \right\rangle \\ &= i^{-1} \left\langle \begin{array}{c} \nearrow \\ \searrow \\ + \\ Q \\ - \end{array} \right\rangle + i \left\langle \begin{array}{c} \nearrow \\ \searrow \\ - \\ Q \\ + \end{array} \right\rangle + \left\langle \begin{array}{c} \nearrow \\ \searrow \\ + \\ Q \\ + \end{array} \right\rangle \\ &= i^{-1} \left\langle \begin{array}{c} \nearrow \\ \searrow \\ + \\ Q \\ - \end{array} \right\rangle + i \left\langle \begin{array}{c} \nearrow \\ \searrow \\ - \\ Q \\ + \end{array} \right\rangle \\ &\quad - \left\langle \begin{array}{c} \nearrow \\ \searrow \\ + \\ Q \\ + \end{array} \right\rangle - \left\langle \begin{array}{c} \nearrow \\ \searrow \\ - \\ Q \\ + \end{array} \right\rangle \\ &\quad + \left\langle \begin{array}{c} \nearrow \\ \swarrow \end{array} \right\rangle. \end{aligned}$$

The last step follows from the identity

$$\left\langle \begin{array}{c} \nearrow \\ \searrow \\ + \\ Q \\ + \end{array} \right\rangle = \left\langle \begin{array}{c} \nearrow \\ \swarrow \end{array} \right\rangle - \left\langle \begin{array}{c} \nearrow \\ \swarrow \\ - \\ Q \\ + \end{array} \right\rangle.$$

I now assert the following two identities:

$$i^{-1} \left\langle \begin{array}{c} + \\ \overleftarrow{\longleftrightarrow} \\ + \end{array} \right\rangle - \left\langle \begin{array}{c} + \\ \downarrow \\ + \end{array} \right\rangle = 0$$

and

$$i \left\langle \begin{array}{c} \overleftarrow{\longleftrightarrow} \\ \overleftarrow{\leftarrow} \\ + \end{array} \right\rangle - \left\langle \begin{array}{c} + \\ \downarrow \\ + \end{array} \right\rangle = 0.$$

[These are trivial to verify by checking cases. Thus

$$\left\langle \begin{array}{c} Q^+ \\ O^+ \end{array} \right\rangle = i^2 = -1$$

$$\left\langle \begin{array}{c} + \\ \text{S} \\ + \end{array} \right\rangle = i$$

$$\Rightarrow i^{-1} \left\langle \begin{array}{c} O^+ \\ O^+ \end{array} \right\rangle - \left\langle \begin{array}{c} + \\ \text{S} \\ + \end{array} \right\rangle = -i^{-1} - i = 0$$

and

$$\left\langle \begin{array}{c} \overrightarrow{\longleftrightarrow} \\ \text{O}^+ \end{array} \right\rangle = i \left\langle \begin{array}{c} \overrightarrow{\longleftrightarrow} \\ + \end{array} \right\rangle$$

$$\left\langle \begin{array}{c} \text{S} \\ + \end{array} \right\rangle = \left\langle \begin{array}{c} \overrightarrow{\longleftrightarrow} \\ + \end{array} \right\rangle$$

$$\Rightarrow i^{-1} \left\langle \begin{array}{c} \overrightarrow{\longleftrightarrow} \\ \text{O}^+ \end{array} \right\rangle - \left\langle \begin{array}{c} + \\ \text{S} \\ + \end{array} \right\rangle = 0.]$$

Note that we must consider separately those cases when one of the local arcs is part of the input/output string.

As a result of these last two identities, we conclude that

$$\left\langle \begin{array}{c} \overleftarrow{\text{D}} \\ \text{C} \end{array} \right\rangle = \left\langle \begin{array}{c} \text{D} \\ \downarrow \\ \text{C} \end{array} \right\rangle.$$

This completes the proof. //

We have now completed the proof that $\langle K \rangle = \sum_{\sigma} \langle K | \sigma \rangle i^{\parallel \sigma \parallel}$ is a regular isotopy invariant of single input/output tangles. It follows that $\nabla_K = (iq^{-1})^{-\text{rot } K} \langle K \rangle$ is an invariant of ambient isotopy for these tangles, and that

1. $\nabla_{\text{tangle}} - \nabla_{\text{tangle}} = z \nabla_{\text{tangle}}$
2. $\nabla_{\text{tangle}} = \nabla_{\text{tangle}} + \nabla_{\text{tangle}} = 1$
- $\nabla_{\text{circle}} = 0$.

Having verified these axioms, we see that $\nabla_K = \nabla_K^+ \nabla_{\text{tangle}} + \nabla_K^- \nabla_{\text{tangle}}$ being free (by using 1. and 2. recursively) of ∇_{tangle} or ∇_{tangle} , it follows that $\nabla_K^+ = \nabla_K^- = \nabla_K$. In other words, the coefficients of $\langle \text{tangle} \rangle$ and $\langle \text{tangle} \rangle$ in the state expansion of $\langle K \rangle$ are identical. For calculation this means that we can assume that all states have (say) positive input and positive output:



Since we already know that these axioms will compute the Conway polynomial of \bar{K} where \bar{K} is the closure of the tangle K :



we conclude that ∇_K is indeed the Conway polynomial.

Finally, we remark that ∇_K will give the same answer from a knot \bar{K} for any choice of associating a tangle K to \bar{K} by dropping a segment. I assume this segment is dropped to produce a planar tangle in the form , fitting our constructions. The independence of segment removal again follows from the corresponding skein calculation using properties 1. and 2.

In fact, the model's independence of this choice can be verified directly and combinatorially in the form of the state summation. I leave this as an exercise for the reader.

Example.

	$\leftrightarrow i$	$\langle \text{---} + \rangle = 1$
	$\leftrightarrow q^2 i^{-1}$	$\langle \text{---} - \rangle = 0$
	$\leftrightarrow 0$	



$$\langle L \rangle = i + q^2 i^{-1}$$

$$\text{rot}(L) = -1$$

$$\begin{aligned}\nabla_L &= (iq^{-1})^{\text{rot}(L)} \langle L \rangle \\ &= iq^{-1}(i + q^2 i^{-1}) \\ &= -q^{-1} + q\end{aligned}$$

$$\nabla_L = z.$$

There are many mysteries about this model for the Conway polynomial. Not the least of these is the genesis of its R -matrix. In the Appendix we classify all spin preserving solutions of the Yang-Baxter Equation for the spin set $\mathcal{I} = \{-1, +1\}$. Such solutions must take the form

$$\begin{aligned}R = \ell &\left[\begin{array}{c} + \\ \downarrow \end{array} \right] \left[\begin{array}{c} - \\ \downarrow \end{array} \right] \left[\begin{array}{c} + \\ \downarrow \end{array} \right] \left[\begin{array}{c} + \\ \downarrow \end{array} \right] \left[\begin{array}{c} + \\ \downarrow \end{array} \right] \left[\begin{array}{c} - \\ \downarrow \end{array} \right] \\ &+ d \begin{array}{c} \diagup \\ \diagdown \end{array} + s \begin{array}{c} \diagdown \\ \diagup \end{array}\end{aligned}$$

for some coefficients ℓ, r, p, n, d , and s . We show in the appendix that the following relations are necessary and sufficient for R to satisfy the YBE:

$r\ell d = 0$
$r\ell s = 0$
$r\ell(\ell - r) = 0$
$p^2\ell = p\ell^2 + \ell ds$
$n^2\ell = n\ell^2 + \ell ds$
$p^2r = pr^2 + rds$
$n^2r = nr^2 + rds$

One natural class of solutions arises from $d = s = 1, r = 0$. Then the conditions reduce to:

$$p^2\ell = p\ell^2 + \ell$$

$$n^2\ell = n\ell^2 + \ell.$$

Assuming n and p invertible, we get $\ell = p - p^{-1}$ and $\ell = n - n^{-1}$. Thus $p - p^{-1} = n - n^{-1}$.

$$\Leftrightarrow np^2 - n = n^2p - p$$

$$\Leftrightarrow \boxed{np(p-n) = n-p}.$$

Thus we see that there are two basic solutions:

1. $n = p$
2. $np = -1$.

The first gives the R -matrix

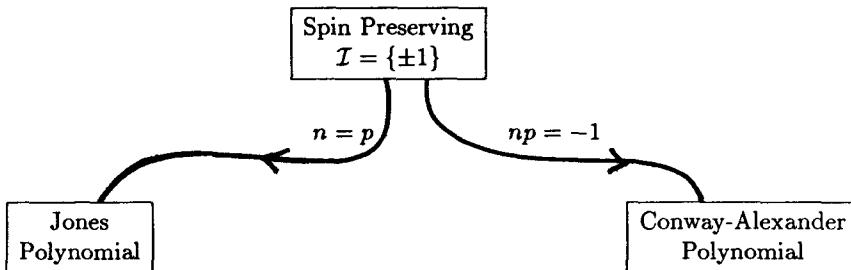
$$R = (p - p^{-1}) \begin{pmatrix} + \\ - \end{pmatrix} \begin{pmatrix} + \\ - \end{pmatrix} = \begin{pmatrix} + \\ - \end{pmatrix} + \begin{pmatrix} * \\ \times \end{pmatrix}$$

and gives the Jones polynomial as in section 11⁰.

The second gives the R -matrix

$$R = (p - p^{-1}) \begin{pmatrix} + \\ - \end{pmatrix} = \begin{pmatrix} + \\ - \end{pmatrix} + \begin{pmatrix} * \\ \times \end{pmatrix}$$

and yields the Conway-Alexander polynomial as we have seen in this chapter. Thus from the viewpoint of the YBE we have the diagram



In this sense the Jones and Conway polynomials have equal footing. But in another sense they are very different and the difference is apparent in the state models.

Let $J(K)$ denote the oriented state model for the Jones polynomial as constructed in section 11⁰. Then

$$J(K) = \sum_{\sigma} J(K|\sigma) t^{||\sigma||}$$

where t is an algebraic variable. Let $A\langle K \rangle$ denote the model in this chapter. Then

$$A\langle K \rangle = \sum_{\sigma} A\langle K | \sigma \rangle i^{\|\sigma\|}$$

where $i^2 = -1$.

That the loop term is numerical ($i^{\|\sigma\|}$) for the Conway polynomial, and algebraic ($t^{\|\sigma\|}$) for the Jones polynomial, is the first visible difference between these models.

There are many results known about the Alexander polynomial via its classical definitions. The classical Alexander polynomial is denoted $\Delta_K(t)$ and it is related to the Conway version as $\Delta_K(t) \doteq \nabla_K(t^{1/2} - t^{-1/2})$ where \doteq denotes equality up to sign and powers of t . Since the model $A\langle K \rangle$ of this section is expressed via $z = q - q^{-1}$ we have that $\Delta_K(q^2) \doteq A\langle K \rangle(q)$. One problem about this model is particularly interesting to me. We know [FOX1] that if K is a **ribbon knot**, then $\Delta_K(t) \doteq f(t)f(t^{-1})$ where $f(t)$ is a polynomial in t . Is there a proof of this fact using the state model of this chapter? A ribbon knot is a knot that bounds a disk immersed in three space with only **ribbon singularities**. In a ribbon singularity the disk intersects itself transversely, matching two arcs. One arc is an arc interior to the disk; the other goes from one boundary point to another.

In the next section we review some classical knot theory and show how this model of the Alexander-Conway polynomial is related to the Burau representation of the Artin braid group.

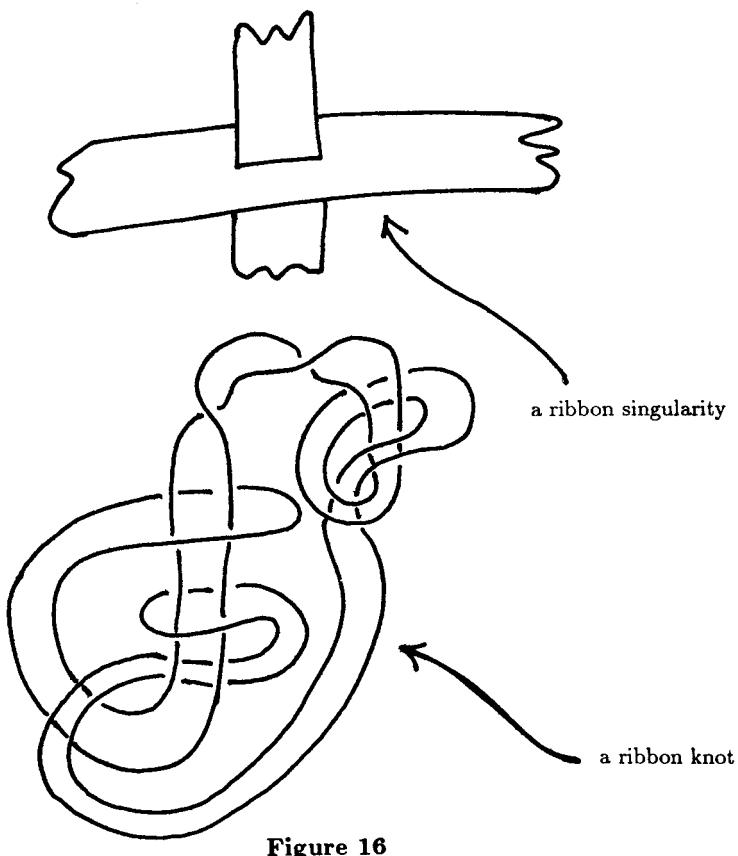
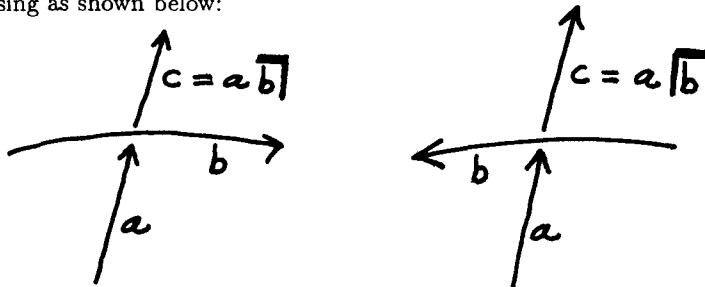


Figure 16

13⁰. Knot-Crystals - Classical Knot Theory in Modern Guise.

This lecture is a short course in classical knot theory - including the fundamental group of the knot (complement), and the classical approach to the Alexander polynomial. I say modern guise, because we shall begin with a construction, the **crystal**, $C(K)$, that generalizes the fundamental group **and** the quandle. The quandle [JOY] is a recent (1979) invariant algebra that may be associated with the diagram of a knot or link. David Joyce [JOY] proved that the quandle classifies the **unoriented type** of the knot. (Two knots are unoriented equivalent if there is a homeomorphism of S^3 - possibly orientation reversing - that carries one knot to the other. Diagrammatically, two diagrams K and L are unoriented equivalent if K is ambient isotopic to either L or L^* , where L^* denotes the mirror image of L .)

Before giving the definition of the crystal, I need to introduce some algebraic notation. In particular, I shall introduce an operator notation, $a\blacksquare$ (read “a cross”), and explain how to use it to replace non-associative formalisms by non-commutative operator formalisms. Strictly speaking, the cross illustrated above (\blacksquare) is a **right cross**. We shall also have a **left cross** (\blacksquare^l : as in $\blacksquare^l a$) and the two modes of crossing will serve to encode the two oriented types of diagrammatic crossing as shown below:

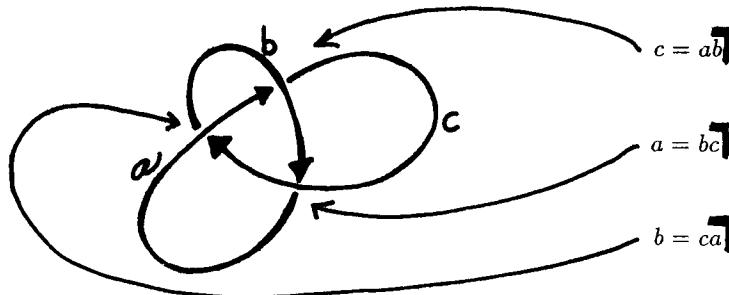


In this formalism, each arc in the oriented link diagram K is labelled with a letter (a, b, c, \dots). In this mode I shall refer to the arc as an **operand**. The arc can also assume the **operator mode**, and will be labelled by a left or right cross ($a\blacksquare$ or $\blacksquare^l a$) according to the context (left or right handed crossing).

As illustrated above, we regard the arc c emanating from an undercrossing as the result of the overcrossing arc b , acting (\blacksquare or \blacksquare^l) on the incoming undercrossing

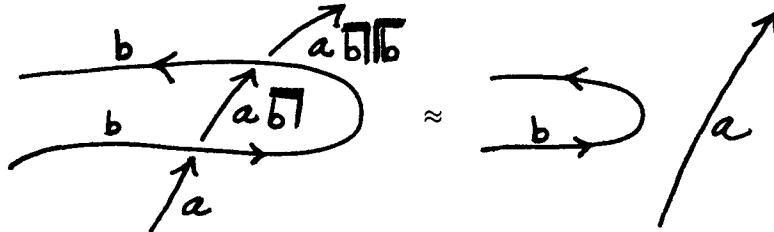
arc a . Thus we write $c = ab\overline{1}$ or $c = \overline{a}b\overline{1}$ according to whether the crossing is of right or left handed type.

At this stage, these notations do not yet abide in an algebraic system. Rather, they constitute a **code** for the knot. For example, in the case of the trefoil, we obtain three code-equations as shown below.



What is the simplest algebra compatible with this notation and giving a regular isotopy invariant of oriented knots and links?

In order to answer this question, let's examine how this formalism behaves in respect to the Reidemeister moves. First, consider the type II move:



From this we see that type II invariance corresponds to the rules

$$ab\overline{1}b = a \quad \text{and} \quad a\overline{b}\overline{1}b = a$$

for any a and b . We also see that in order to form an algebra it will be necessary to be able to multiply x and $y\overline{1}$ to form $xy\overline{1}$ and to multiply x and $y\overline{1}$ to form $x\overline{y}\overline{1}$ for any two elements x and y in the algebra.

Example.



The stipulation that we can form xy for any two elements x and y means that xy does not necessarily have a direct geometric interpretation on the given diagram. Note also that the multiplication, indicated by juxtaposition of symbols, is assumed to be **associative and non-commutative**.

Since the knot diagrams do not ever ask us to form products of the form xy where x and y are uncrossed, we shall not assume that such products live in the crystal. However, since products of the form $a_0 b_1 \overline{b}_2 \overline{b}_3 \dots \overline{b}_k$ occur naturally with a_0 uncrossed, it **does** make sense to consider isolated products of crossed elements such as

$$a \overline{b} c \overline{d}$$

Call these **operator products**. Let 1 denote the empty word. The subalgebra of the crystal $C(K)$ generated by operator products will be called the **operator algebra** $\pi(K) \subset C(K)$.

We write the crystal as a union of $\pi(K)$ and $C_0(K)$, the **primary crystal**. The primary crystal consists of all elements of the form $a\beta$ where a is uncrossed and β belongs to the operator algebra. Thus the operator algebra acts on the primary crystal by right multiplication, and the primary crystal maps to the operator algebra via crossing.

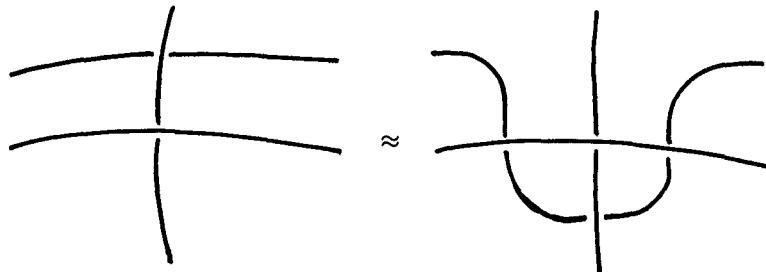
Since we want $ab\overline{b} = a$ for all a and b in the crystal, we shall take as the first axiom (labelled II. for obvious reasons) that

$$\text{II. } b \overline{b} = \overline{b} b = 1$$

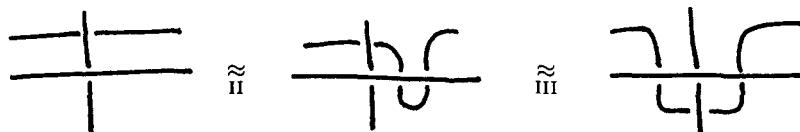
for all $b \in C(K)$. Note that this equality refers to $\pi(K)$, and that therefore $\pi(K)$ is a group. We shall see that, with the addition of one more axiom, $\pi(K)$ is isomorphic with the classical fundamental group of the link K .

In order to add the remaining axiom, we must analyze the behavior under the

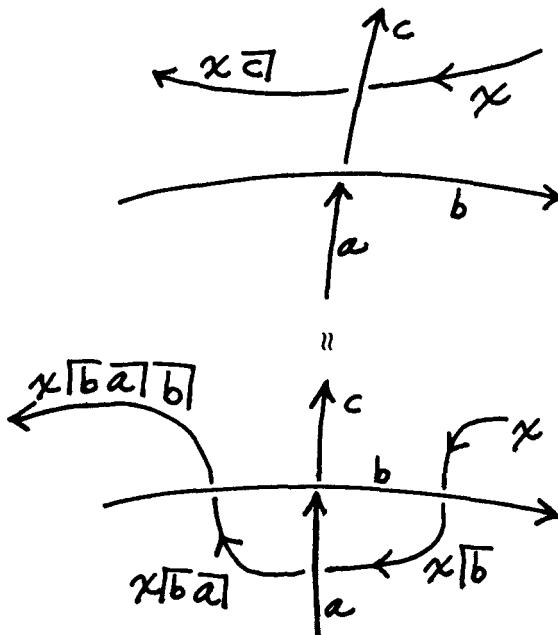
type III move. I shall use the following move, dubbed the detour:



It is easy to see that, in the presence of the type II move, the detour is equivalent to the type III move:



Now examine the algebra of the detour:



Thus, we see that the axiom should be: When $c = ab\bar{b}$, then $\bar{c} = \bar{b} \bar{a} \bar{b}$.

Or, more succinctly:

$$D. \quad ab = [b \ a \ b] \quad (\text{D for detour})$$

There are obvious variants of this rule, depending upon orientations. Thus

$$\begin{aligned}\boxed{a b} &= \boxed{b a} \\ \boxed{a \boxed{b}} &= \boxed{b} \boxed{a} \\ \boxed{\boxed{a} b} &= \boxed{b} \boxed{\boxed{a} b}\end{aligned}$$

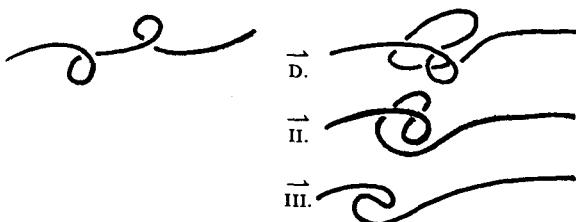
From the point of view of crossing, the detour axiom is a depth reduction rule. It takes an expression of (nesting) depth two ($\overline{a} \overline{b}$) and replaces it by an expression of depth one ($\overline{b} \overline{a} \overline{b}$). Note the pattern: The crossed expression (\overline{b}) inside is duplicated on the right in the same form as its inside appearance, and duplicated on the left in reversed form. The crossed expression (\overline{b}) is removed from the middle, leaving the a with a cross standing over it whose direction corresponds to the original outer cross on the whole expression.

These rules, $a \mid a = 1$, $a \mid b = b \mid a$, capture subtle aspects of regular homotopy in algebraic form. For example:



$$\left\{ \begin{array}{l} a \overline{a} \overline{\overline{a} \overline{a}} = a \overline{a} \overline{\overline{a} \overline{a} \overline{a}} \\ \text{D.} \\ \qquad \qquad \qquad = a \overline{a} \overline{a} \\ \text{II.} \\ \qquad \qquad \qquad = a. \\ \text{II.} \end{array} \right\}$$

This algebraic equivalence is exactly matched by the following performance of the Whitney trick, using the detour:



Another look at the algebraic version of the detour shows its relationship with the fundamental group. For in the Wirtinger presentation of $\pi_1(S^3 - K)$ [ST] there is one relation for each crossing and one generator for each arc. The relation is $\gamma = \beta^{-1} \alpha \beta$ when the crossing is positive as shown below:

$$\gamma = \beta^{-1} \alpha \beta.$$

If we are working in the crystal with crossing labels as shown below

$$\overbrace{\quad}^{\alpha} \overbrace{\quad}^{\beta} \xrightarrow{\gamma} \left\{ \gamma = \beta^{-1} \alpha \beta, \right.$$

then we have the relation on operators

$$c = [b \ a] b,$$

This is recognizable as

$$\gamma = \beta^{-1} \alpha \beta$$

with $\gamma = c$, $\alpha = a$, $\beta = b$. This is the standard relation at a crossing in the Wirtinger presentation of the fundamental group of the link complement. Thus the **group of operators**, $\pi(K)$, is isomorphic to the fundamental group of the link complement.

It should be clear by now how to make a **formal definition of a crystal**. We have the notations a and $[a]$ for maps $f, g : C_0 \rightarrow \pi$. $f(x) = a$, $g(x) = [a]$. π is a group with binary operation denoted $x, y \mapsto xy$. π acts on C_0 on the right via

$$C_0 \times \pi \rightarrow C_0$$

$$a, \lambda \mapsto a\lambda$$

such that $(a\lambda)\mu = a(\lambda\mu)$. And it is given that

II. $\overline{x} \overline{|} \overline{x} = 1 \in \pi$ for all $x \in C_0 \cup \pi$.

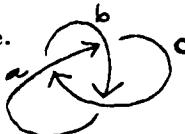
D. $\overline{x} \overline{|} \overline{y} = \overline{y} \overline{|} \overline{x}$ (and the other three variants of this equation) for all $x, y \in C_0 \cup \pi$.

Definition 13.1. A **crystal** is a set $C = C_0 \cup \pi$ with maps $\overline{\lrcorner}, \overline{\lrcorner} : C_0 \rightarrow \pi$ such that C_0 and π are disjoint, π is a group acting on C_0 , and the maps from C_0 to π satisfy the conditions II. and D. above. Note that D. (above) becomes $\overline{a\alpha} = \alpha^{-1}\overline{a}\alpha$ and $\overline{aa} = \alpha^{-1}\overline{a}\alpha$ for $\alpha \in \pi, a \in C_0$.

Definition 13.2. Two crystals C and C' are **isomorphic** if there is a 1-1 onto map $\Phi : C \rightarrow C'$ such that $\Phi(C_0) = C'_0$, $\Phi(\pi) = \pi'$ and

1. $\overline{\Phi(x)} = \Phi(\overline{x})$
 $\overline{\Phi(x)} = \Phi(\overline{|x})$ for all $x \in C_0$.
2. $\Phi(x\alpha) = \Phi(x)\Phi(\alpha)$ for all $x \in C_0, \alpha \in \pi$
3. $\Phi | \pi$ is an isomorphism of groups.

Definition 13.3. Let K be an oriented link diagram. We associate the **link crystal** $C(K)$, to K by assigning one generator of $C_0(K)$ for each arc in the diagram, and one relation (of the form $c = ab$ or $c = a\overline{b}$) to each crossing. The resulting structure is then made into a crystal in the usual way (of universal algebra) by allowing all products (as specified above) and taking equivalence classes after imposing the axioms.

Example.  $C(K) = (a, b, c \mid c = ab, b = ca, a = bc)$.

Here I use the abbreviation

$$C(K) = (a_1, a_2, \dots, a_n \mid r_1, r_2, \dots, r_m)$$

to denote the free crystal on a_1, \dots, a_n modulo the relations r_1, \dots, r_m .

Because it contains information about how the fundamental group $\pi(K)$ acts on the “peripheral” elements $C_0(K)$, the crystal has more information about the link than the fundamental group alone. In fact, if two knots K and K' have

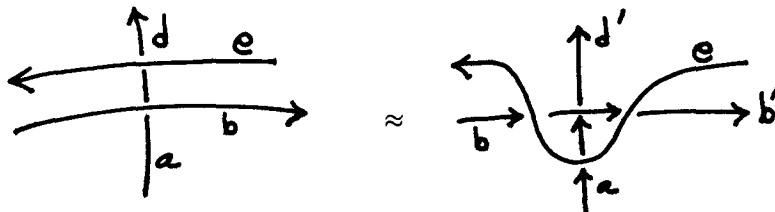
isomorphic crystals, then K is ambient isotopic to either K' or the mirror image K'^* . This follows from our discussion below, relating the crystal to David Joyce's Quandle [JOY]. But before we do this, we will finish the proof of the

Theorem 13.4. Let K and K' be diagrams for two oriented links. If K is regularly isotopic to K' then $C(K)$ is isomorphic to $C(K')$. If K is ambient isotopic with K' , then $\pi(K)$ and $\pi(K')$ are isomorphic.

Proof. Since $a \overline{a} = \overline{a} \overline{a} = \overline{a}$ we see



that $\pi(K)$ is invariant under Reidemeister moves of type I. [$\mathcal{O}(K)$ is not invariant under type I, since we do not insist that aa is equal to a .] It remains to check invariance for the “over” version of the detour.



Referring to the diagram above, we have $d = ab \overline{e} e$ and $d' = a \overline{e} b \overline{e} e$, $b' = b \overline{e} e$. Thus

$$\begin{aligned} d' &= a \overline{e} b \overline{e} e = a \overline{e} e \overline{e} b \overline{e} e \\ &= ab \overline{e} e \\ d' &= d. \end{aligned}$$

This shows the remaining case (modulo a few variations of orientation) for the invariance of $C(K)$ under the moves of type II and III. Hence $C(K)$ is a regular isotopy invariant. //

The Quandle.

Let K be an oriented link, and let $C(K)$ be its crystal. Define binary operations $*$ and $\bar{*}$ on $C_0(K)$ as follows:

$$a * b = ab$$

$$a \bar{*} b = a \overline{b}.$$

Lemma 13.5. $(a * b) \bar{*} b = a$
 $(a * b) * c = (a * c) * (b * c)$

(The same identities hold if $*$ and $\bar{*}$ are interchanged.)

Proof.

$$\begin{aligned}(a * b) \bar{*} b &= ab\overline{\overline{b}} = a \\(a * b) * c &= ab\overline{c} \\(a * c) * (b * c) &= ac\overline{b}\overline{c} \\&= ac\overline{c}\overline{b}\overline{c} \\&= ab\overline{c} \\&= (a * b) * c.\end{aligned}$$

This completes the proof. //

Let $Q(K)$ be the quotient of $(C_0(K), *, \bar{*})$ with the additional axioms $a * a = a$, $a \bar{*} a = a$ for all $a \in C_0(K)$. Thus $Q(K)$ is an algebra with two binary operations that is an **ambient isotopy invariant of K** (we just took care of the type I move by adding the new axioms.)

It is easy to see that $Q(K)$ is the quotient of the set of all formal products of arc labels $\{a_1, \dots, a_n\}$ for K under $*$ and $\bar{*}$ (these are non-associative products) by the **axioms of the quandle [JOY]**.

1. $a * a = a$, $a \bar{*} a = a$ for all a .
2. $(a * b) \bar{*} b = a$
 $(a \bar{*} b) * b = a$ for all a and b .
3. $(a * b) * c = (a * c) * (b * c)$
 $(a \bar{*} b) \bar{*} c = (a \bar{*} c) \bar{*} (b \bar{*} c).$

Thus $Q(K)$ is identical with Joyce's quandle. This shows that the crystal $C(K)$ classifies oriented knots up to mirror images.

The quandle is a non-associative algebra. In the crystal, we have avoided non-associativity via the operator notation:

$$\begin{aligned}(a * b) * c &= ab\overline{c} \\a * (b * c) &= ab\overline{c}\overline{b}\end{aligned}$$

A fully left associated product in the quandle (such as $((a * b) * c) * d$) corresponds to an operator product of depth 1:

$$((a * b) * c) * d = ab\overline{c}\overline{d}$$

while other associations can lead to a nesting of operators such as

$$a * (b * (c * d)) = a \overline{b} \overline{c} \overline{d} \overline{\overline{d}}$$

It might seem that the quandle is very difficult to deal with because of this multiplicity of associated products. However, we see that the second basic crystal axiom ($\overline{a\alpha} = \alpha^{-1}\overline{a}$, $a \in C$, $\alpha \in \pi$) shows at once that

Theorem 13.6. (Winker [WIN]) Any quandle product can be rewritten as a left-associated product.

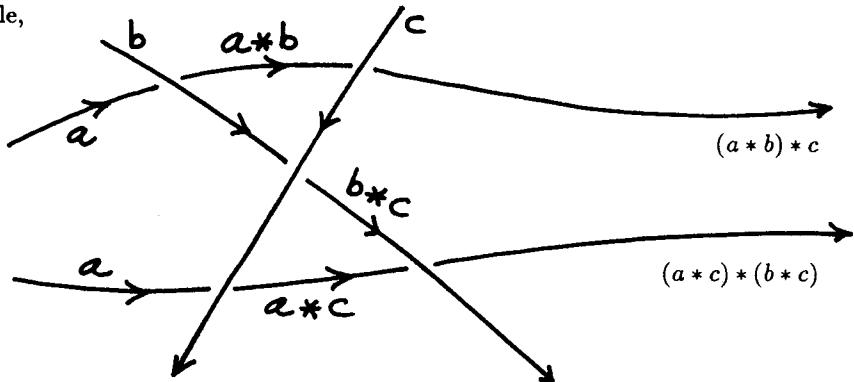
This theorem is the core of Winker's Thesis. It provides canonical representations of quandle elements and permits one to do extensive analysis of the structure of knot and link quandles. Winker did not have the crystal formalism, and so proved this result directly in the quandle. The crystalline proof is given below.

Proof of Winker's Theorem. It is clear that the rules $\overline{a\alpha} = \alpha^{-1}\overline{a}$, α and $\overline{f\alpha} = \alpha^{-1}\overline{f}\alpha$ for $a \in C_0$, $\alpha \in \pi$ give depth reduction rules for any formal product in C_0 . When a product has been reduced to depth 0 or depth 1 then it corresponds to a left-associated product in the quandle notation. This completes the proof.//

Example.

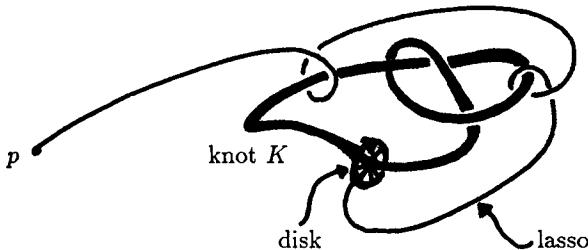
$$\begin{aligned} a * (b * c) &= a \overline{b} \overline{c} \\ &= a \overline{c} \overline{b} \overline{c} \\ &= ((a * c) * b) * c \\ a * (b * (c * d)) &= a \overline{b} \overline{c} \overline{d} \\ &= a \overline{b} \overline{d} \overline{c} \overline{d} \\ &= a \overline{d} \overline{c} \overline{d} \overline{b} \overline{d} \overline{c} \overline{d} \\ &= (((((a * d) * c) * d) * b) * d) * c. \end{aligned}$$

Remark. The quandle axioms are also images of the Reidemeister moves. For example,

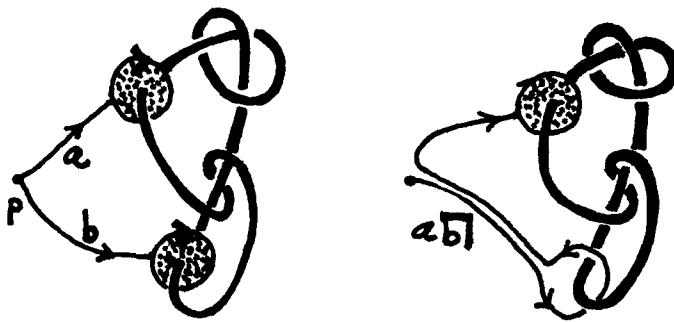


Thus the distributivity $(a * b) * c = (a * c) * (b * c)$ expresses this form of the type III move.

Remark. The quandle has a nice homotopy theoretic interpretation (see [JOY] or [WIN]). The idea is that each element of $Q(K)$ is represented by a lasso consisting of a disk transverse to the knot that is connected by a string attached to the boundary of the disk to the base point p :

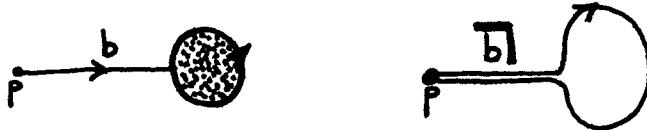


Two lassos are multiplied $(a * b)$ by adding extra string to a via a path along b [Go from base point to the disk, around the disk in the direction of its orientation for $\bar{*}$ (against for $*$) and back to p along this string, then down a 's string to attach the disk!].



$$a * b \equiv ab\overline{1}$$

Note that in this interpretation we can think of $a * b = ab\overline{1}$ where $\overline{1}$ is the element in the fundamental group of $S^3 - K$ corresponding to b via [go down string, around disk and back to base point].



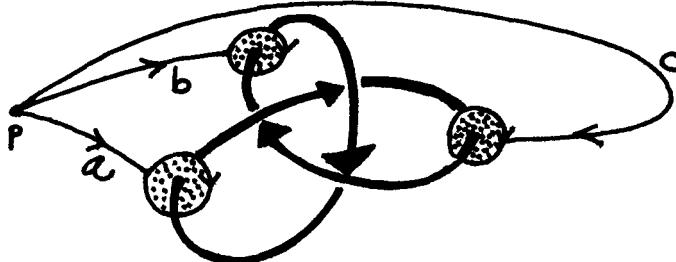
The string of a lasso, b , can entangle the link in a very complex way, but the element $\overline{1}$ (or \overline{b}) in the group π is a conjugate of one of the peripheral generators of π . Such a generator is obtained by going from base point to an arc of the link, encircling the arc once, and returning to base point. For diagrams, we can standardize these peripheral generators by

- 1) choosing a base point in the unbounded region of the diagram,

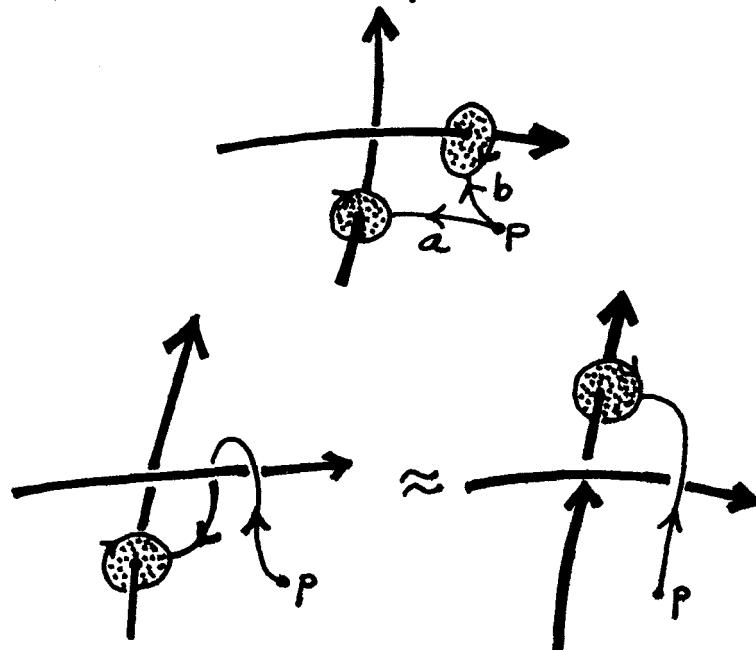
and

- 2) stipulating that the string of the lasso goes over all strands of the diagram that it interacts with.

With these conventions, we get the following generators for the trefoil knot T :



The geometric definition of $a * b = ab\overline{b}$ now has a homotopical interpretation:



We see that the relation at a crossing ($c = a * b$) is actually a description of the existence of a homotopy from the lasso $a * b$ to the lasso c . In terms of the fundamental group, this says that $\overline{a} = \overline{b}^{-1} \overline{a} \overline{b} = \overline{b} \overline{a} \overline{b}$ (as we have seen algebraically in the crystal). Sliding the lasso gives a neat way to picture this relation in the fundamental group.

The homotopy interpretation of the quandle includes the relations $a * a = a$ and $a \bar{*} a = a$ as part of the geometry.

Crystalline Examples.

1. The simplest example of a crystal is obtained by defining $a\bar{b} = 2b - a = a\bar{b}$ where $a, b, \dots \in \mathbb{Z}/m\mathbb{Z}$ for some modulus m . Note that $a\bar{b}\bar{b} = a\bar{b}\bar{b} = a$ and that $x\bar{a}\bar{b} = 2(a\bar{b}) - x = 2(2b - a) - x = 4b - 2a - x = 2b - (2a - (2b - x)) = x\bar{b}\bar{a}\bar{b}$. In this case, the associated group structure π is the additive group of $\mathbb{Z}/m\mathbb{Z}$.

For a given link K , we can ask for the least modulus m ($m \neq 1$) that is compatible with this crystal structure and compatible with crossing relations for the diagram K . Thus [using $\bar{a} = \bar{a} = \bar{a}$ in this system] for the trefoil we have



$$\begin{aligned}c &= a\bar{b} = 2b - a \\b &= c\bar{a} = 2a - c \\a &= b\bar{c} = 2c - b\end{aligned}$$

$$\Rightarrow \begin{aligned}-a + 2b - c &= 0 \\2a - b - c &= 0 \\-a - b + 2c &= 0\end{aligned}$$

$$\left[\begin{array}{ccc} -1 & 2 & -1 \\ 2 & -1 & -1 \\ -1 & -1 & 2 \end{array} \right] \xrightarrow{\text{row}} \left[\begin{array}{ccc} -1 & 2 & -1 \\ 0 & -3 & 0 \\ 0 & 0 & 0 \end{array} \right] \xrightarrow{\text{col}} \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{array} \right]$$

(doing integer-invertible row and column operations on this system). Hence $m(K) = 3$.

The least modulus for the trefoil is three. This corresponds to the labelling



$$(2 \cdot 0 - 2 \equiv 1 \pmod{3}).$$

Thus the three-coloration of the trefoil corresponds to the $(\mathbb{Z}/3\mathbb{Z})$ -crystal. $m(K) = 3$ implies that the trefoil is non-trivial. (Note that in the $(\mathbb{Z}/m\mathbb{Z})$ -crystal, $a\bar{a} = 2a - a = a$ hence $m(K)$ is an ambient isotopy invariant of K .)

2. One way to generalize Example 1 is to consider the most general crystal such that $\bar{a} = \bar{a}$. Thus we require that $a\bar{b}\bar{b} = a$ and that $\bar{a}\bar{b} = \bar{b}\bar{a}\bar{b}$ for all $a, b \in \mathcal{C}_0$. Call such a crystal a **light crystal** (it corresponds to the **involuntary quandle**

[JOY]). Let $\mathcal{CL}(K)$ denote the light crystal corresponding to a given knot or link K . Note that $\mathcal{CL}(K)$ does not depend upon the orientation of K .

In the case of the trefoil, we have

$$\begin{aligned}
 & c = a\bar{b} \\
 & b = c\bar{a} \\
 & a = b\bar{c} \\
 \Rightarrow & a\bar{c} = b\bar{c} \quad \bar{c} = b \\
 & b\bar{a} = c\bar{a} \quad \bar{a} = c \\
 & c\bar{b} = a\bar{b} \quad \bar{b} = a
 \end{aligned}$$

$x * y$	a	b	c
a	a	c	b
b	c	a	a
c	b	a	a

$$x * y = xy$$

This calculation shows that $\mathcal{CL}(K)$ is finite (for the trefoil) with three elements in $\mathcal{C}_0\mathcal{L}(K) = \{a, b, c\}$, and $\pi\mathcal{L}(K) \simeq \mathbf{Z}/3\mathbf{Z}$. In general, the involuntary quandle is not finite. Winker [WIN] shows that the first example of an infinite involuntary quandle occurs for the knot 8_{16} . Relatively simple links, such as

It is instructive to compute the light crystal (involuntary quandle) for the figure eight knot E . It is finite, but has five elements in $\mathcal{C}_0\mathcal{L}(E)$, one more element than there are arcs in the diagram. This phenomenon (more algebra elements than diagrammatic arcs) shows how **the crystal (or quandle) is a generalization of the idea of coloring the arcs of the diagram**. In general there is a given supply of colors, but a specific configuration of the knot or link uses only a subset of these colors.

3. One of the most important crystal structures is the **Alexander Crystal** (after Alexander [ALEX2]). Let \mathcal{M} be any module over the ring $\mathbb{Z}[t, t^{-1}]$. Given $a, b \in \mathcal{M}$, define $a\overline{b}$ and $a\overline{b}$ by the equations

$$\begin{aligned} a\overline{b} &= ta + (1-t)b \\ a\overline{b} &= t^{-1}a + (1-t^{-1})b. \end{aligned}$$

It is easy to verify the crystal axioms for these equations:

$$\begin{aligned} a\overline{b}\overline{b} &= t^{-1}(ta + (1-t)b) + (1-t^{-1})b \\ &= a + (t^{-1}-1)b + (1-t^{-1})b \\ &= a. \end{aligned}$$

$$\begin{aligned} a\overline{b}\overline{c} &= ta + (1-t)bc \\ &= ta + (1-t)(tb + (1-t)c) \\ &= ta + (t-t^2)b + (1-t)^2c. \end{aligned}$$

$$\begin{aligned} a\overline{c}\overline{b}\overline{c} &= t(a\overline{c}\overline{b}) + (1-t)c \\ &= t(ta\overline{c} + (1-t)b) + (1-t)c \\ &= t(t(t^{-1}a + (1-t^{-1})c) + (1-t)b) + (1-t)c \\ &= ta + (t^2-t)c + (t-t^2)b + (1-t)c \\ &= ta + (t-t^2)b + (1-t)^2c. \end{aligned}$$

Thus $\overline{b}\overline{b} = 1$ and $\overline{b}\overline{c}\overline{b} = \overline{c}\overline{b}\overline{c}$. Note that if $\mathcal{A}(\mathcal{M})$ denotes the Alexander crystal of the module \mathcal{M} , then $\mathcal{C}_0(\mathcal{M}) = \mathcal{M}$ and $\pi(\mathcal{M})$ is the group of automorphisms of \mathcal{M} generated by the maps $\overline{b}, \overline{b} : \mathcal{M} \rightarrow \mathcal{M}$, $a\overline{b} = \overline{b}(a)$, $a\overline{b} = \overline{b}(a)$.

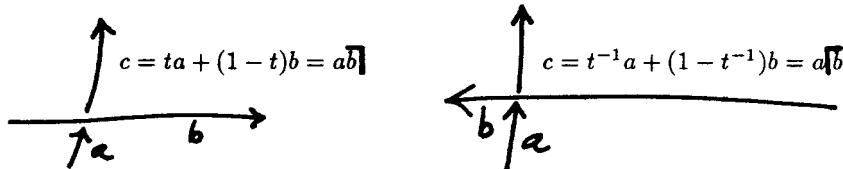
Note also that

$$a\overline{a} = ta + (1-t)a = a$$

and

$$a\overline{a} = t^{-1}a + (1-t^{-1})a = a.$$

Thus each element $a \in \mathcal{M}$ corresponds to an automorphism $a\bar{\square}: \mathcal{M} \rightarrow \mathcal{M}$ leaving a fixed. From the point of view of the knot theory, $\mathcal{A}(\mathcal{M})$ will give rise to an ambient isotopy invariant module $\mathcal{M}(K)$ (called the Alexander Module [CF], [FOX2]) if we define $\mathcal{M}(K)$ to be the module over $\mathbb{Z}[t, t^{-1}]$ generated by the edge labels of the diagram K , modulo the relations



For a given knot diagram, any one of the crossing relations is a consequence of all the others. As a result, we can calculate $\mathcal{M}(K)$ for a knot diagram of n crossings by taking $(n - 1)$ crossing relations.

Example.

$$c = ab\bar{b} = ta + (1 - t)b$$

$$b = ca\bar{a} = tc + (1 - t)b$$

$$a = bc\bar{b} = tb + (1 - t)c$$

Here $\mathcal{M}(T)$ is generated by the relations

$$ta + (1 - t)b - c = 0$$

$$(1 - t)a - b + tc = 0$$

$$-a + tb + (1 - t)c = 0$$

$$\Rightarrow \begin{bmatrix} t & 1-t & -1 \\ 1-t & -1 & t \\ -1 & t & 1-t \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

and we see that the first relation is the negative of the sum of the second two relations.

A closer analysis reveals that the determinant of any $(n-1) \times (n-1)$ minor of this relation matrix is a generator of the ideal in $\mathbb{Z}[t, t^{-1}]$ of Laurent polynomials $f(t)$ such that $f(t) \cdot m = 0$ for any $m \in \widetilde{\mathcal{M}}$. This determinant is determined up to sign and power of t and it is called the **Alexander polynomial** $\Delta(t)$. The Alexander polynomial is an ambient isotopy invariant of the knot. Here $\widetilde{\mathcal{M}} = \mathcal{M}/\langle a \rangle$ where $\langle a \rangle$ is the sub-module generated by any basis element a .

In the case of the trefoil, we have

$$\begin{aligned}\Delta_T(t) &\doteq \text{Det} \begin{pmatrix} -1 & t \\ t & 1-t \end{pmatrix} = (t-1)-t^2 \\ &\doteq t^2 - t + 1.\end{aligned}$$

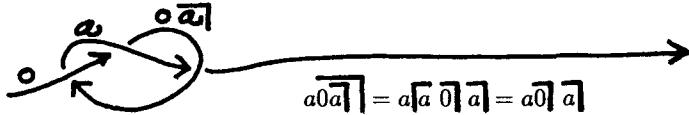
(\doteq denotes equality up to sign and power of t).

Think of the Alexander polynomial as a generalization of the modulus of a knot or link - as described in the first crystalline example.

For small examples the Alexander polynomial can be calculated by putting the relations directly on the diagram - first splitting the diagram as a 2-strand tangle and assigning an input and an output value of 0:



For example:



Note

$$\begin{aligned}x[0]a &= tx + (1-t)0 = tx \\ x[a-bar] &= t^{-1}x + (1-t^{-1})0 = t^{-1}x.\end{aligned}$$

Thus

$$\begin{aligned}a[0]a-bar &= t(ta) + (1-t)a \\ &= (t^2 - t + 1)a \\ &= \Delta(t)a.\end{aligned}$$

The requirement that $\Delta(t)a = 0$ shows that $\Delta(t) = t^2 - t + 1$ is the Alexander polynomial.

We produced the Alexander Crystal as a rabbit from a hat. In fact, the following result shows that it arises as the unique linear representation of the crystal axioms.

Theorem 13.7. Let \mathcal{M} be a free module of rank ≥ 3 over a commutative ring R . Suppose that \mathcal{M} has a crystal structure with operations defined by the formulas

$$ab\boxed{1} = ra + sb$$

$$a\boxed{b} = r'a + s'b$$

where $r, s, r', s' \in R$ and $a, b \in \mathcal{M}$. If r, s, r', s' are invertible elements of R , then we may write $r = t, s = (1 - t), r' = t^{-1}, s' = (1 - t^{-1})$. Thus, this linear crystal representation necessarily has the form of the Alexander Crystal.

Proof. We may assume (by the hypothesis) that a and b are independent over R . By the crystal structure, $ab\boxed{1}\boxed{b} = a$. Hence $a = (ra + sb)\boxed{b} = rr'a + (r's + s')b$. Therefore $rr' = 1$ and $r's + s' = 0$. The same calculation for $a\boxed{b}\boxed{1}\boxed{b} = a$ yields the equations $r'r = 1$ and $rs' + s = 0$. Thus r and r' are invertible, and $r' = r^{-1}$ while $r^{-1}s + s' = 0, rs' + s = 0$. Therefore $s' = -r^{-1}s, r' = r^{-1}$. Now apply the equation $xab\boxed{1} = x\boxed{b} a\boxed{1} \boxed{b}$:

$$\begin{aligned} x\overline{a\boxed{b}\boxed{1}} &= x\overline{ra+sb} = rx + s(ra + sb) \\ x\boxed{b} \overline{a\boxed{1}\boxed{b}} &= r(r(r'x + s'b) + sa) + sb \\ &= rx + r^2s'b + rsa + sb \\ &= rx + (r^2s' + s)b + rsa \\ &= rx + (-rs + s)b + rsa. \quad (s' = -r^{-1}s) \end{aligned}$$

Therefore, we obtain the further condition:

$$s^2 = -rs + s$$

$$\Rightarrow s = 1 - r.$$

Thus, if $r = t$, then $s = 1 - t, r' = t^{-1}, s' = 1 - t^{-1}$. This completes the proof. //

Remark. This theorem shows that the structure of the classical Alexander module arises naturally and directly from basic combinatorial considerations about colored states for link diagrams. The reader may enjoy comparing this development with the classical treatments using covering spaces ([FOX], [LK3], [ROLF], [B2]).

The Burau Representation.

We have already encountered the Alexander polynomial in the framework of the Yang-Baxter models. There is a remarkable relationship between the Alexander module and the R -matrix for the statistical mechanics model of the Alexander polynomial.

In order to see this relationship, I need to first extract the **Burau representation of the braid group** from the Alexander crystal:

$$\begin{array}{ccc} \text{Diagram showing two strands } a \text{ and } b \text{ crossing.} & & \text{Diagram showing two strands } a \text{ and } b \text{ crossing.} \\ \text{The strands } a \text{ and } b \text{ are labeled at the top.} & & \text{The strands } a \text{ and } b \text{ are labeled at the top.} \\ \text{The strands cross, and arrows indicate the strands pass over each other.} & & \text{The strands cross, and arrows indicate the strands pass over each other.} \\ a \overline{b} = t^{-1}a + (1 - t^{-1})b & & ab = ta + (1 - t)b \end{array}$$

Let a and b now denote two basis elements for a vector space V . Define $T : V \rightarrow V$ by $T(b) = ab = ta + (1 - t)b$ and $T(a) = b$. Note that, with respect

$$\begin{array}{ccc} \text{Diagram showing two strands } a \text{ and } b \text{ crossing.} & & \text{Diagram showing two strands } x \text{ and } y \text{ crossing.} \\ \text{The strands } a \text{ and } b \text{ are labeled at the top.} & & \text{The strands } x \text{ and } y \text{ are labeled at the top.} \\ \text{The strands cross, and arrows indicate the strands pass over each other.} & & \text{The strands cross, and arrows indicate the strands pass over each other.} \\ T(a) & & T^{-1}(x) \\ T(b) & & T^{-1}(y) \end{array}$$

to this basis, T has the matrix (also denoted by T)

$$T = \begin{bmatrix} 0 & t \\ 1 & (1-t) \end{bmatrix}$$

and that

$$T^{-1} = \begin{bmatrix} 1 - t^{-1} & 1 \\ t^{-1} & 0 \end{bmatrix}.$$

To obtain a representation of the n -strand braid group B_n , let

$$\rho(\sigma_1) = \begin{bmatrix} [T] & & & \textcircled{O} & & \\ & 1 & & & & \\ & & 1 & & & \\ \textcircled{O} & & & \ddots & & \\ & & & & & 1 \end{bmatrix}$$

$$\rho(\sigma_2) = \begin{bmatrix} 1 & & & \textcircled{O} & & \\ & [T] & & 1 & & \\ & & 1 & & & \\ \textcircled{O} & & & \ddots & & \\ & & & & & 1 \end{bmatrix}$$

$$\rho(\sigma_{n-1}) = \begin{bmatrix} 1 & & & \circ \\ & \ddots & & \\ \circ & & 1 & \\ & & & [T] \end{bmatrix}$$

where the large matrices are $n \times n$, and $[T]$ takes up a 2×2 block. ($\rho(\sigma_i^{-1})$ is obtained by replacing T by T^{-1} in the blocks.)

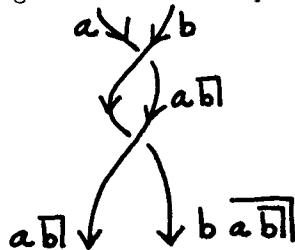
It is easy to see (from our discussion of invariance for the crystal) that the mapping $\rho : B_n \rightarrow GL(n; \mathbb{Z}[t, t^{-1}])$ is a representation of the braid group. This is the **Bureau representation**.

One can use this representation to calculate the Alexander polynomial.

Proposition 13.8. Let $w \in B_n$ be an element of the n -strand Artin braid group. Let $\rho(w) \in GL(n; \mathbb{Z}[t, t^{-1}])$ denote the matrix representing w under the Bureau representation. Let $A(t) = \rho(w) - I$ where I is an $n \times n$ identity matrix. Then $A(t)$ is an Alexander matrix for the link \bar{w} obtained by closing the braid w . (An Alexander matrix is a relation matrix for the Alexander Module $\mathcal{M}(\bar{w})$ as defined in Example 3 just before this remark.)

Remark. This proposition means (see e.g. [FOX]) that for $K = \bar{w}$, $\Delta_K(t)$ is the generator of the ideal generated by all $(n-1) \times (n-1)$ minors of $A(t)$. In the case under consideration, $\Delta_K(t)$ is equal (up to an indeterminacy of sign and power of t) to any $(n-1) \times (n-1)$ minor of $A(t)$. We write $\Delta_K(t) \doteq \text{Det}(A'(t))$ where $A'(t)$ denotes such a minor and \doteq denotes equality up to a factor of the form $\pm t^\ell (\ell \in \mathbb{Z})$.

I omit a proof of this proposition, but point out that it is equivalent to stating that $A(t) - I$ is a form of writing the relations in \mathcal{M} (as described for the crystal) that ensue from closing the braid. For example:



$$\left. \begin{aligned} & ta + (1-t)b \\ & tb + (1-t)(ta + (1-t)b) \\ & \quad \parallel \\ & (t-t^2)a + (t+(1-t)^2)b \\ & \quad \parallel \\ & (t-t^2)a + (t^2-t+1)b \end{aligned} \right\}$$

$$\rho(\sigma_1) = \begin{bmatrix} 0 & t \\ 1 & (1-t) \end{bmatrix}$$

$$\rho(\sigma_1^2) = \rho(\sigma_1)^2 = \begin{bmatrix} 0 & t \\ 1 & (1-t) \end{bmatrix} \begin{bmatrix} 0 & t \\ 1 & (1-t) \end{bmatrix} = \begin{bmatrix} t & t-t^2 \\ (1-t) & t^2-t+1 \end{bmatrix}.$$

We see that $\rho(\sigma_1^2)$ catalogues the mapping from the top to the bottom of the braid.
On closing the braid, we have

$$\begin{aligned} & a = a \bar{b} = ta + (1-t)b \\ & b = b \bar{a} = (t-t^2)a + (t^2-t+1)b \end{aligned}$$

These relations become the matrix $A(t) = \rho(w) - I$. $(w = \sigma_1^2)$

In the example we have

$$A(t) = \begin{bmatrix} t-1 & t-t^2 \\ 1-t & t^2-t \end{bmatrix}.$$

Hence (taking 1×1 minors) we have

$$\Delta \bar{w} \doteq t-1.$$

To continue this example, let's take $K = \sigma_1^3$. Then

$$\begin{aligned} \rho(\sigma_1^3) &= \rho(\sigma_1)^3 = \begin{bmatrix} t & t-t^2 \\ (1-t) & t^2-t+1 \end{bmatrix} \begin{bmatrix} 0 & t \\ 1 & 1-t \end{bmatrix} \\ &= \begin{bmatrix} t-t^2 & t^2+(1-t)(t-t^2) \\ t^2-t+1 & t(1-t)+(t^2-t+1)(1-t) \end{bmatrix} \\ &= \begin{bmatrix} t-t^2 & t-t^2+t^3 \\ t^2-t+1 & t^2-t^3-t+1 \end{bmatrix}. \end{aligned}$$

Thus $A_K(t) = \rho(\sigma_1)^3 - I = \begin{bmatrix} t - t^2 - 1 & t - t^2 + t^3 \\ t^2 - t + 1 & t^2 - t^3 - t \end{bmatrix}$. Hence $\Delta_K(t) \doteq t^2 - t + 1$ as before.

Exercise. It has only recently (Spring 1990) been shown (by John Moody [MD]) that the Burau representation is not faithful. That is, there is a non-trivial braid b such that $\rho(b)$ is the identity matrix, where $\rho(b)$ is the Burau representation described above. Moody's example is the commutator:

$$[765432'7'7'667765'4'32'1'23'456'7'7'6'6'7723'4'5'6'7', 7'7'87788778]$$

Here 7 denotes σ_7 , 7' denotes σ_7^{-1} and $[a, b] = aba^{-1}b^{-1}$. Verify that ρ is the identity matrix on Moody's nine-strand commutator, and show that the braid is non-trivial by closing it to a non-trivial link.

Deriving an R -matrix from the Burau Representation.

This last section of the lecture is devoted to showing a remarkable relationship between the Burau representation and the R -matrix that we used in section 12⁰ to create a model for the Alexander-Conway polynomial. We regard the Burau representation as associating to each braid generator σ_i , a mapping $\rho(\sigma_i) : V \rightarrow V$ where V is a module over $\mathbb{Z}[t, t^{-1}]$. Let $\wedge^*(V)$ denote the exterior algebra on V . This means that $\wedge^*(V) = \wedge^0(V) \oplus \wedge^1(V) \oplus \dots \oplus \wedge^d(V)$ where $d = \dim(V/\mathbb{Z}[t, t^{-1}])$ and $\wedge^0(V) = \mathbb{Z}[t, t^{-1}]$ while $\wedge^k(V)$ consists in formal products $v_{i_1} \wedge \dots \wedge v_{i_k}$ of basis elements $\{v_1, \dots, v_d\}$ of V and their sums with coefficients in $\mathbb{Z}[t, t^{-1}]$. These wedge products are associative, anti-commutative ($v_i \wedge v_j = -v_j \wedge v_i$) and linear ($(v + w) \wedge z = (v \wedge z) + (w \wedge z)$). We shall see that the R -matrix emerges when we prolong ρ to $\widehat{\rho}$ where $\widehat{\rho} : B_n \rightarrow \text{Aut}(\wedge^*(V))$. I am indebted to Vaughan Jones for this observation [JO8]. The relationship leads to a number of interesting and (I believe) important questions about the interplay between classical topology and state models arising from solutions to the Yang-Baxter Equation.

Here is the construction: First just consider the Burau representation for 2-strand braids. Then, if V has basis $\{a, b\}$ ($a = v_1, b = v_2$) then we have

$$\begin{aligned} \rho(\sigma)(a) &= b \\ \rho(\sigma)(b) &= ta + (1 - t)b. \end{aligned}$$

In $\wedge^*(V)$, we have the basis $\{1, a, b, a \wedge b\}$ and $\widehat{\rho}(\sigma)$ acts via the formulas:

$$\begin{aligned}\widehat{\rho}(\sigma)(1) &= 1 \\ \widehat{\rho}(\sigma)(a) &= b \\ \widehat{\rho}(\sigma)(b) &= ta + (1-t)b \\ \widehat{\rho}(\sigma)(a \wedge b) &= (\rho(\sigma)(a)) \wedge (\rho(\sigma)(b)) = (-t)a \wedge b.\end{aligned}$$

Let R denote the matrix of $\widehat{\rho}(\sigma)$ with respect to the basis $\{1, a, b, a \wedge b\}$. Then

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & t & 0 \\ 0 & 1 & (1-t) & 0 \\ 0 & 0 & 0 & (-t) \end{bmatrix}.$$

We can see at once, using the criteria of the Appendix that R is a solution to the Yang-Baxter Equation. To see how it fits into 2-spin (\pm) spin-preserving models, let

$$\begin{aligned}++ &\leftrightarrow 1 \\ + - &\leftrightarrow a \\ - + &\leftrightarrow b \\ -- &\leftrightarrow a \wedge b\end{aligned}$$

so that

$$R = \begin{array}{c|c|c|c|c} & ++ & +- & -+ & -- \\ \hline ++ & 1 & 0 & t & -t \\ \hline +- & 0 & 1 & 1-t & \\ \hline -+ & t & 1-t & -t & \\ \hline -- & -t & & & \end{array}$$

Now refer to the Appendix to see that R satisfies the Yang-Baxter Equation.

On the other hand, we can change basis by first replacing t by t^{-2} and then multiplying all rows by t :

$$\begin{array}{c|c|c|c|c} & 1 & 0 & t & \\ \hline 1 & & & & \\ \hline 0 & & & & \\ \hline t & & & & \\ \hline 1-t & & & & \\ \hline -t & & & & \end{array}$$

$$\begin{array}{c} \leftrightarrow \\ \boxed{\begin{array}{|c|c|c|c|} \hline 1 & 0 & t^{-2} & \\ \hline & 1 & 1-t^{-2} & -t^{-2} \\ \hline \end{array}} \\ \leftrightarrow \\ \boxed{\begin{array}{|c|c|c|c|} \hline t & 0 & t^{-1} & \\ \hline & t & t-t^{-1} & -t^{-1} \\ \hline \end{array}}
 \end{array}$$

Now a straightforward basis change on the middle 2×2 matrix

$$\left(\begin{pmatrix} 1 & 0 \\ 0 & t^{-1} \end{pmatrix} \begin{pmatrix} 0 & t^{-1} \\ t & x \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & x \end{pmatrix} \right)$$

yields the matrix

$$\boxed{\begin{array}{|c|c|c|c|} \hline t & 0 & 1 & \\ \hline & 1 & t-t^{-1} & -t^{-1} \\ \hline \end{array}}$$

This is exactly the Yang-Baxter matrix that we used in section 12⁰ to create a model for the Alexander-Conway polynomial.

Exercise. Using the FRT construction (section 10⁰), show that the bialgebra associated with the R -matrix

$$R = \begin{pmatrix} t & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & (t-t^{-1}) & 0 \\ 0 & 0 & 0 & -t^{-1} \end{pmatrix}$$

for $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has relations

$$\left\{ \begin{array}{ll} ba = -t^{-1}ab & db = tbd \\ ca = -t^{-1}ac & dc = tcd \\ bc = cb & \\ b^2 = c^2 = 0 & \\ ad - da = (t^{-1} - t)bc & \end{array} \right.$$

and comultiplication

$$\Delta(a) = a \otimes a + b \otimes c$$

$$\Delta(b) = a \otimes b + b \otimes d$$

$$\Delta(c) = c \otimes a + d \otimes c$$

$$\Delta(d) = c \otimes b + d \otimes d.$$

(See [LK23], p.314 and also [GE]. Compare with [LEE1] and [LK18], [MUK2].)

Remarks and Questions. I leave it to the reader to explore the full representation $\hat{\rho} : B_n \rightarrow \text{Aut}(\wedge^* V)$ and how it is supported by the matrix R ($\hat{\rho}(\sigma_i) = I \otimes I \otimes \dots \otimes I \otimes R \otimes I \otimes \dots \otimes I$ where I is the identity matrix - for appropriate bases.)

The state model of section 12⁰ and the determinant definitions of the Alexander polynomial (related to the Burau representation) are, in fact, intertwined by these remarks. See [LK18] and [JA8] for a complete account. The key to the relationship is the following theorem.

Theorem 13.9. Let A be a linear transformation of a finite dimensional vector space V . Let $\hat{A} : \wedge^*(V) \rightarrow \wedge^*(V)$ be the extension of A to the exterior algebra of V . Let $S : \wedge^*(V) \rightarrow \wedge^*(V)$ be defined by $S | \wedge^k(V)(x) = (-\lambda)^k x$. Then

$$\text{Det}(A - \lambda I) = \text{Tr}(\mathcal{S}\hat{A}).$$

In other words, the characteristic polynomial of a given linear transformation can be expressed as the trace of an associated transformation on the exterior algebra of the vector space for A . This trace can then be related to our statistical mechanics model for the Alexander polynomial.

In [LK18] H. Saleur and I show that this interpretation of the state model for the Alexander polynomial fits directly into the pattern of the free-fermion model in statistical mechanics. This means that the state summation can be rewritten as a discrete Berezin integral with respect to the link diagram. This expresses the Alexander polynomial as a determinant in a new way. Murakami [MUK2] has observed that these same methods yield Yang-Baxter models for the multi-variable Alexander-Conway polynomials for colored links. (Compare [MUK3].)

In general, knot polynomials defined as Yang-Baxter state models do not have interpretations as annihilators of modules (such as the Alexander module). A reformulation of the state models of section 12⁰ in this direction could pave the way for an analysis of classical theorems about the Alexander polynomial in a statistical mechanics context, and to generalizations of these theorems for the Jones polynomial and its relatives.

Exercise. (The Berezin Integral). The Berezin integral [RY] is another way to work with Grassman variables. If θ is a single Grassman variable, then by definition

$$\int d\theta 1 = 0$$

while

$$\int d\theta \theta = 1$$

and $\theta d\theta = -d\theta \theta$ if it comes up. Since $\theta^2 = 0$, the most general function of θ , $F(\theta)$ has the form $F(\theta) = a + b\theta$. Therefore,

$$\int d\theta F(\theta) = b = dF/d\theta,$$

a strange joke.

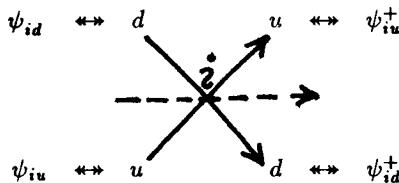
Let $x^1, x^2, \dots, x^n; y^1, y^2, \dots, y^n$ be distinct anti-commuting Grassman variables. Thus $(x^i)^2 = (y^i)^2 = 0$ while $x^i x^j = -x^j x^i$, $y^i y^j = -y^j y^i$ for $i \neq j$, and $x^i y^j = -y^j x^i$. Let M be an $n \times n$ matrix with elements in a commutative ring. Let $x^t = (x^1, \dots, x^n)$ and $y^t = (y^1, \dots, y^n)$ so that y is a column matrix. Show that

$$\text{DET}(M) = (-1)^{(n(n+1)/2)} \int dx^1 \dots dx^n dy^1 \dots dy^n e^{x^t M y}$$

where this is a Berezin integral.

Exercise. This exercise gives the bare bones of the relationship [LK18] between the free fermion model [SAM], [FW] and the Alexander-Conway state model of section 12⁰. Attach Grassman variables to the edges of a universe (shadow of a

link diagram) as follows:



Thus ψ^+ goes out, ψ goes in. The u and d designations refer to up and down with respect to the crossing direction

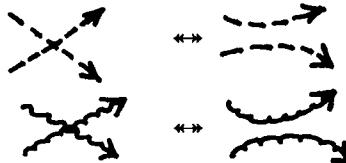
A state of the diagram consists in assigning +1 or -1 to each edge of the diagram. An edge labelled -1 is said to be fermionic. An edge labelled +1 is said to be bosonic. Assume that vertex weights are assigned as shown below



Here the dotted line denotes +1 and the wavy line denotes -1. If K is the diagram let $\langle K \rangle$ denote the state summation

$$\langle K \rangle = \sum_{\sigma} \langle K | \sigma \rangle (-1)^{\mathcal{L}(\sigma)}$$

where σ runs over the states described above, $\langle K | \sigma \rangle$ denotes the product of vertex weights in a given state, and $\mathcal{L}(\sigma)$ is the number of fermionic loops in a given state. Loops are counted by the usual separation convention:



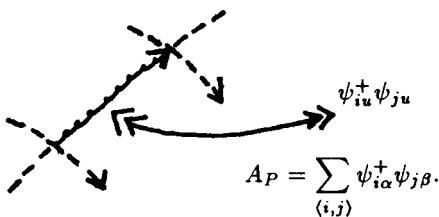
Show that, up to a constant factor, the Alexander polynomial state model of section 12⁰ can be put in this form (with, of course, different vertex weights corresponding to the two types of crossing.). Check that for the Alexander (Burau) R -matrix the vertex weights satisfy

$$w_1 w_2 = w_5 w_6 - w_3 w_4.$$

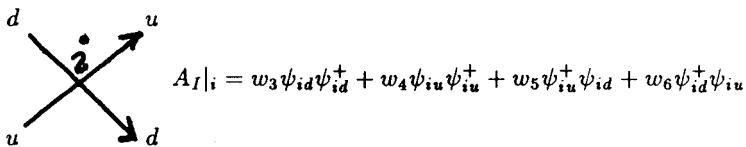
Now suppose that the weights in the model are just known to satisfy the “free-fermion condition” $w_1 w_2 = w_5 w_6 - w_3 w_4$ and suppose that $w_2 = -1$ (as can always be arranged). Define Grassman forms

$$d\psi d\psi^+ = \prod_i d\psi_i d\psi_i^+$$

and $A = A_I + A_P$. The forms A_I (interaction) and A_P (propagator) are explained below:



The term A_P is the sum over all oriented edges in K of the products of fermionic creation and annihilation operators for that edge.



A_I is the sum, over all vertices, of the interaction terms described above.

Now assume that K is a one-input, one-output tangle universe such as



with fermionic input and output. Show that

$$\langle K \rangle = \int d\psi d\psi^+ e^A$$

where the integral denotes a Berezin integral. Show that this implies that there exists a matrix \mathbf{A} such that

$$\langle K \rangle = \text{Det}(\mathbf{A}).$$

14⁰. The Kauffman Polynomial.

The Kauffman polynomial ([LK8], [JO5], [TH3], [TH4]) (Dubrovnik version) is defined as a regular isotopy invariant D_K of unoriented links K . $D_K = D_K(z, a)$ is a 2-variable polynomial satisfying:

1. $K \approx K' \Rightarrow D_K = D_{K'}.$
2. $D \cancel{\times} - D \cancel{\times} = z(D \cancel{\wedge} - D \cancel{\circ} \cancel{\circ})$
3. $D \bullet = \mu = ((a - a^{-1})/z) + 1$
 $D \cancel{\circ} = aD$
 $D \cancel{\circ} \cancel{\circ} = a^{-1}D$

The corresponding normalized invariant, Y_K , of ambient isotopy (for oriented links) is given by the formula

$$Y_K = a^{-w(K)} D_K$$

where $w(K)$ denotes the writhe of K .

This polynomial has a companion that I denote L_K . L_K is a regular isotopy invariant satisfying the identities

$$\begin{aligned} L \cancel{\times} + L \cancel{\times} &= z(L \cancel{\wedge} + L \cancel{\circ} \cancel{\circ}) \\ L \cancel{\circ} &= aL \quad , \quad L \bullet = \lambda = \left(\frac{a + a^{-1}}{z} \right) - 1 \\ L \cancel{\circ} \cancel{\circ} &= a^{-1}L \end{aligned}$$

The L -polynomial specializes at $a = 1$ to the Q -polynomial of [BR] a 1-variable invariant of ambient isotopy. As usual, we can normalize L to obtain a 2-variable invariant of ambient isotopy:

$$F_K(z, a) = a^{-w(K)} L_K.$$

In fact, the F and Y polynomials are equivalent by a change of variables. The result is (compare [LICK2]):

Proposition 14.1.

$$L_K(z, a) = (-1)^{c(K)} i^{w(K)} D_K(-iz, ia)$$

and

$$F_K(z, a) = (-1)^{c(K)} Y_K(-iz, ia)$$

where $c(K)$ denotes the number of components of K , and $w(K)$ is the writhe of K . (The first formula is valid for unoriented links K and any choice of orientation for K to compute $w(K)$ - since $i^{w(K)}$ is independent of the choice of orientation. The second formula depends upon the choice of orientation.)

We shall leave the proof of this proposition to the reader. However, some discussion and examples will be useful.

Discussion of D and L .

First note that we have taken the convention that $D(O) = \mu$, $L(O) = \lambda$ where these loop values are fundamental to the polynomials in the sense that the value of the polynomial on the disjoint union $O \amalg K$ of a loop and a given link K is given by the product of the loop value and the value of K ;

$$D(O \amalg K) = \mu D(K)$$

$$L(O \amalg K) = \lambda L(K).$$

It is easy to see that the bracket polynomial (see section 3⁰):

$$\begin{aligned} \langle \text{X} \rangle &= A \langle \text{X} \rangle + A^{-1} \langle \text{O} \rangle \\ \langle \text{O} \rangle &= -A^2 - A^{-2} \end{aligned}$$

is a special case of both L and D . Hence the *Jones polynomial is a special case of both F and Y* . In fact, we have

$$\begin{aligned} \langle \text{X} \rangle - \langle \text{X} \rangle &= (A - A^{-1})[\langle \text{X} \rangle - \langle \text{O} \rangle] \\ \langle \text{O} \rangle &= (-A^3)\langle \text{X} \rangle \\ \langle \text{O} \rangle &= (-A^{-3})\langle \text{X} \rangle. \end{aligned}$$

Hence $\langle K \rangle(A) = D_K(A - A^{-1}, -A^3)$. Similarly,

$$\langle \text{X} \rangle + \langle \text{Y} \rangle = (A + A^{-1})[\langle \text{X} \rangle + \langle \text{Y} \rangle].$$

Hence $\langle K \rangle(A) = L_K((A + A^{-1}), -A^3)$. From this, it is easy to verify direct formulas for the Jones polynomial as a special case of F and of Y . Thus we know that

$$V_K(t) = f_K(t^{-1/4})/f_O(t^{-1/4})$$

where

$$f_K(A) = (-A^3)^{-w(K)} \langle K \rangle(A).$$

Thus

$$f_K(A) = Y_K(A - A^{-1}, -A^3)$$

and

$$f_K(A) = F_K(A + A^{-1}, -A^3).$$

Therefore

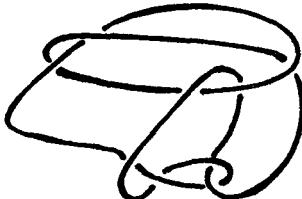
$$V_K(t) = \frac{Y_K(t^{-1/4} - t^{1/4}, -t^{-3/4})}{(-t^{-1/2} - t^{1/2})}$$

and

$$V_K(t) = \frac{F_K(t^{-1/4} + t^{+1/4}, -t^{-3/4})}{(-t^{-1/2} - t^{1/2})}.$$

The polynomials F and Y are pretty good at detecting chirality. They do somewhat better than the Homfly polynomial on this score. See [LK8] for more information along these lines. The book [LK3] contains a table of L -polynomials (normalized so that the unknot has value 1) for knots up to nine crossings. Thistlethwaite [TH5] has computed these polynomials for all knots up to twelve crossings.

Here are two examples for thinking about chirality: The knot 9_{42}



9_{42}

is chiral. Its chirality is *not* detected by either F or by the Homfly polynomial.

The knot 10_{48}



10_{48}

is chiral. Its chirality is detected by F (and Y) but not by the Homfly polynomial.

Birman-Wenzl Algebra.

We have mentioned that there is an algebra (the Hecke algebra) related to the Homfly polynomial. The Hecke algebra \mathcal{H}_n is a quotient of the $\mathbb{Z}[z, z^{-1}, a, a^{-1}]$ -group algebra of the braid group B_n by the relations $\sigma_i - \sigma_i^{-1} = z$ (or $\sigma_i^2 = 1 + z\sigma_i$). These relations are compatible with the corresponding polynomial relation, $H_{\text{X}} - H_{\text{X}} = zH_{\text{Y}}$, and so we get “traces” defined on the Hecke algebra by evaluating the H -polynomial (the un-normalized Homfly polynomial) on braids lifting elements in the Hecke algebra. Another route to the construction of the Homfly polynomial was to directly construct these traces at the algebra level (see [JO3]).

An analogous idea leads to algebras associated with the versions of the Kauffman polynomial. This is the Birman-Wenzl Algebra [B1]. For example, if we are using the Dubrovin polynomial then the basic relation is $D_{\text{X}} - D_{\text{X}} = z(D_{\text{X}} - D_{\text{Y}})$ and this corresponds formally to a relation in the braid monoid algebra:

$$\sigma_i - \sigma_i^{-1} = z(U_i - 1)$$

where U_i is a cup-cap combination in the $i, i+1$ -th place as we discussed in section 7⁰. This algebra involves understanding how the braiding and Temperley-Lieb generators U_i are interrelated. (See [LK8].)

Jaeger's Theorem.

While the Homfly polynomial and the Kauffman polynomial are distinct - with different topological properties, there is a very beautiful relationship between them - due to Francois Jaeger [JA9], and also observed in a special case by Reshetikhin [RES].

Jaeger shows that the Y -polynomial can be obtained as a certain weighted sum of Homfly polynomials on links associated with a given link K . This is a kind of state expansion for Y_K over states that are evaluated via the Homfly polynomial P_K !

The idea behind Jaeger's construction is very simple. Consider the following formalism:

$$\left[\begin{array}{c} \diagup \\ \diagdown \end{array} \right] = z \left[\left[\begin{array}{c} \diagup \\ \diagdown \end{array} \right] - \left[\begin{array}{c} \uparrow \\ \downarrow \end{array} \right] \right] + \left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right] + \left[\begin{array}{c} \nwarrow \\ \swarrow \end{array} \right] + \left[\begin{array}{c} \nearrow \\ \swarrow \end{array} \right] + \left[\begin{array}{c} \nwarrow \\ \nearrow \end{array} \right] \quad (*)$$

The small diagrams in this formula are intended to stand for parts of a larger diagram (differing only at the indicated site) *and* for the corresponding polynomial evaluation. The oriented diagrams will be evaluated by an oriented polynomial (Homfly polynomial) coupled with data about rotation numbers - we shall specify the evaluations below. Viewed as a state expansion, (*) demands states whose arrow configurations are globally compatible as oriented link diagrams. For example,

$$\left[\infty \right] = z \left[\left[\begin{array}{c} \diagup \\ \diagdown \end{array} \right] - \left[\begin{array}{c} \uparrow \\ \downarrow \end{array} \right] \right] + \left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right] + \left[\begin{array}{c} \nwarrow \\ \swarrow \end{array} \right].$$

A *state* for the expansion $[K]$ is obtained by splicing some subset of the crossings of K *and* choosing an orientation of the resulting link. The formula (*) gives vertex weights for each state. Thus we have,

$$[K] = \sum_{\sigma} [K|\sigma][\sigma]$$

where $[K|\sigma]$ is the product of these vertex weights ($\pm z$ or 1), and $[\sigma]$ is an as-yet-to-be-specified evaluation of an oriented link σ .

Define $[\sigma]$ by the formula

$$[\sigma] = (ta^{-1})^{\text{rot}(\sigma)} R_\sigma(t - t^{-1}, a)$$

where $\text{rot}(\sigma)$ denotes the rotation number (Whitney degree) of the oriented link σ , and $R_\sigma(z, a)$ is the regular isotopy version of the Homfly polynomial defined by:

$$\left\{ \begin{array}{l} R \overrightarrow{\nearrow} - R \overleftarrow{\nearrow} = zR \overrightarrow{\rightarrow} \\ R \overrightarrow{\partial} = aR \overrightarrow{\rightarrow} \\ R \overleftarrow{\partial} = a^{-1}R \overrightarrow{\rightarrow} \\ R \circ = \left(\frac{a - a^{-1}}{z} \right), \quad \left(R_{OIK} = \left(\frac{a - a^{-1}}{z} \right) R_K \right). \end{array} \right.$$

Theorem 14.2. (Jaeger). With the above conventions, $[K]$ is a Homfly polynomial expansion of the Dubrovnik version of the Kauffman polynomial:

$$[K](t - t^{-1}, a) = D_K(t - t^{-1}, t^{-1}a^2).$$

Proof (Sketch). First note that this model satisfies the Dubrovnik polynomial exchange identity:

$$\begin{aligned} [\cancel{\times}] &= z \left([\cancel{\leftarrow}] - [\cancel{\uparrow\downarrow}] \right) + [\cancel{\times}] + [\cancel{\times}] + [\cancel{\times}] + [\cancel{\times}] \\ [\times] &= z \left([\cancel{\uparrow\downarrow}] - [\cancel{\leftarrow}] \right) + [\cancel{\times}] + [\cancel{\times}] + [\cancel{\times}] + [\cancel{\times}]. \end{aligned}$$

Hence

$$\begin{aligned}
 [\times] - [\times] &= z \left(\left[\begin{smallmatrix} \leftarrow \\ \rightarrow \end{smallmatrix} \right] - \left[\begin{smallmatrix} \uparrow \\ \downarrow \end{smallmatrix} \right] - \left[\begin{smallmatrix} \downarrow \\ \uparrow \end{smallmatrix} \right] + \left[\begin{smallmatrix} \leftarrow \\ \rightarrow \end{smallmatrix} \right] \right) \\
 &\quad + \left[\begin{smallmatrix} \nearrow \\ \nwarrow \end{smallmatrix} \right] - \left[\begin{smallmatrix} \nearrow \\ \searrow \end{smallmatrix} \right] + \left[\begin{smallmatrix} \searrow \\ \nwarrow \end{smallmatrix} \right] - \left[\begin{smallmatrix} \nearrow \\ \nearrow \end{smallmatrix} \right] \\
 &\quad + \left[\begin{smallmatrix} \nwarrow \\ \nwarrow \end{smallmatrix} \right] - \left[\begin{smallmatrix} \nwarrow \\ \nwarrow \end{smallmatrix} \right] + \left[\begin{smallmatrix} \nearrow \\ \nearrow \end{smallmatrix} \right] - \left[\begin{smallmatrix} \nwarrow \\ \nwarrow \end{smallmatrix} \right] \\
 &= z \left(\left[\begin{smallmatrix} \leftarrow \\ \rightarrow \end{smallmatrix} \right] - \left[\begin{smallmatrix} \uparrow \\ \downarrow \end{smallmatrix} \right] - \left[\begin{smallmatrix} \downarrow \\ \uparrow \end{smallmatrix} \right] + \left[\begin{smallmatrix} \leftarrow \\ \rightarrow \end{smallmatrix} \right] \right) \\
 &\quad + z \left[\begin{smallmatrix} \uparrow \\ \rightarrow \end{smallmatrix} \right] - z \left[\begin{smallmatrix} \downarrow \\ \downarrow \end{smallmatrix} \right] + z \left[\begin{smallmatrix} \leftarrow \\ \leftarrow \end{smallmatrix} \right] \\
 &\quad - z \left[\begin{smallmatrix} \uparrow \\ \uparrow \end{smallmatrix} \right] \quad (\text{by Conway identity for state evaluations}) \\
 &= z \left(\left[\begin{smallmatrix} \leftarrow \\ \rightarrow \end{smallmatrix} \right] + \left[\begin{smallmatrix} \leftarrow \\ \leftarrow \end{smallmatrix} \right] + \left[\begin{smallmatrix} \rightarrow \\ \rightarrow \end{smallmatrix} \right] + \left[\begin{smallmatrix} \leftarrow \\ \leftarrow \end{smallmatrix} \right] \right) - \\
 &\quad z \left(\left[\begin{smallmatrix} \uparrow \\ \downarrow \end{smallmatrix} \right] + \left[\begin{smallmatrix} \downarrow \\ \uparrow \end{smallmatrix} \right] + \left[\begin{smallmatrix} \downarrow \\ \downarrow \end{smallmatrix} \right] + \left[\begin{smallmatrix} \uparrow \\ \uparrow \end{smallmatrix} \right] \right) \\
 &= z \left(\left[\begin{smallmatrix} \sim \\ \sim \end{smallmatrix} \right] - \left[\begin{smallmatrix}) \\ (\end{smallmatrix} \right] \right).
 \end{aligned}$$

Note that the value of the loop in this model is given by

$$\begin{aligned}
 [\circlearrowleft] &= \left[\begin{smallmatrix} \circlearrowleft \\ \circlearrowleft \end{smallmatrix} \right] + \left[\begin{smallmatrix} \circlearrowleft \\ \circlearrowright \end{smallmatrix} \right] \\
 &= (ta^{-1})^{-1}R \left[\begin{smallmatrix} \circlearrowleft \\ \circlearrowleft \end{smallmatrix} \right] + (ta^{-1})R \left[\begin{smallmatrix} \circlearrowleft \\ \circlearrowright \end{smallmatrix} \right] \\
 &= (t^{-1}a + ta^{-1}) \left(\frac{a - a^{-1}}{t - t^{-1}} \right) \\
 &= \frac{t - t^{-1} + t^{-1}a^2 - ta^{-2}}{t - t^{-1}} \\
 [\circlearrowleft] &= 1 + \frac{t^{-1}a^2 - ta^{-2}}{t - t^{-1}}.
 \end{aligned}$$

This is the correct μ -value for $D_K(t - t^{-1}, t^{-1}a^2)$.

It is easy to check that

$$\begin{aligned}
 [\overline{\sigma}] &= a^2t^{-1} \left[\begin{smallmatrix} \sim \\ \sim \end{smallmatrix} \right] \\
 [\bar{\sigma}] &= a^{-2}t \left[\begin{smallmatrix} \sim \\ \sim \end{smallmatrix} \right]
 \end{aligned}$$

by the same sort of calculation.

Finally, it is routine (but lengthy) to check that $[K]$ is a regular isotopy invariant (on the basis of the regular isotopy invariance of R_σ). I shall omit this verification. This completes the sketch of the proof. //

Discussion. This model of the Dubrovnik polynomial tells us that there is a certain kind of Yang-Baxter model for D_K that is based on the Yang-Baxter model we already know for specializations of R_K . We have

$$D_K(t - t^{-1}, a^2 t^{-1}) = \sum_{\sigma} [K|\sigma](t^{-1}a)^{\text{rot}(\sigma)} R_{\sigma}(t - t^{-1}, a)$$

where the states σ are obtained from the unoriented diagram K by splicing a subset of crossings, and orienting the resulting link. The product of vertex weights $[K|\sigma]$ is a product of $\pm z, 1$ according to the expansion rule (*):

$$\left[\begin{array}{c} \diagup \\ \diagdown \end{array} \right] = z \left(\left[\begin{array}{c} \leftarrow \\ \rightarrow \end{array} \right] - \left[\begin{array}{c} \uparrow \\ \downarrow \end{array} \right] \right) + \left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right] + \left[\begin{array}{c} \nwarrow \\ \nearrow \end{array} \right] + \left[\begin{array}{c} \nwarrow \\ \searrow \end{array} \right].$$

If we let $a = t^{n+1}$, then we can replace $R_{\sigma}(t - t^{-1}, t^{n+1})$ by the Yang-Baxter state model of section 11⁰. This gives a rather special set of models of D_K whose properties deserve closer investigation.

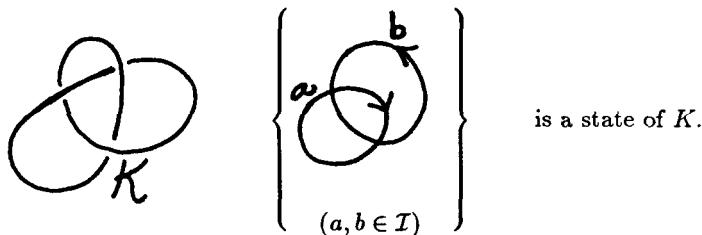
Research Problem. Investigate the models for D_K described above!

A Yang-Baxter Model and the Quantum Group for $SO(n)$.

There is another path to Yang-Baxter models for D_K that arises directly from the Yang-Baxter solution that we have used for specializations of the Homfly polynomial. Of course the model we have just indicated (using the Homfly expansion for the Dubrovnik polynomial) can be also seen as based on this Yang-Baxter solution. Here we study a simpler state structure, and we shall - in the process - uncover a significant solution to the Yang-Baxter Equation that is related to the group $SO(n)$.

The idea is as follows. Let K be an unoriented link. Define a *state* σ of K to be a labelled diagram σ that is obtained from K by first splicing some subset of crossings of K and projecting the rest so that the resulting diagram σ has components that are Jordan curves in the plane (no self-crossings). Each

component is then assigned an orientation and a label from a given ordered index set \mathcal{I} . Thus if K is a trefoil, then



We shall assume that if $a \in \mathcal{I}$ then $-a \in \mathcal{I}$ ($\mathcal{I} \subset \mathbb{R}$, the real numbers) and that a state obtained by switching both orientation and label ($a \mapsto -a$) is equivalent to the original state.

$$\begin{array}{c} \longrightarrow \\ a \end{array} \equiv \begin{array}{c} \longleftarrow \\ -a \end{array}$$

This means that at a splice we can assume that each state has local form since

$$\begin{array}{c} \nearrow a \\ \swarrow -b \end{array} \equiv \begin{array}{c} \nearrow a \\ \swarrow b \end{array}$$

With these assumptions, we can define a loop-count $\|\sigma\|$ just as in the oriented case:

$$\|\sigma\| = \sum_{C \in \text{components}(\sigma)} \text{rot}(C) \cdot \text{label}(C)$$

and, given an oriented R -matrix

$$\begin{array}{ccc} \cancel{\nearrow \searrow} & \leftrightarrow & \begin{array}{c} a \\ b \\ \nearrow c \\ \swarrow d \end{array} \leftrightarrow R_{cd}^{ab} \\ \cancel{\nearrow \searrow} & & \begin{array}{c} a \\ b \\ \swarrow c \\ \nearrow d \end{array} \leftrightarrow \bar{R}_{cd}^{ab} \end{array}$$

we can define $[K|\sigma]$ to be the product of the vertex weights assigned by $R(\bar{R})$ to the crossings in the diagram.

The assumption that the states have either a splice or a projection at each crossing, means that we are here assuming that R and \bar{R} are non-zero only

when $a = c$, $b = d$ or $a = d$, $b = c$. In fact, I shall use R , \bar{R} in the form

$$\begin{aligned} R &= (q - q^{-1}) \downarrow < \downarrow + q \quad \downarrow = \downarrow + \overleftarrow{\times} \\ \bar{R} &= (q^{-1} - q) \downarrow > \downarrow + q^{-1} \quad \downarrow = \downarrow + \overrightarrow{\times} \end{aligned}$$

our familiar solutions from previous sections.

With this choice, we define the state summation

$$[K] = \sum_{\sigma} [K|\sigma] q^{\|\sigma\|}.$$

We now show that an appropriate choice of the index set \mathcal{I} makes $[K]$ into a series of models for the Dubrovnik polynomial.

Before specializing \mathcal{I} , it is easy to see that $[K]$ satisfies the correct exchange identity, for

$$\begin{aligned} [\overleftarrow{\times}] &= (q - q^{-1}) \left[\overleftarrow{\downarrow} \right] + q \left[\overrightarrow{\downarrow} \right] + [\times] \\ &\quad + (q^{-1} - q) \left[\downarrow > \downarrow \right] + q^{-1} \left[\downarrow = \downarrow \right] \\ [\times] &= (q^{-1} - q) \left[\overleftarrow{\downarrow} \right] + q^{-1} \left[\overrightarrow{\downarrow} \right] + [\times] \\ &\quad + (q - q^{-1}) \left[\downarrow < \downarrow \right] + q \left[\downarrow = \downarrow \right]. \end{aligned}$$

Hence

$$\begin{aligned} [\overleftarrow{\times}] - [\times] &= (q - q^{-1}) \left(\left[\overleftarrow{\downarrow} \right] + \left[\overrightarrow{\downarrow} \right] + \left[\overrightarrow{\downarrow} \right] \right) + \\ &\quad (q^{-1} - q) \left(\left[\downarrow > \downarrow \right] + \left[\downarrow < \downarrow \right] + \left[\downarrow = \downarrow \right] \right) \\ &= (q - q^{-1}) \left(\left[\overleftarrow{\downarrow} \right] - \left[\downarrow \right] \left[\downarrow \right] \right). \end{aligned}$$

There are two cases for the index set \mathcal{I} . For n odd we take

$\mathcal{I}_n = \{-n, -n+2, \dots, -1, \tilde{0}, +1, 3, \dots, n-2, n\}$ and for n even we take

$\mathcal{I}_n = \{-n, -n+2, \dots, -2, 0, \tilde{0}, 2, 4, \dots, n\}$.

Note the presence of an extra zero in each index set. The extra zeros act numerically like ordinary zeros. This $\tilde{q}^0 = 1$. Note also that for odd n the index $\tilde{0}$ is special in that it is spaced by one unit from its neighbors, while all the other indices have spacing equal to two.

These new zeros will be handled by the

Zero Rules

1. If n is odd, then SHIFT the occurrence of  to an occurrence of , and assign it a vertex weight of 1. [For regular indices the crossed line configuration receives weight zero unless the two lines are labelled by distinct labels.]
2. For n even, treat the extra zero as an extra copy of the zero index.

As we shall see, these zero-rules are needed to ensure the invariance of the model under regular isotopy. The rules also make the model multiplicative under the type I move. We shall look at some of these issues shortly, but first it is worthwhile extracting a solution of the Yang-Baxter Equation from this model.

Since the model takes the form

$$\left[\begin{array}{c} \times \\ \times \end{array} \right] = z \left[\begin{array}{c} \downarrow \leftarrow \downarrow \\ \downarrow \end{array} \right] - z \left[\begin{array}{c} \rightarrow \\ \rightarrow \end{array} \right] + q \left[\begin{array}{c} \downarrow = \downarrow \\ \downarrow \end{array} \right] + q^{-1} \left[\begin{array}{c} \rightarrow \rightarrow \\ \rightarrow \end{array} \right] + \left[\begin{array}{c} \times \\ \times \end{array} \right]$$

it is tempting to think that the associated scattering matrix must be

$$z \downarrow \leftarrow \downarrow - z \rightarrow \rightarrow + q \downarrow = \downarrow + q^{-1} \rightarrow \rightarrow + \times$$

with conventions for handing  corresponding to the zero rules. However, this matrix does not satisfy the Yang-Baxter Equation! The correct scattering matrix is

$$M = z \left[\begin{array}{c} \downarrow \leftarrow \downarrow \\ \downarrow \end{array} - q \left(\frac{i-j}{2} \right) \begin{array}{c} i \rightarrow \\ j \rightarrow \end{array} \right] + q \left(\begin{array}{c} \downarrow \leftarrow \downarrow \\ (\neq 0) \end{array} \right) = \left(+ q^{-1} \begin{array}{c} \rightarrow \rightarrow \\ \rightarrow \end{array} + \begin{array}{c} \times \# \\ \times \# \end{array} + \begin{array}{c} \tilde{\delta} \\ \tilde{\delta} \end{array} \right)$$

(for the case n odd). Here I have included the Zero Rules in the diagrammatic notation, and an extra factor of $q^{\frac{i-j}{2}}$ on the horizontal split. The extra factor is a rotation number compensation.

In order to see the genesis of this compensation, consider the form of the model:

$$[K] = \sum_{\sigma} [K|\sigma] q^{\|\sigma\|}.$$

In this model it is no longer the case that the factor $q^{\|\sigma\|}$ remains constant for all the states corresponding to a given triangle move (fixed inputs and outputs). The simplest instance (and generic instance) of this change is seen in the difference between rotation numbers for



For example,



$$\begin{aligned} \text{rot} &= 0 \\ \|\sigma\| &= 0 \end{aligned}$$



$$\begin{aligned} \text{rot} &= 1 \\ \|\sigma'\| &= a = (a - (-a))/2. \end{aligned}$$

$$\frac{a \rightarrow}{-a \rightarrow} \equiv \frac{a \rightarrow}{-a \rightarrow}$$

The general compensation is given by a power of q as indicated in the diagram below:

$$\frac{a \rightarrow}{b \rightarrow} \leftrightarrow q^{(a-b)/2}.$$

Here is a second example:

$$\begin{aligned} &\left. \begin{array}{c} a \rightarrow \\ b \rightarrow \end{array} \right\} \frac{a-b}{2} \\ &\left. \begin{array}{c} -b \rightarrow \\ -a \rightarrow \end{array} \right\} \frac{-b+a}{2} \end{aligned}$$

$$\|\sigma''\| = a - b = \left(\frac{a-b}{2} \right) + \left(\frac{-b+a}{2} \right).$$

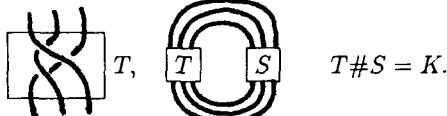
It follows that if M satisfies the QYBE then $[K]$ is invariant under the III(A) triangle move. Thus we have,

Proposition 14.3. Let $[K]$ be defined for unoriented diagrams as described above. If the matrix M is a solution to QYBE, then $[K]$ is invariant under the type IIIA move.

Proof. In this model, we have defined

$$[K] = \sum_{\sigma} [K|\sigma] q^{\|\sigma\|}$$

where $[K|\sigma]$ is the given product of vertex weights for oriented states described prior to this proof. Let T denote the tangle corresponding to a configuration for the type III move, and let S denote a tangle representing the rest of K . Thus $K = T \# S$ in the sense of connecting the tangles as shown below:



(S actually lives all over the plane exterior to T .)

Any state σ of K can be regarded as the $\#$ -connection of states σ_T and σ_S of T and S respectively. In order to prove invariance under the type III move we must consider all states σ_T and show that for any given state σ_S , the sum

$$\sum_{\sigma_T \in S(T, \sigma_S)} [K|\sigma] q^{\|\sigma\|}$$

(where $\sigma = \sigma_T \# \sigma_S$ and $S(T, \sigma_S)$ denotes all states of T whose end-conditions match the end-conditions of σ_S) is invariant under a triangle move applied to T . Now we have seen that for any state $\sigma = \sigma_T \# \sigma_S$, the norm, $\|\sigma\|$, is a sum of

contributions of the form $(a - b)/2$ for each occurrence of

in σ_T (horizontal with respect to T 's vertical lines) plus a part that is independent of the choice of σ_T . Thus

$$\|\sigma\| = \left[\sum_{\text{in } \sigma_T} \left(\frac{a - b}{2} \right) \right] + c$$

where c is independent of σ_T . This means that

$$\sum_{\substack{\sigma \in S(T, \sigma_S) \\ \sigma = \sigma_T \# \sigma_S}} [K|\sigma] q^{\|\sigma\|} = \left[\sum_{\sigma_T} [K|\sigma] \prod q^{(\alpha - b)/2} \right] q^c.$$

Thus, we must verify that the summation

$$\sum_{\sigma_T} [K|\sigma] \prod q^{(a-b)/2}$$

is invariant under the triangle move. If the terms $q^{(a-b)/2}$ are regarded as extra vertex weights, then this invariance follows at once from knowing that the matrix M is a solution to the Yang-Baxter Equation. This completes the proof of the proposition. //

In fact, with the zero-rules in effect, M does satisfy the Yang-Baxter Equation:

Proposition 14.4. Let the index set be taken in the form given in the preceding discussion, with the surrounding conventions for zeros. Then M is a solution to the QYBE, where M is the matrix we have described diagrammatically as

$$M = z \left[\begin{array}{c} \downarrow \leftarrow \\ \downarrow \end{array} \right] \left(-q^{(i-j)/2} \begin{array}{c} \overset{\leftarrow}{\nearrow} \\ \overset{\rightarrow}{\nwarrow} \end{array} \right) + q \left[\begin{array}{c} \downarrow \leftarrow \\ \downarrow \end{array} \right] \left(+ q^{-1} \begin{array}{c} \overset{\leftarrow}{\nearrow} \\ \overset{\rightarrow}{\nwarrow} \end{array} \right) + \cancel{\text{X}} + \cancel{\text{X}} + \cancel{\text{X}}.$$

Discussion. Before proving this proposition, it is worthwhile delineating the standard matrix form for M : Let f_j^i denote an elementary matrix: f_j^i has a 1 in the (i, j) place and 0 elsewhere. Here the indices i, j belong to a given index set \mathcal{I} . Taking the case n odd, so that $\mathcal{I} = \{-n, -n + 2, \dots, -1, \tilde{0}, 1, 3, \dots, n\}$, we then have

$$M = z \left[\sum_{i < j} f_i^i \otimes f_j^i - q^{(i-j)/2} \sum_{i < j} f_{-j}^i \otimes f_{-i}^{-i} \right] \\ + q \sum_{i \neq 0} f_i^i \otimes f_i^i + q^{-1} \sum_{i \neq 0} f_{-i}^i \otimes f_i^{-i} \\ + \sum_{i \neq j, -j} f_j^i \otimes f_i^j + f_0^{\tilde{0}} \otimes f_0^{\tilde{0}} \\ (z = q - q^{-1}).$$

I have arranged the indices on the elementary matrices so that they correspond to our abstract tensor conventions. Thus

$$\begin{array}{c} \text{a} \\ \swarrow \quad \downarrow \\ \text{a} \end{array} \equiv f_a^a$$

$$\begin{array}{c} \leftarrow \quad \leftarrow \\ \downarrow \quad \downarrow \\ (\neq 0) \end{array} \equiv \sum_{i < j} f_i^i \otimes f_j^j, \quad \begin{array}{c} \rightarrow \quad \rightarrow \\ \downarrow \quad \downarrow \\ (\neq 0) \end{array} \equiv \sum_{i \neq 0} f_i^i \otimes f_i^i$$

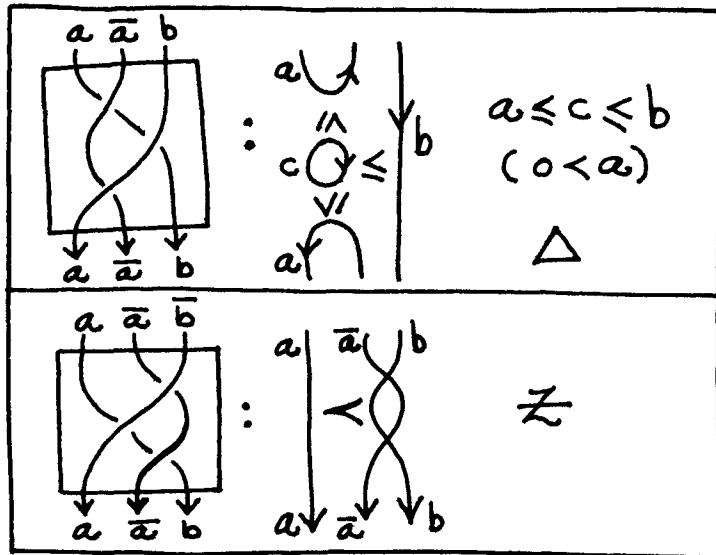
$$\begin{array}{c} \cancel{\times} \\ \cancel{\#} \end{array} \equiv \sum_{i \neq j, -j} f_j^i \otimes f_i^j.$$

With these conventions, we see that M corresponds directly to the matrices cited in [TU1], [RES], [JI2]. In particular, [RES] gives an independent verification that M satisfies the QYBE via the quantum group for $SO(n)$. It would be interesting to understand how the rotation compensation is related to this representation theory.

In the proof to follow, and the remaining verifications, we return to the abstract tensor form for M . For diagrammatic analysis, this shorthand has clear advantages.

Proof of 14.3. I will consider only the case n odd, and verify two key cases. In the following, the notation \bar{a} will be used for the negative of a : $\bar{a} = -a$.

Case 1. Assume that $0 < a < b$.



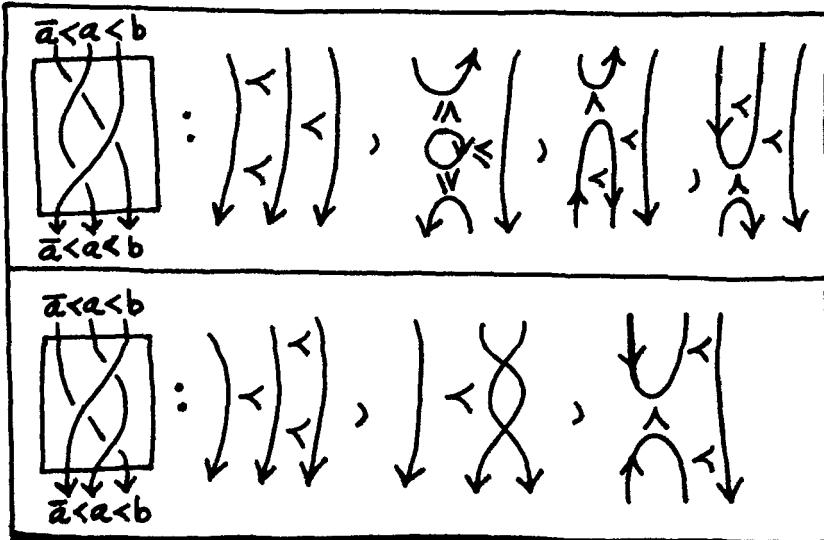
In the top part we have to sum over all the contributions where the center loop has spin c with $a \leq c \leq b$. Call this total contribution Δ . The bottom part contributes a factor of z . Summing up for Δ , we get

$$\begin{aligned}
 \Delta &= q^{-2}z + (-z)^2z \sum_{a < c < b} q^{(a-c)} + (-z)^2q(q^{-2})^{k+1}, (k = (b-a-2)/2) \\
 &= q^{-2}z + z^2q(1-q^2)[q^{-2} + (q^{-2})^2 + \dots + (q^{-2})^k] + z^2q(q^{-2})^{k+1} \\
 &= q^{-2}z + z^2q(1-q^{-2})[(q^{-2} - (q^{-2})^{k+1})/(1-q^{-2})] + z^2q(q^{-2})^{k+1} \\
 &= q^{-2}z + z^2q(q^{-2} - (q^{-2})^{k+1}) + z^2q(q^{-2})^{k+1} \\
 &= q^{-2}z + z^2q^{-1} \\
 &= (q^{-2} + zq^{-1})z \\
 &= (q^{-2} + 1 - q^{-2})z
 \end{aligned}$$

$$\Delta = z.$$

Thus M satisfies the Yang-Baxter Equation for this choice of indices.

Case 2.



Here it is not hard to see that the extra spin contribution inserted at $\tilde{0}$ by the Zero-Rules is just what is needed to compensate for the apparent extra term in the top part. Thus top and bottom contribute equally.

These two cases are representative of the manner in which this matrix M is a solution to the Yang-Baxter Equation. With the remaining cases left to the reader, this completes the proof. //

Finally, we have

Theorem 14.5. With the index set I_n taken as in the preceding discussion (and the Zero Rules) the model $[K]$ is an invariant of regular isotopy. Furthermore these models are multiplicative under the type 1 move with

$$\begin{aligned} [\cancel{\sigma}] &= q^{n+1} [\sim] \\ [\check{\sigma}] &= q^{-n+1} [\sim]. \end{aligned}$$

Since $[\cancel{\sigma}] - [\check{\sigma}] = z([\cancel{\sim}] - [\check{\sim}])$ (by definition of the model), we conclude that these models give an infinite set of models for the Dubrovnik version of the Kauffman polynomial.

Proof. I will omit the verification of invariance under the move IIB. This leaves us to check multiplicativity under the type I move. Again, we omit the proof for n even, and take the index set to be $\mathcal{I} = \{-n, -n+2, \dots, -1, \tilde{0}, 1, 3, 5, \dots, n\}$ for n odd. In fact, I shall take the case $\mathcal{I} = \{-1, \tilde{0}, 1\}$ and work it out in detail:

In the diagram below, I have indicated the four types of contribution to $[\overleftarrow{\partial}]$ and their respective vertex weights.

$$[\overleftarrow{\partial}] : \begin{array}{c} \text{Diagram 1: } \xrightarrow{\substack{a \\ \downarrow \\ b}} \\ \text{Diagram 2: } \xrightarrow{\substack{a \\ \parallel \\ a}} \\ \text{Diagram 3: } \xrightarrow{\substack{a \\ \curvearrowright \\ \tilde{0}}} \\ \text{Diagram 4: } \cancel{\xrightarrow{\substack{a \\ \curvearrowright \\ \tilde{0}}}} \end{array} \quad \begin{array}{c} zq^{-b} \\ (a \neq \tilde{0}) \\ qq^{-a} \\ -z \\ 1 \end{array}$$

Note that we are crucially using the zero-rules.

The bookkeeping for $\mathcal{I} = \{-1, \tilde{0}, 1\}$ then gives

$$[\overleftarrow{\partial}] = [\overleftarrow{\partial}_1] + [\overleftarrow{\partial}_2] + [\overleftarrow{\partial}_3].$$

We have

$$\begin{aligned} [\overleftarrow{\partial}_1] &= \left[\begin{array}{c} \xrightarrow{\substack{a \\ \downarrow \\ b}} \\ \text{Diagram 1} \end{array} \right] = q^2 \left[\xrightarrow{\substack{a \\ \downarrow \\ b}} \right] \\ [\overleftarrow{\partial}_2] &= \left[\begin{array}{c} \xrightarrow{\substack{a \\ \parallel \\ a}} \\ \text{Diagram 2} \end{array} \right] + \left[\begin{array}{c} \xrightarrow{\substack{a \\ \curvearrowright \\ \tilde{0}}} \\ \text{Diagram 3} \end{array} \right] \\ &= (zq + 1) \left[\xrightarrow{\substack{a \\ \curvearrowright \\ \tilde{0}}} \right] \\ &= (q^2 - 1 + 1) \left[\xrightarrow{\substack{a \\ \curvearrowright \\ \tilde{0}}} \right] \\ &= q^2 \left[\xrightarrow{\substack{a \\ \curvearrowright \\ \tilde{0}}} \right] \\ [\overleftarrow{\partial}_3] &= \left[\begin{array}{c} \cancel{\xrightarrow{\substack{a \\ \curvearrowright \\ \tilde{0}}}} \\ \text{Diagram 4} \end{array} \right] + \left[\begin{array}{c} \cancel{\xrightarrow{\substack{a \\ \curvearrowright \\ \tilde{0}}}} \\ \text{Diagram 4} \end{array} \right] + \left[\begin{array}{c} \cancel{\xrightarrow{\substack{a \\ \curvearrowright \\ \tilde{0}}}} \\ \text{Diagram 4} \end{array} \right] + \left[\begin{array}{c} \cancel{\xrightarrow{\substack{a \\ \curvearrowright \\ \tilde{0}}}} \\ \text{Diagram 4} \end{array} \right] \\ &= (zq + z + qq^{-1} - z) \left[\xrightarrow{\substack{a \\ \curvearrowright \\ \tilde{0}}} \right] \\ &= (q^2 - 1 + q - q^{-1} + 1 - q + q^{-1}) \left[\xrightarrow{\substack{a \\ \curvearrowright \\ \tilde{0}}} \right] \\ &= q^2 \left[\xrightarrow{\substack{a \\ \curvearrowright \\ \tilde{0}}} \right]. \end{aligned}$$

Thus $[\overbrace{\text{---}}] = q^2 [\text{---}]$ for $\mathcal{I} = \{-1, \tilde{0}, 1\}$. A similar calculation shows that for $\mathcal{I} = \mathcal{I}_n$,

$$\begin{aligned} [\overbrace{\text{---}}] &= q^{n+1} [\text{---}] \\ [-\overbrace{\text{---}}] &= q^{-n-1} [\text{---}]. \end{aligned}$$

This completes the proof. //

Summary.

We started by assuming the existence of a state model $[K]$ for D_K satisfying the expansion

$$\begin{aligned} [\times] &= z \left[\downarrow \leftarrow \downarrow \right] + q \left[\downarrow = \downarrow \right] + [\times] \\ &\quad - z \left[\rightarrow \rightarrow \right] + q^{-1} \left[\rightarrow \rightarrow \right] \end{aligned}$$

and using the vertex weights for the $SL(n)$ R -matrix

$$\overbrace{\text{---}} = z \left[\downarrow \leftarrow \downarrow \right] + q \left[\downarrow = \downarrow \right] + [\times].$$

We found that this led directly to a state model that indeed works (for special index sets)

$$[K] = \sum_{\sigma} [K|\sigma] q^{\|\sigma\|}$$

where the states are obtained by splicing or projecting the crossings of K to obtain a collection of inter-crossing Jordan curves. Orienting these curves, and labelling them from the index set yields the state σ . $[K|\sigma]$ is the product of vertex weights from the oriented R -matrix, and $\|\sigma\|$ denotes the sum of $\text{rot}(\ell)$ label(ℓ) over Jordan curves ℓ in σ . A second set of solutions to the Yang-Baxter Equation emerges from the structure of this model viz:

$$\begin{aligned} M &= z \left[\downarrow \leftarrow \downarrow + q^{(i-j)/2} \overbrace{\overset{i}{\rightarrow}}_{j} \right] \\ &\quad + q \left[\begin{array}{c} \downarrow = \downarrow \\ (\#o) \end{array} \right] + \left[\begin{array}{c} \rightarrow \rightarrow \\ (\#o) \end{array} \right] + [\times \#] + [\times \times] \end{aligned}$$

(with special index set as specified in the previous discussion).

This identifies these models with the models constructed by Turaev in [TU1]. He begins with the solutions M identified as originating in the work of Jimbo [JI2]. In [RES] Reshetikhin shows that the solutions M emerge from the analysis of the quantum group for $SO(n)$. It would be very nice to see how the $SL(n)$ and $SO(n)$ solutions are related via the quantum groups. There may be a story at this level that is in parallel with the relationship that we have seen in the state models. In particular, there should be a quantum group interpretation for the rotation-compensation factor $q^{(i-j)/2}$. Finally, the matrix M and its inverse \bar{M} can be used to create a specific representation of the Birman-Wenzl algebra [B1]. This is manifest from the structure of the model described here.

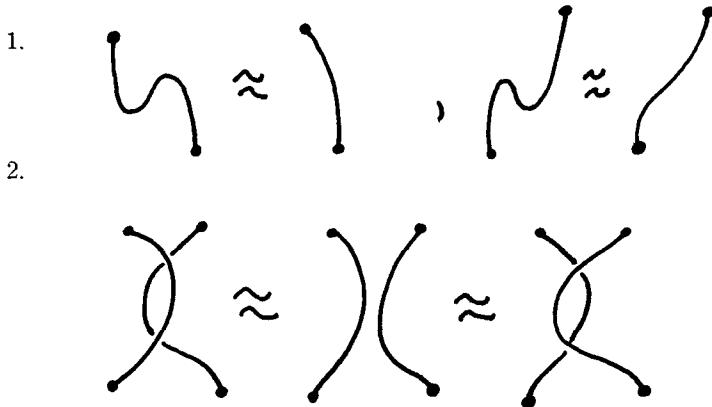
15⁰. Oriented Models and Piecewise Linear Models.

Recall from section 9⁰ our treatment of unoriented abstract tensor models for link invariants. The invariant is viewed as a generalized vacuum-vacuum amplitude with creation, annihilation and braiding operators:

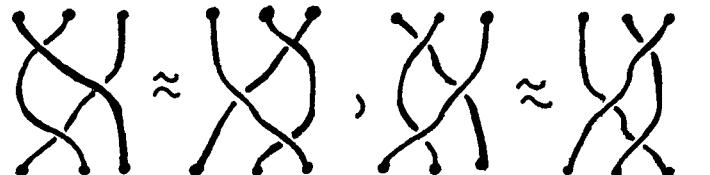
$$\begin{array}{c} \text{Diagram: } \text{a} \cup \text{b} \\ \text{Diagram: } \text{a} \cap \text{b} \end{array} \longleftrightarrow M_{ab}, \quad \begin{array}{c} \text{Diagram: } \text{a} \cup \text{b} \\ \text{Diagram: } \text{a} \cap \text{b} \end{array} \longleftrightarrow M^{ab}$$

$$\begin{array}{c} \text{Diagram: } \text{a} \times \text{b} \\ \text{Diagram: } \text{a} \times \text{b} \end{array} \longleftrightarrow R_{cd}^{ab}, \quad \begin{array}{c} \text{Diagram: } \text{a} \times \text{b} \\ \text{Diagram: } \text{a} \times \text{b} \end{array} \longleftrightarrow \bar{R}_{cd}^{ab}$$

The topology underlying this viewpoint has as its underpinning the generalized Reidemeister moves that take into account maxima and minima in diagrams written with respect to a height function. These moves are:



3.



4.



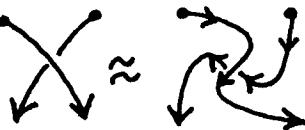
They translate into algebraic conditions on M and R :

1. $M_{ai}M^{ib} = \delta_a^b$, $M^{ai}M_{ib} = \delta_b^a$
2. R and \bar{R} are inverse matrices.
3. Yang-Baxter Equation for R and \bar{R} .
4. $\bar{R}_{cd}^{ab} = M_{ci}R_{dj}^{ia}M^{jb}$
 $\bar{R}_{cd}^{ab} = M^{ai}R_{ic}^{bj}M_{jd}$.

We have also remarked that in the case of the Drinfeld double construction, condition 4. translates to a condition about the antipode (and its representations) for this Hopf algebra. (Theorem 10.3).

A similar set of conditions holds for regular isotopy invariants of oriented link diagrams. It is worth recording these conditions here and comparing them both with Hopf algebra structures and with other forms of these models.

The generalization of the Reidemeister moves for the oriented case reads as follows:

1.  (and reverse orientations)
2. 
3. 
4.  (and variations).

Note that in the oriented case there are left and right oriented creations and annihilations:  and a medley of braiding operators



The twist moves (4.) coupled with the two types of type 2. moves, let us retain the simple Yang-Baxter relations 3. (The other versions of the triangle move then follow from these relations.) The algebra relations are then the direct (abstract tensor) transcriptions of 1. \rightarrow 4.

In this way, we can consider oriented models that take the form of vacuum-vacuum amplitudes.

Some features of the oriented amplitudes serve to motivate models that we have already discussed. For example, it is clear from this general framework that

the amplitude of a closed loop will, in general, depend upon its rotational sense in the plane. Thus, we expect that $\langle \bullet \circ \rangle \neq \langle \circ \bullet \rangle$ in general and that the specific values will be related to properties of the creation and annihilation matrices.

To make the story even more specific, I shall show that all of our previous models can be seen as oriented amplitudes and that they can also be seen as piecewise linear (pl) models (to be defined below). We can then raise the question of the relationship of oriented amplitudes and pl state models. First, we have a description of the pl models.

Piecewise Linear State Models.

The models I am about to describe are designed for piecewise linear knot diagrams. In the piecewise linear category I first described models of this type in [LK10]. These were designed to give a model in common for both the bracket polynomial and a generalization of the Potts model in statistical mechanics. Vaughan Jones [JO4] gave a more general version of vertex models for link invariants using the smooth category. The smooth models involve integrating an angular parameter along the link diagram. Here I give a piecewise linear version of Jones' models (see also [LK14], [LK23]) where the angular information is concentrated at the vertices of the diagram. In statistical mechanics these sorts of models appear in the work of Perk and Wu [PW] and Zamolodchikov [Z3] (and perhaps others).

A piecewise linear link diagram is composed of straight line segments so that the vertices are either **crossings** (locally 4-valent) or **corners** (2-valent). The diagram is oriented, and we assume the usual two types of local crossing. A matrix $S_{cd}^{ab}(\theta)$ is associated with each crossing as indicated below:

$$\Leftrightarrow S_{cd}^{ab}(\theta) = R_{cd}^{ab} \lambda^{(d-a)\theta/2\pi}.$$

Here θ is the angle in radians between the two crossing segments measured in the counter-clockwise direction. It is assumed that R_{cd}^{ab} is a solution to the Yang-Baxter Equation for $a, b, c, d \in \mathcal{I}$ (\mathcal{I} a specified index set) where $\mathcal{I} \subseteq \mathbf{R}$. Furthermore, we assume that R is **spin-preserving** in the sense that $R_{cd}^{ab} = 0$ unless $a + b = c + d$.

To the reverse crossing we associate $\bar{S}_{cd}^{ab}(\theta)$ as shown below:

$$\leftrightarrow \bar{S}_{cd}^{ab}(\theta) = \bar{R}_{cd}^{ab} \lambda^{(d-a)\theta/2\pi}.$$

Here \bar{R} is the inverse matrix for R in the sense that $R_{ij}^{ab} \bar{R}_{cd}^{ij} = \delta_c^a \delta_d^b$. Note that the angular part of the contribution is independent of the type of crossing.

Since the model is to be piecewise linear, we also have vertices of valence two. These acquire vertex weights according to the angle between the segments incident to the vertex:



The **partition function** (or **amplitude**) for a given piecewise linear link diagram K is then defined by the summation $\langle K \rangle = \sum_{\sigma} [K|\sigma]$ where $[K|\sigma]$ denotes the product of these vertex weights and a **state** (**configuration**) σ is an assignment of spins to the edges of the link diagram (each edge extends from one vertex to another) such that spins are constant at the corners (two-valent vertices) and preserved ($a + b = c + d$) at the crossings.

We can rewrite this product of vertex weights as $[K|\sigma] = \langle K|\sigma \rangle \lambda^{\|\sigma\|}$ where $\langle K|\sigma \rangle$ is the product of R values from the crossings, and $\|\sigma\|$ is the sum of angle exponents from all the crossings and corners. Thus

$$\langle K \rangle = \sum_{\sigma} \langle K|\sigma \rangle \lambda^{\|\sigma\|}.$$

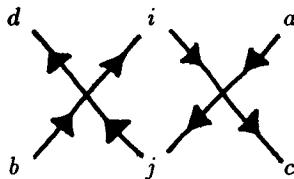
A basic theorem [JO4] gives a sufficient condition for $\langle K \rangle$ to be an invariant of regular isotopy:

Theorem 15.1. If the matrix R_{cd}^{ab} satisfies the Yang-Baxter Equation and if R also satisfies the cross-channel inversion

$$\sum_{i,j} \bar{R}_{jc}^{ia} R_{id}^{jb} \lambda^{((c-i)/2 + (d-j)/2)} = \delta_d^a \delta_c^b$$

then the model $\langle K \rangle$, as described above, is an invariant of regular isotopy.

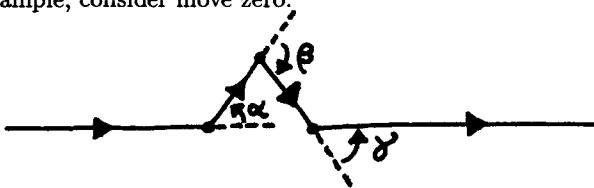
Remark. The cross-channel condition corresponds to invariance under the reversed type II move. The diagram below shows the pattern of index contractions corresponding to $\bar{R}_{jc}^{ia} R_{id}^{jb}$.



Before proving this theorem some discussion is in order.

Piecewise Linear Reidemeister Moves. These moves are shown in Figure 17. (As usual the moves in the figure are representatives, and there are a few more cases - such as switching a crossing in the type III move - that can also be illustrated.) In order to work with the invariance of the model under these moves, we have to keep track of the angles.

For example, consider move zero:



We have that the sum of the angles $\alpha + \beta + \gamma$, equals zero. (Since, in the Euclidean plane, the sum of the angles of a triangle is π .) Thus the product of vertex weights for the above configuration is the same as the product for a straight segment.

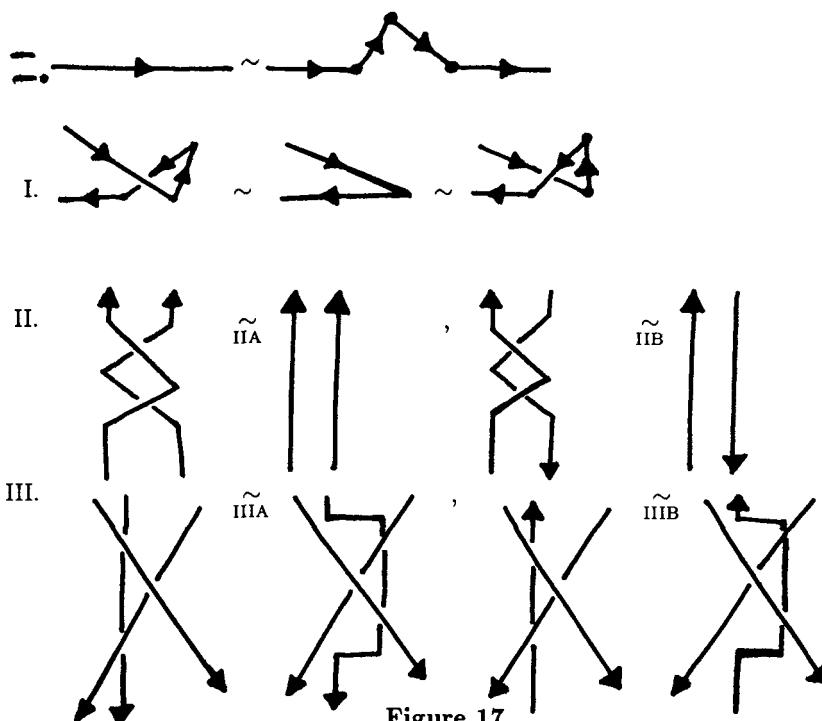


Figure 17

It is not hard to see that it is sufficient to assume that the main (input and output) lines for the type II move are parallel. The angular contributions involve some intricacy. To illustrate this point, and to progress towards the proof of the theorem, let's prove invariance of $\langle K \rangle$ under the move IIA.

Lemma 15.2. If the matrices R_{cd}^{ab} and \bar{R}_{cd}^{ab} are inverses in the sense that $R_{ij}^{ab} \bar{R}_{cd}^{ij} = \delta_c^a \delta_d^b$, then the model $\langle K \rangle$ (using angular vertex weights as described above) satisfies invariance under the pl move IIA.

Proof. We know that $\langle K \rangle$ has a summation of the form

$$\langle K \rangle = \sum_{\sigma} \langle K | \sigma \rangle \lambda^{\|\sigma\|}$$

where $\langle K|\sigma \rangle$ is the product of the vertex weights coming from the matrices R and \bar{R} , and $\|\sigma\|$ is the sum of the angular contributions for a state σ . Since R and \bar{R} are given to be inverse matrices, it will suffice to show that $\|\sigma\| = \|\sigma'\|$ for any states σ, σ' with given inputs a and b and outputs e and f ($a + b = e + f$ since the model is spin preserving) in the type IIA configuration shown in Figure 18.

Let Δ denote the sum of the angular contributions from the part of the diagram indicated in Figure 18. Then we shall prove that $\Delta = 0$

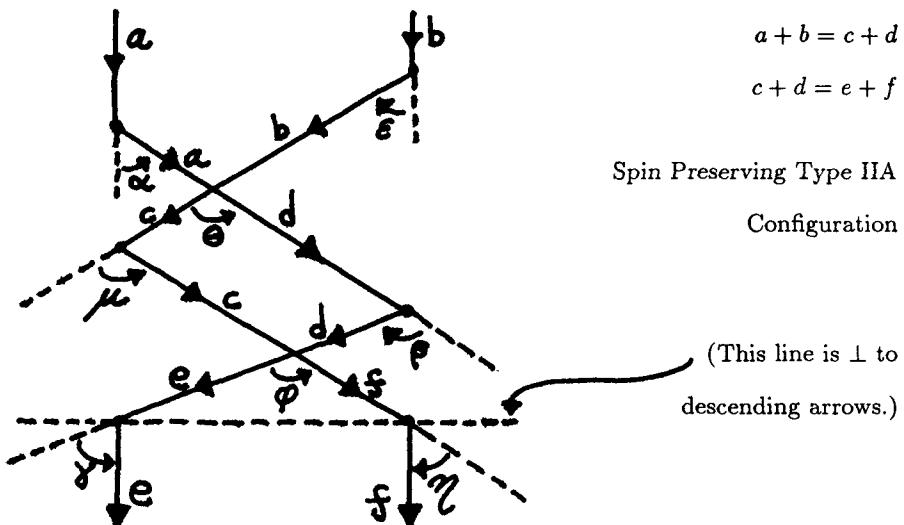


Figure 18

From Figure 18 it follows that

$$(*) \begin{cases} \alpha + \beta + \gamma &= 0 \\ \epsilon + \mu + \eta &= 0 \\ \theta + \phi &= \mu - \beta \\ a + b = c + d &= e + f \end{cases}$$

Thus

$$\Delta = a\alpha + b\epsilon + (d - a)\theta + \mu c + d\beta + (f - c)\phi + \gamma e + \eta f.$$

To show that $\Delta = 0$, our strategy is to use the equations (*) to simplify and reduce this formula for Δ . Thus

$$\begin{aligned}
 \Delta &= a\alpha + d\beta + e\gamma + (d - a)\theta + b\epsilon + c\mu + f\eta + (f - c)\phi \\
 &= a\alpha + (a + b - c)\beta + (a + b - f)\gamma + (f - c)\phi + b\epsilon \\
 &\quad + (a + b - d)\mu + (a + b - e)\eta + (d - a)\theta \\
 &= (b - c)\beta + (b - f)\gamma + (f - c)\phi + (a - d)\mu + (a - e)\eta + (d - a)\theta \\
 &= (d - a)\beta + (b - f)\gamma + (f - c)\phi + (a - d)\mu + (a - e)\eta + (d - a)\theta \\
 &= (d - a)(\beta + \theta - \mu) + (b - f)\gamma + (f - c)\phi + (a - e)\eta \\
 &= (a - d)\phi + (b - f)\gamma + (f - c)\phi + (a - e)\eta \\
 &= (c - b)\phi + (f - c)\phi + (b - f)\gamma + (a - e)\eta \\
 &= (f - b)\phi + (b - f)\gamma + (f - b)\eta \\
 &= (f - b)(\phi - \gamma - \eta) \\
 &= (f - b)(0)
 \end{aligned}
 \tag{By Figure 18}$$

$$\therefore \Delta = 0.$$

//

An exactly similar calculation shows that in the case of the IIB move we have $\Delta/2\pi = [(c - b) + (f - d)]/2$ where the spins are labelled as in Figure 19

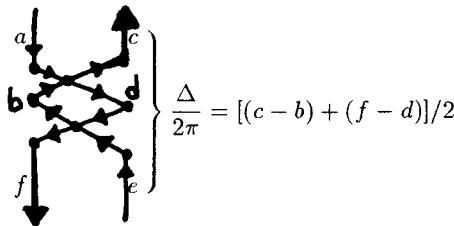
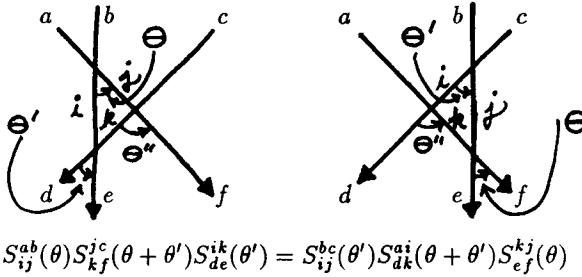


Figure 19

This explains the exponents in the cross-channel inversion statement for Theorem 15.1.

Proof of Theorem 15.1. By the Lemmas and discussion prior to this proof, it suffices to show that $\langle K \rangle$ is invariant under the move (IIIA) (Recall that IIA, IIB, IIIA together imply IIIB as in section 11⁰.).

Now refer to Figure 20 and note that because of the standard angle form of the type III moves it will suffice to prove that the sum of the angle contributions are the same for the simplified diagram shown in Figure 20.



Quantum Yang-Baxter Equation with Rapidity

Figure 20

In Figure 20 we see that for a given state σ with local spins $\{a, b, c, i, j, k, d, e, f\}$ the left hand diagram contribution to $2\pi\|\sigma\|$ by these spins is

$$\begin{aligned}\Delta &= (j - a)\theta + (e - i)\theta' + (f - j)(\theta + \theta') \\ &= (f - a)\theta + (f - j + e - i)\theta'.\end{aligned}$$

The right hand diagram contributes

$$\begin{aligned}\Delta' &= (j - b)\theta' + (k - a)(\theta + \theta') + (f - k)\theta \\ &= (f - a)\theta + (k - a + j - b)\theta'.\end{aligned}$$

Since $k + j = e + f$, we have $f - j + e - i = k - e + e - i = k - i$. Since $a + b = i + j$, we have $k - a + j - b = k + b - i - b = k - i$. Thus $\Delta = \Delta'$. Since $S_{cd}^{ab}(\theta) = R_{cd}^{ab} \lambda^{(d-a)\theta/2\pi}$, we see that this angle calculation, plus the fact that R satisfies the Yang-Baxter Equation implies that the model $\langle K \rangle$ is invariant under move IIIA. This completes the proof. //

Remark. In the course of this proof we have actually shown that if R_{cd}^{ab} satisfies the Yang-Baxter Equation, then $S_{cd}^{ab}(\theta) = R_{cd}^{ab}\lambda^{(d-a)\theta/2\pi}$ satisfies the Yang-Baxter Equation with rapidity parameter θ :

$$S_{ij}^{ab}(\theta)S_{kj}^{jc}(\theta + \theta')S_{de}^{ik}(\theta') = S_{ij}^{bc}(\theta')S_{dk}^{ai}(\theta + \theta')S_{ef}^{kj}(\theta).$$

In physical situations, from which this equation has been abstracted, the rapidity θ stands for the momentum difference between the particles involved in the interaction at this vertex.

A word of background about this rapidity parameter is in order. The underlying assumption is that the particles are interacting in the context of special relativity. This means that the momenta of two interacting particles can be shifted by a Lorentz transformation, and the resulting scattering amplitude must remain invariant. In the relativistic context (with light speed equal to unity) we have the fundamental relationship among energy (E), momentum (p) and mass (m):

$$E^2 - p^2 = m^2.$$

Letting mass be unity, we have

$$E^2 - p^2 = 1$$

and we can represent $E = \cosh(\theta)$, $p = \sinh(\theta)$. The Lorentz transformation shifts theta by a fixed amount. In light-cone coordinates (see section 9⁰ of Part II) the Energy-Momentum has the form $[e^\theta, e^{-\theta}]$, and this is transformed via $[A, B] \mapsto [Ae^\phi, Be^{-\phi}]$ under the Lorentz transformation. Consequently, if two particles have respective **rapidities** (the θ parameter) θ_1 and θ_2 , then the difference $\theta_1 - \theta_2$ is Lorentz invariant. The S-matrix must be a function of this difference of rapidities.

Now consider the diagram in Figure 20. Since momentum is conserved in the interactions, we see that for the oriented triangle interaction, the middle term has a rapidity difference that is the sum of the rapidity differences of the other two terms. Euclid tells us that (by the theorem on the exterior angle of a triangle) the angles between the corresponding lines satisfy just this relation.

Thus we identify rapidity differences with the angles between interaction lines, and obtain the correct angular relations from the geometry of the piecewise linear

diagrams. This angular correspondence is remarkable, and it beckons for a deeper physical interpretation. This is a mystery of a sort because the Yang-Baxter Equation with rapidity parameter is the original form of the equation both in the context of field theory, and in the context of statistical mechanics. Some of the simplest solutions contain a nontrivial rapidity parameter. For example, let

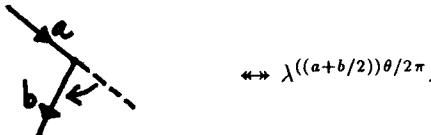
$$S_{cd}^{ab} = \downarrow \downarrow + \theta \times = \delta_c^a \delta_d^b + \theta \delta_d^a \delta_c^b.$$

It is a nice exercise to check that $S_{cd}^{ab}(\theta)$ satisfies the QYBE with rapidity θ . Nevertheless, we do not know how to use this S to create a link invariant! The seemingly more complex form $S_{cd}^{ab}(\theta) = R_{cd}^{ab} \lambda^{(d-a)\theta/2\pi}$ is just suited to the knot theory.

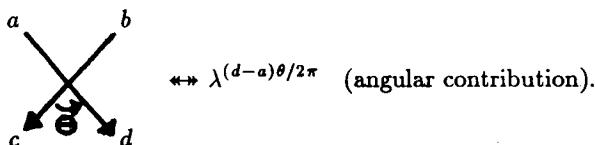
Remark. To apply Theorem 15.1 one must adjust λ so that the condition for cross-channel unitarity is satisfied. We already know good examples of solutions to this problem from section 11⁰ where

$$R_{cd}^{ab} = (q - q^{-1}) \downarrow \downarrow + q \downarrow \downarrow = \downarrow \downarrow + \times$$

and $\lambda = q$ for the index set $\mathcal{I}_n = \{-n, -n+2, \dots, n-2, n\}$. In this case it is easy to see that the state-evaluations match those of the model in section 11⁰. There we regarded states σ as composed of loops of constant spin and each loop received (in $\|\sigma\|$) the product of the spin and rotation numbers. **The local angles in the pl model add up to give the same result.** To see this, first consider a 2-valent vertex with spin input a and output b : (We usually take $a = b$). Set the vertex weight to be $\lambda^{((b+a)/2)\theta/2\pi}$. This gives the usual result when $a = b$.



Now consider the angular contribution at a crossing:



Here we assume that $a + b = c + d$. Thus if we break the crossing we find

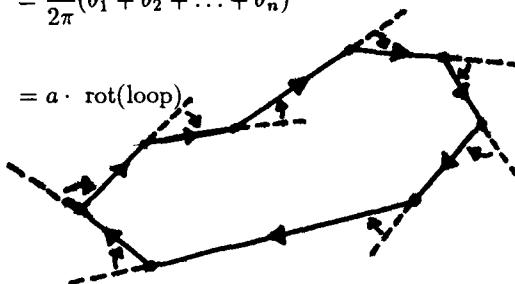
$$\begin{aligned} & \leftrightarrow \lambda^{((a+c)/2)(-\theta/2\pi)} \lambda^{((b+d)/2)\theta/2\pi} \\ &= \lambda^{(((d-a)+(b-c))/2)\theta/2\pi} \\ &= \lambda^{(d-a)\theta/2\pi}. \end{aligned}$$

The angular contribution at a crossing is the same as the sum of the angular contributions from the corners obtained by splicing that crossing, and using the generalized vertex weights for the corners.

In the R -matrix of section 11⁰ we have a correspondence between such splicings and the states as loops of constant spin. (Note that if $a = d$ and $b = c$ then $(d - a)\theta/2\pi = 0$.) Thus each loop receives the value (a = spin value of the loop)

$$\| \text{loop} \| = \frac{a}{2\pi} \times (\text{the sum of the angles at the corners})$$

$$= \frac{a}{2\pi} (\theta_1 + \theta_2 + \dots + \theta_n)$$



And this is exactly the evaluation we used in section 11⁰. This completes the verification that the models of section 11⁰ are special cases of this form of pl state model. It also shows how the pl state model can be construed as a vacuum-vacuum expectation, since we have already explained how the models of section 11⁰ can be cast in this form.

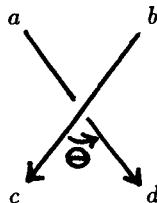
A Remark on the Unoriented Model.

We have seen in section 14⁰ how there arise naturally-state models for (specializations of) the Kauffman polynomial by “symmetrizing” the oriented state models for the Homfly polynomial. This led to a Yang-Baxter solution that was built from the $SL(n)$ Yang-Baxter solutions for the Homfly models. We saw how certain factors in the new Yang-Baxter solution were explained as rotation compensations in the construction of the unoriented model.

The piecewise linear approach using a rapidity parameter θ shows us the pattern of this tensor in a new light. Suppose we have

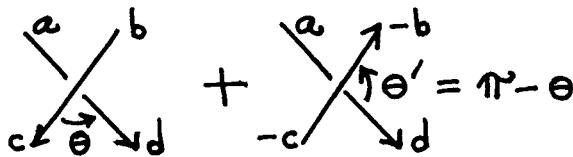
$$S_{cd}^{ab}(\theta) = R_{cd}^{ab} \lambda^{(d-a)\theta/2\pi}$$

corresponding to



and suppose that we build

$$T_{cd}^{ab}(\theta) \leftrightarrow$$



$$T_{cd}^{ab} = R_{cd}^{ab} \lambda^{\frac{(d-a)\theta}{2\pi}} + \bar{R}_{d-b}^{-c a} \lambda^{\frac{(-b+c)\theta'}{2\pi}}.$$

(We are using the convention that $\overset{a}{\rightarrow} \equiv \overset{-a}{\leftarrow}$.) Then

$$T_{cd}^{ab}(\theta) = R_{cd}^{ab} \lambda^{(d-a)\theta/2\pi} + \bar{R}_{d-b}^{-ca} \lambda^{(-b+c)(\pi-\theta)/2\pi}.$$

Here $a+b=c+d$ (i.e. we assume that R is spin-preserving). Hence $d-a=b-c$. Thus

$$\begin{aligned} T_{cd}^{ab}(\theta) &= R_{cd}^{ab} \lambda^{(d-a)\theta/2\pi} + \bar{R}_{d-b}^{-ca} \lambda^{(d-a)(\theta-\pi)/2\pi} \\ &= \left[R_{cd}^{ab} + \bar{R}_{d-b}^{-ca} \lambda^{(a-d)/2} \right] \lambda^{(d-a)(\theta/2\pi)}. \end{aligned}$$

Thus

$$T_{cd}^{ab}(\theta) = \hat{R}_{cd}^{ab} \lambda^{(d-a)\theta/2\pi}$$

where

$$\hat{R}_{cd}^{ab} = R_{cd}^{ab} + \bar{R}_{d-b}^{-ca} \lambda^{(a-d)/2}.$$

This tensor \hat{R} is exactly the tensor that we used in section 14⁰ with $R = (q - q^{-1}) \downarrow \leftarrow \zeta + q \downarrow \rightarrow = \zeta + \text{X}$, $\lambda = q$ and a special index set to produce our version of the Turaev models for the Kauffman polynomial. (\hat{R} is called M in section 14⁰.)

It is a good research problem to determine general conditions on R that will guarantee that \hat{R} is a solution to the Yang-Baxter Equation. With this problem we close the lecture!

16⁰. Three Manifold Invariants from the Jones Polynomial.

The purpose of this section is to explain how invariants of three dimensional manifolds can be constructed from the Jones polynomial. This approach, due to Reshetikhin and Turaev [RT2], instantiates invariants first proposed by Edward Witten [WIT2] in his landmark paper on quantum field theory and the Jones polynomial. (See section 17⁰ for a description of Witten's point of view.)

The approach of Reshetikhin and Turaev uses quantum groups to construct the invariants. The invariants themselves are averages of link polynomials - adjusted so that the resulting summation is unchanged under the Kirby moves [KIRB]. We use the theorem of Dehn and Lickorish [LICK1] that represents three-dimensional manifolds by surgery on framed links in the three-sphere. Kirby [KIRB] gave a set of moves that can be performed on such links so that the classification of three dimensional manifolds is reduced to the classification of framed links up to ambient isotopy augmented by these moves.

In fact, I shall concentrate on a version of the Reshetikhin-Turaev invariant for the case of the classical Jones polynomial. In this case, Lickorish [LICK3] has given a completely elementary proof of the existence of the invariant based upon the bracket polynomial (section 3⁰ and [LK4]) and certain properties of the Temperley-Lieb algebra in its diagrammatic form (as explained in section 7⁰). Thus the present section only requires knowledge of the bracket polynomial for the construction of the three-manifold invariant.

I begin by describing the Kirby Calculus and the formula for the invariant. We then backtrack to fill in background on the Temperley-Lieb algebra.

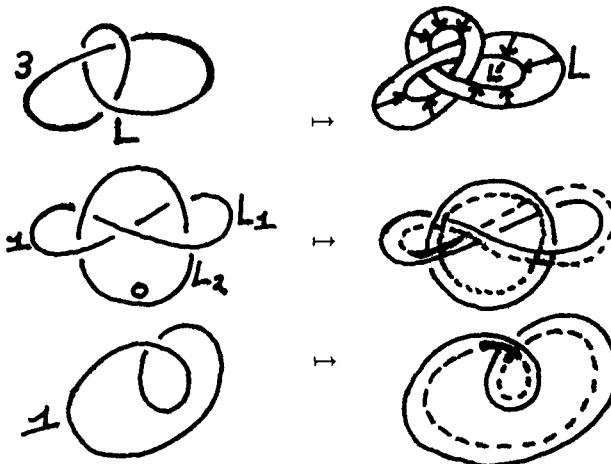
Framed Links and Kirby Calculus.

Consider a single curve K embedded in three-dimensional space. We may equip this curve K with a normal vector field, and if the lengths of the normal vectors are small, then their tips will trace out a second embedded curve - moving near the first curve and winding about it some integral number of times. If K' is that second curve, endowed with an orientation parallel to that of K , then the winding number is equal to the linking number $\ell k(K, K')$. This number $n = \ell k(K, K')$ is called the **framing number** of K . Conversely, if we assign an

integer n to K , then one can construct the normal vector field and curve K' so that $n = \ell k(K, K')$. The ambient isotopy class of K' in the complement of K is uniquely determined by this process.

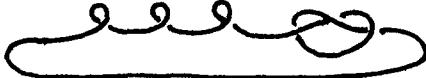
By definition a **framed link** L is a link L together with an assignment of integers, one to each component of L . These integers can be used to describe normal framings of the components as indicated above. In working with link diagrams there is a natural framing that is commonly referred to as the **blackboard framing**: Assign to each component L_i of L the number $n_i = w(L_i)$ where $w(L_i)$ denotes the writhing number of the diagram L_i . That is, $w(L_i)$ is the sum of the signs of the crossings of L_i . (Note that this sum is independent of the orientation assigned to L_i , and hence makes sense for unoriented links.)

A normal vector field that produces the blackboard framing is obtained by using the plane on which the link is drawn. That is, except for the crossings, the vectors lie in the plane, and hence the push-off components L'_i are obtained by drawing essentially parallel copies of the L_i – as shown below:



In each of these examples, I indicate the framing numbers on the link components, but note that via $w(L_i) = n_i$ these numbers are intrinsically determined by the diagram. In fact, any framing can be obtained as a blackboard framing since we can modify the writhe of a component by adding positive or negative curls. The

0-framed trefoil is the blackboard framing of the diagram



I denote a framing of the link L by the notation (L, f) where $f(L_i)$ is the integer assigned to the i -th component of L .

As we shall see, the study of three-dimensional manifolds is equivalent to a particular technical study of the properties of framed links. To be precise, Lickorish [LICK1] proved that every compact oriented three-dimensional manifold can be obtained via surgery on a framed link (L, f) . Kirby [KIRB] gave a set of moves on framed links and proved that two three-manifolds are homeomorphic if and only if the corresponding links can be transformed into one another by these moves. Our aim here is to describe this surgery and the moves, and to show how the Jones polynomial can be used to produce invariants of three dimensional manifolds.

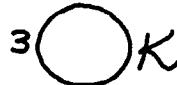
Surgery.

The basic idea of surgery is as follows. Suppose that we are given a 3-manifold M^3 bounding a four-dimensional manifold W^4 . Given an embedding $\alpha : S^1 \times D^2 \rightarrow M$ we can use this embedding to attach a handle, $D^2 \times D^2$, to W^4 . That is, the boundary of $D^2 \times D^2$ (D^2 is a two-dimensional disk) is the union of $S^1 \times D^2$ and $D^2 \times S^1$. We attach $S^1 \times D^2$ to M via α , and obtain a new four-manifold \overline{W}^4 with boundary \overline{M}^3 . One says that \overline{M}^3 is obtained via **surgery on M^3 along the framed link $\alpha(S^1 \times 0) \subset M$** . The framing is given by the twisting of the standard longitude $\alpha(S^1 \times 1) \subset M$. That is, we have $K = \alpha(S^1 \times 0)$, $K' = \alpha(S^1 \times 1)$ and the framing number is $n = \ell k(K, K')$. (This linking number happens in the solid torus $\alpha(S^1 \times D^2)$.)

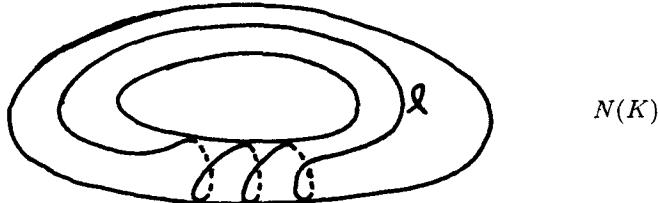
It is worthwhile having a direct description of \overline{M}^3 . It is easy to see from our description of handle-attachment that \overline{M}^3 is obtained from $M^3 - \text{Interior}(\alpha(S^1 \times D^2))$ by attaching $D^2 \times S^1$ along its boundary $S^1 \times S^1$ via the map α . This means that the twisting of the original **longitude** $S^1 \times 1 \subset S^1 \times D^2$ is now matched with a **meridian** $(S^1 \times 1) \subset D^2 \times S^1$ on the new attaching torus. In other words, **we cut-out the interior of $\alpha(S^1 \times D^2)$ and paste back a copy of $D^2 \times S^1$, matching the meridian of $D^2 \times S^1$ to the (twisted by framing**

number) longitude on the boundary torus in M .

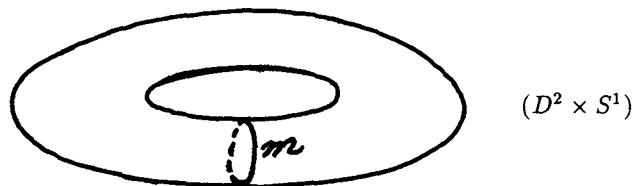
Example 1. $K = \alpha(S^1 \times 0) \subset S^3$, unknotted and framing number is 3.



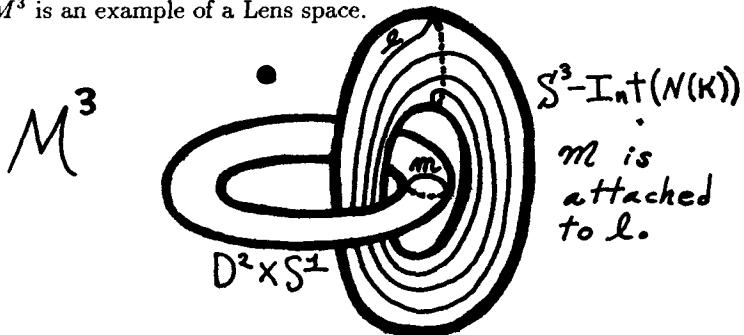
Let $N(K)$ denote $\alpha(S^1 \times D^2)$ and draw the longitude ℓ on the boundary of $N(K)$:
 $\ell = \alpha(S^1 \times 1)$



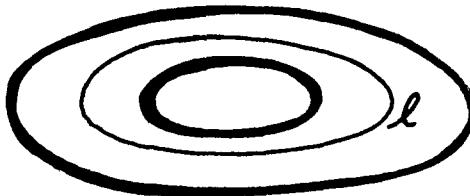
Let M^3 denote the manifold obtained by framed surgery along K with this framing. Then M is obtained from $S^3 - \text{Int}(N(K))$ by attaching $D^2 \times S^1$ so that the meridian $m = (D^2 \times 1)$ is attached to ℓ :



Since m bounds a disk in $(D^2 \times S^1)$, it is easy to see that M^3 has first homology group $\mathbf{Z}/3\mathbf{Z}$. M^3 is an example of a Lens space.



Example 2. Let this be the same as Example 1, but with framing number 0. Then ℓ is a simple longitude,



$$N(K)$$

and it is easy to see that M^3 is homeomorphic to $S^2 \times S^1$. (The longitude bounds a disk in $S^3 - \text{Int}(N(K))$ and m bounds a disk in $D^2 \times S^1$. With m and ℓ identified, we get a union of two disks forming a 2-sphere $S^2 \subset M^3$. The family of 2-spheres obtained by the family of longitudes $S^1 \times e^{i\theta}$ sweeps out M^3 , creating a homeomorphism of M^3 and $S^2 \times S^1$.)

Kirby Moves.

We now describe two operations on framed links such that $\partial W_L \simeq \partial W_{L'}$ if and only if we can pass from L to L' by a sequence of these operations. (Notation: ∂W^4 denotes the boundary of the four manifold W^4 . W_L^4 denotes the four manifold obtained from the 4-ball D^4 by doing handle attachments along the components of the framed link $L \subset S^3$.)

O₁. **Add or subtract from L an unknotted circle with framing 1 or -1. This circle is separated from the other components of L by an embedded S^2 in S^3 .**

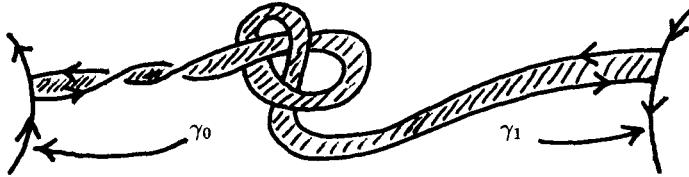
This operation corresponds in W_L to taking the connected sum with (or splitting off) a copy of the complex projective space \mathbf{CP}^2 with positive or negative orientation. ($\mathbf{CP}^2 = D^4 \cup E$ where $\partial E = S^3$ and $E \rightarrow S^2$ is the D^2 bundle associated with the Hopf map $H : S^3 \rightarrow S^2$ [HOP]. One can show that E is the result of adding a handle along an unknotted circle in S^3 with +1 framing. This is a translation of the well-known property of the Hopf map that for $p \neq q$ in S^2 , $H^{-1}\{p, q\}$ is the link .)

\mathcal{O}_2 . Given two components L_i and L_j in L , we “add” L_i to L_j as follows:

1. Push L_i off itself (missing L) using the framing f_i on L_i to obtain L'_i with $\ell k(L_i, L'_i) = f_i$.
2. Change L by replacing L_j with $\tilde{L}_j = L'_i \#_b L_j$ where b is any band missing the rest of L .

The band connected sum $L'_i \#_b L_j$ is defined as follows: Let γ_0, γ_1 be two knots in S^3 . Let $b : I \times I \rightarrow S^3$ be an embedding of $[0, 1] \times [0, 1]$ such that $b(I \times I) \cap \gamma_i = b(i \times I)$ for $i = 0, 1$. Then

$$\gamma_0 \#_b \gamma_1 = \gamma_0 \cup \gamma_1 - b(\partial I \times I) \cup b(I \times \partial I).$$



This second move corresponds in W_L to sliding the j -th handle over the i -th handle via the band b . In order to compute the framing \tilde{f}_j of \tilde{L}_j , consider the intersection form on the second homology group $H_2(W_L; \mathbf{Z})$. If we orient each L_k then they determine a basis \mathcal{B} for $H_2(W_L; \mathbf{Z})$ denoted by $\{[L_1], \dots, [L_r]\}$. The matrix A_L of the intersection form on $H_2(W_L; \mathbf{Z})$ in this basis has the framing numbers f_k down the diagonal, and entries $a_{ij} = \ell k(L_i, L_j)$ for $i \neq j$. The move \mathcal{O}_2 corresponds to either adding $[L_i]$ or subtracting $[L_i]$ from $[L_j]$, depending on whether the orientations on L_i and L_j correspond under b . Thus

$$\begin{aligned}\tilde{f}_j &= (\pm [L_i] + [L_j]) \cdot (\pm [L_i] + [L_j]) \\ &= [L_i] \cdot [L_i] + [L_i] \cdot [L_j] \bullet 2[L_i] \cdot [L_j].\end{aligned}$$

Hence

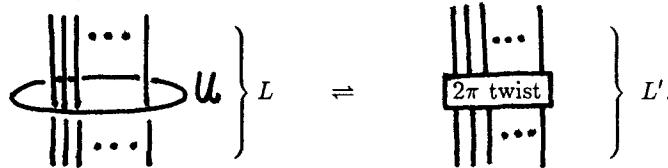
$$\boxed{\tilde{f}_j = f_i + f_j \pm 2a_{ij}}.$$

Call two framed links L and L' ∂ -equivalent if we can obtain L' from L by a sequence of the operations \mathcal{O}_1 and \mathcal{O}_2 (plus ambient isotopy). We write $L \sim_{\partial} L'$.

Theorem 16.1. (Kirby-Craggs). Given two framed links L and L' , then $L \underset{\partial}{\sim} L'$ if and only if ∂W_L is diffeomorphic to $\partial W_{L'}$ (preserving orientations). (Diffeomorphic and piecewise linearly homeomorphic are equivalent in this dimension.)

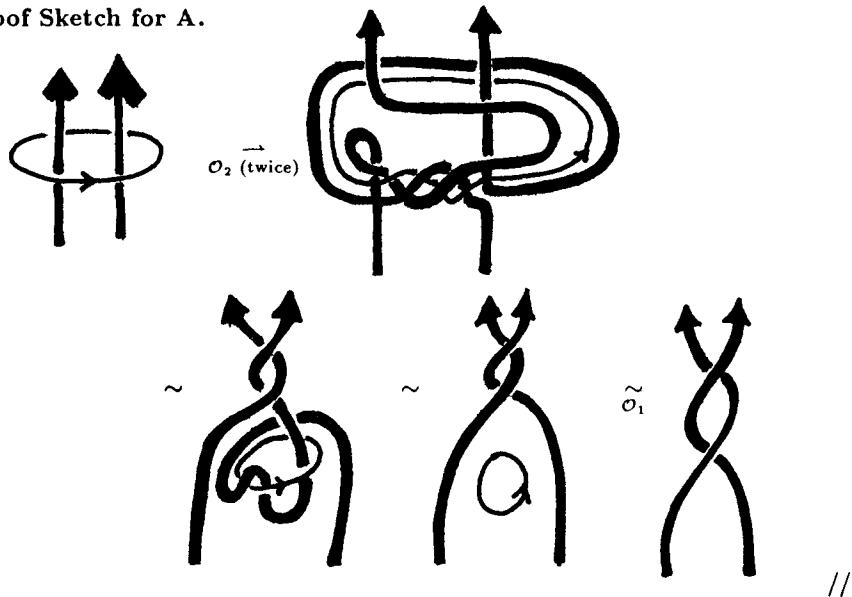
The next proposition gives a fundamental consequence of the Kirby moves:
A cable of lines encircled by an unknotted loop of framing ± 1 is ∂ -equivalent to a twisted cable without this loop.

Proposition A. Let L and L' be identical except for the parts shown in the figure below. Then $L \underset{\partial}{\sim} L'$. Here U is an unknot with framing -1 which disappears in L' , and the box in L' denotes a full (2π) right handed twist. The framing on L'_i is given by $f'_i = f_i + 1$.



Proposition B. Same as Proposition A for U with framing $+1$. The box then denotes one full left hand twist and $f'_i = f_i - 1$.

Proof Sketch for A.



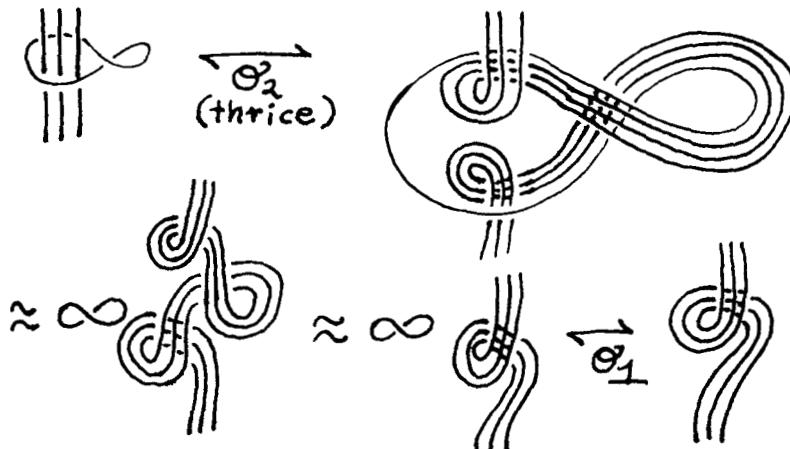
Since this proposition is so important, it is worthwhile discussing the proof sketch once again in the language of blackboard framings and regular isotopy of diagrams. In blackboard framing we have $f_i = w(L_i)$, the writhe of the i -th component. The equation for change of framing under operation \mathcal{O}_2 is (as remarked above) $\tilde{f}_j = f_i + f_j \pm 2a_{ij}$. Hence we require

$$w(\tilde{L}_j) = w(L_i) + w(L_j) \pm 2\ell k(L_i, L_j)$$

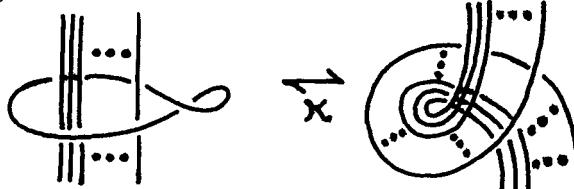
and this equation is true whenever the band consists of parallel strands that contribute no extra writhe or linking! Thus we can express the Kirby calculus in terms of the blackboard framing and regular isotopy of diagrams by simply stipulating that band-sums are to be replaced by recombinations:



Note that the parallel push-off from L_i creates a copy of L_i which then undergoes recombination with L_j to form \tilde{L}_j . Processes of reproduction and recombination generate the underpinning of the Kirby calculus. [These formal similarities with processes of molecular biology deserve further study.] Now, in blackboard framing, let's return to the proof-sketch using, for illustration, a triple strand:



The generalization to n -strands is the move κ illustrated below.

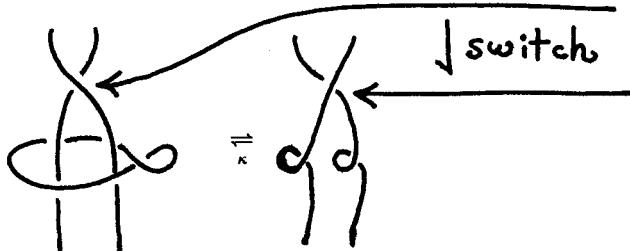


The 2π -twist is replaced by a flat curl on the cable of parallel strands.

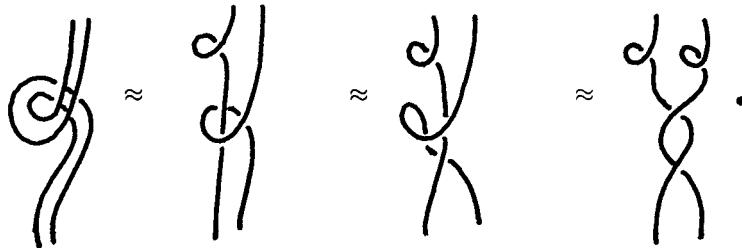
We now move on to the beautiful theorem of Fenn and Rourke [FR]. They prove that the move \mathcal{O}_1 together with the move κ of Proposition A/B generate the Kirby calculus. This means that the move κ (above) (and its mirror image) together with blackboard \mathcal{O}_1 ($\infty \rightleftharpoons$ (nothing)) generate \mathcal{O}_2 , and hence the entire Kirby calculus.

In order to see why we can generate move \mathcal{O}_2 using only \mathcal{O}_1 and the κ -move, we first observe that κ lets us switch a crossing:

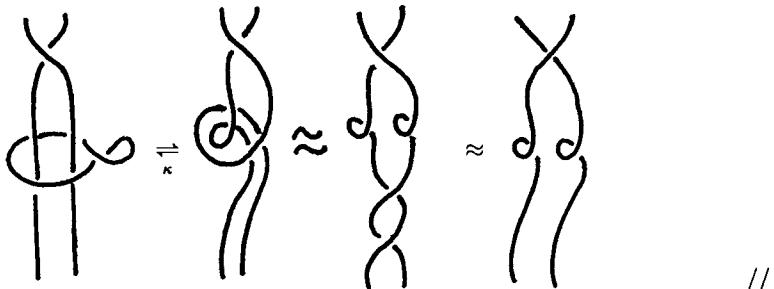
Lemma 16.2.



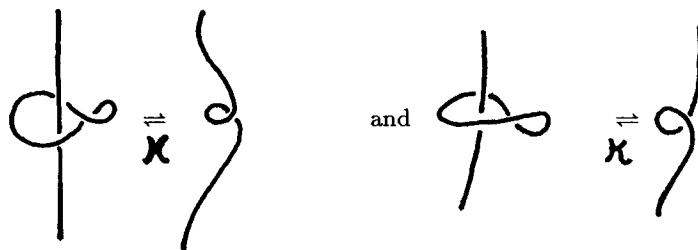
Proof. First note that



Thus

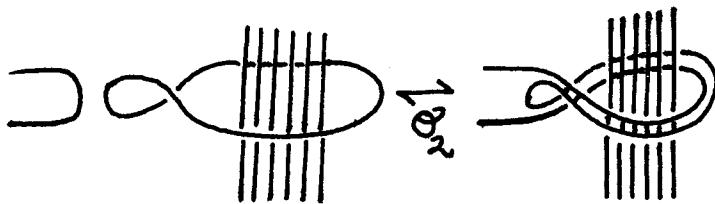


This lemma implies that we may take all the components in our link L to be **unknotted**. Furthermore, since



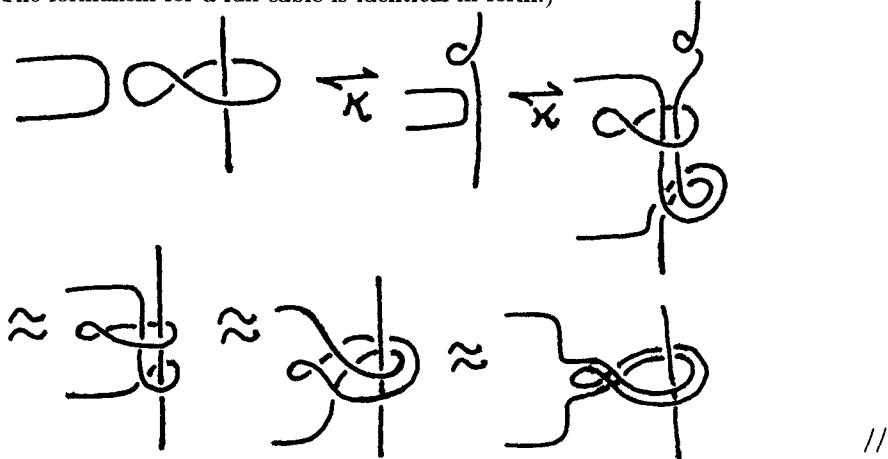
we can change the writhe of a single component until it is $+1$ or -1 or 0 .

Thus it will suffice to use κ and \mathcal{O}_1 to obtain \mathcal{O}_2 for the situation.

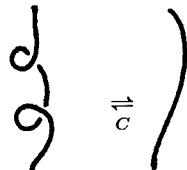


Here is a demonstration of this maneuver: (I replace the cable by a single strand.

The formalism for a full cable is identical in form.)



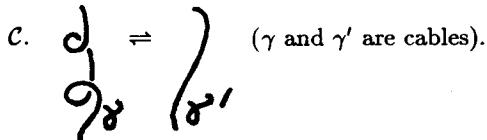
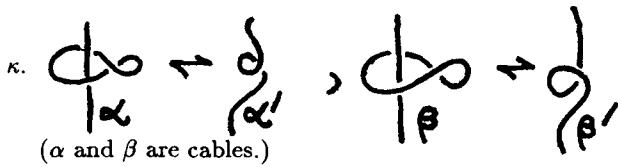
This completes our diagrammatic description of Kirby calculus. However, we must add one more curl-move:



With this move added, two unknotted curves are equivalent if and only if they have the same writhe (as desired). Thus our calculus is generated by

O_0 . regular isotopy

O_1 . $\infty \Rightarrow (\text{blank}) \Leftarrow \infty$



It is an interesting exercise to verify that the β -version of κ (above) follows from \mathcal{O}_0 , $\mathcal{O}_1, \mathcal{C}$ and the α -version of κ . This gives us a minimal set of moves for investigating invariance.

Construction of the Three-Manifold Invariant.

We wish to use the bracket polynomial (which is an invariant of regular isotopy) and create from it a new regular isotopy invariant that is also invariant under the κ -move and the \mathcal{C} -move (a suitable normalization will take care of \mathcal{O}_1).

Since we must examine the behavior of the bracket on cables, it is best to take care of the \mathcal{C} -move first.

Proposition 16.3. Let  denote an i -strand parallel cable. Let $\langle K \rangle$ denote the bracket polynomial of section 3⁰. Then

$$\left\langle \begin{array}{c} \text{---} \\ | \\ \text{---} \end{array} \right\rangle = \left\langle \begin{array}{c} \text{---} \\ | \\ \text{---} \end{array} \right\rangle.$$

Hence the bracket polynomial itself satisfies move C .

Remark. Recall that the bracket is determined by the equations

$$\begin{aligned}\langle \infty \rangle &= A \langle \approx \rangle + A^{-1} \langle \circlearrowleft C \rangle \\ \langle OK \rangle &= (-A^2 - A^{-2}) \langle K \rangle \\ \langle O \rangle &= -A^2 - A^{-2} = \delta\end{aligned}$$

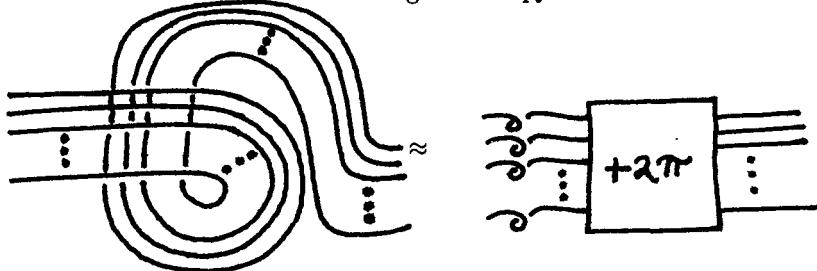
and that it is an invariant of regular isotopy of unoriented link diagrams. It follows directly from these defining equations that (for single strands)

$$\left\langle \begin{array}{c} \text{---} \\ | \\ \text{---} \end{array} \right\rangle = \left\langle \begin{array}{c} \text{---} \\ | \\ \text{---} \end{array} \right\rangle = (-A^3) \left\langle \begin{array}{c} \text{---} \\ | \\ \text{---} \end{array} \right\rangle$$

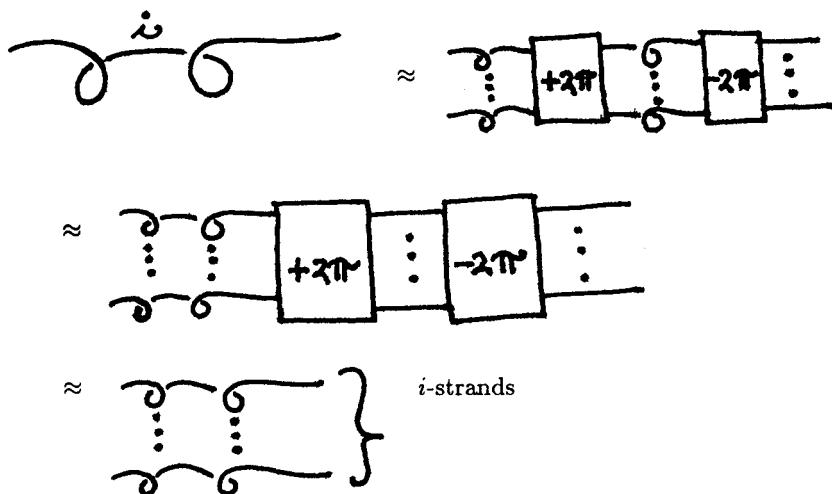
and

$$\left\langle \begin{array}{c} \text{---} \\ | \\ \text{---} \end{array} \right\rangle = \left\langle \begin{array}{c} \text{---} \\ | \\ \text{---} \end{array} \right\rangle = (-A^{-3}) \left\langle \begin{array}{c} \text{---} \\ | \\ \text{---} \end{array} \right\rangle.$$

Proof. First recall that we have the regular isotopy



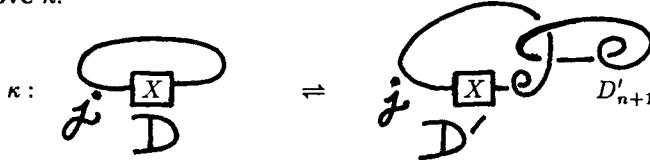
Thus we have the consequent regular isotopy



Since $\langle \text{---} \text{---} \rangle = \langle \sim \rangle$ on single strands, and $\langle K \rangle$ is an invariant of regular isotopy, it follows at once that $\langle \text{---} \text{---} \text{---} \text{---} \rangle = \langle \sim \sim \sim \sim \rangle$ as desired. //

As a result of Proposition 16.3, we can forget about move C and concentrate on making some combination of bracket evaluations invariant under the **basic**

Kirby move κ :



(This is just a rewrite of our previous version of κ . The j refers to a cable of j strands. If D has n components, then D' has $(n + 1)$ components and D'_{n+1} is the new component.)

As it stands, the bracket will **not** be invariant under κ , but we shall consider the bracket values on cables

$$\hat{T}_{i+j} = \text{Diagram with } i+j \text{ strands} \approx \text{Diagram with } i \text{ strands and a cable } j.$$

Let $T_{i+j} = \langle \hat{T}_{i+j} \rangle$. Thus

$$T_1 = \langle \text{Diagram with } 2 \text{ strands} \rangle = -A^3$$

$$T_2 = \langle \text{Diagram with } 3 \text{ strands} \rangle = \langle \text{Diagram with } 2 \text{ strands and a cable } 1 \rangle$$

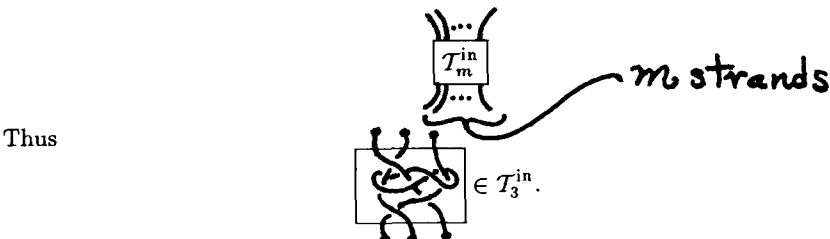
$$= (-A^3)^2 \langle \text{Diagram with } 2 \text{ strands} \rangle$$

$$= +A^6(-A^4 - A^{-4})$$

$$T_2 = -A^{10} - A^2.$$

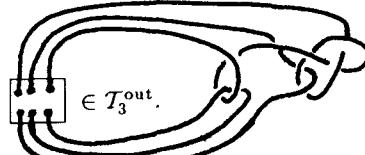
Of course, $T_0 = \langle \text{empty link} \rangle = 1/\delta$.

We are now in position to state the basic technique for obtaining invariance under the Kirby move κ . First consider two species of tangles that we shall denote by T_m^{in} and T_m^{out} . T_m^{in} consists in tangles with m inputs and m outputs so that the tangle is confined to the **interior** of a given tangle-box in the plane.



Thus

Similarly, T_m^{out} consists of tangles restricted to the **outside** of this same box. For example,



We have a pairing

$$\langle \ , \ \rangle : T_m^{\text{in}} \times T_m^{\text{out}} \rightarrow \mathbf{Z}[A, A^{-1}]$$

defined by the formula

$$\langle x, y \rangle = \langle xy \rangle.$$

The second bracket denotes the regular isotopy invariant bracket polynomial, and the product xy denotes the link that results from attaching the free ends of $x \in T_m^{\text{in}}$ to the corresponding free ends of $y \in T_m^{\text{out}}$.

In this context it is convenient to consider formal linear combinations of elements in T_m^{in} and T_m^{out} (with coefficients in $\mathbf{Z}[A, A^{-1}]$). We can then write (by definition)

$$\langle X, aY + bZ \rangle = a\langle X, Y \rangle + b\langle X, Z \rangle$$

and

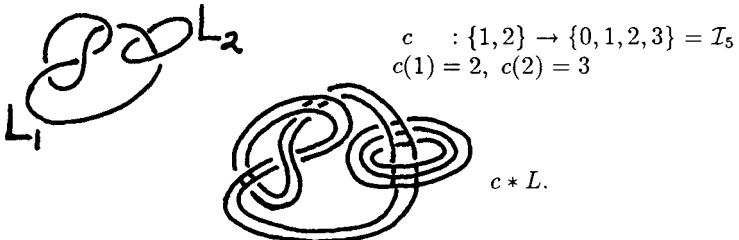
$$\langle aX + bW, Z \rangle = a\langle X, Z \rangle + b\langle W, Z \rangle$$

when $a, b \in \mathbf{Z}[A, A^{-1}]$, $X, W \in T_m^{\text{in}}$, $Y, Z \in T_m^{\text{out}}$. This makes $\langle \ , \ \rangle$ a bilinear form on these modules of linear combinations. In the following discussion, I shall let T_m^{in} and T_m^{out} denote this extension of tangles to linear combinations of tangles.

With the help of this language, we shall now investigate when the following form of regular isotopy invariant, denoted $\ll K \gg$, is also invariant under the Kirby κ -move:

Let \mathcal{I}_r be a fixed finite index set of the form $\mathcal{I}_r = \{0, 1, 2, \dots, r-2\}$ for $r \geq 3$ an integer. (The convenience of $r-2$ will appear later.) Let $\mathcal{C}(n, r)$ denote the set of all functions $c : \{1, 2, \dots, n\} \rightarrow \mathcal{I}_r$ where $n \geq 1$ is an integer. If $\{1, 2, \dots, n\}$ are the names of the components of a link L , then $c(1), c(2), \dots, c(n)$ assign labels to these components from the set \mathcal{I}_r .

Given an n -component link diagram L with components L_1, L_2, \dots, L_n and given $c \in \mathcal{C}(n, r)$, let $c * L$ denote the diagram obtained from L by replacing L_i by c_i parallel planar copies of L_i . Thus



Let $\{\lambda_0, \lambda_1, \dots, \lambda_{r-2}\}$ be a fixed set of scalars. (Read scalar as an element of the complex numbers \mathbf{C} - assuming we are evaluating knot polynomials at complex numbers.) Define $\ll L \gg$ via the formula

$$\begin{aligned}\ll L \gg &= \ll L \gg(A, r, \lambda_0, \dots, \lambda_{r-2}) \\ \ll L \gg &= \sum_{c \in \mathcal{C}(n, r)} \lambda_{c(1)} \lambda_{c(2)} \dots \lambda_{c(n)} \langle c * L \rangle.\end{aligned}$$

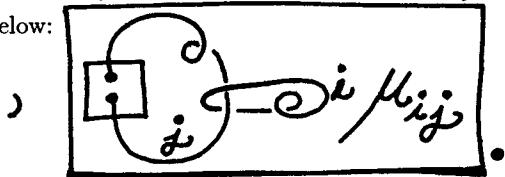
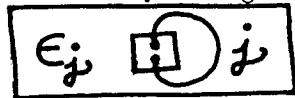
Thus $\ll L \gg$ is a weighted average of bracket evaluations $\langle c * L \rangle$ for all the blackboard cables that can be obtained from L by using \mathcal{I}_r . By using the blackboard framing, as described in this section, we can regard $\ll L \gg$ as a regular isotopy invariant of framed links.

We write the Kirby move in a form that can be expressed in products of tangles:

$$\begin{array}{ccc} \text{Diagram } L & \xrightleftharpoons[\kappa]{j} & \text{Diagram } L'_{n+1} \end{array}$$

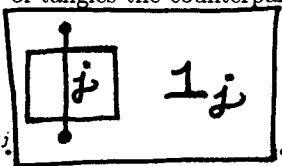
(L has n -components.)

Thus we have, in tangle language, $L = X\epsilon_j$, $L' = X\mu_{ij}$ where the tangles ϵ_j and μ_{ij} are defined by the diagrams below:



Thus $\epsilon_j \in T_j^{\text{out}}$ and $\mu_{ij} \in T_j^{\text{out}}$.

I add to this rogue's gallery of tangles the counterpart of ϵ_j in T_j^{in} , and call it 1_j :



Note that $\langle 1_j, \epsilon_j \rangle = \langle 1_j \epsilon_j \rangle = \delta_j^j$.

We can now examine the behavior of $\ll L \gg$ under the Kirby move:

Hypothesis \mathcal{H} . This hypothesis assumes that for all j and for all $x \in T_j^{\text{in}}$

$$\langle x, \epsilon_j \rangle = \left\langle x, \sum_{i=0}^{r-2} \lambda_i \mu_{ij} \right\rangle.$$

We can write hypothesis \mathcal{H} symbolically by

$$\mathcal{H}: \quad \boxed{j} \cong \sum_{i=0}^{r-2} \lambda_i \boxed{\mu_{ij}} .$$

Proposition 16.4. Under the assumption of hypothesis \mathcal{H} :

(i) For all j , $\delta^j = \sum_{i=0}^{r-2} \lambda_i T_{i+j}$ where $T_{i+j} = \langle \boxed{\mu_{ij}} \rangle$.

(ii) $\ll L \gg$ is invariant under the Kirby move κ , and

$$\ll L \amalg \textcircled{0} \gg = \tau \ll L \gg$$

(τ denotes $\sum_{i=0}^r \lambda_i \bar{T}_i$ where $\bar{T}_i = \langle \boxed{\mu_{ij}} \rangle$ and II denotes disjoint union.)

(iii) $\tau^{(\sigma+\nu-n)/2} \ll L \gg$ is an invariant of the 3-manifold obtained by framed surgery on L . Here σ and ν are the signature and nullity of the linking matrix associated with the n -component link L .

Proof. In this context, let $L' = x\mu_{1y}$ and $L = x\epsilon_j$, as explained above. Also, given $c : \{1, \dots, n\} \rightarrow \mathcal{I}_r$, let $c * x$ denote the result of replacing the components of the tangle x by $c(i)$ parallels for the i -th component. Then if $x \in \mathcal{T}_j^{\text{in}}$ we have $c * x \in \mathcal{T}_{j'}^{\text{in}}$ where j' is the sum of $c(K)$ where K runs over the components of x that share input or output lines on the tangle. If we need to denote dependency on c , I shall write $j' = j(c)$. For $c \in \mathcal{C}(n, r)$ I shall abbreviate λ_c for $\lambda_{c(1)}\lambda_{c(2)} \dots \lambda_{c(n)}$.

$$\begin{aligned}
\ll L' \gg &= \ll x\mu_{1j} \gg \\
&= \sum_{c \in \mathcal{C}(n+1, r)} \lambda_{c(1)} \dots \lambda_{c(n)} \lambda_{c(n+1)} \langle c * [x\mu_{1j}] \rangle \\
&= \sum_{c \in \mathcal{C}(n, r)} \lambda_{\bar{c}} \sum_{i=0}^{r-2} \lambda_i \langle (c * x)(\mu_{ij}(c)) \rangle \\
&= \sum_{c \in \mathcal{C}(n, r)} \lambda_{\bar{c}} \left\langle c * x, \sum_{i=0}^{r-2} \lambda_i \mu_{ij(c)} \right\rangle \\
&= \sum_{c \in \mathcal{C}(n, r)} \lambda_{\bar{c}} \langle c * x, \epsilon_j(c) \rangle \quad (\text{hypothesis } \mathcal{H}) \\
&= \sum_{c \in \mathcal{C}(n, r)} \lambda_{\bar{c}} \langle c * (x\epsilon_j) \rangle \\
&= \sum_{c \in \mathcal{C}(n, r)} \lambda_{\bar{c}} \langle c * L \rangle \\
&= \ll L \gg.
\end{aligned}$$

This demonstrates the first part of (ii). To see the second part, let

$$L' = L \amalg \text{S}$$

Then

$$\begin{aligned}
 \ll L' \gg &= \sum_{c \in \mathcal{C}(n, r)} \lambda_c^{-} \sum_{i=0}^{r-2} \lambda_i \langle c * L \text{ II } \text{ (double circle) } i \rangle \\
 &= \sum_{c \in \mathcal{C}(n, r)} \lambda_c^{-} \sum_{i=0}^{r-2} \lambda_i \langle c * L \rangle \langle \text{ (double circle) } i \rangle \\
 &= \left[\sum_{i=0}^{r-2} \lambda_i \langle \text{ (double circle) } i \rangle \right] \sum_{c \in \mathcal{C}(n, r)} \lambda_c^{-} \langle c * L \rangle \\
 \ll L' \gg &= \tau \ll L \gg .
 \end{aligned}$$

Now, turn to part (i): We have $T_{i+j} = \langle 1_j \mu_{ij} \rangle$. Therefore,

$$\begin{aligned}
 \delta^j &= \langle 1_j \epsilon_j \rangle = \langle 1_j, \epsilon_j \rangle \\
 &= \left\langle 1_j, \sum_{i=0}^{r-2} \lambda_i \mu_{ij} \right\rangle \quad (\text{by } \mathcal{H}) \\
 \therefore \delta^j &= \sum_{i=0}^{r-2} \lambda_i T_{i+j}.
 \end{aligned}$$

Finally, to prove (iii) note that the $n \times n$ linking matrix of L has signature σ and nullity ν , so $\frac{1}{2}(n - \sigma - \nu)$ is the number of negative entries in a diagonalization of the matrix. That number is unchanged under the κ -move and increases by one when we add . Hence (iii) follows directly from (ii). This completes the proof of the proposition. //

Now examine (i) of the above proposition. The set of equations

$$\delta^j = \sum_{i=0}^{r-2} \lambda_i T_{i+j} \quad j = 0, 1, 2, \dots$$

is an apparently overdetermined system. Nevertheless, as we shall see for the bracket, they can be solved when $A = e^{\pi i / 2r}$ and the result is a set of λ 's satisfying hypothesis \mathcal{H} and giving the desired three-manifold invariant. Before proceeding to the general theory behind this, we can do the example for $r = 3$ (and 4) explicitly. However, it is important to note that our tangle modules can be factored by

relations corresponding to the generating formulas for the bracket:

$$\begin{aligned}\langle \text{X} \rangle &= A \langle \text{C} \rangle + A^{-1} \langle \text{DC} \rangle \\ \langle \text{LO} \rangle &= \delta \langle L \rangle.\end{aligned}$$

This means that we can replace the tangle modules T_m^{in} and T_m^{out} by their quotients under the equivalence on tangle diagrams generated by

$$(i) \text{X} = A \text{C} + A^{-1} \text{DC}$$

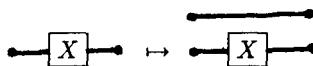
$$(ii) \text{OD} = \delta D.$$

We shall do this and use the notation V_m^{in} and V_m^{out} for these quotient modules.

With this convention, and assuming that A has been given a value in \mathbb{C} , we see that V_m^{in} and V_m^{out} are finitely generated over \mathbb{C} . For example, V_2^{in} has generators



Note that V_m^{in} and V_m^{out} are abstractly isomorphic modules. I shall use V_m for the corresponding abstract module and use pictures written horizontally to illustrate V_m . Thus V_2 has generators $=$ and DC (1 and e_1). We map $V_m \hookrightarrow V_{m+1}$ by adding a string on the top:



V_3 is generated by

$$\mathcal{B} = \left\{ \begin{array}{ccccc} = & \text{DC} & \text{DC} & \text{DC} & \text{DC} \\ 1 & e_1 & e_2 & \alpha & \beta \end{array} \right\}$$

That is, every element of V_3 is a linear combination of these elements. Under tangle multiplication V_m becomes an algebra over \mathbb{C} . This is the Temperley-Lieb algebra. It first arose, in a different form, in the statistical mechanics

of the Potts model (see [BA1]). This diagrammatic interpretation is due to the author ([LK4], [LK8], [LK10]) and is related to the way Jones constructed his V -polynomial [JO2] as we have explained in section 7⁰. Multiplicatively, V_m is generated by the elements $\{1, e_1, e_2, \dots, e_{m-1}\}$ where the e_i 's are obtained by hooking together the i -th and $(i+1)$ -th points on each end of the tangle:

$$\left\{ \begin{array}{c} \text{---} \\ \vdots \\ \text{---} \\ \text{---} \end{array} \right. \left. \begin{array}{c} \text{---} \\ \vdots \\ \text{---} \\ \text{---} \end{array} \right\} \leftrightarrow \{e_i\}.$$

It is easy to see that

$$\left\{ \begin{array}{ll} e_i^2 & = \delta e_i, & i = 1, \dots, m-1 \\ e_i e_{i+1} e_i & = e_i, & i = 1, \dots, m-2 \\ e_i e_{i-1} e_i & = e_i, & i = 0, 1, \dots, m-1 \\ e_i e_j & = e_j e_i, & |i-j| > 1 \end{array} \right\} \quad (*)$$

and one can take V_m as the algebra over \mathbf{C} with these relations (δ has a chosen complex value for the purpose of discussion).

Remark. In Jones' algebra $e_i^2 = e_i$ while $e_i e_{i\pm 1} e_i = \delta^{-1} e_i$. This is obtained from the version given here by replacing e_i by $\delta^{-1} e_i$.

In V_3 we have

$$\begin{aligned} e_1^2 &= \overbrace{\text{---}}^1 \text{---} = 0 \text{---} = \delta e_1 \\ e_1 e_2 e_1 &= \overbrace{\text{---}}^1 \overbrace{\text{---}}^2 \overbrace{\text{---}}^1 \approx \overbrace{\text{---}}^2 \text{---} = e_1. \end{aligned}$$

Note also that in V_3 ,

$$\begin{aligned} e_1 e_2 &= \overbrace{\text{---}}^1 \overbrace{\text{---}}^2 = \alpha \\ e_2 e_1 &= \overbrace{\text{---}}^2 \overbrace{\text{---}}^1 = \beta. \end{aligned}$$

We have been concerned with the pairing $\langle , \rangle : T_m^{\text{in}} \times T_m^{\text{out}} \rightarrow \mathbf{C}$, $\langle x, y \rangle = \langle xy \rangle$. With respect to the Temperley-Lieb algebra this pairing can be written

$$\begin{aligned} \langle , \rangle &= V_m \times V_m \rightarrow \mathbf{C} \\ \langle x, y \rangle &= \langle xy \rangle \end{aligned}$$

where it is understood that x is identified with the corresponding element of T_m^{in} and \hat{y} , with the corresponding element of T_m^{out} . Thus

$$\langle \alpha, \beta \rangle = \left\langle \begin{array}{c} \text{---} \\ \text{---} \end{array}, \begin{array}{c} \text{---} \\ \text{---} \end{array} \right\rangle$$

$$= \left\langle \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{array} \right\rangle = \delta^3.$$

In general,

$$\langle \begin{array}{c} \text{---} \\ \text{---} \end{array}, \begin{array}{c} \text{---} \\ \text{---} \end{array} \rangle = \left\langle \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{array} \right\rangle$$

$$= \left\langle \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{array} \right\rangle.$$

Thus, we can identify $x\hat{y}$ as the link obtained from the tangle product xy by closing with an identity tangle.

Example. The matrix for the pairing $V_2 \times V_2 \rightarrow \mathbf{C}$ is determined by

$$\begin{array}{c|cc} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} \\ \text{---} & \text{---} & \text{---} \end{array} \leftrightarrow \begin{array}{c|c|c} & \delta^2 & \delta \\ \hline \delta & \delta & \delta^2 \end{array}$$

Note that this matrix is **singular** (for $\delta \neq 0$) when $\delta^4 - \delta^2 = 0 \Leftrightarrow \delta^2 - 1 = 0 \Leftrightarrow (-A^2 - A^{-2})^2 = 1 \Leftrightarrow A^4 + A^{-4} + 2 = 1$. Thus, we have \langle , \rangle singular when $A = e^{i\pi/2r}$ for $r = 3$. For this same $r = 3$, we can consider the equations

$$\delta^j = \sum_{i=0}^{r-2} \lambda_i T_{i+j}$$

first for $j = 0, \dots, r-2$ and then for arbitrary j . For $j = 0, 1$, the system reads:

$$1 = \lambda_0 T_0 + \lambda_1 T_1$$

$$\delta = \lambda_0 T_1 + \lambda_1 T_2$$

and we know that

$$T_0 = 1/\delta$$

$$T_1 = -A^3$$

$$T_2 = -A^{10} - A^2.$$

Thus

$$\begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} = \frac{1}{T_0 T_2 - T_1^2} \begin{bmatrix} T_2 & -T_1 \\ -T_1 & 1/\delta \end{bmatrix} \begin{bmatrix} 1 \\ \delta \end{bmatrix}.$$

If $A = e^{i\pi/6}$ then

$$\begin{aligned} T_0 T_2 - T_1^2 &= (A^{10} + A^2)/2 + 1 \\ &= (-A^4 + A^2)/2 + 1 \\ &= (1/2) + 1 \\ &= 3/2 \\ &\neq 0. \end{aligned}$$

Hence we can solve for λ_0, λ_1 . While it may not be obvious that these λ 's provide a solution to the infinite system $\delta^j = \lambda_0 T_j + \lambda_1 T_{j+1} \quad j = 0, 1, 2, \dots$, this is in fact the case! The singularity of the pairing \langle , \rangle at $e^{\pi i/2r}$ is crucial, as we shall see.

Remark. The Temperley-Lieb algebra V_m has the Catalan number

$d(m) = (\frac{1}{m+1}) \binom{2m}{m}$ dimension as a vector space over \mathbf{C} ; and for generic δ , $d(m)$ is the dimension of V_m over \mathbf{C} . These generators can be described diagrammatically as the result of connecting two parallel rows of m points so that all connecting arcs go between the rows, points on the same row can be connected to each other and no connecting arcs intersect one another. Thus for $m = 3$, we have



the familiar basis for V_3 . For convenience, I shall refer to the non-identity generators of V_m : $\{e_1, e_2, \dots, e_{m-1}, e_m, \dots, e_{d(m)}\}$ where $e_m, \dots, e_{d(m)}$ is a choice of labels for the remaining elements beyond the standard multiplicative generators e_1, \dots, e_{m-1} . This notation will be used in the arguments to follow.

Theorem 16.5. Let $A = e^{\pi i/2r}$, $r = 3, 4, 5, \dots$. Let V_m denote the Temperley-Lieb algebra for this value of $\delta = -A^2 - A^{-2}$. (Hence $\delta = -2 \cos(\pi/r)$.) Then there exists a unique solution to the system $\delta^j = \sum_{i=0}^{r-2} \lambda_i T_{i+j}$, $j = 0, 1, 2, \dots$

In fact

- (i) The matrix $M = (M_{ij})$ with $M_{ij} = T_{i+j-2}$, $1 \leq i, j \leq r$ is nonsingular for $A = e^{\pi i/2r}$.

and

- (ii) The pairing $\langle \cdot, \cdot \rangle : V_m \times V_m \rightarrow \mathbf{C}$ is degenerate for $A = e^{\pi i/2r}$ and $m \geq r-1$.

In particular there exists a dependence so that $\langle \cdot, m \rangle$ is a linear combination of pairings with the nonidentity elements e_i , $i = 1, \dots, d(m)$.

Conditions (i) and (ii) imply the first statement of the theorem.

This theorem plus Proposition 16.4 implies an infinite set of 3-manifold invariants, one for each value of $r = 3, 4, 5, \dots$. Before proving the theorem, I will show with two lemmas that conditions (i) and (ii) do imply that $\lambda_0, \dots, \lambda_{r-2}$ solve the infinite system

$$\delta^j = \sum_{i=0}^{r-2} \lambda_i T_{i+j}.$$

This makes it possible for the reader to stop and directly verify conditions (i) and (ii) for small values of r ($r = 3, 4, 5, 6$) if she so desires.

Assuming condition (i) we know that there exists a unique system $\lambda_0, \dots, \lambda_{r-2}$ to the restricted system of equations

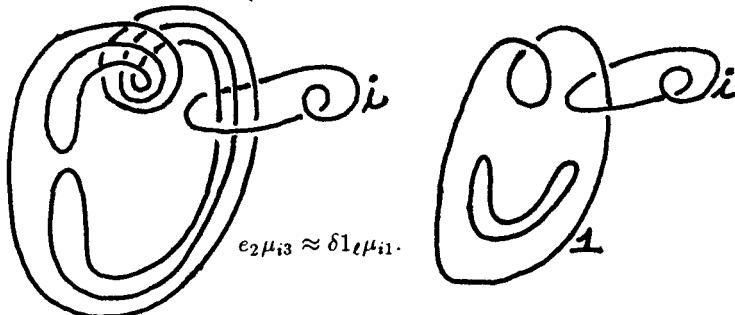
$$\delta^j = \sum_{i=0}^{r-2} \lambda_i T_{i+j}, \quad j = 0, 1, \dots, r-2.$$

Let $\phi_j = \sum_{i=0}^{r-2} \lambda_i \mu_{ij} \in \mathcal{T}_j^{\text{out}}$ for any j . Thus $T_{i+j} = \langle 1_j \mu_{ij} \rangle = \langle 1_j, \mu_{ij} \rangle$, and therefore $\delta^j = \langle 1_j, \phi_j \rangle$ for $j = 0, 1, \dots, r-2$. We wish to show that this last equation holds for all j .

Lemma 16.6. $\langle x, \phi_j \rangle = \langle x, \epsilon_j \rangle$ for all $x \in V_j^{\text{in}}$, $j = 0, 1, \dots, r-2$.

Proof. Since $\langle 1_j, \epsilon_j \rangle = \delta^j$, the lemma is true for $x = 1_j$ by condition (i). Let e_k be any other generating element of V_j^{in} . (V_j^{in} is T_j^{in} modulo the Temperley-Lieb relations.) e.g.

$$e_k = \begin{array}{c} \text{Diagram of } e_k \text{ as a planar graph with vertices and edges} \\ \dots \quad \dots \end{array} \Rightarrow \langle e_k, \mu_{ij} \rangle = \langle 1_\ell, \mu_{ij} \rangle \delta^s \text{ with } \ell < j.$$



By this basic annular isotopy, we conclude that $\langle e_k, \mu_{ij} \rangle = \delta^\ell \delta^s = \delta^j$ for each nonidentity Temperley-Lieb generator. The lemma follows at once from this observation. //

Lemma 16.7. Assuming conditions (i) and (ii) as described in the statement of (16.5), $\langle x, \epsilon_j \rangle = \langle x, \phi_j \rangle$ for all $x \in V_j^{\text{in}}$ and for all $j = 0, 1, 2, \dots$.

Proof. The proof is by induction on j . We assume that 16.7 is true for all integers less than j . By 16.6 we can assume that $j > r - 2$, since the hypothesis of induction is satisfied at least up to $r - 2$. By condition (ii) we can assume that there are coefficients α_k so that

$$\langle 1_j, - \rangle = \sum_{k=1}^{d(j)} \langle \alpha_k e_k, - \rangle$$

on V_j^{out} . Hence

$$\begin{aligned}
 \langle 1_j, \phi_j \rangle &= \sum_{k=1}^{d(j)} \alpha_k \langle e_k, \phi_j \rangle \\
 &= \sum_{k=1}^{d(j)} \alpha_k \delta^{\ell(\kappa)} \langle 1_{j(\kappa)}, \phi_{j(\kappa)} \rangle \quad (\text{annular isotopy}) \\
 &= \sum_{k=1}^{d(j)} \alpha_k \delta^{\ell(\kappa)} \langle 1_{j(\kappa)}, \epsilon_{j(\kappa)} \rangle \quad (\text{induction}) \\
 &= \sum_{k=1}^{d(j)} \alpha_k \delta^{\ell(\kappa)} \delta^{j(\kappa)} \\
 &= \sum_{k=1}^{d(j)} \alpha_k \langle e_k, \epsilon_j \rangle \\
 \therefore \langle 1_j, \phi_j \rangle &= \langle 1_j, \epsilon_j \rangle.
 \end{aligned}$$

By using the annular isotopy for the other generators of V_j^{in} we complete this argument by induction. Hence $\langle x, \epsilon_j \rangle = \langle x, \phi_j \rangle$ for all j . //

Thus we have shown that conditions (i) and (ii) imply the existence of solutions $\lambda_0, \dots, \lambda_{r-2}$ to the system $\delta^j = \sum_{i=0}^{r-2} \lambda_i T_{i+j}$ and hence have shown the existence of 3-manifold invariants for $A = e^{\pi i/2r}$. It remains to prove conditions (i) and (ii) of 16.5. This requires an excursion into the structure of the Temperley-Lieb algebra. This analysis of the Temperley-Lieb algebra will produce the following result:

Theorem 16.8. Let V_m denote the m -strand Temperley-Lieb algebra with multiplicative generators $1_m, e_1, e_2, \dots, e_{m-1}$ as described in this section. Let δ denote the loop value for this algebra (assumed by convenience to be a complex number). Then there exist elements $f_n \in V_m$ for $0 \leq n \leq m-1$ such that

- (i) $f_0 = 1_m$
- (ii) $f_n = f_{n-1} - \mu_n f_{n-1} e_n f_{n-1}$ for $n \geq 1$
- (iii) $e_i f_n = 0$ for $i \leq n$

(iv) $(e_{n+1}f_n)^2 = \mu_{n+1}^{-1}e_{n+1}f_n$ for $n \leq m - 2$

(v) $\text{tr}(f_n) = \langle f_n e_n \rangle = \delta^{m-n-1} \Delta_{n+1}$.

Here μ_n is defined recursively by the formulas $\mu_1 = \delta^{-1}$, $\mu_n = (\delta - \mu_{n-1})^{-1}$, and $\Delta_n = \prod_{k=1}^n (\delta - 2 \cos(k\pi/(n+1)))$. //

We shall postpone discussion of the details of 16.8 to the end of this section. However, it is worth at least writing down f_1 : According to 16.8(ii) we have $f_1 = 1_m - \delta^{-1}e_1$. Thus, in V_2 we have $f_1 = \text{---} - \delta^{-1}\text{---}$ and

$$\begin{aligned} f_1^2 &= (1 - \delta^{-1}e_1)(1 - \delta^{-1}e_1) \\ &= 1 - 2\delta^{-1}e_1 + \delta^{-2}e_1^2 \\ &= 1 - 2\delta^{-1}e_1 + \delta^{-2}e_1 \\ &= 1 - \delta^{-1}e_1 \\ \therefore f_1^2 &= f_1. \end{aligned}$$

Also, $e_1 f_1 = e_1 - \delta^{-1}e_1^2 = 0$, and

$$\begin{aligned} \text{tr}(f_1) &= \left\langle \text{---} \right\rangle - \delta^{-1} \left\langle \text{---} \right\rangle = \delta^2 - 1 \\ &= \delta^{2-1-1}(\delta^2 - 1). \quad (m = 2, n = 1). \end{aligned}$$

Note that by 16.8, $\Delta_2 = (\delta - 2 \cos(\pi/3))(\delta - 2 \cos(2\pi/3)) = (\delta - 2(1/2))(\delta - 2(-1/2)) = (\delta - 1)(\delta + 1) = \delta^2 - 1$.

Remark on the Trace. In the statement of 16.8 I have used the trace function $\text{tr} : V_m \rightarrow \mathbf{C}$. It is defined by the formula, $\text{tr}(x) = \langle x e_m \rangle$. We take the trace of an element of the Temperley-Lieb algebra by closing the ends of the corresponding tangle via the identity element, and then computing the bracket of the link so obtained.

The next result is a direct corollary of 16.8. Among other things, it provides the desired degeneracy of the form $V_m \times V_m \rightarrow \mathbf{C}$ for $A = e^{i\pi/2r}$, $r \geq 3$ and $m \geq r - 1$. In other words, 16.8 implies condition (ii) of 16.5.

Corollary 16.9. Let $A = e^{\pi i/2r}$ for $r \geq 3$. Then

- (i) If $m \leq r - 2$ then there exists an element $p(m) \in V_m$ such that $\langle -, p(m) \rangle = 1'_m$. [Here $1'_m$ is the element of V_m^* dual to $1_m \in V_m$. That is $1'_m(x)$ equals the coefficient of 1_m in an expansion of x in the usual basis.]
- (ii) If $m \geq r - 1$, then the bilinear form $\langle , \rangle : V_m \times V_m \rightarrow \mathbf{C}$ is degenerate, and there exists an element $q(m) \in V_m$ such that $\langle -, q(m) \rangle = \langle -, 1_m \rangle$ and $q(m)$ is in the subalgebra generated by $\{e_1, e_2, \dots, e_{m-1}\}$.

Proof of 16.9. Recall from 16.8 that $\text{tr}(f_n) = \delta^{m-n-1} \Delta_{n+1}$ for $f_n \in V_m$, and that $\Delta_n = \prod_{k=1}^n (\delta - 2 \cos(k\pi/(n+1)))$. If $A^2 = e^{i\theta}$, then $\delta = -A^2 - A^{-2} = -2 \cos(\theta)$. Thus $\Delta_n = \prod_{k=1}^n -2(\cos(\theta) + \cos(k\pi/(n+1)))$. The hypothesis of 16.9 is $\theta = \pi/r$ for $r \geq 3$. Hence $\Delta_1 \Delta_2 \dots \Delta_{r-2} \neq 0$, but $\Delta_{r-1} = 0$.

For $1 \leq m \leq r-2$ the element $f_{m-1} \in V_m$ is defined since $\Delta_1 \Delta_2 \dots \Delta_{m-1} \neq 0$. By 16.8, $e_i f_{m-1} = 0$ for $i = 1, 2, \dots, m-1$. Since the products of the e_i 's generate a basis (other than 1_m) for V_m as a vector space over \mathbf{C} , we see that we can define $p(m) = \Delta_m^{-1} f_{m-1}$. This ensures that $p(m)$ projects everything but the coefficient of 1_m to zero, and

$$\begin{aligned} \langle 1_m, p(m) \rangle &= \langle 1_m, \Delta_m^{-1} f_{m-1} \rangle \\ &= \Delta_m^{-1} \text{tr}(1_m f_{m-1}) \\ &= \Delta_m^{-1} \text{tr}(f_{m-1}) \\ &= \Delta_m^{-1} \Delta_m \quad (\text{by 16.8}) \\ &= 1. \end{aligned}$$

This completes the proof of 16.9 (i).

Now suppose that $m \geq r - 1$. Then $\Delta_1 \Delta_2 \dots \Delta_{r-2} \neq 0$ so $f_{r-2} \in V_{r-1}$ is well-defined. Now f_{r-2} projects any basis element of V_{r-1} generated by e_1, \dots, e_{r-2} to zero, and $\langle 1_{r-1}, f_{r-2} \rangle = \text{tr}(f_{r-2}) = \Delta_{r-1} = 0$. Hence $\langle -, f_{r-2} \rangle : V_{r-1} \rightarrow \mathbf{C}$ is the zero map. For $m \geq r - 1$ we have the standard inclusion $V_{r-1} \hookrightarrow V_m$ obtained by adding $m - r + 1$ parallel arcs above the given element of V_{r-1} (in the convention

of writing the diagrams of elements of V_m horizontally). It is then easy to see that $\langle -, f_{r-2} \rangle : V_m \rightarrow \mathbf{C}$ is the zero map for $m \geq r - 1$. Hence (since $f_{r-2} \neq 0$) the bilinear form is degenerate. Let $q(m) = 1_m - f_{r-2}$. This completes the proof of 16.9. //

Here is an example to illustrate 16.9 in the case $r = 4$. We have $f_1 = 1 - \delta^{-1}e_1$ and $e_1f_1 = 0$.

$$\begin{aligned} f_2 &= f_1 - \mu_2 f_1 e_2 f_1, \quad \mu_2 = (\delta - \mu_1)^{-1} \\ &= (1 - \delta^{-1}e_1)(1 - \mu_2 e_2(1 - \delta^{-1}e_1)) \\ &= (1 - \delta^{-1}e_1)(1 - \mu_2 e_2 + \mu_2 \delta^{-1}e_2 e_1) \\ &= 1 - \mu_2 e_2 + \mu_2 \delta^{-1}e_2 e_1 \\ &\quad - \delta^{-1}e_1 + \delta^{-1}\mu_2 e_1 e_2 - \mu_2 \delta^{-2}e_1 e_2 e_1 \\ &= 1 - (\delta^{-1} + \mu_2 \delta^{-2})e_1 - \mu_2 e_2 \\ &\quad + \mu_2 \delta^{-1}e_2 e_1 + \delta^{-1}\mu_2 e_1 e_2 \\ \therefore f_2 &= 1 - \delta(\delta^2 - 1)^{-1}e_1 - \delta(\delta^2 - 1)^{-1}e_2 \\ &\quad + (\delta^2 - 1)^{-1}e_1 e_2 + (\delta^2 - 1)^{-1}e_2 e_1. \end{aligned}$$

V_3 has basis $\{1, e_1, e_2, e_1 e_2, e_2 e_1\}$. Since $\delta = -2 \cos(\pi/4) = -\sqrt{2}$, $f_2 \neq 0$. However, $\text{tr}(e_1) = \text{tr}(e_2) = \delta^2$, while $\text{tr}(e_1 e_2) = \text{tr}(e_2 e_1) = \delta$. Thus $\text{tr}(f_2) = (\delta/(\delta^2 - 1))(\delta^4 - 3\delta^2 + 2) = 0$, and $e_1 f_2 = e_2 f_2 = 0$. The element $q(3)$ has the formula

$$\begin{aligned} q(3) &= \left(\frac{1}{\delta^2 - 1} \right) [\delta e_1 + \delta e_2 - e_1 e_2 - e_2 e_1] \\ q(3) &= \sqrt{2} e_1 + \sqrt{2} e_2 - e_1 e_2 - e_2 e_1. \end{aligned}$$

The beauty of 16.8 and 16.9 is that they provide specific constructions for these projections and degeneracies of the forms.

Since 16.9 implies condition (ii) of 16.5 (the degeneracy of the bilinear form) we have only to establish the uniqueness of solutions $\lambda_0, \dots, \lambda_{r-2}$ of the equations

$$\delta^j = \sum_{i=0}^{r-2} \lambda_i T_{i+j}, \quad j = 0, 1, 2, \dots, r-2$$

in order to establish the existence of the three-manifold invariant. (Of course we will need to prove the algebraic Theorem 16.8.) That is, we must prove

Theorem 16.10. When $A = e^{\pi i/2r}$, $r \geq 3$, then the matrix $\{T_{i+j} : 0 \leq i, j \leq r-2\}$ is nonsingular.

Proof. Suppose that $\{T_{i+j}\}$ is singular. This means that there exist $\mu_i \in \mathbb{C}$, not all zero, such that

$$\sum_{i=0}^{r-2} \mu_i T_{i+j} = 0 \text{ for } j = 0, 1, \dots, r-2.$$

Whence

$$\sum_{i=0}^{r-2} \mu_i \langle \text{Diagram } i \rangle = 0.$$

Now consider the functional on tangles given by the formula

$$\sum_{i=0}^{r-2} \mu_i \langle \text{Diagram } i \rangle.$$

Exactly the same “annulus trick” as we used in Lemma 16.6 then shows that this functional is identically zero on tangles. Hence

$$\sum_{i=0}^{r-2} \mu_i \langle \text{Diagram } i \rangle = 0.$$

Thus

$$\sum_{i=0}^{r-2} \mu_i \langle \text{Diagram } i \rangle = 0.$$

This means that the matrix

$$\left\{ \langle \text{Diagram } i \rangle : 0 \leq i, j \leq r-2 \right\}$$

is singular. Hence there exist constants $\nu_0, \nu_1, \dots, \nu_{r-2}$ (not all zero) such that

$$\sum_{i=0}^{r-2} \nu_i \langle \text{Diagram } i \rangle = 0.$$

Now apply the annulus trick once more to conclude that

$$\sum_{i=0}^{r-2} \nu_i \langle \text{Diagram } i \rangle = 0,$$

and hence

$$\sum_{i=0}^{r-2} \nu_i \left\langle \begin{array}{c} \text{square} \\ \text{---} \\ \text{---} \end{array} \bigcirc \overset{i}{\text{---}} \right\rangle = 0$$

for all $j = 0, \dots, r-2$ as a functional on tangles. In particular, we can put into this functional the projector element $p(j)$ of 16.9. Then

$$\left\langle \begin{array}{c} p(j) \\ \text{---} \end{array} \bigcirc \overset{i}{\text{---}} \right\rangle$$

is equal to $1'_j(x_j^i)$, the coefficient of 1_j in the basis expansion for

$$x_j^i = \bigcirc \overset{i}{\text{---}} \in V_j.$$

Note that since $\bigcirc \overset{1}{\text{---}} = \bigcirc \text{---}$ we actually have $x_j^i = (x_j)^i$, the i -th power of the element

$$x_j = \bigcirc \overset{1}{\text{---}} \in V_j.$$

Therefore, to complete this analysis we need to compute $1'_j(x_j)$. (Note that $1'_j(x_j^i) = [1'_j(x_j)]^i$.) It is easy (bracket exercise!) to check by induction that if $A^2 = e^{i\theta}$, then $1'_j(x_j) = -2 \cos(j+1)\theta$ for all j . Hence, in our case $e^{i\theta}$ is a primitive $2r$ -th root of unity, whence $\{\cos((j+1)\theta) \mid j = 0, 1, \dots, r-2\}$ are all distinct. The matrix $\{(-2 \cos((j+1)\theta))^i \mid 0 \leq i, j \leq r-2\}$ is a Vandermonde matrix, hence it has nonzero determinant. This contradicts the assumption that the original matrix was singular. //

The Temperley-Lieb Algebra.

It remains for us to prove Theorem 16.8. This is actually an elementary exercise in induction using the relations in the Temperley-Lieb algebra. Rather than do all the details, let's analyze the induction step to see why the relation $\mu_{n+1} = (\delta - \mu_n)^{-1}$ is needed, and see where the polynomials Δ_n come from. Thus we have $f_0 = 1$, $f_n = f_{n-1} - \mu_n f_{n-1} e_n f_{n-1}$. We assume inductively that $f_n^2 = f_n$ and that $(e_{n+1} f_n)^2 = \mu_{n+1}^{-1} e_{n+1} f_n$ ($n \leq m-2$ for $f_n \in V_m$), and $e_i f_n = 0$ for $i \leq n$. Then

$$\begin{aligned}
 f_{n+1}^2 &= (f_n - \mu_{n+1} f_n e_{n+1} f_n)^2 \\
 &= f_n^2 + \mu_{n+1}^2 f_n e_{n+1} f_n^2 e_{n+1} f_n \\
 &\quad - \mu_{n+1} f_n^2 e_{n+1} f_n - \mu_{n+1} f_n e_{n+1} f_n^2 \\
 &= f_n + \mu_{n+1}^2 f_n (e_{n+1} f_n)^2 - 2\mu_{n+1} f_n e_{n+1} f_n \\
 &= f_n + \mu_{n+1} f_n e_{n+1} f_n - 2\mu_{n+1} f_n e_{n+1} f_n \\
 &= f_{n+1}.
 \end{aligned}$$

Thus the crux of the matter is in the induction for $(e_{n+2} f_{n+1})^2$. We have

$$\begin{aligned}
(e_{n+2}f_{n+1})^2 &= (e_{n+2}(f_n - \mu_{n+1}f_n e_{n+1}f_n))^2 \\
&= (e_{n+2}f_n - \mu_{n+1}e_{n+2}f_n e_{n+1}f_n)^2 \\
&= e_{n+2}f_n e_{n+2}f_n \\
&\quad + \mu_{n+1}^2 e_{n+2}f_n e_{n+1}f_n e_{n+2}f_n e_{n+1}f_n \\
&\quad - \mu_{n+1}e_{n+2}f_n e_{n+2}f_n e_{n+1}f_n \\
&\quad - \mu_{n+1}e_{n+2}f_n e_{n+1}f_n e_{n+2}f_n \\
&= e_{n+2}^2 f_n^2 \\
&\quad + \mu_{n+1}^2 f_n e_{n+2} e_{n+1} e_{n+2} f_n^2 e_{n+1} f_n \\
&\quad - \mu_{n+1} e_{n+2}^2 f_n^2 e_{n+1} f_n \\
&\quad - \mu_{n+1} f_n e_{n+2} e_{n+1} e_{n+2} f_n^2 \\
&\quad (\text{Elements more than 1 unit apart in the } T - L \\
&\quad \text{algebra commute with one another.}) \\
&= \delta e_{n+2} f_n \\
&\quad + \mu_{n+1}^2 f_n e_{n+2} f_n e_{n+1} f_n \\
&\quad - \mu_{n+1} \delta e_{n+2} f_n e_{n+1} f_n \\
&\quad - \mu_{n+1} f_n e_{n+2} f_n \\
&= (\delta - \mu_{n+1}) e_{n+2} f_n + (\mu_{n+1}^2 - \mu_{n+1} \delta) e_{n+2} f_n e_{n+1} f_n \\
&= e_{n+2} (\delta - \mu_{n+1}) (f_n - \mu_{n+1} f_n e_{n+1} f_n) \\
&= (\delta - \mu_{n+1}) e_{n+2} f_{n+1}.
\end{aligned}$$

Thus we require that

$$\mu_{n+2}^{-1} = (\delta - \mu_{n+1}), \text{ or}$$

$$\mu_1 = 1/\delta$$

$$\mu_{n+1} = 1/(\delta - \mu_n).$$

We have

$$\begin{aligned}\mu_1 &= 1/\delta \\ \mu_2 &= \delta/(\delta^2 - 1) \\ \mu_3 &= (\delta^2 - 1)/(\delta^3 - \delta - 1)\end{aligned}$$

and generally,

$$\mu_{n+1} = \Delta_n / \Delta_{n+1}$$

where

$$\begin{aligned}\Delta_0 &= 1 \\ \Delta_1 &= \delta \\ \Delta_{n+2} &= \delta \Delta_{n+1} - \Delta_n.\end{aligned}$$

It is easy to see from this definition that

$$\Delta_n = \prod_{k=1}^n (\delta - 2 \cos(k\pi/(n+1))).$$

(HINT: Let $\delta = x + x^{-1}$, then

$$\begin{aligned}\Delta_n &= (x + x^{-1})\Delta_{n-1} - \Delta_{n-2} \\ &\Rightarrow \Delta_n - x\Delta_{n-1} = x^{-1}(\Delta_{n-1} - x\Delta_{n-2}) \\ &\Rightarrow \Delta_n - x\Delta_{n-1} = x^{-n} \\ &\Delta_n - x^{-1}\Delta_{n-1} = x^n \\ &\Rightarrow (x - x^{-1})\Delta_n = x^{n+1} - x^{-(n+1)}.\end{aligned}$$

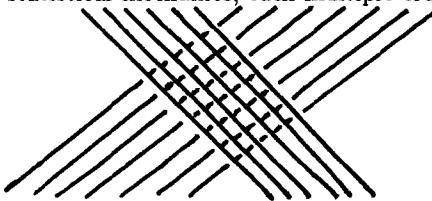
I leave the calculation

$$\text{tr}(f_n) = \delta^{m-n-1} \Delta_{n+1}$$

as an exercise.

This completes the proof of Theorem 16.8, and hence the construction of the three-manifold invariant is now complete. //

Discussion. I recommend that the reader consult the papers [LICK3] of Lickorish, on which this discussion was based. The papers of Reshetikhin and Turaev ([RES], [RT1], [RT2]) are very useful for a more general viewpoint involving quantum groups. See also the paper by Kirby and Melvin [KM] for specific calculations for low values of r and a very good discussion of the quantum algebra. From the viewpoint of statistical mechanics, each multiple crossing



is a (twisted) version of a Potts model (see section 8⁰ of Part II) interconnected with the rest of the knot diagram. This suggests that features of these invariants for $r \rightarrow \infty$ should be related to properties of the continuum limit of the Potts model. In particular the conjectural [AW4] relationships of the Potts model and the Virasoro algebra may be reflected in the behavior of these invariants.

17⁰. Integral Heuristics and Witten's Invariants.

We have, so far, considered various ways to build link invariants as the combinatorial or algebraic analogues of partition functions or vacuum-vacuum expectations. These models tend to assume the form

$$\langle K \rangle = \sum_{\sigma} \langle K | \sigma \rangle \lambda^{\parallel \sigma \parallel}$$

where $\langle K | \sigma \rangle$ is a product of vertex weights, and σ runs over a (large) finite collection of states or configurations of the system associated with the link diagram. In this formulation, the summation $\sum_{\sigma} \lambda^{\parallel \sigma \parallel}$ is analogous to the bare partition function, and $\langle K \rangle$ is the analogue of an expectation value for the link diagram K . The link diagram becomes an observable for a system of states associated with its projection.

It is natural to wonder whether these finite summations can be turned into integrals. We shall see shortly that exactly this does happen in Witten's theory - using the Chern-Simons Lagrangian. The purpose of this section is to give a heuristic introduction to Witten's theory. In order to do this it is helpful to first give the form of Witten's definition and then play with this form in an elementary way. We then add more structure. Thus we shall begin the heuristics in **Level 0**; then we move to **Level 1** and beyond.

The Form of Witten's Definition.

Witten defines the link invariant for a link $K \subset M^3$, M^3 a compact oriented three-manifold, via a functional integral of the form

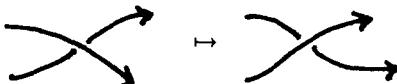
$$Z_K = \int dA e^{\mathcal{L}} \tau_K$$

where A runs over a collection of gauge fields on M^3 . A gauge field (also called a gauge potential or gauge connection) is a Lie-algebra valued field on the three-manifold. \mathcal{L} denotes the integral, over M^3 , of a suitable multiple of the trace of a differential form (the Chern-Simons form). This differential form looks like $A \wedge dA + \frac{2}{3} A \wedge A \wedge A$. Finally, τ_K denotes the product of traces assigned to each component of K . If K has one component, then

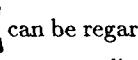
$$\tau_K = \text{Trace} \left[P \exp \left(\oint_K A \right) \right],$$

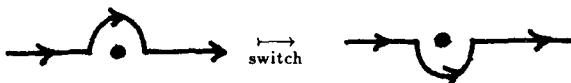
the trace of the path-ordered exponential of the integral of the field A around the closed circuit K . Now before going into the technicalities of these matters, we will look at the surface structure. For small loops K , the function τ_K measures curvature of the gauge field A [exactly how it does this will be the subject of the discussion at Level 1]. The Chern-Simons Lagrangian \mathcal{L} also contains information about curvature. This information is encoded in the way \mathcal{L} behaves when the gauge field is made to vary in the neighborhood of a point $x \in M^3$. [At Level 1, we shall see that the curvature tensor arises as the variation of the Chern-Simons Lagrangian with respect to the gauge field.] These two modes of curvature measurement interact in the integral Z_K to give rise to the link invariant.

We can begin to see this matter intuitively by looking at two issues from the knot theory. One issue is the behavior of Z_K on a “curl”  The other issue is the behavior of Z_K on a crossing switch



Since the curl involves a small loop, it is natural to expect that τ_{curl} will differ from τ_{crossing} by some factor involving curvature. These factors must average out to a constant multiple when the integration is performed.

The difference between  and  can be regarded as a small loop encircling one of the lines. Thus if we write one line perpendicular to the page (as a dot \bullet), then the crossing switch has the appearance



and the difference is a tiny loop around the line normal to the page:



Thus, we expect the curvature to be implicated in both the framing behavior and the change under crossing switch. [This approach to understanding the Witten integral is due to Lee Smolin [LS1], the author [LK29] and P. Cotta Ramusino,

Maurizio Martellini, E. Guadagnini and M. Mintchev [COTT].]

Level 0.

This level explores a formalism of implementing the ideas related to curvature that we have just discussed. Here the idea is to examine the simplest possible formalism that can hold the ideas. As we shall see, this can be done very elegantly. But this is a case of form without content ("beautifully written but content free"). There is an advantage to pure form - it makes demands on the imagination, and maybe there is a - yet unknown - content that fits this form. We shall see.

To work. Let us try to obtain a formal model for the regular isotopy version of the homfly polynomial. That is, let's see what we can do to obtain a functional on knots and links that behaves according to the pattern:

$$\left\{ \begin{array}{l} \nabla \text{X} = z \nabla \text{Y} \\ \nabla \text{Y} = a \nabla \text{Z} \\ \nabla \text{Z} = a^{-1} \nabla \text{Y} \end{array} \right. \quad \left. \right\}$$

To this end we shall write

$$\nabla_K = \int dA e^{\mathcal{L}} \tau_K$$

and make the following assumptions:

$$(1) \frac{\delta \mathcal{L}}{\delta A} = F$$

$$(2) \tau_{\text{X}} - \tau_{\text{Y}} = -z F \tilde{\tau}_{\text{Y}}$$

$$(3) \tau_{\text{Y}} = -a F \tilde{\tau}_{\text{Z}}$$

$$\tau_{\text{Z}} = -a^{-1} F \tilde{\tau}_{\text{Y}}$$

$$(4) \delta \tilde{\tau} / \delta A = \tau$$

Let it be understood that F represents (the idea of) curvature - so that (1), (2) and (3) express the relationship of curvature to the Lagrangian \mathcal{L} , to the switching relation and to the presence of curls. In this Level 0 treatment we use the formalism of freshman calculus, differentiating as though A were a single variable.

Nevertheless, we do take into account the fact that a curve of finite size will not pinpoint the curvature of the field at its center. Thus the local skein relation

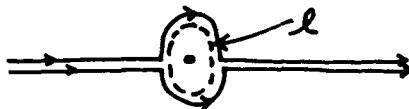
(2) involves a functional $\tilde{\tau}$ so that $\tau_{\text{---}} - \tau_{\text{---}} = -zF\tilde{\tau}_{\text{---}}$. Strictly speaking, one should think of F as measuring curvature at the switch-point p :



Thus

$$\tau_{\text{---}} - \tau_{\text{---}} = -zF(p)\tilde{\tau}_{\text{---}}$$

We have "rolled up" the extra geometry of the small finite loop ℓ



into the function $\tilde{\tau}$ and let the curvature evaluation happen at the center point p . Now in fact we can see approximately what this assumption (4) ($\delta\tilde{\tau}/\delta A = \tau$) means, for

$$\tau_{\text{---}} - \tau_{\text{---}} = -zF(p)d\alpha\tau_{\text{---}}$$

where $d\alpha$ is a function of the area enclosed by the loop ℓ . Thus, to first order of approximation, we are assuming a proportionality

$$\tau\delta A = d\alpha\delta\tau$$

where δA represents the variation of the gauge field in the neighborhood of the line, and $d\alpha$ represents a small area normal to the line. Thus the assumption $\delta\tilde{\tau}/\delta A = \tau$ is an **assumption of a correspondence between field change and geometry near the line**.

With these assumptions in place, and assuming that boundary terms in the integration by parts vanish, we can prove the global skein relation and curl identities:

Proposition 17.1. Under the assumptions (1), (2), (3) and (4) listed above, the formal integral $\nabla_K = \int dA e^{\mathcal{L}} \tau_K$ satisfies the following properties:

(i)

$$\nabla_{\begin{array}{c} \nearrow \\ \searrow \end{array}} - \nabla_{\begin{array}{c} \nearrow \\ \swarrow \end{array}} = z \nabla_{\begin{array}{c} \nearrow \\ \searrow \end{array}}$$

(ii)

$$\nabla_{\begin{array}{c} \nearrow \\ \searrow \end{array}} = a \nabla$$

$$\nabla_{\begin{array}{c} \nearrow \\ \circlearrowleft \end{array}} = a^{-1} \nabla$$

Proof. To show (i), we take the difference and integrate by parts.

$$\begin{aligned} \nabla_{\begin{array}{c} \nearrow \\ \searrow \end{array}} - \nabla_{\begin{array}{c} \nearrow \\ \swarrow \end{array}} &= \int dA e^{\mathcal{L}} \tau_{\begin{array}{c} \nearrow \\ \searrow \end{array}} - \int dA e^{\mathcal{L}} \tau_{\begin{array}{c} \nearrow \\ \swarrow \end{array}} \\ &= \int dA e^{\mathcal{L}} [\tau_{\begin{array}{c} \nearrow \\ \searrow \end{array}} - \tau_{\begin{array}{c} \nearrow \\ \swarrow \end{array}}] \\ &= \int dA e^{\mathcal{L}} (-z F \tilde{\tau}_{\begin{array}{c} \nearrow \\ \searrow \end{array}}) \end{aligned} \tag{2}$$

$$\begin{aligned} &= -z \int dA (e^{\mathcal{L}} F) \tilde{\tau}_{\begin{array}{c} \nearrow \\ \searrow \end{array}} \\ &= -z \int dA \frac{\delta e^{\mathcal{L}}}{\delta A} \tilde{\tau}_{\begin{array}{c} \nearrow \\ \searrow \end{array}} \end{aligned} \tag{1}$$

$$= z \int dA e^{\mathcal{L}} \frac{\delta \tilde{\tau}_{\begin{array}{c} \nearrow \\ \searrow \end{array}}}{\delta A} \quad (\text{integration by parts})$$

$$= z \int dA e^{\mathcal{L}} \tau_{\begin{array}{c} \nearrow \\ \searrow \end{array}} \tag{4}$$

$$= z \nabla_{\begin{array}{c} \nearrow \\ \searrow \end{array}}.$$

Part (ii) proceeds in a similar manner. We shall verify the formula for $\nabla \circlearrowleft$.

$$\begin{aligned}\nabla \circlearrowleft &= \int dA e^{\mathcal{L}} \tau \circlearrowleft \\ &= \int dA e^{\mathcal{L}} (-a F \tilde{\tau} \circlearrowright)\end{aligned}\tag{3}$$

$$\begin{aligned}&= -a \int dA (e^{\mathcal{L}} F) \tilde{\tau} \circlearrowright \\ &= -a \int dA \left(\frac{\delta e^{\mathcal{L}}}{\delta A} \right) \tilde{\tau} \circlearrowright\end{aligned}\tag{1}$$

$$\begin{aligned}&= a \int dA e^{\mathcal{L}} \frac{\delta \tilde{\tau} \circlearrowright}{\delta A} \\ &= a \int dA e^{\mathcal{L}} \tau \circlearrowleft \\ &= a \nabla \circlearrowleft.\end{aligned}\tag{4}$$

This completes the proof. //

Discussion. Notice that in the course of this heuristic we have seen that a plausible condition on the relation of field behavior to geometry can lead the integral over such fields to average out all the local (curvature) behavior and to give global skein and framing identities. It would be interesting to follow up these ideas in more detail, as they suggest a particular physical interpretation of the meaning of the skein and framing identities. In fact, much of this elementary story does go over to the case of the full three-dimensional formalism of gauge fields. This we'll see in Level 1.

Remark. Note that in this picture of the behavior of $\nabla_K = \int dA e^{\mathcal{L}} \tau_K$, we see that ∇_K is not an ambient isotopy invariant. Its values depend crucially on the choice of framing, here indicated by the blackboard framing of link diagrams.

Remark. It appears that the best way to fulfill the promise of this heuristic is to use the full apparatus of gauge theory. Nevertheless, this does not preclude the possibility of some other solution to our conditions (1) \rightarrow (4).

In any case, it is fun to play with the Level 0 heuristic. For example, we can expand

$$\tau_K = e^{\oint_K A} = \sum_{n=0}^{\infty} \left[\iint \cdots \int_{K^n} A^n \right] / n!.$$

Thus

$$\begin{aligned}\nabla_K &= \sum_{n=0}^{\infty} \frac{1}{n!} \int dA e^{\mathcal{L}} \iint \cdots \int_{K^n} A^n \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \iint \cdots \int_{K^n} \left[\int dA e^{\mathcal{L}} A^n \right].\end{aligned}$$

Thus

$$\nabla_K = \sum_{n=0}^{\infty} \frac{1}{n!} \iint \cdots \int_{K^n} \langle A(x_1) \dots A(x_n) \rangle$$

where $\langle A(x_1) \dots A(x_n) \rangle$ denotes the correlation values of the product of these fields for n points on K . Thus, in this expansion the link invariant becomes an infinite sum of Feynman-diagram like evaluations of “self-interactions” of the link K . (See [BN], [GMM].)

Level 1.

Level 0 had the advantage that we could think in freshman calculus level. In order to do Level 1, a leap is required, and we must review some gauge theory. So let us begin by recalling $SU(2)$:

$SU(2)$ is the group of unitary, determinant 1 matrices over the complex numbers. Thus $U \in SU(2)$ has the form $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $ad - bc = 1$ and $\bar{d} = a$, $\bar{c} = -b$ where the bar denotes complex conjugation.

Let U^* denote the conjugate transpose of U , and note that

$$U = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} \text{ with } a\bar{a} + b\bar{b} = 1$$

so that

$$UU^* = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} \begin{pmatrix} \bar{a} & -b \\ \bar{b} & a \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1.$$

If we write $U = e^{i\eta}$ where η is a 2×2 matrix, then the conditions $UU^* = 1$ and $\text{Det}(U) = 1$ become $\eta = \eta^*$ and $\text{tr}(\eta) = 0$ (where tr denotes the usual matrix trace). A matrix of trace zero has the form

$$\eta = \begin{pmatrix} x & y \\ x & -x \end{pmatrix} = x \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + y \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + z \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

and the demand $\eta = \eta^*$ gives rise to the Pauli matrices

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

whose linear combinations give rise to the matrices η .

It is customary to write η in the form $\eta = -\theta(\hat{n} \cdot \tau)/2$ where $\hat{n} = (n_1, n_2, n_3)$ is a real unit vector, and $\hat{n} \cdot \tau = n_1 \tau_1 + n_2 \tau_2 + n_3 \tau_3$. Then

$$e^{i\eta} = e^{-i\theta(\hat{n} \cdot \tau)/2} = \lim_{N \rightarrow \infty} \left(1 - \frac{i}{2} \frac{\theta}{N} \hat{n} \cdot \tau \right)^N$$

and one regards $1 - \frac{i}{2}(d\theta)\hat{n} \cdot \tau$ as an infinitesimal unitary transformation. The Pauli matrices $\tau_a/2$ ($a = 1, 2, 3$) are called the “infinitesimal generators” of $SU(2)$. Note that they obey the commutation relations $[\frac{1}{2}\tau_a, \frac{1}{2}\tau_b] = i\epsilon_{abc}\frac{1}{2}\tau_c$ where ϵ_{abc} denotes the alternating symbol on three indices.

In general with a Lie group of dimension r one has an associated Lie algebra (the infinitesimal generators) with basis elements t_1, t_2, \dots, t_r , and commutation relations

$$[t_a, t_b] = if_{abc}t_c$$

where one sums over the repeated index c . The structure constants f_{abc} themselves assemble into a matrix representation of the Lie algebra called the **adjoint representation**. This is defined via $ad(A)(B) = [A, B]$ for A and B in the Lie algebra of G . Thus, let $T_a = ad(t_a)$, and note that

$$T_a(t_b) = [t_a, t_b] = if_{abc}t_c.$$

Hence $(T_a)_{bc} = if_{abc}$ with respect to the basis $\{t_1, t_2, \dots, t_r\}$.

In the case of $SU(2)$, the adjoint representation gives the matrices

$$T_1 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, T_2 = \frac{1}{2} \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, T_3 = \frac{1}{2} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

These represent infinitesimal rotations about the three spatial axes in \mathbf{R}^3 . These generators are normalized so that $\text{Trace}(T_a T_b) = \frac{1}{2}\delta_{ab}$ where δ_{ab} denotes the Kronecker delta.

With these remarks about $SU(2)$ in hand, let us turn to the beginnings of gauge theory: First, recall Maxwell's equation for the magnetic field $\vec{B} : \nabla \cdot \vec{B} = 0$. That the magnetic field is divergence-free leads us to write it as the curl of another field \vec{A} . Thus $\vec{B} = \nabla \times \vec{A}$ where \vec{A} is called the vector potential. Since $\nabla \cdot (\nabla \times \vec{A}) = 0$ for any \vec{A} , this ensures that \vec{B} is divergence-free. Since the curl of a gradient is zero, we can change \vec{A} by an arbitrary gradient of a scalar field without affecting the value of the \vec{B} -field. Thus, if $\vec{A}' = \vec{A} + \nabla \Lambda$, then $\nabla \times (\vec{A}') = \nabla \times \vec{A} = \vec{B}$. In the same way, the curl equation for the electric field is $\nabla \times \vec{E} = -\partial \vec{B} / \partial t$, whence $\nabla \times (\vec{E} + \partial \vec{A} / \partial t) = 0$. This suggests that the combination of the electric field and the time-derivative of the vector potential should be written as a gradient as in $\vec{E} + \partial \vec{A} / \partial t = \nabla V$. Here V is called the scalar potential. If $\vec{A}' = \vec{A} + \nabla \Lambda$, then we must set $V' = V - \partial \Lambda / \partial t$ in order to preserve the electric field.

In tensor form, one has the electromagnetic tensor

$$F^{\mu\nu} = \partial^\nu A^\mu - \partial^\mu A^\nu$$

constructed from the four-vector potential $A^\mu = (V; \vec{A})$. This tensor is unchanged by the **gauge transformation** $A^\mu \mapsto A^\mu - \partial^\mu \Lambda$ where Λ is any differentiable function of these coordinates.

This principle of gauge invariance provides a close tie between the formalisms of electromagnetic and quantum theory. Recall that a state in quantum mechanics is described by a complex-valued wave function $\psi(x)$ and that quantum mechanical observables involve integrations of the form $\langle E \rangle = \int \psi^* E \psi$. Such an average is invariant under a global phase change $\psi(x) \mapsto e^{i\theta} \psi(x)$. Thus it is natural to ask what happens to the quantum mechanics if we set $\psi'(x) = e^{i\theta(x)} \psi(x)$. We see at once that the derivatives of $\psi'(x)$ involve more than a phase change:

$$\partial_\mu \psi'(x) = e^{i\theta(x)} [\partial_\mu \psi(x) + i(\partial_\mu \theta(x)) \psi(x)].$$

If we replace the derivative ∂_μ by the gauge-covariant derivative

$$\mathcal{D}_\mu = \partial_\mu + ieA_\mu$$

(Think of e as charge and A_μ as a four-vector potential for an electromagnetic field associated with $\psi(x)$.), and if we assume that A_μ transforms via $A'_\mu \mapsto A'_\mu =$

$A_\mu - (1/e)\partial_\mu\theta(x)$ when $\psi \mapsto \psi' = e^{i\theta(x)}\psi$, then we have $\mathcal{D}_\mu\psi'(x) = e^{i\theta(x)}\mathcal{D}_\mu\psi(x)$. In this way, the principle of gauge invariance leads naturally to an interrelation of quantum mechanics and electromagnetism. This is the present form of a unification idea that originated with Herman Weyl [WEYL].

The Bohm-Aharanov Effect.

If the wave function $\psi^0(x, t)$ is a solution to the Schrödinger equation in the absence of a vector potential, then the solution in the presence of a vector potential will be $\psi(x, t) = \psi^0(x, t)e^{iS/\hbar}$ where $S = e \oint dx \cdot \vec{A}$ (See [Q], p. 43.). At first it appears that since the new solution differs from the old by only a phase factor, the potential has no observable influence.

Imagine that a coherent beam of charged particles is split by an obstacle and forced to travel in two paths on opposite sides of a solenoid. After passing the solenoid the beams are recombined and the resulting interference pattern is the sum of the wave functions for each path. If current flows in the solenoid, a magnetic field \vec{B} is created that is essentially confined to the solenoid. Since $\nabla \times \vec{A} = \vec{B}$ we see by Stoke's Theorem that the line integral of \vec{A} around a closed loop containing the solenoid will be nonzero (and essentially constant - equal to the flux of \vec{B}). The interference in the two components of the wave function is then determined by the phase difference $(e/\hbar) \oint dx \cdot \vec{A}(x)$ consisting in this integral around a closed loop. The upshot is that the vector potential **does** have observable significance in the quantum mechanical context! Furthermore, the significant phase factor is the Wilson loop

$$\exp\left[(-ie/\hbar) \oint dx^\mu A_\mu\right].$$

This sets the stage for gauge theory and topology.

Parallel Transport and Wilson Loops.

The previous discussion has underlined the physical significance of the integral $R(C; A) = e^{i \int_{P_1}^{P_2} A \cdot dx}$ where A is a vector field. We find that when A undergoes a gauge transformation

$$A \mapsto A' = A - \nabla\Lambda$$

then

$$R(C; A') = e^{-i\Lambda(P_2)} R(C; A) e^{i\Lambda(P_1)}.$$

We can regard this integration as providing a parallel transport for the wave function ψ between the points p_1 and p_2 . Specifically, suppose that we consider the displacement

$$p_1 = x, \quad p_2 = x + dx.$$

Then

$$R(C; A) = 1 + iA \cdot dx.$$

The transport for the wave function $\psi(x)$ to the point $x + dx$ is given by $\psi^t(x) = (1 + iA \cdot dx)\psi(x)$. Thus, we define the covariant derivative $D\psi$ via

$$D\psi = \frac{\psi(x + dx) - \psi^t(x)}{dx}$$

Thus

$$D\psi = \frac{\psi(x + dx) - \psi(x) + \psi(x) - \psi^t(x)}{dx}.$$

Hence

$$D\psi = (\nabla - iA)\psi.$$

(This agrees (conceptually) with our calculation prior to the discussion of the Aharonov-Bohm effect.)

We see that if $\psi(x)$ transforms under gauge transformation as

$$\psi(x) \mapsto \psi'(x) = e^{-i\Lambda(x)}\psi(x),$$

then

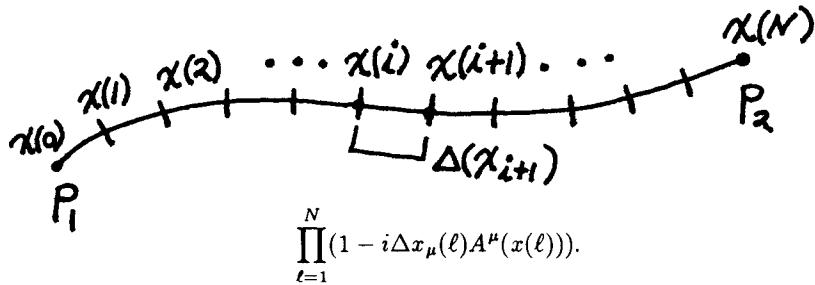
$$\begin{aligned} D'\psi'(x) &= (\nabla - iA')e^{-i\Lambda(x)}\psi(x) \\ &= (\nabla - iA + i\nabla\Lambda)e^{-i\Lambda(x)}\psi(x) \\ &= e^{-i\Lambda(x)}(-i\nabla\Lambda + \nabla - iA + i\nabla\Lambda)\psi(x) \\ &= e^{i\Lambda}(\nabla - iA)\psi \\ D'\psi' &= e^{-i\Lambda}D\psi. \end{aligned}$$

With these remarks, we are ready to generalize the entire situation to an arbitrary gauge group. (The electromagnetic case is that of the gauge group $U(1)$ corresponding to the phase $e^{i\theta}$.)

Thus we return now to a Lie algebra represented by Hermitian matrices $[T_a, T_b] = i\epsilon_{abc}T_c$. Group elements take the form $U = e^{-i\Lambda^a T_a}$ (sum on a) with Λ^a real numbers.

In this circumstance, the gauge field is a generalization of the vector potential. This gauge field depends on an **internal index a** (corresponding to a decomposition of the wave function into an ensemble of wave functions $\psi_a(x)$) and has values in the Lie algebra: $A_\mu(x) = T_a A_\mu^a(x)$. Thus $A(x) = (A_1(x), A_2(x), A_3(x))$ giving a Lie algebra valued field on three-dimensional space. The infinitesimal transport is given by $R(x + dx, x; A) = 1 - idx^\mu A_\mu(x)$.

In order to develop this transport for a finite trajectory it is necessary to take into account the non-commutativity of the Lie algebra elements. Thus we must choose a partition of the path C from p_1 to p_2 and take an ordered product:



The limit (assuming it exists) of this ordered product, as $\Delta x(\ell) \rightarrow 0$ is by definition the **path-ordered exponential**

$$R(C; A) = \mathbf{P} \exp \left(-i \int_\alpha^\beta A(x) \cdot dx \right).$$

We want a notion of gauge transformation $A \mapsto A'$ so that

$$R(C; A') = U(\beta) R(C; A) U^{-1}(\alpha).$$

(This generalizes the abelian case.)

To determine the transformation law for A , consider the infinitesimal case.

We have

$$\begin{aligned}
 1 - i dx^\mu A'_\mu &= U(x + dx)(1 - i dx^\mu A_\mu)U^{-1}(x) \\
 &= U(x + dx)U^{-1}(x) - i dx^\mu [U(x + dx)A_\mu U^{-1}(x)] \\
 &= (U(x) + U'(x)dx)U^{-1}(x) - i dx^\mu [(U(x) + U'(x)dx)A_\mu U^{-1}(x)] \\
 &= 1 + [(\partial_\mu U)dx^\mu]U^{-1} - i dx^\mu [(U + \partial_\mu U dx^\mu)A_\mu U^{-1}] \\
 &\approx 1 + ((\partial_\mu U)U^{-1})dx^\mu - i dx^\mu [UA_\mu U^{-1}] \\
 &= 1 + ((\partial_\mu U)U^{-1} - i UA_\mu U^{-1})dx^\mu \\
 &= 1 - i dx^\mu [i(\partial_\mu U)U^{-1} + UA_\mu U^{-1}].
 \end{aligned}$$

Thus

$$A'_\mu = UA_\mu U^{-1} + i(\partial_\mu U)U^{-1}.$$

This equation defines the appropriate generalization of gauge transformation in the non-abelian context.

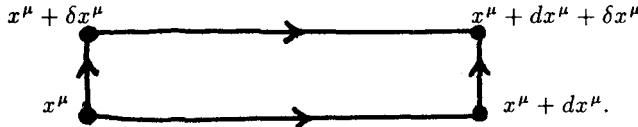
With this interpretation of covariant derivative and parallel transport via Wilson loops, we can recover the gauge theoretic analog of curvature via parallel transport around a small loop. That is, given the gauge potential

$$A_\mu = A_\mu(x) = A_\mu^a(x)T_a$$

let the **curvature tensor** $F_{\mu\nu}$ be defined by the formula

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu]$$

where $[A_\mu, A_\nu] = A_\mu A_\nu - A_\nu A_\mu$. Consider an infinitesimal circuit \square :



Let $R_\square = R(\square; A) = \mathbf{P} \exp(i \oint Adx)$ be the Wilson loop for this circuit. Then

Lemma 17.2. With \square as above,

$$R(\square; A) = \exp\{i F_{\mu\nu} dx^\mu \delta x^\nu\}.$$

Proof.

$$\begin{aligned} R_{\square} &= R(x, x + dx) \cdot R(x + dx, x + dx + \delta x) \\ &= R(x + dx + \delta x, x + \delta x) R(x + \delta x, x). \end{aligned}$$

$$\begin{aligned} \text{Thus } R(x, x + dx)R(x + dx, x + dx + \delta x) \\ &= \exp(iA_\mu(x)dx^\mu) \exp(iA_\nu(x + dx)\delta x^\nu) \\ &= \exp(i[A_\mu dx^\mu + A_\nu \delta x^\nu + \partial_\mu A_\nu dx^\mu \delta x^\nu] \\ &\quad - \frac{1}{2}[A_\mu, A_\nu]dx^\mu \delta x^\nu). \end{aligned}$$

Here we are using the (easily checked) identity

$$e^{\lambda A} e^{\lambda B} = e^{\lambda(A+B)+(\lambda^2/2)[A,B]} + \mathcal{O}(\lambda^3) \text{ for matrices } A \text{ and } B \text{ and scalar } \lambda.$$

Similarly, the return part of the loop yields

$$\begin{aligned} R(x + dx + \delta x, x + \delta x)R(x + \delta x, x) \\ &= \exp(-iA_\mu(x + \delta x)dx^\mu) \exp(-iA_\nu(x)\delta x^\nu) \\ &= \exp(-i(A_\mu dx^\mu + \partial_\nu A_\mu dx^\mu \delta x^\nu + A_\nu \delta x^\nu) \\ &\quad - \frac{1}{2}[A_\mu, A_\nu]dx^\mu \delta x^\nu). \end{aligned}$$

Multiplying these two pieces, we obtain

$$\begin{aligned} R_{\square} &= \exp(i[\partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu]]dx^\mu \delta x^\nu) \\ &= \exp(iF_{\mu\nu}dx^\mu \delta x^\nu). \end{aligned}$$

This completes the proof. //

Remark. Two directions of expression are useful for this curvature. If we express the gauge potential as the 1-form $A = A_\mu^a T_a dx^\mu$ (sum on a and μ) then $A = A_\mu dx^\mu$ and

$$dA - iA \wedge A = \frac{1}{2}(\partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu])dx^\mu \wedge dx^\nu = \frac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu.$$

Thus the curvature can be expressed neatly in terms of differential forms. In this language it is customary to rewrite so that the i is not present, but we shall not

do this since it is most natural to have the i as an expression of the transport $\psi(x + dx) = (1 + iA_\mu dx^\mu)\psi(x)$.

The other direction is to explicitly compute the terms of the curvature - using the Lie algebra: $[T_a, T_b] = if_{abc}T_c$ (sum on c). For this formalism, assume the f_{abc} is anti-symmetric in a, b, c . Then

$$\begin{aligned} F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu] \\ &= \partial_\mu A_\nu^a T_a - \partial_\nu A_\mu^b T_b - i[A_\mu^a T_a, A_\nu^b T_b] \\ &= (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a)T_a - iA_\mu^a A_\nu^b [T_a, T_b] \\ &= (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a)T_a + A_\mu^a A_\nu^b f_{abc} T_c \\ &= (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a)T_a + A_\mu^a A_\nu^b f_{cab} T_c \\ &= (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a)T_a + A_\mu^b A_\nu^c f_{abc} T_a \\ F_{\mu\nu} &= (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a + A_\mu^b A_\nu^c f_{abc})T_a. \end{aligned}$$

Therefore, we define $F_{\mu\nu}^a$ as the coefficient of T_a in the summation above:

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + A_\mu^b A_\nu^c f_{abc}.$$

The Chern-Simons Form.

We now turn to the Chern-Simons form, and its relationship with curvature. In the language of differential forms the Chern-Simons form is given by the formula

$$CS = A \wedge dA - \frac{2i}{3} A \wedge A \wedge A$$

where $A = A_\mu^a(x)T_a dx^\mu$ is a gauge potential on three-dimensional space. Here we assume that the T_a are Lie algebra generators, and that

- (1) $\text{tr}(T_a T_b) = (\frac{1}{2})\delta_{ab}$ where δ_{ab} denotes the Kronecker delta, and tr is the usual matrix trace.
- (2) $[T_a, T_b] = \sum_c i f_{abc} T_c$ with f_{abc} anti-symmetric in a, b, c .
 $([T_a, T_b] = T_a T_b - T_b T_a).$

(Later in the chapter, we restrict to the special case of the Lie algebra of $SU(N)$ in the fundamental representation.)

The Chern-Simons form occurs in a number of contexts. An important fact, for our purposes, that I shall not verify is that, given a three-dimensional manifold

M^3 without boundary, and a gauge transformation $g : M^3 \rightarrow G$ (G denotes the Lie group in this context.) then the integral changes by a multiple of an integer. More precisely, if \mathcal{CS}^g denotes the result of applying the gauge transformation to \mathcal{CS} , then

$$\int_{M^3} \text{tr}(\mathcal{CS}^g) = \int_{M^3} \text{tr}(\mathcal{CS}) + 8\pi^2 n(g)$$

where $n(g)$ is the degree of the mapping $g : M^3 \rightarrow G$. See [JACI] for the proof of this fact.

Thus, it is sufficient to exponentiate the integral to obtain a gauge invariant quantity:

$$\exp\left(\frac{ik}{4\pi} \int_{M^3} \text{tr}(\mathcal{CS})\right).$$

We let $\mathcal{L}_M = \mathcal{L}_M(A) = \int_{M^3} \text{tr}(\mathcal{CS})$. Then $\exp((ik/4\pi)\mathcal{L}_M)$ is gauge invariant for all integers k .

Witten's definition for invariants of links and 3-manifolds is then given by the formula

$$W_K = \int dA e^{(ik/4\pi)\mathcal{L}_M} \prod_{\lambda} \text{tr}\left(P \exp\left(i \oint_{K_\lambda} A\right)\right)$$

(The product is taken over link components K_λ .) where the integration is taken over all gauge fields on the three-manifold M^3 , modulo gauge equivalence. The existence of an appropriate measure on this moduli space is still an open question (See [A2].). Nevertheless, it is possible to proceed under the assumption that such a measure does exist, and to investigate the formal properties of the integral.

One of the most remarkable formal properties of Witten's integral formula is the interplay between curvature as detected by the Wilson lines and curvature as seen in the variation of the Chern-Simons Lagrangian $\mathcal{L}_M(A)$. Specifically, we have the

Proposition 17.3. Let $\mathcal{CS} = A \wedge dA - i\frac{2}{3}A \wedge A \wedge A$ and

$$\mathcal{L}_{M^3}(A) = \int_{M^3} \text{tr}(\mathcal{CS})$$

where M^3 is a compact three-manifold and A is a gauge field as described above. Then the curvature of the gauge field at a point $X \in M^3$ is determined by the

variation of \mathcal{L} via the formula

$$F_{\mu\nu}^a = \epsilon_{\mu\nu\lambda} \delta\mathcal{L}/\delta A_\lambda^a.$$

The $\epsilon_{\mu\nu\lambda}$ is the “epsilon” on three indices with value $\epsilon_{123} = +1$, value ± 1 for $\mu\nu\lambda$ distinct according to the sign of the permutation, and value 0 otherwise. The derivative indicated in this formula is a functional derivative taken with respect to the gauge field varying in a neighborhood of the point X . The curvature tensor is given explicitly as

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + A_\mu^b A_\nu^c f_{abc}.$$

Proof. In order to prove this proposition, we first obtain a local coordinate expression for the Chern-Simons form.

$$\begin{aligned} \mathcal{CS} &= A \wedge dA - (2i/3)A \wedge A \wedge A \\ A &= A_k^a T_a dx^k \\ \Rightarrow dA &= \partial_j A_k^a T_a dx^j \wedge dx^k \\ A \wedge dA &= A_i^a T_a \partial_j A_k^b T_b dx^i \wedge dx^j \wedge dx^k \\ &= A_i^a \partial_j A_k^b T_a T_b dx^i \wedge dx^j \wedge dx^k \\ A \wedge dA &= \epsilon^{ijk} A_i^a A_j^b A_k^c T_a T_b T_c dx^1 \wedge dx^2 \wedge dx^3. \end{aligned}$$

Similarly $A \wedge A \wedge A = \epsilon^{ijk} A_i^a A_j^b A_k^c T_a T_b T_c dx^1 \wedge dx^2 \wedge dx^3$. Thus

$$\mathcal{CS} = \epsilon^{jkt} \left[A_j^a \partial_k A_t^b T_a T_b - \frac{2i}{3} A_j^a A_k^b A_t^c T_a T_b T_c \right] dv$$

where $dv = dx^1 \wedge dx^2 \wedge dx^3$.

Now, we re-express $A \wedge A \wedge A$ using Lie algebra commutators:

$$\begin{aligned} A \wedge A \wedge A &= \epsilon^{jkl} A_j^a A_k^b A_l^c T_a T_b T_c \\ &= \sum_{\pi \in S_3} \text{sgn}(\pi) A_{\pi_1}^a A_{\pi_2}^b A_{\pi_3}^c T_a T_b T_c \end{aligned}$$

where S_3 denotes the set of permutations of $\{1, 2, 3\}$ and $\text{sgn}(\pi)$ is the sign of the permutation π . Thus

$$A \wedge A \wedge A = \sum_{\pi \in \text{Perm}\{\alpha, \beta, \gamma\}} \text{sgn}(\pi) A_1^\alpha A_2^\beta A_3^\gamma T_{\pi(\alpha)} T_{\pi(\beta)} T_{\pi(\gamma)}.$$

Here we still sum over a, b, c and use the fact that A_μ^a is a commuting scalar. Thus

$$\begin{aligned} A \wedge A \wedge A &= A_1^a A_2^b A_3^c \sum_{\pi \in \text{Perm}\{a,b,c\}} \text{sgn}(\pi) T_{\pi_a} T_{\pi_b} T_{\pi_c} \\ &= A_1^a A_2^b A_3^c \{[T_a, T_b] T_c - [T_a T_c] T_b + [T_b T_c] T_a\}. \end{aligned}$$

Now use the fact that $\text{tr}(T_a T_b) = (\delta_{ab})/2$ to find that

$$\begin{aligned} \text{tr}([T_a, T_b] T_c) &= \sum_k \text{tr}(i f_{abk} T_k T_c) \\ &= \sum_k (i/2) f_{abk} \delta_{kc} \\ &= \left(\frac{i}{2}\right) f_{abc}. \end{aligned}$$

Thus

$$\begin{aligned} \text{tr}(A \wedge A \wedge A) &= \left(\frac{i}{2}\right) A_1^a A_2^b A_3^c [f_{abc} - f_{acb} + f_{bca}] \\ &= \left(\frac{i}{4}\right) A_1^a A_2^b A_3^c \left[\begin{array}{ccc} f_{abc} & -f_{acb} & f_{bca} \\ -f_{bac} & f_{cab} & -f_{cba} \end{array} \right] \end{aligned}$$

since f_{abc} is antisymmetric in a, b, c . Therefore, $\text{tr}(A \wedge A \wedge A) = \frac{i}{4} \epsilon^{ijk} A_j^a A_k^b A_\ell^c f_{abc}$.

Hence

$$\text{tr}(\mathcal{CS}) = \frac{\epsilon^{jkl}}{2} \left[\sum_a A_j^a \partial_k A_\ell^a + \frac{1}{3} A_j^a A_k^b A_\ell^c f_{abc} \right].$$

We are now ready to compute $\delta \mathcal{L}/\delta A_j^a$ where $\mathcal{L} = \int_M \text{tr}(\mathcal{CS})$. Note that

$$\begin{aligned} \frac{\delta}{\delta A_\ell^r} [\epsilon^{ijk} A_i^a A_j^b A_k^c f_{abc}] \\ &= \epsilon^{\ell jk} A_j^b A_k^c f_{rbc} + \epsilon^{i\ell k} A_i^a A_k^c f_{arc} + \epsilon^{ij\ell} A_i^a A_j^b f_{abr} \\ &= 3\epsilon^{\ell jk} A_j^b A_k^c f_{rbc} \end{aligned}$$

(since both ϵ and f are antisymmetric), and

$$\begin{aligned} \frac{\delta}{\delta A_\ell^r} \left(\epsilon^{ijk} \sum_a A_i^a \frac{\partial A_k^a}{\partial x^j} \right) \\ &= \epsilon^{ijk} \sum_a \left[\frac{\delta A_i^a}{\delta A_\ell^r} \partial_j A_k^a + A_i^a \frac{\partial_j [\delta A_k^a]}{\delta A_\ell^r} \right] \\ &= \epsilon^{\ell jk} \partial_j A_k^r + \epsilon^{ij\ell} A_i^r \frac{\partial_j [\delta A_k^r]}{\delta A_\ell^r}. \end{aligned}$$

In integrating over M^3 , we assume that it is valid to integrate by parts and that boundary terms in that integration vanish. Thus

$$\begin{aligned} & \int_{M^3} \frac{\delta}{\delta A_i^r} \left(\epsilon^{ijk} \sum_a A_i^a \partial_j A_k^a \right) \\ &= \int_{M^3} \epsilon^{ijk} \partial_j A_k^r - \int_{M^3} \epsilon^{ijr} \partial_j A_i^r \\ &= 2 \int_{M^3} \epsilon^{ijk} \partial_j A_k^r. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\delta \mathcal{L}}{\delta A_i^a} &= \int_{M^3} \left[\epsilon^{ijk} \partial_j A_k^a + \frac{1}{2} \epsilon^{ijk} A_j^b A_k^c f_{abc} \right] \\ \epsilon_{rsi} \frac{\delta \mathcal{L}}{\delta A_i^a} &= \int_{M^3} (\partial_r A_s^a - \partial_s A_r^a) + \frac{1}{2} [A_r^b, A_s^c] f_{abc} \\ &= \int_{M^3} F_{rs}^a. \end{aligned}$$

Note, however that it is understood in taking the functional derivative that we multiply by a delta function centered at the point $x \in M^3$. Thus

$$\epsilon_{\mu\nu\lambda} \frac{\delta \mathcal{L}}{\delta A_\lambda^a(x)} = F_{\mu\nu}^a(x). \quad //$$

We are now ready to engage Level 1 of the integral heuristics. Let's summarize facts and notation. Let K be a knot, and let $\langle K | A \rangle$ denote the value of the Wilson loop for the gauge potential A taken around K . Thus

$$\langle K | A \rangle = \text{tr} \left(\mathbf{P} \exp \left(i \int_K A_\nu^a(x) T_a dx^\nu \right) \right).$$

We can write the Wilson loop symbolically as

$$\langle K | A \rangle = \prod_x (1 + i A_\nu^a(x) T_a dx^\nu)$$

where it is understood that this product stands for the limit

$$\lim_{n \rightarrow \infty} \prod_{k=1}^n (1 + i A_\nu^a(x_k) T_a dx_k^\nu)$$

where $\{x_1, x_2, \dots, x_k\}$ is a partition of the curve K . With

$$\mathcal{L}(A) = \mathcal{L}_{M^3} = \int_{M^3} \mathcal{CS}(A)$$

(\mathcal{CS} denotes the Chern-Simon form.) we have the functional integral

$$W_K = \int dA e^{(ik/4\pi)\mathcal{L}}(K|A)$$

and, if K has components K_1, \dots, K_n then

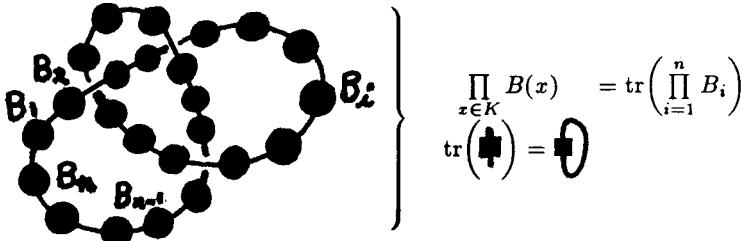
$$\langle K|A \rangle = \prod_{i=1}^n \langle K_i|A \rangle.$$

We shall determine a difference formula for $W_K - W_{K'}$ that is valid in the limit of large k and infinitesimally close strings in the crossing exchange. Upon applying this difference formula to the gauge group $SU(N)$ in the fundamental representation, the familiar Homfly skein identity will emerge.

Before calculating, it will help to think carefully about the schematic picture of this model. First of all, the Wilson loop, written in the form

$$\langle K|A \rangle = \prod_{x \in K} (1 + iA_\nu(x)dx^\nu) = \prod_{x \in K} B(x)$$

is already a trace, since it is a circular product of matrices:



In this sense, the functional integral is similar to our previous state models in that its evaluation involves traced matrix products along the lines of the knot or link. But since the link is in three-space or in a three-manifold M^3 , there is no preference for crossings. Instead, the matrices $B(x) = (1 + iA_\nu(x)dx^\nu)$ are arrayed all along the embedded curves of K . Each matrix $B(x)$ detects the local behavior

of the gauge potential, and their traced product is the action of the link as an observable for this field.

It is definitely useful to think of these matrices as arrayed along the line and multiplied via our usual diagrammatic tensor conventions. In particular, if we differentiate the Wilson line with respect to the gauge, then the result is a **matrix insertion in the line**:

$$\begin{aligned} \frac{\delta\langle K|A\rangle}{\delta A_\nu^a(x)} &= \frac{\delta}{\delta A_\nu^a(x)} \prod_{y \in K} (1 + iA_\nu^a(y)T_a dy^\nu) \\ &= iT_a dx^\nu \langle K|A\rangle \\ \frac{\delta\langle K|A\rangle}{\delta A_\nu^a(x)} &= i dx^\nu T_a \langle K|A\rangle \\ \frac{\delta}{\delta A_\nu^a(x)} \left[\begin{array}{c} \text{Diagram of a loop with points } y \\ \text{and a point } x \text{ marked with an arrow} \end{array} \right] &= (idx^\nu) \left[\begin{array}{c} \text{Diagram of a loop with points } y \\ \text{and a point } x \text{ marked with an arrow} \\ \text{with a circle around the point } x \text{ labeled } T_a \end{array} \right]. \end{aligned}$$

In this functional differentiation, the Lie algebra matrix T_a is inserted into the line at the point x . In writing $idx^\nu T_a \langle K|A\rangle$ it shall be understood that T_a is to be so inserted. Thus $dx^\nu T_a \mathcal{O}$ means that T_a is inserted into the product \mathcal{O} at the point x . Such insertions can be indicated more explicitly, but at the cost of burdening the notation.

Curvature Insertion.

By Proposition 17.3 we know that $F_{\mu\nu}^a = \epsilon_{\mu\nu\lambda} \delta\mathcal{L}/\delta A_\lambda^a$, and by Lemma 17.2 we know that the curvature $F_{\mu\nu}^a$ can also be interpreted as the valuation of a small Wilson loop. More precisely, 17.2 tells us that

$$\begin{aligned} \langle \square|A \rangle &= \text{tr} R(\square; A) = \text{tr} \exp(iF_{\mu\nu} dx^\mu dx^\nu) \\ &= \text{tr}(1 + iF_{\mu\nu} dx^\mu dx^\nu) = \text{tr}(1 + iF_{\mu\nu}^a T_a dx^\mu dx^\nu). \end{aligned}$$

It follows, that if we change the Wilson line by a small amount in the neighborhood of a point x , then the new line will differ from the old line by an insertion of $1 + iF_{\mu\nu} dx^\mu dx^\nu$ at x . Consequently, the difference between old and new lines,

denoted by $\delta\langle K|A \rangle$, is obtained by the curvature insertion

$$\delta\langle K|A \rangle = i dx^\mu dx^\nu F_{\mu\nu}^a T_a \langle K|A \rangle.$$

Once again, this is an insertion of the Lie algebra element T_a at x , the point of variation, multiplied by the curvature and by the infinitesimal area traced out by the moving curve.

We now work out the effect of this variation of the Wilson line on W_K .

Notation. Let $\langle K \rangle$ denote W_K . This follows our state-model conventions, but is at variance with statistical mechanics, where $\langle K \rangle$ is the quotient of W_K and $W_\phi = \int_M dA e^{(i/4\pi)\mathcal{L}_M}$. In this notation we have

$$\begin{aligned} \delta\langle K \rangle &= \int dA e^{(ik/4\pi)\mathcal{L}} \delta\langle K|A \rangle \\ &= \int dA e^{(ik/4\pi)\mathcal{L}} i dx^\mu dx^\nu F_{\mu\nu}^a T_a \langle K|A \rangle \\ &= \int dA e^{(ik/4\pi)\mathcal{L}} \left(\epsilon_{\mu\nu\lambda} \frac{\delta \mathcal{L}}{\delta A_\lambda^a(x)} \right) i dx^\mu dx^\nu T_a \langle K|A \rangle \\ &= \frac{4\pi}{k} \int dA \frac{\delta(e^{(ik/4\pi)\mathcal{L}})}{\delta A_\lambda^a(x)} \epsilon_{\mu\nu\lambda} dx^\mu dx^\nu T_a \langle K|A \rangle \\ &= -\frac{4\pi}{k} \int dA e^{(ik/4\pi)\mathcal{L}} \sum_a \epsilon_{\mu\nu\lambda} dx^\mu dx^\nu T_a \frac{\delta}{\delta A_\lambda^a(x)} \langle K|A \rangle \\ &= -\frac{4\pi i}{k} \int dA e^{(ik/4\pi)\mathcal{L}} [\epsilon_{\mu\nu\lambda} dx^\mu dx^\nu dx^\lambda] \left[\sum_a T_a T_a \right] \langle K|A \rangle. \end{aligned}$$

Thus we have shown

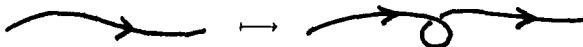
Proposition 17.4. With $\langle K \rangle$ denoting the functional integral

$$\langle K \rangle = \int dA e^{(ik/4\pi)\mathcal{L}_M(A)} \langle K|A \rangle$$

where $\langle K|A \rangle$ denotes the product of Wilson loops for the link K , the variation in $\langle K \rangle$ corresponding to an infinitesimal deformation of the Wilson line K in the neighborhood of a point x is given by the formula,

$$\delta\langle K \rangle = \frac{-4\pi i}{k} \int dA e^{(ik/4\pi)\mathcal{L}} [\epsilon_{\mu\nu\lambda} dx^\mu dx^\nu dx^\lambda] \left[\sum_a T_a T_a \right] \langle K|A \rangle.$$

Discussion. In order to apply this variational formula, it is necessary to interpret it carefully. First of all, note that $\sum(x) = \epsilon_{\mu\nu\lambda} dx^\mu dx^\nu dx^\lambda$ is the volume element in three-space, with $dx^\mu dx^\nu$ corresponding to the **area** swept out by the deformed curve, and dx^λ to the **tangent direction** of the original curve. Thus a “flat” deformation (e.g. one that occurs entirely in a plane) will have zero variation. On the other hand $\delta\langle K \rangle$ will be nonzero for a twisting deformation such as



This is where the framing information appears in the functional integral. We see that for planar link diagrams **the functional integral is necessarily an invariant of regular isotopy**.

The term $\sum_a T_a T_a$ is the Casimir operator in the Lie algebra. We will use its special properties for $SU(N)$ shortly.

This entire discussion is only accurate for k large, and in this context it is convenient to normalize the volume form $[\epsilon_{\mu\nu\lambda} dx^\mu dx^\nu dx^\lambda]$ so that its values are $+1, -1$ or \emptyset . That is, I shall only use cases where the $\mu\nu\lambda$ frame is degenerate or orthogonal, and then multiply the integral by an (implicit) normalization constant. The normalized variation will be written:

$$\delta\langle K \rangle = \frac{-4\pi i}{k} \int dA e^{(ik/4\pi)\mathcal{L}} \sum(x) \left[\sum_a T_a T_a \right] \langle K | A \rangle$$

with $\sum(x) = \pm 1$ or \emptyset .

A Generalized Skein Relation.

We are now in a position to apply the variational equation of 17.4 to obtain a formula for the difference $\langle \text{X} \text{X} \rangle - \langle \text{X} \text{X} \text{X} \rangle$. I call this a generalized skein relation for the invariant $\langle K \rangle$. The result is as follows.

Theorem 17.5. Let $\langle K \rangle$ denote the functional integral invariant of regular isotopy defined $\langle K \rangle = \int dA e^{(ik/4\pi)\mathcal{L}} \langle K | A \rangle$ as described above. Let $\{T_a\}$ denote the generators for the Lie algebra in the given representation of the gauge group G . Then for large k , the following switching identity is valid

$$\langle \text{X} \text{X} \rangle - \langle \text{X} \text{X} \text{X} \rangle = z \langle \text{O} \text{C} \text{X} \rangle$$

where

$$(i) \ z = e^{-(2\pi i/k)} - e^{(2\pi i/k)} \simeq -4\pi i/k$$

(ii) $\mathcal{C} = \sum_a T_a T'_a$ denotes the insertion of T_a at x in the line and a second T_a insertion in the line . (Each segment receives a separate insertion.)

Proof. Write $\langle \text{---} \rangle - \langle \text{---} \rangle = \Delta_+ - \Delta_-$ where $\Delta_+ = \langle \text{---} \rangle - \langle \text{---} \rangle$ and $\Delta_- = \langle \text{---} \rangle - \langle \text{---} \rangle$.

The notation denotes the result of replacing the crossing by a graphical vertex. Let x denote the vertex:



In order to apply 17.4 to the case of a deformation from to (or to) we must reformulate the proof to include the conditions of self-crossing, since both parts of the Wilson loop going through x will participate in the calculation. We can assume that only one segment moves in the deformation. Therefore, the first part of the calculation, giving

$$\delta \langle K | A \rangle = i dx^\mu dx^\nu F_{\mu\nu}^a T_a \langle K | A \rangle$$

refers to a T_a insertion in the moving line. After the integration by parts, the functional derivative $\delta \langle K | A \rangle / \delta A_\lambda^a(x)$ applies to both segments going through x . However, the resulting volume for $\sum(x)$ for the differentiation in the direction of the moving line is zero - since we assume that the line is deformed parallel to its tangential direction. Thus if we write $\langle K | A \rangle = \mathcal{O} \mathcal{O}'$ where \mathcal{O} denotes that part of the Wilson loop (as a product of matrices $(1 + i dx^\mu A^\mu)$) containing the moving segment and \mathcal{O}' the stationary segment, then

$$\frac{\delta \langle K | A \rangle}{\delta A_\lambda^a(x)} = i dx^\mu T_a \mathcal{O} \mathcal{O}' + i dx'^\mu T_a \mathcal{O} \mathcal{O}'$$

where the two terms refer to insertions at x in \mathcal{O} and \mathcal{O}' respectively. Since the functional integration wipes out the first term, we conclude that

$$\delta \langle K \rangle = \frac{-4\pi i}{k} \int dA e^{(ik/4\pi)\mathcal{C}} \sum(x) \left[\sum_a T_a T'_a \right] \langle K | A \rangle$$

where T_a is inserted in the moving line, while $T'_a = T_a$ is inserted in the stationary line.

Now, in fact, we are considering two instances of $\delta\langle K \rangle$, namely Δ_+ and Δ_- as described at the beginning of this proof. Since the volume element $\sum(x)$ is positive for Δ_+ and negative for Δ_- , we find that $\Delta_- = -\Delta_+$ and that $\Delta_+ - \Delta_- = 2\Delta_+$. This difference is the significant one for the calculation. By convention, we take the full volume element from  to  to be +1 so that

$$\left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle - \left\langle \begin{array}{c} \searrow \\ \nearrow \end{array} \right\rangle = \frac{-4\pi i}{k} \int dA e^{(ik/4\pi)\mathcal{L}} \left[\sum_a T_a T'_a \right] \langle K | A \rangle.$$

This formula is the conclusion of the theorem, written in integral form. This completes the proof. //

We now give two applications of Theorem 17.5. The first is the case of an abelian gauge (say $G = U(1)$). In this case the Lie algebra elements are commuting scalars. Hence the formula of 17.5 becomes

$$\left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle - \left\langle \begin{array}{c} \searrow \\ \nearrow \end{array} \right\rangle = \frac{-4\pi i c}{k} \left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle$$

where c is a constant, and there is no extra matrix insertion at the crossing. Now note that with $\Delta_+ = -\Delta_-$ (see the Proof of 17.5) we have $\Delta_+ + \Delta_- = 0$. Hence $0 = \left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle + \left\langle \begin{array}{c} \searrow \\ \nearrow \end{array} \right\rangle - 2\left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle$. Therefore, in abelian gauge,

$$\left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle - \left\langle \begin{array}{c} \searrow \\ \nearrow \end{array} \right\rangle = \frac{-2\pi i c}{k} \left[\left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle + \left\langle \begin{array}{c} \searrow \\ \nearrow \end{array} \right\rangle \right].$$

Hence

$$\left(1 + \frac{2\pi i c}{k}\right) \left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle - \left(1 - \frac{2\pi i c}{k}\right) \left\langle \begin{array}{c} \searrow \\ \nearrow \end{array} \right\rangle = 0$$

or

$$\left\langle \begin{array}{c} \nearrow \\ \searrow \end{array} \right\rangle = e^{(4\pi i c/k)} \left\langle \begin{array}{c} \searrow \\ \nearrow \end{array} \right\rangle,$$

in this large k approximation.

Since switching a crossing just changes the functional integral by a phase factor, we see that linking and writhing numbers appear naturally in the abelian

context: Letting $x = e^{-4\pi i c/k}$, we have that $\langle \text{X} \rightarrow \text{X} \rightarrow \rangle = x \langle \text{X} \rightarrow \text{X} \rightarrow \rangle$. By assumption in the heuristic, there is an $\alpha \in \mathbb{C}$ such that

$$\langle \text{O} \rightarrow \rangle = \alpha \langle \text{C} \rightarrow \rangle$$

and

$$\langle \text{O} \leftarrow \rangle = \alpha^{-1} \langle \text{C} \rightarrow \rangle.$$

Thus α is determined by the equation

$$\begin{aligned} \langle \text{O} \rightarrow \rangle &= x \langle \text{O} \rightarrow \rangle \\ \Rightarrow \alpha &= x \alpha^{-1} \\ \Rightarrow \alpha^2 &= x \\ \Rightarrow \alpha &= x^{1/2}. \end{aligned}$$

We also have $\langle K \otimes \rangle = \delta \langle K \rangle$ for some δ , and can assume that $\langle \text{O} \circlearrowleft \rangle = \delta$ as well. With this in mind, it is easy to see that for $K = K_1 \cup K_2 \cup \dots \cup K_n$, a link of n components

$$\langle K \rangle = \left(\sum_{x \neq j} \ell k(K_i, K_j) + \sum_i w(K_i)/2 \right) \delta^n$$

where ℓk denotes linking number and w denotes writhe.

All of these remarks can be extended to normalized versions for framed links in three-dimensional space (or an arbitrary three-manifold), but it is very interesting to see how simply the relationship of the abelian gauge with linking numbers appears in this heuristic.

The second application is the $SU(N)$ gauge group in the fundamental representation. Here we shall find the Homfly polynomial, and for $N = 2$ the original Jones polynomial. For $N = 2$, the fundamental representation of $SU(2)$ is given by $T_a = \sigma_a/2$ where $\sigma_1, \sigma_2, \sigma_3$ are the Pauli matrices.

$$\sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

Note that $[\sigma_a, \sigma_b] = i \epsilon_{abc} \sigma_c$ and that $\text{tr}(T_a T_b) = (\delta_{ab})/2$. In order to indicate the form of the fundamental representation for $SU(N)$, here it is for $SU(3)$:

$$\begin{aligned}\lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \lambda_3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \\ \lambda_5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ \lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.\end{aligned}$$

Then $T_a = \lambda_a/2$.

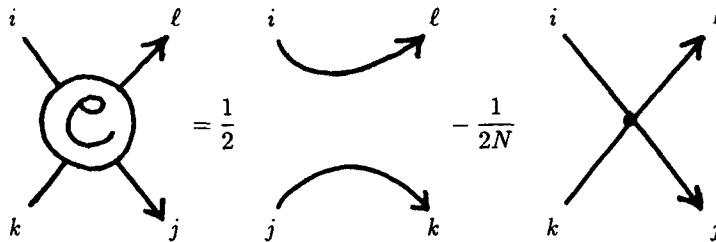
The analog for $SU(N)$ should be clear. The nondiagonal matrices have two nonzero entries of i and $-i$ or 1 and -1 . The diagonal matrices have a string of k 1's, followed by $-k$, followed by zeroes - and a suitable normalization so that $\text{tr}(T_a T_b) = \frac{1}{2} \delta_{ab}$ as desired.

Now we use the **Fierz identity** for $SU(N)$ in the fundamental representation:

$$\sum_a (T_a)_{ij} (T_a)_{kl} = \frac{1}{2} \delta_{il} \delta_{jk} - \frac{1}{2N} \delta_{ij} \delta_{kl}.$$

(**Exercise:** Check this identity.)

Applying this identity to 17.5, we find $\mathcal{C} = \sum_a T_a T_a'$



Hence, the generalized skein identity

$$\langle \overleftrightarrow{\gamma} \rangle - \langle \overleftrightarrow{\gamma} \rangle = \frac{-4\pi i}{k} \langle \circlearrowleft c \circlearrowright \rangle$$

becomes, for $SU(N)$,

$$\langle \overleftarrow{\text{X}} \rangle - \langle \overrightarrow{\text{X}} \rangle = \frac{-2\pi i}{k} \langle \overleftrightarrow{\text{X}} \rangle - \frac{2\pi i}{Nk} \langle \overrightarrow{\text{X}} \rangle.$$

Using $\langle \overrightarrow{\text{X}} \rangle = \frac{1}{2}(\langle \overleftarrow{\text{X}} \rangle + \langle \overrightarrow{\text{X}} \rangle)$, as in the abelian case, we conclude

$$\left(1 - \frac{\pi i}{Nk}\right) \langle \overleftarrow{\text{X}} \rangle - \left(1 + \frac{\pi i}{Nk}\right) \langle \overrightarrow{\text{X}} \rangle = \frac{-2\pi i}{k} \langle \overleftrightarrow{\text{X}} \rangle.$$

Now we use this identity to determine α such that $\langle \overrightarrow{\text{O}} \rangle = \alpha \langle \overleftrightarrow{\text{O}} \rangle$: Let $\beta = 1 - \pi i/Nk$, $z = -2\pi i/k$. Then the identity above is equivalent to

$$\beta \langle \overleftarrow{\text{O}} \rangle - \beta^{-1} \langle \overrightarrow{\text{O}} \rangle = z \langle \overleftrightarrow{\text{O}} \rangle.$$

Hence

$$\beta \langle \overrightarrow{\text{O}} \rangle - \beta^{-1} \langle \overleftarrow{\text{O}} \rangle = z \langle \overleftrightarrow{\text{O}} \rangle.$$

The extra loop has value N since (compare with the Fierz identity) we are tracing an $N \times N$ identity matrix. Thus

$$\beta \alpha \langle \overleftrightarrow{\text{O}} \rangle - \beta^{-1} \alpha^{-1} \langle \overleftrightarrow{\text{O}} \rangle = zN \langle \overleftrightarrow{\text{O}} \rangle$$

or

$$\begin{aligned} \beta \alpha - (\beta \alpha)^{-1} &= \left(1 - \frac{\pi i N}{k}\right) - \left(1 + \frac{\pi i N}{k}\right) \\ \beta \alpha - (\beta \alpha)^{-1} &= x^N - x^{-N} \end{aligned}$$

with $z = x - x^{-1}$. Thus $\beta \alpha = x^N$.

It is then easy to see that for the writhe-normalized

$$P_K = \alpha^{-w(K)} \langle K \rangle$$

we have

$$(\beta \alpha) P_{\overleftarrow{\text{X}}} - (\beta \alpha)^{-1} P_{\overrightarrow{\text{X}}} = z P_{\overleftrightarrow{\text{X}}}.$$

Thus $x^N P_{\overleftarrow{\text{X}}} - x^{-N} P_{\overrightarrow{\text{X}}} = (x - x^{-1}) P_{\overleftrightarrow{\text{X}}}$. This gives the specialization of the

Homfly polynomial. Note that they correspond formally to the types of specialization available from the state models built via the Yang-Baxter Equation and the $SL(N)$ quantum groups in Chapter 11. A similar relationship holds for the Kauffman polynomial and the $SO(N)$ gauge in the fundamental representation. See [WIT3].

Beyond Integral Heuristics.

The next stage beyond these simple heuristics is to consider the large k limit for the three-manifold functional integrals

$$Z_{M^3} = \int dA e^{(ik/4\pi)\mathcal{L}_{M^3}}.$$

In Witten's paper [WIT2] he shows how this leads to the Ray-Singer torsion (analytic Reidemeister torsion) for the three-manifold M^3 , and to eta invariants related to the phase factor. See also [BN] and [GMM] for a discussion of the perturbation expansion of the functional integral. Quantization leads to relationships with conformal field theory: See Witten [WIT2] and also [MS], [CR1], [CR2]. Witten explains his beautiful idea that can rewrite, via surgery on links in the three-manifold M^3 , the functional integral as a sum over link invariants in the three-sphere. This expresses Z_{M^3} in a form that is essentially equivalent to our descriptions of three-manifold invariants in section 16⁰.

It is worth dwelling on the idea of this surgery reformulation, as it leads directly to the patterns of topological quantum field theory and grand strategies for the elucidation of invariants. The surgery idea is this: Suppose that M' is obtained from M by surgery on an embedded curve $K \subset M$. Then we are looking at a way to change the functional integral $Z_M = \int dA e^{(ik/4\pi)\mathcal{L}_M}$ by changing the three-manifold via surgery. We already know other ways to change the functional integral relative to an embedded curve, namely by adding a Wilson loop along that curve. Therefore Witten suggested writing $Z_{M'}$ as a sum of integrals of the form

$$Z_M(K, G_\lambda) = \int dA e^{(ik/4\pi)\mathcal{L}_M} \langle K|A \rangle_\lambda$$

where $\langle K|A \rangle$ denotes the value of the Wilson loop in the representation G_λ of the gauge group G . Applied assiduously, this technique re-writes Z_M as a sum of link

invariants of the surgery curves in S^3 . Much remains to be understood about this point of view, but the specific constructions of Reshetikhin, Turaev and Lickorish's work with the bracket (see Chapter 16) show that this approach via surgery to the three-manifold invariants is highly non-trivial.

The idea of looking at the functional integral on a decomposition of the three-manifold is very significant. Consider a Heegaard decomposition of the three-manifold M^3 . That is, let $M^3 = M_1^3 \cup_{\psi} M_2^3$ where M_1^3 and M_2^3 are standard solid handlebodies with boundary a surface F and pasting map $\psi : F \rightarrow F$. The diffeomorphism ψ is used to glue M_1 and M_2 to form M . Then we can write (formally)

$$Z_M = \langle M_1 | \psi | M_2 \rangle$$

in the sense that these data are sufficient to do the functional integral and that $\langle M_1 |$ can be regarded as a **functional** on pairs (ψ, M_2) ready to compute an invariant by integrating over $M_1 \cup_{\psi} M_2$. In this sense, the functional integral leads to the notion of a Hilbert space \mathcal{H} of functionals $\langle M_1 |$ and a dual space \mathcal{H}^* with its ket $|M_2\rangle$. A handlebody M_1 gives rise to a "vacuum vector" $\langle M_1 | \in \mathcal{H}$ and the three-manifold invariant (determined up to a phase) is the inner product $\langle M_1 | \psi | M_2 \rangle$ essentially determined by the surface diffeomorphism $\psi : F \rightarrow F$.

Thus formally, we see that the Hilbert space \mathcal{H} depends only on F and that there should be an intrinsic description of the three-manifold invariant via $\mathcal{H}(F)$, an inner product structure and the pasting data ψ . This goal has been accomplished via the use of conformal field theory on the surface. (See [CR1], [CR2], [CLM], [KO2], [SE], [MS].) The result is a definition of these invariants that does not directly depend upon functional integrals. Nevertheless, the functional integral approach is directly related to the conformal field theory, and it provides an organizing center for this entire range of techniques.

With these remarks, I reluctantly bring Part I to a close. The journey into deeper mysteries of conformal field theory and the functional integral must wait for the next time.

Remark. Figure 21 illustrates the form of the basic integration by parts maneuver that informs this section. This maneuver is shown in schematized form.

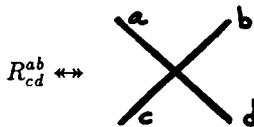
$F_{\mu r}^a \equiv \text{Diagram } 1$, $E_{\mu r \lambda} \equiv \text{Diagram } 2$ $\delta \mathcal{L} / \delta A_{\mu i}^a \equiv \text{Diagram } 3$, $\int \text{Diagram } 4 = - \int \text{Diagram } 5$ $\text{Diagram } 6 = \text{Diagram } 7 \iff F_{\mu r}^a = E_{\mu r \lambda} \delta \mathcal{L} / \delta A_{\lambda}^a$
$dx^i \equiv \Delta \text{ or } \Upsilon$, $T_a \equiv \square \text{ or } T$ $R \iff \text{Wilson Loop}$, $\delta R = \text{Diagram } 8 R$ $\text{Diagram } 9 = \Upsilon \square R$, $\langle \kappa \rangle = \int e^{k\mathcal{L}} R$
$\delta \langle \kappa \rangle = \int e^{k\mathcal{L}} \delta R = \int e^{k\mathcal{L}} \text{Diagram } 10 R$ $= \int e^{k\mathcal{L}} \text{Diagram } 11 R = \frac{1}{k} \int \text{Diagram } 12 e^{k\mathcal{L}} R$ $= -\frac{1}{k} \int e^{k\mathcal{L}} \text{Diagram } 13 R = -\frac{1}{k} \int e^{k\mathcal{L}} \text{Diagram } 14 R$ $\Rightarrow \langle \rightarrow \rightarrow \rangle - \langle \leftarrow \leftarrow \rangle = \bar{z} \langle \text{Diagram } 15 \rangle$

Figure 21

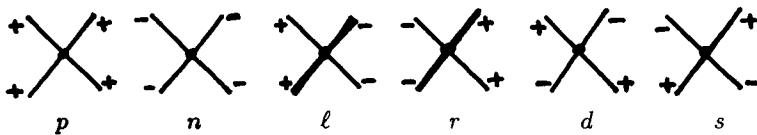
18⁰. Appendix – Solutions to the Yang-Baxter Equation.

The purpose of this appendix is to outline the derivation of conditions for a two-index spin-preserving solution to the Yang-Baxter Equation.

Consider a solution $R = R_{cd}^{ab}$ of the Yang-Baxter Equation (without rapidity parameter) where the indices all belong to the set $\mathcal{I} = \{-1, +1\}$. Assume also that the solution is **spin-preserving** in the sense that $R_{cd}^{ab} \neq 0$ only when $a+b = c+d$. There are then exactly six possible choices of spin at a vertex



The six possibilities are shown below



The lower-case letters (p, n, l, r, d, s) next to each local state designate the corresponding vertex weight for the matrix R . Thus we can write R in matrix form as

$$R = \begin{array}{c|cccc} & \text{--} & \text{-+} & \text{+-} & \text{++} \\ \text{--} & n & 0 & 0 & 0 \\ -+ & 0 & r & d & 0 \\ +- & 0 & s & l & 0 \\ ++ & 0 & 0 & 0 & p \end{array}$$

The Yang-Baxter Equation (without rapidity parameter) reads

$$\sum_{i,j,k} R_{ij}^{ab} R_{kf}^{jc} R_{de}^{ik} = \sum_{i,j,k} R_{ij}^{bc} R_{dk}^{ai} R_{ef}^{kj}.$$

The main result of this appendix is the

Theorem. $R = \begin{bmatrix} n & 0 & 0 & 0 \\ 0 & r & d & 0 \\ 0 & s & \ell & 0 \\ 0 & 0 & 0 & p \end{bmatrix}$ is a solution to the Yang-Baxter Equation if and only if the following conditions are satisfied:

$$\left\{ \begin{array}{l} rld = 0 \\ rls = 0 \\ r\ell(\ell - r) = 0 \\ p^2\ell = pl^2 + lds \\ n^2\ell = nl^2 + lds \\ p^2r = pr^2 + rds \\ n^2r = nr^2 + rds \end{array} \right\} (*)$$

Proof (sketch). As we have noted above, the Yang-Baxter Equation has the form

$$\mathcal{L}_{def}^{abc} = \mathcal{R}_{def}^{abc}$$

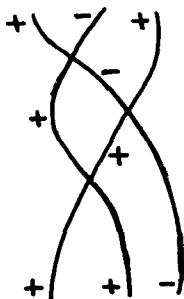
where \mathcal{L} and \mathcal{R} denote the left and right halves of the equation and the indices $a, b, c, d, e, f \in \mathcal{I} = \{-1, +1\}$. Thus YBE in our case reduces to a set of 64 specific equations, one for each choice of a, b, c, d, e, f . These are further reduced by the requirement of spin-preservation. Thus, for any given choice of a, b, c, d, e, f we must, to construct \mathcal{L} , find i, j, k so that

$$\left. \begin{array}{lcl} a+b & = & i+j \\ j+c & = & k+f \\ i+k & = & d+e \end{array} \right\} \mathcal{L}$$

Let $(a, b, c/i, j, k/d, e, f)$ denote a specific solution to this condition. Since a configuration for \mathcal{L} is also a configuration for \mathcal{R} (turn it upside down!), it suffices to enumerate the 9-tuples for \mathcal{L} , in order to enumerate the equation. Here is the list of 9-tuples.

- | | |
|--|---|
| 0. $(+++ / + ++ / + ++)$ | 4. $\left\{ \begin{array}{l} (-+ + / -++ / -++) \\ (-+ + / +-- / -++) \end{array} \right\}$ |
| 1. $(++-/+++ / +++)$
$(++-/++- / +-+)$
$(++-/++- / -++)$ | $\left\{ \begin{array}{l} (-++ / +-- / +-+) \\ (-++ / -++ / +-+) \\ (-++ / + - / +-+) \end{array} \right\}$ |
| 2. $\left\{ \begin{array}{l} (+-+/+-- / +-+) \\ (+-+/--+ / +-+) \\ (+-+/ -++ / +-+) \\ (+-+/+-- / -++) \\ (+-+/+ - / +-+) \end{array} \right\}$ | 5. $\left\{ \begin{array}{l} (-+-/-++ / -++) \\ (-+-/ +-- / -++) \\ (-+-/ +-- / +-+) \\ (-+-/ -++ / +-+) \\ (-+-/ -+ - / -++) \end{array} \right\}$ |
| 3. $\left\{ \begin{array}{l} (+--/+-- / +-+) \\ (+--/-++ / +-+) \\ (+--/ -++ / +-+) \\ (+--/ +-- / -++) \\ (+--/ -+ - / -++) \end{array} \right\}$ | 6. $(---/+--- / ---)$
$(---/ -++ / -+-)$
$(---/ -+ - / +--)$ |
| | 7. $(---/ --- / ---)$ |

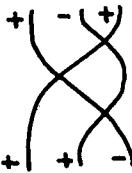
All the equations can be read from this list of admissible 9-tuples. For example, suppose that we want the equations for $(a, b, c) = (+, -, +)$ and $(d, e, f) = (+, +, -)$. Then the list tells us that the only \mathcal{L} configuration available is $(-+/+ - / + + -)$ corresponding to



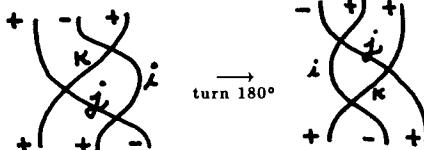
with product of
vertex weights ℓ_{sp} .

This look-up involves checking entries that begin with $(- + +)$. Now for the

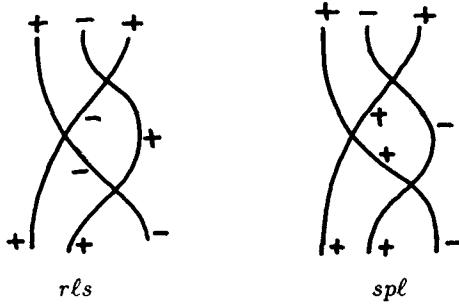
\mathcal{R} -configurations, we want



and this corresponds to the \mathcal{L} -configuration



We find $(-++/+--/+--)$ and $(-++/-++/+--)$ with configurations and vertex weights:



Thus this specific equation $\mathcal{L} = \mathcal{R}$ is

$$\ellsp = rls + spl$$

or

$$0 = rls.$$

Using this table of allowable configurations it is an easy task to enumerate the full list of equations. There are many repetitions, and the complete list of equations is exactly as given in the statement of the theorem. This completes the proof. //

Remark. The six basic configurations for these solutions are in correspondence with the six basic local configurations of the six-vertex model [BA1] in statistical mechanics. In particular, they correspond to the local **arrow configurations** of the **ice-model**. In the ice-model the edges of the lattice (or more generally, a

4-valent graph) are given orientations so that **at each vertex two arrows go in and two arrows go out**. This yields six patterns



By choosing a direction (e.g. “up” as shown below) and labelling each line + or – as it goes with or against this direction, we obtain the original spin-preserving labellings.



Jones and Alexander Solutions.

We leave the full consequences of the theorem to the reader. But it should be noted that the special case $r = 0$, $\ell \neq 0$, $d = s = 1$ yields the equations

$$\begin{aligned} p^2 &= p\ell + 1 \\ n^2 &= n\ell + 1 \end{aligned}$$

Thus assuming $p \neq 0 \neq n$, we have $p - p^{-1} = n - n^{-1}$. This has **two** solutions: $p = n$ and $pn = -1$. As we have previously noted (section 8⁰) $p = n$ gives the YB solution for the $SL(2)$ quantum group and the Jones polynomial, while $pn = -1$ gives the YB solution derived from the Burau representation and yielding the Alexander-Conway polynomial (sections 11⁰ and 12⁰).

PART II

PART II. KNOTS AND PHYSICS – MISCELLANY.

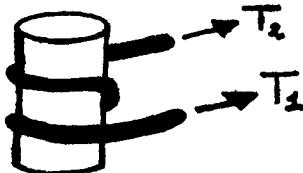
This half of the book is devoted to all manner of speculation and rambling over the subjects of knots, physics, mathematics and philosophy. If all goes well, then many tales shall unfold in these pages.

1⁰. Theory of Hitches.

This section is based on the article [BAY].

We give a mathematical analysis of the properties of **hitches**. A **hitch** is a mode of wrapping a rope around a post so that, with the help of a little friction, the rope holds to the post. And your horse does not get away.

First consider simple wrapping of the rope around the post in coil-form:



Assume that there are an integral number of windings. Let tensions T_1 and T_2 be applied at the ends of the rope. Depending upon the magnitudes (and relative magnitudes) of these tensions, the rope may slip against the post.

We assume that there is some friction between rope and post. It is worth experimenting with this aspect. Take a bit of cord and a wooden or plastic rod. Wind the cord one or two times around the rod. Observe how easily it slips, and how much tension is transmitted from one end of the rope to the other. Now wind the cord ten or more times and observe how little slippage is obtained – practically no counter-tension is required to keep the rope from slipping.

In general, there will be no slippage in the T_2 -direction so long as

$$T_2 \leq \kappa T_1$$

for an appropriate constant κ . This constant κ will depend on the number of windings. The more windings, the larger the constant κ .

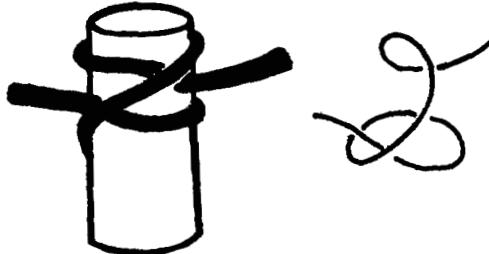
A good model is to take κ to be an exponential function of the angle (in radians) that the cord is wrapped around the rod, multiplied by the coefficient of friction between cord and rod. For simplicity, take the coefficient of friction to be unity so that

$$\kappa = e^{\theta/2\pi}$$

where θ is the total angle of rope-turn about the rod.

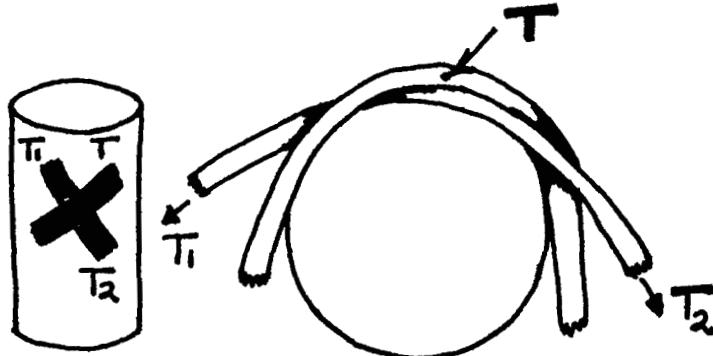
Thus, for a single revolution we need $T_2 \leq eT_1$ and for an integral number n of revolutions we need $T_2 \leq e^n T_1$ to avoid slippage.

A real hitch has "wrap-overs" as well as windings:



Clove Hitch

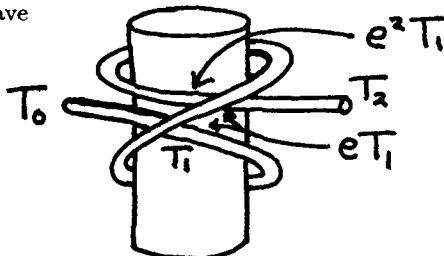
Here, for example, is the pattern of the clove hitch. In a wrap-over, under tension, the top part squeezes the bottom part against the rod.



Hold Fast $T_2 \leq T_1 + uT$

This squeezing produces extra protection against slippage. If, at such a wrap-over point, the tension in the overcrossing cord is T , then the undercrossing cord will hold-fast so long as $T_2 \leq T_1 + uT$ where u is a certain constant involving the friction of rope-to-rope, and T_2 and T_1 are the tensions on the ends of the undercrossing rope at its ends.

With these points in mind, we can write down a series of inequalities related to the crossings and loopings of a hitch. For example, in the case of the clove hitch we have



$$\begin{aligned} T_1 &\leq T_0 + ueT_1 \\ T_2 &\leq e^2T_1 + ueT_1 \end{aligned}$$

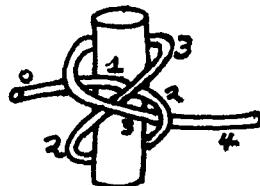
that the equations necessary to avoid slippage are:

$$T_1 \leq T_0 + ueT_1$$

$$T_2 \leq e^2T_1 + ueT_1.$$

Since the first inequality holds whenever $ue > 1$ or $u > 1/e$, we see that the clove hitch will not slip no matter how much tension occurs at T_2 just so long as the rope is sufficiently rough to allow $u > 1/e$.

This explains the efficacy of this hitch. Other hitches can be analyzed in a similar fashion.



$$\begin{aligned} T_1 &\leq T_0 + ueT_3 \\ T_2 &\leq T_1 + uT_3 \\ T_3 &\leq eT_2 + ueT_3 \\ T_4 &\leq eT_3 + uT_2 \end{aligned}$$

In this example, if we can solve

$$T_1 \leq ueT_3$$

$$T_2 \leq T_1 + uT_3$$

$$T_3 \leq eT_2 + ueT_3$$

then the hitch will hold.

T_1	T_2	T_3
1	0	$-ue$
-1	1	$-u$
0	$-e$	$1 - ue$

In matrix form, we have

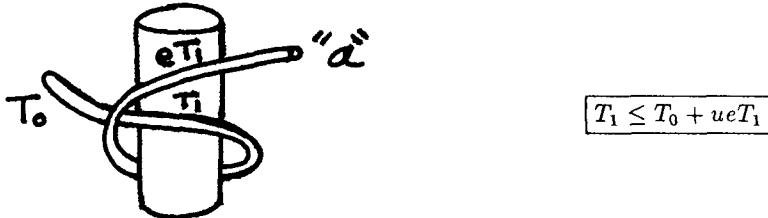
$$\begin{bmatrix} 1 & 0 & -ue \\ -1 & 1 & -u \\ 0 & -e & 1 - ue \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} \leq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

The determinant of this matrix is

$$1 - ue(2 - e).$$

Thus the critical case is $u = 1/e(2 + e)$. For $u > 1/e(2 + e)$, the hitch will hold.

Remark. Let's go back to the even simpler "hitch":



Our abstract analysis would suggest that this will hold if $ue > 1$. However, there is no stability here. A pull at "a" will cause the loop to rotate and then the "u-factor" disappears, and slippage happens. A pull on the clove hitch actually tightens the joint.

This shows that in analyzing a hitch, we are actually taking into account some properties of an already-determined-stable mechanical mechanism that happens to be made of rope. [See also Sci. Amer., Amateur Sci., Aug. 1983.]

There is obviously much to be done in understanding the frictional properties of knots and links. These properties go far beyond the hitch to the ways that ropes interplay with one another. The simplest and most fascinating examples are

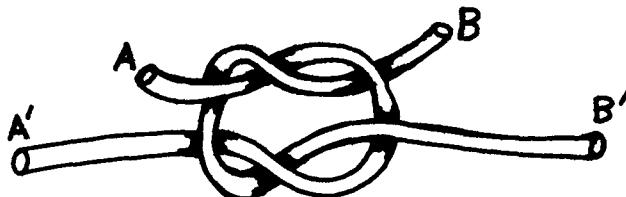
the **square knot** and the **granny knot**. The square knot pulls in under tension, each loop constricting itself and the other - providing good grip:



Square Knot

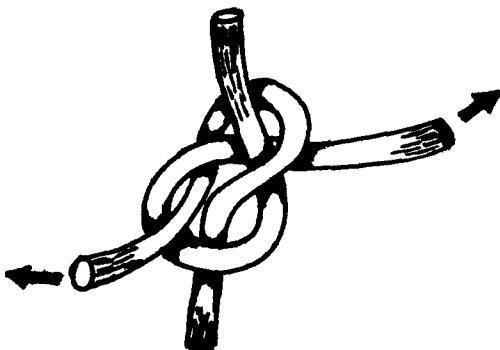
Construct this knot and watch how it grips itself.

The **granny** should probably be called the **devil**, it just won't hold under tension:



Granny Knot

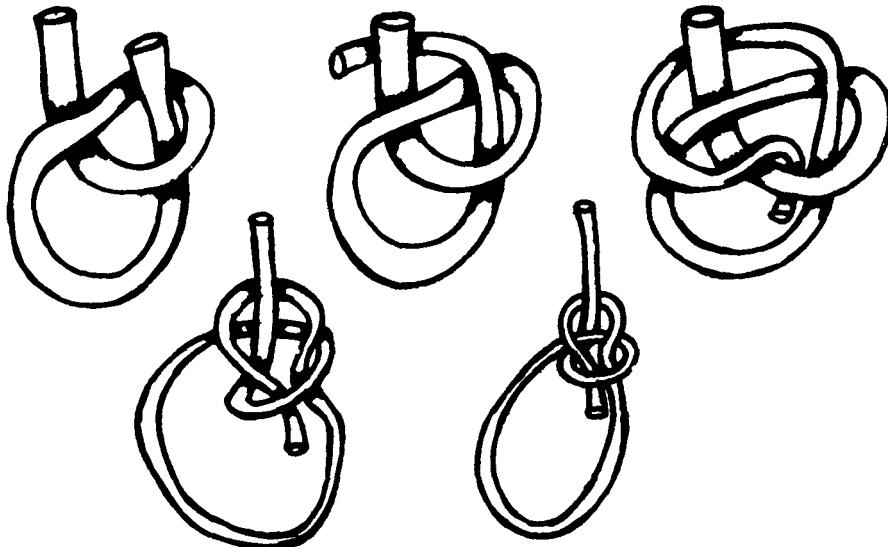
Try it! Ends *A* and *B*, are twisted perpendicular to ends *A'* and *B'* and the rope will feed through this tangle if you supply a sufficient amount of tension.



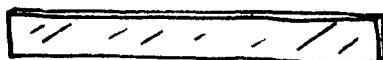
The fact of the matter is that splices and hitches are fantastic sorts of mechanical devices. Unlike the classical machine that is composed of well-defined parts that interact according to well-understood rules (gears and cogs), the sliding interaction of two ropes under tension is extraordinary and interactive, with tension, topology and the system providing the form that finally results.

Clearly, here is an arena where a deeper understanding of **topology instantiated in mechanism** is to be desired. But here we are indeed upon untrodden ground. The topology that we know has been obtained at the price of initial abstraction from these physical grounds. Nevertheless, it is the intent of these notes to explore this connection.

We do well to ponder the knot as whole system, decomposed into parts only via projection, or by an observer attempting to ferret out properties of the interacting rope. Here is a potent metaphor for observation, reminding us that the decompositions into parts are in every way our own doing - through such explications we come to understand the whole.



Tying the Bowline

2⁰. The Rubber Band and Twisted Tube.

(one ordinary supercoil)



“all rolled up”



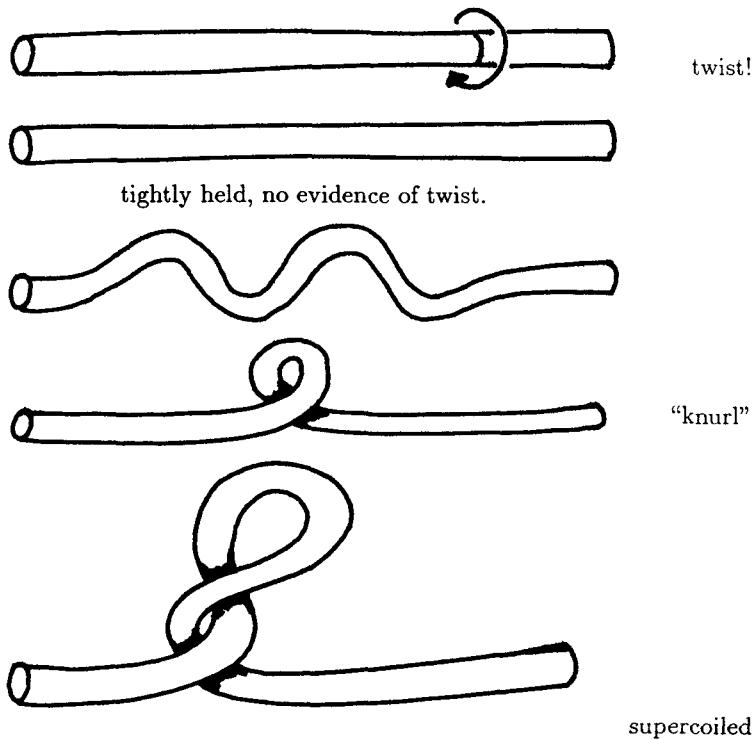
“the first “knurls” ”



all “knurled up”

If you keep twisting the band it will "knurl", a term for the way the band gets in its own way after the stage of being all rolled up. A knurl is a tight super-coil and if you relax the tension () on the ends of the band, knurls will "pop" in as you feel the twisted band relax into some potential energy wells. It then takes a correspondingly long pull and more energy to remove the knurls.

The corresponding phenomena on a twisted - spring-loaded tube are even easier to see:



In all these cases, it is best to do the experiment. The interesting phenomenon is that once the first knurl has formed, it takes a lot of force to undo it due to the

interaction of the tube against itself.

The energetics of the situation demand much experimentation. I am indebted to Mr. Jack Armel for the following experiment.

The Armel Effect. Take a flat rubber band. Crosscut it to form a rubber strip. Paint one side black. Tape one end to the edge of a table. Take the other end in your hand and twist gently counterclockwise until the knurls hide one color. Repeat the experiment, but turn the band clockwise. Repeat the entire experiment using a strip of paper.

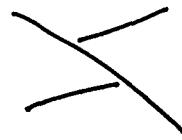
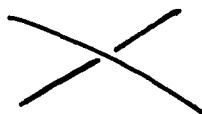
3^o. On a Crossing.

This section is a musing about the structure of a diagrammatic crossing.



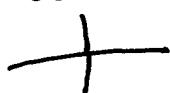
This (locally) consists in three arcs: The overcrossing arc, and two arcs for the undercrossing line. These two are drawn near the overcrossing line, and on opposite sides of it – indicating a line that continuously goes under the overcrossing line.

What happens if we vary the drawing? How much lee-way is there in this visual convention?

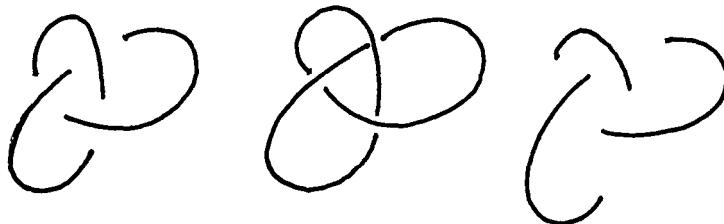


Obviously not much lee-way in regard to the meeting of the arcs for opposite sides.

And the gap??



Here there seems to be a greater latitude.



Too much ambiguity could get you into trouble.



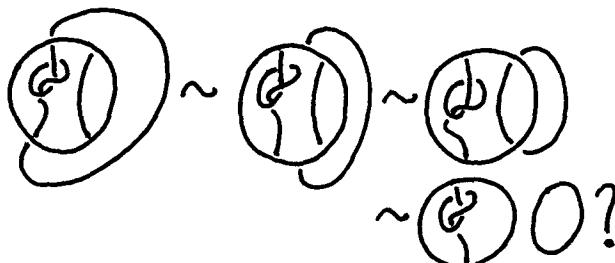
A typical arc in the diagram begins at one undercross and travels to the next. Along the way it is the overcrossing line for any other arcs. Could we generalize

knot theory to include "one-sided originations"?



What are the Reidemeister moves for the likes of these diagrams?

Will you allow:



What about

Surely

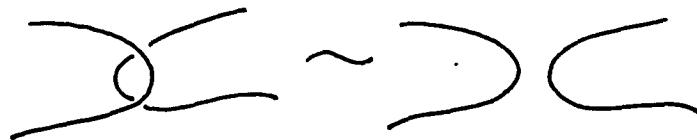


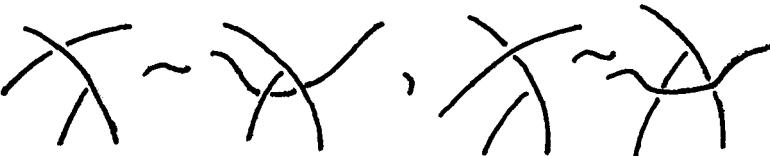
must be forbidden!

Would you like some rules?? Ok. How about

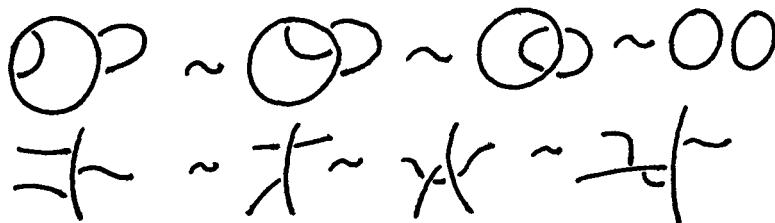
- 1.

- 2.

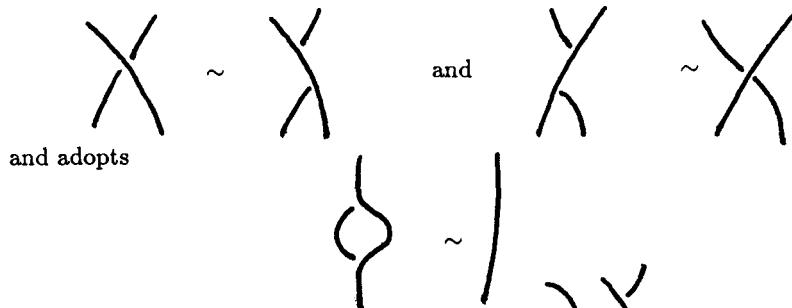


3. 
4. 
5. 

Call this **Imaginary Knot Theory**. Invariants anyone? (Compare with [K] and also with [TU4].)

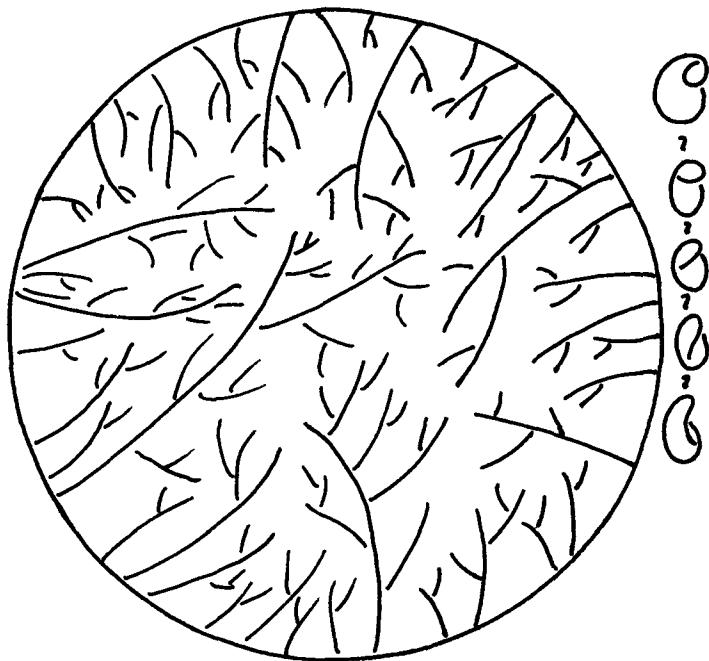


This idea of loosening the restrictions on a crossing has been independently invented in braid form by M. Khovanov [K]. In his system one allows



This makes the individual three-vertex elements   into analogs of three-vertices in recoupling theory (compare sections 12⁰ and 13⁰ of Part II). Outside

of braids, it seems worthwhile to include the equivalence  as well. We formalize this as **slide equivalence** in the next section.



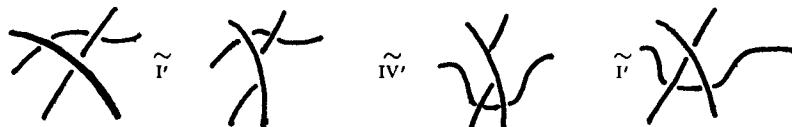
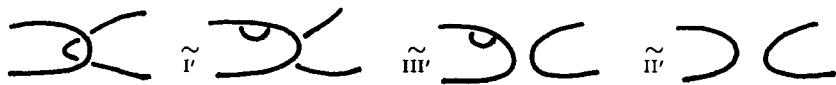
$$\begin{array}{ccccccc} \circ & \sim & \circ & \sim & \circ & \sim & \circ \\ \circ & \sim & \circ & \sim & \circ & \sim & \circ \\ \circ & \sim & \circ & \sim & \circ & \sim & \circ \\ \circ & \sim & \circ & \sim & \circ & \sim & \circ \\ \circ & \sim & \circ & \sim & \circ & \sim & \circ \end{array}$$

4^o. Slide Equivalence.

Let's formalize **slide equivalence** for unoriented diagrams as described in the last section. Let slide equivalence (denoted \sim) be generated by:

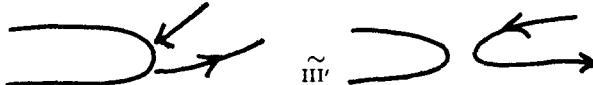


Note that the usual Reidemeister moves of type II and III are consequences of the axioms of slide equivalence:



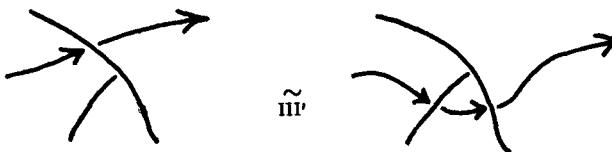
The oriented versions of the moves for slide equivalence should be clear. Moves I' and II' hold for any choice of orientations. Moves III' and IV' require that paired

arcs have compatible orientations. Thus



\tilde{III}'

and



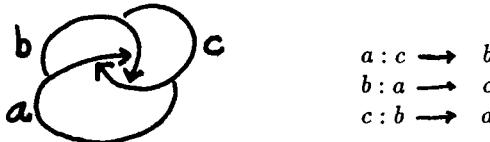
\tilde{III}''

Definition. A category $[M] C$ is said to be a **mirror category** if every object in C is also a morphism in C and every morphism in C is a composition of such objects.

We see that any oriented slide diagram can be regarded as a diagram of objects and morphisms for a mirror category. Thus



$$\begin{aligned} a : b &\rightarrow b \\ b : a &\rightarrow a \end{aligned}$$



$$\begin{aligned} a : c &\rightarrow b \\ b : a &\rightarrow c \\ c : b &\rightarrow a \end{aligned}$$

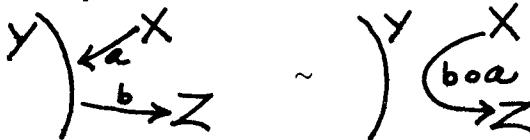
Of course the diagrams, being in the plane, have extra structure. Nevertheless, each of the slide-moves (I' to IV') expresses a transformation of diagrams that makes sense categorically. Thus I' asserts that



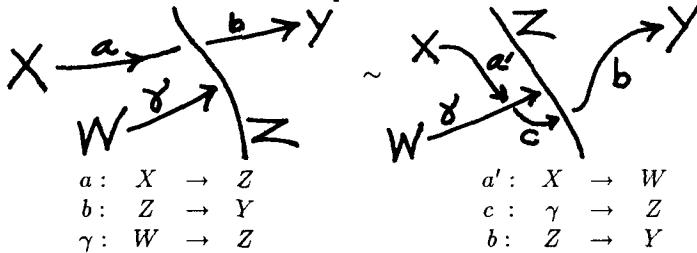
$$\begin{aligned} a : X &\rightarrow Y \\ b : Y &\rightarrow Z \end{aligned}$$

morphisms can be composed without diagrammatic obstruction. II' allows the insertion or removal of self-maps that are not themselves target objects. III' is

another form of composition.



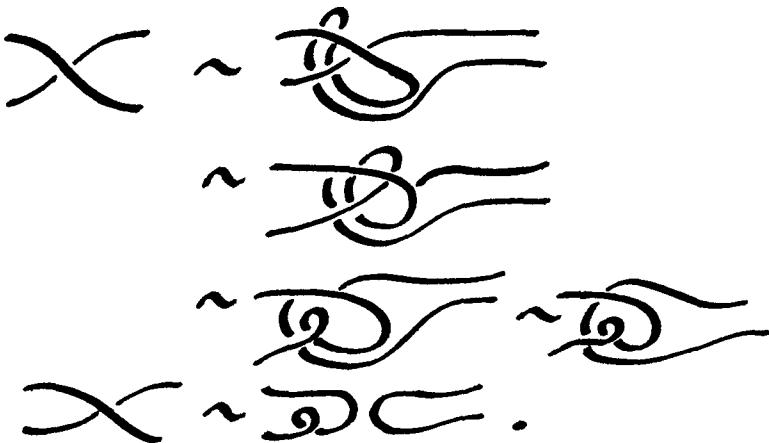
IV' gives different forms of local compositions as well:



Thus the abstract version of IV' in mirror category language is that given morphisms $a : X \rightarrow Z$, $b : Z \rightarrow Y$ and an object-morphism $\gamma : W \rightarrow Z$ then there exist morphisms $c : \gamma \rightarrow Z$, $a' : X \rightarrow \gamma$ such that $b \circ c \circ a' = b \circ a$.

This suggests defining a **special mirror category** as one that satisfies the various versions of IV' and the analog of II'. The slide diagrams are depictions of the generating object-morphisms for special mirror categories of planar type.

To return to diagrams in the category of side-equivalence, it is a very nice problem to find invariants. This is open territory. (Use orientation!)



5⁰. Unoriented Diagrams and Linking Numbers.

Let's note the following: **The linking number of a two-component link in absolute value depends only on the unoriented link.**

Yet we normally go through all the complexities of putting down orientations to determine the linking number. In this section I will show how to compute the absolute value of the linking number from an unoriented diagram. Let K be any unoriented link diagram. Let $K^{\mathcal{O}}$ be the result of orienting K . Thus \mathcal{O} belongs to one of the $2^{|K|}$ orientations of K where $|K|$ denotes the number of components of K .

Define $[K]$ by the formula

$$[K] = \sum_{\mathcal{O}} \alpha^{w(K^{\mathcal{O}})}.$$

Thus $[K]$ is the sum over all possible orientations of K of a variable α raised to the writhe of K with that orientation.

Remark. It is useful for display purposes to use a bracket in many different contexts and to denote different polynomials. Let us adopt the convention that a given bracket is specified by its last definition, unless otherwise specified.

Certainly, $[K] = \sum_{\mathcal{O}} \alpha^{w(K^{\mathcal{O}})}$ is a regular isotopy invariant.

If K_0 is a specific orientation of K then $\Lambda(K_0) = \alpha^{-w(K_0)}[K_0]$ will be an ambient isotopy invariant of the oriented link K_0 . Now

$$\begin{aligned} \Lambda(K_0) &= \sum_{\mathcal{O}} \alpha^{w(K^{\mathcal{O}})-w(K_0)} \\ &= 1 + \sum_{\mathcal{O} \neq \mathcal{O}_0} \alpha^{w(K^{\mathcal{O}})-w(K_0)} \end{aligned}$$

where \mathcal{O}_0 denotes the orientation of K_0 .

To see what this means, note the following: if $K_0 = L_1 \cup L_2 \cup \dots \cup L_n$ (its link components), then

$$w(K_0) = \sum_{i=1}^n w(L_i) + \sum_{i < j} 2\ell k(L_i, L_j).$$

Hence if K^O results in the changing of orientations on some subset of the L_i , we can denote this by

$$K^O = L_1^{\epsilon_1} \cup L_2^{\epsilon_2} \cup \dots \cup L_n^{\epsilon_n}$$

where $\epsilon_i = \pm 1$ (1 for same, -1 for changed). But $w(L_i^+) = w(L_i^-)$, hence

$$w(K^O) - w(K_0) = \sum_{i < j} 2(\ell k(L_i^{\epsilon_i}, L_j^{\epsilon_j}) - \ell k(L_i, L_j))$$

and

$$\text{i) } \left\{ \begin{array}{l} \text{or} \\ \epsilon_i = -1, \epsilon_j = +1 \\ \epsilon_i = +1, \epsilon_j = -1 \end{array} \right\} \Rightarrow \ell k(L_i^{\epsilon_i}, L_j^{\epsilon_j}) = -\ell k(L_i, L_j). \text{ Thus}$$

$$w(K^O) - w(K_0) = -4 \sum_{\substack{i < j \\ \epsilon_i \neq \epsilon_j}} \ell k(L_i^{\epsilon_i}, L_j^{\epsilon_j}).$$

Thus

$$\Lambda(K_0) = 1 - 4 \sum_{L, L' \subset K^O} \ell k(L, L')$$

where L, L' are a pair of different components of K^O such that exactly one of them has orientation reversed from the reference K_0 .

This is an exact description of the associated ambient isotopy invariant to $[K]$.

The simplest case is a link of two components $K = K_1 \cup K_2$. Then there are four possible orientations.

$$\begin{array}{lll} K_0 &= K_1^+ \cup K_2^+ & \text{Say } w(K_0) = w_0 \\ K_1 &= K_1^- \cup K_2^- & \Rightarrow \quad w(K_1) = w(K_0) \\ K_2 &= K_1^- \cup K_2^+ & \quad \quad \quad w(K_2) = w(K_3) \\ K_3 &= K_1^+ \cup K_2^- & \end{array}$$

Thus

$$\begin{aligned} [K] &= 2(\alpha^{w(K_1^+ \cup K_2^+)} + \alpha^{w(K_1^- \cup K_2^-)}) \\ \Rightarrow [K] &= 2(\alpha^{d_1} + \alpha^{d_2}) \end{aligned}$$

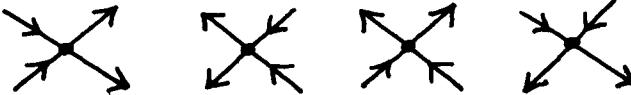
where

$$|d_1 - d_2| = 4|\ell k(K_1^+, K_2^+)|.$$

Thus, for a two-component link, $[K]$ calculates (via the difference of two exponents) the absolute linking number (≥ 0) of the two curves.

What I want to do now is show how $[K] = \sum_{\sigma} \alpha^{w(K^\sigma)}$ can be calculated recursively from an unoriented link diagram without assigning any orientations.

In order to accomplish this feat, first extend the writhe to link diagrams containing a (locally) 4-valent graphical vertex in the oriented forms



Do this by summing over ± 1 contributions from the crossings. Thus

$$w\left(\text{Diagram}\right) = +1.$$

Definition 5.1. $\left[\begin{array}{c} \diagup \\ \diagdown \end{array}\right]_+ = \sum_{\sigma \in \mathcal{O}_+(\diagup \diagdown)} \alpha^{w(K^\sigma)}$

$$\left[\begin{array}{c} \diagup \\ \diagdown \end{array}\right]_- = \sum_{\sigma \in \mathcal{O}_-(\diagup \diagdown)} \alpha^{w(K^\sigma)}$$

where $\mathcal{O}_+(\diagup \diagdown)$ denotes all orientations of the link that give a locally + orientation of the two lines at the site indicated. Thus

$$\begin{aligned} \mathcal{O}_+(\diagup \diagdown) &\leftrightarrow \diagup \cup \diagdown \\ \mathcal{O}_-(\diagup \diagdown) &\leftrightarrow \diagdown \cup \diagup \end{aligned}$$

where this diagram indicates all orientations of K with this local configuration at the crossing under surveillance.

Proposition 5.2. $\left[\begin{array}{c} \diagup \\ \diagdown \end{array}\right]_+ = \frac{\alpha}{2} \left(\left[\begin{array}{c} \diagup \\ \diagdown \end{array}\right] + \left[\begin{array}{c} \diagdown \\ \diagup \end{array}\right] - \left[\begin{array}{c} \diagup \\ \diagup \end{array}\right] \right)$

$$\left[\begin{array}{c} \diagup \\ \diagdown \end{array}\right]_- = \frac{\alpha^{-1}}{2} \left(\left[\begin{array}{c} \diagup \\ \diagdown \end{array}\right] - \left[\begin{array}{c} \diagdown \\ \diagup \end{array}\right] + \left[\begin{array}{c} \diagup \\ \diagup \end{array}\right] \right)$$

Proof. Consider the possibilities.

$$\begin{array}{c}
 \text{X} \iff \begin{array}{c} \nearrow \searrow \\ \downarrow \end{array}, \begin{array}{c} \nearrow \swarrow \\ \downarrow \end{array}, \begin{array}{c} \nearrow \swarrow \\ \uparrow \end{array}, \begin{array}{c} \nearrow \searrow \\ \uparrow \end{array} \\
 \text{---} \iff \begin{array}{c} \rightarrow \leftarrow \\ \uparrow \end{array}, \begin{array}{c} \leftarrow \rightarrow \\ \uparrow \end{array}, \begin{array}{c} \rightarrow \leftarrow \\ \downarrow \end{array}, \begin{array}{c} \leftarrow \rightarrow \\ \downarrow \end{array} \\
) (\iff \begin{array}{c} \uparrow \uparrow \\ \downarrow \downarrow \end{array}, \begin{array}{c} \downarrow \downarrow \\ \uparrow \uparrow \end{array}, \begin{array}{c} \uparrow \uparrow \\ \downarrow \uparrow \end{array}, \begin{array}{c} \downarrow \downarrow \\ \uparrow \downarrow \end{array}
 \end{array}$$

$$\begin{aligned}
 \text{Therefore } & \text{X} + \text{---} -) (\iff 2(\text{---} + \text{---}) \\
 & \Rightarrow [\text{X}] + [\text{---}] - [)()] = 2\alpha^{-1}[\text{X}]_+
 \end{aligned}$$

and similarly

$$[\text{X}] - [\text{---}] + [)()] = 2\alpha^{+1}[\text{X}]_- \quad //$$

Corollary 5.3.

$$(a) [\text{X}] = \left(\frac{\alpha - \alpha^{-1}}{2}\right)[\text{---}] + \left(\frac{\alpha^{-1} - \alpha}{2}\right)[)()] + \left(\frac{\alpha + \alpha^{-1}}{2}\right)[\text{X}].$$

(b) If G is a planar graph with 4-valent vertices then

$$[G] = 2^{|G|}$$

where $|G|$ denotes the number of knot-theoretic circuits in G (i.e. the number of components in the link that would be obtained by creating a crossing at each vertex).

Proof. (a) follows from the previous proposition, coupled with the fact that $[\mathbf{X}] = [\mathbf{X}]_+ + [\mathbf{X}]_-$. (b) follows from the fact that a link with r components has 2^r orientations. //

Remark. Note that $[O] = 2$. The formulas of this corollary give a recursive procedure for calculating $[K]$ for any unoriented K . Hence we have produced the desired mode of computing absolute linking number from an unoriented diagram.

$$[\mathbf{X}] = \left(\frac{\alpha - \alpha^{-1}}{2}\right)[\mathbf{\Xi}] + \left(\frac{\alpha^{-1} - \alpha}{2}\right)[\mathbf{O}](\mathbf{)} + \left(\frac{\alpha + \alpha^{-1}}{2}\right)[\mathbf{X}]$$

$$[\mathbf{X}] = \left(\frac{\alpha^{-1} - \alpha}{2}\right)[\mathbf{\Xi}] + \left(\frac{\alpha - \alpha^{-1}}{2}\right)[\mathbf{O}](\mathbf{)} + \left(\frac{\alpha + \alpha^{-1}}{2}\right)[\mathbf{X}].$$

It follows from these equations that

$$\begin{aligned} [\mathbf{X}] - [\mathbf{X}] &= \left(\frac{\alpha - \alpha^{-1}}{2}\right)([\mathbf{\Xi}] - [\mathbf{O}](\mathbf{)}) \\ [\mathbf{\Xi}] &= \alpha[\mathbf{\sim}], \quad [\mathbf{O}] = \alpha^{-1}[\mathbf{\sim}]. \end{aligned}$$

Thus $[K]$ of this section is a special case of the Dubrovnik polynomial discussed in section 14⁰ of Part I.

We can also view the expansion formulas for $[K]$ in terms of a **state summation** of the form

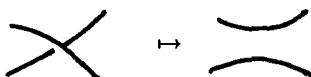
$$[K] = \sum_G [K|G] 2^{|G|}$$

where each **state** is obtained from the diagram for K as follows:

1. Replace each crossing \mathbf{X} of K by one of the three local configurations



labelled as indicated. Note that the label A corresponds to the A -split



2. The resulting labelled state G is a planar graph (with 4-valent vertices).

Let $[K|G]$ denote the product of the labels for G .

Let $|G|$ denote the number of **crossing circuits** in G , where a crossing circuit is obtained by walking along G and **crossing at each crossing**, e.g.

e.g. $\left| \begin{array}{c} \text{Diagram} \\ \text{of a graph} \end{array} \right| = 3.$

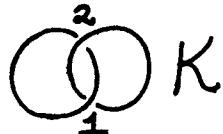
It then follows from our discussion that if we set

$$\left. \begin{aligned} C &= \frac{1}{2}(\alpha + \alpha^{-1}) & \Leftrightarrow & \times \\ B &= \frac{1}{2}(\alpha^{-1} - \alpha) & \Leftrightarrow & \circ \\ A &= \frac{1}{2}(\alpha - \alpha^{-1}) & \Leftrightarrow & \approx \end{aligned} \right\} \times$$

then $[\times] = A[\approx] + B[\circ] + C[\times]$,

$$[K] = \sum_G [K|G] 2^{|G|}.$$

I will leave the details of verification as an exercise. However, here is an example:



	A	B	C
A	$\circ\circ$	$\circ\circ$	$\circ\circ$
B	$\circ\circ$	$\circ\circ$	$\circ\circ$
C	$\circ\circ$	$\circ\circ$	$\circ\circ$

Rows are labelled with the state choice for vertex 1, columns for vertex 2.

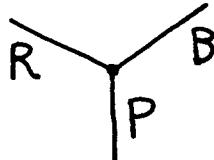
Thus

$$\begin{aligned}
 \sum_G [K|G] 2^{|G|} &= A^2 \cdot 2^2 + AB \cdot 2^1 + AC \cdot 2^1 \\
 &\quad + BA \cdot 2^1 + B^2 \cdot 2^2 + BC \cdot 2^1 \\
 &\quad CA \cdot 2^1 + CB \cdot 2^1 + C^2 \cdot 2^2 \\
 &= 4(A^2 + B^2 + C^2) + 4(AB + AC + BC) \\
 &= (\alpha - \alpha^{-1})^2 + (\alpha^{-1} - \alpha)^2 + (\alpha + \alpha^{-1})^2 \\
 &\quad + (\alpha^{-1} - \alpha)(\alpha - \alpha^{-1}) + (\alpha - \alpha^{-1})(\alpha + \alpha^{-1}) \\
 &\quad + (\alpha^{-1} - \alpha)(\alpha + \alpha^{-1}) \\
 &= (\alpha - \alpha^{-1})^2 + (\alpha + \alpha^{-1})^2 \\
 &= 2(\alpha^2 + \alpha^{-2}) = [K] \text{ as expected!}
 \end{aligned}$$

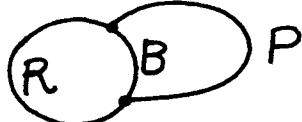
$$\left\{ \begin{array}{l} \text{OK, } [K] = 2(\alpha^{d_1} + \alpha^{d_2}) \\ |\text{rk}(K)| = \frac{1}{4} |d_1 - d_2| \end{array} \right\}$$

6⁰. The Penrose Chromatic Recursion.

A diagram recursion very closely related to the formalism of 5^0 occurs in the number of edge 3-colorings for planar trivalent graphs. We wish to color the edges with three colors - say red (R), blue (B) and purple (P) so that **three distinct colors occur at each vertex**.



For example,



is such a coloring. In [PEN1] Penrose gives the following recursion formula:

$$[\text{X}] = [\text{X}(\text{)}] - [\text{X}(\text{)}]$$

where $[G] = 3^{|G|}$ for a 4-valent planar graph G , and $|G|$ denotes the number of crossing circuits of G . For example,

$$[\text{O}] = [\text{OO}] - [\infty] = 3^2 - 3 = 6.$$

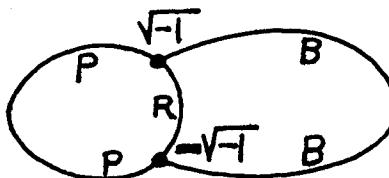
The proof of this recursion formula is actually quite simple. I will give a state summation that models it.

In the plane, we may distinguish two types of oriented vertex.



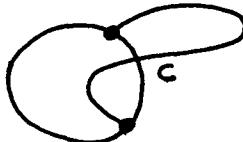
These two types correspond to the cyclic order RBP or BRP (counterclockwise) at the vertex. I shall label a vertex of type RBP by $\sqrt{-1}$ and a vertex of type BRP by $+\sqrt{-1}$.

Thus,



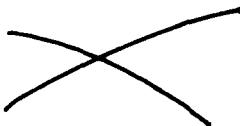
We shall need to distinguish between maps with and without singularities.

Thus



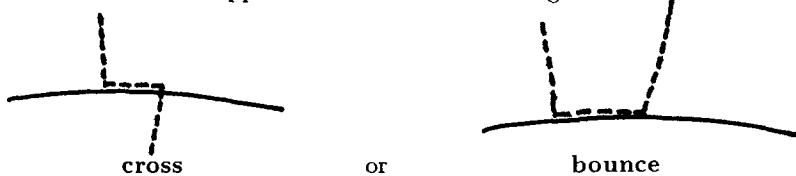
has a singularity (the fourfold (valent) crossing at c).

Definition 6.1. A **cast** is any (possibly singular) embedding of a graph (with trivalent vertices) into the plane. All cast singularities are ordinary double points as in



A **planar cast** or **map** is a cast that is free of singularities.

Now there is a nice way to think about an edge coloring of a cast by the three colors R, B, P. This is called a **formation**. (The terminology is due to G. Spencer-Brown [SB1].) In a formation we have **red** (R) circuits and **blue** (B) circuits. Two circuits of opposite color can share an edge and



Here I let denote **red** (R) and I let denote **blue** (B). I also regard the superpositions of red and blue as purple. Thus



is **purple** (P). For planar casts it is a very easy matter to draw formations. Just draw a disjoint collection of red circles and interlace them with blue circles (crossing and bouncing).

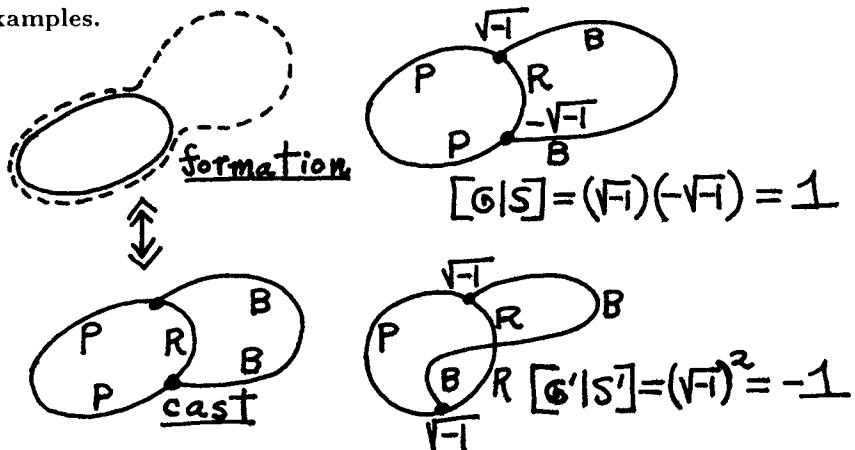
Thus for a **map** (planar cast) a formation corresponds to an edge coloring and it is a certain collection of Jordan curves in the plane. It is easy to see how to formate any cast if you are given a 3-coloring for it. (Check this!)

Definition 6.2. Let G be a cast, and S a 3-coloring of the edges of G (three distinct colors per vertex). Define the **value** of S with respect to G , denoted $[G|S]$, by the formula

$$[G|S] = \prod_{j=1}^n \epsilon(j)$$

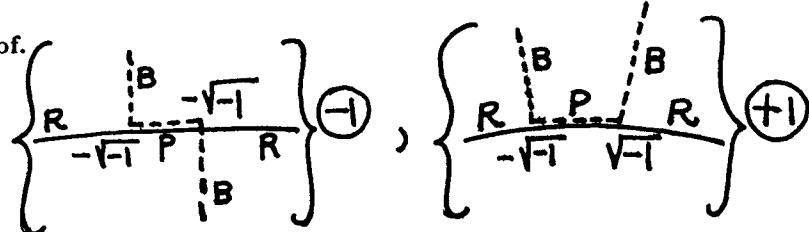
where $\{1, 2, \dots, n\}$ denotes the vertices of G , and $\epsilon(j) = \pm\sqrt{-1}$ is the label assigned to this vertex by the coloring S .

Examples.



Proposition 6.3. Let M be a planar cast, and S a 3-coloring of M . Then $[M|S] = 1$. That is, the product of the imaginary values assigned to the vertices of M by the coloring S is equal to 1.

Proof.



Take the formation \mathcal{F} associated to S . Note that bounces contribute $(\sqrt{-1})(-\sqrt{-1}) = +1$ while crossings contribute $(\pm\sqrt{-1})(\pm\sqrt{-1}) = -1$. Thus $[G|S] = (-1)^c$ where c is the number of crossings between red curves and blue curves in the formation. But c is even by the Jordan curve theorem. //

Definition 6.4. Let $[G] = \sum_S [G|S]$ where G is any cast, and S runs over all 3-colorings of G .

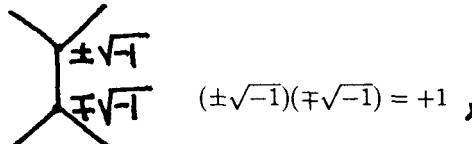
Proposition 6.5. If G is a planar cast, then $[G]$ equals the number of distinct 3-colorings of the edges of G with three colors per vertex.

Proof. By the previous proposition, when G is planar, then $[G|S] = +1$ for each S . This completes the proof. //

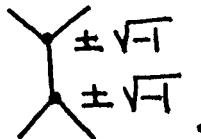
Proposition 6.6. Let G be any cast. Let  denote a specific edge of G and let  and  denote the two casts obtained by replacing this edge as indicated. Then

$$[\text{ } \text{ } \text{ } \text{ }] = [\text{ } \text{ } \text{ } \text{ }] - [\text{ } \text{ } \text{ } \text{ }].$$

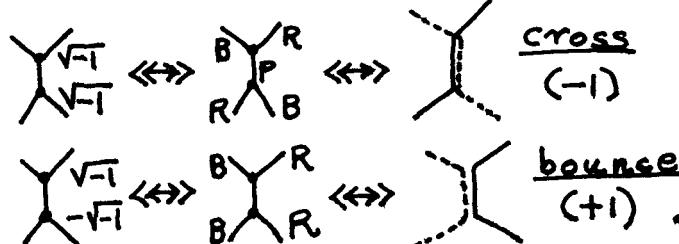
Proof. Let S_+ denote the states (colorings) S contributing to the edge  in the form



and let S_- denote the states S contributing to the edge in the form



Then



Thus the sum over all states becomes the sum over crossing types plus bounces.

$$\therefore [\text{X}] = -[\text{X}] + [()()],$$

since S_+ is in 1-1 correspondence with states of $)()$ and S_- is in 1-1 correspondence with states of X . //

Examples.

$$[\text{O}] = [00] - [\infty] = 3^2 - 3 = 6$$

$$[\text{G}] = [\text{G}] - [\Theta] = 3 - 3^2 = -6.$$

Note that $[G]$ does not necessarily count the number of colorings for a non-planar cast.

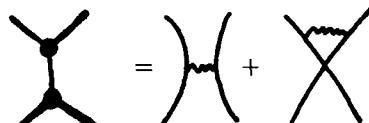
Theorem 6.7. (Kempe). Coloring a planar cast with 3 colors on the edges is equivalent to coloring the associated map with 4 colors so that regions sharing a boundary receive different colors.

Proof. Let W be a fourth color and let $\mathbf{K} = \{R, B, P, W\}$ with group structure given by: $W = 1$, $R^2 = B^2 = P^2 = W$, $RB = P$, $BR = P$.

Choose any region of the edge-colored map and color it W . Now color an adjacent region X so that $WY = X$ where Y is the color of their common boundary. Continue in this fashion until done. //

A Chromatic Expansion.

There is a version of this formalism that yields the number of edge-three-colorings for any cubic graph. (See [JA6] for a different point of view.) To see this, let  denote two lines that receive different colors, and let  also denote lines that receive different colors. Then we can write



in the sense that whenever either of the graphs on the right is colored with distinct

local colors, then we can amalgamate to form a coloration of the graph on the left.

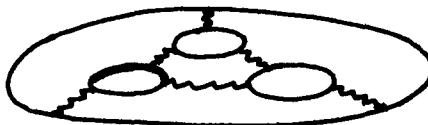
$$\text{Diagram: } \textcircled{1} = \textcircled{0} + \textcircled{\infty} \\ = 3 \cdot 2 + 0 = 6.$$

Since loops can not be colored differently from themselves, many states in this expansion contribute zero. Let us call this state summation the **chromatic expansion** of a cubic graph. Note how the dumbbell behaves:

$$\text{Diagram: } \textcircled{1} = \textcircled{0} + \textcircled{\infty} = 0 + 0 = 0.$$

In fact, it is easy to see that the 4-color theorem is equivalent to the following statement about Jordan curves in the plane: Let J be a disjoint collection of Jordan curves in the plane. Assume that J is decorated with prohibition markers  (locally as shown, and never between a curve and itself). Then either the loops of J can be colored with three colors in the restrictions indicated by the prohibition markers or some subset of markers can be switched ( $\xrightarrow{\text{switch}}$ ) to create a colorable configuration.

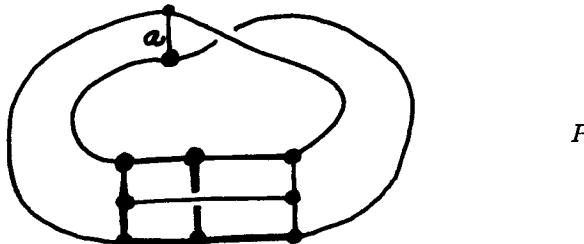
For example, let J be as shown below



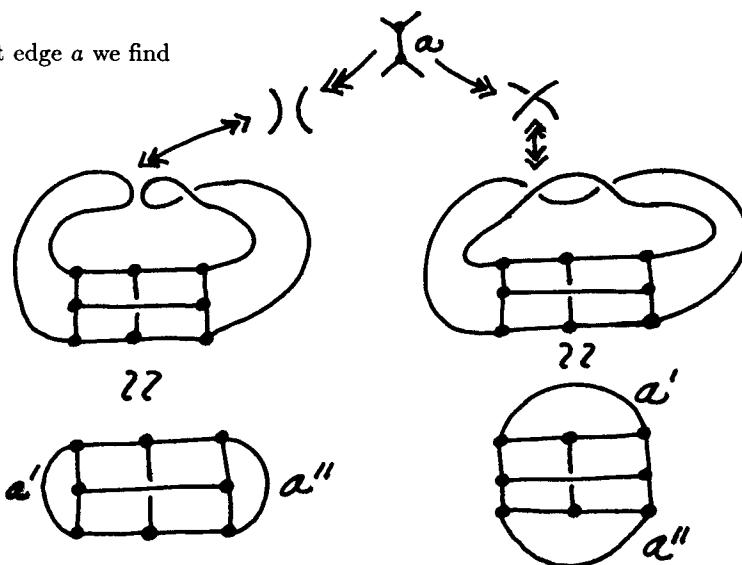
Then J has no coloring, but J' , obtained by switching all the markers, can be colored.

From the chromatic expansion we see that if  represents an edge in any minimal uncolorable graph then both  and  must receive the same colors in any coloring of these smaller graphs. The simplest known (non-planar)

example is the Petersen graph shown below:

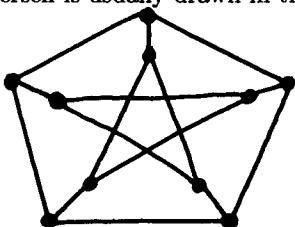


At edge a we find



and in each case a'' and a' are forced to receive the same color.

Remark. The Petersen is usually drawn in the pentagonal form [SA]:



7⁰. The Chromatic Polynomial.

The chromatic polynomial is a polynomial $K(G) \in \mathbf{Z}[x]$ associated to any graph G such that $K(G)(x)$ is the number of vertex colorings of G with x colors. (In a vertex coloring two vertices are colored differently whenever they are connected by an edge.)

Examples.

$$K\left(\begin{array}{c} \bullet \\ \end{array}\right) = x$$

$$K\left(\begin{array}{c} \bullet - \bullet \\ \end{array}\right) = x(x-1) = x^2 - x$$

$$K\left(\begin{array}{c} \bullet \\ \bullet \bullet \\ \end{array}\right) = 0$$

$$K\left(\begin{array}{c} \bullet \\ \bullet - \bullet \\ \end{array}\right) = x(x-1)(x-2) = x^3 - 3x^2 + 2x.$$

Note that, by convention, we take $K(G) = 0$ if G has any self-loops.

To begin this tale, we introduce the well-known [WH2] recursive formula for $K(G)$:

$$\begin{aligned} K\left(\begin{array}{c} \rightarrow - \leftarrow \\ \end{array}\right) &= K\left(\begin{array}{c} \rightarrow \\ \end{array}\right) - K\left(\begin{array}{c} \leftarrow \\ \end{array}\right) \\ K(G) &= K(G - a) - K(G/a). \end{aligned}$$

This formula states that the number of colorings of a graph G with a specified edge a is the difference of the number of colorings of $(G - a)$ (G with a deleted) and (G/a) (G with a collapsed to a point). This is the logical identity:

$$\text{Different} = \text{All} - \text{Same}.$$

The formula may be used to recursively compute the chromatic polynomial.

For example:

$$\begin{aligned}
 K\left(\text{graph}\right) &= K\left(\text{graph}\right) - K\left(\text{graph}\right) \\
 &= K\left(\text{graph}\right) - K\left(\text{graph}\right) - \left(K\left(\text{graph}\right) - K\left(\text{graph}\right)\right) \\
 &= \left\{ \begin{array}{ll} K\left(\text{graph}\right) & -K\left(\text{graph}\right) \\ -K\left(\text{graph}\right) & +K\left(\text{graph}\right) \\ -K\left(\text{graph}\right) & +K\left(\text{graph}\right) \\ -K\left(\text{graph}\right) & +K\left(\text{graph}\right) \end{array} \right\} \\
 &= \left\{ \begin{array}{ll} x^3 & -x^2 \\ -x^2 & +x \\ -x^2 & +x \\ -x & +x \end{array} \right\} = x^3 - 3x^2 + 2x \\
 \therefore K\left(\text{graph}\right) &= x(x-1)(x-2).
 \end{aligned}$$

In this computation we have used the **edge-bond**  to indicate a collapsed edge. Thus

$$K\left(\text{graph}\right) = K\left(\text{graph}\right) - K\left(\text{graph}\right)$$

As the middle section of the computation indicates, the recursion results in a sum of values of K on the **bond-graphs** obtained from G by making a choice for every edge to either delete it or to bond it. Clearly, the contribution of such a graph is $(-1)^{\#\text{(bonds)}} x^{\#\text{(components)}}$.

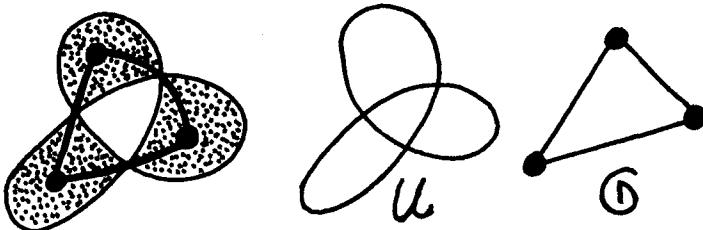
Thus, for any graph G we may define a **bond state** B of G to be the result of choosing, for each edge to delete it or to bond it. Let $\|B\|$ denote the number of components of B and $i(B)$ the number of bonds.

Thus B  is a bond state of  , and $\|B\| = 2$, $i(B) = 2$. We have shown that the chromatic polynomial is given by the formula

$$K(G) = \sum_B (-1)^{i(B)} x^{\|B\|}.$$

The first topic we shall discuss is a reformulation of this chromatic formula in terms of states of a universe (planar 4-valent graph) associated with the graph G . In fact, there is a one-to-one correspondence between universes and planar graphs. The correspondence is obtained as follows (see Figure 7.1).

- (i) Given a universe U , we obtain an associated graph G by first shading U in checkerboard fashion so that the unbounded region is colored white. (Call the colors white and black.) For each black region (henceforth called shaded) place a vertex in its interior. Join two such vertices by an edge whenever they touch at a crossing of U . Thus

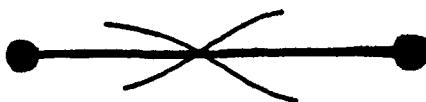


A graph G and its universe U .

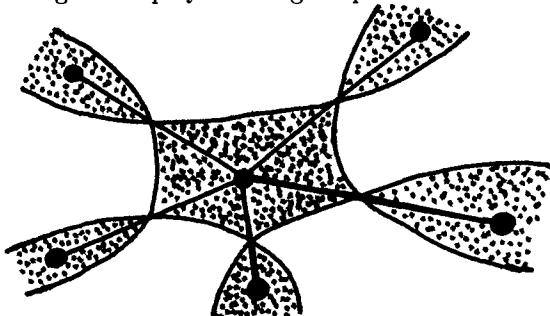
Figure 7.1

Compare this description with Figure 7.1.

- (ii) Conversely, given a graph, we obtain a universe by first placing a crossing on each edge in the form

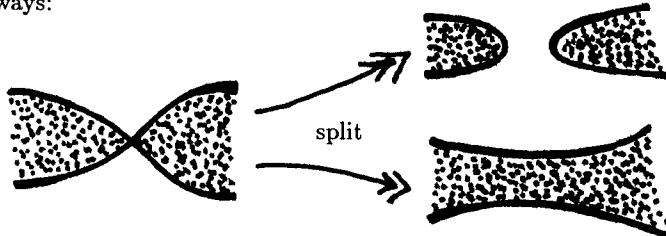


and then connecting these up by following the pattern indicated below for a single vertex:



Exercise. Formulate this correspondence precisely, taking care of examples such as

(In the literature, the universe associated with a planar graph is sometimes referred to as its **medial graph**.) By replacing the graph G by its associated universe U , we obtain a reformulation of the chromatic polynomial. Instead of deleting and bonding edges of the graph, we split the crossings of U in the two possible ways:



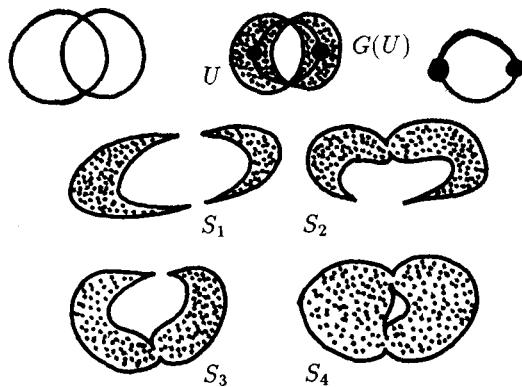
Since coloring the vertices of the graph G corresponds to coloring the shaded regions of U so that **any two regions meeting at a crossing receive different colors**, we see that the chromatic identity becomes:

$$K\left(\text{graph with crossing}\right) = K\left(\text{two separate regions}\right) - K\left(\text{region with bridge}\right)$$

By splitting the crossings so that the crossing structure is still visible we can classify **exterior and interior vertices of a chromatic state**. A chromatic state of a shaded universe U is a choice of splitting at each vertex (crossing). A split vertex is **interior** if it locally creates a connected shading, and it is exterior if there is a local disconnection. See Figure 7.2.



Figure 7.2

Chromatic States of U

$$\|S_1\| = 2, \|S_2\| = 1, \|S_3\| = 1, \|S_4\| = 1$$

$$i(S_1) = 0, i(S_2) = 1, i(S_3) = 1, i(S_4) = 2$$

$$\begin{aligned} K(U) &= \sum_S (-1)^{i(S)} x^{\|S\|} \\ &= (-1)^0 x^2 + (-1)^1 x + (-1)^1 (x) + (-1)^2 x \end{aligned}$$

$$\therefore K(U) = x^2 - x$$

Figure 7.3

In Figure 7.3 we have listed all the chromatic states for a small universe U . Since interior vertices correspond to bonded edges in the associated graph we see that the coloring formula for the universe is:

$$K(U) = \sum_S (-1)^{i(S)} x^{\parallel S \parallel} \quad (*)$$

where $i(S)$ is the number of interior vertices in the chromatic state S , and $\parallel S \parallel$ is the number of shaded components in S .

The subject of the rest of this section is **an easily programmable algorithm for computing the chromatic polynomial for planar graphs**. (Of course the graphs must be relatively small.)

Later we shall generalize this to the dichromatic polynomial and relate it to the Potts model in statistical physics. There the algorithm is still useful and can be used to investigate the Potts model. On the mathematical side, this formulation yields a transparent explanation of the interrelationship of the Potts model, certain algebras, and analogous models for invariants of knots and links. There is a very rich structure here, all turning on the coloring problem.

Turning now to the algorithm, the first thing we need is **a method to list the chromatic states**. Since each state is the result of making a binary choice (of splitting) at each vertex of U , it will be sufficient to use any convenient method for listing the binary numbers from zero (0) to $2^n - 1$ (n = the number of vertices in U). In particular, it is convenient to use the **Gray code**. The Gray code lists binary numbers, changing only one digit from step to step. Figure 7.4 shows the first terms in the Gray code.

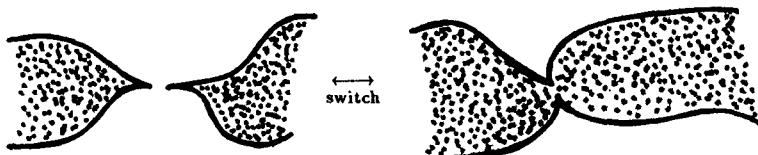
			0	
			1	
		1	1	
		1	0	
	1	1	0	
	1	1	1	
	1	0	1	
	1	0	0	
1	1	0	0	
1	1	0	1	
1	1	1	1	
1	1	1	0	
1	0	1	0	
1	0	1	1	
1	0	0	1	
1	0	0	0	
1	1	0	0	
:				

Gray Code

Figure 7.4

The code alternates between changing the right-most digit and changing the digit just to the left of the **right-most string** of the forms $1\ 0\ 0 \dots 0$.

Since we are changing one site (crossing) at a time, the signs in this expansion will alternate $+ - + - + \dots$ (These are the signs $(-1)^{\iota(S)}$ in the formula (*).) The crucial item of information for each state is $\|S\|$, the number of shaded components. Thus we need to determine how $\|S\|$ and $\|S'\|$ differ when S' is obtained from S by switching the split at one vertex.



We have

Lemma 7.1. Let S and S' be chromatic states of a universe U . Suppose that S' is obtained from S by switching the split at one vertex v .

Then

- 1) $\|S'\| = \|S\| - 1$ when v is an exterior vertex incident to disjoint cycles of S .
- 2) $\|S'\| = \|S\| + 1$ when v is an interior vertex incident to a single cycle of S .
- 3) $\|S'\| = \|S\|$ in any other case.

(Refer to Figure 7.5.)

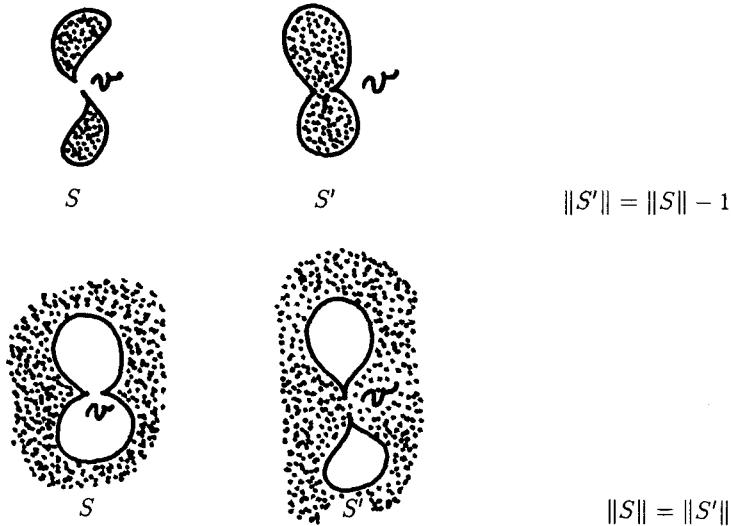


Figure 7.5

It should be clear from Figure 7.5 what is meant by incidence (to) cycles of S . Any state S is a disjoint union of closed curves in the plane. These are the cycles of S . Any vertex is met by either one or two cycles.

We see from this lemma that the construction of an algorithm requires that we know how many cycles there are in a given state. In fact, it is a simple matter

to code a state by a list of its cycles (plus a list of vertex types). Just list in clockwise order the vertices of each cycle. For example, (view Figure 7.6),

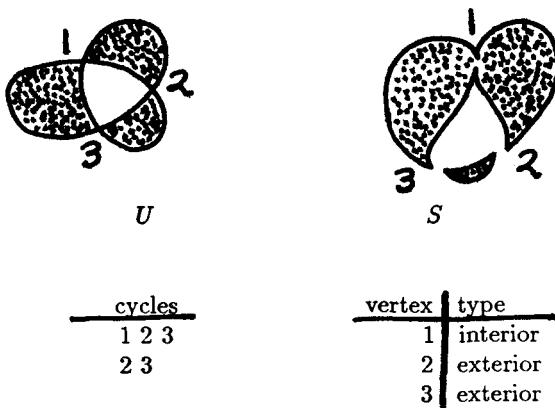


Figure 7.6

In Figure 7.6 we have illustrated a single trefoil state S , the list of cycles and the list of vertex types. From cycle and vertex information alone, we can deduce that switching vertex 3 reduces the number of components by 1 (It is an exterior vertex incident to two cycles.).

Thus we have given the outline of a computer-algorithm for computing the chromatic polynomial for a planar graph. The intermediate stages of this algorithm turn out to be as interesting as its end result! In order to discuss this I give below an outline form of the algorithm adapted to find the chromatic number of the graph:

1. Input the number of colors (NUM) (presumably four).
2. Input state information for one state S : number of vertices and their types, list of cycles, number of shaded regions.
3. Initialize the SUM to 0. (SUM will be the partial value of the state summation.)

4. $\text{SUM} = \text{SUM} + (-1)^{i(S)}(\text{NUM}^{\|S\|})$. (Here $i(S)$ denotes the number of interior vertices in S , $\|S\|$ is the number of shaded regions.)
5. Print $\log(\text{SUM})$ on a graph of \log (iterations of this algorithm) versus $\log(\text{SUM})$.
6. Use the **gray code** to choose the vertex V to be switched next. (If all states have been produced, stop!)
7. Use the choice V of step 6 and switch the state S to a new state by re-splicing the cycles and changing the designation of the vertex v . (interior \rightleftarrows exterior). Let the new state also be called S .
8. Go to step 4.

This algorithm depends upon the choice of initial state only in its intermediate behavior. I find that the easiest starting state is that state S_0 whose cycles are the boundaries of the shaded regions of the universe U . Then all vertices are external and $\|S_0\| = (\text{NUM}^M)$ where M equals the number of shaded regions. This is the maximal value for S_0 .

By analyzing the gray code, one can see that under this beginning the intermediate values of SUM are always **non-negative**. That they are always **positive** is a conjecture generalizing the four color theorem (for $\text{NUM} = 4$).

In fact, we find empirically that the graph of the points of the $\log - \log$ plot for 2^K iterations $K = 1, 2, \dots$ is approximately linear (these values correspond to chromatic numbers of a set of subgraphs of the dual graph $\Gamma(U)$)! Clearly much remains to be investigated in this field.

Figure 7.7 illustrates a typical plot produced by this algorithm. (See [LK10].)

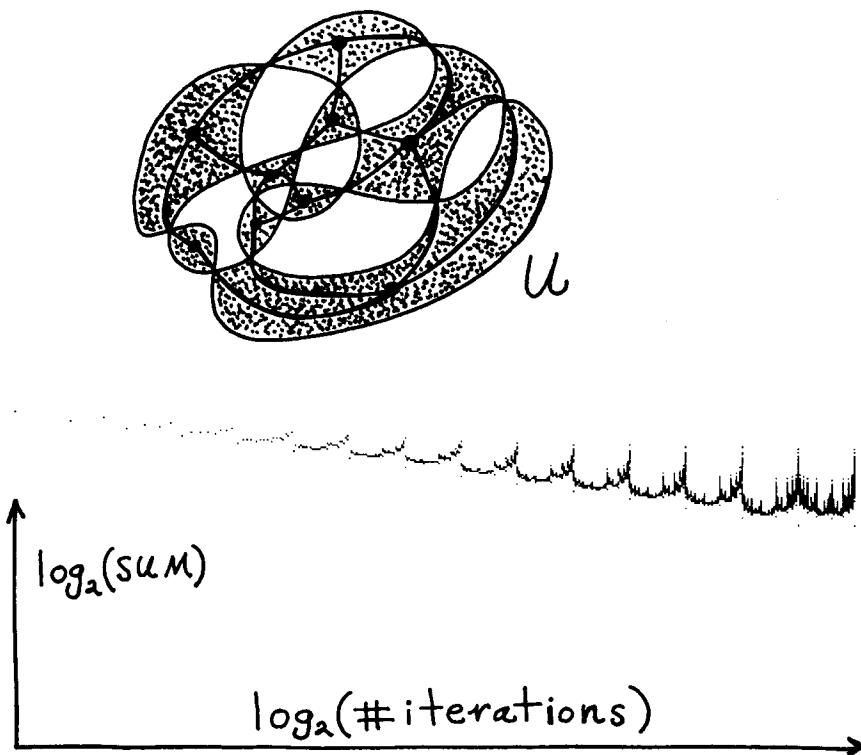


Figure 7.7

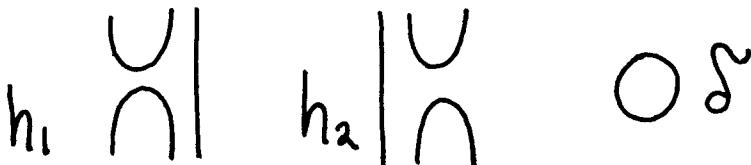
The minima in this plot correspond to chromatic numbers of subgraphs (i.e. they occur at 2^K iterations). The peaks correspond to those states where the gray code has disconnected most of the regions (as it does periodically in the switching process, just before switching a new bond.).

It is amusing and very instructive to use the algorithm, and a least squares approximation to the minima, as a “chromatic number predictor.”

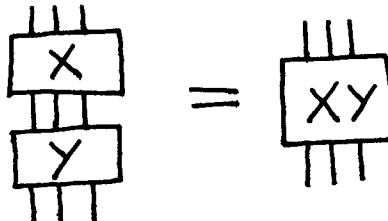
8⁰. The Potts Model and the Dichromatic Polynomial.

In this section we show how the partition function for the Potts model (See [BA2].) can be translated into a bracket polynomial state model. In particular we show how an operator algebra (first constructed by Temperley and Lieb from the physics of the Potts model) arises quite naturally in computing the dichromatic polynomial for universes (four-valent planar graphs).

Before beginning a review of the Potts model, here is a review of the operator algebra:



The diagrams h_1 and h_2 represent the two generators of the 3-strand diagram algebra D_3 . The diagrams are multiplied by attachment:



Elsewhere, we have used e_i for h_i and called this algebra (directly) the Temperley-Lieb algebra. Strands may be moved topologically, but they are not allowed to move above or below horizontal lines through the **fixed** end-points of the strands. Multiplication by δ is accomplished by disjoint union.

In D_n the following relations hold:

$$\left\{ \begin{array}{ll} h_i^2 = \delta h_i & i = 1, 2, \dots, n-1 \\ h_i h_{i+1} h_i = h_i & \\ h_{i+1} h_i h_{i+1} = h_{i+1} & \\ h_i h_j = h_j h_i, & |i-j| > 1 \end{array} \right\}$$

See Figure 8.1 for an illustration.

Diagram Algebra Relations

Figure 8.1

The Potts Model.

In this model we are concerned with calculating a partition function associated with a planar graph G . In general for a graph G with vertices i, j, \dots and edges $\langle i, j \rangle$ a partition function has the form $Z_G = \sum_S e^{-E(S)/kT}$ where S runs over states of G , $E(S)$ is the energy of the state S , k is Boltzmann's constant and T is the temperature.

This presumes that there is a physical system associated to the graph and a collection of combinatorially described states S , each having a well-defined energy. The partition function, if sufficiently well-known, contains much information about the physical system. For example, the probability for the system to be in the state S is given by the formula $p(S) = e^{-E(S)/kT}/Z$.

This exponential model for the probability distribution of energy states is based on a homogeneity assumption about the energy: If $\hat{p}(E)$ denotes the probability of occurrence of a state of energy E then $\hat{p}(E+K)/\hat{p}(E'+K) = \hat{p}(E)/\hat{p}(E')$

for a pair of energy levels E and E' . Furthermore, states with the same energy level are equally probable.

The Boltzmann distribution follows from these assumptions. The second part of the assumption is an assumption that the system is in equilibrium. The homogeneity really has to do with the observer's inability to fix a "base-point" for the energy.

In the Potts model, the states involve choices of q values ($1, 2, 3, \dots, q$ say) for each vertex of the graph G . Thus if G has N vertices, then there are q^N states. The q values could be spins of particles at sites in a lattice, types of metals in an alloy, and so on. The choice of spins (we'll call $1, \dots, q$ the spins) at the vertices is free, and the energy of a state S with spin S_i at vertex i is taken to be given by

$$E(S) = \sum_{\langle i,j \rangle} \delta(S_i, S_j).$$

Here we sum over all edges in the graph. And $\delta(x, y)$ is the Kronecker delta:

$$\delta = (x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}.$$

Thus, the partition function for the Potts model is given by

$$\begin{aligned} Z_G &= \sum_S e^{K \sum_{\langle i,j \rangle} \delta(S_i, S_j)} \quad (K = -1/kT) \\ &= \sum_S \prod_{\langle i,j \rangle} e^{K \delta(S_i, S_j)} \\ \therefore Z_G &= \sum_S \prod_{\langle i,j \rangle} (1 + v \delta(S_i, S_j)) \\ &\quad \text{where } v = e^K - 1 \\ &\quad (\text{i.e. } e^{K \delta(x,y)} = 1 + v \delta(x, y)). \end{aligned}$$

This last form of the partition function is particularly convenient because it implies the structure:

$$\begin{aligned} Z(\text{---}) &= Z(\text{---}) + vZ(\text{---}) \\ Z(\bullet \sqcup H) &= qZ(H). \end{aligned}$$

Here  denotes the graph G () after deletion of one edge, while  means the graph obtained by collapsing this edge.

Hence, $Z(G)$ is a **dichromatic polynomial** of G in variables v, q .

Remark. In some versions of the Potts model the value of v is taken to vary according to certain edge types. Here we have chosen the simplest version.

Remark. One reason to calculate $Z(G)$ is to determine critical behavior corresponding to phenomena such as phase transitions (liquids to gas, ice to liquid, ...). Here one needs the behavior for large graphs. The only thoroughly analyzed cases are for low q (e.g. $q = 2$ in Potts is the Ising model) and planar graphs.

Remark. For $v = -1$, $Z(G)$ is the chromatic polynomial (variable q). Physically, $v = -1$ means $-1 = e^{-1/kT} - 1 \Rightarrow e^{-1/kT} = 0$. Hence $T = 0$.

Dichromatic Polynomials for Planar Graphs.

Just as we did for the chromatic polynomial, we can replace a planar graph with its associated universe (the medial graph) and then the dichromatic polynomial can be computed recursively on the shaded universe via the equations:

- (i) $Z\left(\text{graph with two regions}\right) = Z\left(\text{graph with one region}\right) + vZ\left(\text{graph with two regions}\right)$
- (ii) $Z\left(\text{graph with one region} \sqcup \text{graph with one region}\right) = qZ\left(\text{graph with one region}\right)$. Here  stands for any connected shaded region, and \sqcup denotes disjoint union.

Defining, as before, a **chromatic state** S to be any splitting of the universe U so that every vertex has been split, then **with the shading of U** we have the numbers

$$\|S\| = \text{number of shaded regions of } S,$$

and

$$i(S) = \text{number of interior vertices of } S.$$

Then, as a state summation,

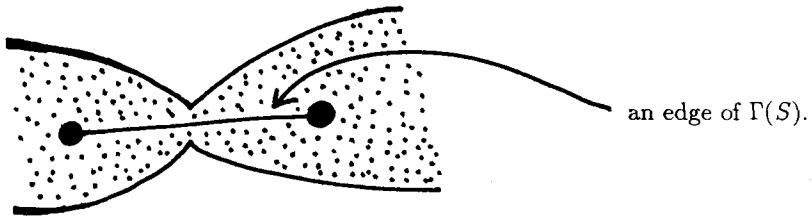
$$Z(U) = \sum_S q^{\|S\|} v^{i(S)}.$$

We will now translate this version of the dichromatic polynomial so that the shaded universe is replaced by an alternating link diagram and the count $\|S\|$ of shaded regions is replaced by a component (circuit) count for S .

Proposition 8.1. Let U be a universe (the shadow of a knot or link diagram). Let U be shaded in checkerboard fashion. Let N denote the number of shaded regions in U (These are in one-to-one correspondence with the vertices of graph $\Gamma(U)$). Hence we say N is the number of dual vertices.). Let S be a chromatic state of U with $\|S\|$ connected shaded regions, and $i(S)$ internal vertices. Let $|S|$ denote the number of circuits in S (i.e. the number of boundary cycles for the shaded regions in S). Then

$$\|S\| = \frac{1}{2}(N - i(S) + |S|).$$

Proof. Associate to each state S a graph $\Gamma(S)$: The vertices of $\Gamma(S)$ are the vertices of $\Gamma(U)$, one for each shaded region of U . Two vertices of $\Gamma(S)$ are connected by an edge exactly when this edge can be drawn in the gap of an interior vertex of S . Thus



Thus the edges of $\Gamma(S)$ are in one-to-one correspondence with the interior vertices of S . $\Gamma(S)$ is a planar graph, and we assert that the number of faces of $\Gamma(S)$ is $|S| - \|S\| + 1$.

To see this last assertion, note that each cycle in $\Gamma(S)$ surrounds a white island interior to one of the black components of S . Thus we may count faces of $\Gamma(S)$ by counting circuits of S , but those circuits forming outer boundaries to shaded parts of S must be discarded. This accounts for the subtraction of $\|S\|$. We add 1 to count the unbounded face.

Clearly, $\Gamma(S)$ has $\|S\|$ components. Therefore we apply the Euler formula for

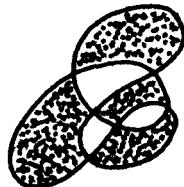
planar graphs:

$$(\text{Vertices}) - (\text{Edges}) + (\text{Faces}) = (\text{Components}) + 1$$

$$N - i(S) + (|S| - \|S\| + 1) = \|S\| + 1.$$

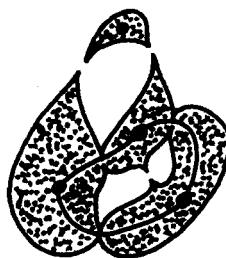
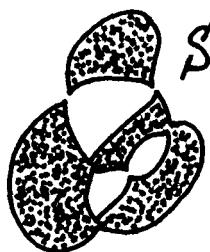
Hence $N - i(S) + |S| = 2\|S\|$. This completes the proof. //

Example.



$$N = 4$$

U



$$\|S\| = 2$$

$$|S| = 3$$

$$i(S) = 3$$



$$\Gamma(S)$$

Thus

$$\frac{N - i(S) + |S|}{2} = \frac{4 - 3 + 3}{2} = 2 ,$$

$$\|S\| = 2.$$

Note that in the dichromatic polynomial, S contributes the term $q^{\|S\|} v^{i(S)} = q^2 v^3$ here. Since the number of v -factors corresponds to the positive energy contributions in the Potts model, we see that these contributions correspond to the clumps of vertices in $\Gamma(S)$ that are connected by edges. These vertices all mutually agree on choice of spin state.

Now the upshot of Proposition 8.1 is that we can replace $\|S\|$ by $|S|$ in our formulas! Let's do so!

$$\begin{aligned} Z(U) &= \sum_S q^{\|S\|} v^{i(S)} \\ &= \sum_S q^{1/2(N-i(S)+|S|)} v^{i(S)} \\ \therefore Z(U) &= q^{N/2} \sum_S (q^{1/2})^{|S|} (q^{-1/2} v)^{i(S)} \end{aligned}$$

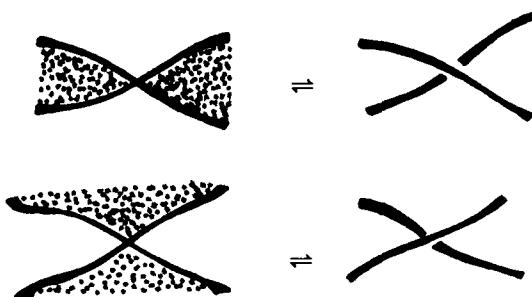
Let

$$W(U) = \sum_S (q^{1/2})^{|S|} (q^{-1/2} v)^{i(S)}.$$

Thus

$$Z(U) = q^{N/2} W(U).$$

We are now ready to complete the reformulation of the dichromatic polynomial. The shading of the universe will be translated into crossings of a link diagram as indicated below (Figure 8.2):



Crossing Transcriptions of Local Shading

Figure 8.2

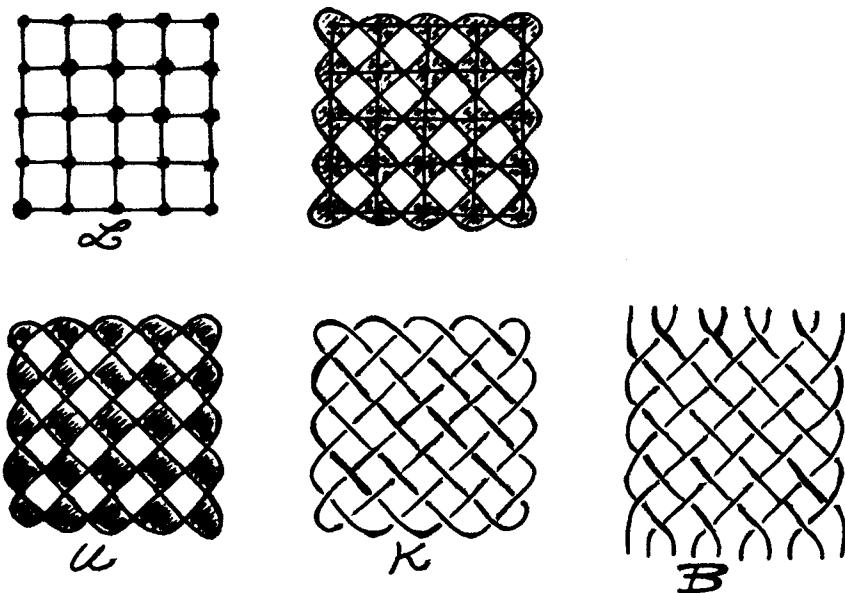


Figure 8.3

Since W satisfies

$$W\left(\text{Diagram L}\right) = W\left(\text{Diagram M} \cup \text{Diagram N}\right) + (q^{-1/2}v)W\left(\text{Diagram O}\right)$$

$$W\left(\text{Diagram O} \cup \text{Diagram K}\right) = q^{1/2}W(K).$$

(Note the extra component is not shaded!) We can write the

W -Axioms (The Potts Bracket).

1. Let K be any knot or link diagram. Then $W(K) \in \mathbf{Z}[q^{1/2}, q^{-1/2}, v]$ is a well-defined function of q and v .

$$2. \quad W(\text{ } \bowtie \text{ }) = W(\text{ } \supset \text{ } \subset \text{ }) + (q^{-1/2}v)W(\text{ } \times \text{ } \times \text{ })$$

$$W(\text{ } \bowtie \text{ }) = (q^{-1/2}v)W(\text{ } \supset \text{ } \subset \text{ }) + W(\text{ } \times \text{ } \times \text{ })$$

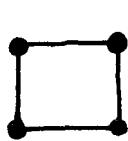
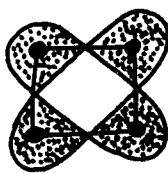
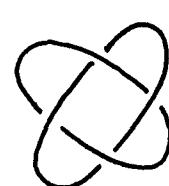
$$3. \quad W(\text{ } \bullet \text{ } \sqcup \text{ } \mathcal{K} \text{ }) = q^{1/2}W(K)$$

$$4. \quad W(\text{ } \bullet \text{ }) = q^{1/2}.$$

and the Z -Definition: $Z(K) = q^{N/2}W(K)$ where N is the number of shaded regions in K . (Compare with [LK10].)

According to our discussion, the dichromatic polynomial (hence the Potts partition function) for a lattice \mathcal{L} can be computed by forming a link diagram $K(\mathcal{L})$ obtained by first forming the medial graph of \mathcal{L} , regarding it as a shaded universe U , and configuring the knot diagram $K(\mathcal{L})$ according to the rule of Figure 8.2. In Figure 8.3 we have indicated these constructions for a 5×5 lattice \mathcal{L} .

Here is a worked example for the 2×2 lattice:

 \mathcal{L}  U  $K = K(\mathcal{L})$

(a) Direct computation of $Z(\mathcal{L}) = Z_1$.

$$\begin{aligned}
 Z_1\left(\begin{array}{c} \text{square} \end{array}\right) &= Z_1\left(\begin{array}{c} \text{L-shaped} \end{array}\right) + vZ_1\left(\begin{array}{c} \Delta \end{array}\right) \\
 &= Z_1\left(\begin{array}{c} \bullet \\ \text{L-shaped} \end{array}\right) + vZ_1\left(\begin{array}{c} \text{L-shaped} \end{array}\right) \\
 &\quad + vZ_1\left(\begin{array}{c} \curvearrowleft \\ \text{L-shaped} \end{array}\right) + v^2Z_1\left(\begin{array}{c} \bullet \\ \text{circle} \end{array}\right) \\
 &= (q+v)Z_1\left(\begin{array}{c} \text{L-shaped} \end{array}\right) + vZ_1\left(\begin{array}{c} \curvearrowleft \\ \text{L-shaped} \end{array}\right) + v^2Z_1\left(\begin{array}{c} \text{circle} \end{array}\right) \\
 &= (q+v)(q+v)^2q + v(q+v)^2q + v^2(q+v)q \\
 &= (q+v)^3q + (q+v)^2vq + (q+v)v^2q \\
 &= (q^3 + 3q^2v + 3qv^2 + v^3)q \\
 &\quad + (q^2 + 2qv + v^2)vq + q^2v^2 + v^3q \\
 &= q^4 + 3q^3v + 3q^2v^2 + qv^3 \\
 &\quad + q^3v + 2q^2v^2 + qv^3 + q^2v^2 + qv^3 \\
 \therefore Z\left(\begin{array}{c} \text{square} \end{array}\right) &= q^4 + 4q^3v + 6q^2v^2 + 3qv^3
 \end{aligned}$$

(b) Computing $W(K)$ first.

In order to compute W , we use the rules

$$\begin{aligned}
 W\left(\begin{array}{c} \text{X} \end{array}\right) &= W\left(\begin{array}{c} \text{D} \text{ C} \end{array}\right) + xW\left(\begin{array}{c} \text{Y} \end{array}\right) \\
 W\left(\begin{array}{c} \text{O} \sqcup \text{L} \end{array}\right) &= yW(L), \quad W(\bullet) = y
 \end{aligned}$$

with $x = q^{-1/2}v$, $y = q^{1/2}$.

Before, doing the 2×2 lattice, let's get some useful formulas:

$$\left\{ \begin{array}{l} W(\text{---}) = W(\text{---}) + xW(\text{---}) \\ \quad = W(\text{---}) + xyW(\text{---}) \\ \therefore W(\text{---}) = (1+v)W(\text{---}) = (1+xy)W(\text{---}). \\ \\ W(\text{---}) = W(\text{---}) + xW(\text{---}) \\ \quad = yW(\text{---}) + xW(\text{---}) \\ \therefore W(\text{---}) = (q^{1/2} + q^{-1/2}v)W(\text{---}) = (x+y)W(\text{---}). \end{array} \right.$$

Returning to our lattice/knot diagram K , we have

$$\begin{aligned}
W(\text{X}) &= W(\text{Y}) + xW(\text{Z}) \\
&= (x+y)^3 W(\text{O}) + x[W(\text{E}) + xW(\text{D})] \\
&= (x+y)^3 y + x(x+y)^2 y + x^2 W(\text{C}) \\
&= (x+y)^3 y + x(x+y)^2 y + \\
&\quad + x^2 \left(W(\text{B}) + xW(\text{C}) \right) \\
&= (x+y)^3 y + x(x+y)^2 y + x^2((x+y)y + x(x+y)y) \\
\therefore W(\text{X}) &= (x+y)^3 y + (x+y)^2 xy + (x+y)(x^2 y + xy).
\end{aligned}$$

We leave as an exercise for the reader to show that $Z(\mathcal{L}) = q^{N/2}W(K(\mathcal{L}))$ for this case!

Comment. Obviously, this translation of the dichromatic polynomial into computations using knot diagrams does not directly make calculations easier. It does however reveal the geometry of the Temperley-Lieb algebra in relation to the dichromatic polynomial (via the Potts model).

I will show how the W -formulation of the dichromatic polynomial naturally gives rise to a mapping of braids into the Temperley-Lieb algebra.

The idea is as follows: If we expanded W on a braid diagram out into the corresponding states, we would find ourselves looking at an element of the Temperley-Lieb algebra. For example, if we were calculating $W(\beta)$ for

$$\beta = \text{Diagram}$$

then one possible local state configuration is

$$\text{Diagram}$$

and this is $h_1 h_2 h_1$ in D_3 . In general, if we expand (for example)

$$W = \left(\begin{array}{c} \beta \\ \end{array} \right) = W(\sigma_1 \sigma_2 \sigma_1)$$

we can obtain the expansion algebraically by replacing

$$\begin{array}{ll} \sigma_i & \text{by } I + xh_i \\ \sigma_i^{-1} & \text{by } xI + h_i \end{array}$$

and formally multiplying these replacements.

For example, in the above case we have

$$\begin{aligned} & (I + xh_1)(I + xh_2)(I + xh_1) \\ &= (I + xh_1 + xh_2 + x^2h_1h_2)(I + xh_1) \\ &= I + xh_1 + xh_2 + x^2h_1h_2 \\ &\quad + xh_1 + x^2h_1^2 + xh_2h_1 + x^3h_1h_2h_1 \\ &= I + 2(xh_1) + (xh_2) + (x^2h^2 + 1) + (x^2h_1h_2) \\ &\quad + (xh_2h_1) + (x^3h_1h_2h_1). \end{aligned}$$

	$\leftrightarrow h_1$		$\leftrightarrow h_2$		$\leftrightarrow h_1^2$
	$\leftrightarrow h_1h_2$		$\leftrightarrow h_2h_1$		$\leftrightarrow h_1h_2h_1$

Now it is **not** the case that W is a topological invariant of braids (when we put a braid in W we mean it to indicate a larger diagram containing this pattern), however this correspondence does respect the relations in the Temperley-Lieb algebra.

Thus

$$W(h_1 h_2 h_1) = W\left[\begin{array}{c} \text{U} \\ \text{G} \\ \text{N} \end{array}\right] = W\left[\begin{array}{c} \text{U} \\ \text{N} \end{array}\right] = W(h_1)$$

and

$$W(h_1^2) = W\left[\begin{array}{c} \text{U} \\ \text{G} \\ \text{N} \end{array}\right] = yW\left[\begin{array}{c} \text{U} \\ \text{N} \end{array}\right] = yW(h_1)$$

($h_1^2 = \delta h_1$, let $W(\delta X) = yW(X)$ be the algebraic version of $W(\text{O} \sqcup \text{X}) = yW(X)$. Therefore let $\delta = y$ here.) What is the image algebra here? Formally, it is $\mathcal{U}_n = \mathbb{Z}[x, y](D_n)$. That is, we are formally adding and multiplying elements of D_n with coefficients from $\mathbb{Z}[x, y]$ - the polynomials in x and y .

In creating this formulation we have actually reformulated the calculation of the rectangular lattice Potts model entirely into an algebra problem about \mathcal{U}_n . To see this, think of the link diagram $K(\mathcal{L})$ of the lattice \mathcal{L} as a braid that has been closed to form $K(\mathcal{L})$ by adding maxima and minima:



$$P(\beta) = K(\mathcal{L})$$

This way of closing a braid β to form $P(\beta)$ is called a **plat**.

Now note. Let R denote the diagram

$$\begin{array}{ccccc} \cup & \cup & \dots & \cup & \equiv h_1 h_3 h_5 \dots h_{2n-1} \in D_{2n} \\ \cap & \cap & \dots & \cap & \end{array}$$

Then, if β is any element of B_{2n} we have $R\beta R = P(\beta)R$.

Proof.

$$R\beta R = P(\beta)R. \quad //$$

The upshot of this formula is that all of the needed component counting will happen **automatically** if we substitute for β the product expression in \mathcal{U}_n . Let $X(\beta)$ denote the corresponding element of \mathcal{U}_n :

$$\begin{aligned} X : \widehat{\mathcal{B}}_n &\rightarrow \mathcal{U}_n \\ \left\{ \begin{aligned} X(\sigma_i) &= \mathcal{I} + x h_i \\ X(\sigma_i^{-1}) &= x \mathcal{I} + h_i \\ X(\mathcal{I}) &= \mathcal{I} \end{aligned} \right. \\ \{X(ab) = X(a)X(b)\}. \end{aligned}$$

Here $\widehat{\mathcal{B}}_n$ is the collection of braid diagrams, described by braid words. On the diagrammatic side we are not including ambient isotopy. On the algebraic side $\widehat{\mathcal{B}}_n$ is the **free group** on $\sigma_1, \dots, \sigma_{n-1}, \sigma_1^{-1}, \dots, \sigma_{n-1}^1$, \mathcal{I} (modulo $\sigma_i\sigma_j = \sigma_j\sigma_i$ for $|i - j| > 1$). Then if $R = h_1 h_3 \dots h_{2n-1}$ and $\beta \in \widehat{\mathcal{B}}_{2n}$ we have

$$RX(\beta)R = W(P(\beta))R,$$

a completely algebraic calculation of W using the operator algebra \mathcal{U}_n .

Example.  $= P\left(\begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array}\right) = P(\sigma_2 \sigma_1^{-1} \sigma_3^{-1} \sigma_2)$

$$\beta = \sigma_2 \sigma_1^{-1} \sigma_3^{-1} \sigma_2$$

$$\begin{aligned} X(\beta) &= (\mathcal{I} + xh_2)(x\mathcal{I} + h_1)(x\mathcal{I} + h_3)(\mathcal{I} + xh_2) \\ &= (x\mathcal{I} + x^2h_2 + h_1 + xh_2h_1)(x\mathcal{I} + h_3)(\mathcal{I} + xh_2) \\ &= x^2 + x^3h_2 + xh_1 + x^2h_2h_1 + xh_3 + x^2h_2h_3 + h_1h_3 + xh_2h_1h_3(\mathcal{I} + xh_2) \end{aligned}$$

$$\begin{aligned} X(\beta) &= x^2 + x^3h_2 + xh_1 + x^2h_2h_1 \\ &\quad + xh_3 + x^2h_2h_3 + h_1h_3 + xh_2h_1h_3 \\ &\quad + x^3h_2 + x^4h_2^2 + x^2h_1h_2 + x^2h_2h_1h_2 \\ &\quad + x^2h_3h_2 + x^3h_2h_3h_2 + xh_1h_3h_2 \\ &\quad + x^2h_2h_1h_3h_2. \end{aligned}$$

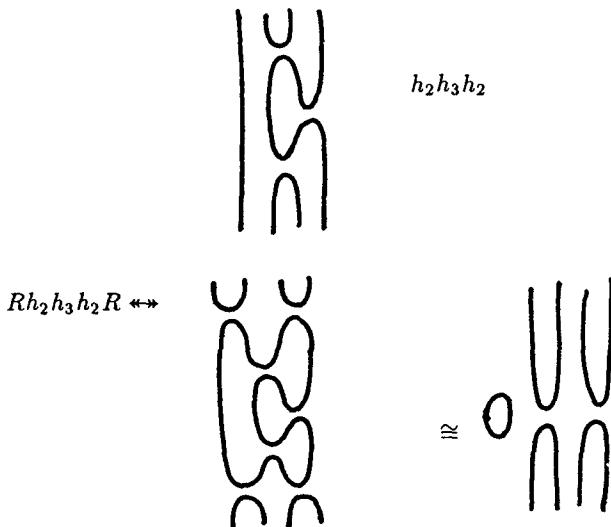
These correspond to the sixteen states of K . Let's take one term and look at it algebraically and geometrically. For this problem, $\beta \in B_4$ and

$$R = h_1h_3 \leftrightarrow \begin{array}{cc} \cup & \cup \\ \cap & \cap \end{array}$$

Take $RX(\beta)R$ and look at the term

$$\begin{aligned} R[x^2h_2h_3h_2]R \\ &= x^2h_1h_3h_2h_3h_2h_1h_3 \\ &= x^2h_1h_3h_2h_1h_3 \\ &= x^2h_1h_3h_2h_3h_1 \\ &= x^2h_1h_3h_1 \\ &= x^2h_1^2h_3 \\ &= (x^2y)h_1h_3 \\ &= (x^2y)R. \end{aligned}$$

Geometrically:



and the extra component gives rise to the y in the algebra.

To summarize, we have shown that for any braid diagram $\beta \in \widehat{B}_{2n}$, the W -polynomial is given algebraically via

$$W(P(\beta))R = RX(\beta)R.$$

This yields an operator-algebraic formulation of the Potts model that is formally isomorphic to that given by Temperley and Lieb (see [BA1]). Our method constructs the operator algebra purely geometrically. In fact, this formulation makes some of the statistical mechanical features of the model quite clear. For example, in the anti-ferromagnetic case of a large rectangular lattice one expects the critical point to occur when there is a symmetry between the partition function on the lattice and the dual lattice. In the bracket reformulation of the Potts model this corresponds to having $W(\bowtie) = W(\times)$ and this occurs when $q^{-(1/2)}v = 1$, Hence the critical temperature occurs (conjecturally) at

$$e^{(1/kT)} - 1 = \sqrt{q}$$

or

$$T = \frac{1}{k \ln(1 + \sqrt{q})} .$$

Obviously, more work needs to be done in this particular interconnection of knots, combinatorics and physics.

Exercise. Show that the dichromatic polynomial of any planar graph G can be expressed via the Temperley-Lieb algebra by showing that $Z_G = q^{N/2} \{\tilde{K}(G)\}$ where $\{L\} = W_L$ is the Potts bracket, N is the number of vertices of G and $\tilde{K}(G)$ is a link in plat form. (HINT: Obtain $\tilde{K}(G)$ from the alternating medial link $K(G)$ by arranging maxima and minima.) Consequences of this exercise will appear in joint work of Hubert Saleur and the author. Compare with [SAL2].

9⁰. Preliminaries for Quantum Mechanics, Spin Networks and Angular Momentum.

This section is preparatory to an exposition of the Penrose theory of spin-networks. These networks involve an evaluation process very similar to our state summations and chromatic evaluations. And by a fundamental result of Roger Penrose, there is a close relation to three-dimensional geometry in the case of certain large nets. This is remarkable: a three-dimensional space becomes the appropriate **context** for the interactive behaviors in a large, purely combinatorially defined network of spins.

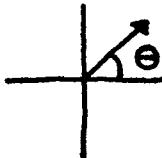
We begin with a review of spin in quantum mechanics.

Spin Review. (See [FE].)

Recall that in the quantum mechanical framework, to every **process** there corresponds an **amplitude**, a complex number ψ . ψ is usually a function of space and time: $\psi(x, t)$. With proper normalization, the probability of a process is equal to the absolute square of this amplitude:

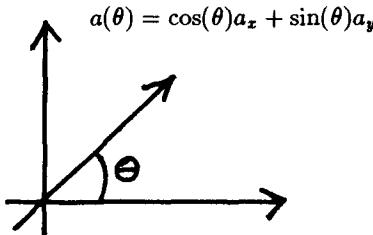
$$\text{prob}(\psi) = \psi^* \psi$$

where ψ^* denotes the complex conjugate of ψ . I am using the term process loosely, to describe a condition that includes a physical process **and** a mode of observation. Thus if we observe a beam of polarized light, there is an amplitude for observing polarization in a given direction θ (This could be detected by an appropriately oriented filter.).



Consider a beam of light polarized in a given direction. Let the axis of an analyzer (polaroid filter or appropriate prism) be placed successively in two perpendicular directions x, y to measure the number of photons passing a corresponding lined-up polarizer. [x and y are orthogonal to the direction of the beam.] Let a_x

and a_x be the amplitudes for these directions. Then, if the analyzer is rotated by θ degrees, with respect to the x -direction, the amplitude becomes the superposition



This result follows from the two principles:

- 1) Nature has no preferred axis. [Hence the results must be invariant under rotation.]
- 2) Quantum amplitudes for a mixed state undergo linear superposition.

The polarizer experiment already contains the essence of quantum mechanics. Note, for example, that it follows from this description that the amplitude for the process

x -filter followed by y -filter

is zero, while the process

x -filter / 45° -filter / y -filter

has a non-zero amplitude in general. Two polarizing filters at 90° to one another block the light. Insert a third filter between them at an angle, and light is transmitted.

The two principles of **non-preferential axis** and **linear superposition** are the primary guides to the mathematics of the situation.

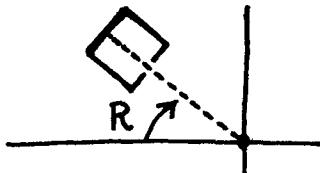
In the more general situation of a beam of particles there may be a vector of amplitudes

$$\vec{a} = (a_1, a_2, \dots, a_n).$$

Rotation of the analyzer will produce a new amplitude

$$\vec{a}' = D(R)\vec{a}$$

where R denotes the rotation in three-dimensional space of the analyzer.



If S and R are two rotations, let SR denote their composition (first R , then S). We require that $D(SR) = D(S)D(R)$ and that $D(R)$ and $D(S)$ are linear transformations.

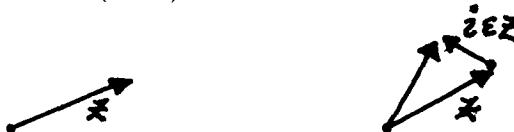
Linearity follows from the superposition principle, and composition follows essentially from independence of axis. That is, suppose \vec{a} is the amplitude for a standard direction and that we rotate by R to get $\vec{a}' = D(R)\vec{a}$ and then rotate by S to get $\vec{a}'' = D(S)D(T)\vec{a}$. This is the same experiment as rotating by SR and so $\vec{a}'' = D(SR)\vec{a}$ also, whence $D(SR) = D(S)D(R)$. (This is true up to a phase factor.)

In order to discuss the possibilities for these representations, recall that over the complex numbers a rotation of the plane can be represented by

$$e^{i\theta} = \lim_{n \rightarrow \infty} (1 + i\theta/n)^n.$$

Note that this is a formal generalization of the real limit $e^x = \lim_{n \rightarrow \infty} (1 + x/n)^n$, and note that

$$(1 + i\epsilon)z = z + i\epsilon z$$



Multiplication of a complex number z by $(1 + i\epsilon)$ has the effect of rotating it by approximately ϵ radians in a counter-clockwise direction in the complex plane. Thus we can regard $(1 + i\epsilon)$ as an "infinitesimal rotation".

Following this lead, we shall try

$$D(\epsilon/x) = 1 + i\epsilon M_x$$

$$D(\epsilon/y) = 1 + i\epsilon M_y$$

$$D(\epsilon/z) = 1 + i\epsilon M_z$$

where these represent small angular rotations about the x , y and z axes respectively. For this formalism, $D(\theta/z)$ denotes the matrix for a rotation by angle θ about the z -axis, and so

$$\begin{aligned} D(\theta/z) &= \lim_{\epsilon \rightarrow 0} (1 + i\epsilon M_z)^{\theta/\epsilon} \\ &= (e^{iM_z})^\theta \\ D(\theta/z) &= e^{i\theta M_z}. \end{aligned}$$

For a given vector direction \vec{v} , we have $D(\theta/\vec{v}) = e^{i\theta(\vec{v} \cdot \vec{M})}$ where

$$\begin{aligned} \vec{M} &= (M_x, M_y, M_z) \\ \vec{v} \cdot \vec{M} &= v_x M_x + v_y M_y + v_z M_z. \end{aligned}$$

Proposition 9.1. $D(R)$ is a unitary matrix. That is, $D(R)^{-1} = D(R)^*$ where $*$ denotes conjugate transpose.

Proof. It is required that $\psi\psi^*$ be invariant under rotation. But if $\psi' = D\psi$, then

$$\begin{aligned} \psi'\psi'^* &= D\psi(D\psi)^* = \sum_i (D\psi)_i (\psi^* D^*)_i \\ &= \sum_i \left(\sum_j D_{ji} \psi_j \right) \left(\sum_k \psi_k^* D_{ki}^* \right) \\ &= \sum_{jk} \sum_i D_{ji} \bar{D}_{ik} \psi_j \bar{\psi}_k \\ &= \sum_{j,k} (DD^*)_{jk} \psi_j \bar{\psi}_k. \end{aligned}$$

In order for this last formula to equal $\psi\psi^*$ for all possible, ψ we need $DD^* = I$.

//

Since $D = e^{i\theta M}$ is unitary, we conclude that

$$1 = DD^* = e^{i\theta(M - M^*)}.$$

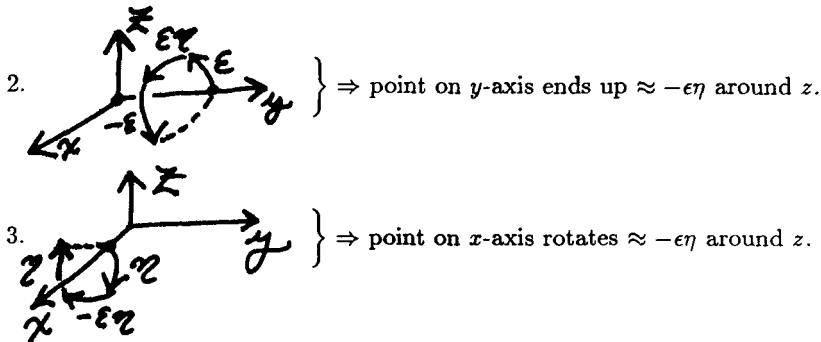
Thus $M = M^*$ and so M is Hermitian.

Thus the matrices M are candidates for **physical observables** in the quantum theory. [Since they have real eigenvalues, and the quantum mechanical model for observation is $M\psi = \lambda\psi$ for some amplitude ψ .]

Consequently, we want to know more about M_x , M_y and M_z . In particular, we want their eigenvalues and how they commute.

To obtain commutation relations for M_x and M_y , consider a rotation by ϵ about x -axis followed by rotation by η about y -axis followed by rotation by $-\epsilon$ about x -axis followed by rotation by $-\eta$ about y -axis.

- Up to 1st order the z -axis is left fixed.



Thus

$$\begin{aligned}
 (1 - i\epsilon\eta M_z) &\approx D(-\eta/y)D(-\epsilon/x)D(\eta/y)D(\epsilon/x) \\
 &= (1 - i\eta M_y)(1 - i\epsilon M_x)(1 + i\eta M_y)(1 + i\epsilon M_x) \\
 \\
 &\Rightarrow -iM_z = -M_y M_x - M_y M_x + M_y M_x + M_x M_y.
 \end{aligned}$$

So

$$M_x M_y - M_y M_x = -iM_z.$$

Because of this minus sign it is convenient to rewrite to the conjugate representation. Then we have:

$M_x M_y$	$-$	$M_y M_x$	$=$	iM_z	(*)
$M_y M_z$	$-$	$M_z M_y$	$=$	iM_x	
$M_z M_x$	$-$	$M_x M_z$	$=$	iM_y	

Eigenstructure.

Let

$$\begin{aligned} M^2 &= M_x^2 + M_y^2 + M_z^2 \\ M_- &= M_x - iM_y \\ M_+ &= M_x + iM_y. \end{aligned}$$

Note that M^2 commutes with M_x , M_y and M_z . (Exercise - use the commutation relations (*).)

Proposition 9.2. $M_x M_- = M_-(M_z - 1)$
 $M_x M_+ = M_+(M_z + 1)$.

Proof.

$$\begin{aligned} M_x M_- &= M_x(M_x - iM_y) \\ &= M_x M_x - iM_x M_y \\ &= M_x M_z + iM_y - iM_z M_y \\ &= (M_x - iM_y)(M_z - 1) \\ \therefore M_x M_- &= M_-(M_z - 1) \\ M_x M_+ &= M_x(M_x + iM_y) \\ &= M_x M_z + iM_y + iM_z M_y \\ &= (M_x + iM_y)(M_z + 1) \\ \therefore M_x M_+ &= M_+(M_z + 1). \end{aligned}$$

//

Now suppose that $M_x a^{(m)} = m a^{(m)}$. That is, suppose that $a^{(m)}$ is an eigenvector of $a^{(m)}$ with eigenvalue m . Then

$$\begin{aligned} M_x M_- a^{(m)} &= M_-(M_x - 1)a^{(m)} \\ &= M_-(m - 1)a^{(m)} \\ &= (m - 1)M_- a^{(m)}. \end{aligned}$$

Thus $M_- a^{(m)}$ becomes a new eigenvector of M_x with eigenvalue $(m - 1)$.

We can define

$$ca^{(m-1)} = M_- a^{(m)}$$

with the constant c to be determined.

Assume that $a^{(m)}$ is normalized for each m so that $1 = \sum_n a_n^{(m)} a_n^{(m)*}$. Then

$$1 = \frac{1}{cc^*} \sum_n [M_- a^{(m)}]_n [M_- a^{(m)}]_n^*$$

Hence

$$\begin{aligned} 1 &= \frac{1}{cc^*} \sum_n [M_- a^{(m)}]_n^* [M_- a^{(m)}]_n \\ &= \frac{1}{cc^*} \sum_n [a^{(m)*}]_n [M_-^* M_- a^{(m)}]_n \\ &= \frac{1}{cc^*} \sum_n [a^{(m)*}]_n [M_+ M_- a^{(m)}]_n. \end{aligned}$$

Now

$$\begin{aligned} M_+ M_- &= (M_x + iM_y)(M_x - iM_y) \\ &= M_x^2 + M_y^2 + i(M_y M_x - M_x M_y) \\ &= M_x^2 + M_y^2 + i(-iM_z) \\ &= M_x^2 + M_y^2 + M_z \\ M_+ M_- &= M^2 - M_z^2 + M_z. \end{aligned}$$

We may choose $a^{(m)}$'s so that

$$M^2 a^{(m)} = k a^{(m)}$$

(since M^2 commutes with the other operators) for a fixed k , and ask for the resulting range of m -values.

$$\begin{aligned} c^* c &= \sum_n a_n^{(m)*} (M_+ M_-) a_n^{(m)} \\ \left(\begin{array}{rcl} M_- a^{(m)} &=& ca^{(m-1)} \\ M_+ a^{(m-1)} &=& c^* a^{(m)} \end{array} \right) \\ \Rightarrow |c|^2 &= k - m^2 + m. \end{aligned}$$

Thus $|c| = (k - m(m-1))^{1/2}$. From this it follows that $k = j(j+1)$ and $2j$ is an integer.

Proof. Let $m = -j$ be the lowest state - meaning that $M_- a^{(m)} = 0$. This implies $c = 0$ for $m = -j$ whence $0 = k - (-j)(-j - 1)$ whence $k = j(j + 1)$.

By the same argument, if the largest value of m is j' then $k = j'(j' + 1)$. Thus $j = j'$ and $2j$ is integer. //

Example. $j = 1/2$.

Then there are two states. Let

$$a^{(1/2)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad a^{(-1/2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\begin{aligned} c &= [j(j + 1) - m(m - 1)]^{1/2} \\ &= \left[\frac{1}{2} \left(\frac{3}{2} \right) - \frac{1}{2} \left(-\frac{1}{2} \right) \right]^{1/2} = [4/4]^{1/2} = 1 \\ \therefore c &= 1. \end{aligned}$$

Thus

$$\left. \begin{aligned} M_- \begin{pmatrix} 1 \\ 0 \end{pmatrix} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ M_- \begin{pmatrix} 0 \\ 1 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned} \right\} \Rightarrow M_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

$$\left. \begin{aligned} M_z \begin{pmatrix} 1 \\ 0 \end{pmatrix} &= \frac{1}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ M_z \begin{pmatrix} 0 \\ 1 \end{pmatrix} &= -\frac{1}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{aligned} \right\} \Rightarrow M_z = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Similarly, $M_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and

$$\left. \begin{aligned} M_x &= \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{1}{2} \sigma_x \\ M_y &= \frac{1}{2} \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} = \frac{1}{2} \sigma_y \\ M_z &= \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \frac{1}{2} \sigma_z \end{aligned} \right\} \boxed{\begin{array}{lcl} M_- &=& M_x - iM_y \\ M_+ &=& M_x + iM_y \end{array}}$$

where $\sigma_x, \sigma_y, \sigma_z$ are the Pauli matrices satisfying

$$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$$

$$\sigma_x \sigma_y = -\sigma_y \sigma_x = i\sigma_z$$

Quick Review of Wave Mechanics.

$$\psi = Ae^{i(kx - \omega t)}$$

$$\frac{\partial \psi}{\partial t} = -i\omega \psi \quad \boxed{E = h\nu = \hbar\omega}$$

$$(h = \hbar/2\pi)$$

$$i\hbar \frac{\partial \psi}{\partial t} = E\psi \quad \boxed{p = \hbar k}$$

$$-i\hbar \frac{\partial \psi}{\partial x} = \hbar k\psi = p\psi.$$

De Broglie suggested (1923) that particles such as electrons could be regarded as having wave-like properties, and his fundamental associations were

$$E = \hbar\omega$$

$$p = \hbar k$$

for energy E and momentum p . These quantities can be combined in a simple wave-function $\psi = e^{i(kx - \omega t)}$.

This led Schrödinger in the same year to postulate the equation

$$\boxed{i\hbar \frac{\partial \psi}{\partial t} = H\psi}$$

where $H = p^2/2m + V$ represents total energy, kinetic plus potential. And one associated the operator

$$\frac{\hbar}{i} \frac{\partial}{\partial x}$$

to momentum. So

$$H = \frac{(-i\hbar)^2}{2m} \frac{\partial^2}{\partial x^2} + V = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V.$$

At about the same time, Heisenberg invented matrix mechanics, based on non-commutative algebra. Heisenberg's mechanics is based on an algebra where position x and momentum p no longer commuted, but rather

$$xp - px = \hbar i.$$

The formal connection of this algebra and the Schrödinger mechanics is contained in the association of operators (such as $(\hbar/i) \partial/\partial x$ to p) for physical quantities.

Note that

$$\begin{aligned} & \frac{\hbar}{i} \frac{\partial}{\partial x}(xf) - x \frac{\hbar}{i} \frac{\partial}{\partial x}(f) \\ &= \frac{\hbar}{i} \left[f + x \frac{\partial f}{\partial x} - x \frac{\partial f}{\partial x} \right] = \frac{\hbar}{i} f. \end{aligned}$$

Thus $xp - px = \hbar i$ in the Schrödinger theory as well.

One consequence is that position and momentum are not simultaneously observable. Observability is associated with a Hermitian operator, and its eigenbehavior.

It took some years, and debate before Max Born made the interpretation that ψ measures a “complex probability” for the event - with $\psi^* \psi$ as the real probability.

To return to the subject of rotations, let us consider the role of angular momentum in quantum mechanics. Let

$$\begin{aligned} \vec{r} &= ix + jy + kz \\ \vec{p} &= ip_x + jp_z + kp_z \end{aligned}$$

where i, j, k denotes the standard basis for \mathbf{R}^3 . The classical angular momentum is given by $\vec{L} = \vec{r} \times \vec{p}$.

$$\vec{L} = \vec{r} \times \vec{p} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ x & y & z \\ p_x & p_y & p_z \end{vmatrix}.$$

In quantum mechanics, $\vec{p} = i\hbar \vec{\nabla}$. Thus

$$\begin{aligned} L_x &\leftrightarrow \frac{\hbar}{i} \left(y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right) \\ L_y &\leftrightarrow -\frac{\hbar}{i} \left(x \frac{\partial}{\partial z} - z \frac{\partial}{\partial x} \right) \\ L_z &\leftrightarrow \frac{\hbar}{i} \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) \\ L^2 &= |\vec{L} \cdot \vec{L}| = L_x^2 + L_y^2 + L_z^2. \end{aligned}$$

Let $[a, b] = ab - ba$. Then

$$\left. \begin{array}{lcl} [L_x, L_y] & = & i\hbar L_z \\ [L_y, L_z] & = & i\hbar L_x \\ [L_z, L_x] & = & i\hbar L_y \end{array} \right\} \vec{L} \times \vec{L} = i\hbar \vec{L}.$$

$[L^2, L_z] = 0$, so L^2 commutes with \vec{L} .

The formalism for angular momentum in quantum mechanics is identical to the formalism for unitary representation of rotational invariance. The operators for angular momentum correspond to rotational invariance of ψ .

Digression on Wave Mechanics.

Note that $\sin(X + Y) + \sin(X - Y) = 2 \sin(X) \cos(Y)$.

$$\boxed{\sin\left(\frac{2\pi}{\lambda}(x - ct)\right)} \quad \begin{aligned} \lambda &= \text{wave-length} \\ c &= \text{wave-velocity} \end{aligned}$$

$$\sin\left(\frac{2\pi}{\lambda}(x - ct)\right) = \sin(kx - wt)$$

$$k = 2\pi/\lambda, \quad w = 2\pi c/\lambda = 2\pi\nu.$$

Let

$$u = a \left[\underbrace{\sin(kx - wt)}_{X+Y} + \underbrace{\sin(k'x - w't)}_{X-Y} \right]$$

$$\Rightarrow X = \left(\frac{k+k'}{2} \right)x - \left(\frac{w+w'}{2} \right)t$$

$$Y = \left(\frac{k-k'}{2} \right)x - \left(\frac{w-w'}{2} \right)t$$

$$u = a \sin \left[\left(\frac{k+k'}{2} \right)x - \left(\frac{w+w'}{2} \right)t \right] \cos \left[\left(\frac{k-k'}{2} \right)x - \left(\frac{w-w'}{2} \right)t \right].$$

$$\text{If } k - k' = \delta k, w - w' = \delta w \text{ then } u \approx \left[a \cos \left(\frac{\delta k}{2}x - \frac{\delta w}{2}t \right) \right] \sin(kx - wt)$$



$$V_g = \text{group velocity} = \delta w / \delta k$$

$$\therefore V_g = \frac{d(c/\lambda)}{d(1/\lambda)} = \frac{dc/d\lambda}{\lambda d(1/\lambda)} + c = -\lambda \frac{dc}{d\lambda} + c.$$

De Broglie postulated $E = h\nu = \hbar w$ and, by analogy with photon,

$p = E/c = \frac{h\nu}{c} = \frac{h}{\lambda} = \frac{\hbar 2\pi}{\lambda} = \hbar k$. With these conventions we have

$$V_g = \frac{d(c/\lambda)}{d(1/\lambda)} = \frac{d\nu}{d(1/\lambda)} = \frac{d(h\nu)}{d(h/\lambda)}$$

$$\therefore V_g = dE/dp$$

Compare this with the classical situation:

$$p = mv$$

$$E = \frac{1}{2}mv^2 + U \quad (U = \text{ potential})$$

$$\frac{dE}{dp} = \frac{1}{m} \frac{dE}{dv} = \frac{1}{m}(mv + 0) = v.$$

De Broglie's identification of energy $E = \hbar w$ and momentum $p = \hbar k$ creates a correspondence of classical velocity with the group velocity of wave packets.

Digression on Relativity.

In this digression I want to show how the pattern

$$\left. \begin{array}{l} \sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1 \\ \sigma_x \sigma_y = \sqrt{-1} \sigma_z \\ \sigma_y \sigma_z = \sqrt{-1} \sigma_x \\ \sigma_z \sigma_x = \sqrt{-1} \sigma_y \end{array} \right\}$$

arises in special relativity.

In special relativity there are two basic postulates:

1. Physical laws take the same form in two reference frames that are moving at constant velocity with respect to one another.
2. The speed of light as observed by persons in two inertial frames (i.e. moving at constant velocity with respect to one another) is the same.

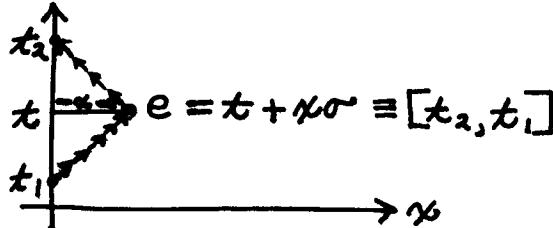
The first postulate of special relativity is almost a definition of physical law. The second postulate is far less obvious. It is justified by experiment (the Michelson-Morley experiment) and the desire to have Maxwell's equations for electromagnetism satisfy the first postulate. [In the Maxwell theory there is a traveling wave solution that has velocity corresponding to the velocity of light. This led to the identification of light/radio waves and the waves of this model. If the Maxwell equations indeed satisfy 1., then 2. follows.]

Now let's consider events in special relativity. For a given observer, an event may be specified by two times: $[t_2, t_1] = e$. The two times are

$$t_1 = \text{time signal sent}$$

$$t_2 = \text{time signal received.}$$

In this view, the observer sends out a light signal at time t_1 . The signal reflects from the event (as in radar), and is received by the same observer (transmitted along his/her world-line) at time t_2 .



I shall assume (by convention) that the speed of light equals unity. Then the world paths of light are indicated by 45° lines (to the observer's (vertical) time line t). Thus the arrow pathways in the diagram above indicate the emission of a signal at time t_1 , its reception (reflection) at the event at time t , and its reception (after reflection) by the observer at time t_2 . With $c = \text{light speed} = 1$, we have that

$$t_2 - t_1 = 2x$$

$$t_2 + t_1 = 2t$$

where x and t are the space and time coordinates of the event e . Thus $[t_2, t_1] = [t + x, t - x]$ gives the translation from light-cone (radar) coordinates to spacetime coordinates.

Regard the ordered pair $[a, b]$ additively so that: $[a, b] + [c, d] = [a + c, b + d]$ and $k[a, b] = [ka, kb]$ for a scalar k . Then

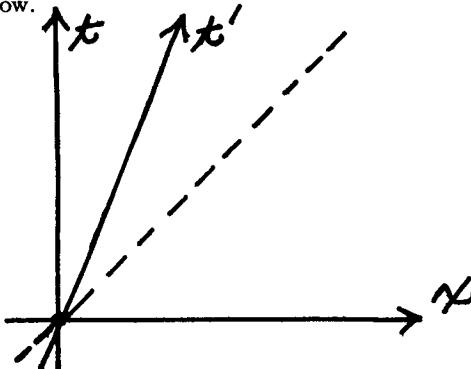
$$[t_2, t_1] = t[1, 1] + x[1, -1].$$

Let 1 abbreviate $[1, 1]$ and σ abbreviate $[1, -1]$. Then

$$e = [t_2, t_1] = t1 + x\sigma.$$

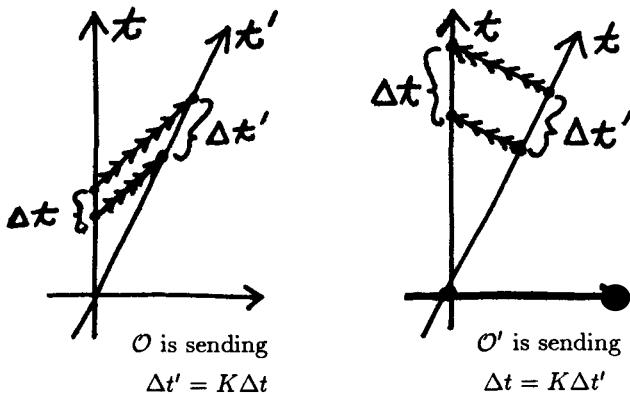
This choice and translation of coordinates will be useful to us shortly.

Now let's determine how the light-cone coordinates of an event are related for two reference systems that move at constant velocity with respect to one another. In our representation, the world-line for a second inertial observer appears as shown below.



A coordinate (x, t) on the line labelled t' denotes the position and time as measured by \mathcal{O} of a "stationary" observer \mathcal{O}' in the second coordinate system. Note that the relative velocity of this observer is $v = x/t$. Since the velocity is constant, the world-line appears straight. (Note also that a plot of distance as measured by \mathcal{O}' would not necessarily superimpose on the \mathcal{O}' diagram at 90° to the line t' .)

Thus we can draw the system \mathcal{O} and \mathcal{O}' in the same diagram, and consider the pictures of signals sent from \mathcal{O} to \mathcal{O}' :

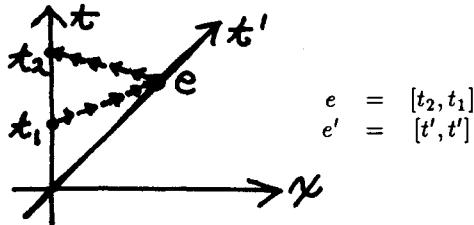


The diagrams above tell the story. If \mathcal{O} sends two signals to \mathcal{O}' at a time interval Δt , then \mathcal{O}' will receive them at a time interval $K\Delta t$. The constant K is defined by

$$\boxed{\frac{\Delta \text{ (time received)}}{\Delta \text{ (time sent)}} = K}.$$

Here we use the first principle of relativity to deduce that $\mathcal{O} \rightarrow \mathcal{O}'$ (\mathcal{O} sending to \mathcal{O}') and $\mathcal{O}' \rightarrow \mathcal{O}$ have the **same constant K** . In fact, K is a function of the relative velocity of \mathcal{O} and \mathcal{O}' . (See [BO]. This is Bondi's K -calculus.)

Let's find the value of K .



Consider a point e on \mathcal{O}' 's time-line. Then $e = [t_2, t_1]$ from \mathcal{O} 's viewpoint and $e' = [t', t']$ represents the same event from the point of view of \mathcal{O}' (just a time, no distance).

Then we have

$$t' = Kt_1$$

$$t_2 = Kt'$$

(by relativity and the assumption that \mathcal{O} and \mathcal{O}' have synchronized clocks at the origin).

Thus $t_2 = K^2 t_1$ and we know that

$$t_2 = t + x$$

$$t_1 = t - x$$

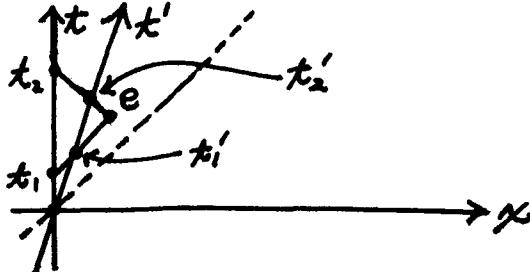
where t and x are the time-space coordinates of e from \mathcal{O} . Thus

$$\begin{aligned} t + x &= K^2(t - x) \\ \Rightarrow 1 + (x/t) &= K^2(1 - (x/t)) \end{aligned}$$

Here $1 + v = K^2(1 - v)$ where v is the relative velocity of \mathcal{O} and \mathcal{O}' . Thus

$$K = \sqrt{\frac{1+v}{1-v}}.$$

We now use K to determine the general transformation of coordinates:



The event e is now off \mathcal{O}' 's time-line and \mathcal{O} sends a signal at t_1 that passes \mathcal{O}' at t'_1 (\mathcal{O}' measurement), reflects from e , passes \mathcal{O}' at t'_2 , is received by \mathcal{O} at t_2 . Then we have (using the same assumptions)

$$t'_1 = Kt_1$$

$$t_2 = Kt'_2$$

Hence

$$[t'_2, t'_1] = [K^{-1}t_2, Kt_1].$$

This is the Lorentz transformation in light-cone coordinates.

With the help of the velocity formula for K and the identifications

$$\begin{array}{rcl} t_2 & = & t + x \\ t_1 & = & t - x \end{array} \qquad \begin{array}{rcl} t'_2 & = & t' + x' \\ t'_1 & = & t' - x' \end{array}$$

it is an easy matter to transform this version of the Lorentz transformation into the usual coordinates. I prefer the following slightly more algebraic approach [LK17], [LK30].

1: Regard $[K^{-1}t_2, Kt_1]$ as a special case of a product structure given by the formula

$$[A, B] * [C, D] = [AC, BD].$$

Thus $[K^{-1}, K]*[t_2, t_1] = [K^{-1}t_2, Kt_1] = [t'_2, t'_1]$ so that the Lorentz transformation is expressed through this product. I call this the **iterant algebra** [LK17].

2: Rewrite

$$\begin{aligned}[K^{-1}, K] &= [X + T, -X + T] \\ &= T1 + X\sigma.\end{aligned}$$

Then

$$\begin{aligned}T + X\sigma &= \frac{1 + (X/T)\sigma}{\sqrt{1 - (X/T)^2}} \\ (T^2 - X^2 = K^{-1}K = 1) &\Rightarrow \sqrt{T^2 - X^2} = 1 \\ \Rightarrow T\sqrt{1 - (X/T)^2} &= 1\end{aligned}$$

and $v = -(X/T)$ because

$$\begin{aligned}K &= \sqrt{\frac{1+v}{1-v}} \\ \Rightarrow T + X &= \sqrt{\frac{1-v}{1+v}} \\ T - X &= \sqrt{\frac{1+v}{1-v}} \\ \Rightarrow \frac{1+(X/T)}{1-(X/T)} &= \frac{\sqrt{(1-v)/(1+v)}}{\sqrt{(1+v)/(1-v)}} = \frac{1-v}{1+v} \\ \therefore [K^{-1}, K] &= \boxed{\frac{1-v\sigma}{\sqrt{1-v^2}}}\end{aligned}$$

Using this iterant algebra, we have

$$1^2 = [1, 1] * [1, 1] = [1, 1] = 1$$

$$\sigma^2 = \sigma * \sigma = [1, -1] * [1, -1] = [1, 1] = 1$$

$$1 * \sigma = \sigma * 1 = \sigma.$$

Thus

$$\begin{aligned}[t'_2, t'_1] &= [K^{-1}, K] * [t_2, t_1] \\ \Updownarrow \\ t' + x'\sigma &= \left(\frac{1 - v\sigma}{\sqrt{1 - v^2}} \right) * (t + x\sigma).\end{aligned}$$

Thus

$$\boxed{\begin{aligned} t' &= (t - xv)/\sqrt{1 - v^2} \\ x' &= (x - vt)/\sqrt{1 - v^2}\end{aligned}}.$$

This is the classical form of the Lorentz transformation with light speed equal to unity.

Remark. We are accustomed to deriving the Lorentz transformation and then noting that $t'^2 - x'^2 = t^2 - x^2$ and that the group of Lorentz transformations is the group of linear transformations leaving the form $t^2 - x^2$ invariant. Here we knew that $t'_2 t'_1 = K^{-1} t_2 K t_1 = t_2 t_1$ early on, and hence that

$$\begin{aligned}(t' + x')(t' - x') &= (t + x)(t - x) \\ \therefore t'^2 - x'^2 &= t^2 - x^2.\end{aligned}$$

This digression on relativity connects with our previous discussion of spin and angular momentum as follows: **The algebra of the Pauli matrices arises naturally in relation to the Lorentz group.** To see this we must consider what we have done so far in the context of a Euclidean space of spatial directions and displacements. So let us suppose that spacetime has dimension $n + 1$ with n dimensions of space. Let σ_0 denote the temporal direction and let $\sigma_1, \sigma_2, \dots, \sigma_n$ be a basis of unit spatial directions. Thus we identify

$$\sigma_1 = (1, 0, 0, \dots, 0)$$

$$\sigma_2 = (0, 1, 0, \dots, 0)$$

$$\vdots$$

$$\sigma_n = (0, 0, 0, \dots, 1).$$

Let $\sigma = a_1\sigma_1 + a_2\sigma_2 + \dots + a_n\sigma_n$ denote a Euclidean unit spatial direction so that $a_1^2 + a_2^2 + \dots + a_n^2 = 1$.

We are given that

$$1) \sigma_i^2 = 1 \quad i = 1, \dots, n$$

$$2) \sigma^2 = 1$$

by our assumption that $t\sigma_0 + x\sigma_i$ ($i = 1, \dots, n$) and $t\sigma_0 + x\sigma$ represent points in spacetime with these given directions. (Any given, spatial direction produces a two-dimensional spacetime that must obey the laws we have set out for such structures.)

Now multiply out the formula for σ^2 :

$$\begin{aligned} \sigma^2 &= (a_1\sigma_1 + \dots + a_n\sigma_n)(a_1\sigma_1 + \dots + a_n\sigma_n) \\ \sigma^2 &= \sum_{i=1}^n a_i^2 \sigma_i^2 + \sum_{i \neq j} a_i a_j \sigma_i \sigma_j \\ &= \sum_{i=1}^n a_i^2 + \sum_{i < j} a_i a_j (\sigma_i \sigma_j + \sigma_j \sigma_i) \\ \therefore \sigma^2 &= 1 + \sum_{i < j} a_i a_j (\sigma_i \sigma_j + \sigma_j \sigma_i). \end{aligned}$$

Since $\sigma^2 = 1$ for all choices of (a_1, a_2, \dots, a_n) with $a_1^2 + \dots + a_n^2 = 1$, this calculation implies that

$$\sigma_i \sigma_j = -\sigma_j \sigma_i \text{ for } i \neq j.$$

Let's look at the case of 4-dimensional spacetime, and then we have

$$e = t\sigma_0 + x\sigma_1 + y\sigma_2 + z\sigma_3.$$

If we take $\sigma_1, \sigma_2, \sigma_3$ as Pauli matrices then,

$$\sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The spacetime point $e = t\sigma_0 + x\sigma_1 + y\sigma_2 + z\sigma_3$ is then represented by a 2×2 Hermitian matrix as:

$$\begin{aligned} e &= t \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + x \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + y \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} + z \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ e &= \begin{pmatrix} t+x & iy+z \\ -iy+z & t-x \end{pmatrix}. \end{aligned}$$

Note that the determinant of e is the spacetime metric:

$$\text{Det}(e) = \begin{vmatrix} t+x & iy+z \\ -iy+z & t-x \end{vmatrix}$$

$$\text{Det}(e) = t^2 - x^2 - y^2 - z^2.$$

This is a (well-known) that leads to a very perspicuous description of the full group of Lorentz transformations.

That is, we wish to consider the construction of linear transformations of $\mathbf{R}^4 = \{(t, x, y, z)\}$ that leave the quadratic form $t^2 - x^2 - y^2 - z^2$ invariant. Let's use the Hermitian representation,

$$e = \begin{pmatrix} A & W \\ \bar{W} & B \end{pmatrix}$$

where A and B are real numbers, and W is complex, with conjugate \bar{W} . Let P be any complex 2×2 matrix with $\text{Det}(P) = 1$. The group of such matrices is called $SL(2, \mathbf{C})$. Then $(PeP^*)^* = Pe^*P^* = PeP^*$. Hence PeP^* is Hermitian. Furthermore,

$$\begin{aligned} \text{Det}(PeP^*) &= \text{Det}(P)\text{Det}(e)\text{Det}(P^*) \\ &= \text{Det}(P)\text{Det}(e)\text{Det}(P) \\ &= \text{Det}(e). \end{aligned}$$

Thus, if \mathcal{L} denotes the group of Lorentz transformations, then we have exhibited a $2 - 1$ map $\pi : SL(2, \mathbf{C}) \rightarrow \mathcal{L}$.

This larger group of Lorentz transformations includes our directional transforms

$$[A, B] \mapsto [K^{-1}A, KB].$$

Let's see how:

$$[A, B] = [T + X, T - X] \leftrightarrow \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

$$\therefore [K^{-1}A, KB] \leftrightarrow \begin{bmatrix} K^{-1}A & 0 \\ 0 & KB \end{bmatrix} = \begin{bmatrix} 1/\sqrt{K} & 0 \\ 0 & \sqrt{K} \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} 1/\sqrt{K} & 0 \\ 0 & \sqrt{K} \end{bmatrix}.$$

Quaternionic Representation.

The Pauli matrices and the quaternions are intimately related. In the quaternions we have three independent orthogonal unit vectors i, j, k with $i^2 = j^2 = k^2 = ijk = -1$ where the vectors $1, i, j, k$ form a vector space basis for 4-dimensional real space \mathbf{R}^4 . This gives \mathbf{R}^4 the structure of an associative division algebra. The translation between Pauli matrices and quaternions arises via:

$$\begin{aligned} i \leftrightarrow -\sqrt{-1}\sigma_1 &= -\sqrt{-1} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} -\sqrt{-1} & 0 \\ 0 & \sqrt{-1} \end{pmatrix} \\ j \leftrightarrow -\sqrt{-1}\sigma_2 &= -\sqrt{-1} \begin{pmatrix} 0 & \sqrt{-1} \\ -\sqrt{-1} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ k \leftrightarrow -\sqrt{-1}\sigma_3 &= -\sqrt{-1} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\sqrt{-1} \\ -\sqrt{-1} & 0 \end{pmatrix} \\ 1 \leftrightarrow 1 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ \therefore a + bi + cj + dk \leftrightarrow &\begin{pmatrix} a - b\sqrt{-1} & c - d\sqrt{-1} \\ -c - d\sqrt{-1} & a + b\sqrt{-1} \end{pmatrix} = \begin{pmatrix} Z & W \\ -\bar{W} & \bar{Z} \end{pmatrix} \end{aligned}$$

With this identification, we see that the **unit quaternions** are isomorphic to

$$SU(2) = \left\{ \begin{pmatrix} Z & W \\ -\bar{W} & \bar{Z} \end{pmatrix} \mid \begin{array}{l} Z\bar{Z} + W\bar{W} = 1 \\ Z, W \text{ complex numbers} \end{array} \right\}.$$

The upshot of this approach is that we can represent spacetime points quaternionically. It is convenient, for this, to change conventions and use

$$\sqrt{-1}T + iX + jY + kZ = p.$$

Then

$$\begin{aligned} p\bar{p} &= (\sqrt{-1}T + iX + jY + kZ)(\sqrt{-1}T - iX - jY - kZ) \\ &= -T^2 + X^2 + Y^2 + Z^2. \end{aligned}$$

Note that we are here working in complexified quaternions with $\sqrt{-1}$ commuting with i, j and k .

Since $p\bar{p}$ represents the Lorentz metric, we can represent Lorentz transformations quaternionically via

$T(p) = gpg$ where $g = t + \sqrt{-1}v$, $\|v\|^2 - t^2 = 1$

$$\begin{aligned}
T(p)\overline{T(p)} &= qpg\overline{(gpg)} \quad \boxed{g\bar{g} = 1} \\
&= gpg \bar{g} \bar{p} \bar{g} \\
&= gp(\sqrt{-1}t + v)(\sqrt{-1}t - v)\bar{p} \bar{g} \\
&= gp(-t^2 + \|v\|^2)\bar{p} \bar{g} \\
&= gp \bar{p} \bar{g} \\
&= (p\bar{p})g\bar{g} \\
\therefore T(p)\overline{T(p)} &= p\bar{p}.
\end{aligned}$$

We must verify here under what circumstances $T(p)$ is a point in spacetime.

Note that

$$g = t + \sqrt{-1}v = t + v_1\sigma_1 + v_2\sigma_2 + v_3\sigma_3$$

and if $g^2 = (t^2 + v \cdot v) + \sqrt{-1}(2tv)$ actually represents our Lorentz boost, then gpg will perform the boost for $p = \sqrt{-1}t + q$, q in the direction v . If $q \perp v$ then $gqg = q$. Compare this discussion with the treatment of reflections in the next section.

10⁰. Quaternions, Cayley Numbers and the Belt Trick.

First we will develop the quaternions by reflecting on reflections. Then come string tricks, and interlacements among the topological properties of braids and strings, and algebras.

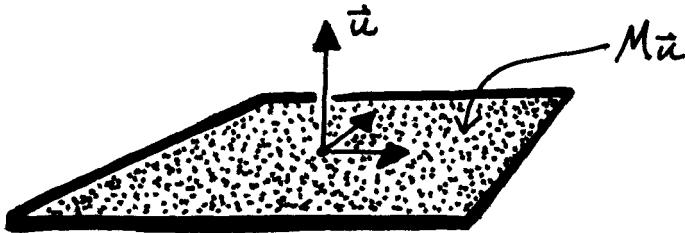
So we begin with the algebra of reflections in \mathbf{R}^3 . Let $\vec{u} \in \mathbf{R}^3$ be a unit vector. By \mathbf{R}^3 we mean Euclidean three-dimensional space. Thus

$$\mathbf{R}^3 = \{(x, y, z) | x, y, z \text{ real numbers}\}$$

$$\vec{u} = (u_1, u_2, u_3)$$

$$\vec{u} \text{ a unit vector means } \|\vec{u}\| = \sqrt{u_1^2 + u_2^2 + u_3^2} = 1 .$$

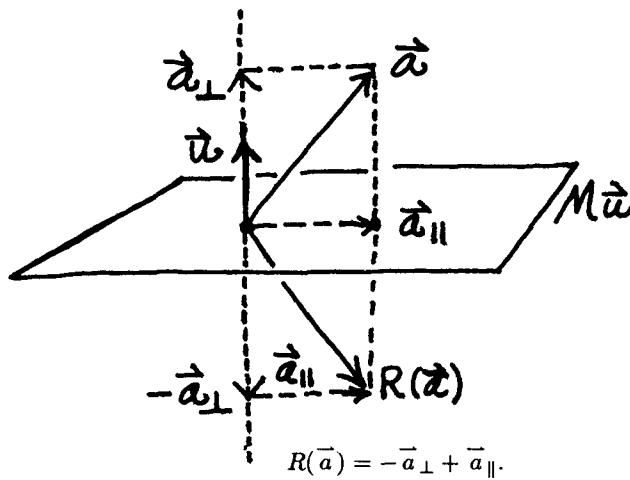
We have the inner product $\vec{a} \cdot \vec{b} = a_1 b_1 + a_2 b_2 + a_3 b_3$. Each unit vector \vec{u} determines a plane $M\vec{u}$ that is perpendicular to it:



In order to use $M\vec{u}$ as a mirror, that is in order to reflect a vector \vec{a} with respect to it, we first write $\vec{a} = \vec{a}_\perp + \vec{a}_\parallel$ where \vec{a}_\perp is perpendicular to $M\vec{u}$ and \vec{a}_\parallel is parallel to $M\vec{u}$. Then

$$R(\vec{a}) = -\vec{a}_\perp + \vec{a}_\parallel$$

accomplishes the reflection.



And we also see that

$$\vec{a}_{\perp} = (\vec{a} \cdot \vec{u})\vec{u}$$

(since \vec{u} is a unit vector). Thus

$$\vec{a}_{\parallel} = \vec{a} - (\vec{a} \cdot \vec{u})\vec{u}$$

and $R(\vec{a}) = \vec{a} - 2(\vec{a} \cdot \vec{u})\vec{u}$. This gives a specific formula for reflection in the plane perpendicular to the unit vector \vec{u} .

We wish to represent reflections by an algebra structure on the points of three-space.

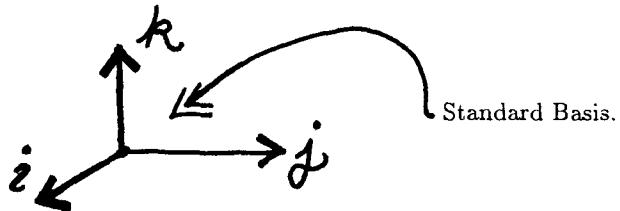
As geometry is translated into algebra the geometric roles (a little geometry in the algebra) become inverted. That which is mirrored in geometry becomes a (notational) mirror in algebra.

In symbols:

$$R(X) = uXu.$$

Here X (which is acted on by the mirror in geometry) becomes a notational

mirror, reflecting u on either side.



We want uXu to be the reflection of X in the plane perpendicular to u .

Therefore:

$$\begin{cases} kkk = -k \\ kjk = j \\ kik = i \end{cases}$$

Let's assume that this algebra is **associative** (i.e. $(ab)c = a(bc)$ for all a, b, c) and that **non-zero vectors have inverses**. Then multiplying $kkk = -k$ by k^{-1} we have

$$kk = -1$$

$$kj k = j$$

$$kik = i.$$

Therefore, multiplying the second two equations by k , we have

$$jk = -kj$$

$$ik = -ki.$$

The same argument can be applied to reflections in the planes perpendicular to i and j . Hence $i^2 = j^2 = k^2 = -1$, and, in fact, by the same argument $u^2 = -1$ for any unit vector u .

This may seem a bit surprising, since $-1 \notin \mathbf{R}^3$! In fact, we see that for this scheme to work, it is necessary to use $\mathbf{R}^4 = \{(x, y, z, t)\}$ with $(0, 0, 0, -1)$ representing the "scalar" -1 .

What about ij ? Is this forced by our scheme? With the given assumptions it

is not, however one way to fulfill the conditions is to use:

$$\begin{aligned} ij &= k \\ jk &= i \\ ki &= j \end{aligned}$$

the familiar vector-cross-product pattern. This algebra

$$i^2 = j^2 = k^2 = ijk = -1$$

is **Hamilton's Quaternions**. It produces an algebra structure on $\mathbf{H} = \mathbf{R}^4 = \{xi + yj + zk + t\}$ that is associative and so that non-zero elements have inverses. In fact, writing

$$q = t + ix + jy + kz$$

$$\bar{q} = t - ix - jy - kz$$

then $q\bar{q} = \bar{q}q = t^2 + x^2 + y^2 + z^2$.

An element of the form $ia + jb + kc \in \mathbf{R}^3$ is a **pure quaternion** (analogous to a pure imaginary complex number). Let

$$X = iX_1 + jX_2 + kX_3$$

$$Y = iY_1 + jY_2 + kY_3.$$

Then the quaternionic product is given by the formula

$$\begin{aligned} XY &= -X_1Y_1 - X_2Y_2 - X_3Y_3 \\ &\quad + \begin{vmatrix} Y_1 & X_2 \\ Y_1 & X_2 \end{vmatrix} k - \begin{vmatrix} X_1 & X_3 \\ Y_1 & Y_3 \end{vmatrix} j + \begin{vmatrix} X_2 & X_3 \\ Y_2 & Y_3 \end{vmatrix} i \\ \therefore XY &= -(X \cdot Y) + (X \times Y) \end{aligned}$$

where $X \times Y$ is the vector cross product in \mathbf{R}^3 .

Note that, as we have manufactured them, there is a 2-dimensional sphere's worth of pure quaternionic square roots of -1 . That is, if $U \in \mathbf{R}^3$ then

$$UU = -U \cdot U + U \times U$$

$$\therefore U^2 = -U \cdot U$$

$$\therefore U^2 = -1 \Leftrightarrow U \cdot U = 1$$

$$\therefore U^2 = -1 \Leftrightarrow U \in S^2.$$

Here S^2 denotes the set of points in \mathbf{R}^3 at unit distance from the origin.

Any quaternion is of the form $a + bu$ where $a, b \in \mathbf{R}$ and $u \in S^2$. This gives \mathbf{H} the aspect to a kind of “spun” complex numbers since

$\mathbf{H}_u = \{a + bu | a, b \in \mathbf{R}, u \text{ fixed}\}$ is isomorphic to $\mathbf{C} = \{a + bi | a, b \in \mathbf{R}, i = \sqrt{-1}\}$. Multiplication within \mathbf{H}_u is ordinary complex multiplication, but multiplication between different \mathbf{H}_u depends upon the quaternionic multiplication formulas.

Another important property of quaternionic multiplication is that it is **norm-preserving**. That is, $\|qq'\| = \|q\| \|q'\|$ where $\|q\| = q\bar{q}$. This is, of course easy to see since

$$\begin{aligned}\|qq'\| &= (qq')\overline{(qq')} \\ &= (qq')(\bar{q'}\bar{q}) \\ &= q(q'\bar{q'})\bar{q} \\ &= (q\bar{q})(q'\bar{q'}) \\ \therefore \|qq'\| &= \|q\| \|q'\|.\end{aligned}$$

Here we have used an easily verified fact: $\overline{qq'} = \bar{q'}\bar{q}$ and the associativity of the multiplication.

In fact one way of thinking about how the quaternions were produced is that we started with the complex numbers $\mathbf{C} = \{a + bi\}$ and we tried to add a new imaginary j ($j^2 = -1$) so that $jZ = \bar{Z}j$, ($Z = a + ib, \bar{Z} = a - ib$) and so that the multiplication was associative. (Then $k = ij$ and $k^2 = (ij)(ij) = i(ji)j = i(\bar{i}j)j = i(-ij)j = (i(-i))(jj) = (+1)(-1) \therefore k^2 = -1$, recovering the quaternions.)

That this procedure treads on dangerous ground is illustrated when we try to repeat it! Try to add an element J to \mathbf{H} so that the new algebra is associative, $JA = \bar{A}J \forall A \in \mathbf{H}$, and J is invertible. Then

$$\begin{aligned}
J(AB) &= \overline{(AB)}J \\
&= (\overline{B} \overline{A})J \\
&= \overline{B}(\overline{A}J) \\
&= \overline{B}(J\bar{\bar{A}}) \\
&= \overline{B}(JA) \\
&= (\overline{B}J)A \\
&= (J\bar{\bar{B}})A \\
&= (JB)A \\
J(AB) &= J(BA) \\
J^{-1}(J(AB)) &= J^{-1}(J(BA)) \\
(J^{-1}J)(AB) &= (J^{-1}J)(BA) \\
\therefore AB &= BA.
\end{aligned}$$

Since this is to hold for any quaternions A and B and since $AB \neq BA$ in general in \mathbf{H} we see that **associativity is too much to demand of this extension of the quaternions.**

There does exist an algebra structure on eight-dimensional space $\mathbf{H} + J\mathbf{H}$ such that $J^2 = -1$, $JZ = \bar{Z}J$, called the **Cayley numbers**. It is necessarily non-associative. How can one determine the Cayley multiplication?? We now initiate a Cayley digression.

Cayley Digression.

First a little abstract work. Let's assume that we are working with a normed algebra \mathcal{A} . We are given Euclidean space \mathbf{R}^n having a multiplicative structure $\mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}^n$ and a notion of conjugation $Z \mapsto \bar{Z}$ so that $\bar{\bar{Z}} = Z$. (For \mathbf{R} itself, $\bar{r} = r$.) Assume that the Euclidean norm is given by $\|Z\| = Z\bar{Z}$ and that **the multiplication is norm-preserving**

$$\|ZW\| = \|Z\| \|W\|.$$

(This no longer follows from $\|Z\| = Z\bar{Z}$ since we can't assume associativity.)

Define a bilinear form

$$[Z, W] = \frac{1}{2}(Z\bar{W} + \bar{Z}W).$$

(We assume that $\overline{Z + W} = \bar{Z} + \bar{W}$ also.) And finally assume that the norm on the algebra is non-degenerate in the sense that

$$\boxed{\begin{array}{rcl} [Z, X] & = & 0 \quad \forall X \\ \Rightarrow Z & = & 0 \end{array}}.$$

Lemma 10.1. $2[Z, W] = \|Z + W\| - \|Z\| - \|W\|$.

Proof. Easy.

Proposition 10.2. (J. Conway [CON1]).

- (1) $[ac, bc] = [a, b][c]$
- (2) $[ac, bd] + [ad, bc] = 2[a, b][c, d]$
- (3) $[ac, b] = [a, b\bar{c}]$.

Here a, b, c, d are any elements of the algebra \mathcal{A} and $[a] = \|a\|$.

Proof.

$$\begin{aligned} [(a + b)c] &= [ac + bc] = [ac] + [bc] + 2[ac, bc] \\ &\quad \| \\ [a + b][c] &= ([a] + [b] + 2[a, b])[c] \\ \Rightarrow [ac, bc] &= [a, b][c], \text{ proving (1).} \end{aligned}$$

To prove (2) replace c by $(c + d)$ in (1):

$$\begin{aligned} [a, b][c + d] &= [a(c + d), b(c + d)] = [ac + ad, bc + bd] \\ &\quad \| \\ [a, b]([c] + [d] + 2[c, d]). \end{aligned}$$

Continue expanding and collect terms. For (3) note that $\bar{c} = 2[c, 1] - c$, $(2[c, 1] - c = (c \cdot \bar{1} + \bar{c} \cdot 1) - c = \bar{c})$ and let $d = 1$ in (2). Then

$$\begin{aligned} [ac, b] &= 2[a, b][c, 1] - [a, bc] \\ &= [a, b(2[c, 1] - c)] \\ &= [a, b\bar{c}]. \end{aligned}$$

This completes the proof of the proposition.

Conway's proposition shows that we can use the properties

$$\begin{aligned}
 [ab] &= [a][b] \\
 [a, a] &= [a] \\
 [a, b][c] &= [ac, bc] \\
 2[a, b][c, d] &= [ac, bd] + [ad, bc] \\
 [ac, b] &= [a, b\bar{c}] \\
 ([a, bc] &= [a\bar{c}, b], [ac, b] = [c, \bar{a}b], \text{etc.})
 \end{aligned}$$

plus the non-degeneracy

$$[Z, t] = [W, t] \quad \forall t \Rightarrow Z = W$$

to ferret out the particular properties of the multiplication. Just watch!

$$(1) \quad \bar{\bar{x}} = x$$

Proof.

$$\begin{aligned}
 [x, t] &= [1, t\bar{x}] \\
 &= [1 \cdot \bar{\bar{x}}, t] \\
 \therefore [x, t] &= [\bar{\bar{x}}, t] \\
 \therefore x &= \bar{\bar{x}}.
 \end{aligned}$$

//

$$(2) \quad \bar{a}\bar{b} = \bar{b}\bar{a}$$

Proof.

$$\begin{aligned}
 [ab, t] &= [a, t\bar{b}] \\
 &= [\bar{t}a, \bar{b}] \\
 \therefore [ab, t] &= [\bar{t}, \bar{b}\bar{a}] \\
 [ab, t] &= [1(ab), t] \\
 &= [1, t(\bar{a}\bar{b})] = [\bar{t}, \bar{a}\bar{b}] \\
 \therefore [\bar{t}, \bar{b}\bar{a}] &= [\bar{t}, \bar{a}\bar{b}] \\
 \therefore \bar{b}\bar{a} &= \bar{a}\bar{b}
 \end{aligned}$$

//

Now suppose that H is a subalgebra on which $[,]$ is non-degenerate, and that there exists $J \perp H$ (\perp means $[h, J] = 0 \quad \forall h \in H$ whence $h\bar{J} + J\bar{h} = 0$ whence (assuming for a moment $\bar{J} = -J$) $J\bar{h} = hJ$. This was our previous extension idea. Here it is embodied geometrically as perpendicularity.). We assume $[J] = 1$.

Theorem 10.3. Under the above conditions $H + JH$ is a subalgebra of the given algebra \mathcal{A} , and $H \perp JH$. Furthermore, if $a, b, c, d \in H$ then

$$(a + Jb)(c + Jd) = (ac - d\bar{b}) + J(cb + \bar{a}d).$$

Remark. If H is \mathbf{H} , the quaternions, then this specifies the Cayley multiplication.

Proof. Note $\bar{J} = 2[J, 1] - J = -J$ since $J \perp H \quad (1 \in H)$.

$$\begin{aligned} [a, Jb] &= [a\bar{b}, J] = 0 \quad \forall a, b \in H \\ \therefore H &\perp JH. \end{aligned}$$

Also

$$\begin{aligned} [Ja, Jb] &= [J][a, b] = [a, b] \\ \therefore JH &\text{ is isometric to } H. \end{aligned}$$

Claim 1. $aJ = J\bar{a} \quad \forall a \in H$

Proof.

$$\begin{aligned} [aJ, t] &= [J, \bar{a}t] \\ &= 2[J, \bar{a}][1, t] - [Jt, \bar{a}] \\ ([ac, bd] &= 2[a, b][c, d] - [ad, bc]) \\ \therefore [aJ, t] &= -[Jt, \bar{a}] \\ &= -[t, \bar{J}\bar{a}] \\ &= [t, J\bar{a}] \\ &= [J\bar{a}, t]. \end{aligned}$$

Claim 2. $(Jb)c = J(cb)$

Proof.

$$\begin{aligned}
 [(Jb)c, t] &= [Jb, t\bar{c}] \\
 &= [\bar{b}J, t\bar{c}] \\
 &= -[\bar{b}\bar{c}, tJ] \quad (\text{as above}) \\
 &= [(\bar{b}\bar{c})J, t] \\
 \therefore (Jb)c &= (\bar{b}\bar{c})J \\
 &= J(\overline{\bar{b}\bar{c}}) \\
 \therefore (Jb)c &= J(cb).
 \end{aligned}$$

//

Claim 3. $a(Jb) = J(\bar{a}b)$

Proof.

$$\begin{aligned}
 [a(Jb), t] &= [Jb, \bar{a}t] \\
 &= -[Jt, \bar{a}b] \quad (\text{by perpendicularity}) \\
 &= [t, J(\bar{a}b)] \\
 \therefore a(Jb) &= J(\bar{a}b).
 \end{aligned}$$

//

Claim 4. $(Ja)(Jb) = -b\bar{a}$.

Proof.

$$\begin{aligned}
 [(Ja)(Jb), t] &= [Jb, (\overline{Ja})t] \\
 &= [\bar{b}J, (\overline{Ja})t] \\
 &= -[\bar{b}t, (\overline{Ja})J] \\
 &= -[(\bar{b}t)\bar{J}, \overline{Ja}] \\
 &= -[(\bar{b}t)\bar{J}, \bar{a}\bar{J}] \\
 &= -[\bar{b}t, \bar{a}]\bar{J} \\
 &= -[t, b\bar{a}]
 \end{aligned}$$

$$\therefore (Ja)(Jb) = -b\bar{a}.$$

This completes the proof. //

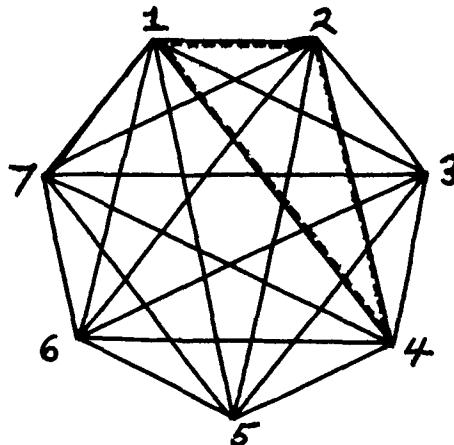
Applying this construction iteratively, beginning with the reals \mathbf{R} , we find

$$\begin{array}{ccccccc} \mathbf{R} & \hookrightarrow & \mathbf{C} & \hookrightarrow & \mathbf{H} & \hookrightarrow & \mathbf{K} \\ \text{reals} & & \text{complexes} & & \text{quaternions} & & \text{Cayley numbers} \end{array}$$

At each stage some element of abstract structure is lost. Order goes on the passage from \mathbf{R} to \mathbf{C} . Commutativity disappears on going from \mathbf{C} to \mathbf{H} . The Cayley numbers \mathbf{K} lose associativity. Finally, the process of the theorem, when applied to \mathbf{K} itself simply collapses. Thus $\mathbf{R} \subset \mathbf{C} \subset \mathbf{H} \subset \mathbf{K}$ are the only non-degenerate normed algebras. (This was originally proved in [HU]).

The situation is deeper than that. One can ask for division algebra (not necessarily normed) structures on Euclidean spaces, and again the only dimensions are 1, 2, 4 and 8. But now the result requires deep algebraic topology (see [A1] for an account).

Exercise.



Show that there is a basis for \mathbf{K} , $\{1, e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$, with

$$e_i^2 = -1$$

$$e_i e_j = -e_j e_i \quad i \neq j$$

and

$$\begin{array}{ccccc} \text{1} & \curvearrowright & \text{2} & \curvearrowright & \text{3} \\ e_4 \leftarrow e_2 = e_4 & & e_5 \leftarrow e_3 = e_5 & & e_6 \leftarrow e_4 = e_6 \\ \text{4} & \curvearrowright & \text{5} & \curvearrowright & \text{6} \\ e_7 \leftarrow e_5 = e_7 & & e_1 \leftarrow e_7 = e_1 & & \dots \\ \text{7} & \curvearrowright & \text{1} & \curvearrowright & \text{3} \\ e_3 \leftarrow e_1 = e_3 & & & & \end{array}$$

$$\boxed{e_i e_{i+1} = e_{i+3}} \text{ (indices modulo 7)}$$

This ends our digression into the Cayley numbers.

Rotations and Reflections.

So far we have used the quaternions to represent reflections. If $R_u : \mathbf{R}^3 \rightarrow \mathbf{R}^3$ is the reflection in the plane perpendicular to the unit vector $u \in S^2$, then

$$R(p) = R_u(p) = upu$$

is the quaternion formula for the reflection.

We have verified this for the frame $\{i, j, k\}$. To see it in general, let $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ be a right-handed frame of orthogonal unit vectors in \mathbf{R}^3 . Then we know (using $\vec{u}_i \cdot \vec{u}_j = 0$ $i \neq j$, $\vec{u}_i \cdot \vec{u}_i = 1$, $\vec{u}_1 \times \vec{u}_2 = \vec{u}_3$ etc...) that $\vec{u}_1^2 = \vec{u}_2^2 = \vec{u}_3^2 = \vec{u}_1 \vec{u}_2 \vec{u}_3 = -1$ as quaternions. Hence the formula will hold in any frame, hence for any unit vector u .

Recall that we also know that the reflection is given by the direct formula

$$R(p) = p - 2(u \cdot p)u,$$

as shown at the beginning of this section. Therefore

$$\begin{aligned} p - 2(u \cdot p)u &= upu \\ &= (-u \cdot p + u \times p)u \\ &= -(u \cdot p)u - (u \times p) \cdot u + (u \times p) \times u \\ \therefore p - 2(u \cdot p)u &= -(u \cdot p)u + (u \times p) \times u \\ \therefore \boxed{p - (u \cdot p)u = (u \times p) \times u}. \end{aligned}$$

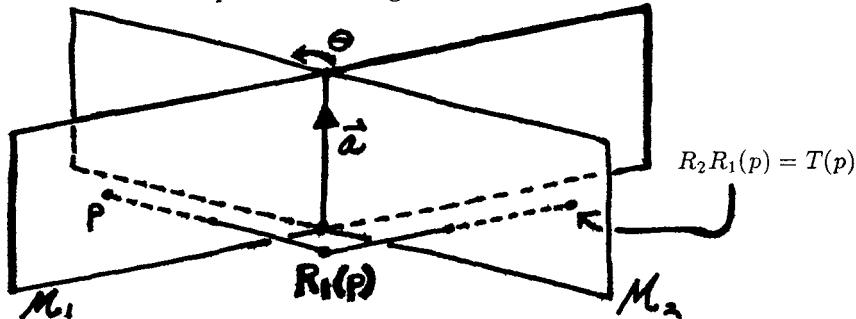
This is a “proof by reflection” of this well-known formula about the triple vector product.

Remark. It is important to note that the vector cross product is a **non-associative** multiplication. Thus

$$\begin{aligned} i \times (i \times j) &= i \times k = -j \\ (i \times i) \times j &= 0 \times j = 0. \end{aligned}$$

The associativity of the quaternion algebra gives rise to specific relations about vector cross products.

We now turn from reflections to rotations. Every rotation of three dimensional space is the resultant of two reflections. If the planes M_1, M_2 make an angle θ with each other, then the rotation resulting from the composition of reflecting in M_1 and then in M_2 is a rotation of angle 2θ about the axis of intersection of the two planes. View Figure 10.1.



$T = R_2R_1$ is rotation about \vec{a} by 2θ

Figure 10.1

Now let's examine the algebra of the product of two reflections. Let $u, v \in S^2$. Then the vector $u \times v$ is the axis of intersection of the planes M_u, M_v of vectors perpendicular to u and to v respectively. Thus we expect $Rv \circ Ru = T$ to be a rotation about $u \times v$ with angle 2θ for $\theta = (\text{angle between } u \text{ and } v) = \arccos(u \cdot v)$. And we have

$$\begin{aligned} T(p) &= Ru(Rv(p)) \\ &= u(vpv)u \\ T(p) &= (uv)p(vu). \end{aligned}$$

Let $g = uv = -(u \cdot v) + u \times v$ and note that $vu = -(v \cdot u) + v \times u = +\bar{uv} = \bar{g} = g^{-1}$. Thus $T(p) = gp\bar{g} = gpg^{-1}$ where

$$g = -\cos(\theta) - \sin(\theta)w,$$

$w = (u \times v)/\|u \times v\|$. The global minus sign is extraneous since

$$gpg^{-1} = (-g)p(-g)^{-1}.$$

Hence

Proposition 10.4. Let $g = (\cos \theta) + (\sin \theta)u$ where $u \in S^2$ is a unit quaternion. Then $T(p) = gpg^{-1} = gp\bar{g}$ defines a mapping $T : \mathbf{R}^3 \rightarrow \mathbf{R}^3$ (\mathbf{R}^3 is the set of pure quaternions), and **T is a rotation with axis u and angle of rotation about u equal to 2θ .**

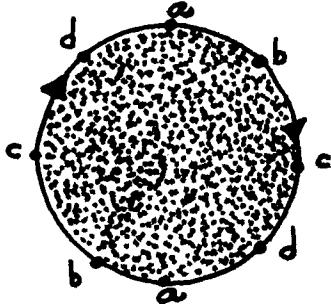
Put a bit more abstractly, we have shown that there is a **2 to 1 mapping** $\pi : S^3 \rightarrow SO(3)$ given by $\pi(g)(p) = gpg^{-1}$. Here S^3 represents the set of all unit length quaternions in $\mathbf{H} = \mathbf{R}^4$. This is the three-dimensional sphere

$$S^3 = \{ai + bj + ck + t | a^2 + b^2 + c^2 + t^2 = 1\}.$$

Any element of S^3 can be represented as $g = e^{u\theta} = (\cos \theta) + (\sin \theta)u$ with $u \in S^2 \subset \mathbf{R}^3$. Since $\|qq'\| = \|q\| \|q'\|$ for any quaternions $q, q' \in \mathbf{H}$, we see that S^3 is closed under multiplication. Hence S^3 is a group. This is the Lie group structure on S^3 . $SO(3)$ is the group of orientation preserving orthogonal linear transformations of \mathbf{R}^3 – hence the group of rotations of \mathbf{R}^3 .

The mapping $\pi : S^3 \rightarrow SO(3)$ is $2 \rightarrow 1$ since $\pi(g) = \pi(-g)$. We conclude that, **topologically, $SO(3)$ is the result of identifying antipodal points in S^3 .** In topologists parlance this says that $SO(3)$ is homeomorphic to \mathbf{RP}^3 = the real projective three space = the space of lines through the origin in $\mathbf{R}^4 \cong$ the three dimensional sphere S^3 modulo antipodal identifications.

Remark. In general, \mathbf{RP}^n is the space of lines in \mathbf{R}^{n+1} (through the origin) and there is a double covering $\pi : S^n \rightarrow \mathbf{RP}^n$. Historically, the term projective space comes from projective geometry. Thus $\mathbf{RP}^2 = S^2 /$ (antipodal identifications) is homeomorphic to a disk D^2 whose boundary points are antipodally identified.



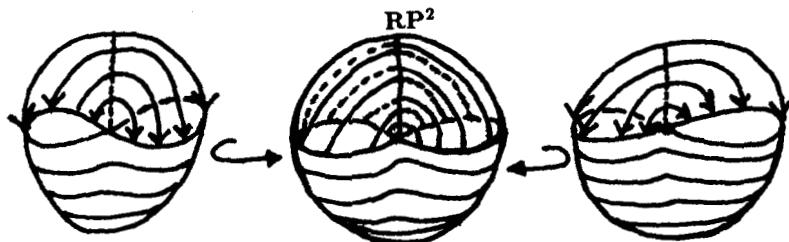
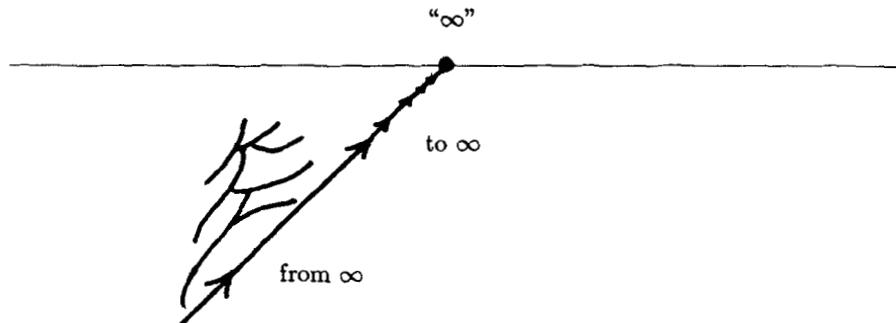


Figure 10.2

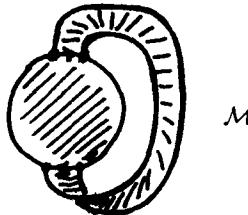
This is a strange space, but arises naturally in projective geometry as the **completion of a plane by adding a circle of points at infinity** so that travel outward along any ray through the origin will go through ∞ and return along the negative ray.



Exercise. Show directly that $SO(3) \cong D^2 / \sim$ where D^3 represents a three-dimensional ball, and \sim denotes antipodal identification of boundary points. [HINT: Let D^3 have radius π . Associate to $\vec{v} \in D^3$ a rotation (right-handed) about \vec{v} of $\|\vec{v}\|$ radians.]

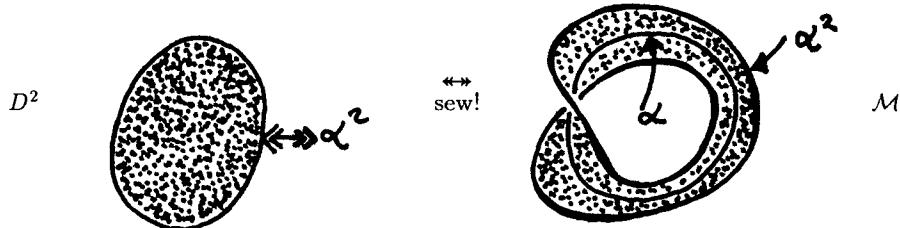
Let's return for a moment to the projective plane \mathbf{RP}^2 . This is a closed surface, like the surface of a sphere or of a torus, but (it can be proved) there is no way to embed this surface into three dimensional space without introducing

singularities. If we start to sew the boundary of a disk to itself by antipodal identifications, some of the difficulties show themselves:



Begin the process by identifying two intervals results in a Möbius strip \mathcal{M} as above. In the projective plane, if you go through infinity and come back, you come back with a twist! In projective three-space, the same process will turn your handedness. This consideration is probably the origin of the Möbius band. The Möbius band is a creature captured from infinity, and brought to live in bounded space.

To complete \mathcal{M} to \mathbf{RP}^2 , all you must do is sew a disk along its boundary circle to the single boundary component of the Möbius \mathcal{M} . (Figure 10.3).



Instructions for producing \mathbf{RP}^3

Figure 10.3

This description, $\mathbf{RP}^3 = \mathcal{M} \cup D^2$, gives us a very good picture of the topology of \mathbf{RP}^3 . In particular it lets us spot a loop α on \mathbf{RP}^2 that can't be contracted, but $\alpha^2 = \text{"the result of going around on } \alpha \text{ twice"}$ is contractible. The loop α is the core of the band \mathcal{M} . α^2 is the boundary of \mathcal{M} (you go around \mathcal{M} twice when traversing the boundary). The curve α^2 is contractible, because it is identified with the boundary of the sewing disk D^2 . (α is a generator of $\pi_1(\mathbf{RP}^2)$ and $\pi_1(\mathbf{RP}^2) \cong \mathbf{Z}_2$.)

The same effect happens in \mathbf{RP}^3 , and hence in $SO(3)$. There is a loop α on $SO(3)$ that is not contractible, but α^2 is contractible. This loop can be taken to be $\alpha(t) = r(2\pi t, \vec{n})$ where $r(\theta, \vec{n})$ is a rotation by angle θ about the axis \vec{n} (call it the north pole). Note that each $\alpha(t)$ is a rotation and that $\alpha(0) = \alpha(1) =$ identity.

Given this loop α of rotations we can examine its geometry and topology by creating an image of the action of $\alpha(t)$ as t ranges from 0 to 1. To accomplish this, let $\hat{\alpha} : S^2 \times [0, 1] \rightarrow S^2 \times [0, 1]$ be defined by

$$\hat{\alpha}(p, t) = (\alpha(t)p, t).$$

If you visualize $S^2 \times [0, 1]$ as a 3-dimensional annulus with a large S^2 (2-dimensional sphere) on the outside, and a smaller S^2 on the inside (See Figure 10.4.) then I can illustrate

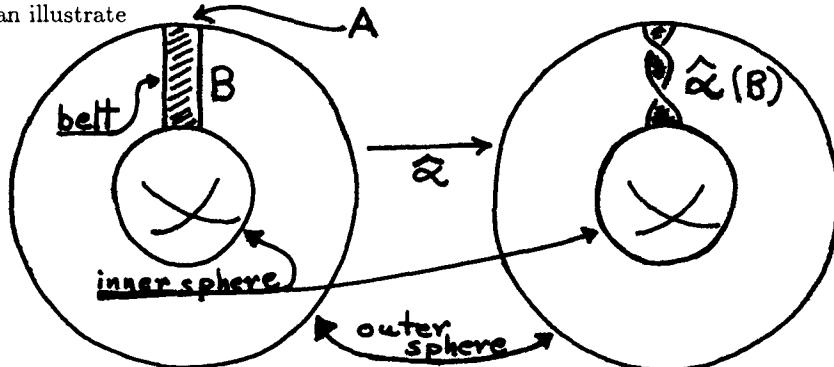


Figure 10.4

the action of $\hat{\alpha} : S^2 \times [0, 1] \rightarrow S^2 \times [0, 1]$ by considering the image of a band $B \subset S^2 \times [0, 1]$ so that $B = A \times [0, 1]$ where A is an arc on S^2 passing through the north pole. Since $\alpha(t)$ is a rotation about the north pole by $2\pi t$, this band receives a 2π twist under $\hat{\alpha}$. And the fact that $\hat{\alpha}^2$ can be deformed to the identity map implies that $\hat{\alpha}^2(B)$ (= band with 4π -twist) can be deformed in $S^2 \times I$, keeping the ends of the band fixed, so that all twist is removed from the band.

This is a “theoretical prediction” based on the topological structure of $SO(3)$. In fact our prediction can be “experimentally verified” as shown in Figure 10.5.

This is known to topologists as the **belt trick** and to physicists as the **Dirac string trick**.

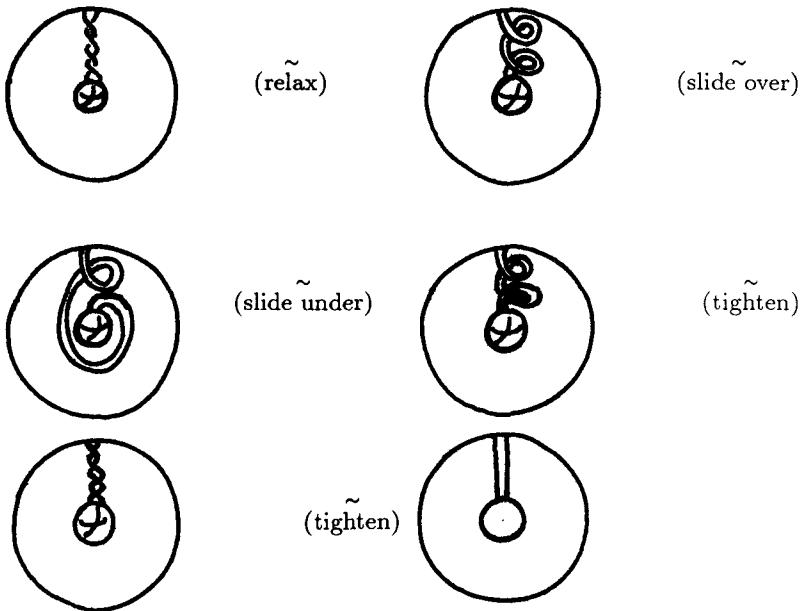


Figure 10.5

Another way of putting the string trick is this. Imagine a sphere suspended in the middle of a room attached by many strings to the floor, ceiling and walls. Turning the ball by 2π will entangle the strings, and turning it by 4π will apparently entangle them further. However with the 4π turn, all the strings can be disentangled without removing their moorings on the walls or the sphere, and without moving the sphere!

A particle with spin is something like this ball attached to its surroundings by string. It's amplitude changes under a 2π rotation, and is restored under the 4π rotation. Compare with Feynman [FE1].

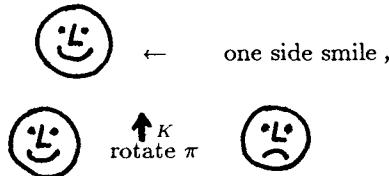
It is interesting to note that the structure of the quaternion group

$$i^2 = j^2 = k^2 = ijk = -1$$

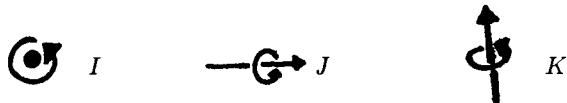
can be visualized by adding strings to a disk-representation of the Klein 4-group

$$I^2 = J^2 = K^2 = 1, \quad IJ = JI = K.$$

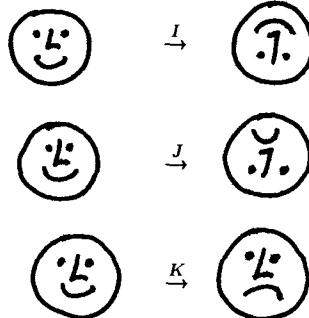
To see this let's take a disk



and one side frown. Let I denote a rotation of π about an axis perpendicular to this page. Let J be a π rotation about a horizontal axis and K a π rotation about a vertical axis.



Then, applying these operations to the disk we have



and it is easy to see that $I^2 = J^2 = K^2 = 1$, $IJ = JI = K$ in the sense that these equalities mean identical results for the disk.

If we now add strings to the disk, creating a puppet, then a 2π rotation leaves tangled strings. Let i, j, k denote the same spatial rotations as applied to our puppet.

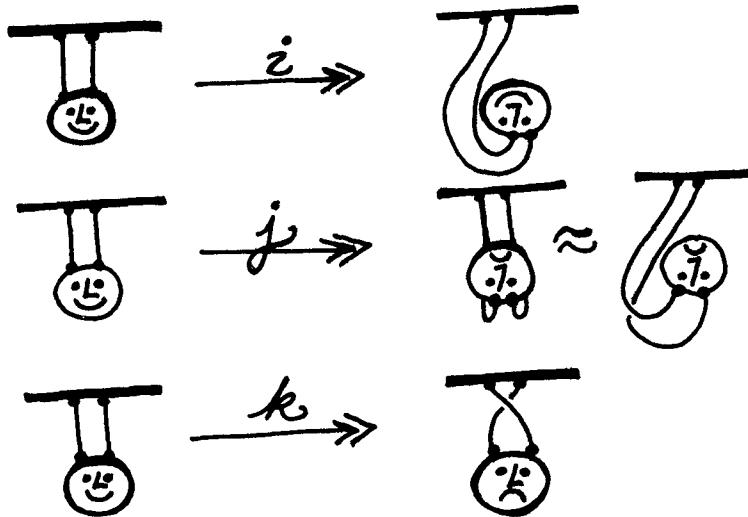
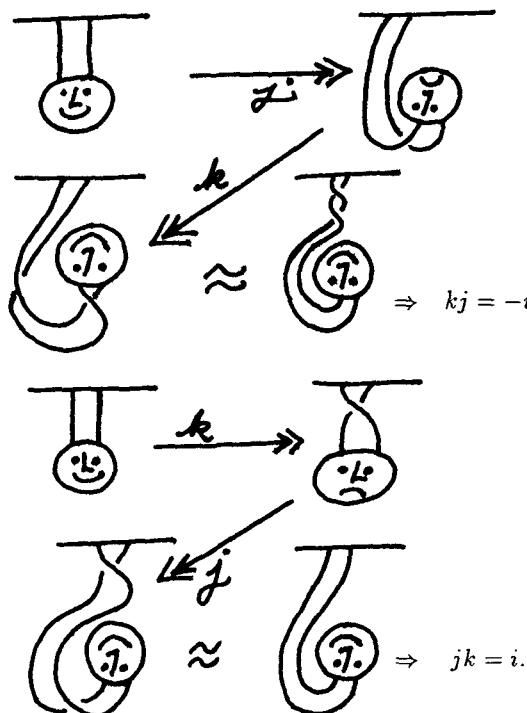


Figure 10.6

As we see, k^2 is no longer 1 since there is a 2π twist on the strings. The same goes for i^2, j^2 . Thus we may identify a $(\pm 2\pi)$ twist on the strings as -1 and set $i^2 = j^2 = k^2 = -1$ since by the belt trick $(i^2)^2 = (j^2)^2 = (k^2)^2 = 1$.

If a $\pm 2\pi$ twist is -1 , then we can also identify a $\pm \pi$ twist with $\pm \sqrt{-1}$! And this is being done with all three entities i, j, k . Note that $\bar{i} = -i$, $\bar{j} = -j$, $\bar{k} = -k$ are all obtained either by reversing the rotation, switching crossings, or adding a $\pm 2\pi$ twist on the strings.

The other quaternion relations are present as well:



Thus the puppet actually reproduces the quaternion relations. In its way this diagrammatic situation is analogous to our diagram algebra where a different string trick produces the relations in the Temperley-Lieb algebra.

Another comment about possible extensions of the string trick arises in connection with the Cayley numbers. As we have seen, the Cayley multiplicative structure is given by starting with $i, j, k, -1$ and adding J so that $Jz = \bar{z}J$, $J^2 = -1$. **Is there a string trick interpretation of the Cayley numbers?**

Return to Proposition 10.4.

Let's return to Proposition 10.4. This shows how to represent rotations of $\mathbf{R}^3 = \{xi + yj + zk | x, y, z \text{ real}\}$ via quaternions. For $g = e^{u\theta} = (\cos \theta) + (\sin \theta)u$, u a unit length quaternion in \mathbf{R}^3 , the mapping $T : \mathbf{R}^3 \rightarrow \mathbf{R}^3$, $T(p) = gp\bar{g}$, is a rotation about the axis u by 2θ . One of the charms of this representation is that it lets us compute the axis and rotation angle of the composite of two rotations. For, given

two rotations $g = e^{u\theta}$ and $g' = e^{u'\theta'}$. Then $T_{g'}(T_g(p)) = g'gp\bar{g}g' = (g'g)p\bar{(g'g)}$. Thus $T_{g'} \circ T_g = T_{g'g}$.

This means that if $g'g$ is rewritten into the form $e^{u''\theta''}$ then u'' will be the axis of the composite, and $2\theta''$ will be the angle of rotation. For example,

$$g = e^{k\pi/4} = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} k$$

is a rotation of $\pi/2$ about the k axis. And

$$g' = e^{j\pi/4} = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} j$$

corresponds to a $\pi/2$ rotation about the j -axis. Thus

$$\begin{aligned} g'g &= \left(\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} j\right) \left(\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} k\right) \\ &= \frac{1}{2}(1+j)(1+k) \\ &= \frac{1}{2}(1+j+k+jk) \\ &= \frac{1}{2}(1+j+k+i) \\ \therefore g'g &= \frac{1}{2} + \frac{\sqrt{3}}{2} \left(\frac{i+j+k}{\sqrt{3}} \right) \\ \therefore g'g &= e^{((i+j+k)/\sqrt{3})\pi/3}. \end{aligned}$$

Hence the composite of two $\pi/2$ rotations about the orthogonal axes j and k is a rotation of $2\pi/3$ around the diagonal axis. Figure 10.7 illustrates this directly via a drawing of a rotated cube.

This ends our chapter on the quaternions, Cayley numbers and knot and string interpretations. We showed how these structures arise naturally via looking for a model of the algebraic structure of reflections in space. In this model, the geometry is mirrored in the algebra: the geometric point being mirrored becomes notational mirror for the representation of the mirror plane: $R(p) = upu$, u a unit

vector perpendicular to the mirror plane.

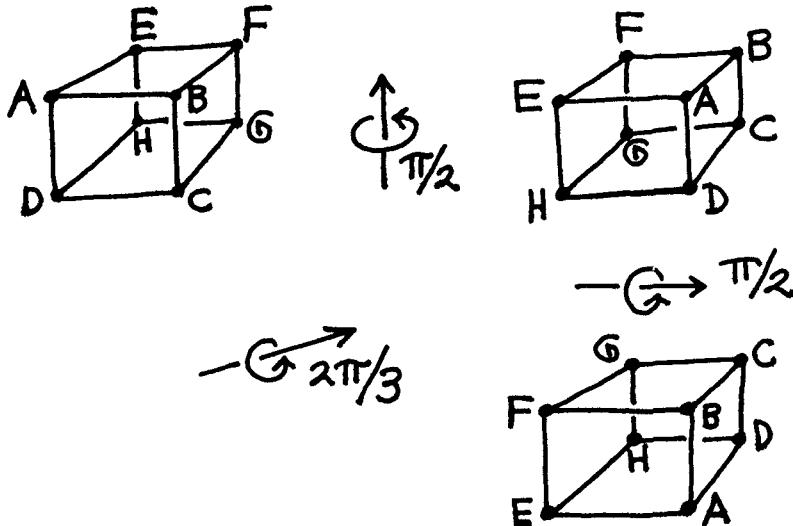


Figure 10.7

Epilog. This chapter has been a long excursion, from reflections to normed algebras, quaternions, string tricks and back to reflections. We have touched upon the quaternions and their role in spin and special relativity (section 9^o Part II). The role of the Cayley numbers in physics is an open question, but the quaternions, invented long before relativity, make a perfect fit with relativity, electromagnetism and the mathematics of spin.

Exercise. [SI]. Let $D = \sqrt{-1} \frac{\partial}{\partial t} + \vec{\nabla}$ where $\vec{\nabla} = i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z}$. Here i, j, k generate the quaternions, and $\sqrt{-1}$ is an extra root of -1 that commutes with i, j and k . Let $\psi = H + \sqrt{-1} E$ where $E = E_x i + E_y j + E_z k$, $H = H_x i + H_y j + H_z k$ are fields on \mathbf{R}^3 . (i) Show that the equation $D\psi = 0$, interpreted in complexified quaternions is identical to the vacuum Maxwell equations. (ii) Determine the non-vacuum form of Maxwell's equations in this language. (iii) Prove the relativistic

invariance of the Maxwell equations by using the quaternions. (iv) Investigate the nonlinear field equations ([HAY], [STEW]) $D\psi = \frac{1}{2}\psi\sqrt{-1}\psi^*$ where $\psi^* = H - \sqrt{-1}E$. (v) Let $K = K_1 + \sqrt{-1}K_2$ be a constant where K_1 and K_2 are linear combinations of i, j and k . Show that the equation $D\psi = K\psi$ is equivalent to the Dirac equation [DIR].

$$\Psi = \vec{H} + \sqrt{-1} \vec{E}$$

$$D = \sqrt{-1} \frac{\partial}{\partial t} + \nabla$$

$$D\Psi = \frac{1}{\sqrt{-1}} \rho + \vec{\sigma}$$

III

$$\nabla \cdot \vec{E} = \rho, \quad \nabla \cdot \vec{H} = 0$$

$$\frac{\partial \vec{H}}{\partial t} = -\nabla \times \vec{E}$$

$$\frac{\partial \vec{E}}{\partial t} + \vec{\sigma} = \nabla \times \vec{H}$$

11⁰. The Quaternion Demonstrator.

As we have discussed in section 10⁰ of Part II, The Dirac String Trick ([FE1], [BAT]) illustrates an elusive physical/mathematical property of the spin of an electron. An observer who travels around an electron (in a proverbial thought experiment) will find that the wave function of the electron has switched its phase for each full turn of the observer. See [AH] for a discussion of this effect. Here we repeat a bit of the lore of section 10⁰, and continue on to an application of the string trick as a quaternion demonstrator.

It is a fact of life in quantum mechanics that a rotation of 2π does not bring you back to where you started. Nevertheless, in the case of the spin of the electron (more generally for spin 1/2), a 4π rotation does return the observer to the original state. These facts, understandable in terms of the need for unitary representations of rotations of three space to the space of wave functions, seem very mysterious when stated directly in the language of motions in three dimensional space.

In order to visualize these matters, Dirac found a simple demonstration of a topological analog of this phenomenon using only strings, or a belt. As the **Belt Trick** this device is well-known to generations of topologists. In fact, a closely related motion, the **Plate Trick** (or Philippine Wine Dance) has been known for hundreds of years. (See [BE], p.122.)

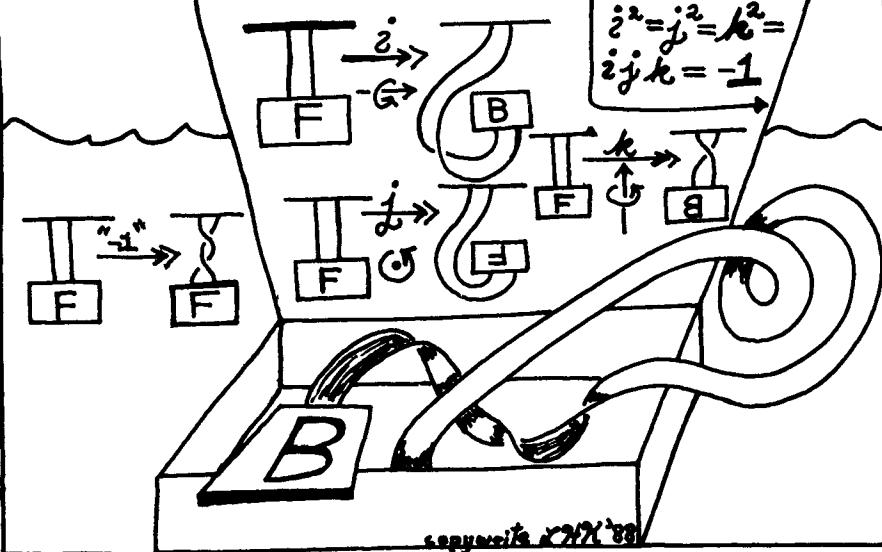
Note that the trick shows that a 4π twist in the belt can be removed by a topological ambient isotopy that leaves the ends of the belt fixed. We have illustrated the fixity of the ends of the belt by attaching them to two spheres. The belt moves in the complement of the spheres. This trick is easy to perform with a real belt, using an accomplice to hold one end, and one's own two hands to hold the other end as the belt is passed around one end.

It is also well-known that the belt trick illustrates the fact that the fundamental group of the space of orientation preserving rotations of three dimensional space ($SO(3)$) has order two. This is, in turn, equivalent to the fact that $SO(3)$ is double covered by the unit quaternions, $SU(2)$.

It is the purpose of this section to explore the belt trick and the group $SU(2)$. I will show how a natural generalization of the belt trick gives rise to a topological/mechanical model for the quaternion group, and how subgroups of $SU(2)$ are

The

Quaternion Demonstrator



Twist the card.

Then move the ribbon
Around the card

To simplify.

i Sometimes the card
 j can spin 'round and
 k 'round, while the ribbon
 j spontaneously unwinds.

< Brought to you by the number 3. >

visualized by attaching a belt to an object in three space. This generalizes the belt trick to a theorem, (the **Belt Theorem** - see below) about symmetry and $SU(2)$. The Belt Theorem is a new result.

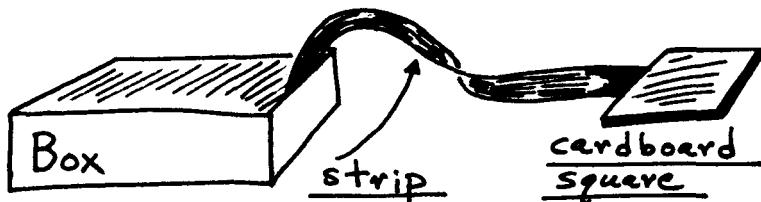
Finally, we show how to make a computer graphics model of the belt trick. This model is actually a direct visualization of the homotopy of the generator of $\pi_1(SO(3))$ with its inverse.

The Quaternion Demonstrator.

In order to create a mechanical/topological demonstration of the quaternion group, follow these instructions.

1. Obtain a cardboard box, a small square of cardboard and a strip of paper about a foot long and an inch wide.
2. Tape one end of the strip to the cardboard box, and tape the other end of the strip to the square of cardboard.

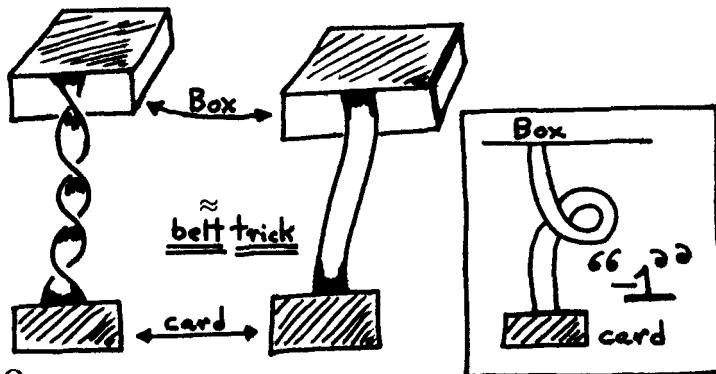
Your **demonstrator** should appear as shown below.



Before we construct the quaternions it is useful to do the belt trick and then to construct a square root of minus one. After a few square roots of minus one have appeared (one for each principal direction in space), we shall find that the quaternions have been created.

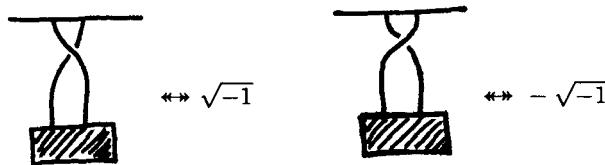
Belt Trick.

To demonstrate the belt trick, set the cardboard box of the demonstrator on a table, pick up the card, and begin with an untwisted belt (The paper strip shall be called the belt.). Twist the card by 720 degrees, and demonstrate the belt trick as we did in Section 10⁰.



Minus One.

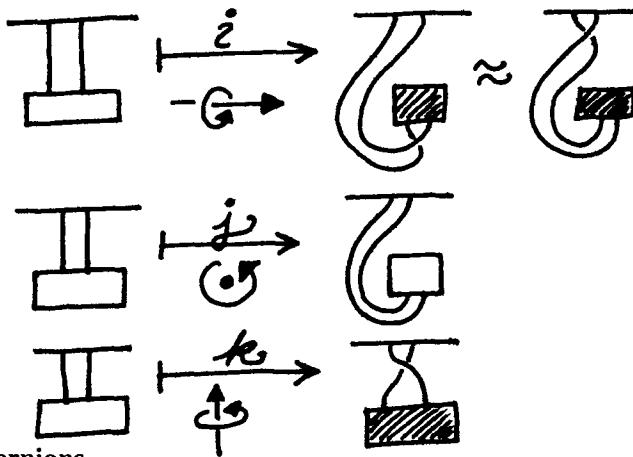
Since the belt trick disappears a twist of 720 degrees, it follows that a **twist of 180 degrees is the analog of the square root of minus one**. Four applications of the 180 degree twist bring you back to the beginning (with a little help from topology). If this is so, then a **twist of 360 degrees must be the analog of minus one**. Of course you may ask of the 360 degree twist - which one, right or left? But the belt trick shows that there is **only one 360 degree twist**. The left and right hand versions can be deformed into one another.



The presence of a 360 degree twist on the belt indicates multiplication by minus one.

Many Square Roots of Minus One.

There are many square roots of minus one. Just rotate the card around any axis by 180 degrees. Doing this four times will induce the belt trick and take you back to start. In particular, consider the rotations about three perpendicular axes as shown below.

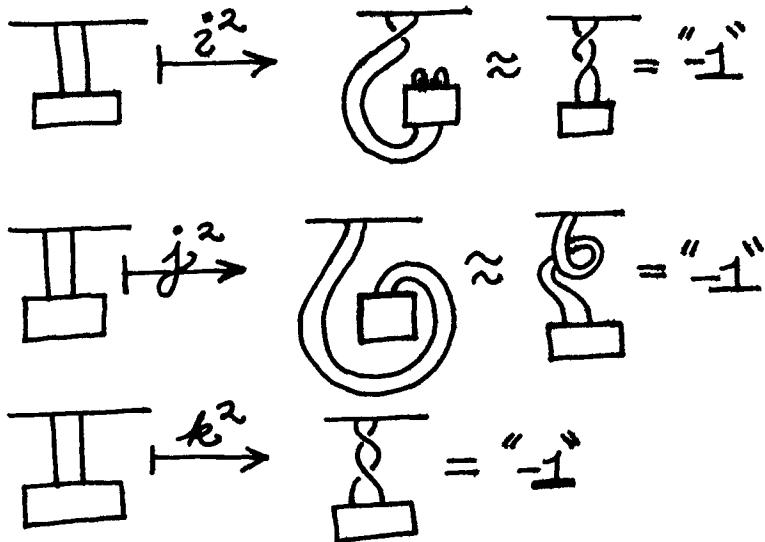


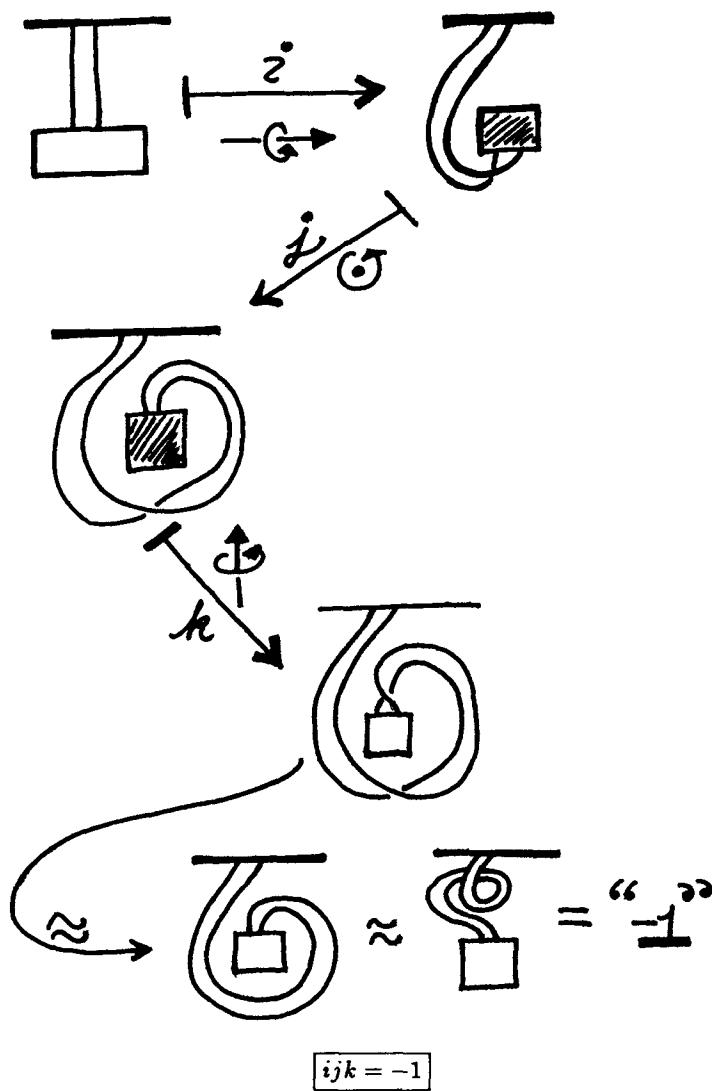
The Quaternions.

Having labelled these three basic roots of minus one i, j, k we find that

$$i^2 = j^2 = k^2 = ijk = -1.$$

These are the defining relations for the quaternion group.





Thus we have made the promised construction of a simple mechanical demonstrator of the quaternions.

The Belt Theorem.

The demonstrator is a special case of a more general result.

Let G be any subgroup of $SO(3)$. Regard $SU(2)$ as the set of unit quaternions: $a + bi + cj + dk$ with a, b, c, d real and $a^2 + b^2 + c^2 + d^2 = 1$.

Let $p : S^3 = SU(2) \rightarrow SO(3)$ be the two fold homomorphism defined by $p(g)(v) = gvg^{-1}$ where v is a point in Euclidean three space – regarded as the set of quaternions of the form $ai + bj + ck$ with a, b, c real numbers.

Then the preimage $\widehat{G} = p^{-1}(G)$ is a subgroup of $SU(2)$ that has twice as many elements as G . \widehat{G} is called the **double group** of G . For example, the double group of the Klein Four Group, K , (i.e. the group of symmetries of a rectangle in three space) is the quaternion group, H , generated by i, j and k in $SU(2)$.

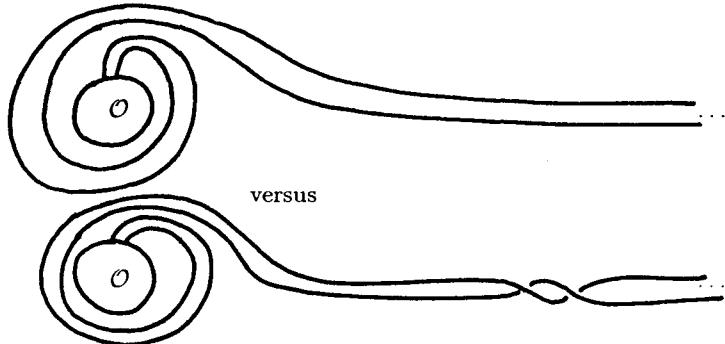
We have seen that the quaternions arise from the Klein Four Group, by attaching a band to the rectangle whose symmetries generate K . With this attachment, the rotations all double in value (2π is non-trivial and of order two), and the resulting group of rotations coupled with the sign indicated on the band is isomorphic with the quaternions H .

The Belt Theorem. Let A be an object embedded in three space, and let $G(A)$ denote the subgroup of $SO(3)$ of rotations that preserve A .

$G(A) = \{s \text{ in } SO(3) | s(A) = A\}$. Then the double group $\widehat{G}(A) \subset SU(2)$ is realized by attaching a band to A and following the prescription described above. That is, we perform a rotation on A and catalog the state of the band topologically. Two rotations of A are $\widehat{\sim}$ -equivalent if they are identical on A , and they produce the same state on the band. Then the set of $\widehat{\sim}$ -equivalence classes of rotations of A is isomorphic to the double group $\widehat{G}(A)$.

Proof of Belt Theorem. Let $\pi : SU(2) \rightarrow SO(3)$ be the covering homomorphism as described in Part II, section 10⁰. Given an element $\hat{g} \in \widehat{G} \subset SU(2)$ where G is the symmetry group of an object $\mathcal{O} \subset \mathbf{R}^3$, let $g = \pi(\hat{g})$ and let β denote the band that is attached to \mathcal{O} . We assume that the other end of β is “attached to infinity” for the purpose of this discussion. That is, \mathcal{O} is equipped with an infinite band with one end attached to \mathcal{O} . Twisting \mathcal{O} sends some curling onto the band.

Each position of \mathcal{O} in space curls the band, but by the belt trick we may compare 2π and 4π rotations of \mathcal{O} and find that they differ by a 2π twist in the band “at infinity”. That is, we can make the local configurations of (\mathcal{O}, β) the same for g and g followed by 2π , but the stationary part of the band some distance away from \mathcal{O} will have an extra twist in the second case.



As a result, we get a doubling of the set of spatial configurations of \mathcal{O} to the set of configurations of (\mathcal{O}, β) . If \mathcal{S} is the original set of configurations, let $\widehat{\mathcal{S}}$ be the set of configurations for (\mathcal{O}, β) . Then \mathcal{S} is in 1-1 correspondence with G , hence $\widehat{\mathcal{S}}$ is in 1-1 correspondence with \widehat{G} . It is easy to see that the latter correspondence is an isomorphism of groups. //

Programming the Belt Trick.

The program illustrates a belt that stretches between two concentric spheres in three dimensional space. A belt will be drawn that is the image under rotations of an untwisted belt that originally stretches from the north pole of the larger sphere to the north pole of the smaller sphere.

Call this belt the **original belt**. Upon choosing an axis of rotation, the program draws a belt that is obtained from the original belt via rotation by $2\pi t$ around the axis of the ***t*-shell**. The ***t*-shell** (with $0 < t < 1$) is a sphere concentric to the two boundary spheres at $t = 0$ and $t = 1$. Thus, at the top ($t = 0$), the belt is unmoved. It then rotates around the axis and is unmoved at the bottom ($t = 1$) – the inner concentric sphere.

For each choice of axis we get one image belt. The axis with $u_1 = 0$, $u_2 = 1$,

$u_3 = 0$ gives a belt with a right-handed 2π twist. As we turn the axis through π in the $u_1 - u_2$ plane, the belt undergoes a deformation that moves it around the inner sphere and finally returns it to its original position with a reversed twist of 2π .

The set of pictures of the deformation (stereo pairs after the program listing) were obtained by running the program for just such a set of axes. The program itself uses the quaternions to accomplish the rotation. Thus we are using the quaternions to illustrate themselves!

```

10 PRINT“DIRAC STRING TRICK!”
15 PRINT“COPYRIGHT(1989)L.H.KAUFFMAN - KNOTS INC.” 
20 PRINT“INPUT AXIS U1i+U2j+U3k”
25 PI=3.141592653886
27 U1=0:U2=1:U3=0:‘DEFAULT AXIS’
30 INPUT“U1”;U1:INPUT“U2”;U2:INPUT“U3”;U3
40 U=SQR(U1*U1+U2*U2+U3*U3)
50 U1=U1/U:U2=U2/U:U3=U3/U
60 PRINT“THE TWISTED BELT”:CLS
65 ‘SOME CIRCLES
70 FOR I=0 to 100
75 A=COS(2*PI*I/100):B=SIN(2*PI*I/100)
80 PSET(120+100*A,125-100*B)
90 PSET(120+20*A,125-20*B)
95 PSET(370+100*A,125-100*B)
97 PSET(370+20*A,125-20*B)
100 NEXT I
105 ‘DRAW BELT
110 FOR I=0 TO 50
120 FOR J=0 TO 20
130 X = -10+J
140 R = 100-I*(8/5)
150 Y+SQR(R*R-X*X)
160 Z=0
170 P1=X:P2=Y:P3=Z
180 IF I=0 OR I=50 THEN 210
190 TH=(PI/50)*I
200 GOSUB 2000:X=Y:Y=Z:Z=W
210 PSET(120+X,125-Y)
215 PSET(370+X+.2*Z,125-Y)
220 NEXT J
230 NEXT I

```

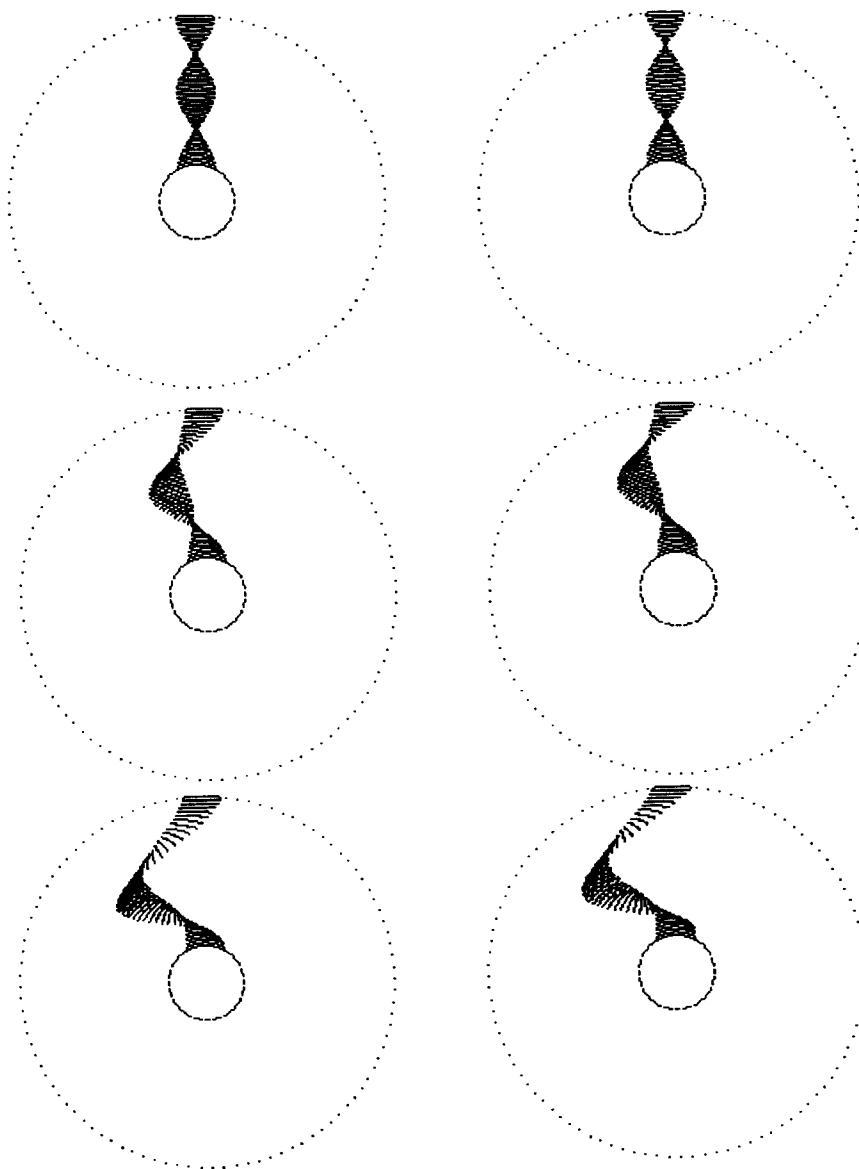
```

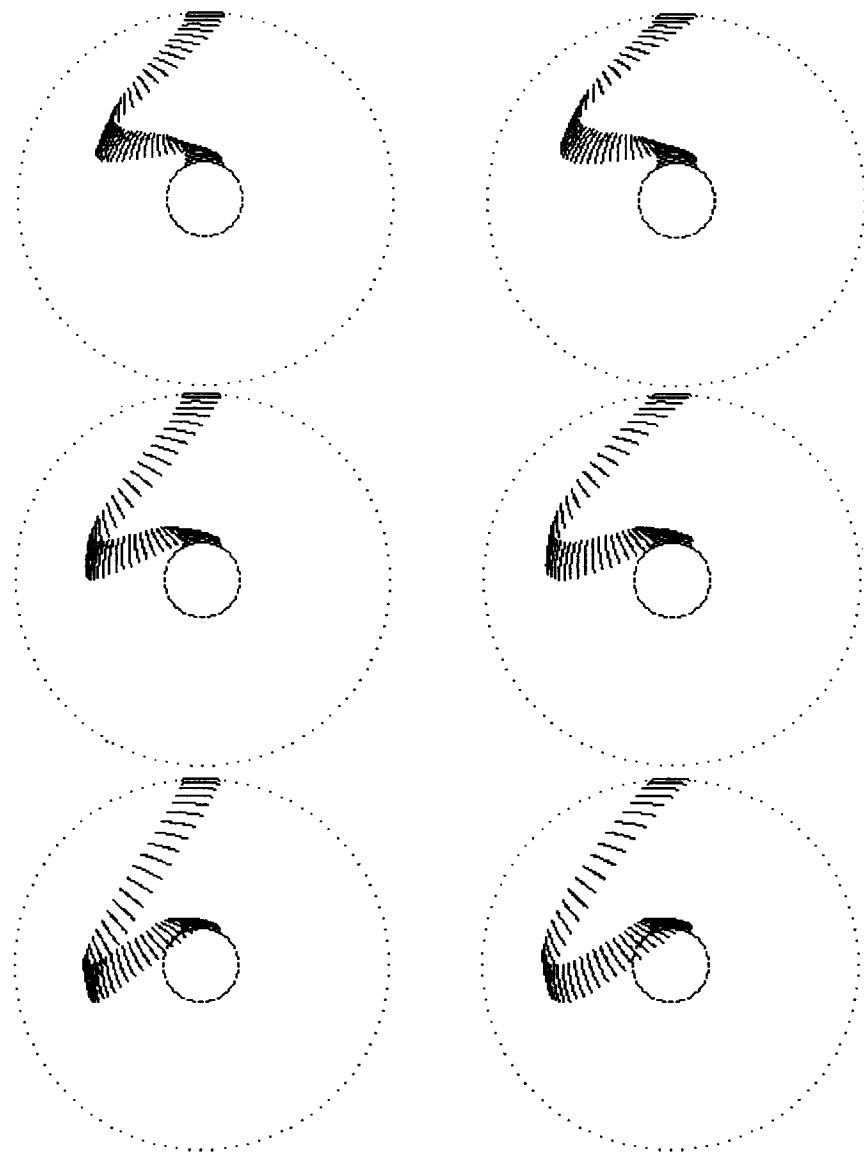
235 A$=INKEY$:IF A$="" THEN 235
237 GOTO 10
240 END
1000 'THE QUATERNION SUBROUTINE!
1020 'INPUT A+Bi+Cj+Dk
1040 'INPUT E+Fi+Gj+Hk
1060 X=A*E-B*D-C*G-D*H
1070 Y=B*E+A*D-G*C+H
1080 Z=C*E+D*F+A*G-B*H
1090 W=D*E-C*F+B*G+A*H
1100 RETURN
2000 'ROTATION ROUTINE
2020 'INPUT AXIS U1i+U2j+U3k
2030 'INPUT ANGLE TH=(1/2)rotation angle
2040 'INPUT POINT P1i+P2j+P3k
2050 P=COS(TH):Q=SIN(TH)
2060 A=P:B=Q*U1:C=Q*U2:D=Q*U3
2070 E=0:F=P1:G=P2:H=P3
2080 GOSUB 1000:'QMULT
2090 E=P:F=-B:G=-C:H=-D
2100 A=X:B=Y:C=Z:D=W
2110 GOSUB 1000:'QMULT
2120 RETURN

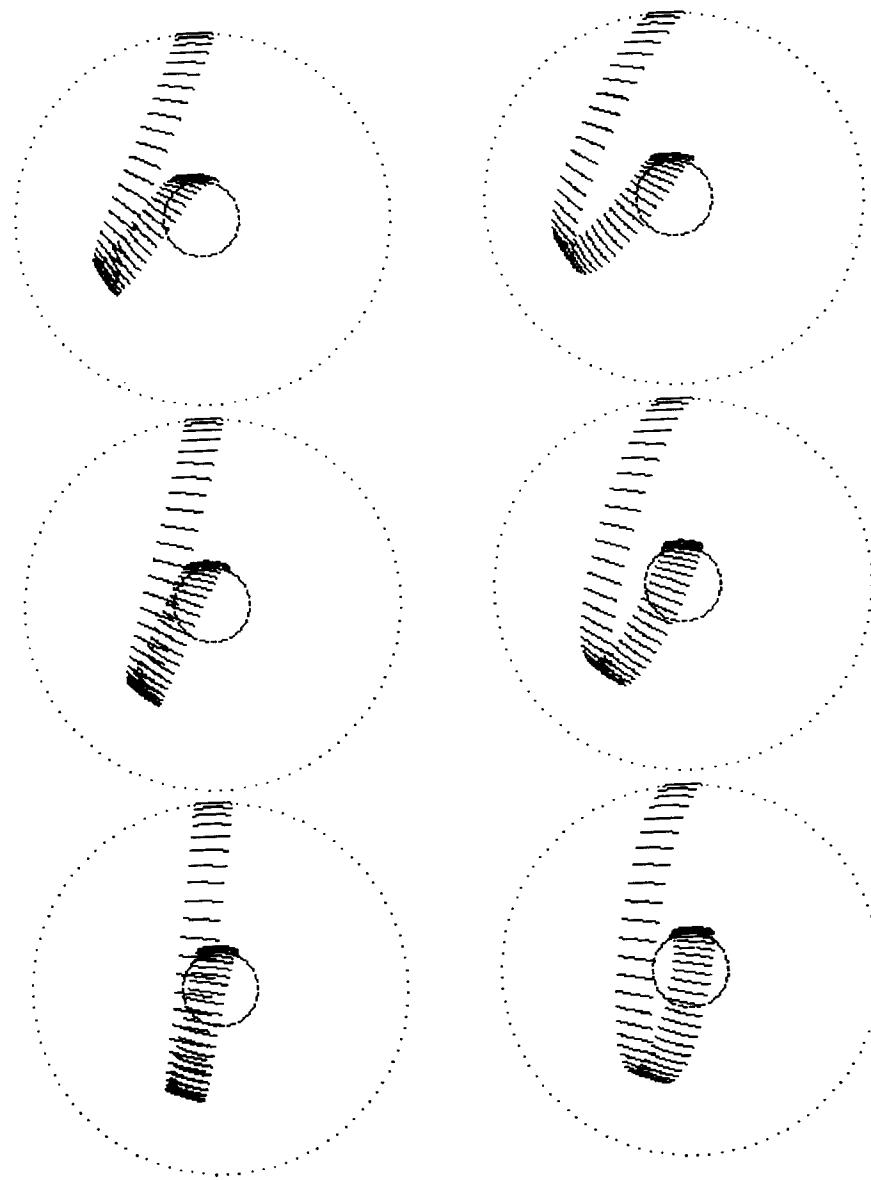
```

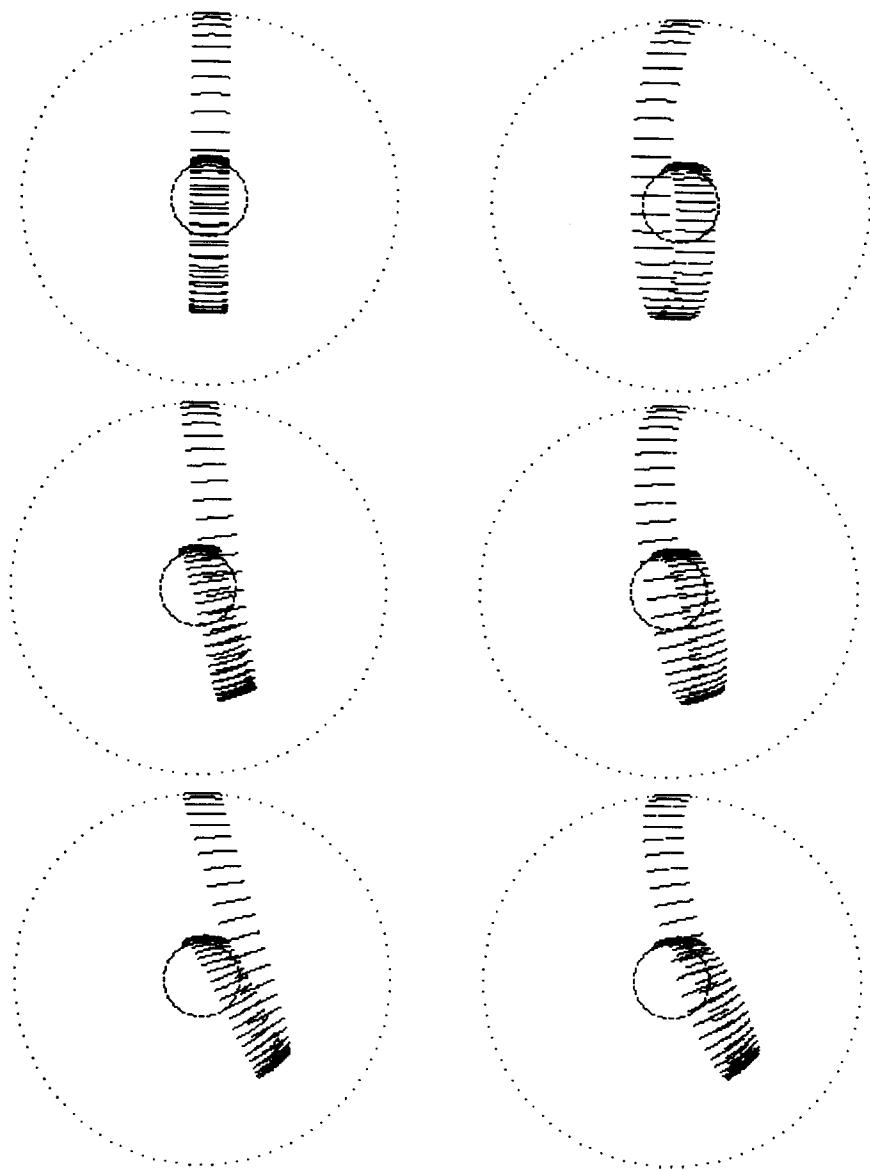
The Quaternionic Arm.

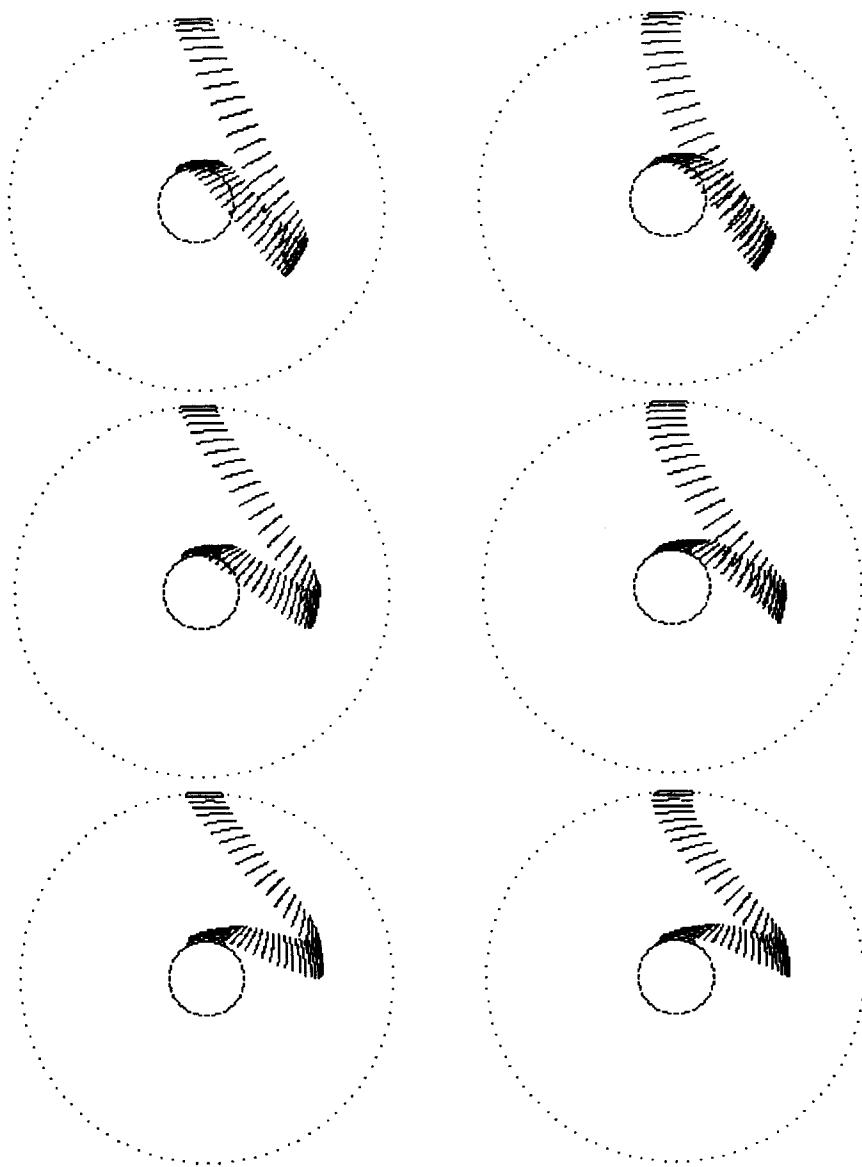
In an extraordinary conversation about quaternions, quantum mechanics, quantum psychology [O], quantum logic and the martial art Wing Chun, the author and Eddie Oshins discovered that the basic principle of the quaternion demonstrator can be done with only a single human arm. Hold your (right) arm directly outward perpendicular to your body with the palm up. Turn your arm by 180 degrees. This is *i*. Now bend your elbow until your hand points to your body. This is *j*. Now push your hand out in a stroke that first moves your palm downward - until your arm is once again fully extended from your body with your palm up. This is *k*. The entire movement produces an arm that is twisted by 360 degrees. Whence $ijk = -1$.

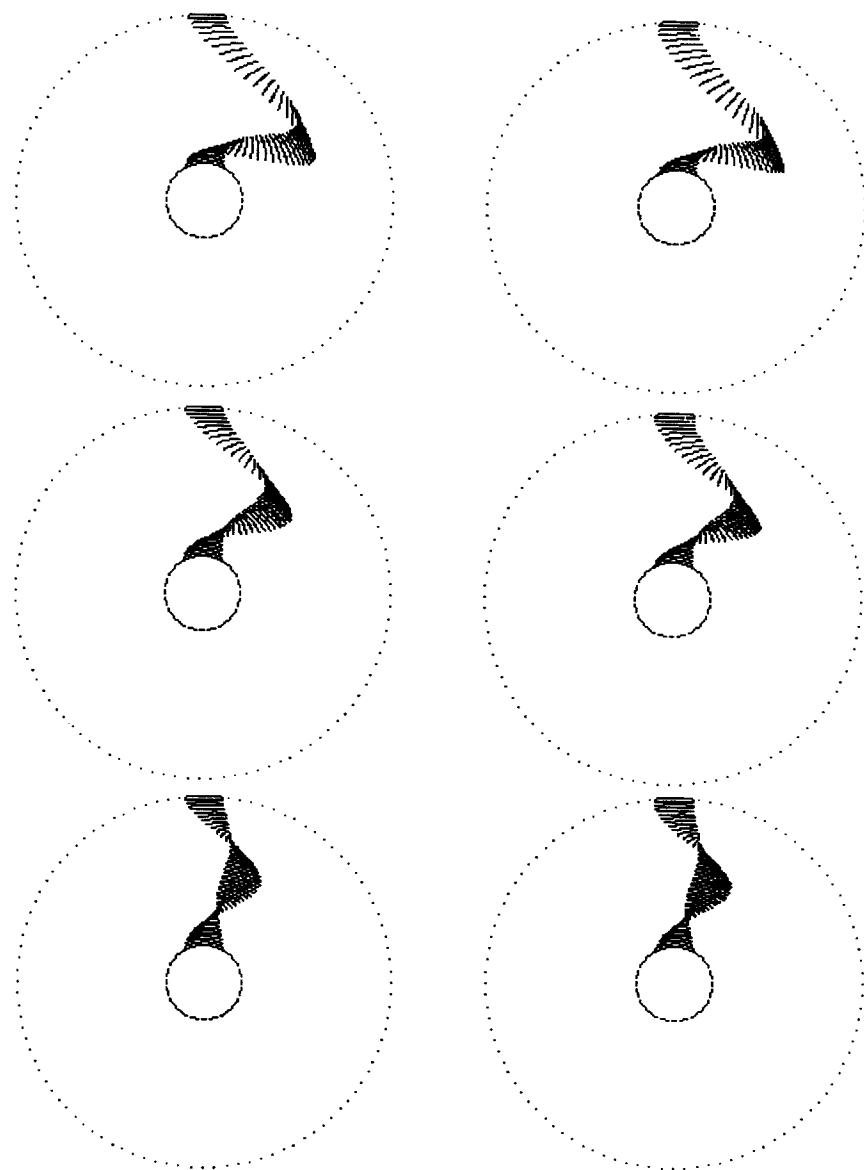












12⁰. The Penrose Theory of Spin Networks.

We begin with a quote from the introduction to the paper “Combinatorial Quantum Theory and Quantized Directions” [PEN4] by Roger Penrose. For the references to spin nets see [HW].

“According to conventional quantum theory, angular momentum can take only integral values (measured in units of $\hbar/2$) and the (probabilistic) rules for combining angular moments are of combinatorial nature. . . . In the limit of large angular momenta, we may . . . expect that the quantum rules for angular momentum will determine the geometry of directions in space. We may imagine these directions to be determined by a number of spinning bodies. The angles between their axes can then be defined in terms of the probabilities that their total angular momenta (i.e. their “spins”) will be increased or decreased when, say, an electron is thrown from one body to another. In this way the geometry of directions may be built up, and the problem is then to see whether the geometry so obtained agrees with what we know of the geometry of space and time.”

The Penrose spin networks are designed to facilitate calculations about angular momentum and $SL(2)$. Perhaps the most direct route into this formalism is to give part of the group-theoretic motivation, then set up the formal structure.

Recall that a spinor is a vector in two complex variables, denoted by ψ^A ($A = 1, 2$). Elements $U \in SL(2)$ act on spinors via

$$(U\psi)^A = \sum_B U_B^A \psi^B.$$

Defining the conjugate spinor by

$$\psi_A^* = \epsilon_{AB} \psi^B$$

(ϵ_{AB} is the standard epsilon: $\epsilon_{12} = 1 = -\epsilon_{21}$, $\epsilon_{11} = \epsilon_{22} = 0$. Einstein summation convention is operative throughout.), we have the natural $SL(2)$ invariant inner product

$$\psi\psi^* = \psi^A \epsilon_{AB} \psi^B.$$

In order to diagram this inner product, let

$$\psi^A \leftrightarrow \boxed{A}$$

Then it is natural to lower the index via

$$\psi_B^* \leftrightarrow B$$

so that

$$\psi\psi^* \leftrightarrow \boxed{} \quad \boxed{}$$

with the cap $A \cap B$ representing the epsilon matrix ϵ_{AB} .

Unfortunately (perhaps fortunately!) this notation is not topologically invariant since if $\cap \leftrightarrow \epsilon_{AB}$ and $\cup \leftrightarrow \epsilon^{AB}$ then

$$A \cap B \cup C \leftrightarrow \epsilon_{AB} \epsilon^{BC} = -\delta_A^C \leftrightarrow -A \cap C$$

where δ_A^C denotes the Kronecker delta. We remedy this situation by replacing

$$\left\{ \begin{array}{l} A \cap B \leftrightarrow \sqrt{-1}\epsilon_{AB} \\ C \cup D \leftrightarrow \sqrt{-1}\epsilon^{CD} \end{array} \right\}$$

the epsilons by their multiples by the square root of -1 .

This gives a topologically invariant diagrammatic theory, but translations are required in carrying $SL(2)$ -invariant tensors into the new framework. For example, the new loop value is

$$A \cap B = (\sqrt{-1})^2 + (-\sqrt{-1})^2 = -2,$$

whence the name negative dimensional tensors.

Even with this convention, we still need to add minus signs for each crossing in order to have

$$\begin{aligned} \text{Diagram with two strands crossing} &= -\epsilon_{cd}\delta_b^c\delta_a^d = -\epsilon_{ba} = \epsilon_{ab} \\ &= \text{Diagram with a loop}. \end{aligned}$$

Thus  $= -\delta_a^a\delta_c^b$ by definition. With these conventions, we have the basic binor identity:

$$\text{Diagram X} + \text{Diagram U} + \text{Diagram C} = 0.$$

Proof.

$$\begin{aligned} \text{Diagram X} + \text{Diagram U} + \text{Diagram C} &= -\delta_d^a\delta_c^b + \epsilon^{ab}\epsilon_{cd}(\sqrt{-1})^2 + \delta_c^a\delta_d^b \\ &= (\delta_c^a\delta_c^b - \delta_d^a\delta_d^b) - \epsilon^{ab}\epsilon_{cd} \\ &= 0 \end{aligned}$$

//

As aficionados of the bracket (Part I) we recognize the possibility of a topological generalization via

$$\text{Diagram X} = A \text{ Diagram U} + A^{-1} \text{ Diagram C}$$

$$\text{Diagram O} = -A^2 - A^{-2},$$

but this is postponed to the next section.

The next important spin network ingredient is the **antisymmetrizer**. This is a diagram sum associated to a bundle of lines, and it is denoted by



where N denotes the number of lines. The antisymmetrizer is defined by the formula

$$\text{Diagram with } N \text{ lines} = \sum_{\sigma \in S_N} \text{sgn}(\sigma) \text{ Diagram with } \sigma \text{ lines}$$

where σ runs over all permutations in S_N (the set of permutations on N distinct objects), and the σ in the box denotes the diagrammatic representation of this permutation as a braid projection. Thus

$$\boxed{\#}^N = N! \#^N, \text{ and}$$

$$\# = || - X$$

$$= || + \begin{smallmatrix} U \\ N \end{smallmatrix} +) (\quad [X + \begin{smallmatrix} U \\ N \end{smallmatrix} +) (= \phi]$$

$$\# = 2)(+ \begin{smallmatrix} U \\ N \end{smallmatrix}.$$

$$\# = ||| - X | - | X - X + X + X$$

$$= ||| + \begin{smallmatrix} U \\ N \end{smallmatrix} + ||| + | \begin{smallmatrix} U \\ N \end{smallmatrix} + ||| - X$$

$$+ \begin{smallmatrix} U \\ N \end{smallmatrix} + \begin{smallmatrix} U \\ N \end{smallmatrix} + | \begin{smallmatrix} U \\ N \end{smallmatrix} + |||$$

$$+ \begin{smallmatrix} U \\ N \end{smallmatrix} + \begin{smallmatrix} U \\ N \end{smallmatrix} + | \begin{smallmatrix} U \\ N \end{smallmatrix} + |||$$

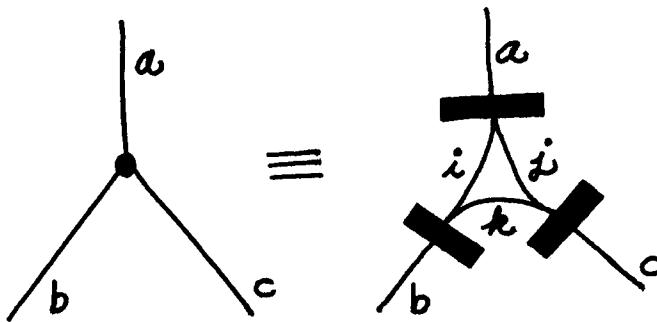
$$= 5||| + 3 \begin{smallmatrix} U \\ N \end{smallmatrix} + 3 | \begin{smallmatrix} U \\ N \end{smallmatrix} + \begin{smallmatrix} U \\ N \end{smallmatrix} + \begin{smallmatrix} U \\ N \end{smallmatrix} - X$$

$$= 5||| + 3 \begin{smallmatrix} U \\ N \end{smallmatrix} + 3 | \begin{smallmatrix} U \\ N \end{smallmatrix} + \begin{smallmatrix} U \\ N \end{smallmatrix} + \begin{smallmatrix} U \\ N \end{smallmatrix}$$

$$+ \begin{smallmatrix} U \\ X \end{smallmatrix} + X \quad [\begin{smallmatrix} U \\ X \end{smallmatrix} = \begin{smallmatrix} U \\ N \end{smallmatrix}]$$

$$\# = 6||| + 4 \begin{smallmatrix} U \\ N \end{smallmatrix} + 4 | \begin{smallmatrix} U \\ N \end{smallmatrix} + 2 \begin{smallmatrix} U \\ N \end{smallmatrix} + 2 \begin{smallmatrix} U \\ N \end{smallmatrix}.$$

These antisymmetrizers are the basic ingredients for making spin network calculations of Clebsch-Gordon coefficients, $3j$ and $6j$ symbols, and other apparatus of angular momentum. In this framework, the 3-vertex is defined as follows:



Here a, b, c are positive integers satisfying the condition that the equations $i + j = a, i + k = b, j + k = c$ can be solved in non-negative integers. Think of spins being apportioned in this way in interactions along the lines. This vertex gives the spin-network analog of the quantum mechanics of particles of spins b and c interacting to produce spin a .

Formally, the $6j$ symbols are defined in terms of the re-coupling formula

$$\begin{array}{c} b \\ \diagdown \\ a \end{array} \otimes \begin{array}{c} c \\ \diagup \\ d \end{array} = \sum_i \left\{ \begin{array}{ccc} a & b & i \\ c & d & j \end{array} \right\} \begin{array}{c} b \\ \diagup \\ a \end{array} \otimes \begin{array}{c} c \\ \diagdown \\ d \end{array}$$

To obtain a formula for these recoupling coefficients, one can proceed diagrammatically as follows.

1.

$\mu = \frac{\mu}{\delta_b} \Rightarrow \mu = \frac{s}{\delta_b}$

$$2. \quad \text{Diagram} = \sum_i \left\{ \begin{matrix} a & b & i \\ c & d & j \end{matrix} \right\} \text{Diagram}$$

$$\xrightarrow{(1.)} \text{Diagram} = \left\{ \begin{matrix} a & b & i \\ c & d & j \end{matrix} \right\} \text{Diagram}$$

$$\Rightarrow \text{Diagram} = \left\{ \begin{matrix} a & b & i \\ c & d & j \end{matrix} \right\} \text{Diagram}$$

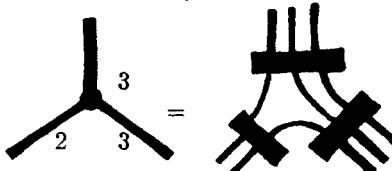
$$\Rightarrow \text{Diagram} = \left\{ \begin{matrix} a & b & i \\ c & d & j \end{matrix} \right\} \frac{\text{Diagram}}{i!}$$

$$\Rightarrow \left\{ \begin{matrix} a & b & i \\ c & d & j \end{matrix} \right\} = \frac{\text{Diagram}}{\text{Diagram}}$$

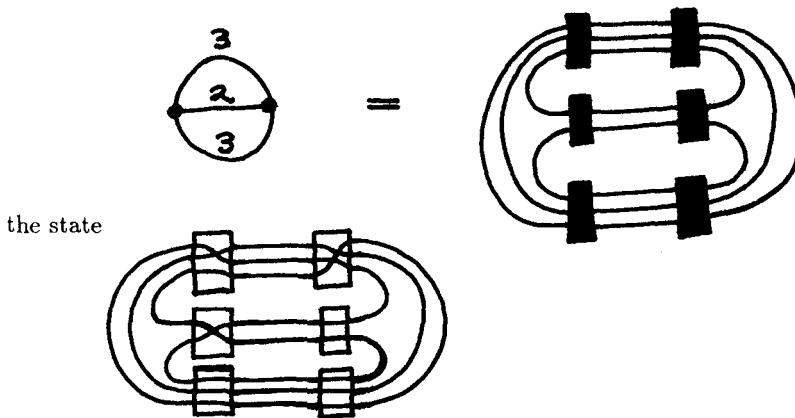
The final formula determines the $6j$ symbol in terms of spin network evaluations of some small nets.

This formulation should make clear what we mean by a spin network. It is a graph with trivalent vertices and spin assignments on the lines so that each 3-

vertex is admissible. Each vertex comes equipped with a specific cyclic ordering and is embedded in the plane to respect this ordering. A network with no free ends is evaluated just like the bracket except that each antisymmetrizer connotes a large sum of different connections. For example,



and this expands into $(2!)(3!)^2 = 72$ local configurations. The rule for expanding the antisymmetrizers takes care of the signs. Each closed loop in a given network receives the value (-2) . Thus for



contributes $(-1)^4(-2)^2$ since the total permutation sign is $(-1)^4$ (from four crossings) and the two loops each contribute (-2) .

A standard spin network convention is to ignore self-crossings of the trivalent graph. Then the norm, $\|G\|$, of a closed spin network G is given by the formula

$$\|G\| = \sum_{\sigma} (-1)^{\chi(\sigma)} (-2)^{\|\sigma\|}$$

where σ runs over all states in the sense (of replacing the bars of the antisymmetrizers by specific connections), and $\chi(\sigma)$ is the total number of crossings (within the bars) in the state. $\|\sigma\|$ is the number of closed loops.

Just as with the bracket, we can also evaluate by using the binor expansion

$$X = - \begin{array}{c} \text{X} \\ \diagup \quad \diagdown \end{array} -) (. \text{ Thus}$$

$$-2 = \begin{array}{c} 8 \\ \diagup \quad \diagdown \end{array}$$

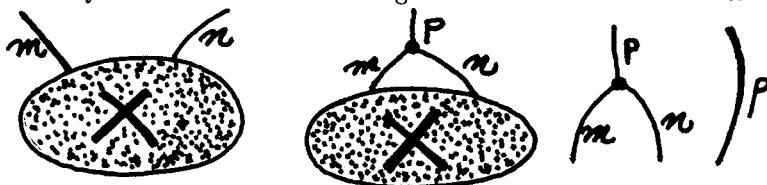
$$\text{and } 8 = - \begin{array}{c} 8 \\ \diagup \quad \diagdown \end{array} = -(-2)^2 - (-2) = -4 + 2 = -2.$$

Note also, that the antisymmetrizers naturally satisfy the formula

$$\begin{array}{c} n \\ \diagup \quad \diagdown \end{array} = n! \begin{array}{c} n \\ \diagup \quad \diagdown \end{array} \cdot$$

The Spin-Geometry Theorem.

Penrose takes the formalism of the spin networks as a possible combinatorial basis for spacetime, and proceeds as follows [PEN4]: "Consider a situation in which a portion of the universe is represented by a network having among its free ends, one numbered m and one numbered n . Suppose the particles or structures represented by these two ends combine together to form a new structure.

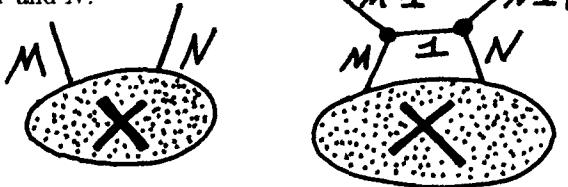


We wish to know, for any allowable p , what is the probability that the angular momentum number of this new structure be p ." Penrose takes this probability to be given by the formula

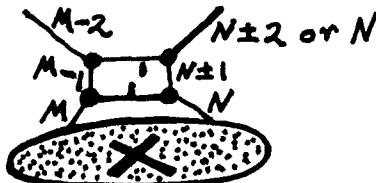
$$P = \frac{\begin{array}{c} p \\ \diagup \quad \diagdown \end{array}}{\begin{array}{c} m/m \\ \diagup \quad \diagdown \end{array} \quad \begin{array}{c} p \\ \diagup \quad \diagdown \end{array}}$$

where the norm of a network with free ends is the norm of two copies of that net plugged into one another. (analog of $\psi\psi^*$).

"Consider now the situation below involving two bodies with large angular momenta M and N .



We define the **angle** between the axes of the bodies in terms of relative probability of occurrence of $N + 1$ and $N - 1$ respectively, in the "experiment" shown above. ... To eliminate the possibility that part of the probability be due to ignorance, we envisage a repetition of the experiment as shown below.



If the probability given in the second experiment is essentially unaffected by the result of the first experiment then we say that the angle θ between the axes of the bodies is well-defined and is determined by this probability. It then turns out that

$$\left| \left| \begin{array}{c} \text{M-2} \\ \text{M-1} \\ \text{N} \end{array} \right| \right| \cong \frac{1}{2} \cos(\theta)$$

and from this ... it is possible to show that the angles obtained in this way satisfy the same laws as do angles in a three-dimensional Euclidean space." This is the Spin-Geometry Theorem.

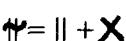
Calculating Vector Coupling Coefficients.

Here I show how to translate a standard problem about angular momenta into the spin network language. Let J_1, J_2, J_3 be the standard generators for the Lie algebra of $SU(2)$ (compare section 9⁰, Part II).

$$J_1 = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad J_2 = \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad J_3 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Let $J_+ = J_1 + iJ_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $J_- = J_1 - iJ_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ be the raising and lowering operators, so that for $u = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $d = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ we have $J_3 u = \left(\frac{1}{2}\right)u$, $J_3 d = \left(-\frac{1}{2}\right)d$ and $J_+ d = u$, $J_+ u = 0$, $J_- u = d$, $J_- d = 0$. Let u and d be diagrammed via

$$u \leftrightarrow \text{↑}, \quad d \leftrightarrow \text{↓}$$

Let  denote **symmetrization**. Thus . Then, letting

$$\tilde{J} = (J \otimes 1 \otimes 1 \otimes \dots \otimes 1) + (1 \otimes J \otimes 1 \otimes \dots) + \dots,$$

we let

$$V(k, \ell) = \underbrace{\text{↑↑ } \dots \text{↑↑}}_k \text{---} \text{---} \text{---} \underbrace{\text{↓↓ } \dots \text{↓↓}}_\ell$$

and get

$$J_3(V(k, \ell)) = \left(\frac{k - \ell}{2}\right)V(k, \ell)$$

in the corresponding tensor product. This yields an irreducible representation of $\mathfrak{sl}(2)$ with eigenvalues $\{-m/2, \dots, m/2\}$ and $m = (k + \ell)/2$. We write

$$|jm\rangle = \underbrace{\text{↑↑ } \dots \text{↑↑}}_r \text{---} \text{---} \text{---} \underbrace{\text{↓↓ } \dots \text{↓↓}}_s$$

$$|jm\rangle = \text{Diagram with arrows and a label } 2j$$

with $m = \frac{r-s}{2}$, $j = \frac{r+s}{2}$. Then $J_3|jm\rangle = m|jm\rangle$, and the raising and lowering operators move these guys around.

We wish to compute the vector coupling coefficient, or Wigner coefficient giving the amplitude for three spins to combine to zero. This is denoted by

$$\begin{bmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{bmatrix}.$$

In order to compute this amplitude, we form the product state

$|j_1 m_1\rangle |j_2 m_2\rangle |j_3 m_3\rangle$ and apply antisymmetrizers to reduce it to a scalar.

Question. Why do we apply ϵ_{ab} ($\Pi_{ab} = \epsilon_{ab}$) (i.e. antisymmetrizers) to reduce these tensors to scalars?

Answer. To combine to zero involves spin conservation. Thus the possibilities are

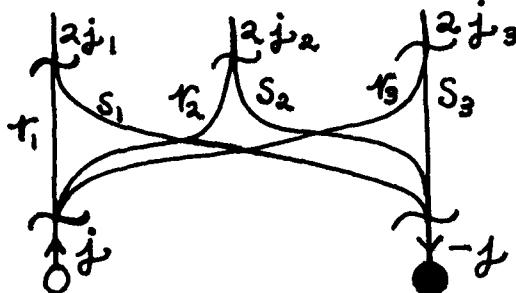
$$\uparrow_{(\text{time's arrow})} \quad +\frac{1}{2} \quad \quad -\frac{1}{2} \quad \quad \text{and} \quad \quad -\frac{1}{2} \quad \quad +\frac{1}{2}.$$

And remember, the difference between these two is the difference in phase of the wave function. Thus this change in sign

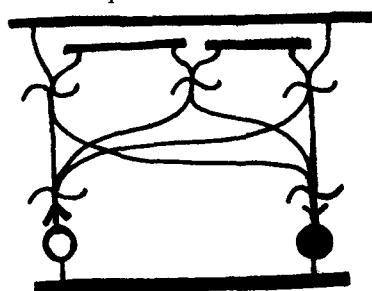
$$\Pi_{ab} = -\Pi_{ba}$$

corresponds to the change in phase when you walk around a spin ($\frac{1}{2}$) particle (Compare with sections 9⁰, 10⁰ and 11⁰ of Part II.).

The diagram representing the structure of three spins of total spin zero is as shown below



We wish to compute the contraction



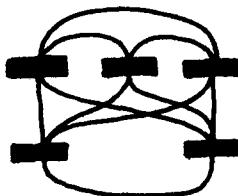
$$(m_K = (r_K - \delta_K)/2) \\ K = 1, 2, 3$$

$$= \begin{bmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{bmatrix}$$

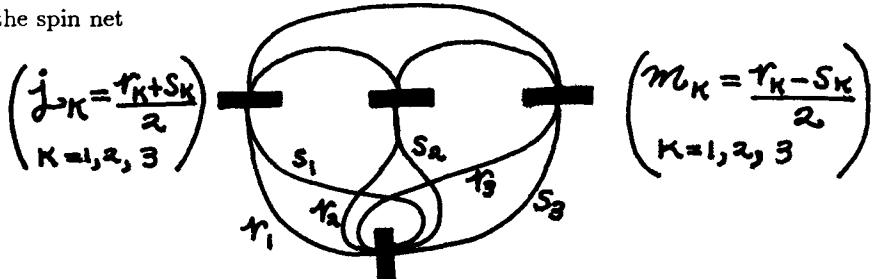
In order to translate this into a spin network, we replace Π by $\sqrt{-1}\Pi = \cup$ and Π by $\sqrt{-1}\Pi = \cap$, and **symmetrizers become anti-symmetrizers** (by the

convention of minus signs for crossings). Thus

$$\begin{bmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{bmatrix} =$$



This reduces the calculation of the vector coupling coefficient to the evaluation of the spin net



See [HW] for a discussion of combinatorial techniques of evaluation.

The Chromatic Method.

Here is a summary of Moussouris' chromatic method of spin net evaluation. Recall that we have defined the norm $\|G\|$ of a closed network to be

$$\|G\| = \sum_{\vec{\sigma}} \text{sgn}(\vec{\sigma})(-2)^{\|\vec{\sigma}\|}$$

where $\vec{\sigma}$ is a “state” of the network consisting in specific choices for each antisymmetrizer. Let $\text{sgn}(\vec{\sigma})$ denote the product of the permutation signs induced at each antisymmetrizer, and $\|\vec{\sigma}\|$ denote the number of loops in $\vec{\sigma}$. For example,

$$\begin{aligned} \| \text{---} \circ \circ \circ \| &= \| \circ \circ \circ - \circ \circ \circ - \circ \circ \circ + \circ \circ \circ + \circ \circ \circ - \circ \circ \circ \| \\ &= (-2)^3 - (-2)^2 - (-2)^2 + (-2) + (-2) - (-2)^2 \\ &= -8 - 4 - 4 - 2 - 2 - 4 \\ &= -24. \end{aligned}$$

Note that in this language, we can think of the network and its state expansion as synonymous.

In order to facilitate computations it is convenient to define a polynomial $P_G(\delta)$ as follows

$$P_G(\delta) = \left(\sum_{\vec{\sigma}} \text{sgn}(\vec{\sigma}) \delta^{\|\vec{\sigma}\|} \right) / \Pi e!$$

where $\Pi e!$ denotes the product of the factorial of the multiplicities assigned to the edges of G . Now note that in these strand nets the edges enter antisymmetrizer bars in a pattern



where a, b, c, d denote the multiplicities of the lines. The term $\Pi e!$ denotes the product of factorials of these multiplicities. Of course the standard spin network norm occurs for $\delta = -2$.

If $\delta = N$ (a positive integer), we can regard the strands as labelled from an index set of “colors” $\mathcal{I} = \{1, 2, \dots, N\}$. Then a non-zero state corresponds to an N coloring of the (individual strands) so that for each strand entering a bar there is exactly one strand of the same color leaving the bar. (Note that N must be greater than or equal to the maximum multiplicity entering an antisymmetrizer in order to use this set up.)

Thus for a given choice of permutations at the bars, we can count cycles in the resulting graph, and stipulate that **cycles sharing a bar have different colors**.

For a state $\vec{\sigma}$, let $c_i(\vec{\sigma})$ denote the number of cycles in $\vec{\sigma}$ of type i (assuming that we have classified the possible cycles into **types** so that cycles in a given type can receive the same color). The possibility of this type classification rests on the observation: **All terms in the state summation for $P_G(\delta)$ with a given set of cycle numbers $\{c_i\}$ have the same sign.**

Proof. The sign is the parity of the number of crossings that occur at the bars. The total number of crossings between different cycles is even. (Recall that for spin nets we do not count crossings away from bars.) Hence the total parity of the

bar-crossings equals the parity of the “spurious” intersections of different cycles away from the bars. The number of spurious intersections is

$$\sum_{e_i \cap e_j} e_i \cdot e_j - \sum_{\text{self-intersecting cycles}} c_k^2.$$

Hence the overall parity is determined by $\{c_k\}$. //

Note that for each allowed coloring of the cycles, the number of terms in the sum is $\Pi c!$ (as defined above). Thus we have proved the

Theorem 12.1. The loop polynomial $P_G(\delta)$ of the strand net G is given by the formula (for $\delta = N$ a positive integer)

$$P_G(\delta) = \sigma \sum_{\{c_i\}} \epsilon K(\{c_i\}, \delta)$$

where

$$\begin{aligned} \sigma &= (-1)^{\sum_{e_i \cap e_j} (e_i \cdot e_j)} \\ \epsilon &= (-1)^{\sum_{\text{self intersections}} c_k \cdot c_k} \end{aligned}$$

and $K(\{c_i\}, N)$ is equal to the number of allowed colorings of the configuration of $\{c_i\}$ cycles. The sum is over all non-negative cycle numbers $\{c_i\}$ such that for each edge $e_i = \sum c_j$ where the c_j catalog those cycles that pass through e_i .

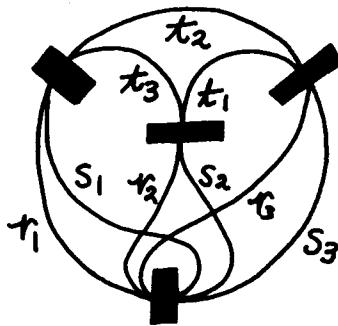
Remark. Knowing $P_G(\delta)$ as a polynomial in δ via its evaluations at positive integers, allows the evaluation at $\delta = -2$. We get an integer for $P_G(-2)$ since any polynomial that is integer valued at positive integers is integer valued at negative integers as well. In fact we will use the following:

$$\binom{N}{K} = \frac{N(N-1)\dots(N-K+1)}{K!}$$

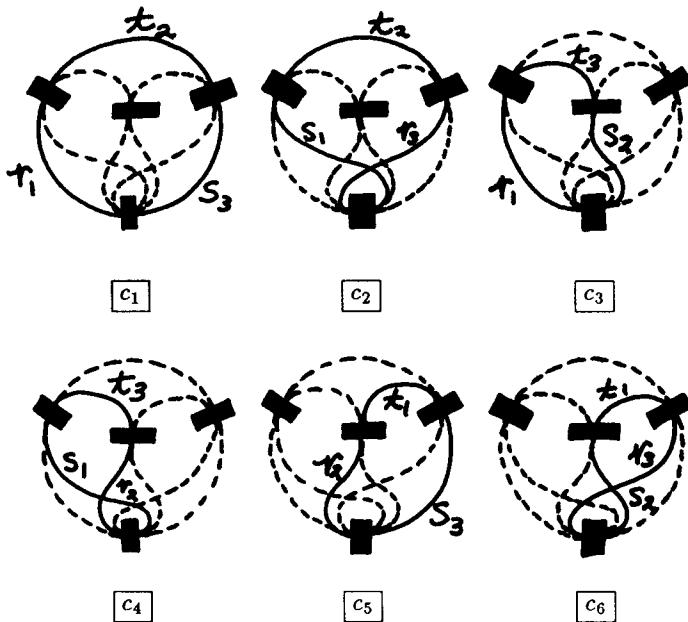
$$\Rightarrow \binom{-N}{K} = \frac{(-N)(-N-1)\dots(-N-K+1)}{K!} \in \mathbb{Z}.$$

Note that $(-N)! = (-1)^N (2N-1)!/(N-1)!$.

Example. Here is an application of the theorem. Let G be the network shown below.



This net has 9 edges, 4 bars and 6 cycle types as shown below



Thus, we have

$$t_1 = c_5 + c_6 \quad s_1 = c_2 + c_4 \quad r_1 = c_1 + c_3$$

$$t_2 = c_1 + c_2 \quad s_2 = c_3 + c_6 \quad r_2 = c_4 + c_5$$

$$t_3 = c_3 + c_4 \quad s_3 = c_1 + c_5 \quad r_3 = c_2 + c_6$$

and, letting $\pi = s_1r_2 + s_1r_3 + r_3s_2 + c_2 + c_4 + c_6$, we have $(-1)^\pi$ is the sign of

the corresponding state. The equations above let us solve for all the c_i in terms of $z = c_1$:

$$\begin{array}{ll} c_1 = z & c_5 = s_3 - z \\ c_2 = t_2 - z & c_6 = t_1 - s_3 + z \\ c_3 = r_1 - z & \\ c_4 = t_3 - r_1 + z & \end{array}$$

Thus

$$\pi = s_1 r_2 + s_1 r_3 + r_3 s_2 + (t_2 - z) + (t_3 - r_1 + z) + (t_1 - s_3 + z)$$

$$\pi = s_1 r_2 + s_1 r_3 + r_3 s_2 + t_1 + t_2 + t_3 - s_3 - r_1 + z.$$

Here $K(\{c_i\}, N)$ equals the number of ways to distribute N objects in 6 bins with c_i objects in each bin. Hence, letting $J = \sum c_i = t_1 + t_2 + t_3$,

$$K(\{c_i\}, N) = \frac{N!/(N - J)!}{c_1! c_2! c_3! c_4! c_5! c_6!}.$$

The chromatic sum is over the range of z giving non-negative values for all c_i .

Thus, at $N = -2$, $N!/(N - J)! = (-2)(-3)\dots(-J - 1) = (-1)^J(J + 1)!$. Let $q = \pi - z$ (above). Then

$$\begin{aligned} (\pi e!) P_G(-2) = \\ (-1)^q (J + 1)! \sum_z \frac{(-1)^z}{z! (t_2 - z)! (r_1 - z)! (t_3 - r_1 + z)! (s_3 - z)! (t_1 - s_3 + z)!} \end{aligned}$$

Exercise. Recall that the network G of this example is derived from the recoupling network for spins $\left[\begin{smallmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{smallmatrix} \right]$ as described in the first part of this section. Re-express the evaluation in terms of the j 's and m 's and compare your results with the calculations of Clebsch-Gordon coefficients in any text on angular momentum.

13⁰. Q-Spin Networks and the Magic Weave.

The last section has been a quick trip through the formalism and interpretation of the Penrose spin networks. At the time they were invented, these networks seemed a curious place to **begin** a theory. Their structure was obtained by a process of **descent** from the apparatus of quantum mechanical angular momentum.

Yet the basis of the spin network theory is the extraordinarily simple binor identity

$$\text{X} + \text{U} + \text{J} = 0,$$

and we know this to be a special case of the bracket polynomial identity

$$\text{X} = A \text{U} + A^{-1} \text{J}$$

(brackets removed for convenience). In the theory of the bracket polynomial the loop value is $\delta = -A^2 - A^{-2}$, hence binors occur for $A = -1$. We have already explained in Part I (section 9⁰) how the bracket is related to the quantum group $SL(2)q$ for $\sqrt{q} = A$.

Thus, it should be clear to the reader that there is a possibility for generalizing the Penrose spin nets to q -spin nets based upon the bracket identity. In this section I shall take the first steps in this direction, and outline what I believe will be the main line of this development. Even in their early stages, the q -spin networks are very interesting. For example, we shall see that the natural generalization of the anti-symmetrizer leads to a “magic weave” that realizes the special projection operators in the Temperley-Lieb algebra (see section 16⁰ of Part I). We then use the q -spin nets to construct Turaev-Viro invariants of 3-manifolds. ([TU2], [LK19], [LK20].)

Before beginning the technicalities of q -spin nets, I want to share foundational thoughts about the design of such a theory and its relationship with quantum mechanics and relativity. This occurs under the heading – **network design**. The reader interested in going directly into the nets can skip across to the sub-heading **q -spin nets**.

Network Design – Distinction as Pregeometry.

What does it mean to produce a combinatorial model of spacetime? One thinks of networks, graphs, formal relationships, a fugue of interlinked constructions. Yet what will be fundamental to such an enterprise is a point of view, or a place to stand, from which the time space unfolds. Before constructing, before speaking and even before logic stands a single concept, and that is the concept of **distinction**. I take this concept first in ordinary language where it has a multitude of uses and a curious circularity. For there can be no seeing without a seer, no speaking without a speaker. We, who would discuss this concept of distinction, become distinguished through the conversation of distinction that is us. Distinction is a concept that requires its own understanding. Fortunately, there are examples: Draw a distinction; draw a circle in the plane; delineate the boundary of this room; indicate the elements of the set of prime integers. Speech brings forth an architecture of distinctions and a framework for clarity and for action.

The specific models we use for space and spacetime are extraordinary towers of distinctions. Think of creating the positive integers from the null set (Yea, think of creating the null set from the void!), the rationals from the integers, the reals from limits of sequences of rationals, the coordinate spaces, the metrics and all the semantics of interpretation - just to arrive at Minkowski spacetime and the Lorentz group! It is no wonder that there is a desire for other ways to tell this story. (Compare [F12], [KVA], [LK17], [SB2].)

There are other ways. At the risk of creating a very long digression, I want to indicate how the **Lorentz group arises naturally in relation to any distinction**. We implicitly use the Lorentz group in the ordinary world of speech, thought and action. Recall from section 9⁰ of Part II that in light cone coordinates for the Minkowski plane the Lorentz transformation has the form $\mathcal{L}[A, B] = [KA, K^{-1}B]$ where K is a non-zero constant derived from the relative velocity of the two inertial frames. In fact, it is useful to remember that for time-space coordinates (t, x) , the corresponding light-cone coordinates are $[t + x, t - x]$ (for light speed = 1). Thus $(t - x)$ can be interpreted as the time of emission of a light signal and $(t + x)$ as the time of reception of that signal after it has reflected from an event at the point (t, x) .

Writing $[t+x, t-x]$ as $[t+x, t-x] = t \cdot 1 + x\sigma$ where $1 = [1, 1]$ and $\sigma = [1, -1]$ we recover standard coordinates and the Pauli algebra. That is, it is natural to define

$$\begin{aligned}[X, Y][Z, W] &= [XZ, YW] \\ [X, Y][Z, W] &= [X+Z, Y+W]\end{aligned}$$

in the light-cone coordinates. Then $\mathcal{L}[A, B] = [K, K^{-1}][A, B]$, and the Lorentz transformation is a product in this algebra. In this algebra structure

$\sigma^2 = [1, -1][1, -1] = [1, 1] = 1$, and in order to obtain a uniform algebraic description over arbitrary space directions, we conclude that $\sigma^2 = 1$ for any unit direction. This entails a Clifford algebra structure on the space of directions [LK17]. For example, if $\sigma_x, \sigma_y, \sigma_z$ denote unit perpendicular directions in three space with $\sigma_x\sigma_y = -\sigma_y\sigma_x$, $\sigma_x\sigma_z = -\sigma_z\sigma_x$, $\sigma_y\sigma_z = -\sigma_z\sigma_y$ and $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$, then $(a\sigma_x + b\sigma_y + c\sigma_z)^2 = a^2 + b^2 + c^2$ for scalars a, b, c commuting with $\sigma_x, \sigma_y, \sigma_z$. Hence, any unit direction has square 1.

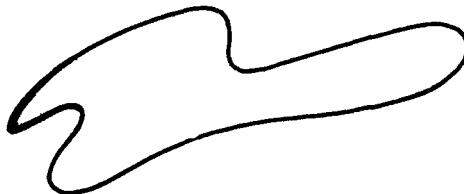
By this line of thought we are led to the minimal Pauli representation of a spacetime point as a Hermitian 2×2 matrix H :

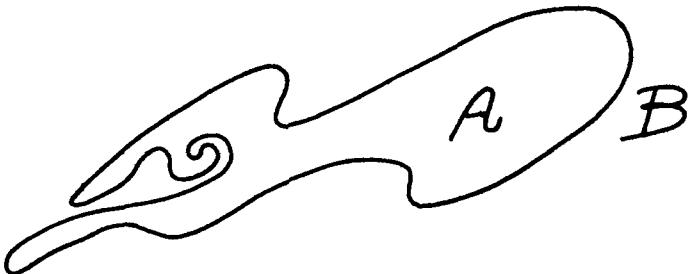
$$H = \begin{bmatrix} T + X & Y + \sqrt{-1}Z \\ Y - \sqrt{-1}Z & T - X \end{bmatrix} = T + X\sigma_x + Y\sigma_y + Z\sigma_z$$

and $\|H\|^2 = \text{Det}(H) = T^2 - X^2 - Y^2 - Z^2$.

All this comes from nothing but light cone coordinates and the principle of relativity. I will descend **underneath** this structure and look at the distinctions out of which it is built. (See [LK30].)

Therefore, begin again. Let a distinction be given





with sides labelled A and B . Call this distinction $[A, B]$. Consider A and B as **evaluations** of the two sides of the distinction. (In our spacetime scenario the distinction has the form [**after**, **before**], and numerical times are the evaluations.) Now suppose another observer gives evaluations $[A', B']$ with $A' = \rho A$, $B' = \lambda B$ for non-zero ρ and λ . Then

$$[\rho, \lambda] = \sqrt{\rho\lambda}[K, K^{-1}]$$

where $K = \sqrt{\rho/\lambda}$.

$$[A', B'] = \sqrt{\rho\lambda}[KA, K^{-1}B].$$

Thus, **up to a scale factor the numerical transformation of any distinction is a Lorentz transformation**. In this sense the Lorentz group operates in all contexts of evaluation. The Lorentz group arises at once along with the concept of (valued) distinction.

We should look more closely to see if the Pauli algebra itself is right in front of us in the first distinction. Therefore, consider once more a distinction with sides labelled A and B : I will suppose only the most primitive of evaluations - namely $+1$ and -1 . Thus the space of evaluations consists in the primitive distinction $-1/+1$ or unmarked/mark. By choosing a mark we can represent -1 as an

un-marked side and +1 as a marked side.



In any case there are now two basic operations:

- (1) **Change** the status of a side.
- (2) **Interchange** the relative status of two sides.

I represent **change** by two operations $p[A, B] = [+A, -B]$ (change inside) and $q[A, B] = [-A, +B]$ (change outside). The interchange is given by the formula $\Phi[A, B] = [B, A]$. We have $pq[A, B] = [-A, -B] = -[A, B]$. Hence $pq = -1$, or $q = -p$. Thus -1 denotes the operation changing the status of two sides. For convenience, I let \bar{p} denote $-p$. Of course 1 denotes the identity operation. We then have 1, p , Φ and $p\Phi$ with $p\Phi[A, B] = [+B, -A]$. Hence

$$(p\Phi)^2 = -1.$$

The distinction (observed) gives rise of its own accord to a square root of negative one, and three elements of square one ($1, p, \Phi$). By introducing an extra commuting $\sqrt{-1}$ (If it exists, why not replicate it?) we get $\sigma_x = p$, $\sigma_y = \Phi$, $\sigma_z = \sqrt{-1}p\Phi$ and the event $T + Xp + Y\Phi + Z\sqrt{-1}p\Phi$ corresponds directly to the Hermitian matrix

$$\begin{aligned} H &= \begin{bmatrix} T + X & Y + \sqrt{-1}Z \\ Y - \sqrt{-1}Z & T - X \end{bmatrix} \\ &= T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + X \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + Y \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \sqrt{-1}Z \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \end{aligned}$$

In fact, these 2×2 matrices

$$1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, p = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}, \Phi = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

represent our operations by left multiplication, and $p\Phi = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

Spacetime arises directly from the operator algebra of a distinction.

The basic Hermitian representation of an event proceeds directly from the idea of a distinction. Also, this point of view shows the curious splitting of 4-space into

$$[T + X, T - X] = T + \sigma X \text{ and } Y + \sqrt{-1}Z.$$

In the form

$$T + Xp + (Y + Z\sqrt{-1}p)\Phi$$

we see that **operationally** the split is between the actions that consider **one side or both sides (change and interchange)** of the distinction.

The important point about this line of reasoning is that, technicalities aside, spacetime is an elementary concept, proceeding outward from the very roots of our being in the world. Resting in this, it will be interesting to see how spacetime arises in the q -spin nets. Since we are building it in from the beginning, it is already there.

In particular, the formula

$$\cancel{\times} = A \cancel{\times} + A^{-1}) ($$

now appears in a new light as a knot theoretic way to write an ordered pair $[A, A^{-1}]$. (And $[A, A^{-1}]$ is a possible Lorentz transformation in light cone coordinates.) Here the crossing indicates the distinction $[A, B]$:

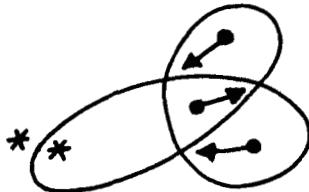


It is not the case at this level that the splices  and  have algebra structure related to the Pauli algebra. However, this **does** occur in the FKT [LK2] model

of the Conway-Alexander polynomial (see also [LK3], [LK9]). In that model, we have

$$\text{Diagram} = t \text{ Diagram} - t^{-1} \text{ Diagram} + \text{Diagram}$$

where the dots indicate the presence of state markers. Here a state of a link shadow consists in a choice of pointers from regions to vertices, with two adjacent regions selected without pointers (the stars):



In a state, each vertex receives exactly one pointer. The vertex weights are

$$\begin{aligned}\left\langle \text{Diagram} \mid \text{Diagram} \right\rangle &= t \\ \left\langle \text{Diagram} \mid \text{Diagram} \right\rangle &= -t^{-1} \\ \left\langle \text{Diagram} \mid \text{Diagram} \right\rangle &= 1 \\ \left\langle \text{Diagram} \mid \text{Diagram} \right\rangle &= 1\end{aligned}$$

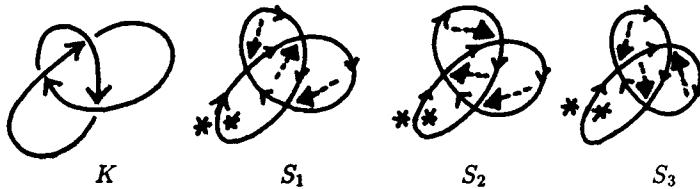
(For a negative crossing t and t^{-1} are interchanged.)

It is easy to prove (by direct comparison of state models) that

$$\nabla_K(2i) = i^{w(K)/2} \langle K \rangle (\sqrt{i})$$

where $\nabla_K = \sum_S \langle K | S \rangle$ is the Conway-Alexander polynomial using this state model, $w(K)$ is the writhe of K , and $\langle K \rangle$ is the bracket polynomial (Part I, section 3⁰).

Example.



$$\begin{aligned}\nabla_K &= \langle K | S_1 \rangle + \langle K | S_2 \rangle + \langle K | S_3 \rangle \\ &= (t)(-t^{-1}) + (-t^{-1})(-t^{-1}) + (t)(t) \\ &= -1 + t^{-2} + t^2 \\ \nabla_K &= (t - t^{-1})^2 + 1\end{aligned}$$

$$\begin{aligned}\nabla_K &= z^2 + 1 \\ \nabla_K(2i) &= -3.\end{aligned}$$

Exercise. Prove that

$$\nabla_K(2i) = i^{w(K)/2} \langle K \rangle (\sqrt{i})$$

where $w(K)$ denotes the writhe of K .

The point in regard to spacetime algebra is this: Let $P = \begin{smallmatrix} \nearrow & \searrow \\ \downarrow & \uparrow \end{smallmatrix}$, $Q = \begin{smallmatrix} \nearrow & \searrow \\ \uparrow & \downarrow \end{smallmatrix}$. Then $P^2 = P$, $Q^2 = Q$ and $PQ = QP = 0$, $P + Q = 1$ by the exclusion principles embedded in the formalism of this state model. Thus

$$A \begin{smallmatrix} \nearrow & \searrow \\ \downarrow & \uparrow \end{smallmatrix} - A^{-1} \begin{smallmatrix} \nearrow & \searrow \\ \uparrow & \downarrow \end{smallmatrix} = AP - A^{-1}Q \equiv [A, A^{-1}]$$

coincides with our spacetime algebra if we take P and Q to be the light cone generators

$$P = \frac{1 + \sigma}{2}, \quad Q = \frac{1 - \sigma}{2}$$

($\sigma^2 = 1$).

In summary, we have seen that elementary concepts of spacetime occur directly represented in the formalisms of the bracket (Jones) and Conway-Alexander polynomials. It will be useful to keep this phenomenon in mind as we construct the q -spin networks.

q -Spin Networks.

The first task is to generalize the antisymmetrizers of the standard theory. These are projection operators in the sense that

$$\begin{array}{c} \text{Diagram with } n \text{ strands crossing} \\ \hline \end{array} = n! \quad \begin{array}{c} \text{Diagram with } n \text{ strands crossing} \\ \hline \end{array}$$

and they kill off the non-identity generators of the Temperley-Lieb algebra

e_1, \dots, e_{n-1} :

$$\begin{array}{c} \text{Diagram with } n \text{ strands crossing} \\ \hline \end{array}_{e_1}, \dots, \begin{array}{c} \text{Diagram with } n \text{ strands crossing} \\ \hline \end{array}_{e_{n-1}}, \text{ since } \begin{array}{c} \text{Diagram with } n \text{ strands crossing} \\ \hline \end{array} = 0 \text{ by antisymmetry.}$$

These remarks mean that the expanded forms (using the binor identity) of the antisymmetrizers are special projection operators in the Temperley-Lieb algebra. For example

$$\begin{array}{c} \text{Diagram with } 3 \text{ strands crossing} \\ \hline \end{array} = \begin{array}{c} \text{Diagram with } 2 \text{ strands crossing} \\ \hline \end{array} + \begin{array}{c} \text{Diagram with } 1 \text{ strand crossing} \\ \hline \end{array}$$

Similarly,

$$\begin{array}{c} \text{Diagram with } 4 \text{ strands crossing} \\ \hline \end{array} = 6 \begin{array}{c} \text{Diagram with } 3 \text{ strands crossing} \\ \hline \end{array} + 4 \begin{array}{c} \text{Diagram with } 2 \text{ strands crossing} \\ \hline \end{array} + 4 \begin{array}{c} \text{Diagram with } 1 \text{ strand crossing} \\ \hline \end{array} + 2 \begin{array}{c} \text{Diagram with } 0 \text{ strands crossing} \\ \hline \end{array} + 2 \begin{array}{c} \text{Diagram with } -1 \text{ strands crossing} \\ \hline \end{array}$$

We shall now see that a suitable generalization produces these operators in the full Temperley-Lieb algebra.

We use the bracket identity $\mathcal{X} = A \begin{array}{c} \text{Diagram with } 1 \text{ strand crossing} \\ \hline \end{array} + A^{-1} \begin{array}{c} \text{Diagram with } -1 \text{ strands crossing} \\ \hline \end{array}$ in place of the binor identity and use loop value $\delta = -A^2 - A^{-2}$. A **q -spin network** is nothing more than a link diagram with special nodes that are designated as anti-symmetrizers (to be defined below). Thus any q -spin network computes its own bracket polynomial.

Define the q -symmetrizer $\begin{array}{c} \text{Diagram with } n \text{ strands crossing} \\ \hline \end{array}$ by the formula

$$\begin{array}{c} \text{Diagram with } n \text{ strands crossing} \\ \hline \end{array} = \sum_{\sigma \in S_n} (A^{-3})^{T(\sigma)} \quad \begin{array}{c} \text{Diagram with } n \text{ strands crossing} \\ \hline \end{array}_{\hat{\sigma}}$$

where $T(\sigma)$ is the minimal number of transpositions needed to return σ to the identity, and $\hat{\sigma}$ is a minimal braid representing σ with all negative crossings, that

is with all crossings of the form shown below with respect to the braid direction



Example 1.

$$\text{Diagram} = \left\| + \bar{A}^3 \text{Diagram} \right\| = \left\| + \bar{A}^{-3} [A \cup + \bar{A}^1] \right\|$$

Hence $\text{Diagram} = (1 + A^{-4}) \left[\text{Diagram} - \delta^{-1} \text{Diagram} \right]$. Here $\delta = -A^2 - A^{-2}$. We recognize $1 - \delta^{-1} e_1$ as the element f_1 of section 16⁰, Part I. It is the first Temperley-Lieb projector.

Define the quantum factorial $[n]!$ by the formula

$$[n]! = \sum_{\sigma \in S_n} (A^{-4})^{T(\sigma)}.$$

Exercise.

$$\begin{aligned} [n]! &= \sum_{\sigma \in S_n} (A^{-4})^{T(\sigma)} \\ \Rightarrow [n]! &= \prod_{k=1}^n \left(\frac{1 - A^{-4k}}{1 - A^{-4}} \right). \end{aligned}$$

In the example, note that $[2]! = 1 + A^{-4}$. Thus

$$\text{Diagram} = [2]! f_1.$$

Recall from section 16⁰, Part I that the Temperley-Lieb projectors are defined inductively via the equations

$$\left\{ \begin{array}{l} f_0 = 1 \\ f_{n+1} = f_n - \mu_{n+1} f_n e_{n+1} f_n \\ \mu_1 = 1/\delta \\ \mu_{n+1} = 1/(\delta - \mu_n) \end{array} \right\}.$$

Thus

$$f_1 = 1 - \delta^{-1} e_1$$

$$f_2 = 1 - \mu_2(e_1 + e_2) + \mu_1\mu_2(e_1e_2 + e_2e_1)$$

$$\mu_1 = 1/\delta$$

$$\mu_2 = \delta/(\delta^2 - 1).$$

Theorem 13.1. The Temperley-Lieb elements f_n with $\delta = -A^2 - A^{-2}$ are equivalent to the q -symmetrizers. In particular, we have the formula

$$f_{(n-1)} = \frac{1}{[n]!} \quad \text{Diagram of } f_{(n-1)}$$

Proof. f_n is characterized by $f_n^2 = f_n$, $e_i f_n = f_n e_i = 0$ $i = 1, \dots, n-1$. That the q -symmetrizer annihilates e_1, \dots, e_{n-1} follows as illustrated in the example below for $n=2$:

$$\text{Diagram} = \bigcup + A^{-3} \text{Diagram} = \bigcup + (A^{-3})(-A^3) \bigcup = 0.$$

This completes the proof. //

I have emphasized the construction of these generalized antisymmetrizers because they clarify knot theoretically the central ingredient for Lickorish's construction of three manifold invariants from the bracket - as described in section 16⁰, Part I. Thus these 3-manifold invariants are, in fact, q -spin network evaluations!

Example.

$$\begin{aligned} \text{Diagram} &= \text{Diagram} + x \text{Diagram} + x \text{Diagram} + x^3 \text{Diagram} \\ &\quad + x^2 \text{Diagram} + x^2 \text{Diagram}. \quad (x = A^{-3}) \end{aligned}$$

Hence

$$\text{Diagram} = [3]! \left[\text{Diagram} - \frac{\delta}{\delta^2 - 1} \left[\text{Diagram} + \text{Diagram} \right] + \frac{1}{\delta^2 - 1} \left[\text{Diagram} + \text{Diagram} \right] \right].$$

Exercise. Verify this formula by direct bracket expansion.

Now we can define the 3-vertex for q -spin nets just as before:

$$\begin{array}{c} a \\ | \\ b \text{---} c \end{array} \equiv \begin{array}{c} a \\ | \\ i \text{---} j \\ | \\ b \times \text{---} k \text{---} c \end{array}$$

$$\begin{array}{lcl} i + j & = & a \\ j + k & = & c \\ i + k & = & b \end{array}$$

with q -symmetrizers in place of antisymmetrizers.

The recoupling theory goes directly over, and this context can be used to define a theory of q -angular momentum and to create invariants of three manifolds. (See [LK19], [LK20]) for our work.

Exercise.

$$\begin{array}{c} a \\ | \\ b \text{---} c \end{array} = \epsilon A^{\left(\frac{b' + c' - a'}{2}\right)} \begin{array}{c} a \\ | \\ b \text{---} c \end{array}$$

$$\epsilon = (-1)^{\frac{b+c-a}{2}}$$

$$x' = x(x+z)$$

Exercise.

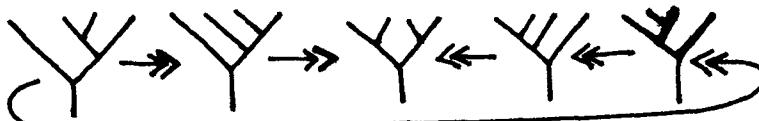
$$\begin{array}{c} a \\ | \\ b \text{---} c \\ | \\ a' \end{array} = \phi \quad \text{if } a \neq a'$$

$$\begin{array}{c} a \\ | \\ b \text{---} c \\ | \\ a' \end{array} = \left[\begin{array}{c} a \\ | \\ b \text{---} c \\ | \\ a' \end{array} \right] - \begin{array}{c} a \\ | \\ a' \end{array}$$

Exercise. Given that

$$\begin{array}{c} b \\ | \\ a \end{array} \quad \begin{array}{c} c \\ | \\ d \end{array} = \sum_{i,j} \left\{ \begin{matrix} a & b & i \\ c & d & j \end{matrix} \right\} \quad \begin{array}{c} b \\ | \\ a \\ | \\ i \\ | \\ c \\ | \\ d \end{array}$$

Show that (Elliot-Biedenharn Identity)



$$\left\{ \begin{matrix} b & c & j \\ d & f & i \end{matrix} \right\} \left\{ \begin{matrix} g & a & k \\ b & j & f \end{matrix} \right\} = \sum_{\ell} \left\{ \begin{matrix} g & a & \ell \\ i & d & f \end{matrix} \right\} \left\{ \begin{matrix} \ell & a & k \\ b & c & i \end{matrix} \right\} \left\{ \begin{matrix} k & c & j \\ d & g & \ell \end{matrix} \right\}.$$

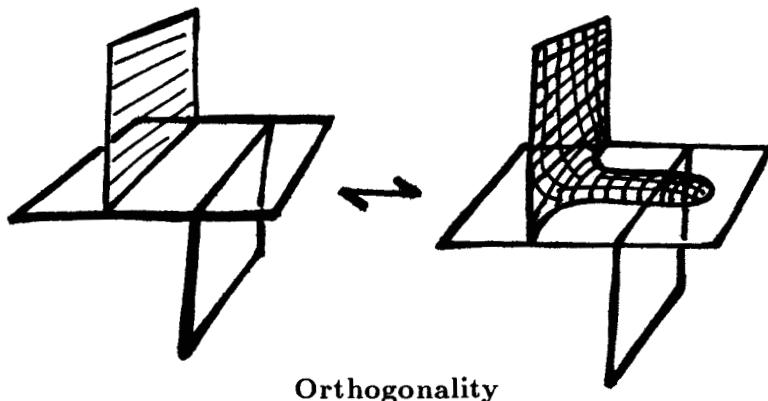
Exercise.

$$\sum_i \left\{ \begin{matrix} a & b & i \\ c & d & j \end{matrix} \right\} \left\{ \begin{matrix} d & \alpha & \kappa \\ b & c & i \end{matrix} \right\} = \delta_{j\kappa}$$

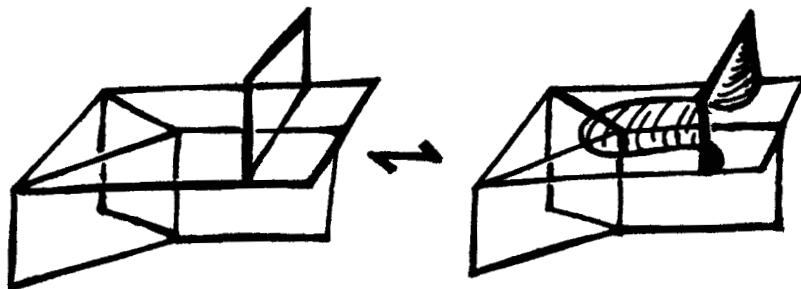
(orthogonality).

Exercise. Read [TU2].

In this paper Turaev and Viro construct invariants of 3-manifolds via a state summation involving the $q - 6j$ symbols. In [LK20] we (joint work with Sostenes Lins) show how to transfer their idea to an invariant based on the re-coupling theory of q -spin networks. The underlying topology for both approaches rests in the **Matveev moves** [MAT] shown below.



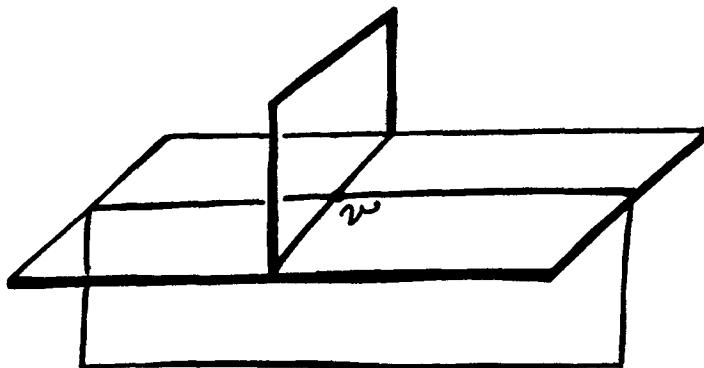
Orthogonality



Biedenharn-Elliott

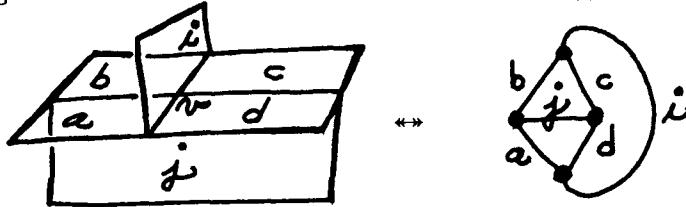
Matveev Moves

These are moves on special spines for three-manifolds that generate piecewise linear homeomorphisms of three-manifolds. In such a spine, a typical vertex appears as shown below with four adjacent one-cells, and six adjacent two-cells. Each one-cell abuts to three two-cells.

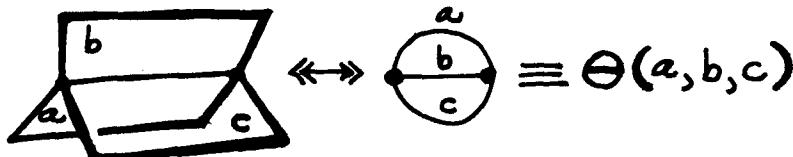


For an integer $r \geq 3$ the **color set** is $C(r) = \{0, 1, 2, \dots, r - 2\}$. A **state at level r** of the three-manifold M is an assignment of colors from $C(r)$ to each of the two-dimensional faces of the spine of M .

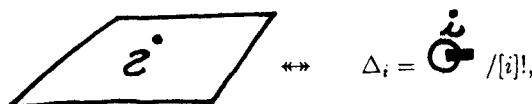
Given a state S of M , assign to each vertex the tetrahedral symbol whose edge colors are the face colors at that vertex as shown below:



Assign to each edge the $q - 3j$ symbol associated with its triple of colors



and assign to each face the Chebyshev polynomial $\Delta_i = (x^{i+1} - x^{-(i+1)})/(x - x^{-1})$ for $x = -A^2$.



whose index is the color of that face.

Then, for $A = e^{i\pi/2r}$, let $I(M, r)$ denote the sum over all states for level r , of the product of vertex evaluations and the face evaluations divided by the edge evaluations **for those evaluations that are admissible and non-zero**. Thus, if any edge has zero evaluation in a state S , then this state is dropped from the summation. For a given r , the edge evaluations are non-zero exactly when $a + b + c \leq 2r - 4$.

$$I(M, r) = \sum_S \frac{\prod_v \text{TET}(v, S) \prod_f \Delta_{S(f)}}{\prod_e \theta(S_a(e), S_b(e), S_c(e))}$$

Here $\text{TET}(v, s)$ denotes the tetrahedral evaluation associated with a vertex v , for the state S . $S(f)$ is the color assigned to a face f and $S_a(e), S_b(e), S_c(e)$ are the triplet of colors associated with an edge in the spine, and all faces and edges are contractible cells.

This is our [LK20] version of the Turaev-Viro invariant. It follows via the orthogonality and Elliot-Biedenharn identities that $I(M, r)$ is invariant under the Matveev moves. The q -spin nets provide an alternative and elementary approach to this subject.

In [TU5] Turaev has announced the following result.

Theorem (Turaev). Let $|M|_q$ denote the Turaev-Viro invariant, and $Z_q(M)$ denote the Reshetikhin-Turaev invariant for $SU(2)_q$. Then

$$|M|_q = Z_q(M) \overline{Z_q(M)}$$

where the bar denotes complex conjugation.

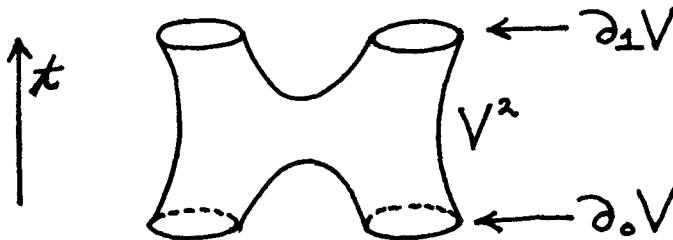
Since $Z_q(M)$ can be taken to be identical to the surgery invariant described in section 16⁰ of Part I, we encourage the (courageous) reader to prove that $I(M, r) = |Z_q(M, r)|^2$.

Exercise. Read [HAS]. These authors suggest a relationship between quantum gravity and the semi-classical limit of $6j$ symbols - hence the title “Spin networks are simplicial quantum gravity”. Investigate this paper in relation to q -spin networks. Compare with the ideas in [MOU] and [CR3].

14⁰. Knots and Strings – Knotted Strings.

It is natural to speculate about knotted strings. At this writing, such thoughts can only be speculation. Nevertheless, there are some pretty combinations of ideas in this domain, and they reflect on both knot theory and physics. (See [ROB], [ZA], [Z2].)

To begin with, the string as geometric entity is a surface embedded in spacetime. Suppose, for the purpose of discussion, that this spacetime is of dimension four. (I shall speculate a little later about associating knots with super strings in 10-space.) A surface embedded in dimension four can itself be knotted – for example, there exist knotted two dimensional spheres in Euclidean four space. However, the most interesting matter for topology is the consideration of a string interaction vertex embedded in spacetime. Abstractly the vertex is a sphere with four punctures, and is pictured as shown below



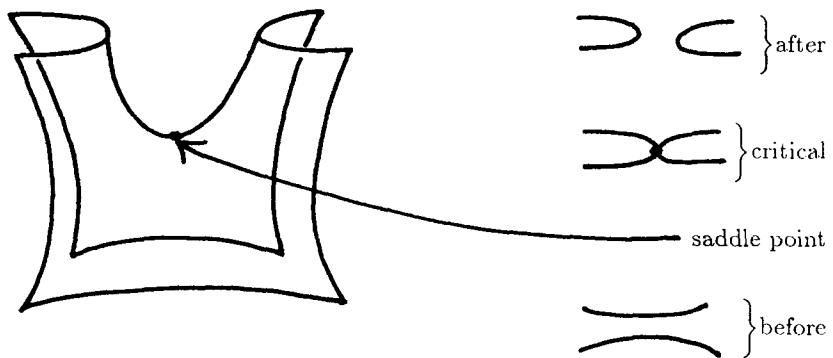
Thus we view the four-vertex as a process where two one-dimensional strings interact, tracing out a surface of genus zero with two punctures at $t = 0$ and two punctures at $t = 1$. Let V^2 denote this surface. Now, letting spacetime be $E^4 = \mathbf{R}^3 \times \mathbf{R} = \{(x, t)\}$, we consider embedding $F : V^2 \rightarrow E^4$ such that $\partial_0 V \hookrightarrow \mathbf{R}^3 \times 0$ and $\partial_1 V \hookrightarrow \mathbf{R}^3 \times 1$. Here $\partial_0 V$ and $\partial_1 V$ are indicated in the diagram above, each of these partial boundaries consists in two circles. The embedding is assumed to be smooth, and we can take this to mean that

- (i) All but finitely many levels $F(V^2) \cap \{(x, t) \mid t \text{ fixed}\} = F_t(V)$ are embeddings of a collection of circles (hence a link) in $\mathbf{R}_t^3 = \{(x, t) \mid x \in \mathbf{R}^3\}$.
- (ii) The critical levels involve singularities that are either **births** (minima), **deaths** (maxima) or **saddle points**.

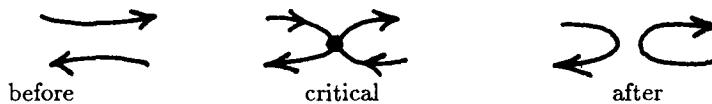
In a birth at level t_0 there is a sudden appearance of a point at level t_0 . The point becomes an unknotted circle in the levels immediately above t_0 . At a maximum or death point a circle collapses to a point and disappears from higher levels.



At a saddle point, two curves touch and rejoin as illustrated below.



We usually want to keep track of curve orientations. Thus the before, critical, after sequence for an oriented saddle appears as:

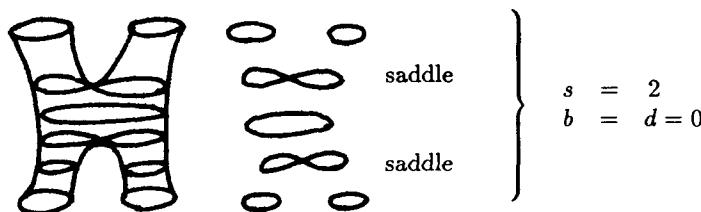


The requirement on the production of a standard vertex is that we

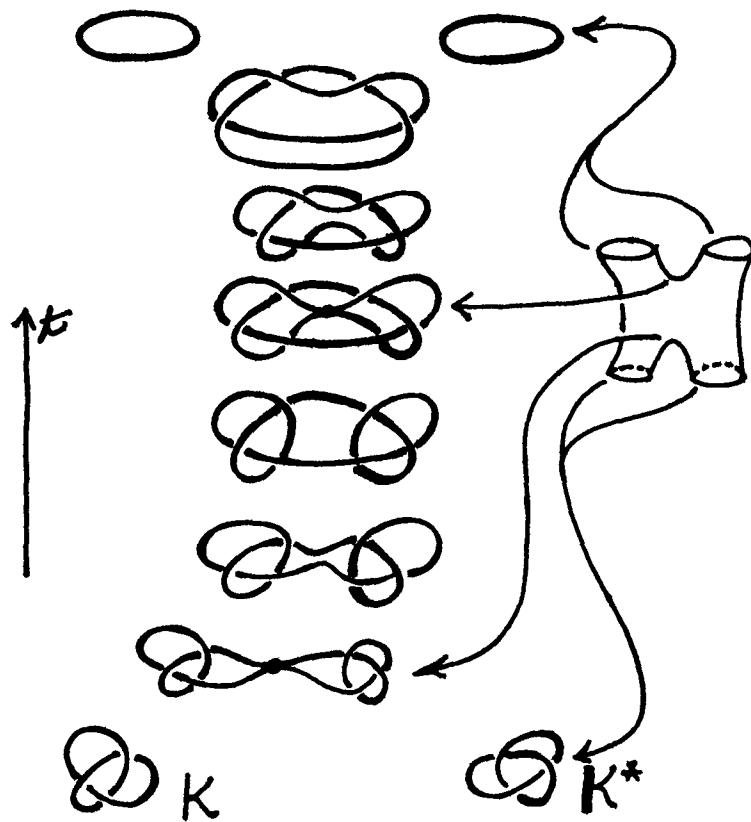
- (i) start with two knots
- (ii) allow a sequence of births, deaths, and saddles that ends in two knots.
- (iii) the genus of the surface so produced is zero.

(As long as the surface is connected, the last requirement is equivalent to the condition that $s = b + d + 2$ where s denotes the number of saddles, b denotes the number of births and d denotes the number of deaths.)

For example, for two unknotted strings the vertex look like



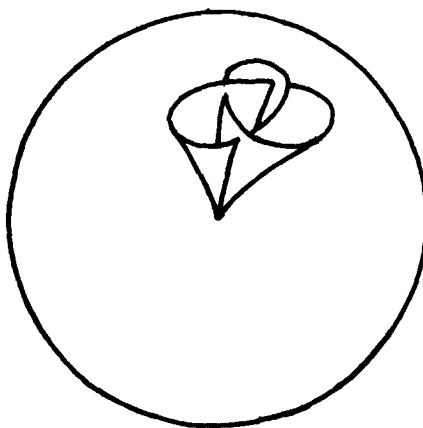
Now the remarkable thing is that two knotted strings can interact at such an embedded four vertex and produce a pair of entirely different knots. For example, the trefoil and its mirror image annihilate each other, producing two unknots.



In general a knot and its mirror image annihilate in this way. Some knots, called slice knots (see [FOX1]) can interact with an unknot and disappear. In general there is a big phenomena of interactions of such knotted strings. The corresponding topological questions are central to the study of the interface between three and four dimensions. This mathematical structure will be inherent in any theory of knotted strings.

Now the primary objection to knotted strings is that the superstring is necessarily living in a space of 10 dimensions, and there is no known theory for associating an embedding of the string in spacetime that is detailed enough to account for knotting. Furthermore, one dimensional curves and two dimensional surfaces are always unknotted in \mathbf{R}^{10} . The easy way out is to simply declare that it is useful to study string like objects in four-dimensional spacetime.

The difficult road is to find a way to incorporate knottedness with the superstring. Moving now to the department of unsupported speculation, here is an idea. Recall that many knots and links can be represented as neighborhood boundaries of algebraic singularities [MI]. For example, the right-handed trefoil knot is the neighborhood boundary of the singular variety $Z_1^2 + Z_2^3 = 0$ in \mathbf{C}^2 . That is $K = \{(z_1, z_2) \mid z_1^2 + z_2^3 = 0 \text{ and } |z_1|^2 + |z_2|^2 = 1\} = \mathcal{L}(K) = V(z_1^2 + z_2^3) \cap S^3$.



In this algebraic context, we can “suspend” K to a manifold of dimension 7 sitting inside the unit sphere in \mathbf{R}^{10} .

$$\Sigma^7(K) = \mathcal{V}(z_1^2 + z_2^3 + (z_3^2 + z_4^2 + z_5^2)) \cap S^9$$

$$\Sigma^7(K) \subset S^9 \subset \mathbf{R}^{10}.$$

It is tantalizing to regard $\Sigma^7(K) \subset \mathbf{R}^{10}$ as the “string”. One would have to rewrite the theory to take into account the vibrational modes of this 7-manifold associated with the knot. The manifold $\Sigma^7(K)$ has a natural action of the orthogonal group $O(3)$ (acting on (z_3, z_4, z_5)) and $\Sigma^7(K)/O(3) \cong D^4$ with $K \subset S^3 \subset D^4$ representing the inclusions of the orbit types. Thus we can recover spacetime (as D^4) and the knot from $\Sigma^7(K)$. This construction is actually quite general (see [LK]) and any knot $K \subset S^3$ has an associated $O(3)$ -manifold $\Sigma^7(K) \subset \mathbf{R}^{10}$. Thus this link-manifold [LK] construction could be a way to interface knot theory and the theory of superstrings.

The manifolds $\Sigma^7(K)$ are quite interesting. For example $\Sigma^7(3, 5)$ is the Milnor exotic 7-sphere [MI]. Here $(3, 5)$ denotes a torus knot of type $(3, 5)$ - that is a knot winding three times around the meridian of a torus while it winds five times around the longitude. Thus we are asking about the vibrational modes of exotic differentiable spheres in \mathbf{R}^{10} .

This mention of exoticity brings to mind another relationship with superstring theory. In [WIT4] Witten has investigated global anomalies in string theory that are associated with exotic differentiable structures on 10 and 11 dimensional manifolds. In [LK22] we investigate the relationship of these anomalies with specific constructions for exotic manifolds. These constructions include the constructions of link manifolds and algebraic varieties just mentioned, but in their 11-dimensional incarnations.

For dimension 10 the situation is different since exotic 10-spheres do not bound parallelizable manifolds. In this regard it is worth pointing out that such **very exotic n -spheres** are classified by $\pi_{n+k}(S^k)/\text{Im}(J)$ where $\pi_{n+k}(S^k)$ denotes a (stable) homotopy group of the sphere S^k and $\text{Im}(J)$ denotes the image of the J -homomorphism

$$J : \pi_n(SO(k)) \rightarrow \pi_{n+k}(S^k)$$

The very exotic spheres live in the darkest realms of homotopy theory. A full discussion of this story appears in [KV]. The J -homomorphism brings us back the Dirac string trick, for the simplest case of it is

$$J : \pi_1(SO(3)) \rightarrow \pi_4(S^3).$$

The construction itself is exactly parallel to our belt twisting geometry of section 11⁰, and here it constructs the generator of $\pi_4(S^3) \cong \mathbf{Z}_2$. To see this, we must define J . Let $g : S^n \rightarrow SO(k)$ represent an element of $\pi_n(SO(k))$. Then we get $g^* : S^n \times S^{k-1} \rightarrow S^{k-1}$ via $g^*(a, b) = g(a)b$. Associated with any map $H : X \times Y \rightarrow Z$ there is a map $\widehat{H} : X * Y \rightarrow S(Z)$ where $X * Y$ denotes the space of all lines joining points of X to points of Y , and $S(Z)$ denotes the suspension of Z obtained by joining Z to two points. In general, $S^n * S^{k-1} \cong S^{n+k}$ and $S(S^{k-1}) \cong S^k$. Thus g^* induces a map $\widehat{g^*} : S^{n+k} \rightarrow S^k$ representing an element in $\pi_{n+k}(S^k)$. This association of g with $\widehat{g^*}$ is the J -homomorphism.

In the case of $\pi_1(SO(3))$ we have $g : S^1 \rightarrow SO(3)$ representing the loop of rotations around the north pole of S^2 . The map

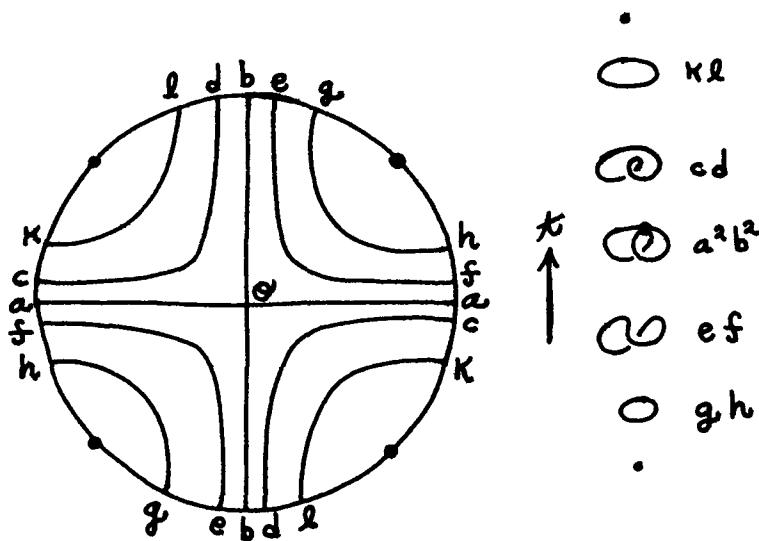
$$g^* : S^1 \times S^2 \rightarrow S^2$$

induces $\widehat{g^*} : S^1 * S^2 \rightarrow S(S^2)$ hence $\widehat{g^*} : S^4 \rightarrow S^3$. This element generates $\pi_4(S^3)$. In the belt trick, we used the same idea to a different end. $g : S^1 \rightarrow SO(3)$ gave rise to $g : I \rightarrow SO(3)$ where I denotes the unit interval and $g(0) = g(1)$. Then $g^* : I \times S^2 \rightarrow I \times S^2$ via $g^*(t, p) = (g(t)p, t)$ and we used g^* to twist a belt stretched between $S^2 \times 0$ and $S^2 \times 1$. The fact that the 720° twisted belt can be undone in $I \times S^2$ without moving the ends is actually a visualization of $\pi_4(S^3) \cong \mathbf{Z}_2$.

Elements of $\pi_{n+k}(S^k)$ in the image of the J -homomorphism can correspond to exotic spheres that bound parallelizable manifolds (such as the Milnor spheres in dimensions 7 and 11). Thus there is a train of relationships among the concepts of knotted strings, high dimensional anomalies, the belt trick and the intricacies of the homotopy groups of spheres.

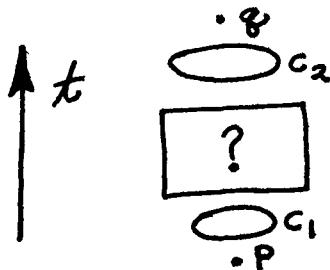
Exercise 14.1. Show that any knot K can interact with its mirror image K^* through an embedded genus zero string vertex in 3 + 1 spacetime to produce two unknots.

Exercise 14.2. The projective plane, \mathbf{P}^2 , is by definition the quotient space of the two-dimensional sphere S^2 ($S^2 = \{(x, y, z) \in \mathbf{R}^3 \mid x^2 + y^2 + z^2 = 1\}$) by the antipodal map $T : S^2 \rightarrow S^2$, $T(x, y, z) = (-x, -y, -z)$. Topologically this is equivalent to forming \mathbf{P}^2 from a disk D^2 by identifying antipodal points on the boundary of the disk. Examine the decomposition of \mathbf{P}^2 indicated by the figure below:

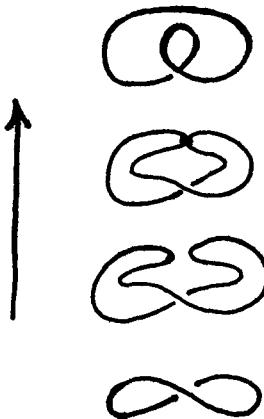


Compare: saddle point self-intersection

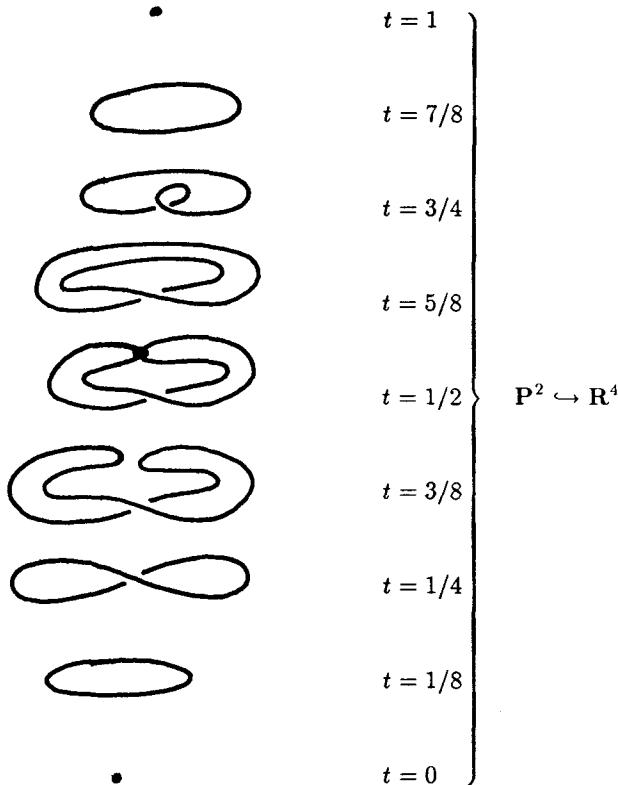
In this figure, points that are identified with one another on the disk boundary are indicated by the same letter. Therefore we see that the two arcs with endpoints c, d form a circle in \mathbb{P}^2 , as do the arcs with endpoints e, f . Similarly, the arcs labelled a, a and b, b each form circles. Thus we see that except for the points p and q the figure indicates a decomposition of \mathbb{P}^2 into circles all disjoint from one another except for the a and b circles at θ . We can therefore imagine \mathbb{P}^2 as described by a birth of the point p ; p grows to become a circle c_1 , c_1 undergoes a saddlepoint singularity to become yet another circle c_2 ; c_2 dies at q .



But how does a circle go through a saddle point with itself to produce a single circle? The answer lies in a twist:



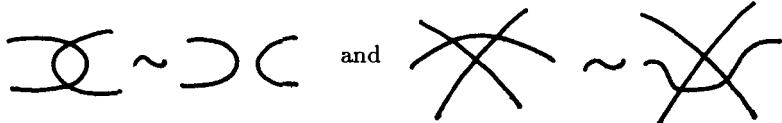
With this idea, we can produce an embedding of P^2 in R^4 as follows.



In this figure we mean each level to indicate the intersection $P^2 \cap (R^3 \times t)$ for $0 \leq t \leq 1$. At $t = 0$ there is a birth (minimum). As t goes from $1/4$ to $3/8$ the curve undergoes an ambient isotopy the curled form at $t = 3/8$. Note that the trace of this ambient isotopy in R^4 is free from singularities. The saddle point occurs at $t = 1/2$.

Corresponding to this embedding of P^2 in R^4 there is an immersion of P^2 in R^3 . An immersion of a surface into R^3 is a mapping that may have self-intersections of its image, but these are all locally transverse surface intersections modeled on the form of intersection of the $x - y$ plane and the $x - z$ plane in R^3 . We can exhibit an immersion by giving a series of levels, each showing an

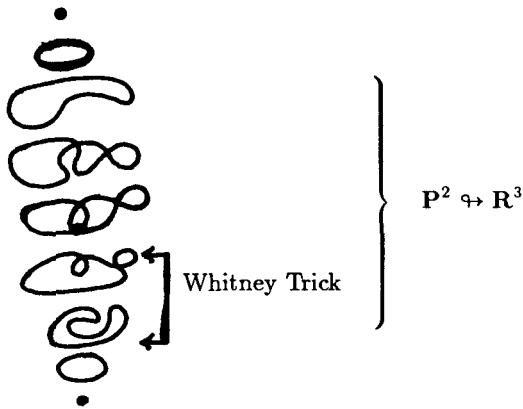
immersion of circles in a plane (possibly with singularities). The entire stack of levels then describes a surface in \mathbf{R}^3 . Instead of using ambient isotopy as we move from plane to plane, we use **regular homotopy**. Regular homotopy is generated by the projections of the Reidemeister II and III moves. Thus



are the basic regular homotopies. The Whitney trick is a fundamental example of a regular homotopy:



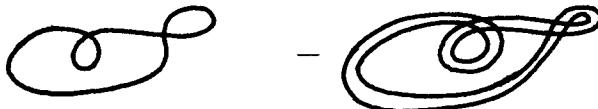
With the help of the Whitney trick we can obtain an immersion of \mathbf{P}^2 in \mathbf{R}^3 :



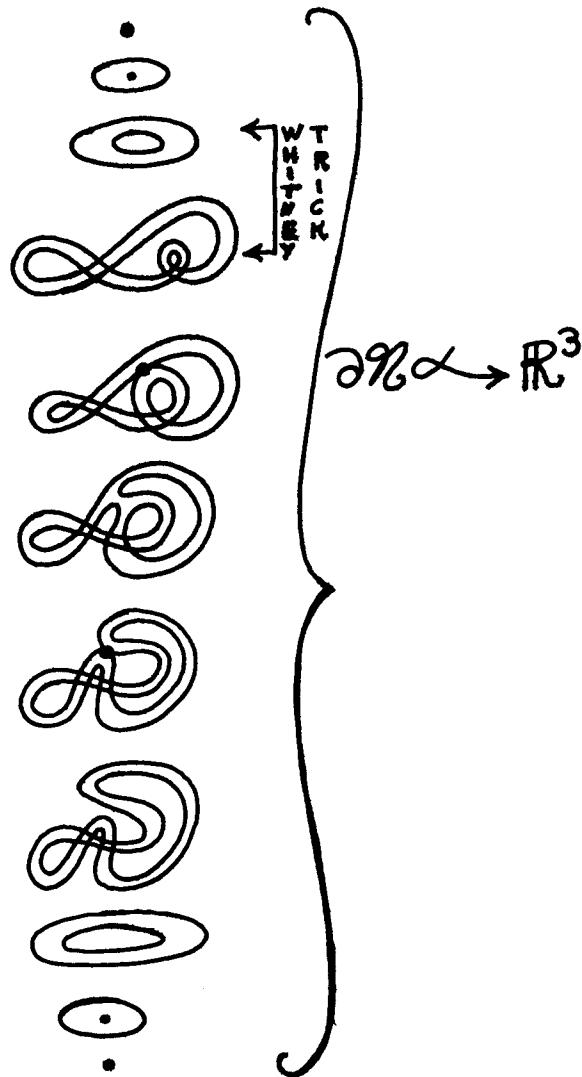
Problem. You might think that the immersion above could be covered by an embedding of \mathbf{P}^2 into \mathbf{R}^4 so that each level is obtained from its ancestors by either minima, maxima, saddle points or **regular isotopy**. Prove that this is not so.

The immersion shown above for $\mathbf{P}^2 \hookrightarrow \mathbf{R}^3$ has a normal bundle. This is the result of taking a neighborhood of $\mathbf{P}^2 \hookrightarrow \mathbf{R}^3$ consisting of small normals to the surface in \mathbf{R}^3 . We can construct this normal bundle \mathcal{N} by replacing each level

curve by a thickened version as in



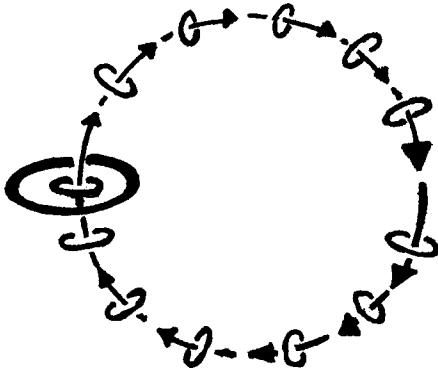
The boundary of \mathcal{N} , $\partial\mathcal{N}$, then appears as a sort of doubled version of the immersion of $P^2 \leftrightarrow \mathbf{R}^3$.



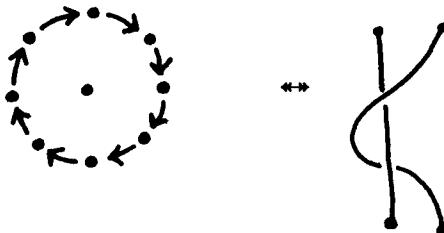
Problem. Show that the boundary of \mathcal{N} , $\partial\mathcal{N}$, is homeomorphic to a two sphere S^2 . Thus the induced immersion $\partial\mathcal{N} \hookrightarrow \mathbf{R}^3$ is an immersion of S^2 in \mathbf{R}^3 . Since the normal bundle has a symmetry (positive to negative normal directions) it is possible to evert this immersion $\alpha : S^2 \hookrightarrow \mathbf{R}^3$. In other words, you can easily turn this sphere inside out by exchanging positive and negative normals. It is not so obvious that the standard embedding of $S^2 \subset \mathbf{R}^3$ can be turned inside out through immersions and regular homotopy. ([FRA]) Nevertheless this is so. The eversion follows from the fact that the immersion $\alpha : S^2 \hookrightarrow \mathbf{R}^3$ described above is regularly homotopic to the standard embedding of S^2 . Prove this last statement by a set of pictures. Think about sphere eversion in the context of knotted strings by lifting the eversion into four-space.

Exercise 14.3. Investigate other examples of knotted string interactions, by finding pairs of knots that can undergo a 4-vertex interaction.

Exercise 14.4. (Motion Groups) Dahm and Goldsmith (See [G] and references therein.) generalize the Artin Braid Group to a **motion group** for knots and links in \mathbf{R}^3 . The idea of this generalization is based upon the braid group as fundamental group of a configuration space of distinct points in the plane. Just as we can think of braiding of points, we can investigate closed paths in the space of configurations of a knot or link in three space. The motion group is the fundamental group of this space of configurations. The simplest example of such a motion is obtained when one unlinked circle passes through another, as shown below:



This is the exact analog of a point moving around another point.



It turns out that the motion group for a collection of n disjoint circles in \mathbb{R}^3 is isomorphic to the subgroup of the group $\text{Aut}(n) = \text{Aut}(F(X_1, \dots, X_n))$ (the group of automorphisms of the free group on generators X_1, \dots, X_n) that is generated by (in the case of $n = 2$ - two strings) T, E and S where $T(x_1) = x_1^{-1}$, $T(x_2) = x_2$, $E(x_1) = x_2$, $E(x_2) = x_1$, $S(x_1) = x_2x_1x_2^{-1}$, $S(x_2) = x_2$. Imbo and Brownstein [IM2] calculate that the motion group for two unknotted strings in \mathbb{R}^3 is

$$\langle T, E, S \mid T^2 = E^2 = (TE)^4 = TST^{-1}S^{-1} = (TESE)^2 = 1 \rangle.$$

This has implications for exotic statistics ([IM1], [LK28]).

Problem: Prove this result and extend it to n strings.

Exercise. Read [JE] and make sense of these ideas (elementary particles as linked and knotted quantized flux) in the light of modern developments. (Compare with [RA].)

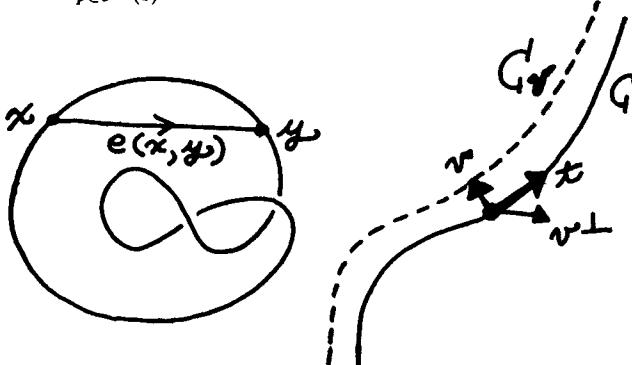
15⁰. DNA and Quantum Field Theory.

One of the most successful relationships between knot theory and DNA research has been the use of the formula of James White [W], relating the linking, twisting and writhing of a space curve. It is worth pointing out how these matters appear in a new light after linking and higher order invariants are re-interpreted a la Witten [WIT2] as path integrals in a topological quantum field theory. This section is an essay on these relationships. We begin by recalling White's formula.

Given a space curve C with a unit normal framing v, v^\perp and unit tangent t (v and v^\perp are perpendicular to each other and to t , forming a differentiably varying frame, $\langle v, v^\perp, t \rangle$, at each point of C .) Let C_ϵ be the curve traced out by the tip of ϵv for $0 < \epsilon \ll 1$. Let $Lk = Lk(C, C_\epsilon)$ be the linking number of C with this displacement C_ϵ . Define the **total twist**, Tw , of the framed curve C by the formula $Tw = \frac{1}{2\pi} \int v^\perp \cdot dv$. Given $(x, y) \in C \times C$, let $e(x, y) = (y - x)/|y - x|$ for $x \neq y$ and note that $e(x, y) \rightarrow t/|t|$ (for t the unit tangent vector to C at x) as x approaches y . This makes e well-defined on all of $C \times C$. Thus we have $e : C \times C \rightarrow S^2$. Let $d\Sigma$ denote the area element on S^2 and define the (spatial) **writhe** of the curve C by the formula

$$Wr = \frac{1}{4\pi} \int_{C \times C} e^* d\Sigma = \frac{1}{4\pi} \int_{z \in S^2} Cr(z) dz.$$

Here $Cr(z) = \sum_{p \in \epsilon^{-1}(z)} J(p)$ where $J(p) = \pm 1$ according to the sign of the Jacobian of e .



It is easy to see, from this description that the writhe coincides with the flat writhe (sum of crossing signs) for a curve that is (like a knot diagram) nearly embedded

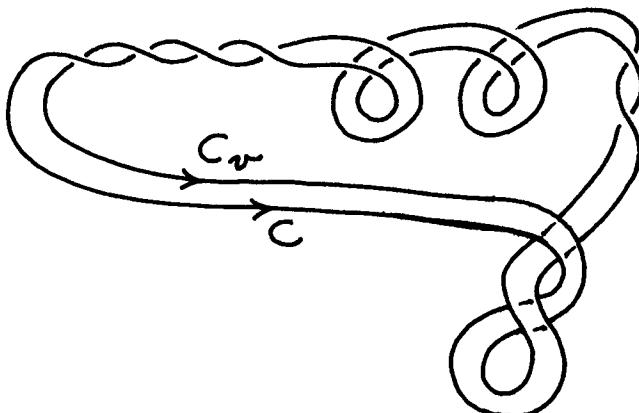
in a single plane.

With these definitions, White's theorem reads

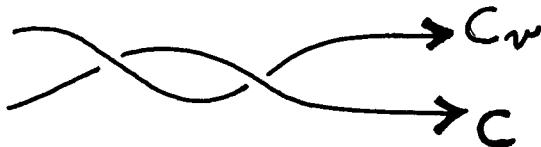
$$Lk = Tw + Wr.$$

This equation is fully valid for differentiable curves in three space. Note that the writhe only depends upon the curve itself. It is independent of the framing. By combining two quantities (twist and writhe) that depend upon metric considerations, we obtain the linking number - a topological invariant of the pair (C, C_v) .

The planar version of White's theorem is worth discussing. Here we have C and C_v forming a pair of parallel curves as in the example below:

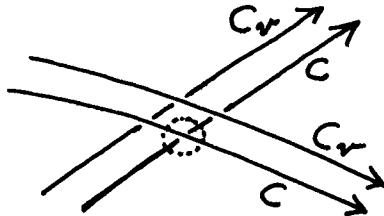


The twisting occurs between the two curves, and is calculated as the sum of $\pm(1/2)$ for each crossing of one curve with the other – in the form



The flat crossings also contribute to the total linking number, but we see that they

also catalogue the flat writhe of C . Thus

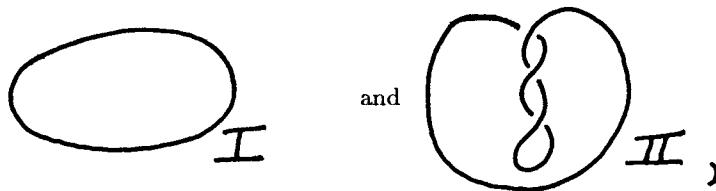


is regarded as a contribution of $(+1)$ to $w(C)$. Thus, we get

$$Lk(C, C_v) = Tw(C, C_v) + Wr(C)$$

where $Wr(C)$ is the sum of the crossing signs of C .

This simple planar formula has been invaluable in DNA research since photomicrographs give a planar view of closed double-stranded DNA. In a given electron micrograph the super-coiling, or writhe, may be visible while the twist is **unobservable**. If we could see two versions of the same closed circular DNA, then invariance of the linking number would allow a deduction of that number! For example, suppose that you observe



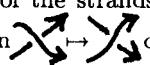
both representing the same DNA. In the first case $Wr_I = 0$. In the second case $Wr_{II} = -3$. Thus

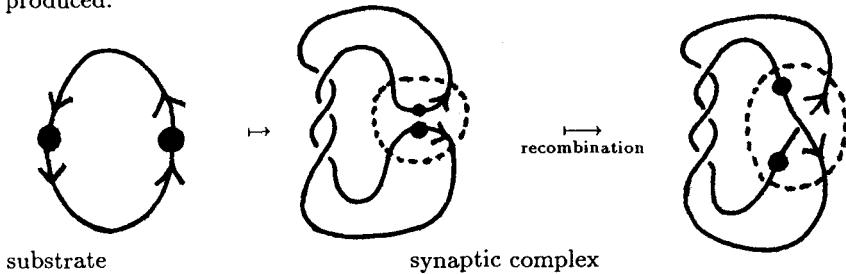
$$LK_I = Wr_I + Tw_I = Tw_I$$

$$LK_{II} = Wr_{II} + Tw_{II} = -3 + Tw_{II}.$$

Since $Lk_I = Lk_{II}$, we conclude that $Tw_{II} = Tw_I + 3$. Since I is relaxed, one can make an estimate based on the DNA geometry of the total twist. Let's say that $Tw_I = 500$. Then, we know that $Tw_{II} = 503$.

In this way the formula of White enables the DNA researcher to deduce quantities not directly observable, and in the process to understand some of the topological and energetical transformations of the DNA. The chemical environment of the DNA can influence the twist, thus causing the molecule to supercoil either in compensation for lowered twist (underwound DNA) or increase of twist (overwound DNA).

On top of this, there is the intricacy of the geometry and topology of recombination and the action of enzymes such as topoisomerase – that change the linking number of the strands ([BCW]). These enzymes perform the familiar switching operation  on interlinked strands of DNA by breaking one strand, and allowing the other strand to pass through it. In recombination a **substrate** with special sites on it twists into a **synaptic complex** with the sites for recombination adjacent to one another. Then in the simplest form of recombination the two sites are replaced by a crossover, and a new (possibly linked) double-stranded form is produced.



(In the biological terminology a link is called a catenane.) Of course, many different sorts of knots and links can be produced by such a process. In its simplest form, this movement from substrate to synaptic complex to product of the recombination takes the form of the combining of two tangles. Thus if

$$T = \boxed{\text{tangle } T}, \quad S = \boxed{\text{tangle } S}, \quad R = \boxed{\text{tangle } R}$$

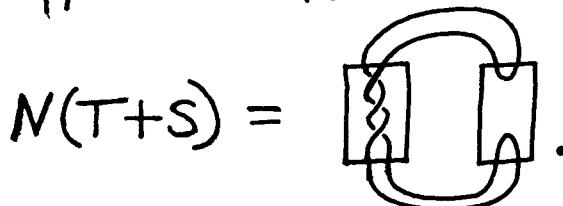
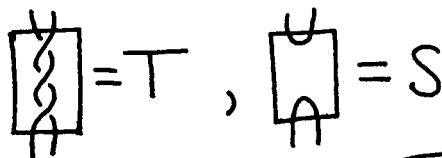
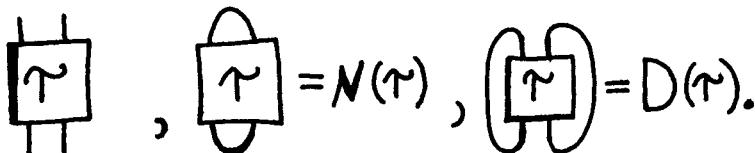
then

$$T + S = \boxed{T} \cup \boxed{S}$$

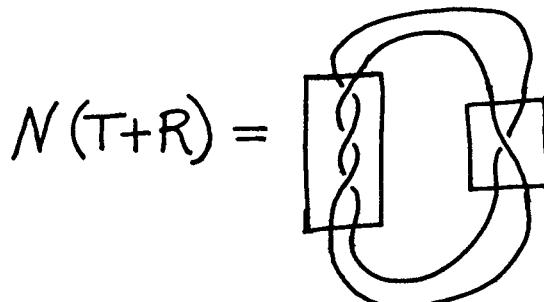
defines

tangle addition, and any four strand tangle τ has a **numerator** $N(\tau)$ and a

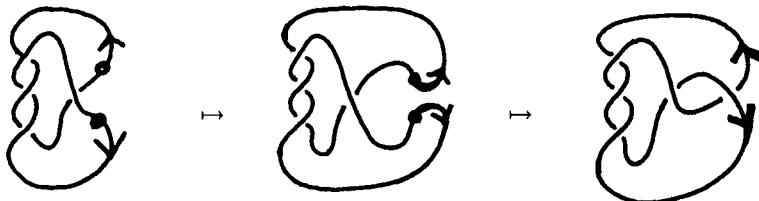
denominator $D(\tau)$:



This is the synaptic complex. The product of the recombination is



A second round of the recombination constitutes $N(T + R + R)$:



And a third round yields $N(T + R + R + R)$.



Thus we have

$$N(T + S) \cong \text{O} \quad \text{Unknot}$$

$$N(T + R) \cong \text{C} \quad \text{Hopf Link}$$

$$N(T + R + R) \cong \text{Figure Eight Knot}$$

$$N(T + R + R + R) \cong \text{Whitehead Link}$$

In experiments carried out with T_{n3} Resolvase [CSS] N. Cozarelli and S. Spengler found that just these products (Hopf Link, Figure Eight Knot, Whitehead Link) were being produced in the first three rounds of site specific recombination. In making a tangle-theoretic model of this process, D. W. Sumners [S] assumed that **T and R are enzyme determined constants independent of the variable geometry of the substrate S** . With this assumption, he was able to prove, using knot theory, that in order to obtain the products as shown above for the first three rounds, and assuming that $S = \text{---} \curvearrowright$, then

$$T = \boxed{\text{X}} \quad \text{and} \quad R = \boxed{\text{X}} .$$

This is a remarkable scientific application of knot theory to deduce the underlying mechanism of the biochemical interaction.

How did this result work? The key lay in Sumners' reduction of the problem to one involving only so-called **rational knots and links**, and then the construction of

a tangle calculus involving this class of links. The rational tangles are characterized topologically by values in the extended rational numbers $Q^* = Q \cup \{1/0 = \infty\}$. An element in Q^* has the form β/α where $\alpha \in N \cup \{0\}$, (N is the natural numbers), and $\beta \in \mathbb{Z}$ with $\gcd(\alpha, \beta) = 1$. Rational tangles themselves are obtained by iterating operations similar to the recombination process itself. Thus

$$\begin{array}{ccc} \text{Diagram A} & = & 0, \quad \text{Diagram B} & = & \infty \\ \text{Diagram C} & = & 1, \quad \text{Diagram D} & = & -1 \\ \text{Diagram E} & \Rightarrow & \text{Diagram F} & = & \text{Diagram G} \quad (\text{rotate by } \pi) \end{array}$$

The inverse of a tangle is obtained by turning it 180° around the left-top to right-bottom diagonal axis. Thus

$$\begin{array}{ccc} 2 & = & \text{Diagram H}, \quad 2^{-1} & = & \text{Diagram I} & = & \text{Diagram J} \\ (-1)^{-1} & = & \text{Diagram K} & = & \text{Diagram L} \end{array}$$

Rational numbers correspond to tangles via the continued fraction expansion. Thus

$$\begin{aligned} \frac{7}{3} &= 2 + \frac{1}{3} = \boxed{2} \quad \boxed{3^{-1}} \\ &= \text{Diagram M} \end{aligned}$$

Since two rational tangles are topologically equivalent if and only if they receive the same fraction in Q^* (See [ES].), it is possible to calculate possibilities for site specific recombination in this category. Here we have an arena in which molecular operations, knot theoretic operations and the **topological information** carried out by a knot or link are in good accord!

This brings us directly to the general question: **What is the nature of the topological information carried by a knot or link?** For biology this

information manifests itself in the history of a recombinant process, or in the architecture of the constituents of a cell. In this book, we have seen that such information is naturally enfolded in a quantum statistical framework. The most succinct statement of the generalized skein relations occurs for us in section 17⁰ of Part I, where - as a consequence of the generalized path integral

$$Z_K = \int dA e^{(ik/4\pi)\mathcal{L}} \langle K|A \rangle$$

where \mathcal{L} is the Chern-Simons Lagrangian, and $\langle K|A \rangle$ is the Wilson loop

$$\langle K|A \rangle = \text{tr} \left(P \exp \left(i \oint_K A \right) \right),$$

we find

where \mathcal{C} denotes an insertion of the Casimir operator for the representation of the Lie algebra in the gauge field theory. The basic change of information in switching a crossing involves the full framework of a topological quantum field theory. Yet conceptually this viewpoint is an enormous advance because it is a context within which one can begin to understand the nature of the information in the knot.

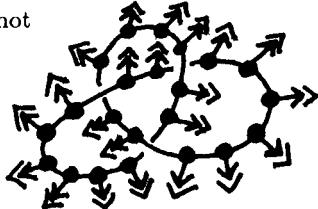
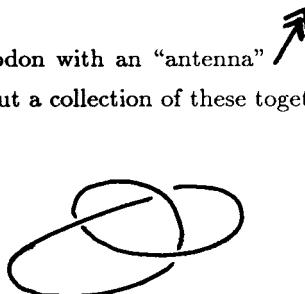
How does this viewpoint influence molecular biology? It may be too early to tell. The topological models used so far in DNA research are naive oversimplifications of the biology itself. We do not yet know how to relate the topology with the information structure and coding of DNA sequencing, let alone with the full complex of cellular architecture and interaction. Furthermore, biology occurs in the physical context of fields and the ever present quantum mechanical underpinnings of the molecules themselves.

In this regard, the Witten functional integral gives a powerful metaphor linking biology and quantum field theory. For remember the nature of the Wilson loop $\langle K|A \rangle = \text{tr}(P \exp(i \oint_K A))$. This quantity is to be regarded as the limit as the number of "detectors" goes to infinity of a product of matrices plucked from the gauge potential A (defined on the three dimensional space). Each detector takes

a matrix of the form $(1 + iAdx)$ at a spatial point x .



I like to think of each detector as a little codon with an “antenna” and legs (—○—) for plugging into nearby codons. Put a collection of these together into a knot

 K 

and you have an element, ready to measure the field and form $\langle K | A \rangle$. The interconnection is the product that computes the Wilson loop.

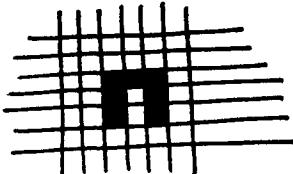
Just so, it may be that the circularly closed loops of knotted and catenated DNA are indeed Wilson loops – receivers of information from the enveloping biological field. If this is so, then knot theory, quantum field theory and molecular biology are fundamentally one scientific subject. Information is never carried only in a discrete mode. Why not regard DNA as the basic transceiver of the morphogenetic field? Each knotted form is a receptor for different sorts of field information, and the knots are an alphabet in the language of the field.

If these ideas seem strange in the biological context, they are not at all unusual for physics. For example, Lee Smolin and Carlo Rovelli ([RS], [LS2]) have proposed a theory of quantum gravity in which knots represent (via Wilson loop integration) a basis of quantum mechanical states in the theory.

Nevertheless, we should pursue the metaphor within the biological context. In this context I take a morphogenetic field to be a field composed of physical fields, but also fields of form and recursion. A simple example will help to elucidate this point of view. Consider John H. Conway’s cellular automaton the “Game of Life”. This is a two dimensional automaton that plays itself out on a rectangular lattice of arbitrary size. Squares in the lattice are either marked # or unmarked . The rules of evolution are:

1. $\blacksquare \rightarrow \square$ at less than two neighbors.
2. $\blacksquare \rightarrow \square$ at more than three neighbors.
3. $\square \rightarrow \blacksquare$ at exactly three neighbors.

The rules are applied simultaneously on the board at each time step. As a result, very simple initial configurations such as



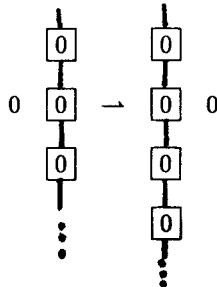
have long and complex lifetimes. At each step in the process a given square is (with its neighbor configuration) a detector of the morphogenetic field that is the rule structure for the automaton. The field is present everywhere on the board, it is a field of form, a non-numerical field. The particular local shapes on the board govern the evolutionary possibilities inherent in the field.

Another example arises naturally in Cymatics [JEN]. In Cymatics one studies the forms produced in materials under the influence of vibration - such as the nodal patterns of sand on a vibrating plate, or the form of a liquid surface under vibration. To an observer, the system appears to decompose into an organism of correlated parts, but these parts are all parts only to that observer, and are in reality, orchestrated into the whole through the vibratory field. Each apparent part is (to the extent that "it" is seen as a part) a locus of reception for information from the field, and a participant in the field of form and local influence generated in the material through the vibration.

One more example will help. This is the instantiation of autopoiesis as described by Varela, Maturana and Uribe in [VMU]. Here one has a cellular space occupied by codons, 0, marked codons, $\boxed{0}$, and catalysts, *. The marked codons can interlink: $\dots \rightarrow \boxed{0} \rightarrow \boxed{0} \rightarrow \boxed{0} \dots$. The catalyst facilitates marking via the combination of two codons to a marked codon: $0\ 0\ * \rightarrow \boxed{0}$. Marked codons may spontaneously decay

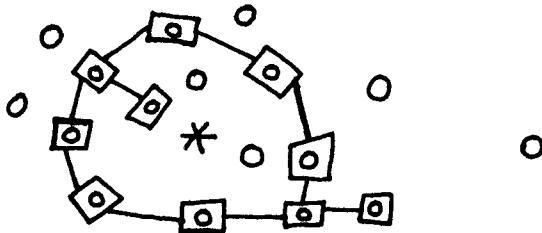
$$\boxed{0} \rightarrow 0\ 0$$

or lose or gain their linkages. Codons can pass through linkage walls:



but catalysts can **not** pass through a linkage.

With these rules, the system is subjected to random perturbations in a soup of codons, marked codons and catalysts. Once closed loops of marked codons enclosing a catalyst evolve, they tend to maintain a stable form



since the catalyst encourages the restoration of breaks in the loop, and has difficulty in escaping since it cannot cross the boundary that the loop creates.

In this example, we see in a microcosm how the cell may maintain its own structure through a sequence of productions that involve the very parts out of which it is constructed. The closing of the loop is essential for the maintenance of the form. It is through the closure of the loop that the cell has stability and hence a value in the system. The sensitivity of the closed loop is that of a unity that maintains its own form.

These examples give context to the possibility that there really is a deep relationship between the Jones polynomial – seen as a functional integral for topological quantum field theory in $SU(2)$ gauge – and the role of knotted DNA in molecular biology. $SU(2)$ can be related to local rotational structures (compare [BAT]) for bodies in three dimensional space (as we have discussed in relation to

the Dirac string trick) and hence there is the possibility of interpreting the Wilson loop in geometric terms that relate directly to biology. This is an idea for an idea.

To make one foray towards a concrete instantiation of this idea, consider the gauge theory of deformable bodies as described in [WIL]. Here one wants to determine the net rotation that results when a deformable body goes through a given sequence of unoriented shapes, in the absence of external forces. That is, given a path in the space of unlocated shapes (This is what a swimmer, falling cat or moving cell has under its control.), what is the corresponding path in the configuration space of located shapes? How does it actually move? In order to make axial comparisons between the unlocated and located shapes we have to give each unlocated shape a standard orientation, and then assign elements of $SO(3)$ relating the actual path to the abstract path. This ends up in a formulation that is precisely an $SO(3)$ gauge theory. For a given sequence of standard shapes $S_0(t)$, one wants a corresponding sequence of oriented shapes $S(t)$ that correspond to the real path of motion. These two paths are related by an equation

$$S(t) = R(t)S_0(t)$$

where $R(t)$ is a path in $SO(3)$. Computing $R(t)$ infinitesimally corresponds to solving the equation

$$\frac{dR}{dt} = R(t)A(t) \quad \left(A(t) = R^{-1}(t)\frac{dR}{dt} \right).$$

Once A is known, the full rotation $R(t)$ can be expressed as a path ordered exponential

$$R(t) = \mathbf{P} \exp\left(\int_{t_0}^t A(t)dt\right).$$

Wilczek defines a gauge potential over **the space of unlocated shapes** S_0 (also denoted by A) via

$$A_{\dot{S}_0}[S_0(t)] \equiv A(t).$$

That is, A is defined on the tangent space to S_0 - it is a vector field with a component for every direction in shape space - and takes values in the Lie algebra of $SO(3)$. As usual, the curvature tensor (or field strength)

$$F_{ij} = \frac{\partial A_i}{\partial w_j} - \frac{\partial A_j}{\partial w_i} + [A_i, A_j]$$

(w_i are a basis of tangent vectors to S_0) determines the rotations due to infinitesimal deformations of S_0 .

In specific cases, the gauge potential can be determined from physical assumptions about the system in question, and the relevant conservation laws. I have described this view of gauge theories for deformable bodies because it shows how close the formalisms of gauge theory, Wilson loops, and path ordered integration are to actual descriptions of the movement of animals, cells and their constituents. We can move to $SU(2)$ gauge by keeping track of orientation-entanglement relations (as in the Dirac string trick), and averages over Wilson loop evaluations then become natural for describing the statistics of motions. This delineates a broad context for gauge field theory in relation to molecular biology.

16⁰. Knots in Dynamical Systems – The Lorenz Attractor.

The illustration on the cover of this book indicates a **Lorenz knot**; that is, it represents a periodic orbit in the system of ordinary differential equations in three dimensional space discovered by Lorenz [LOR]. This system is, quite specifically,

$$X' = \sigma(Y - X)$$

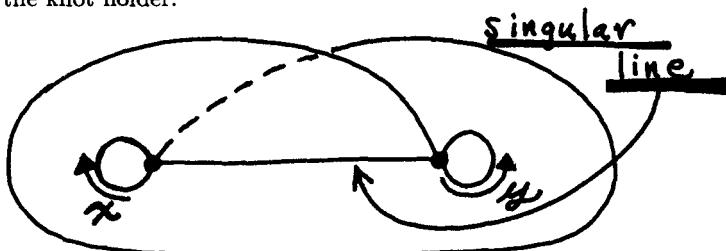
$$Y' = X(r - Z) - Y$$

$$Z' = XY - bZ$$

where σ , r and b are constants. (Lorenz used $\sigma = 10$, $b = 8/3$ and $r = 28$.) The constant r is called the **Reynolds number** of the Lorenz system.

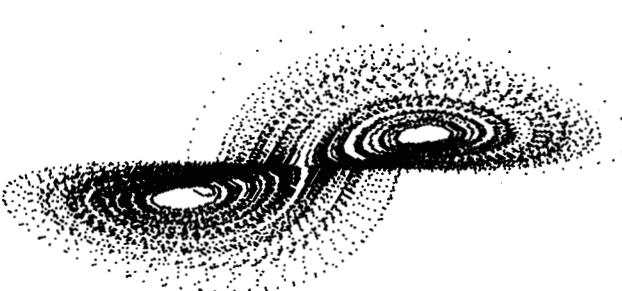
The Lorenz system is the most well-known example of a so-called chaotic dynamical system. It is extremely sensitive to initial conditions, and for low Reynolds number any periodic orbits (if they indeed exist) are unstable. Thus a generic orbit of the system wanders unpredictably, weaving its way back and forth between two unstable attracting centers. The overall shape of this orbit is stable and this set is often called the **Lorenz attractor**. Thus one may speak of orbits in the Lorenz attractor.

A qualitative analysis of the dynamics of the Lorenz attractor by Birman and Williams [B3] shows that periodic orbits (in fact orbits generally) should lie on a singular surface that these authors dub the “knot holder”. See Figure 16.2 for the general aspect of the Lorenz attractor - as a stereo pair, and Figure 16.1 for a picture of the knot holder.

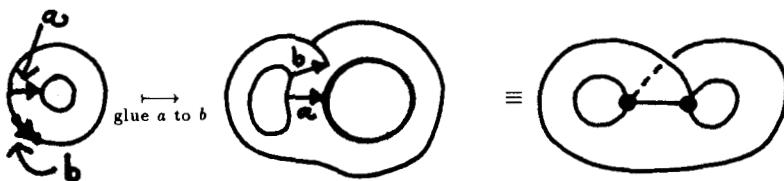


Knot Holder

Figure 16.1

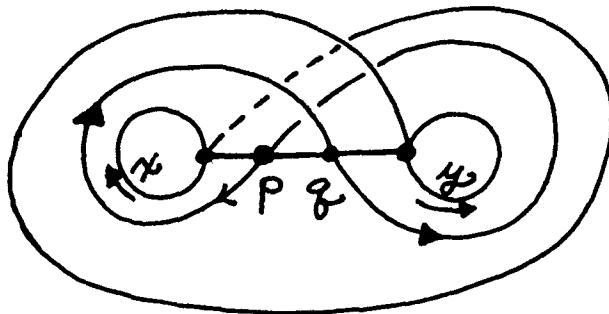
**Lorenz Attractor****Figure 16.2**

The knot-holder can be regarded as the result of gluing part of the boundary of an annulus to a line joining the inner and outer boundaries as shown below:



In Figure 16.1, we have indicated two basic circulations - x and y . The label x denotes a journey around the left-hand hole in the knot-holder, while y denotes a journey around the right-hand hole. Starting from the singular line, x connotes a journey on the upper sheet, while y will move on the lower sheet. If a point is told to perform x starting on the singular line, it will return to that line, and at first turn can return anywhere on the line.

According to the analysis of Birman and Williams, the orbit of a point in the Lorenz attractor is modeled by an infinite sequence of letters x and y . Thus $xy\ xy\ xy\ xy\dots$ denotes the simple periodic orbit shown below.



$$w = xy \ xy \ xy \ xy \dots$$

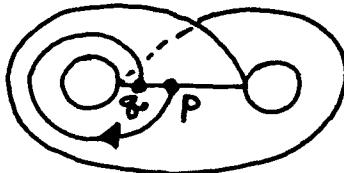
Already, many features of the general situation are apparent here. Note that this orbit intersects the singular line in points p and q and that from p the description is $yx \ yx \ yx \dots$. Thus we can write

$$w_p = xy \ xy \ xy \dots$$

$$w_q = yx \ yx \ yx \dots$$

Note that the list w_p, w_q is in lexicographic order if we take $x < y$ and that $p < q$ in the usual ordering of the unit interval homeomorphic with the singular line.

Are there any other orbits with this description $xy \ xy \ xy \dots$? Consider the experiment of doing x and then returning to a point $q < p$:



It is then not possible to perform y , since this would force the orbit to cross itself transversely (contradicting the well-definedness of the Lorenz vector field).

Therefore for $q < p$, we see that

$$w_p = xx^*$$

$$w_q = x^*$$

and $w_p \leq w_q$. If $w_p = w_q$ then $w_p = w_q = xxxx\dots$ – an infinite spiraling orbit – not realistic for the Lorenz system at $r = 28$, but possible in this combinatorics. Otherwise, we find (e.g.)

$$w_p = xxxx y \dots$$

$$w_q = xx y \dots$$

and $w_p < w_q$. Incidentally, we shall usually regard $xxx\dots$ as a simple periodic orbit encircling the left hole in the knot-holder.

Now, suppose we want a periodic orbit of the form

$$w = (xyxyy)(xyxyy)(xyxyy)\dots$$

Then this will cross the singular line in **five** points and these points will have descriptions corresponding to the five cyclic permutations of the word $xyxyy$:

$$1. \ xyxyy : w_1 = xyxyy \ xyxyy \dots$$

$$4. \ yxyyx : w_4 = yxyyx \ yxyyx \dots$$

$$2. \ xyyxy : w_2 = xyyxy \ xyyxy \dots$$

$$5. \ yyxyx : w_5 = yyxyx \ yyxyx \dots$$

$$3. \ yxyxy : w_3 = yxyxy \ yxyxy \dots$$

I have listed in order the cyclic permutations of $xyxyy$ and labelled each word with the integers 1,2,3,4,5 to indicate the lexicographic order of the words. The orbit must then pass through points p_1, p_2, p_3, p_4, p_5 on the singular line of the knot-holder so that w_i is the orbit description starting from p_i . We need (as discussed above for the simplest example) that $p_1 < p_2 < p_3 < p_4 < p_5$ since, in

lexicographic order, $w_1 < w_2 < w_3 < w_4 < w_5$. See Figure 16.3 for the drawing on the knot-holder.

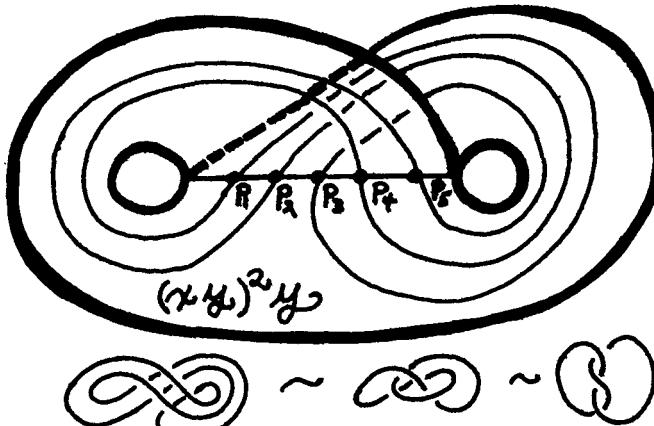


Figure 16.3

These considerations give rise to a simple algorithm for drawing a Lorenz knot corresponding to any non-periodic word in x and y .

- 1) Write an ordered list of the cyclic permutations of the word, by successively transferring the first letter to the end of the word.
- 2) Label the words in the list according to their lexicographic order.
- 3) Place the same labels in numerical order on the singular line of the knot-holder.
4. Starting at the top of the list, connect p_i to p_j by x or y – whichever is the first letter of the word labelled i where words i and j are successive words on the list.

We have done exactly this algorithm to form the knot in Figure 16.3 corresponding to $(xy)^2 y$. Note, as shown in this figure, that $(xy)^2 y$ is indeed knotted. It is a trefoil knot.

Now we can apply the algorithm to the knot on the cover of the book. It is shown as a stereo pair in Figure 16.4.

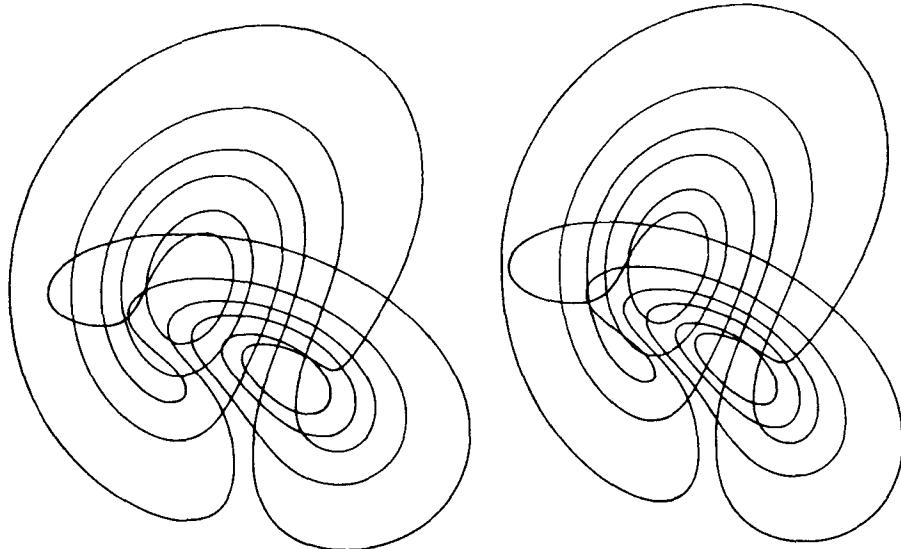


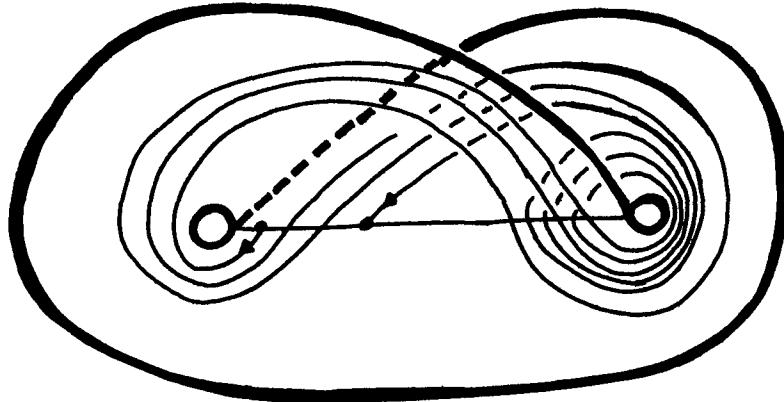
Figure 16.4

This knot is a computer generated Lorenz knot obtained by myself and Ivan Handler. Our method is accidental search. The knot is a periodic orbit for our computer at $r = 60$, $\sigma = 10$, $b = 8.3$. This is not a reproducible result! For a good discussion of the use of a generalization of Newton's method for the production of computer independent Lorenz knots, see [CUR]. Also, [JACK] has a good discussion of the generation of knots at high Reynolds number (say $r \approx 200$). In this range, one gets stable unknotted periodic orbits; and, as the Reynolds number is decreased, these bifurcate in a period-doubling cascade of stable - but hard to observe - knots.

Anyway, getting back to our knot in Figure 16.4, we see easily that it has the word x^3yxy^3xy .

Exercise 16.1. Apply the Lorenz knot algorithm to the word x^3yxy^3xy , and show that the corresponding knot is a torus knot of type $(3, 4)$.

Exercise 16.2. Draw an infinite orbit on the knot-holder with word $w = xyxyyxyyyx \dots$. Figure 16.5 shows the first few turns.



$$xyxy^2xy^3xy^4\dots$$

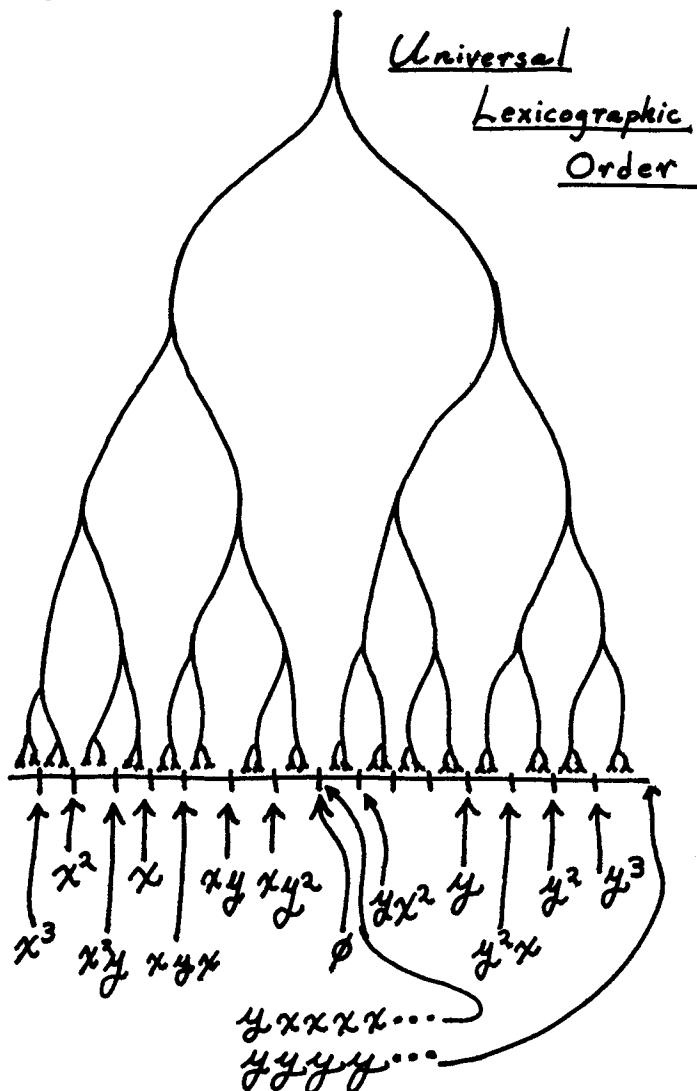
Figure 16.5

Remark. In regard to finite and infinite words, it is interesting to note that there is a correspondence between the lexicographic order of the partial trips on the knot-holder and the representations of numbers in the extended number system of John H. Conway [CON2]. To see this relationship, I use the representation of Conway numbers due to Martin Kruskal. For Kruskal, a Conway number is an ordinal sequence of x 's and y 's. Think of x as **left** and y as **right**. Then any word in x and y gives a sequence of “numbers” where

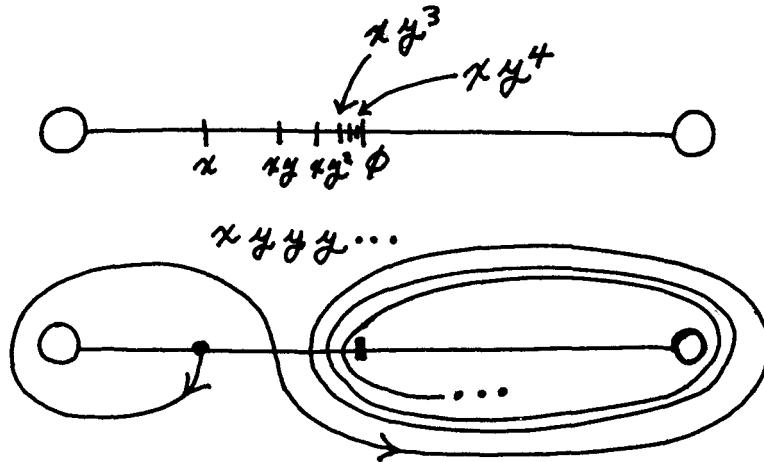
0	\leftrightarrow	\emptyset the empty word
+1	\leftrightarrow	y
+2	\leftrightarrow	yy
-1	\leftrightarrow	x
-2	\leftrightarrow	xx
	\dots	
$-\frac{1}{2}$	\leftrightarrow	xy

In general, x is taken as an instruction to step back midway between the last two created numbers, and y as an instruction to step forward. We always have that $\omega < \tau$ on the Conway line if ω precedes τ lexicographically.

Example 16.1.



Example 16.2. $xyyyyy\dots$ (This is a number infinitely close to zero.)



In general, given any infinite word $w = e_1e_2e_3e_4\dots$ where $e_i = x$ or y we get an infinite sequence of Conway numbers:

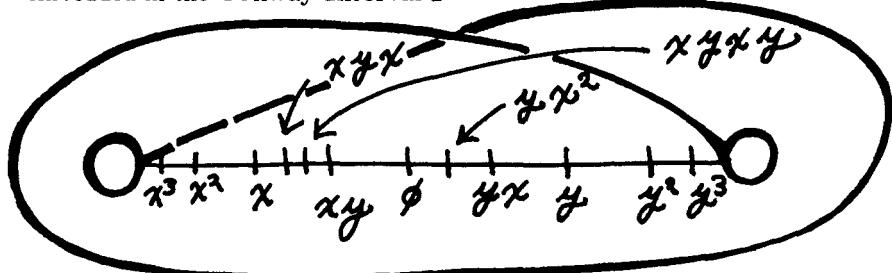
$$w_1 = e_1e_2e_3e_4\dots$$

$$w_2 = e_2e_3e_4\dots$$

$$w_3 = e_3e_4e_5\dots$$

...

embedded in the Conway Interval I



Define the **Extended Lorenz Holder** to be $S^1 \times I \sim$ where S^1 is a Conway circle, and \sim connotes the identification of the Conway annulus $S^1 \times I$ to form the singular line. Then for each w , the points w_1, w_2, w_3, \dots give the symbolic dynamics for a non-self intersecting orbit on the extended Lorenz Holder.

CODA.

In the beginning was the word.
The word became self-referential/periodic.
In the sorting of its lexicographic orders,
The word became topology, geometry and
The dynamics of forms;
Thus were chaos and order
Brought forth together
From the void.

“That from which creation came into being,
Whether it had held it together or it had not
He who watches in the highest heaven
He alone knows, unless...
He does not know.”

(Stanza 7 of **The Hymn of Creation**, RG VEDA 10.129 [DN]).

REFERENCES

- [AR] C. Agnes and M. Rasetti. Piecewise-linear category approach to entanglement. *II. Nuovo Cimento*, vol. 3D, No. 1 (1984), pp. 121–136.
- [AB] Y. Aharonov and D. Bohm. Significance of electromagnetic potentials in quantum theory. *Phys. Rev.*, **115**(1959), pp. 485-491.
- [AH] Y. Aharonov and L. Susskind. Observability of the Sign Change of Spinors under 2π Rotations. *Phys. Rev.*, vol. 158, No. 5. June 1967, pp. 1237.
- [AW1] Y. Akutsu and M. Wadati. Knot invariants and critical statistical systems. *J. Phys. Soc. Japan*, **56**(1987), 839-842.
- [AW2] Y. Akutsu and M. Wadati. Exactly solvable models and new link polynomials. I. N -state vertex models. *J. Phys. Soc. Japan*, **56**(1987), 3039-3051.
- [AW3] Y. Akutsu and M. Wadati. Knots, links, braids and exactly solvable models in statistical mechanics. *Comm. Math. Phys.*, **117**(1988), 243-259.
- [AW4] Y. Akutsu, A. Kuniba and M. Wadati. Virasoro algebra, von Neumann algebra and critical eight-vertex SOS models. *J. Phys. Soc. Japan*, vol. 55, no. 10. October 1986, pp. 3285-3288.
- [AW5] Y. Akutsu and M. Wadati. From solitons to knots and links. *Progress of Theoretical Physics Supplement*, no. 94 (1988), pp. 1-41.
- [ADW1] Y. Akutsu, T. Deguchi and M. Wadati. Exactly solvable models and new link polynomials II. Link polynomials for closed 3-braids. *J. Phys. Soc. Japan*, **56**(1987), 3464-3479.
- [ADW2] Y. Akutsu, T. Deguchi and M. Wadati. Exactly solvable and new link polynomials III. Two-variable polynomial invariants. *J. Phys. Soc. Japan*, **57**(1988), pp. 757-776.
- [ALEX1] J. W. Alexander. A lemma on systems of knotted curves. *Proc. Nat. Acad. Sci., USA* **9**(1923), pp. 93-95.
- [ALEX2] J. W. Alexander. Topological invariants of knots and links. *Trans. Amer. Math. Soc.*, **20**(1923), 275-306.
- [ASH] A. Ashtekar. **New perspectives in canonical gravity**. Bibliopolis (1988).
- [A1] M. F. Atiyah. **K-Theory**. W. A. Benjamin Inc. (1967).
- [A2] M. F. Atiyah. **Geometry of Yang-Mills Fields**. Accademia Nazionale dei Lincei Scuola Normale Superiore - Lezioni Fermiane. Pisa (1979).
- [A3] M. F. Atiyah. **The Geometry and Physics of Knots**. Cambridge University Press. (1990).

- [BE] H. J. Bernstein and A. V. Phillips. Fiber bundles and quantum theory. *Scientific American*, vol. 245, no. 1. July 1981, pp. 122-137.
- [BS] L. Baulieu and I. M. Singer. Topological Yang-Mills symmetry. *Nuclear Phys. B* (Proc. Suppl.) B(1988) pp. 12-19.
- [BA1] R. J. Baxter. **Exactly Solved Models in Statistical Mechanics**. Academic Press (1982).
- [BA2] R. J. Baxter. Chromatic polynomials of large triangular lattices. *J. Phys. A.: Math. Gen.*, **20**(1987), 5241-5261.
- [BN] D. Bar-Natan. Perturbative Chern-Simons Theory. (Preprint 1990).
- [BAT] E. P. Battey-Pratt and T. J. Racey. Geometric Model for Fundamental Particles. *Int. J. Theo. Phys.*, vol. 19, no. 6, (1980), pp. 437-475.
- [BCW] W. R. Bauer, F. H. C. Crick and J. H. White. Supercoiled DNA. *Scientific American*, vol. 243. July 1980, pp. 118-133.
- [BAY] B. F. Bayman. Theory of Hitches, *Amer. J. Physics*, vol. 45, no. 2, (1977), pp. 185-190.
- [B1] J. S. Birman and H. Wenzl. Braids, link polynomials and a new algebra. *Trans. AMS*, **313** (1989), pp. 249-273.
- [B2] J. S. Birman. **Braids, links and mapping class groups**. Annals of Math Study, no. 82. Princeton University Press (1974).
- [B3] J. S. Birman and R. F. Williams. Knotted Periodic Orbits in Dynamical Systems - I. Lorenz's Equations. *Topology*, vol. 22, no. 1, (1983), pp. 47-82.
- [B4] J. S. Birman. On the Jones polynomial of closed 3-braids. *Invent. Math.*, **81**(1985), pp. 287-294.
- [BO] H. Bondi. **Relativity and Common Sense**. Dover Pub. (1977).
- [BR] R. D. Brandt, W. B. R. Lickorish, K. C. Millett. A polynomial invariant for unoriented knots and links. *Invent. Math.*, **84**(1986), 563-573.
- [BUC] G. Buck. Knots as dynamical systems. (Preprint 1990.)
- [BZ] G. Burde and H. Zieschang. **Knots**. deGruyter (1986).
- [BUR] P. N. Burgoyne. Remarks on the combinatorial approach to the Ising problem. *J. Math. Phys.*, vol. 4, no. 10, Oct. 1963.
- [CLM] S. E. Cappell, R. Lee, E. Y. Miller. Invariants of 3-manifolds from conformal field theory. (Preprint 1990.)
- [CG] Casson, A. and C. M. Gordon. On slice knots in dimension three. In **Geometric Topology**, R. J. Milgram ed., pp. 39-53, *Proc. Symp. Pure Math.*, XXXII, Amer. Math. Soc., Providence, R. I.
- [CL] T. P. Cheng and L. F. Li. **Gauge Theory of Elementary Particle Physics**. Clarendon Press - Oxford (1988).

- [CH] G. F. Chew and V. Poenaru. Single-surface basis for topological particle theory. *Phys. Rev. D*, vol. 32, no. 10, pp. 2683-2697.
- [CON] J. H. Conway. An enumeration of knots and links and some of their algebraic properties. **Computational Problems in Abstract Algebra**. Pergamon Press, New York (1970), 329-358.
- [CON1] J. H. Conway. (Lecture at the University of Illinois at Chicago, Spring 1979.)
- [CON2] J. H. Conway. **On Numbers and Games**. Academic Press (1976).
- [COR] E. Corrigan, D. B. Fairlie, P. Fletcher and R. Sasaki. Some aspects of quantum groups and supergroups. (Preprint 1990.)
- [COTT] P. Cotta-Ramusino, E. Guadagnini, M. Martellini and M. Mintchev. Quantum field theory and link invariants. (Preprint 1989).
- [CSS] N. R. Cozzarelli, S. J. Spengler, A. Stasiak. The stereostructure of knots and catenanes produced by phase λ integrative recombination: implications for mechanism and DNA structure. *Cell* 42(1985), pp. 325-334.
- [CR1] L. Crane. Topology of three manifolds and conformal field theories. (Preprint 1989).
- [CR2] L. Crane. 2-D Physics and 3-D Topology. *Commun. Math. Phys.*, **135** (1991), 615-640.
- [CR3] L. Crane. Conformal Field Theory, Spin Geometry and Quantum Gravity. (Preprint 1990).
- [CF] R. H. Crowell, R. H. Fox. **Introduction to Knot Theory**. Blaisdell Pub. Co., (1963).
- [CUR] J. H. Curry. An algorithm for finding closed orbits. *Lecture Notes in Mathematics*, no. 819, Springer-Verlag (1979) pp. 111-120.
- [DN] A. T. de Nicolás. **Meditations through the Rg Veda**. Shambala Press (1978).
- [DIR] P. A. M. Dirac. **The Principles of Quantum Mechanics**. Oxford University Press (1958).
- [DRIN1] V. G. Drinfeld. Quantum Groups, *Proc. Intl. Congress Math.*, Berkeley, Calif. USA (1986), 789-820.
- [DRIN2] V. G. Drinfeld. Hopf algebras and the quantum Yang-Baxter Equation. *Soviet Math. Dokl.*, **32**(1985), no. 1.
- [ES] C. Ernst and D. W. Summers. A calculus for rational tangles: applications to DNA recombination. *Math. Proc. Camb. Phil. Soc.*, **108**(1990), pp. 489-515.
- [EL] B. Ewing and K. Millett. A load balanced algorithm for the calculation of the polynomial knot and link invariants. (Preprint 1989.)
- [FWu] C. Fan and F. Y. Wu. General lattice model of phase transitions. *Phys. Rev. B*, vol. 2, no. 3, (1970), pp. 723-733.

- [FRT] L. D. Faddeev, N. Yu Reshetikhin, L. A. Takhtajan. Quantization of Lie groups and Lie algebras. LOMI Preprint E-14-87, Steklov Mathematical Institute, Leningrad, USSR.
- [FR] R. A. Fenn and C. P. Rourke. On Kirby's calculus of links. *Topology*, **18**(1979), pp. 1-15.
- [FE] R. P. Feynman. **Theory of Fundamental Processes**. W. A. Benjamin Pub., (1961).
- [FE1] R. P. Feynman and S. Weinberg. **Elementary Particles and the Laws of Physics – The 1986 Dirac Memorial Lectures**. Cambridge University Press (1987).
- [FI1] D. Finkelstein. Kinks. *J. Math. Phys.*, **7**(1966), pp. 1218-1225.
- [FI2] D. Finkelstein. Space-Time Code. *Phys. Rev.*, **184**(1969), pp. 1261-1271.
- [FR] D. Finkelstein and J. Rubenstein. Connection between spin, statistics and kinks. *J. Math. Phys.*, **9**(1968), pp. 1762-1779.
- [FOX1] R. H. Fox and J. W. Milnor. Singularities of 2-spheres in 4-space and cobordism of knots. *Osaka J. Math.*, **3**, 257-267.
- [FOX2] R. H. Fox. A quick trip through knot theory. In **Topology of 3-Manifolds**. Ed. by M. K. Fort Jr., Prentice-Hall (1962), pp. 120-167.
- [FRA] G. K. Francis. **A Topologists Picturebook**. Springer-Verlag (1987).
- [FW] J. Franks and R. Williams. Braids and the Jones polynomial. *Trans. Amer. Math. Soc.*, **303** (1987), pp. 97–108.
- [FG] D. S. Freed and R. E. Gompf. Computer calculation of Witten's 3-Manifold Invariant. (Preprint 1990).
- [F] P. Freyd, D. Yetter, J. Hoste, W. B. R. Lickorish, K. C. Millett and A. Ocneanu. A new polynomial invariant of knots and links. *Bull. Amer. Math. Soc.*, **12**(1985), 239-246.
- [FQS] D. Friedan, Z. Qui, S. Shenker. Conformal invariance, unitarity and two-dimensional critical exponents. In **Vertex operators in Mathematics and Physics**. Ed. by J. Lepowsky, S. Mandelstam, and I. M. Singer, pp. 419-450, (1985).
- [FRO] J. Frohlich. Statistics of fields, the Yang-Baxter Equation and the theory of knots and links. (Preprint 1987).
- [FROM] J. Frohlich and P. Marchetti. Quantum field theories of vortices and anyons. (Preprint 1988).
- [FU] B. Fuller. Decomposition of the linking number of a closed ribbon: a problem from molecular biology. *Proc. Natl. Acad. Sci. USA*, vol. 75, no. 8. pp. 3557-3561 (1978).
- [GA] K. Gawedski. Conformal Field Theory. Sem. Bourbaki. 41e année (1988-89), 704, pp. 1-31.

- [GE] M. L. Ge, N. Jing and Y. S. Wu. New quantum group associated with a “non-standard” braid group representation. (Preprint 1990.)
- [GL] M. L. Glasser. Exact partition function for the two-dimensional Ising model. *Amer. J. Phys.*, vol. 38, no. 8, August 1970, pp. 1033-1036.
- [G1] D. L. Goldsmith. The Theory of Motion Groups. *Michigan Math. J.*, 28(1981), pp. 3-17.
- [G2] D. L. Goldsmith. Motion of Links in the Three Sphere, *Math. Scand.*, 50(1982), pp. 167-205.
- [GR] B. Grossman. Topological quantum field theories: relations between knot theory and four manifold theory. (Preprint 1989.)
- [GMM] E. Guadagnini, M. Martellini and M. Mintchev. Perturbative aspects of the Chern-Simons field theory. *Phys. Lett.*, B227 (1989), p. 111.
- [HKC] P. de la Harpe, M. Kervaire and C. Weber. On the Jones polynomial. *L'Enseign. Math.*, 32(1986), 271-335.
- [HA] R. Hartley. Conway potential functions for links. *Comment. Math. Helv.*, 58(1983), pp. 365-378.
- [HAS] B. Hasslacher and M. J. Perry. Spin networks are simplicial quantum gravity. *Phys. Lett.*, vol. 103B, no. 1, July (1981), pp. 21-24.
- [HAY] M. V. Hayes. **A Unified Field Theory**. The Stinehour Press. (1964).
- [HE1] M. A. Hennings. A polynomial invariant for oriented banded links. (Preprint 1989.)
- [HE2] M. A. Hennings. A polynomial invariant for unoriented banded links. (Preprint 1989.)
- [HE3] M. A. Hennings. Hopf algebras and regular isotopy invariants for link diagrams. *Math. Proc. Cambridge Phil. Soc.* **109** (1991), 59-77.
- [HE4] M. A. Hennings. Invariants of links and 3-manifolds obtained from Hopf algebras. (To appear.)
- [HO] C. F. Ho. A new polynomial invariant for knots and links - preliminary report. AMS Abstract, vol. 6, no. 4, Issue **39**(1985), 300.
- [HOP] H. Hopf. Über die abbildungen der drei dimensionalen Sphäre die Kugelfläche. *Math. Ann.*, **104**(1931), pp. 637-665.
- [HOS1] J. Hoste. A polynomial invariant of knots and links. *Pacific J. Math.*, **124**(1986), 295-320.
- [HOS2] J. Hoste and M. Kidwell. Dichromatic link invariants. *Trans. A.M.S.*, vol. 321, no. 1, (1990), pp. 197-229.
- [HOS3] J. Hoste and J. H. Przytycki. Homotopy skein modules of orientable 3-manifolds. *Math. Proc. Camb. Phil. Soc.*, **108**(1990), pp. 475-488.

- [HU] A. Hurwitz. "Über die Composition der quadratischen Formen von Beliebig vielen Variablen," **Nachrichten von der königlichen Gesellschaft der Wissenschaften in Göttingen** (1898), pp. 309-316; *Mathematische Werke*, vol. 11, pp. 565-571. Basel, 1932.
- [HW] L. P. Hughston and R. S. Ward. **Advances in Twistor Theory**. Pitman (1979).
- [IM1] T. D. Imbo, C. S. Imbo, R. S. Mahajan and E. C. G. Sudarshan. A User's Guide to Exotic Statistics. (Preprint 1990).
- [IM2] T. D. Imbo and R. Brownstein. (private communication).
- [IM3] T. D. Imbo, C. S. Imbo and E. C. G. Sudarshan. Identical particles, exotic statistics and braid groups. (Preprint 1990).
- [JACI] R. Jackiw. Topological investigations of quantized gauge theories. In **Current Algebra and Anomalies**. Ed. by S. B. Treiman, R. Jackiw, B. Zumino, E. Witten. (1985), World Sci. Pub., pp. 211-340.
- [JACK] E. A. Jackson. **Perspectives of Nonlinear Dynamics Vols. I and II**. Cambridge University Press (1990).
- [JR] G. Jacucci and M. Rasetti. Exploring the use of artificial intelligence, logic programming, and computer-aided symbolic manipulation in computational physics I. 1. The mathematical structure of phase transition theory. *J. Phys. Chem.*, **91**(1987), pp. 4970-4980.
- [JA1] F. Jaeger. A combinatorial model for the Homfly polynomial. (Preprint 1988).
- [JA2] F. Jaeger. On Tutte polynomials and cycles of plane graphs. *J. Comb. Theo. (B)*, **44**(1988), pp. 127-146.
- [JA3] F. Jaeger. On transition polynomials of 4-regular graphs. (Preprint 1987).
- [JA4] F. Jaeger. Composition products and models for the Homfly polynomial. *L'Enseignement Math.*, t35, (1989), pp. 323-361.
- [JA5] F. Jaeger. Tutte polynomials and link polynomials. *Proc. Amer. Math. Soc.*, **103** (1988), pp. 647-654.
- [JA6] F. Jaeger. On edge colorings of cubic graphs and a formula of Roger Penrose. *Ann. of Discrete Math.*, **41**(1989), pp. 267-280.
- [JA7] F. Jaeger, D. L. Vertigan and D. J. A. Welsh. On the computational complexity of the Jones and Tutte polynomials. *Math. Proc. Camb. Phil. Soc.*, **108**(1990), pp. 35-53.
- [JA8] F. Jaeger, L. H. Kauffman and H. Saleur. (to appear)
- [JA9] F. Jaeger. (Announcement 1989.)
- [JE] H. Jehle. Flux quantization and particle physics. *Phys. Rev. D.*, vol. 6, no. 2 (1972), pp. 441-457.

- [JEN] H. Jenny. **Cymatics**, Basel: Basilius Presse, (1974).
- [JI1] M. Jimbo. A q -difference analogue of $U(q)$ and the Yang-Baxter Equation. *Lect. in Math. Physics*, **10**(1985), 63-69.
- [JI2] M. Jimbo. Quantum R -matrix for the generalized Toda system. *Comm. Math. Phys.*, **102**(1986), 537-547.
- [JO1] V. F. R. Jones. A new knot polynomial and von Neumann algebras. *Notices of AMS*, **33**(1986), 219-225.
- [JO2] V. F. R. Jones. A polynomial invariant for links via von Neumann algebras. *Bull. Amer. Math. Soc.*, **129**(1985), 103-112.
- [JO3] V. F. R. Jones. Hecke algebra representations of braid groups and link polynomials. *Ann. of Math.*, **126**(1987), 335-388.
- [JO4] V. F. R. Jones. On knot invariants related to some statistical mechanics models. *Pacific J. Math.*, vol. 137, no. 2, (1989), 311-334.
- [JO5] V. F. R. Jones. On a certain value of the Kauffman polynomial. (To appear in *Comm. Math. Phys.*).
- [JO6] V. F. R. Jones. Subfactors and related topics. In **Operator algebras and Applications**, vol. 2. Edited by D. E. Evans and M. Takesaki. Cambridge University Press (1988), pp. 103-118.
- [JO7] V. F. R. Jones. Index for subfactors. *Invent. Math.*, **72**(1983), pp. 1-25.
- [JO8] V. F. R. Jones. (Private communication.)
- [JS] A. Joyal and R. Street. Braided monoidal categories. Macquarie reports 86008 (1986).
- [JOY] D. Joyce. A classifying invariant of knots, the knot quandle. *J. Pure Appl. Alg.*, **23**, 37-65.
- [KVA] L. H. Kauffman and F. J. Varela. Form dynamics. *J. Social and Bio. Struct.*, **3**(1980), pp. 171-206.
- [LK] L. H. Kauffman. Link Manifolds. *Michigan Math. J.*, **21**(1974), pp. 33-44.
- [LK1] L. H. Kauffman. The Conway polynomial. *Topology*, **20**(1980), 101-108.
- [LK2] L. H. Kauffman. **Formal Knot Theory**. Princeton University Press, Mathematical Notes #30 (1983).
- [LK3] L. H. Kauffman. **On Knots**. Annals of Mathematics Studies **115**, Princeton University Press (1987).
- [LK4] L. H. Kauffman. State Models and the Jones Polynomial. *Topology*, **26**(1987), 395-407.
- [LK5] L. H. Kauffman. Invariants of graphs in three-space. *Trans. Amer. Math. Soc.*, vol. 311, **2** (Feb. 1989), 697-710.
- [LK6] L. H. Kauffman. New invariants in the theory of knots. (Lectures given in Rome, June 1986). – *Asterisque*, **163-164** (1988), pp. 137-219.

- [LK7] L. H. Kauffman. New invariants in the theory of knots. *Amer. Math. Monthly*, vol. 95, no. 3, March 1988, pp. 195-242.
- [LK8] L. H. Kauffman. An invariant of regular isotopy. *Trans. Amer. Math. Soc.*, vol. 318, no. 2, (1990), pp. 417-471.
- [LK9] L. H. Kauffman. Statistical Mechanics and the Alexander polynomial. *Contemp. Math.*, **96**(1989), Amer. Math. Soc. Pub., pp. 221-231.
- [LK10] L. H. Kauffman. Statistical mechanics and the Jones polynomial. (Proceedings of the 1986 Santa Cruz conference on Artin's Braid Group. AMS Contemp. Math. Series (1989), vol. 78, pp. 263-297. Reprinted in **New Problems and Methods and Techniques in Quantum Field Theory and Statistical Mechanics**. pp. 175-222. Ed. by M. Rasetti. World Scientific Pub. (1990).)
- [LK11] L. H. Kauffman. A Tutte polynomial for signed graphs. *Discrete Applied Math.*, **25**(1989), pp. 105-127.
- [LK12] L. H. Kauffman. State models for knot polynomials - an introduction. (In the Proceedings of the meeting of the Brasilian Mathematical Society - July 1987.)
- [LK13] L. H. Kauffman. State models for link polynomials. *l'Enseignement Mathematique*, t. 36 (1990), pp. 1-37.
- [LK14] L. H. Kauffman. Knot polynomials and Yang-Baxter models. **IXth International Congress on Mathematical Physics – 17-27 July 1988 - Swansea, Wales**, Ed. Simon, Truman, Davies, Adam Hilger Pub. (1989), pp. 438-441.
- [LK15] L. H. Kauffman. Polynomial Invariants in Knot Theory. **Braid Group, Knot Theory and Statistical Mechanics**. Ed. C. N. Yang and M. L. Ge. *World Sci. Pub. Advanced Series in Mathematical Physics*, **9**(1989), 27-58.
- [LK16] L. H. Kauffman. Spin networks and knot polynomials. *Intl. J. Mod. Phys. A.*, vol. 5, no. 1 (1990), pp. 93-115.
- [LK17] L. H. Kauffman. Transformations in Special Relativity. *Int. J. Theo. Phys.*, vol. 24, no. 3, pp. 223-236, March 1985.
- [LK18] L. H. Kauffman and H. Saleur. Free fermions and the Alexander-Conway polynomial. *Comm. Math. Phys.* **141** (1991), 293-327.
- [LK19] L. H. Kauffman. Knots, Spin Networks and 3-Manifold Invariants. In the Proceedings of KNOTS 90 – Osaka Conference, ed. Akio Kawauchi, Walter de Gruyter (1992), pp. 271-285.
- [LK20] L. H. Kauffman and S. Lins. A 3-manifold invariant by state summation. (Preprint 1990).
- [LK21] L. H. Kauffman and S. Lins. Computing Turaev-Viro invariants for 3-manifolds *Manuscripta Math.* **72** (1991), 81-94.

- [LK22] L. H. Kauffman and R. Baadhio. Classical knots, gravitational anomalies and exotic spheres. (To appear.)
- [LK23] L. H. Kauffman. Knots, abstract tensors, and the Yang-Baxter Equation. In **Knots, Topology and Quantum Field Theories** - Proceedings of the Johns Hopkins Workshop on Current Problems in Particle Theory 13. Florence (1989). Ed. by L. Lussana. World Scientific Pub. (1989), pp. 179-334.
- [LK24] L. H. Kauffman and P. Vogel. Link polynomials and a graphical calculus. (Announcement 1987) *J. Knot Theory and its Ramifications* **1**, no. 1, (1992), 59-104.
- [LK25] L. H. Kauffman. Map coloring and the vector cross product. *J. Comb. Theor. B.*, vol. 48, no. 2. April 1990, pp. 145-154.
- [LK26] L. H. Kauffman. From knots to quantum groups (and back). In **Proceedings of the CRM Workshop on Hamiltonian Systems, Transformation Groups and Spectral Transform Methods**. Ed. by J. Harnad and J. E. Marsden. Les Publications CRM (1990), pp. 161-176.
- [LK27] L. H. Kauffman. **Map Reformulation**. Princelet Editions (1986).
- [LK28] C. Anezeris, A. P. Balachandran, L. Kauffman and A. M. Srivastava. Novel statistics for strings and string 'Chern Simons' terms. (Preprint 1990.)
- [LK29] L. H. Kauffman. An integral heuristic. *Intl. J. Mod. Phys. A.*, vol. 5, no. 7, (1990), pp. 1363-1367.
- [LK30] L. H. Kauffman. Special relativity and a calculus of distinctions. In **Proceedings of the 9th Annual International Meeting of the Alternative Natural Philosophy Association - Cambridge, England (September 23, 1987)**. Published by ANPA West, Palo Alto, Calif., pp. 290-311.
- [KA] T. Kanenobu. Infinitely many knots with the same polynomial. *Amer. Math. Soc.*, **97**(1986), pp. 158-161.
- [KE] A. B. Kempe. On the geographical problem of four colors. *Amer. J. Math.*, **2**(1879), pp. 193-200.
- [KV] M. A. Kervaire and J. W. Milnor. Groups of Homotopy Spheres: I. *Ann. of Math.*, vol. 77, no. 3, May 1963, pp. 504-537.
- [K] M. Khovanov. New geometrical constructions in low dimensional topology. (Preprint 1990).
- [KID] M. Kidwell. On the degree of the Brandt-Lickorish-Millett polynomial of a link. *Proc. Amer. Math. Soc.*, **100**(1987), pp. 755-762.
- [KIRB] R. Kirby. A calculus for framed links in S^3 . *Invent. Math.*, **45**(1978), 35-56.
- [KM] R. Kirby and P. Melvin. On the 3-manifold invariants of Reshetikhin-Turaev for $sl(2, \mathbb{C})$. *Invent. Math.* **105** (1991), 473-545.
- [KI] T. P. Kirkman. The enumeration, description and construction of knots with fewer than 10 crossings. *Trans. Royal Soc. Edin.*, **32**(1865), 281-309.

- [KR] A. N. Kirillov and N. Y. Reshetikhin. Representations of the algebra $U_q(sl_2)$, q -orthogonal polynomials and invariants of links. In **Infinite dimensional Lie algebras and groups**, Ed. V. G. Kac, *Adv. Ser. in Math. Phys.*, 7(1988), pp. 285-338.
- [KO1] T. Kohno. Monodromy representations of braid groups and Yang-Baxter Equations. *Ann. Inst. Fourier, Grenoble*, 37, 4 (1987), 139-160.
- [KO2] T. Kohno. Topological invariants for 3-manifolds using representations of mapping class groups I. (Preprint 1990.)
- [KON] M. Kontsevich. Rational conformal field theory and invariants of 3-dimensional manifolds. (Preprint 1988.)
- [KU] P. P. Kulish, N. Yu, Reshetikhin and E. K. Sklyanin. Yang-Baxter Equation and representation theory: I. *Letters in Math. Physics*, 5(1981), 393-403.
- [KUS] P. P. Kulish and E. K. Sklyanin. Solutions of the Yang-Baxter Equation. *J. Soviet Math.*, 19(1982), pp. 1596-1620. (Reprinted in **Yang-Baxter Equation in Integrable Systems**. Ed. by M. Jimbo. World Scientific Pub. (1990).)
- [KAW] A. Kuniba, Y. Akutsu, M. Wadati. Virasoro algebra, von Neumann algebra and critical eight-vertex SOS models. *J. Phys. Soc. Japan*, 55, no. 10 (1986), 3285-3288.
- [KUP1] G. Kuperberg. Involutory Hopf algebras and 3-manifold invariants. (Preprint 1990.)
- [KUP2] G. Kuperberg. The quantum G_2 link invariant. (Preprint 1990.)
- [KMU] B. I. Kurpita and K. Murasugi. On a hierarchical Jones invariant. (Preprint 1990.)
- [LAR] L. A. Lambe and D. E. Radford. Algebraic aspects of the quantum Yang-Baxter Equation. (Preprint 1990.)
- [LAT] R. G. Larson and J. Towber. Braided Hopf algebras: a dual concept to quasitriangularity. (Preprint 1990.)
- [LAW1] R. Lawrence. A universal link invariant using quantum groups. (Preprint 1988.)
- [LAW2] R. Lawrence. A functorial approach to the one-variable Jones polynomial. (Preprint 1990.)
- [LEE1] H. C. Lee. Q -deformation of $sl(2, C) \times Z_N$ and link invariants. (Preprint 1988.)
- [LEE2] H. C. Lee. Tangles, links and twisted quantum groups. (Preprint 1989.)
- [LCS] H. C. Lee, M. Couture and N. C. Schmeling. Connected link polynomials (Preprint 1988).
- [LM1] W. B. R. Lickorish and K. C. Millett. A polynomial invariant for oriented links. *Topology*, 26(1987), 107-141.

- [LM2] W. B. R. Lickorish and K. C. Millett. The reversing result for the Jones polynomial. *Pacific J. Math.*, **124**(1986), pp. 173-176.
- [LICK1] W. B. R. Lickorish. A representation of orientable, combinatorial 3-manifolds. *Ann. Math.*, **76**(1962), pp. 531-540.
- [LICK2] W. B. R. Lickorish. Polynomials for links. *Bull. London Math. Soc.*, **20**(1988), 558-588.
- [LICK3] W. B. R. Lickorish. 3-Manifolds and the Temperley-Lieb Algebra. (Preprint 1990.)
- [LICK4] W. B. R. Lickorish. Calculations with the Temperley-Lieb algebra. (Preprint 1990.)
- [LIP1] A. S. Lipson. An evaluation of a link polynomial. *Math. Proc. Camb. Phil. Soc.*, **100**(1986), 361-364.
- [LIP2] A. S. Lipson. Some more states models for link invariants. *Pacific J. Math.*, (To appear).
- [LIP3] A. S. Lipson. Link signatures, Goeritz matrices and polynomial invariants, *L'Enseignement Math.*, t.36, (1990), pp. 93-114.
- [LIT] C. N. Little. Non-alternate + - knots. *Trans. Royal Soc. Edin.*, **35**(1889), 663-664.
- [LOR] E. N. Lorenz. Nonperiodic flow. *J. Atmos. Sci.*, **20**(1963), pp. 130-141.
- [M] S. Mac Lane. **Categories for the Working Mathematician**. Springer-Verlag (1971). GTM 5.
- [MN] J. M. Maillet and F. W. Nijhoff. Multidimensional integrable lattices, quantum groups and the D -simplex equations. (Preprint – Institute for Non-Linear Studies – Clarkson University – INS #131 (1989).)
- [MAJ] S. Majid. Quasitriangular Hopf algebras and Yang-Baxter equations. (Preprint 1988).
- [MAN1] Yu. I. Manin. Quantum groups and non-commutative geometry. (Lecture Notes, Montreal, Canada, July 1988).
- [MAN2] Yu. I. Manin. Some remarks on Kozul algebras and Quantum groups. *Ann. Inst. Fourier, Grenoble*, **37**, 4 (1987), 191-205.
- [MAR] P. P. Martin. Analytic properties of the partition function for statistical mechanical models. *J. Phys. A: Math. Gen.*, **19**(1986), pp. 3267-3277.
- [MAT] S. V. Matveev. Transformations of special spines and the Zeeman conjecture. *Math. USSR Izvestia* 31: 2 (1988), pp. 423-434.
- [ME] D. A. Meyer. State models for link invariants from the classical Lie groups. Proceedings of Knots 90 – Osaka Conference, ed. Akio Kawauchi, Walter de Gruyter (1992), pp. 559-592.
- [MIE] E. W. Mielke. Knot wormholes in geometrodynamics. *General Relativity and Gravitation*, vol. 8, no. 3, (1977), pp. 175-196.

- [MI] J. W. Milnor. **Singular Points of Complex Hypersurfaces.** *Annals Studies* **61**, Princeton University Press (1968).
- [MD] J. Moody. A link not detected by the Burau matrix. (Preprint 1989.)
- [MS] G. Moore and N. Seiberg. Classical and quantum conformal field theory. *Commun. Math. Phys.*, **123**(1989), pp. 177-254.
- [MO1] H. R. Morton. The Jones polynomial for unoriented links. *Quart. J. Math. Oxford* (2), **37**(1986), pp. 55-60.
- [MO2] H. R. Morton. Seifert circles and knot polynomials. *Math. Proc. Camb. Phil. Soc.*, **99**(1986), pp. 107-109.
- [MO3] H. R. Morton and H. B. Short. The two-variable polynomial for cable knots. *Math. Proc. Camb. Phil. Soc.*, **101**(1987), pp. 267-278.
- [MO4] H. R. Morton and P. M. Strickland. Satellites and surgery invariants. (Preprint 1990.)
- [MOU] J. P. Moussouris. **Quantum Models of Space-Time Based on Recoupling Theory.** (Thesis, Oxford Univ., 1983).
- [MUK1] J. Murakami. The parallel version of link invariants. (Preprint Osaka University (1987).)
- [MUK2] J. Murakami. A state model for the multi-variable Alexander polynomial. (Preprint 1990.)
- [MUK3] J. Murakami. On local relations to determine the multi-variable Alexander polynomial of colored links. (Preprint 1990.)
- [MUR1] K. Murasugi. The Jones polynomial and classical conjectures in knot theory. *Topology*, **26**(1987), 187-194.
- [MUR2] K. Murasugi. Jones polynomials and classical conjectures in knot theory II. *Math. Proc. Camb. Phil. Soc.*, **102**(1987), 317-318.
- [N] M. H. A. Newman. On a string problem of Dirac. *J. London Math. Soc.*, **17**(1942), pp. 173-177.
- [O] E. Oshins. Quantum Psychology Notes. Vol. 1: A Personal Construct Notebook (1987). (Privately distributed.)
- [PEN1] R. Penrose. Applications of negative dimensional tensors. **Combinatorial Mathematics and its Applications.** Edited by D. J. A. Welsh, Academic Press (1971).
- [PEN2] R. Penrose. Theory of quantized directions. (Unpublished.)
- [PEN3] R. Penrose. Angular momentum: an approach to Combinatorial Space-Time. In **Quantum Theory and Beyond.** Ed. T. A. Bastin. Cambridge University Press, (1969).
- [PEN4] R. Penrose. Combinatorial quantum theory and quantized directions. **Advances in Twistor Theory.** Ed. by L. P. Hughston and R. S. Ward. Pitman (1979), pp. 301-307.

- [PW] J. H. H. Perk and F. Y. Wu. Nonintersecting string model and graphical approach: equivalence with a Potts model. *J. Stat. Phys.*, vol. 42, nos. 5/6, (1986), pp. 727-742.
- [PS] J. H. H. Perk and C. L. Schultz. New families of commuting transfer matrices in q -state vertex models. *Physics Letters*, vol. 84A, no. 8, August 1981, pp. 407-410.
- [PO] A. M. Polyakov. Fermi-Bose transmutations induced by gauge fields. *Mod. Phys. Lett.*, A3 (1987), pp. 325-328.
- [PT] J. H. Przytycki and P. Traczyk. Invariants of links of Conway type. *Kobe J. Math.*, 4(1987), pp. 115-139.
- [Q] C. Quigg. **Gauge Theories of the Strong, Weak, and Electromagnetic Interactions**. Benjamin Pub. (1983).
- [RAD] D. E. Radford. Solutions to the quantum Yang-Baxter Equation: the two-dimensional upper triangular case. (Preprint 1990.)
- [RA] M. Rasetti and T. Regge. Vortices in He II, Current Algebras and Quantum Knots. *Physica* 80a (1975), pp. 217-233, Quantum vortices. In **Highlights of Condensed Matter Theory** (1985). Soc. Italiana di Fisica - Bologna, Italy.
- [REI] K. Reidemeister. **Knotentheorie**. Chelsea Publishing Co., New York (1948). Copyright 1932. Julius Springer, Berlin.
- [RES] N. Y. Reshetikhin. Quantized universal enveloping algebras, the Yang-Baxter Equation and invariants of links, I and II. LOMI reprints E-4-87 and E-17-87, Steklov, Institute, Leningrad, USSR.
- [RT1] N. Y. Reshetikhin and V. Turaev. Ribbon graphs and their invariants derived from quantum groups. *Comm. Math. Phys.*, 127(1990), pp. 1-26.
- [RT2] N. Y. Reshetikhin and V. Turaev. Invariants of three manifolds via link polynomials and quantum groups. *Invent. Math.*, 103 (1991), 547-597.
- [ROB] G. D. Robertson. Torus knots are rigid string instantons. University of Durham - Centre for Particle Theory. (Preprint 1989.)
- [ROLF] D. Rolfsen. **Knots and Links**. Publish or Perish Press (1976).
- [ROSS1] M. Rosso. Groupes quantiques de Drinfeld et Woronowicz. (Preprint 1988).
- [ROSS2] M. Rosso. Groupes quantiques et modeles a vertex de V. Jones en theorie des noeuds. *C. R. Acad. Sci. Paris*, t.307 (1988), pp. 207-210.
- [RS] C. Rovelli and L. Smolin. Knot theory and quantum gravity. (Preprint 1988).
- [RY] L. H. Ryder. **Quantum Field Theory**. Cambridge University Press (1985).
- [SA] T. L. Saaty and P. C. Kainen. **The Four Color Problem**. Dover Pub. (1986).

- [SAL1] H. Saleur. Virasoro and Temperley-Lieb algebras. In **Knots, Topology and Quantum Field theories – Proceedings of the Johns Hopkins Workshop on Current Problems in Particle Theory 13**. Florence (1989). Ed. by L. Lussana. World Scientific Pub. (1989), pp. 485-496.
- [SAL2] H. Saleur. Zeroes of chromatic polynomials: a new approach to Beraha conjecture using quantum groups. *Commun. Math. Phys.*, **132**(1990), pp. 657-679.
- [SAM] S. Samuel. The use of anticommuting variable integrals in statistical mechanics. I. The computation of partition functions. II. The computation of correlation functions. III. Unsolved models. *J. Math. Phys.*, **21**(12)(1980), pp. 2806-2814, 2815-2819, 2820-2833.
- [SE] G. Segal. Two-dimensional conformal field theories and modular functors. **IXth International Congress on Mathematical Physics – 17-27 July 1988-Swansea, Wales**. Ed. Simon, Truman, Davies, Adam Hilger Pub. (1989), pp. 22-37.
- [CS] C. Shultz. Solvable q -state models in lattice statistics and quantum field theory. *Phys. Rev. Letters*. **46** no. 10, March 1981.
- [SI] L. Silberstein. **The Theory of Relativity**. MacMillan and Co. Ltd. (1914).
- [SIM] J. Simon. Topological chirality of certain molecules. *Topology*, vol. 25, no. 2, (1986), pp. 229-235.
- [SK] E. K. Sklyanin. Some algebraic structures connected with the Yang-Baxter Equation. Representation of quantum algebras. *Funct. Anal. Appl.*, **17**(1983), pp. 273-284.
- [LS1] L. Smolin. Link polynomials and critical points of the Chern-Simons path integrals. *Mod. Phys. Lett. A.*, vol. 4, no. 12, (1989), pp. 1091-1112.
- [LS2] L. Smolin. Quantum gravity in the self-dual representation. *Contemp. Math.*, **28**(1988).
- [SO1] R. Sorkin. Particle statistics in three dimensions. *Phys. Dev.*, D27, (1983), pp. 1787-1797.
- [SO2] R. Sorkin. A general relation between kink exchange and kink rotation. *Commun. Math. Phys.*, **115**(1988), pp. 421-434.
- [SB1] G. Spencer-Brown. **Cast and Formation Properties of Maps**. Privately distributed manuscript. Deposited in the Library of the Royal Society, London, March 1980.
- [SB2] G. Spencer-Brown. **Laws of Form**. George Allen and Unwin Ltd., London (1969).
- [STEW] R. L. Stewart. Pure field electromagnetism. (Unpublished - 1971).
- [ST] J. Stillwell. **Classical Topology and Combinatorial Group Theory**. Springer-Verlag (1980). GTM 72.

- [S] D. W. Sumners. Untangling DNA. *Math. Intelligencer*. vol. 12, no. 3, (1990), pp. 71-80.
- [TAIT] P. G. Tait. On Knots I, II, III. Scientific Papers, vol. I, Cambridge University Press, London, 1898, 273-347.
- [TEL] H. N. V. Temperley and E. H. Lieb. Relations between the ‘percolation’ and ‘coloring’ problem and other graph-theoretical problems associated with regular planar lattices: some exact results for the ‘percolation’ problem. *Proc. Roy. Soc. Lond. A.*, **322**(1971), pp. 251-280.
- [TEU] R. D. Teuschner. Towards a topological spin-statistics theorem in quantum field theory, (Preprint 1989.)
- [TH1] M. B. Thistlethwaite. Knot tabulations and related topics. *Aspects of Topology*. Ed. I. M. James and E. H. Kronheimer, Cambridge University Press (1985), 1-76.
- [TH2] M. B. Thistlethwaite. A spanning tree expansion of the Jones polynomial. *Topology*, **26**(1987), pp. 297-309.
- [TH3] M. B. Thistlethwaite. On the Kauffman polynomial of an adequate link. *Invent. Math.*, **93**(1988), pp. 285-296.
- [TH4] M. B. Thistlethwaite. Kauffman’s polynomial and alternating links. *Topology*, **27**(1988), pp. 311-318.
- [TH5] M. B. Thistlethwaite. (Unpublished tables.)
- [T] W. Thompson (Lord Kelvin). On vortex atoms. *Philosophical Magazine*, **34** July 1867, pp. 15-24. *Mathematical and Physical Papers*, vol. 4, Cambridge (1910).
- [TR] B. Trace. On The Reidemeister moves of a classical knot. *Proc. Amer. Math. Soc.*, vol. 89, no. 4, (1983).
- [TU1] V. G. Turaev. The Yang-Baxter equation and invariants of links. LOMI preprint E-3-87, Steklov Institute, Leningrad, USSR, *Inventiones Math.*, **92**, Fasc. 3, 527-553.
- [TU2] V. G. Turaev and O. Viro. State sum invariants of 3-manifolds and quantum $6j$ symbols. (Preprint 1990.)
- [TU3] V. G. Turaev. Quantum invariants of links and 3-valent graphs in 3-manifolds. (Preprint 1990.)
- [TU4] V. G. Turaev. Shadow links and face models of statistical mechanics. Publ. Inst. Recherche Math. Advance. Strasbourg (1990).
- [TU5] V. G. Turaev. Quantum invariants of 3-manifolds and a glimpse of shadow topology. (Preprint 1990.)
- [TU6] V. G. Turaev. The Conway and Kauffman modules of the solid torus with an appendix on the operator invariants of tangles. LOMI Preprint E-6-88.

- [TU7] V. G. Turaev. Algebras of loops on surfaces, algebras of knots, and quantization. LOMI Preprint E-10-88. In **Braid Group, Knot Theory and Statistical Mechanics**. Ed. C. N. Yang and M. L. Ge. World Sci. Pub. Advanced Series in Mathematical Physics, **9**(1989), pp. 59-96.
- [TUT] W. T. Tutte. Graph Theory. **Encyclopedia of Mathematics and its Applications**, **21**, Cambridge University Press, (1984).
- [VMU] F. Varela, H. Maturana, R. Uribe. Autopoiesis: the organization of living systems, its characterization and a model. *Biosystems*, **5**:187, (1974).
- [V] P. Vogel. Representations of links by braids: A new algorithm. *Comment. Math. Helvetici*, **65**(1990), pp. 104-113.
- [WE] H. Wenzl. Representations of braid groups and the quantum Yang-Baxter Equation, *Pacific J. Math.*, **145**(1990), pp. 153-180.
- [WEYL] H. Weyl. Gravitation und Elektrizität. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, **26**(1918), pp. 465-480.
- [W] J. H. White. Self-linking and the Gauss integral in higher dimensions. *Amer. J. Math.*, **91**(1969), pp. 693-728.
- [WH1] H. Whitney. On regular closed curves in the plane. *Comp. Math.*, **4**(1937), 276-284.
- [WH2] H. Whitney. A logical expansion in mathematics. . *Bull. Amer. Math. Soc.*, **38**(1932), 572-579.
- [WI] A. S. Wightman. (Private communication.)
- [WIL] F. Wilczek. Gauge theory of deformable bodies. IASSNS-HEP-88/41 (Preprint 1988), in *Intl. Congress on Math. Phys.* Ed. by B. Simon, A. Truman and I. M. Davies. Adam Hilger (1989), pp. 220-233.
- [WZ] F. Wilczek and A. Zee. Linking numbers, spin and statistics of solitons. *Phys. Rev. Lett.*, vol. 51, no. 25, (1983), pp. 2250-2252.
- [WIT1] E. Witten. Physics and geometry. *Proc. Intl. Congress Math.*, Berkeley, Calif., USA (1986), 267-303.
- [WIT2] E. Witten. Quantum field theory and the Jones polynomial. *Comm. Math. Phys.*, **121**(1989), 351-399.
- [WIT3] E. Witten, Gauge theories vertex models and quantum groups. *Nucl. Phys. B*, **330**(1990), pp. 225-246.
- [WIT4] E. Witten. Global gravitational anomalies. *Comm. Math. Phys.*, **100**(1985), pp. 197-229.
- [WIN] S. Winker. **Quandles, Knots Invariants and the N -fold Branched Cover**. Ph.D. Thesis, University of Illinois at Chicago, (1984).
- [WU1] F. Y. Wu. The Potts Model. *Rev. Mod. Phys.*, vol. 54, no. 1, Jan. 1982, pp. 235-268.

- [WU2] F. Y. Wu. Graph theory in statistical physics. In **Studies in Foundations and Combinatorics** - Advances in Mathematics Supplementary Studies. Vol. 1, (1978). Academic Press Inc., pp. 151-166.
- [Y] S. Yamada. The minimum number of Seifert circles equals the braid index of a link. *Invent. Math.*, **89**(1987), pp. 346-356.
- [YGW] K. Yamagishi, M. Ge, Y. Wu. New hierarchies of knot polynomials from topological Chern-Simons gauge theory. *Lett. Math. Phys.*, **19**(1990), pp. 15-24.
- [YE] D. Yetter. Quantum groups and representations of monoidal categories. *Math. Proc. Camb. Phil. Soc.*, **108**(1990), pp. 197-229.
- [ZA] M. Zaganescu. Bosonic knotted strings and cobordism theory. *Europhys. Lett.*, **4(5)**(1987), pp. 521-525.
- [Z1] A. B. Zamolodchikov. Factorized S matrices and lattice statistical systems. *Soviet Sci. Reviews. Part A* (1979-1980).
- [Z2] A. B. Zamolodchikov. Tetrahedron equations and the relativistic S -matrix of straight strings in $2+1$ dimensions. *Comm. Math. Phys.*, **79**(1981), 489-505.

Index

- abelian gauge** 309, 310
- abstract Feynman diagram** 74, 116, 117
- abstract tensor** 104, 137
- achiral** 9, 10, 47
- adequate link** 46
- adjoint representation** 292
- Alexander-Conway polynomial** 51, 52, 54, 172, 174–185, 202–214
- Alexander crystal** 201–204
- Alexander module** 201, 202, 204
- Alexander polynomial** 172, 174–185, 202, 203, 206
- Alexander's Theorem** 91–93
- alternating link** 20, 39–48, 45, 370, 380
- alternating symbol** 126, 130
- ambient isotopy** 18
- amplitude** 111, 118, 237, 239, 381, 382
- ancestor** 25
- angular momentum** 381, 390
- annihilation** 74, 77, 80, 84, 235
- antipode** 141, 142, 158
- antisymmetrizer** 445
- A*-region** 27
- Armel's Experiment** 331
- Artin braid group** 86, 88, 205, 208, 374, 375
- band link** 55, 56
- Belt Theorem** 429, 433
- belt trick** 403, 420, 427, 429, 434, 480
- Berezin integral** 211, 212
- bialgebra** 137, 141, 151, 153
- binors** 125, 128, 445
- Birman-Wenzl algebra** 218, 234
- blackboard framing** 251, 257
- blue (see red and purple)** 22, 347
- Bohm-Aharonov effect** 294
- Boltzmann's constant** 365
- Boltzmann distribution** 366
- bend-graph** 354
- Bondi's K-calculus** 395
- Borromean rings** 11, 12, 20
- Borromean rings (generalized)** 38
- bowline** 4, 328
- bracket polynomial** 25, 28, 33, 43, 49, 74, 117, 124, 134, 216, 371
- braid** 85, 87, 122, 123, 374
- braid axis** 91

braid state 97
B-region 27
bridge 44
Burau representation 186, 205, 206, 208, 213, 320
cap 117
Casimir operator 307, 495
cast 347
Catalan number 272
Cayley numbers 403, 408, 423
channel unitarity 75–77, 109, 111
checkerboard surface 39
Chern-Simons form 54, 285, 299–303
chiral, chirality 9, 46, 47, 66, 217, 218
chromatic expansion 350
chromatic polynomial 353
chromatic state 356
Clebsch-Gordon coefficients 447
closed braid 88
clove hitch 4, 324
coassociative 140
code 187
colors 22
commutation relations 385, 389
comultiplication 136, 144
conformal field theory 314
conjugate braid 94
creation 74, 77, 84, 117, 235, 511
critical temperature 379
cross(right and left) 186
cross-channel unitarity (inversion) 75, 76, 79, 109, 111, 239
crossing 10, 19, 80
crossing (oriented) 19
crossing sign 19
crystal 186, 188, 191, 192, 194
cup 117
curvature 287, 297–299, 499
curvature insertion 305
cycle conditions 168
De Broglie 389, 392
decorate 58
degree 44
devil 327
diagram 8, 10
diagram monoid 100
diagrammatic formal system 18

dichromatic polynomial 364, 367
Dirac equation 426
Dirac string trick 420, 427, 435
distinction 460–466
DNA 488–500
double bracket 264, 265
double construction (quantum double, Drinfeld double) 152–156, 236
double group 433
Dubrovnik polynomial (see Kauffman polynomial) 215
edge 11
electromagnetic tensor 293
electromagnetism 425
Elliot-Biedenharn identity 474
encode 13
epsilon 125–129, 443–445
exchange identity 52
exotic spheres 479, 480
exotic statistics 487
expectation value 285
Faddeev-Reshetikhin-Takhtajan Construction (FRT) 152, 153, 160, 210, 211
Feynman diagram 291
Fierz identity 311
figure eight knot 8, 35
filter 381, 382
formation 347
fragment 117
framed link 56, 251
framing number 250
free fermions 174, 212–214
functional integral 285, 304, 307
fundamental group 186, 191
gauge covariant derivative 293, 295, 297
gauge field (gauge potential) 285, 286, 296, 298, 280, 500
gauge transformation 293, 295, 296, 300
generalized Reidemeister moves 236
generalized skein relation 307, 495
granny knot 7, 8, 327
graph 11, 12, 47
Gray code 358
group velocity 391
handle 254
handle slide 255
Hecke algebra 218
Hermitian 384, 400

heuristics 285
hitches 323
Homfly polynomial 52, 54, 162, 171, 217, 313
Hopf algebra 141, 156, 159, 236
Hopf link 12, 56

ice-model 319
imaginary knots 334
imaginary value 348
infinitesimal rotation 383
insertion 306
integration 24, 285, 287, 289, 304
involuntary quandle 199
iterant algebra 397

J-homomorphism 479, 480
Jones polynomial 49, 52, 54, 74, 82, 85, 216, 217, 250, 320
Jordan curve 15, 28, 41, 222, 351
Jordan Curve Theorem 15, 41

Kauffman polynomial 52, 54, 215–234, 249
Kirby calculus 250, 260
Kirby moves 252, 254–260, 265
knot holder 501
knotted quantized flux 487
knotted strings 475
Kronecker delta 106

lasso 196
Laurent polynomial 49
Lens space 253
Lie algebra 143, 292, 296
Life, Game of (Conway) 496, 497
light 381, 392, 393
light-cone 394
light-cone coordinates 394, 460
light crystal (see involuntary quandle) 199
light speed 393
link 8, 19
linking number 19, 339, 489
longitude 252, 253
Lorentz group 398, 460
Lorentz transformation 396–402, 462, 464
Lorenz attractor (Lorenz knot) 501–509

magic weave 459
map 347
Markov move 94, 95

Markov Theorem 95, 123
Markov trace 86, 96, 123
matrix insertion 305, 308
Matveev moves 471, 474
Maxwell's equations 293, 426
mechanism 328
medial graph 356
meridian 252, 253
meta-time 73, 166
minus one (square root of) 430
mirror 403, 404, 424
mirror category 337
mirror conjecture 48
mirror image 9, 18, 34, 38
Möbius band 418
morphogenetic field 496, 497
motion group 486
move 10, 16
move zero 16

normalized bracket 33
normed algebra 408, 425

observable 384, 390
observation 381
operand 186
operator 186, 188
oriented model 235
oriented triangle move 81

Parallel transport 294
parity 84
partition function 28, 239, 285, 365
path-ordered (exponential) 286, 296
Pauli matrices 292, 310, 388, 398, 461
Penrose chromatic recursion 346
Penrose spin network 125, 134, 443–458
permutation 91
Peterson graph 352
phase (factor) 294
physical knots 4, 323
piecewise linear models 235, 238
piecewise linear Reidemeister moves 240
Potts bracket 371
Potts model 28, 284, 364–367
probability amplitude 118
probability distribution 365

- process** 381
- projective space (projective plane)** 416, 417, 481–485
- purple** (see red and blue) 22, 347

- $q - 6j(q - 3j)$ symbol 471, 473
- q -spin network 459, 466, 467–474
- q -symmetrizer 469
- quandle 186, 193
- quantum factorial 468
- quantum field theory 54, 119, 250, 488
- quantum group 54, 117, 134, 222
- quantum mechanics 381
- quantum universal enveloping algebra 144
- quasi-triangular Hopf algebra 159
- quaternion demonstrator 427–442
- quaternions 401, 403, 406, 421–425, 427, 431

- rapidity parameter 245, 246
- rational tangle 494
- Ray-Singer torsion 313
- recombination 491
- red (see blue and purple) 22, 347
- reduced diagram 44
- reflection (see mirror) 403, 404, 414, 415, 425
- regular isotopy 16, 17, 18, 52, 54, 119, 159, 239, 307
- Reidemeister moves 16, 119, 241
- relativity 392, 425
- Reshetikhin-Turaev invariant 250
- reversing property (of Jones polynomial) 51
- Reynolds number 501
- ribbon knot 184, 185
- ribbon singularity 184, 185
- R -matrix (see Yang-Baxter equation) 109, 111, 114, 117, 124, 130, 208, 244–246, 248
- roadway problem 39–41
- rogue's gallery 6, 266
- rot (rotation number) 163, 164, 178, 219, 223, 247
- rotation 390, 414, 415, 424
- rubber band 329

- Schrödinger equation 294, 389
- semi-oriented polynomial 52
- shadow 11
- sign 3, 19
- signature 266
- skein 57
- skein identity (see exchange identity) 52

skein model 62, 66
 skein relation 287, 307
 skein template algorithm 57–63, 67–73
 skein tree 57, 65
 $SL(2)$ 125, 134, 443, 444
 $SL(2)q$ 117, 136
 slide equivalence 336
 $SO(3)$ 416, 427
 $SO(n)$ 222, 229
 solenoid 294
 spacetime 393, 464
span 46
 spin 381, 425
 Spin-Geometry Theorem 450, 451
 spin network 125, 134, 381, 451, 456
 spinors 128, 443
 splice 25, 84
 split 27, 40
 split of type A 27
 split of type B 27
 square knot 6, 8, 327
 state 24, 25, 28, 108, 162, 213, 219, 222, 223, 238, 343, 349, 365
 state model 28, 167, 174, 233
 statistical mechanics 28, 160, 211, 365–367
 strings 132, 475–480
 string tricks 425
 $SU(2)$ 290, 427, 453
 $SU(N)$ 311
 substrate 491, 493
 superposition 382
 surgery 252, 253, 313
 symmetric group 107
synaptic complex 491, 492

tangle 176, 264
 tangle addition 491
 Temperley-Lieb algebra 99, 102, 219, 250, 269, 272, 275, 281–284, 364, 374, 376, 379, 469
 Temperley-Lieb projector 275, 281–284, 468
 template (see skein-template algorithm) 58, 60
 three-colored 22
 three manifold invariants 250, 261, 266, 267, 474
 time's arrow 74, 112, 116, 117, 132, 166
 topological quantum field theory 119
 torus link 37
 tower construction 85, 86

trace 86, 123, 211, 276
 transposition 91, 107
 trefoil knot 9, 20, 21–24, 35
 triangle invariance 76, 80
 triangles 81
 Turaev-Viro invariant 474
 twist (move) 119–122, 158, 237
 twist number (see writhe) 19, 488
 undercrossing 10
 unitary 77
 unitary transformation 384
 universal R -matrix 147, 148, 152, 154
 universe 335, 364, 368, 370
 unknot 8, 9, 18, 84
 unknot problem 84
 vacuum-vacuum expectation (amplitude) 73, 117, 118, 156, 161, 163, 235, 286
 variation of the Chern-Simons Lagrangian 300
 vector coupling coefficients 451, 452
 vector potential 294
 vertex, vertices 11
 vertex weights 28, 61, 108, 222, 223, 225, 233, 285
 von Neumann algebra 85
 wave mechanics 389, 391
 weaving pattern 11
 Weyl 294
 Whitehead link 36, 493
 White's Theorem 488, 489
 Whitney trick 484
 Wilson loop 294, 297, 303–305, 308, 496, 500, 502
 Winker's Theorem 195
 Wirtinger presentation 191
 Witten's Integral 285, 300, 495
 word 505, 511
 writhe 19, 33, 216, 251, 488–490
 Yang-Baxter commutation relation 125, 149–151, 156
 Yang-Baxter Equation (YBE or QYBE) 104, 110, 111, 117, 122, 125, 182, 183, 225–231, 239, 244, 248, 316–320
 Yang-Baxter model 83, 174, 222
 zero rules 225

APPENDIX

Introduction

This appendix consists of the present introduction plus a reprinting of four articles by the author:

1. Gauss codes, quantum groups and ribbon Hopf algebras [K93]
2. Spin networks, topology and discrete physics [K]
3. (with P. Vogel) Link polynomials and a graphical calculus [KV92]
4. (with J. Goldman) Knots, tangles and electrical networks [GK93]

These articles contain information that extends the material in the first edition of *Knots and Physics*. (We shall refer to the earlier parts of the present book by the abbreviation K&P.)

We first discuss the four articles. They are referred to, respectively, as the “first article”, the “second article”, the “third article” and the “fourth article”. Then we discuss some points about the applications of the techniques in K&P and the third article to Vassiliev invariants. A fifth article on “Knot Theory and Functional Integration” has been added to this third edition.

First Article

The first article draws a line from abstract tensor models for link invariants to the structure of quasi-triangular and ribbon Hopf algebras and the construction of 3-manifold invariants via traces on these Hopf algebras. In the case of finite dimensional Hopf algebras these traces can be given by right integrals on the algebra. We reconstruct Hennings’ invariants of 3-manifolds. Our method for constructing link invariants is of interest as it uses the Hopf algebra structure directly on the link diagrams rather than using a representation of the Hopf algebra. This yields diagrammatic interpretations for many properties of ribbon Hopf algebras. For more information the reader is referred to the joint work of the author with David Radford [KR] where we show that invariants derived in this way are definitely distinct from the class of invariants obtained by Witten, Reshetikhin and Turaev.

Second Article

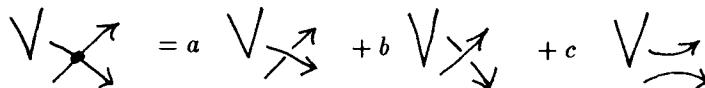
The second article contains an up-date on results related to q -deformed spin networks as discussed in Part II of K&P. In particular, we show how the spin network methods can be used to retrieve both the Turaev-Viro and Witten-Reshetikhin-Turaev invariants in the case of $SU(2)$. Further information about

this point of view can be obtained from the monograph by the author and Sostenes Lins [KL], the book edited by the author and Randy Baadhio in this series [KB], and from papers by W.B.R. Lickorish such as [L], the paper by Justin Roberts [R93], the paper by Louis Crane and David Yetter [CY], and the papers by the author, Louis Crane and David Yetter [CKY1], [CKY2].

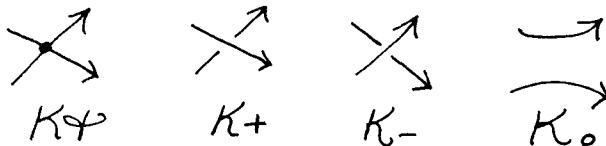
Third Article

The third article is about constructions of rigid vertex invariants of embedded 4-valent graphs in 3-dimensional space. If $V(K)$ is a (Laurent polynomial valued, or more generally — commutative ring valued) invariant of knots, then it can be naturally extended to an invariant of rigid vertex graphs by defining the invariant of graphs in terms of the knot invariant via an “unfolding” of the vertex as indicated below:

$$V(K_{\&}) = aV(K_+) + bV(K_-) + cV(K_0)$$



Here $K_{\&}$ indicates an embedding with a transversal 4-valent vertex ($\&$).



Formally, this means that we define $V(G)$ for an embedded 4-valent graph G by taking the sum over $a^{i^+(S)}b^{i^-(S)}c^{i^o(S)}V(S)$ for all knots S obtained from G by replacing a node of G with either a crossing of positive or negative type, or with a smoothing (denoted o). It is not hard to see that if $V(K)$ is an ambient isotopy invariant of knots, then, this extension is a rigid vertex isotopy invariant of graphs. In rigid vertex isotopy the cyclic order at the vertex is preserved, so that the vertex behaves like a rigid disk with flexible strings attached to it at specific points.

The third article considers this construction of invariants in the context of the skein polynomials and develops algorithms for the calculation and relation of these structures with the Hecke algebra and the Birman Wenzl algebra.

Fourth Article

The fourth article gives an interpretation of the graphical Reidemeister moves in terms of the properties of generalized electrical networks. As a consequence, we define new link invariants (for links in the complement of two special lines, and for tangles) by measuring the generalized conductance between two nodes in the signed graph of the link. In the case of tangles, this invariant can be expressed as a ratio of two Alexander–Conway polynomials evaluated at the point where the Jones polynomial and the Alexander polynomial coincide.

Vassiliev Invariants

There is a rich class of graph invariants that can be studied in the manner of the third article. The *Vassiliev Invariants* ([V90], [BL92], [BN92]) constitute the important special case of these graph invariants where $a = +1$, $b = -1$ and $c = 0$. Thus $V(G)$ is a Vassiliev invariant if

$$V(K_\&) = V(K_+) - V(K_-).$$

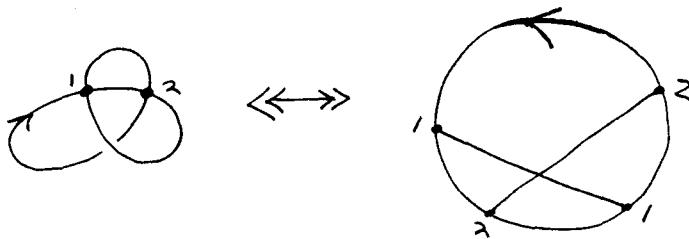
$$\checkmark \nearrow \nwarrow = \checkmark \nearrow \nearrow - \checkmark \searrow \nearrow$$

$V(G)$ is said to be of *finite type k* if $V(G) = 0$ whenever $\#(G) > k$ where $\#(G)$ denotes the number of 4-valent nodes in the graph G . If V is of finite type k , then $V(G)$ is independent of the embedding type of the graph G when G has exactly k nodes. This follows at once from the definition of finite type. The values of $V(G)$ on all the graphs of k nodes is called the *top row* of the invariant V .

Note that an isolated node in a graph embedding causes a zero invariant.

$$\checkmark \nearrow \nearrow = \checkmark \nearrow - \checkmark \searrow = \checkmark - \checkmark = 0.$$

For purposes of enumeration it is convenient to use *chord diagrams* to enumerate and indicate the abstract graphs. A chord diagram consists in an oriented circle with an even number of points marked along it. These points are paired with the pairing indicated by arcs or chords connecting the paired points. See the following figure.



This figure illustrates the process of associating a chord diagram to a given embedded 4-valent graph. Each transversal self intersection in the embedding is matched to a pair of points in the chord diagram.

It follows from the work of Bar-Natan [BN92] and Kontsevich [Ko92] that a set of top row evaluations satisfying a certain condition (the four-term relation) is sufficient to determine a Vassiliev invariant. Thus it is significant to determine such top rows. Furthermore, there is a great abundance of Vassiliev invariants of finite type, enough to build all the known quantum link invariants and many classical invariants as well.

Let us begin by showing the result of Birman and Lin [BL92] that the coefficients of x^i in the Jones polynomial evaluated at e^x are Vassiliev invariants of type i .

Letting $V_K(t)$ denote the original Jones polynomial, we can write $V_K(e^x) = \sum v_i(K)x^i$ with $i = 0, 1, 2, \dots$. We wish to show that $v_i(K)$ is a Vassiliev invariant of finite type i . This means that $v_i(G) = 0$ if $\#G > i$. This, in turn, is equivalent to the statement: x^i divides $V_G(e^x)$ if $\#G = i$. It is this divisibility statement that we shall prove.

In order to see this divisibility, let us apply the oriented state model for the Jones polynomial as derived in Chapter 6 of Part I of K&P. Recall the formulas:

$$\bigvee \begin{smallmatrix} \nearrow \\ \searrow \end{smallmatrix} = -t^{\frac{1}{2}} \bigvee \begin{smallmatrix} \nearrow \\ \nearrow \end{smallmatrix} - t \bigvee \begin{smallmatrix} \nearrow \\ \nearrow \\ \searrow \end{smallmatrix}$$

$$\bigvee \begin{smallmatrix} \nearrow \\ \searrow \end{smallmatrix} = -t^{-\frac{1}{2}} \bigvee \begin{smallmatrix} \nearrow \\ \nearrow \end{smallmatrix} - t^{-1} \bigvee \begin{smallmatrix} \nearrow \\ \nearrow \\ \searrow \end{smallmatrix}.$$

Letting $t = e^x$ and taking the difference, we have

$$\begin{aligned} \text{V} \begin{array}{c} \nearrow \\ \searrow \end{array} &= \text{V} \begin{array}{c} \nearrow \\ \nearrow \end{array} - \text{V} \begin{array}{c} \searrow \\ \nearrow \end{array} = -2 \sinh(x/2) \text{V} \begin{array}{c} \nearrow \\ \nearrow \end{array} \\ &\quad - 2 \sinh(x) \text{V} \begin{array}{c} \nearrow \\ \nearrow \end{array} \end{aligned}$$

from which it immediately follows that x divides $V \begin{array}{c} \nearrow \\ \searrow \end{array}$. Hence x^i divides V_G if G has i nodes. This completes the proof. //

Lie Algebras and the Four Term Relation

Recall that in Chapter 17 of Part I of K&P we derived a difference formula, correct to order $(1/k)$, using the functional integral formulation of link invariants. This formula reads

$$Z(K_+) - Z(K_-) = (-4\pi i/k)Z(T_a T_a K_\#) + O(1/k^2)$$

where $T_a T_a K_\#$ denotes the Casimir insertion into the Wilson line as explained in that chapter. This insertion is obtained by replacing the crossing by transversely intersecting segments, putting a Lie algebra generator on each segment, and summing over the elements of the basis for the Lie algebra. $K_\#$ denotes the diagram obtained from K by replacing the crossing by transversely intersecting segments.

In the discussion in Chapter 17 we assumed that the Lie algebra was classical and that $T_a T_b - T_b T_a = if_{abc} T_c$ (sum on c) where f_{abc} is totally antisymmetric. This assumption simplifies the present discussion as well. See [BN92] for a more general treatment.

Recall that $Z(K)$ is a regular isotopy invariant of knots and links, with framing behaviour

$$\begin{aligned} Z \begin{array}{c} \nearrow \\ \searrow \end{array} &= \alpha Z \begin{array}{c} \nearrow \\ \nearrow \end{array} \\ Z \begin{array}{c} \searrow \\ \nearrow \end{array} &= \bar{\alpha}' Z \begin{array}{c} \nearrow \\ \nearrow \end{array} \end{aligned}$$

for a constant α that depends upon the particular representation of the gauge group. We normalize to get an invariant of ambient isotopy in the usual way

by multiplying by a factor of $\alpha^{-w(K)}$ where $w(K)$ is the writhe of K . In order to simplify this discussion, let us note that we could have taken the definition of Vassiliev invariants in the context of regular isotopy.

We say that V_K is a (*framed*) *Vassiliev invariant* if it satisfies the usual difference formula

$$\vee \begin{array}{c} \nearrow \\ \times \\ \searrow \end{array} = \vee \begin{array}{c} \nearrow \\ \times \\ \searrow \end{array} - \vee \begin{array}{c} \nearrow \\ \times \\ \swarrow \end{array}$$

Finite type is defined exactly as before. Note however, that for framed Vassiliev invariants, it is not necessarily the case that an isolated loop gets the value zero. For example, we have for $Z(K)$ that

$$Z_{\circlearrowleft} = Z_{\rightarrow} - Z_{\leftarrow} = (\alpha - \bar{\alpha}) Z_{\rightarrow}.$$

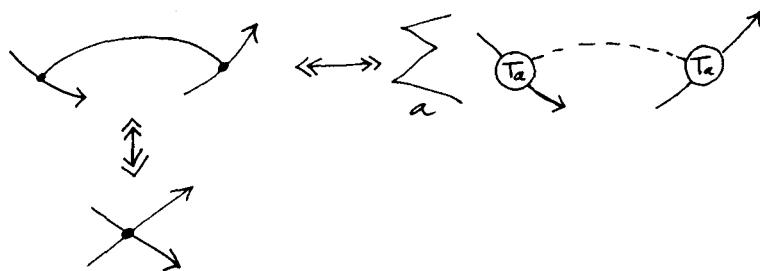
Thinking of $Z(K)$ as a framed Vassiliev invariant we see that the vertex has the structure

$$Z(K_\&) = Z(K_+) - Z(K_-) = (-4\pi i/k) Z(T_a T_a K_\#) + O(1/k^2).$$

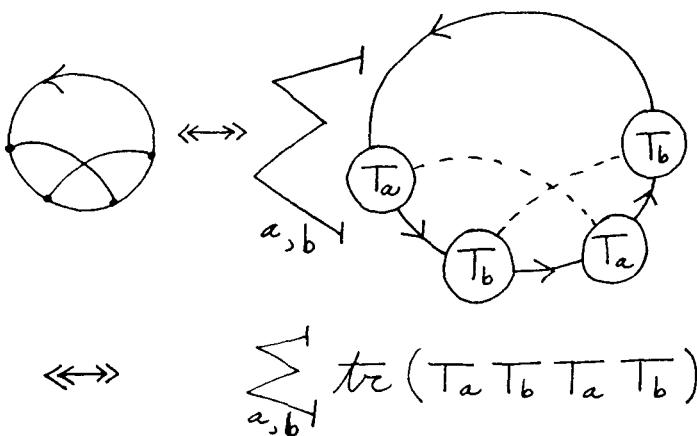
Here $K_\&$ denotes the result of replacing the crossing in K_+ or K_- with a node that represents the graphical node for the framed Vassiliev invariant associated with $Z(K)$.

Thus we see that $(1/k)$ divides $Z(K_\&)$, and hence that $(1/k)^i$ divides $Z(G)$ whenever G has i nodes. It follows that the coefficient $Z_i(K)$ of $(-4\pi i/k)^i$ in this expansion of $Z(K)$ is a (*framed*) Vassiliev invariant of type i .

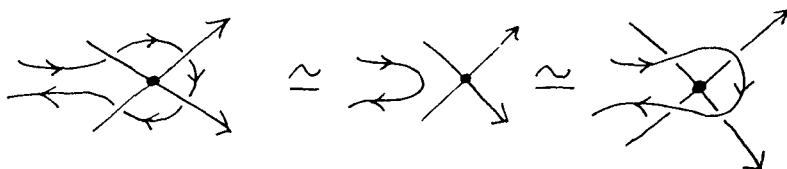
If G has exactly i nodes, and we want to compute $Z_i(G)$, then we need to look at the result of replacing every node of G by the Casimir insertion. The resulting evaluation of the graph G is independent of the embedding of G in three-space. Heuristically, this means that at this level (that is, computing $Z_i(G)$ with G having i nodes) we are no longer computing the functional integral but rather only the trace of the matrix product that is inserted into the Wilson line. The result is most easily said in terms of the chord diagrams. For each pair of matched points in the chord diagram place a pair of Lie algebra generators T_a and T_a . Take the sum over a .



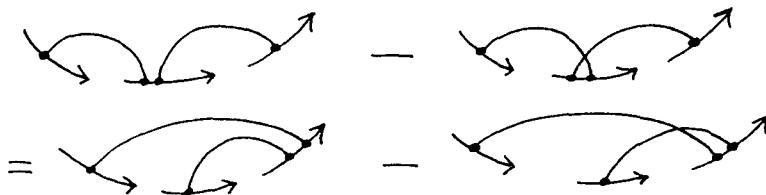
An entire chord diagram gives rise to a sum of products of T 's. Take the sum of the traces of these products. This is an abstract Wilson line (with insertions).



This method of assigning weights to chord diagrams has been studied by Bar-Natan, and he points out that such evaluations automatically satisfy the fundamental 4-term relation [ST92] that a Vassiliev invariant must obey. The 4-term relation is a direct consequence of the topological demand that the switching relations be compatible with the regular isotopy shown below.



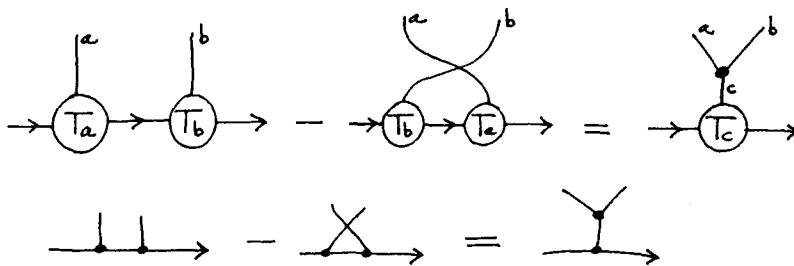
This demand translates into the following type of relation on the top row ($V_i(G)$ when G has i nodes) weightings:



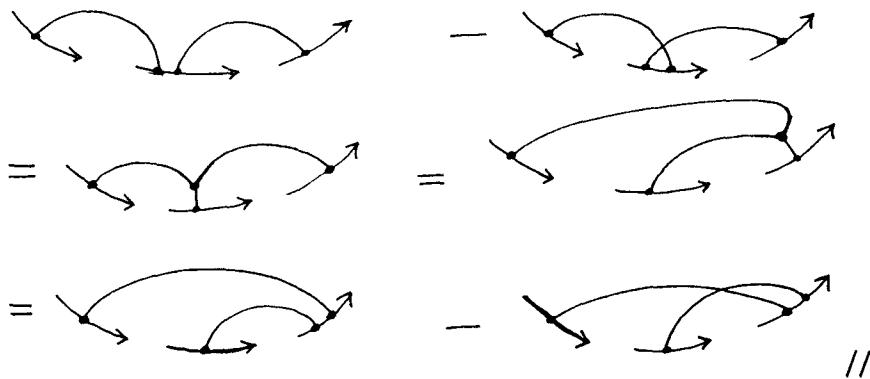
The four term relation is true for the Lie algebra weightings because it follows directly from the formula for the commutator in the Lie algebra, once this formula is translated (as shown below) into diagrammatic language.

Here is the translation of the commutator into diagrams.

$$T_a T_b - T_b T_a = i f_{abc} T_c$$



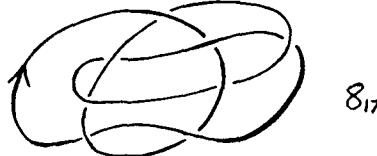
Here is a diagrammatic proof of the 4-term relation from the Lie algebra relation.



The upshot of this analysis is that the structure of Vassiliev invariants is intimately tied up with the structure of the quantum link invariants. At the top row evaluations the relationship of Lie algebra structure and topological invariance is as stark as it can possibly be.

It may be the case that all Vassiliev invariants can be constructed with combinations of quantum link invariants that arise from the classical Lie algebras. At this writing the problem is open.

As a parting shot, we mention one further problem about Vassiliev invariants. It is not known whether there exists an (unframed) Vassiliev invariant that can detect the difference between a knot and its reverse. For example, it is known that the knot 8_{17} shown below



is inequivalent to the knot obtained by reversing its orientation. It is unknown whether there is a Vassiliev invariant that can detect this difference.

References

- [BN92] D. Bar-Natan. On the Vassiliev knot invariants. (preprint 1992).
- [BL92] J. Birman and X.S. Lin. Knot polynomials and Vassiliev's invariants. *Invent. Math.* (to appear).
- [CKY1] L. Crane, L.H. Kauffman and D.N. Yetter. Evaluating the Crane Yetter invariant. *Quantum Topology*. edited by Kauffman and Baadhio, World Scientific.
- [CKY2] L. Crane, L.H. Kauffman and D.N. Yetter. On the failure of the Lickorish encirclement lemma for Temperley-Lieb recoupling theory at certain roots of unity. (preprint 1993).
- [CY93] L. Crane and D.N. Yetter. A categorical construction of 4D topological quantum field theories. *Quantum Topology*. edited by Kauffman and Baadhio, World Scientific.
- [GK93] J. Goldman and L. H. Kauffman. Knots, tangles and electrical networks. Advances in Applied Mathematics, Vol. 14 (1993), 267–306.
- [KB] L.H. Kauffman and R. A. Baadhio. *Quantum Topology*. World Scientific.

- [K] L.H. Kauffman. Spin networks, topology and discrete physics.
(to appear in *Braid Group, Knot Theory and Statistical Mechanics, II*.
ed. by Ge and Yang. World Scientific (1994)).
- [K93] L.H. Kauffman. Gauss codes, quantum groups and ribbon Hopf algebras.
(to appear in *Reviews of Modern Physics*).
- [KL] L.H. Kauffman and S. Lins. *Temperley-Lieb Recoupling Theory and Invariants of 3-Manifolds*. Princeton University Press (1994). (to appear).
- [KR] L.H. Kauffman and D. Radford. On invariants of links and 3-manifolds derived from finite dimensional Hopf algebras. (announcement 1993.
paper to appear).
- [KV92] L.H. Kauffman and P. Vogel. Link polynomials and a graphical calculus.
J. Knot Theory and Its Ramifications, Vol. 1, No. 1
(March 1992), 59–104.
- [KO92] M. Kontsevich. Graphs, homotopical algebra and low dimensional topology
(preprint 1992).
- [L] W.B.R. Lickorish. The skein method for three-manifold invariants.
J. Knot Theory and Its Ramifications, Vol. 2, No. 2 (1993), 171–194.
- [R93] J. Roberts. Skein theory and Turaev-Viro invariants. (preprint 1993).
- [ST92] T. Stanford. Finite-type invariants of knots, links and graphs.
(preprint 1992).
- [V90] V.A. Vassiliev. Cohomology of knot spaces. *Theory of Singularities and its Applications*. ed. by V.I. Arnold. Adv. in Soviet Math.,
Vol. 1, AMS (1990).

Gauss Codes, Quantum Groups and Ribbon Hopf Algebras

By relating the diagrammatic foundations of knot theory with the structure of abstract tensors, quantum groups and ribbon Hopf algebras, specific expressions are derived for quantum link invariants. These expressions, when applied to the case of finite dimensional unimodular ribbon Hopf algebras, give rise to invariants of 3-manifolds.

I. Introduction

In this paper we construct invariants of links in three space and invariants of 3-manifolds via surgery on links. These are quantum link invariants. That is, they are invariants constructed through the use of quantum groups. It is the purpose of the paper not only to produce these invariants, but also to explain just how the concept of a quantum group and the associated concept of ribbon Hopf algebra features in these constructions.

By beginning with certain fundamental structures of link diagrams, it is possible to give very clear motivations for each stage in the process of associating Hopf algebras (aka quantum groups) with link invariants. The basic link diagrammatic structures that we consider are as follows. First, there is the *Gauss code*, a sequence of symbols representing a record of the crossings encountered in taking a trip along the curve of the knot. Knots can be reconstructed from their Gauss codes (Sec. 2), and it is natural to wonder how invariants of knots may be related to the codes themselves. Secondly, there is the structure of the link diagram with respect to a height function in the plane. Construction and isotopy of knots and links can be described in regard to such a function. For construction there are

four basic building blocks (we are concerned here with unoriented knots and links) - minima, maxima and two types of crossing. These building blocks are subject to interpretation as abstract morphisms or abstract tensors. This gives rise to a first description of both particular and universal invariants (Secs. 3 and 4). In the course of this description of link invariants, we give a rigorous formulation of the concept of an *abstract tensor algebra*. Abstract tensors are distinct from tensor calculus or linear tensor algebra; they are a direct abstraction of the actual practice of working with tensors. In this practice, certain entities are manipulated according to algebraic rules while under the constraint of being tied (through common indices) to other such entities. By bringing forth the concept of indices as constraints, we open a field of representation for abstract tensors where the indices are replaced by connective tissue. In this way, two boxes connected by a line or arc can represent two tensors with a common index. While it is usual to sum over a common index in ordinary tensor algebra, this is only one of many representations of abstract tensor algebra. As a result, abstract tensor algebra provides a language that intermediates between linear tensors and the topological/geometrical structure of certain systems. Knots and links are the main examples of such systems in this paper. Section 4 details the use of abstract tensor algebra for knot and link invariants.

In Sec. 5 we begin the transition to Hopf algebras by considering the form of an abstract algebra that would describe the tensor algebra associated with link invariants. Certain features of the abstract algebra are already present in natural algebraic structures on tangles. In particular, one has tangle multiplication, and the parallel doubling of lines in a tangle is analogous to a coproduct. (With proper definitions such doubling will map to a coproduct in an associated Hopf algebra.) Furthermore there is a natural candidate in tangles for an analogue to the antipode in a Hopf algebra. Section 5 defines these notions without reference to a specific representation of the quantum algebra. Rather, the quantum algebra is seen to decorate the lines of the knot diagram, and this notion of decoration is made precise. These results apply directly to finite dimensional algebras. Then a given knot is mapped to a sum of products of elements of the algebra, and an appropriate trace on the algebra will yield an invariant of the knot. This sum of products is obtained from the diagram by reading a generalization of the Gauss code. In this way the Gauss code is fundamental to our constructions.

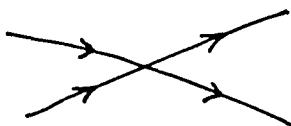
In Sec. 6 we make the precise connections between our formalism and the quasitriangular and ribbon Hopf algebras. Our formalism exhibits ribbon elements and the action of the antipode directly. The fact that the square of the antipode is given by conjugation by a grouplike has a direct interpretation in terms of the immersions shadowed by a link projection and the topology of regular homotopies of immersed curves in the plane. Section 7 puts our results in a categorical framework. Section 8 shows how to apply our formalism in the case of finite dimensional unimodular Hopf algebras to produce Hennings-type invariants of 3-manifolds. Here we use the right integral on the Hopf algebra as a trace. No representation of the Hopf algebra is necessary. This section describes joint work of the author and David Radford. Further details will appear elsewhere. Section 9 is an epilogue.

The present paper completes and considerably extends an earlier version of this work [K7]. This research was partially supported by NSF Grant Number DMS-9205277 and by the Program for Mathematics and Molecular Biology, University of California at Berkeley, Berkeley, California.

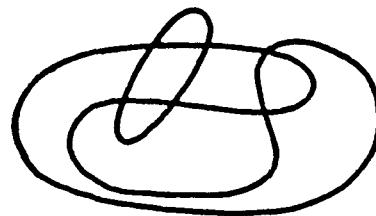
II. Knots and the Gauss Code

This section recalls the method of forming the Gauss code for a knot or link diagram, and one method of decoding the knot or link from the code. The Gauss code (or trip code) is a particularly simple method for transcribing the information in a knot or link diagram into a linear code. We shall see in the next sections that the Gauss code is very closely tied to the construction of link invariants via quantum groups.

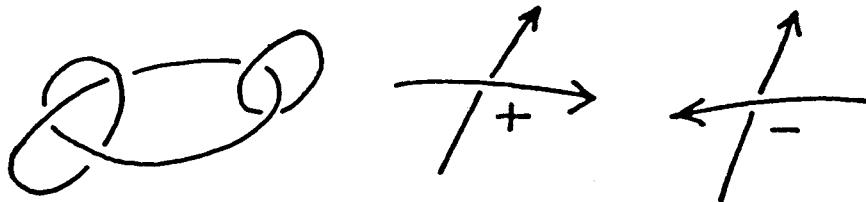
In order to describe the Gauss code, it is convenient to first encode the underlying four-valent planar graph for a link diagram. I shall refer to this graph as a *universe* [K1]. One counts components of a universe in the sense of components of an overlying link diagram. That is, one considers the closed curves that are obtained from a given universe of travelling along it in a pattern that crosses at each crossing as shown below.



The number distinct closed curves of this type is called the *number of components* of the universe. This universe shown below has two components.

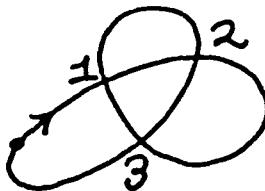


A universe with one component is called a *knot universe*. A *link diagram* is a universe with extra structure at the crossings indicating over and underpasses in the standard convention illustrated below:

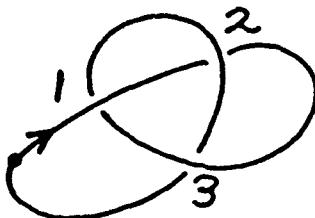


We can also indicate the standard conventions for positive and negative oriented crossings in a link diagram. In an oriented universe or diagram, a direction of travel has been chosen for each component. This gives rise, in a link diagram, to a sign at each crossing. The *sign is positive if the overcrossing line goes from left to right for an observer travelling toward the crossing on the undercrossing line*. If the overcrossing line goes from right to left, then the crossing is negative.

The simplest Gauss code is the code for a single component unoriented universe. Here we choose, as a basepoint, an interior point of an edge of the universe. Then, choosing a direction of travel, we read the names of the crossings in order as we traverse the diagram (crossing at each crossing). Call this traverse, a *trip* on the diagram. For example, the Gauss code for the universe shown below is 123123 where the crossings are labelled 1, 2 and 3.



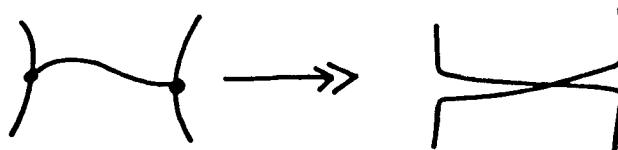
In order to indicate a link diagram that overlies the given universe, it is sufficient to add information at each crossing. One way to add information is to indicate whether the trip goes through the crossing on an overpass (O) or an underpass (U). Thus we get the augmented trip code O1U2O3U1O2U3 for the trefoil diagram shown below. The arrow on diagram indicates the starting point for the trip.



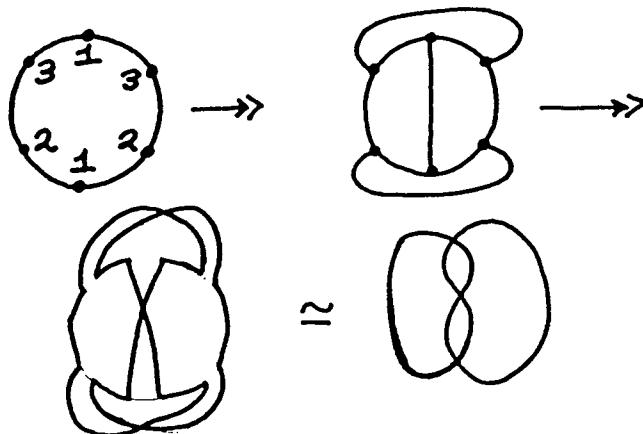
There is a very beautiful method for decoding the link universe from its trip code that is due to Dehn [D] for single component links and to Rosenstiehl and Read [R] in the general case. Rosenstiehl and Read solve completely the problem originally raised by Gauss [G], concerning when an abstract code actually represents a planar universe. The upshot of the procedure that we are about to describe is that, *given the Gauss code of a prime diagram, one can retrieve the knot type of the diagram from the code.* (A diagram is not prime (composite) if it can be decomposed into two diagrams by a simple closed curve in the plane that intersects the original diagram in two points and so that the parts of the diagram that are in the two components delineated by the simple closed curves are each nontrivial. A trivial diagram in this instance is an embedded arc.)

The idea for this decodement is as follows: by splicing the crossings of a knot universe it is possible to obtain (in various ways) a single simple closed curve (Jordan curve) in the plane. This curve has a trip code composed of the same labels as the code for the original universe, but arranged in a

different order. If we know the trip code for the Jordan curve, then we can reconstruct the knot universe by first drawing a circle in the plane with the given order of labels upon it. After drawing this circle, connect pairs of points with the same labels by arcs in such a way that no two arcs intersect one another. This may entail drawing some of the arcs outside and some of the arcs inside the given circle. (When this is impossible, then the code has no planar realisation.) Once the arcs are drawn, they can be replaced by crossings.



For example, suppose that 132123 is the trip code for the Jordan curve. Then we draw the Jordan curve, and reconstruct the knot universe as shown below:



The key to the full decoding procedure is the determination of a Jordan curve code from the trip code for the universe in question. A combinatorial observation about the link diagrams provides this key. In a knot universe,

one way to splice a crossing produces a two component universe, the other a one component universe. A sequence of splices that retains one component will produce the Jordan curve. *Splices that retain a single component reverse the order of the terms in the code between two appearances of the label for the crossing being spliced.* The upshot of this observation is that we can obtain the trip code for the Jordan curve from the trip code for the knot universe by successively choosing each label, and switching the order of terms between the first and second instances of that label. To be systematic, we start from the smallest label and work upwards. For example, if the knot universe has the Gauss code 123123 then this switching procedure will go through the following stages

123123 (flip 1)

132123 (flip 2)

132123 (flip 3)

132123 .

Here there are some redundant stages. A more complex decodment is shown below for a diagram with trip code 123451637589472698.

123451637589472698 (flip 1)

154321637589472698 (flip 2)

154327498573612698 (flip 3)

154375894723612698 (flip 4)

154985734723612698 (flip 5)

158945734723612698 (flip 6)

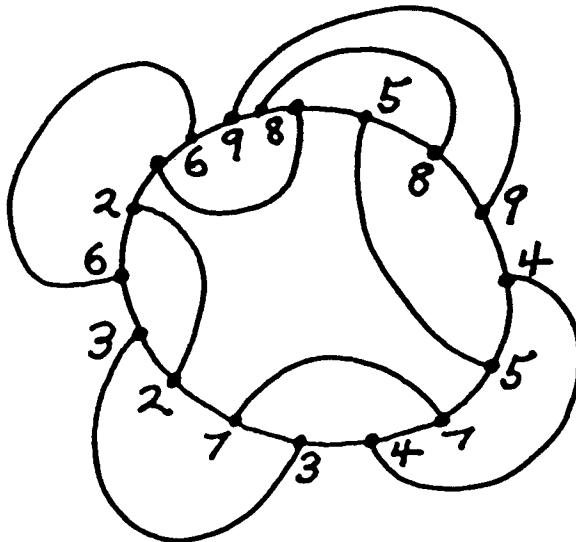
158945734723621698 (flip 7)

158945743723621698 (flip 8)

158961263273475498 (flip 9)

158945743723621698

In this case, the diagram below reveals that the original knot universe code 123451637589472698 can be realised in the plane.



I leave it to the reader to complete the construction of a planar embedding for the original code.

Along with the code, we can also write the changes on the Gauss code that correspond to the Reidemeister moves. Figure 1 shows the moves (one example of each) and the corresponding transformations of the Gauss code. Note that the third Reidemeister move corresponds to a pattern of reversal of order of three pairwise products in the word. Later, we shall see how this pattern is related to the Yang-Baxter Equation.

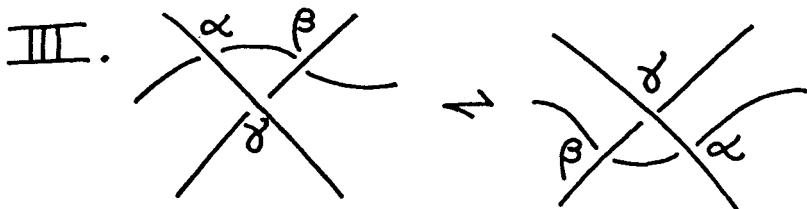
This completes our tour of the basic facts about the Gauss code. The theory of knots and links in three dimensional space is a theory about information in these circular sequential codes. This point of view has been useful for computer computation of link invariants. Here we shall point out its significance for the relationships between link theory and quantum groups.



$$w\alpha\alpha\tau \rightarrow w\tau$$



$$w\beta\alpha\tau\alpha\beta\lambda \rightarrow w\tau\lambda$$



$$w\alpha\beta\tau\alpha\gamma\lambda\beta\delta\mu \rightarrow w\beta\alpha\tau\gamma\alpha\lambda\delta\beta\mu$$

Reidemeister Moves and Corresponding Changes on Gauss Codes

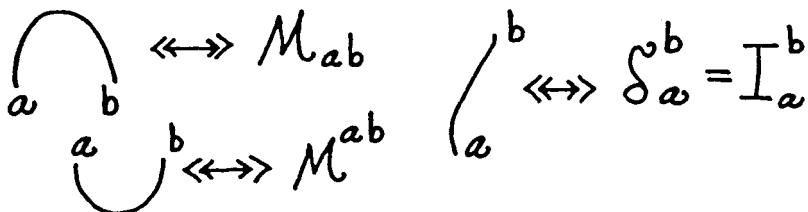
Figure 1

III. Jordan Curves and Immersed Plane Curves

A Jordan curve is a simple closed curve in the plane that is free from self-intersections. We assume that the curve is described by a smooth function so that concepts of general position, tangency and transversality can be applied to it. It then follows from general position arguments that a generic Jordan curve can be decomposed, with respect to a fixed direction, into segments that we call *cups*, *caps* and *arcs*. A cup is a local maximum with

respect to the chosen direction. A cap is a local minimum. An arc is a segment that contains no local maxima or minima, and so that the normals to the chosen direction intersect it in either one point or in no points. (Cups and caps have non-empty intersections of either two points or one point, and only one intersection of one point, that begin the maximum or minimum as the case may be.)

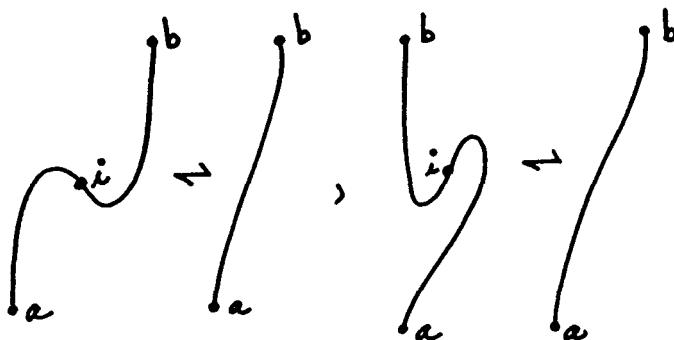
It is convenient to symbolize these cups, caps and arcs by attaching labels to their endpoints, and using the symbols M_{ab} , M^{ab} , l_b^a for caps, cups and arcs respectively. Here the letters a, b denote the labels at the endpoints. See the illustration below. We shall not orient the cups and caps unless it is necessary.



With respect to the chosen direction, ambient isotopy of Jordan curves is generated by the cancellation of adjacent maxima and minima as shown below. This symbolism takes the form shown directly below

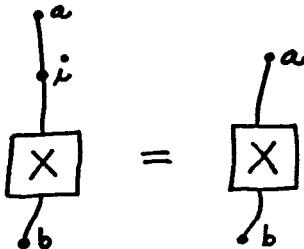
$$M_{ai} M^{ib} = l_b^a$$

$$M^{ai} M_{ib} = l_b^a$$



Since arcs are redundant, appending an arc is an identity operation:

$$l_i^a X_b^i = X_b^a$$



With these rules, the symbolism allows us to transcribe a Jordan curve, and to reduce it to the form of a circle:

$$\begin{aligned} \textcircled{} &\iff \bullet_a \bullet_b \iff M_{ab} M^{ab} \\ \textcircled{a} \curvearrowright b \curvearrowright c \curvearrowright d \bullet_d &\iff M_{ab} M^{bc} M_{cd} M^{ad} \\ &= I_a^c M_{cd} M^{ad} \\ &= M_{ad} M^{ad} \iff \bullet_a \bullet_d \end{aligned}$$

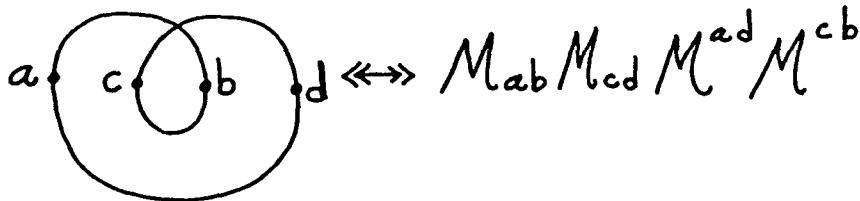
It is interesting to note that there must exist a cancelling pair of maxima and minima in any Jordan curve unless it is in the form of a circle ($M_{ab} M^{ab}$). This is easily seen by thinking about the relation of left and right in the construction of plane curves. For example, in the picture below, a continuation from c either produces a cancelling pair or it generates an endpoint that is to the right of c , and hence to the right of a .



In order for the curve to close up, a cancelling pair must be produced. As a result of this observation about the existence of cancelling pairs, we obtain a proof of the Jordan Curve Theorem [S] in the form: *Any Jordan curve is isotopic to a circle. Hence, any Jordan curve divides the plane into two disjoint connected components, each homeomorphic to a disk.*

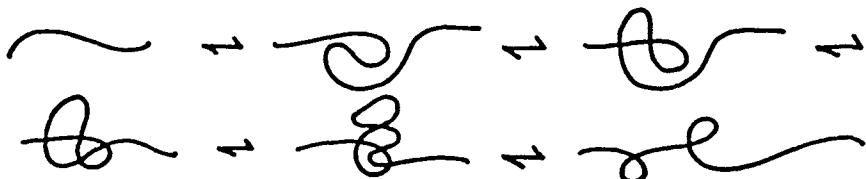
The symbolism we have used for the Jordan curves is reminiscent of the formalism for matrix multiplication. In the next section, we shall see how to represent it via matrices.

Just as in the case of the Gauss code, it is possible to have a word in cups, caps and arcs that does not reduce to a circle. This can be regarded as a plane curve with self-crossings as illustrated below:



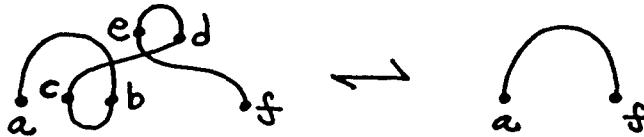
In this case the essential information contained in the min/max code for the curve is the Whitney degree [W] of the immersed curve. This is completely determined once a direction is chosen, and it is easy to compute from the code (augmented by one orientation). The Whitney degree is the total turn of the tangent vector of the immersed plane curve. I shall explain how to compute it from the min/max code in the next few paragraphs.

The important point to understand about the min/max code is that it is invariant under projected moves of Reidemeister type II and III. This is easily seen, and we leave it as an exercise. As a consequence we have the Whitney trick:

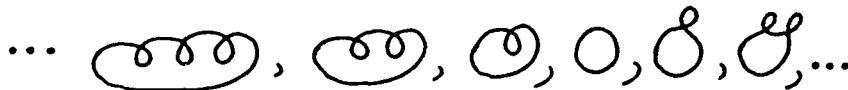


In code the Whitney Trick is the following maneuver:

$$M_{ab} M^{cd} l_c^d M_{ed} l_f^e = M_{ab} M^{db} M_{fd} = M_{ab} l_f^b = M_{af}$$

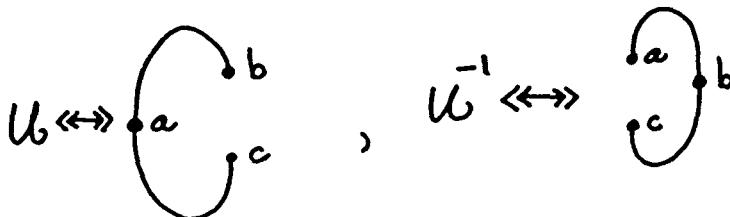


The upshot of the Whitney trick is that any immersed plane curve is regularly homotopic (Regular homotopy is generated by the projected Reidemeister moves of type II and type III.) to a curve that is in one of the standard forms shown below:



This result is known as the Whitney-Graustein Theorem [W].

The code for an immersed curve in the plane is a product of elements U^{-1} and U where these denote elements of the form $U = M_{ab} M^{cb}$ and $U^{-1} = M_{ab} M^{ac}$.



By convention we take U to have Whitney degree 1, and U^{-1} to have Whitney degree -1. Then the Whitney degree of the plane curve with a given min/max code can be read off from a reduced form of the code word.

This completes our description of the min/max coding for immersed plane curves and its relation to the Whitney Graustein Theorem. We next turn to a formalism for regular isotopy invariants of links that involves a mixture of the Gauss code and the min/max code.

IV. The Abstract Tensor Model for Link Invariants

By augmenting the min/max code to include information about the crossing in a link diagram, we obtain a complete symbolic description of the link diagram up to regular isotopy. (Regular isotopy is generated by the Reidemeister moves of type II and III.)

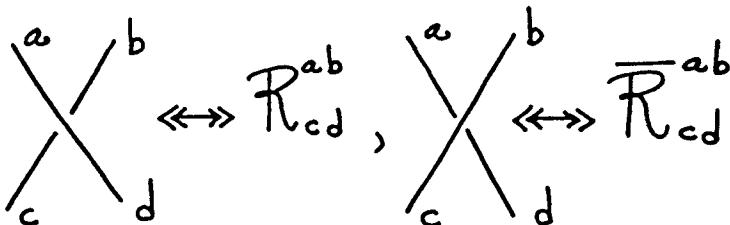
In this section we deal with the abstract form of this symbolism.

Along with the cup, cap and identity arc formalisms of the previous section, we now add symbols for the crossings in the diagram and further relations among these symbols that parallel the Reidemeister moves.

A crossing will be denoted by

$$R_{cd}^{ab} \text{ or } \bar{R}_{cd}^{ab}$$

according to its type with respect to a chosen direction. We shall take this direction to be the vertical direction on the page. Thus the R without a bar refers to a crossing where the overcrossing line proceeds from lower right to upper left, while the barred R refers to a crossing where the overcrossing line proceeds from lower left to upper right. This is illustrated below.



Note that in this abstract tensor algebra individual terms commute with one another. All non-commutativity is contained in the index relations, and these involve the distinctions left, right, up and down in relation to the chosen direction. In the tensor notation, that chosen direction has become the vertical direction of the written page.

Reidemeister moves with respect to a vertical direction take the forms indicated in Fig. 2. In this Figure are shown representatives for the moves that generate regular isotopy. Thus the moves are labelled I, II, III, IV, V where II and III are the usual second and third Reidemeister

moves in this context. Move I asserts the cancellation of adjacent pairs of maxima and minima. The move IV is an exchange of crossing and maxima or minima that must be articulated when working with respect to a given direction. Move V is a direct consequence of move I and move IV; it expresses the geometric “twist” relationship between one crossing type and the other. Each of these moves has an algebraic counterpart in the form of the abstract tensor algebra. A list of these algebraic moves is given below. (Here we have codified just the abstract tensor algebra for unoriented links for the sake of simplicity. The oriented versions are similar but more complicated to write in this formalism.)

The moves are labelled, Cancellation, Inverses, Yang-Baxter Equation, Slide and Twist. The Yang-Baxter Equation is an abstract form of the Yang-Baxter Equation for matrices. One method for representation of the abstract tensor structure is to assume that the indices range over a finite set, and repeated upper and lower indices connote a summation over this set (the Einstein summation convention). With this convention, the axioms become demands on the properties of the matrices.

It is worth noting that the axioms for an abstract tensor algebra and the distinctions between cups, caps and the two forms of interaction are an articulation of the basic distinctions up, down, left and right in an oriented plane.

Definition. The *abstract tensor algebra*, ATA, is the free (multiplicative) monoid and free additive abelian group on the symbols

$$M_{ab}, M^{cd}, I_b^a, R_{cd}^{ab}, \bar{R}_{cd}^{ab}$$

modulo the axioms stated below. The indices on these symbols range over the English alphabet augmented by any extra conventional set of symbols that are appropriate for a given application. Thus elements of ATA are sums of products of symbols modulo commutativity and the relations that are expressed by the axioms. It may be convenient later to tensor ATA with a commutative ring or field to obtain appropriate “scalar coefficients for these symbols.

An *expression* in ATA is a product of the basic symbols that generate ATA. The axioms are rules of transformation on expressions. Two expressions in ATA are equivalent if and only if there is a finite sequence of transformations from one to the other.

Index and Substitution Conventions. Unless it is specified directly that indices are identical it is assumed that they are distinct. This leads to the following conventions for substitution:

1) Since this is an abstract tensor algebra, we regard an index as a *dummy index* when it is repeated more than once. (In the standard model of tensor algebra one sums over all occurrences of such an index.) A dummy index can be changed, so long as all occurrences of it are changed in the same way. Thus $M^{ai}M_{ib}$ has dummy index i , and we can write

$$M^{ai}M_{ib} = M^{aj}M_{jb} \text{ so long as } j \text{ is distinct from } i \text{ and from } a \text{ and } b.$$

Similarly, $M^{ii}R_{ic}^{ai}$ has a dummy index i , and all occurrences of i must be changed simultaneously (to values distinct from a and b).

2) Suppose X is a given expression in ATA. Suppose that X occurs as a subexpression of an expression Y in ATA. Let $Y//X$ denote the expression obtained by erasing X from Y . Suppose that $X = X'$, and that the dummy indices of X' are distinct from the dummy indices of X in Y and the dummy indices of X' are also distinct from all the indices of $Y//X$ that are not dummy indices of X . Then $Y = X'(Y//X)'$ where $(Y//X)'$ is obtained from $Y//X$ by replacing all dummy indices of X in $Y//X$ by the corresponding dummy indices of X' .

For example, $Y = M^{ai}R_{ij}^{cd}M^{ji}$, $X = M^{ai}R_{ij}^{cd} = X' = M^{ak}R_{kj}^{cd}$, then $Y//X = M^{ji}$ and $(Y//X)' = M^{jk}$ and $X'(Y//X)' = M^{ak}R_{kj}^{cd}M^{jk}$.

These rules of substitution amount to little more than the statement that if a dummy index is changed, then the change must be applied to all instances of the dummy index. Furthermore, a dummy index can not be replaced by an index that is already being used in the given expression.

Remark. The point of the abstract tensor symbolism is that it serves as a symbolic code (for link diagrams) that is also subject to numerous algebraic interpretations. While these symbols look like tensors, *there is no summation on the indices*. The *identity of the indices is very important* — two symbols that share an index represent fragments of a diagram that are tied together at the site of this index. Abstract tensor algebra is a significant departure from other forms of abstract algebra. In the abstract tensor algebra there are symbols, separated by typographical distance, that are interconnected by their indices. If two symbols have a common index, then any change in that index in one of them must be accompanied by a

corresponding change in the index of the other symbol. This describes exactly the process we are familiar with in the context of using a “dummy” summation index, but in the abstract tensor algebra there is no summation over a common index (although this will occur in certain representations of the algebra). Anyone who has performed computations using the Einstein summation convention is familiar with the process of abstract tensor algebra. Here this process is given a mathematical definition. (Our definition is very similar to that given for abstract tensor systems for Penrose [P], except that in [P] there is an assumed background of linear algebra.)

Regular Isotopy Axioms for Abstract Tensor Algebra

I. $M_{ai}M^{ib} = l_a^b$, $M^{ai}M_{ib} = l_b^a$ (**Cancellation**)

l_b^a is an identity object in the sense that, given any expression X_c^a in ATA, then $l_b^a X_c^b = X_c^a$ and $l_b^a X_a^c = X_b^c$. Here the object X shares an index with l .

II. $R_{ij}^{ab} \bar{R}_{cd}^{ij} = l_c^a l_d^b$ (**Inverses**)

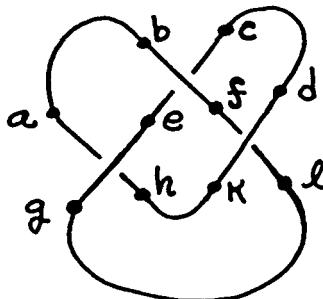
III. $R_{ij}^{ab} R_{kf}^{jc} R_{de}^{ik} = R_{ki}^{bc} R_{df}^{ak} R_{ef}^{ji}$ (**Yang-Baxter Equation**)

IV. $M_{ab} R_{de}^{bc} = \bar{R}_{ad}^{cb} M_{be}$, $R_{cd}^{ab} M^{de} = M^{ac} \bar{R}_{cd}^{be}$ (**Slide Move**)

V. $\bar{R}_{cd}^{ab} = M_{ci} R_{dj}^{ia} M^{jb}$, $\bar{R}_{cd}^{ab} = M^{ac} R_{ic}^{bj} M_{jd}$ (**Twist Move**)

These axioms parallel directly the regular isotopies shown in Fig. 2. The twist move is a consequence of the other four axioms. It is a direct consequence of axioms IV and I.

We translate an (unoriented) link diagram into an expression in the abstract tensor algebra by arranging the diagram for the link so that it decomposes into cups, caps and crossings with respect to the vertical. (I shall refer to the special direction as the vertical.) Label the nodes of the decomposition and write the corresponding tensor. The diagram below illustrates this process for a right handed trefoil diagram. Given a diagram K , we let $T(K)$ denote the corresponding abstract tensor. Thus if K is the trefoil diagram shown below, then $T(K) = M_{ab} M_{cd} R_{ef}^{bc} \bar{R}_{gh}^{ae} \bar{R}_{kl}^{fd} M^{gl} M^{hk}$. The abstract tensor corresponding to a given link diagram (no free ends) will have no free indices. Each upper index in the tensor is uniquely paired with a lower index and vice versa.



$$T(K) = M_{ab}M_{cd}R_{ef}^{bc}R_{gh}^{ae}R_{kl}^{fd}M^{gl}M^{hk}$$

Theorem 4.1. Let K be an (unoriented) link diagram arranged in standard position with respect to the vertical. Let $T(K)$ be an abstract tensor expression (free of identities - l) corresponding to K . Then the diagram K can be retrieved from $T(K)$. (Here it is assumed that the axioms other than the use of the identity $-l$ for the tensor algebra are not applied. That is, $T(K)$ is taken simply as a code for the link diagram.)

Proof. To retrieve the link diagram from $T(K)$ first place crossings corresponding to the R 's in the expression $T(K)$. (I shall say the R 's when referring to R 's or R -bars.) Note that in $T(K)$ every upper index is matched by a unique lower index. Choose any R , and draw in a plane a glyph corresponding to that R . Now choose an index on the R in question and find the other R (there is a unique other) that contains this index. Place another glyph in position with respect to the first so that their indices are matched. Note that the relative placement is unique. For example, the first R may have a lower left index to be matched with an upper right index of the second R . In this case the second R is placed to the lower left of the first. After all the R 's have been successively interconnected in this manner, place the cups and caps corresponding to the M 's. Since the expression is known to come from a link diagram, this is possible, and will give a reconstruction of the diagram. We leave the placement of the cups and caps for last because the sites for their nodes are then chosen by the placements of the R 's. This completes the proof.//

Remark. This proof can be regarded as a generalization of the usual procedure for writing a braid from an expression in the braid group. It is particularly important for our considerations, because of the next result.

Theorem 4.2. Let K and L be two links arranged in standard position with respect to the vertical. Let $T(K)$ and $T(L)$ be corresponding expressions for K and L in the abstract tensor algebra, ATA. Then K and L are regularly isotopic as link diagrams if and only if the expressions $T(K)$ and $T(L)$ are equivalent in the abstract tensor algebra.

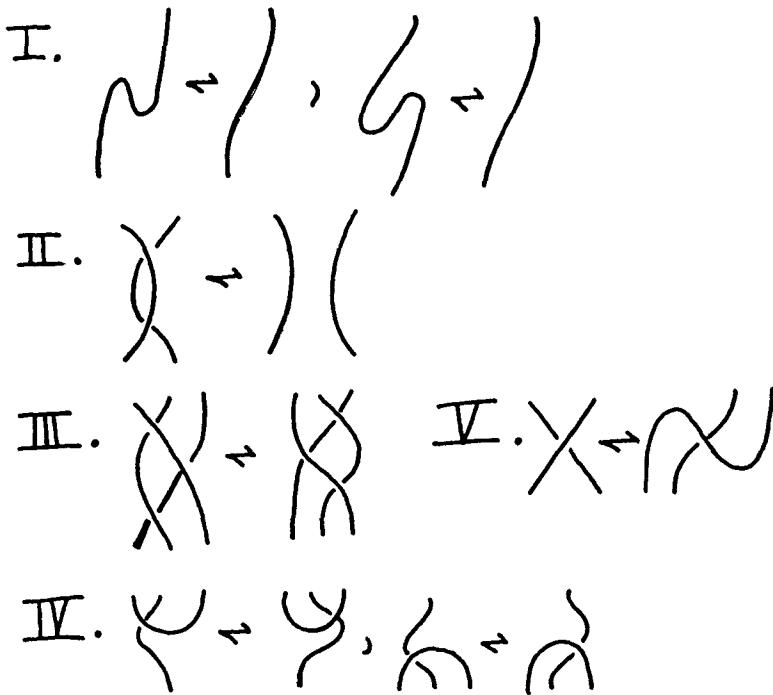


Figure 2

Proof. If K is regularly isotopic to L then there is a sequence of the diagrammatic moves I, II, III, IV (V) taking K to L (see e.g. [K3]). This is the reason for articulating these particular moves with respect to a direction.) By the translation from moves to abstract tensors, this gives a sequence of axiomatic transformations from $T(K)$ to $T(L)$. Hence $T(K)$ and $T(L)$ are equivalent in ATA. Conversely, if $T(K)$ and $T(L)$ are equivalent in ATA, then there is a sequence of axiomatic transformations from $T(K)$ to $T(L)$. Apply Theorem 1 to each expression in this sequence and obtain a regular isotopy from K to L . This completes the proof. //

Remark. We have shown that the association $K \rightarrow T(K)$ of links to expressions in ATA is a faithful representation of the regular isotopy category of links. When ATA is represented by actual matrices, and the Einstein summation convention, then this association gives invariants of knots and links such as the Jones polynomial and its generalizations. We will give specific examples in the next section. At the abstract level, the correspondence gives a “complete invariant” for knots and links. This suggests that if there are sufficiently many representations, then the matrix invariants derived from ATA may be able to distinguish topologically different links. I conjecture that any given matrix invariant will not separate all distinct knots and links, but that the collection of all representations of ATA does have this property.

The Bracket Polynomial

We now discuss the bracket polynomial (see [K5]) in the context of the abstract tensor algebra. Recall that the bracket polynomial is a regular isotopy invariant of link diagrams that satisfies the following two formulas:

$$1. \langle \text{X} \rangle = A \langle \text{ } \rangle + A^{-1} \langle \text{ } \rangle$$

$$2. \langle OK \rangle = d \langle K \rangle, \quad d = -A^2 - A^{-2}$$

$$\langle O \rangle = d$$

In other words, the bracket polynomial for a given link decomposes into a weighted sum of bracket polynomials of links obtained by smoothing any given crossing in the two possible ways illustrated above. The smoothing that receives the A coefficient is the one that joins the two regions obtained by turning the overcrossing line counterclockwise. These rules are sufficient to calculate the bracket polynomial from any finite diagram.

The abstract tensor formulation of the bracket invariant is obtained by adding the following relations to ATA:

$$1. R_{cd}^{ab} = A M^{ab} M_{cd} + A^{-1} l_c^a l_d^b$$

$$\bar{R}_{cd}^{ab} = A^{-1} M^{ab} M_{cd} + A l_c^a l_d^b$$

$$2. M^{ab} M_{ab} = d = -A^2 - A^{-2}$$

(Here we take ATA as a module over $Z[A, A^{-1}]$.)

Call this quotient module $\langle \text{ATA} \rangle$ - the bracket quotient of ATA.

The equivalence classes of link tensors $\langle T(K) \rangle$ in $\langle \text{ATA} \rangle$ are then represented uniquely by the bracket polynomials $\langle K \rangle$.

For example, here is the beginning of the reduction of the abstract tensor corresponding to the right-handed trefoil knot to its bracket polynomial:

$$\begin{aligned} T(K) &= M_{ab} M_{cd} R_{ef}^{bc} \bar{R}_{gh}^{ae} \bar{R}_{kl}^{fd} M^{gl} M^{hk} \\ &= M_{ab} M_{cd} (A M^{bc} M_{ef} + A^{-1} l_e^b l_f^c) (A^{-1} M^{ae} M_{gh} + A l_g^a l_h^e) \\ &\quad (A^{-1} M^{fd} M_{kl} + A l_k^f l_l^d) M^{gl} M^{hk} \end{aligned}$$

While it is obviously very cumbersome to calculate in this formalism, it is interesting to see that it is possible to phrase this chapter of the knot theory as pure algebra.

Example. There is a simple matrix representation of $\langle \text{ATA} \rangle$ given by taking the entries of the matrix $M = \begin{bmatrix} 0 & \sqrt{-1}A \\ -\sqrt{-1}A^{-1} & 0 \end{bmatrix}$ so that $M^{ab} = M_{ab}$ are both interpreted as ranging over the entries of this matrix with a and b belonging to the index set $\{0, 1\}$, and the expressions in ATA are sent to sums of products via the Einstein summation convention. For example we then have $M^{ab} M_{ab}$ represented as the sum of the squares of the entries of the matrix M , and this is exactly $d = -A^2 - A^{-2}$, as desired. With this choice of M , the R -matrix will satisfy the Yang-Baxter equation, and all the other axioms will be satisfied. See [K6] for a more detailed description of this representation and its relation to the $\text{SL}(2)_q$ quantum group.

This example of the quotient module $\langle \text{ATA} \rangle$ raises many questions about the nature of invariants on knots and links. The intrinsic algebra structure of ATA alone is quite complex, and it contains all the problems of the theory of knots and links. In the next section we see how this structure is intimately tied with the notion of a quantum group.

V. From Abstract Tensors to Quantum Algebras

Abstract tensors associate matrix-like objects to the crossings, maxima and minima of a link diagram. (The diagram is arranged with respect to a given direction that we take to be the vertical direction of the page).

The Gauss code (Sec. 2) suggests considering non-commutative algebra elements arrayed along the edges of the link diagram. The natural one-dimensional ordering imposed by the edges keeps track of the order of multiplication.

These two points of view come together when we consider the *algebraic structure of tangle multiplication*. A tangle is a piece of knot or link diagram, confined to a box, with free strand ends emanating from the box. There are no free ends inside the box, but loops are allowed inside. In isotoping a tangle we allow isotopies inside the tangle box, but the strands emanating from the box must remain fixed. Thus two tangles are isotopic if there is an isotopy of one to the other that is restricted to their respective tangle boxes. The boundaries and lines on the boxes are fixed during the isotopy.

Usually, we designate a subset of the free ends as “inputs” to the tangle and another subset as “outputs”. Thus a single strand tangle is a tangle with one input and one output line as illustrated in Fig. 3. In that figure the input is indicated on the bottom of the tangle and the output is indicated on the top.

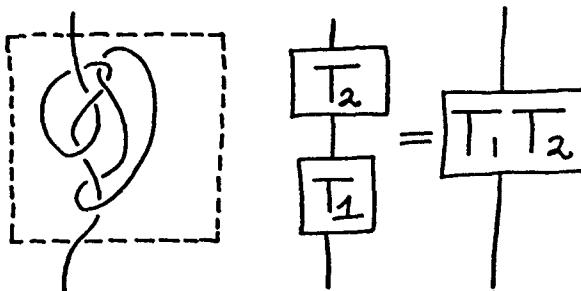


Figure 3

Single strand tangles can be multiplied by attaching the output of one tangle to the input of the other tangle. In this picture the order of multiplication for tangles proceeds from the bottom of the page to the top of the page. See Fig. 3.

Let L denote the set of regular isotopy classes of unoriented single strand tangles. Define $s : L \dashrightarrow L$ by the operation shown in Fig. 4. Call s the *tangle antipode*. In words, $s(A)$ is obtained by reversing the input and the output of the tangle A . This is accomplished by bending the input upwards and the output downwards. It should be clear from Fig. 4 that s is an antimorphism of the multiplicative structure of L . We shall prove this in more generality shortly. Another mapping that it is natural to associate with L is a coproduct map from L to $L^{(2)}$ denoted $\Delta : L \dashrightarrow L^{(2)}$. Here

$L^{(2)}$ is the set of two strand tangles — in the sense of two input strands and two output strands. With the input and output strands ordered from left to right, the two strand tangles have a multiplicative structure via standard tangle multiplication. The same remarks apply to n strand tangles.

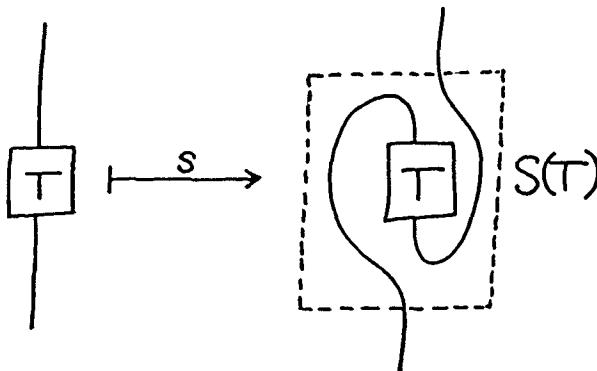


Figure 4

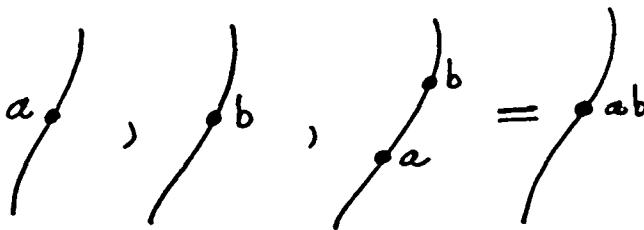
The mapping Δ associates to the tangle A a new tangle obtained from A by replacing each component of A by a two strand parallel cable of that component (parallel meaning parallel with respect to the natural framing in the plane). See Fig. 4.

There is also a natural “forgetful map” $\varepsilon : L \dashrightarrow l$ where l denotes the (one element) algebra generated by the tangle with a single strand. Suitably generalized, this map becomes the counit in the corresponding quantum algebra.

The mappings s and Δ are the tangle theoretic analogues of mappings of an algebra A where $\Delta : A \dashrightarrow A \otimes A$ is an algebra map from A to the tensor product of A with itself. A quasi-triangular Hopf algebra comes equipped with just such structures and we shall see that the tangle algebra is exactly the appropriately intermediary between the abstract tensor algebra and the quasi-triangular Hopf algebra.

Before proceeding of Hopf algebras it is useful to define the concept of algebra elements on the lines of a link diagram. Diagrammatically, this means that we choose a place on the diagram, arranged with respect to the vertical, and we place there an algebra element. If another algebra element is placed above a given one, on the same line, then the combination of the two elements is regarded as their product. Thus we have indicated a, b

and ab in the diagram below. This sort of diagrammatics is quite useful. It requires a mathematical definition such a hybrid of algebra and link diagrams is an unusual amalgam.



The abstract tensor algebra holds the key. Let A be any algebra with an associative, not necessarily commutative multiplication. Let a be an element of A . A *line algebra element* is by definition an abstract tensor of the form $[a]_j^i$ with the rule of combination $[a]_j^i [b]_k^j = [ab]_k^i$. In this notation, a line element is simply an element of the algebra that has been placed inside a bracket whose indices give the placement of this element in the abstract tensor algebra. This formalism is very useful since it gives an exact definition of the desired amalgam. In diagrams we seldom indicate both the tensor indices and the algebra elements since this is hard to read.

At this stage we could give the definition of a quasi-triangular Hopf algebra and explain how the tangles map to such an algebra via suitable assignments of algebra to the lines. However, just the construction of link invariants does not depend upon the full definition of the Hopf algebra, and so we shall begin with a quantum algebra as defined below.

A *quantum algebra* is an algebra A , defined over a commutative ring k and equipped with the following structure:

- 1) There is an invertible map $s : A \rightarrow A$ such that $s(ab) = s(b)s(a)$ for all a and b in A . s is called the *antipode* of A .
- 2) There is an invertible element ρ in $A \otimes A$ that satisfies the (algebraic) Yang-Baxter equation $\rho_{12}\rho_{13}\rho_{23} = \rho_{23}\rho_{13}\rho_{12}$ where the indices refer to the placements of the tensor factors of ρ in the three factors of $A \otimes A \otimes A$ (labelled 1, 2, 3 respectively).
- 3) The relationship between s and the inverse of ρ is given by the formulas: $\rho^{-1} = (s \otimes 1)\rho = (1 \otimes s^{-1})\rho$.

We shall write ρ symbolically as a summation of tensor products of elements of A : $\rho = \sum_e e \otimes e'$.

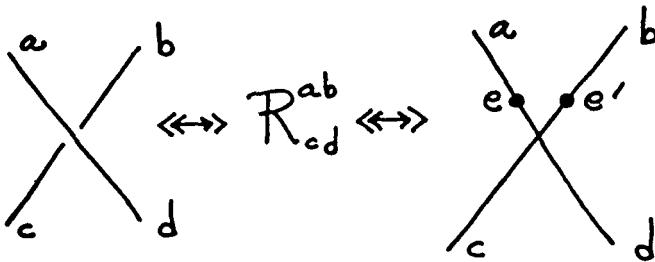
- 4) There is an invertible element G in A such that $s(G) = G^{-1}$, and $s^2(x) = G \times G^{-1}$ for all x in A .

Remark. With the summation understood, we shall often write

$$\rho = e \otimes e' .$$

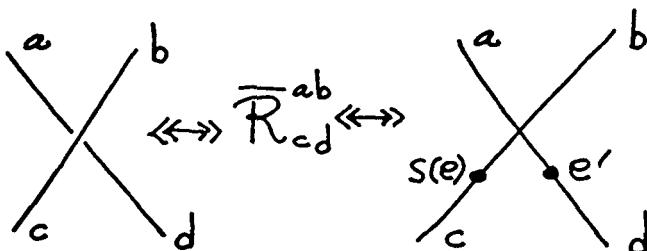
Thus condition 3), part 1, in the quantum algebra reads $s(e)f \otimes e'f' = 1$ if we understand an implicit summation on the paired elements e and e' , f and f' .

Now define the abstract tensor $R_{cd}^{ab} = [e]_d^a [e']_c^b$ (implicit summation on pairs e, e'). It is then easy to see that R_{cd}^{ab} satisfies the ATA Yang-Baxter equation by a direct calculation in the ATA algebra. This calculation corresponds directly to the diagrammatic interpretations in terms of line algebra shown below.



Note that we interpret $[e \otimes e']_{cd}^{ab} = [e]_d^a [e']_c^b$ with the permutation of indices at the bottom just so that the ATA (or knot theoretic) and the algebraic Yang-Baxter equations will correspond to each other.

Define the inverse \bar{R}_{cd}^{ab} by the equation $\bar{R}_{cd}^{ab} = [s(e)]_c^b [e']_d^a$. This corresponds both to the equation for ρ^{-1} in the quantum algebra and to the twist and slide moves in the knot theory and abstract tensor algebra, as shown below after discussing the antipode.



We instantiate the antipode s in the abstract tensor algebra by the following definition of the action of the operators M_{ab} and M^{ab} :

$$M_{ab}[x]_c^b = [s(x)]_a^b M_{bc}$$

$$[x]_b^a M^{bc} = M^{ab} [s(x)]_b^c$$

$$[G]_b^a = M^{ai} M_{bi}$$

$$[G^{-1}]_b^a = M^{ia} M_{ib}$$

The first two equations correspond to diagrams for the antipode as shown below. The equations involving the element G are statements in ATA that correspond to the fourth axiom for the quantum algebra. In other words, the antipode is accomplished on an element in the line algebra by “sliding” it around a maximum or a minimum with a counterclockwise turn.

Now we shall assume that the M 's with upper indices are the inverses of the M 's with lower indices as in the first axiom for ATA. As a result we have the topological diagram moves

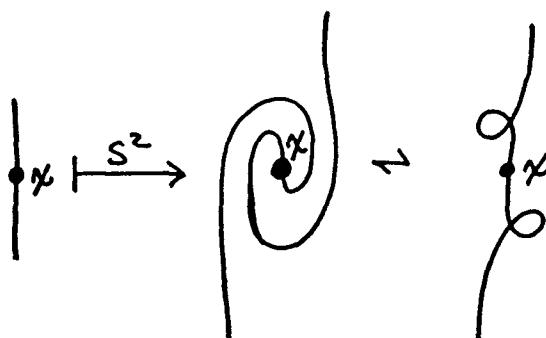
and consequently the antipode is given as left composition with M_{ab} followed by right composition with M^{cd} :

$$[s(x)]_a^d = M_{ab}[x]_a^b M^{cd}$$

Note how our definitions are consistent with the reversing property of s : $s(xy) = s(y)s(x)$.

$$\begin{aligned}
 s(xy) &= \text{Diagram of } xy = \text{Diagram of } x \cdot s(y) = \text{Diagram of } s(x) \cdot s(y) = \text{Diagram of } s(x) \\
 &= s(y)s(x).
 \end{aligned}$$

Finally, note that the square of the antipode is given diagrammatically as shown below:



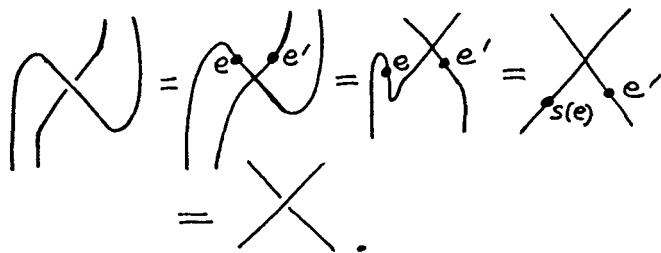
Thus we see that the immersed loops

$$\text{Diagram of } G \iff \text{Diagram of } G^{-1}$$

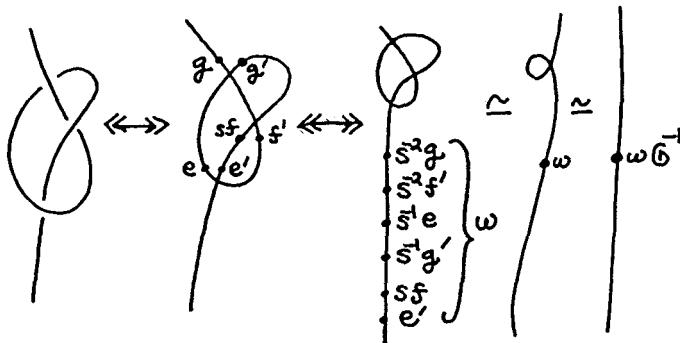
represent the elements G and G^{-1} .

With this representation, it is easy to see that all the axioms of that abstract tensor algebra are satisfied. The only point that is not yet checked is Axiom

V , and this is illustrated below.



Hence, if we decorate a link diagram according to these prescription, then the resulting algebraic elements are preserved under regular isotopy of diagrams. In fact, with the associations of algebra to diagrams as described above, we can actually associate a specific element of the algebra A to each single component tangle. The method of association is as follows: First decorate the tangle with algebra elements at the crossings. Then slide all the algebra to the bottom of the tangle by using the antipode rule for moving algebra elements around the maxima and minima of the diagram. The result is a (sum of) algebra products indicated on a vertical part of the tangle that is followed (above) by an immersion of the tangle. The immersion can be reduced by regular homotopy to a power of G : the exponent is the Whitney degree of the curve obtained by starting at the bottom of the tangle and proceeding to the top. Thus the original tangle becomes an element wG^d where w is this sum of products and d is the given Whitney degree. This process is shown below for the case of a tangle with a trefoil knot in it.



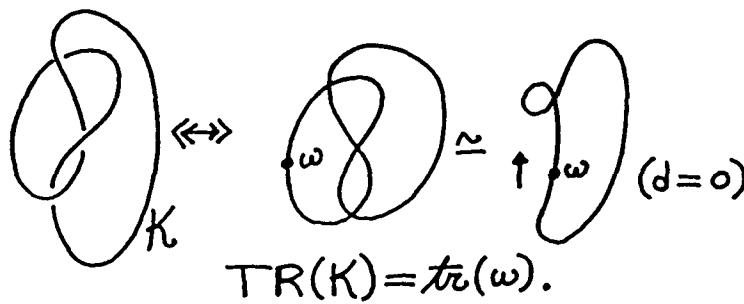
Remark. One may, in principle, extract invariants of knots and links by using a quantum algebra. In order to extract invariants we need a function that assigns scalars in the ring k to elements of the quantum algebra A . Such a function $\text{tr}: A \rightarrow k$ is said to be a *trace* if it satisfies the following properties:

- 1) $\text{tr}(xy) = \text{tr}(yx)$ for all x and y in A .
- 2) $\text{tr}(s(x)) = \text{tr}(x)$ for all x in A . (s is the antipode in A)

For a diagram K consisting of a single component arranged with respect to the vertical, we define $\text{TR}(K)$ as follows:

- 1) Decorate K with quantum algebra as described above.
- 2) Choose a vertical segment of the diagram and slide all the algebra into this vertical segment, obtaining a sum of products w . Let d be the Whitney degree of the part the immersion obtained by traversing the projected diagram from the given segment, choosing the orientation corresponding to going upward at the segment.
- 3) Define $\text{TR}(K)$ by the formula $\text{TR}(K) = \text{tr}(wG^d)$ where tr is a trace function (as defined above) on the algebra A , w is the sum of products obtained in 2), and d is the Whitney degree obtained in 2).

Example. In the case of the trefoil knot, we have



Theorem 5.1. The value of $\text{TR}(K)$ is independent of the choice of vertical segment along the diagram for K at which the algebra has been concentrated. $\text{TR}(K)$ is an invariant of regular isotopy of K .

Proof. It suffices to see that the evaluation of $\text{TR}(K)$ is independent of moving the element w around a maximum or a minimum of the diagram. (I refer to w as an element of the algebra A .) Suppose that in its given location w is coupled with a Whitney degree d . If we move w to the left

across a maximum, then w is replaced by $s(w)$, and d is replaced by $-d$. Thus we must show that $\text{tr}(wG^d) = \text{tr}(s(w)G^{-d})$. Here is the proof of this formula:

$$\text{tr}(wG^d) = \text{tr}(G^d w) = \text{tr}(s(G^d w)) = \text{tr}(s(w)s(G^d)) = \text{tr}(s(w)G^{-d}).$$

This completes the proof for the case of sliding to the left over a maximum. The other cases follow in the same manner. Regular isotopy invariance follows from the discussion preceding the Theorem.//

This description for obtaining link invariants applies directly to some cases of quantum algebras that are not Hopf algebras (e.g. to certain finite dimensional representations of Hopf algebras such as the example given at the end of [K2], which for convenience we repeat below).

Example. Here is a description of a quantum algebra that gives the original Jones polynomial: Let e^1, e^2, e^3 be the following 2×2 matrices.

$$e^1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$e^2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

$$e^3 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

$$e_1 = \begin{pmatrix} t^{-1} & 0 \\ 0 & t \end{pmatrix}$$

$$e_2 = \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix}$$

$$e_3 = \begin{pmatrix} 0 & 0 \\ t(t^{-2} - t^2) & 0 \end{pmatrix}$$

$$R = e_1 \otimes e^1 + e_2 \otimes e^2 + e_3 \otimes e^3$$

$$M = \begin{pmatrix} 0 & it \\ -it^{-1} & 0 \end{pmatrix}$$

$$W = \begin{pmatrix} 0 & it \\ -it^{-1} & 0 \end{pmatrix}$$

$$G = M^T M = \begin{pmatrix} -t^{-2} & 0 \\ 0 & -t^2 \end{pmatrix}$$

This algebra gives rise directly to the bracket polynomial [K2], [K5], for $A = t$, hence it gives the original Jones polynomial after normalization.//

We will continue the discussion by adding extra structure to the quantum algebra so that it is a Hopf algebra, but first it is worth pointing out that the framework for knot invariants that we have constructed up to this point gives a universal knot invariant for unoriented knots embedded in oriented 3-space.

Theorem 5.2. Let UA be a universal quantum algebra (in the sense that the only relations in it are those that are consequences of axioms for a quantum algebra). Let Utr denote a universal trace on UA .

(That is, let Q denote the quotient of the set UA by the equivalence relation generated by $ab \leftrightarrow ba$ and $a \leftrightarrow s(a)$ for a and b in UA , and let $\text{Utr}(x)$ denote the equivalence class of x in Q . Note that this universal trace does not take values in the ring k . Any trace taking values in k will factor through the universal trace by definition of the trace.)

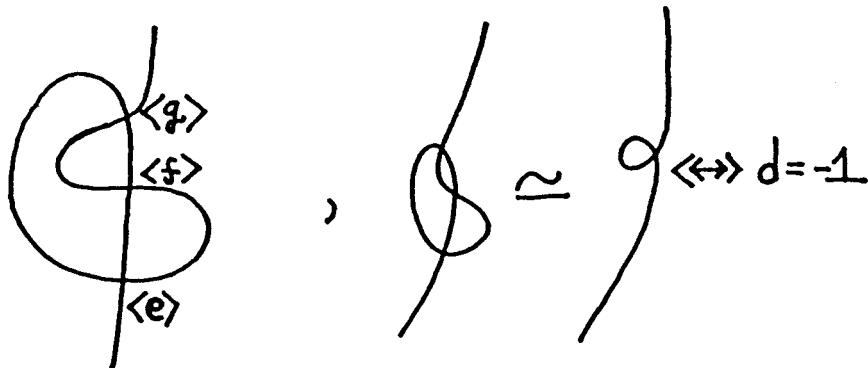
Let K be an unoriented link diagram, and let $\text{UTR}(K)$ be defined by the formula $\text{UTR}(K) = \text{Utr}(wG^d)$ where the element w and the degree d are obtained from K as described in the previous theorem. Then two prime knot diagram K and K' are regularly isotopic if and only if $\text{UTR}(K) = \text{UTR}(K')$.

Proof. The formula $\text{UTR}(K) = \text{Utr}(wG^d)$ contains, by reading it, a Gauss code for a projected link diagram for K . It also contains the Whitney degree of the diagram. These two facts are sufficient to reconstruct the projected diagram including the sense of closure (right or left if the diagram is first drawn as a vertical 1-strand tangle in the plane). Perform this reconstruction, and configure the diagram so that it is aligned with respect to the vertical. Now read the element w , successively placing its labels on the diagram at the crossings, and applying the antipode to the labels in w if one has to traverse a maxima or minima in the diagram to reach the next crossing. The resulting labellings on the diagram then can be decoded into over an undercrossings to obtain a diagram that represents the regular isotopy type that is encoded in the universal trace. Since the equivalence relations on these universal objects correspond exactly to performing regular isotopies on the diagrams, it is sufficient to give such a reconstruction procedure. This completes the proof.//

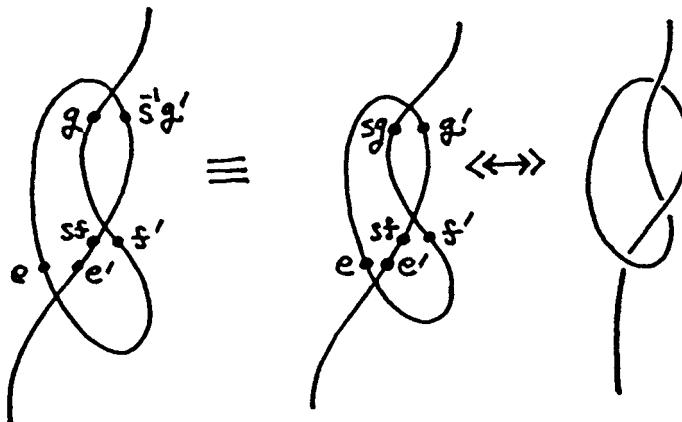
Example. The example below illustrates the reconstruction procedure in the proof above.

$$\tau = e's(f)s^{-1}(g')s^{-1}(e)s^{-2}(f')s^{-2}(g)G^{-1}$$

The Gauss code is $\langle e \rangle \langle f \rangle \langle g \rangle \langle e \rangle \langle f \rangle \langle g \rangle$ where we have indicated the crossings by symbols of the form $\langle x \rangle$. Note that the Whitney degree from bottom to top is $d = -1$.



Now reconfigure with respect to the vertical, add the line algebra and write the corresponding knot diagram.



VI. From Quantum Algebra to Quantum Groups

We now add more structure and show how the quantum algebra specializes to the concept of quantum group (Hopf algebra). The most relevant

Hopf algebra for our purposes is a ribbon Hopf algebra [RT1]. This is a special sort of quasi-triangular Hopf algebra [DR], and we shall define it first. It turns out that ribbons Hopf algebras are intimately connected with regular isotopy classes of link diagrams, as will become apparent from the discussion to follow.

Recall that a *Hopf algebra* A [SW] is a bialgebra over a commutative ring k that is associative, coassociative and equipped with a counit, a unit and an antipode. (k is usually taken to be a field, but we may want to tensor the algebra with a ring on which the field acts. Hence it is useful to designate such a ring from the beginning.)

In order to be an algebra, A needs a multiplication $m : A \otimes A \dashrightarrow A$. It is assumed that M satisfies the associative law: $m(m \otimes 1) = m(1 \otimes m)$.

In order to be a bialgebra, an algebra needs a coproduct $\Delta : A \dashrightarrow A \otimes A$. The coproduct is a map of algebras, and is regarded as the dual of a multiplicative structure. Coassociativity means that $(\Delta \otimes 1)\Delta = (1 \otimes \Delta)\Delta$ holds as compositions of maps.

The unit is a mapping from k to A taking 1 in k to 1 in A , and thereby defining an action of k on A . It will be convenient to just identify the units in k and in A , and to ignore the name of the map that gives the unit.

The counit is an algebra mapping from A to k denoted by $E : A \dashrightarrow k$. The following formulas for the counit dualize the structure inherent in the unit: $(E \otimes 1)\Delta = 1 = (1 \otimes E)\Delta$. Here 1 denotes the identity mapping of A to itself.

It is convenient to write formally $\Delta(x) = \sum x_{(1)} \otimes x_{(2)}$ to indicate the decomposition of the coproduct of x into a sum of first and second factors in the two-fold tensor product of A with itself. We shall adopt the summation convention that $\sum x_{(1)} \otimes x_{(2)}$ can be abbreviated to just $x_{(1)} \times x_{(2)}$. Thus we shall write $\Delta(x) = x_{(1)} \otimes x_{(2)}$.

The antipode is a mapping $s : A \dashrightarrow A$ satisfying the equations $m(1 \otimes s)\Delta(x) = E(x)1$, and $m(s \otimes 1)\Delta(x) = E(x)1$ where 1 denotes the unit of k as identified with the unit of A in these equations. It is a consequence of this definition that $s(xy) = s(yx)$ for all x and y in A .

A *quasitriangular Hopf algebra* A [DR] is a Hopf algebra with an element ρ in $A \otimes A$ satisfying the following equations:

1) $\rho\Delta = \Delta'\rho$ where Δ' is the composition of Δ with the map on $A \otimes A$ that switches the two factors.

2) $\rho_{12}\rho_{13} = (1 \otimes \Delta)\rho$, $\rho_{13}\rho_{23} = (\Delta \otimes 1)\rho$.

These conditions imply that $\rho^{-1} = (1 \otimes s^{-1})\rho = (s \otimes 1)\rho$.

It follows easily from the axioms of the quasitriangular Hopf algebra that ρ satisfies the Yang-Baxter Equation

$$\rho_{12}\rho_{13}\rho_{23} = \rho_{23}\rho_{13}\rho_{12} .$$

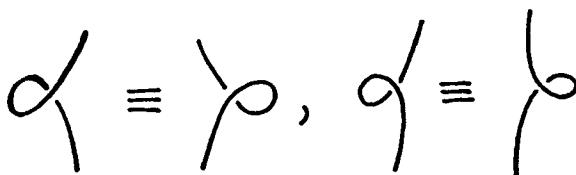
A less obvious fact about quasitriangular Hopf algebras is that there exists an element u such that u is invertible and $s^2(x) = uxu^{-1}$ for all x in A . In fact, we may take $u = \Sigma s(e')e$ where $\rho = \Sigma e \otimes e'$.

An element G in a Hopf algebra is said to be *grouplike* if $\Delta(G) = G \otimes G$ and $E(G) = 1$ (from which it follows that G is invertible and $s(G) = G^{-1}$). A quasitriangular Hopf algebra is said to be a *ribbon Hopf algebra* (See [RT1], [KR1].) if there exists a special grouplike element G such that (with u as in the previous paragraph) $v = G^{-1}u$ is in the center of A and $s(u) = G^{-1}uG^{-1}$.

Note that the condition of being ribbon implies that $s^2(x) = GxG^{-1}$ for all x in A (since $vx = xv$ implies that $uxu^{-1} = GxG^{-1}$ for all x). Note also that $s(v) = v$ (Proof: $s(v) = s(G^{-1}u) = s(u)S(G^{-1}) = G^{-1}uG^{-1}G = G^{-1}u = v$.)

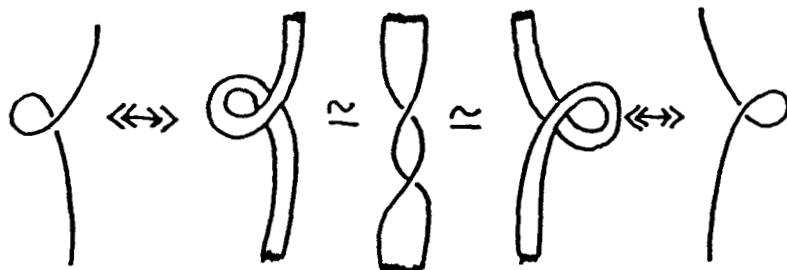
It is with these conditions about the square of the antipode that we make our first return to the link diagrams. Let A be a ribbon Hopf algebra, and regard it as a quantum algebra with the conventions in regard to link diagrams of Sec. 5. In particular, let the special element G in A correspond to M_b, M^{ai} as in Sec. 5. Thus, diagrammatically and in the abstract tensor algebra we have that $s^2(x) = GxG^{-1} = uxu^{-1}$.

As it happens, the notion of ribbon Hopf algebra corresponds to *framed equivalence of diagrams*. Framed equivalence of diagrams is regular isotopy of diagrams modulo the equivalence relation generated by the equivalence indicated below:

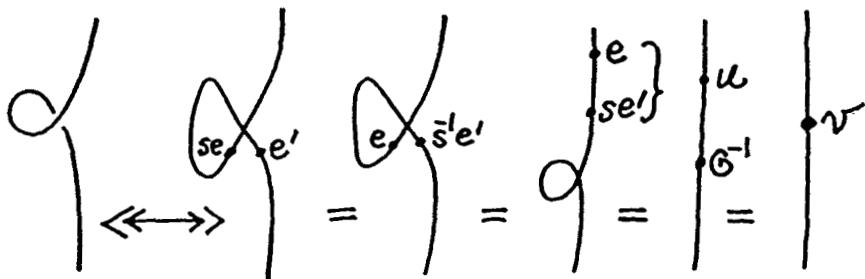


This equivalence of curls puts the framed equivalence of diagrams in one to one correspondence with equivalence classes of framed links [K] (a framed link is the same as an embedding of a band taken up to ambient isotopy)

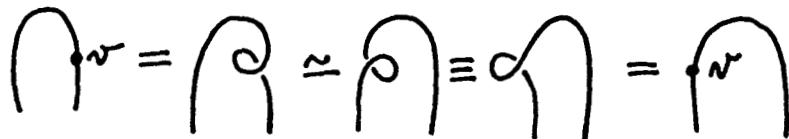
with the diagram representing the link with “blackboard framing” inherited from the plane. Thus we have the correspondences



Now note the algebra element that corresponds to the basic curl:



Since this element is exactly equal to $v = G^{-1}u$ with u the element described above for Ribbon Hopf algebras, we see that the requirement the v belong to the center of A corresponds to the desirable geometrical/topological fact that one can regard such a curl as only an indication of the framing, and hence it can be moved anywhere in the diagram. The isotopy shown below indicates that $s(v) = v$, and this is indeed the case in a ribbon Hopf algebra.

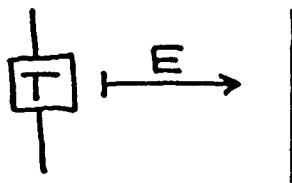


Remark. It is interesting to see how the line algebra dovetails with properties of the abstract algebra. For example, here is a diagrammatic proof of the well-known fact about ribbon elements $v^2 = s(u)u$. This shows that the element $s(u)u$ is in the center of A when A is a ribbon Hopf algebra. ($s(u)u$ is sometimes called the Casimir element of the Hopf algebra.)

Proof.

$$\begin{aligned}
 v &= \left| \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right| = \left| \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right|^u \Rightarrow v = \left| \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right| = \left| \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right|^u \\
 s(u)u &= \left| \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right|^u = \left| \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right|_u \simeq \left| \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right|_u = \left| \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right| = \left| \begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right| = v^2 \\
 \Rightarrow s(u)u &= v^2
 \end{aligned}
 \quad //.$$

We now take up the matter of the counit and the coproduct in the Hopf algebra in relation to the diagrams. First the counit: Let E be defined on single strand tangles by taking the given tangle to a bare straight line tangle:



This corresponds to the fact that the algebra elements corresponding to such tangles satisfy $E(a) = 1$. For single input, single output tangles with more than one component, each closed loop resolves into a sum of traces. On such tangles, E is defined to be identical to the algebraic result of applying E to the algebra elements corresponding to the tangles. For example,

$$\begin{array}{c}
 \text{Diagram showing a tangle transformation:} \\
 \text{Left: A single loop.} \\
 \text{Middle: Two loops labeled } e \text{ and } e' \text{ with strands } f \text{ and } f' \text{ connecting them.} \\
 \text{Equation: } = \left(\begin{array}{c} e \quad e' \\ f \quad f' \end{array} \right) \xrightarrow{\mathbb{E}} \left(\begin{array}{c} E(f'e) \\ tr(f'e'G) \end{array} \right)
 \end{array}$$

Letting $T^{(n)}$ denote the tangles with n inputs and n outputs, we have $\Delta : T^{(1)} \rightarrow T^{(2)}$ defined by taking the (planar) parallel 2-cable of all the strands in the given element of $T^{(1)}$. Thus

$$\begin{array}{ccc}
 \text{Diagram of a complex tangle} & \xrightarrow{\Delta} & \text{Diagram of its parallel 2-cable}
 \end{array}$$

This diagrammatic coproduct is compatible with the coproduct in the Hopf algebra. The verification depends upon quasitriangularity, and we omit it here. The following two examples illustrate the correspondence.

- (i) The diagram below corresponds to the algebraic relation $\Delta(s(x)) = s \otimes s(\Delta'(x))$ where Δ' denotes the composition of Δ with the mapping of $A \otimes A$ that interchanges the two factors.

$$\begin{array}{ccc}
 \text{Diagram:} & = & \text{Diagram:} \\
 \text{Top: } x & & S(x) \\
 \downarrow \Delta & & \downarrow \Delta \\
 \text{Bottom: } x_{(1)} & = & S(x_{(1)}) \\
 & & S(x_{(2)})
 \end{array}$$

(ii) Note (for use in the next section) that there is the following diagrammatic interpretation of the formula $\Delta(u^{-1}) = \rho_{21}\rho_{12}(u^{-1} \otimes u^{-1})$.

$$\Delta(\bar{\omega}') = \text{Diagram} \simeq \text{Diagram} = \frac{\text{Diagram}}{e \cdot e'} = \frac{\text{Diagram}}{f \cdot f'} = \bar{\omega}'$$

VII. Categories

Since the description of invariants given in Secs. 4, 5 and 6 may seem a bit ad hoc, the purpose of this section is to give a quick description of our methods in categorical terms. In these terms, everything appears quite functorial, and we see that we have defined functors between categories of diagrams and categories naturally associated with Hopf algebras.

A link diagram, with or without free ends, that is arranged with respect to a vertical direction can be regarded as a morphism in a tensor category with two generating objects V and k . We think of V as the analog of a vector space or module over k . Thus $k \otimes k = k$, $k \otimes V = V \otimes k = V$, and the products $V, V \otimes V, V \otimes V \otimes V, \dots$ are all distinct. Maxima are morphisms from $V \otimes V$ to k , and denoted $\cap : V \otimes V \dashrightarrow k$. Minima are morphisms from k to $V \otimes V$, and are denoted $\cup : k \dashrightarrow V \times V$. The two types of crossing yield morphisms from $V \otimes V$ to itself, and are denoted by R and \hat{R} , respectively: $R, \hat{R} : V \otimes V \dashrightarrow V \otimes V$. See the diagrams below.

$$\begin{array}{ccc}
 \text{Diagram 1: } & \text{Diagram 2: } & \text{Diagram 3: } \\
 \text{U: } V \otimes V \rightarrow k & \text{U: } k \rightarrow V \otimes V & \text{U: } k \rightarrow V \otimes V \\
 \text{R: } V \otimes V \rightarrow k & \text{R: } V \otimes V \rightarrow k & \text{R: } V \otimes V \rightarrow k
 \end{array}$$

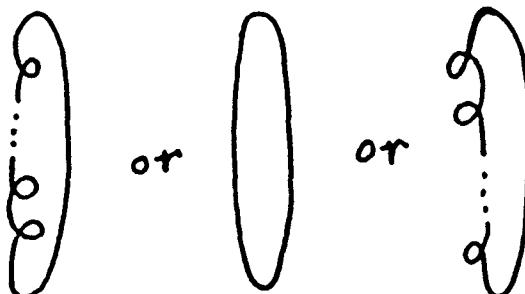
Composition of morphisms corresponds to tying input and output lines together in these diagrams. In this way, each diagram with out any tangle lines becomes a morphism from k to k . Since we call such a diagram a *closed diagram*, we shall call the corresponding morphism from k to k a *closed morphism*. The trace function described in the last section is designed to assign a specific element in a commutative ring (e.g. the complex numbers) to such k to k morphisms. But this puts us ahead of our story.

The morphisms in this regular isotopy category of link diagrams (denoted REG) are assumed to satisfy those relations that correspond to regular isotopy of links with respect to a vertical direction. These identities correspond exactly to the diagrams in Fig. 2.

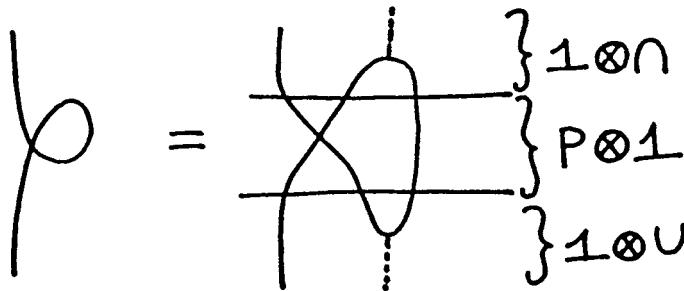
We also use the category of FLAT diagrams. In these the crossing has no distinction between over and under, but there is still a morphism: $P : V \otimes V \dashrightarrow V \otimes V$ corresponding to a crossing. The rules for the category of FLAT diagrams are exactly the same in form as the rules for the category of REG diagrams, except that there are no restrictions about the types of crossing since there is only one type of crossing.

$$\begin{array}{c} \diagup \quad \diagdown \\ \text{V} \otimes \text{V} \\ \uparrow \text{P} \\ \text{V} \otimes \text{V} \end{array}$$

We know from the Whitney-Graustein theorem that FLAT has a particularly simple structure of closed morphisms. In terms of diagrams any such morphism has the form shown below.



It is for this reason that we have singled out the morphisms G and G^{-1} in the previous section. Here $G = (1 \otimes \cap)(P \otimes 1)(1 \otimes \cup)$ as a morphism from V to V . (Note that $V = V \otimes k$.) As we have already remarked, G corresponds to a flat curl, and in the context of labelling the diagram with elements of a quantum group, G corresponds to a grouplike element.



For both categories REG and FLAT we have functors $\Delta : \text{REG} \otimes \text{REG} \dashrightarrow \text{REG}$ and $\Delta : \text{FLAT} \otimes \text{FLAT} \dashrightarrow \text{FLAT}$. On objects W , $\Delta(W) = W \otimes W$. On morphisms, $\Delta(T)$ is obtained by taking the 2-fold parallel cable of each component arc in T .

In order to describe labelled diagrams in categorical language, we begin with a Hopf algebra A , and associate to A a category $C(A)$. The category $C(A)$ has two generating objects V and k with the same properties as the V and k for REG and FLAT. However, k is taken to be identical with the ground field for the Hopf algebra. (This can be generalised if we so desire.) The morphisms of $C(A)$ are the elements of A . Each morphism takes V to V and composition of elements corresponds to multiplication in A . The labelled diagrams of simplest form



exhibit elements of A in the role of morphisms in $C(A)$.

Omitting technical details, we now define categories $C(A)^*\text{REG}$ and $C(A)^*\text{FLAT}$, by adding the morphisms \cup , \cap , R , \hat{R} , or P to $C(A)$ in the form of diagrams as we have already described them in the previous section. Note how \cup and \cap are defined to carry the structure of the antipode of A .

One then sees that we have, in the previous section, defined a functor from REG to $C(A)^*\text{FLAT}$, and then by “moving elements around the diagram”, shown how to reduce closed morphisms in $C(A)^*\text{FLAT}$ to a sum of formal traces of products of elements in A . This functor $F:\text{REG} \dashrightarrow C(A)^*\text{FLAT}$ is the source of the invariants.

VIII. Invariants of 3-Manifolds

The structure we have built so far can be used to construct invariants of 3-manifolds presented in terms of surgery on framed links. We sketch here our technique that simplifies an approach to 3-manifold invariants of Mark Hennings [H2]. The material in this section represents joint work with David Radford ([KR1],[KR2]) and will appear in expanded form elsewhere.

Recall that an element λ of the dual algebra A^* is said to be a *right integral* if $\lambda(x)1 = m(\lambda \otimes 1)(\Delta(x))$ for all x in A . It turns out that for a unimodular (see [LS],[RA3],[SW]) finite dimensional ribbon Hopf algebra A (see example at the end of the section) there is a (unique up to scalar multiplication when k is a field) right integral λ satisfying the following properties for all x and y in A :

- 1) $\lambda(xy) = \lambda(s^2(y)x)$
- 2) $\lambda(gx) = \lambda(s(x))$ where $g = G^2$, G the special grouplike element for the ribbon element $v = G^{-1}u$.

Given the existence of this λ , define a functional $\text{tr}:A \dashrightarrow k$ by the formula $\text{tr}(x) = \lambda(Gx)$.

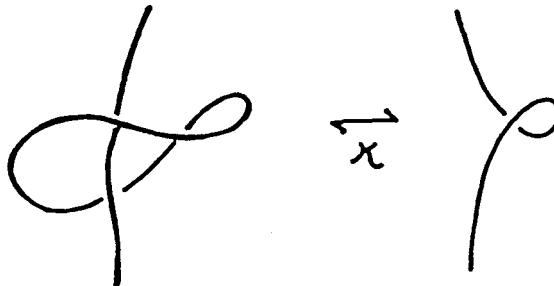
Theorem 8.1. With tr defined as above, then

- 1) $\text{tr}(xy) = \text{tr}(yx)$ for all x, y in A .
- 2) $\text{tr}(s(x)) = \text{tr}(x)$ for all x in A .
- 3) $[m(\text{tr} \otimes 1)(\Delta(u^{-1}))]u = \lambda(v^{-1})v$ where $v = G^{-1}u$ is the ribbon element.

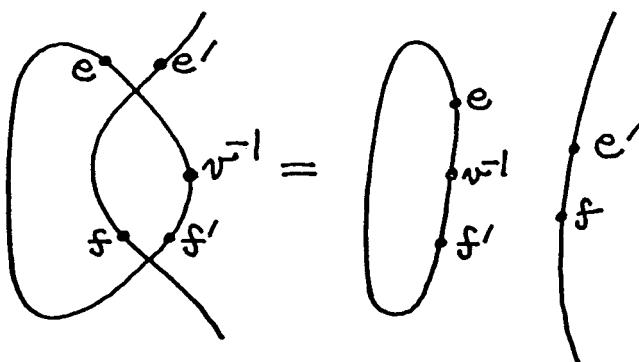
Proof. The proof is a direct consequence of the properties 1) and 2) of λ . Thus $\text{tr}(xy) = \lambda(Gxy) = \lambda(s^2(y)Gx) = \lambda(GyG^{-1}Gx) = \lambda(Gyx) = \text{tr}(yx)$, and $\text{tr}(s(x)) = \lambda(Gs(x)) = \lambda(gG^{-1}s(x)) = \lambda(s(G^{-1}s(x))) = \lambda(s^2(x)s(G^{-1})) = \lambda(s^2(x)G) = \lambda(GxG^{-1}G) = \lambda(Gx) = \text{tr}(x)$. Finally,

$[m(\text{tr} \otimes 1)(\Delta(u^{-1}))]u = G^- [m(\lambda.G \otimes G)(\Delta(u^{-1}))]u = [m(\lambda \otimes 1)(\Delta(Gu^{-1}))]$
 $G^{-1}u = \lambda(Gu^{-1})G^{-1}u = \lambda(v^{-1})v$. This completes the proof. //

The upshot of this Theorem is that for a unimodular finite dimensional Hopf algebra there is a natural trace defined via the existent right integral. Remarkably, this trace is just designed by property 3) of the Theorem to behave well with respect to the Kirby move. (The Kirby move [K] is the basic transformation on framed links that leaves the corresponding 3-manifold obtained by frame surgery unchanged. See [K][KM], [RT2].) This means that a suitably normalized version of this trace on frame links gives an invariant of 3-manifolds. To see how this works, here is a sample Kirby move



The cable going through the loop can have any number of strands. The loop has one strand and the framing as indicated. The replacement on the right hand side puts a 360 degree twist in the cable with blackboard framing as shown above. Here we calculate the case of a single strand cable:



The diagram shows that the trace contribution is (with implicit summation on the repeated primed and unprimed pairs of Yang-Baxter elements)

$$\begin{aligned} & \text{tr}(f'v^{-1}eG^{-1})fe' \\ &= \text{tr}(f'ev^{-1}G^{-1})fe' \\ &= \text{tr}(f'eu^{-1})fe' \\ &= [m(\text{tr} \otimes 1)(f'eu^{-1} \otimes fe'u^{-1})]u \\ &= [m(\text{tr} \otimes 1)(\rho_{21}\rho_{12}(u^{-1} \otimes u^{-1}))]u \\ &= [m(\text{tr} \otimes 1)(\Delta(u^{-1}))]u = \lambda(v^{-1})v. \end{aligned}$$

($\Delta(u^{-1}) = \rho_{21}\rho_{12}(u^{-1} \otimes u^{-1})$ as demonstrated at the end of Sec. 6.)

It follows from this calculation that the evaluation of the lefthand picture is the Kirby move is $\lambda(v^{-1})$ times the evaluation of the right-hand picture. For an n -strand cable we get the same result. Thus a proper normalization of $\text{TR}(K)$ gives an invariant of the 3-manifold obtained by framed surgery on K . More precisely, (assuming that $\lambda(v)$ and $\lambda(v^{-1})$ are non-zero) let

$$\text{INV}(K) = \{[\lambda(v)\lambda(v^{-1})]^{-c(K)/2}[\lambda(v)/\lambda(v^{-1})]^{-\sigma(K)/2}\}\text{TR}(K)$$

where $c(K)$ denotes the number of components of K , and $\sigma(K)$ denotes the signature of the matrix of linking numbers of the components of K (with framing numbers on the diagonal), then $\text{INV}(K)$ is an invariant of the 3-manifold obtained by doing framed surgery on K in the blackboard framing. This is our reconstruction of the form of Hennings invariant [H2].

$U_q(sl_2)'$

We realize the invariant $\text{INV}(K)$ in terms of a specific right integral for $U_q(sl_2)'$ when q is a root of unity.

Here is the definition of $A = U_q(sl_2)'$.

Let t denote a primitive n -th root of unity over a field k . Let $q = t^2$ and $m = \text{order}(t^4) = \text{order}(q^2)$. Thus we assume that $t^{4m} = 1$.

The algebra A is generated by elements a, e, f with the relations

$$a^n = 1, e^m = 0 = f^m$$

$$ae = qea$$

$$af = q^{-1}fa$$

$$[e, f] = ef - fe = (a^2 - a^{-2})/(q - q^{-1})$$

The formulas for the antipode are

$$s(e) = -q^{-1}e, s(f) = -qf, s(a) = a^{-1}.$$

The formulas for the counit are

$$E(e) = E(f) = 0, \quad E(a) = 1.$$

The coproduct is given by

$$\Delta(a) = a \otimes a$$

$$\Delta(x) = x \otimes a^{-1} + a \otimes x, \quad x = e, f.$$

The special grouplike element is $G = a^{-2}$.

Proposition 8.2. The right integral for $A = U_q(sl_2)'$ is $\lambda = \text{CHAR}\{a^{2(m-1)}e^{(m-1)}f^{(m-1)}\}$, where $\text{CHAR}\{X\}$ is the characteristic function of the algebra element X . Thus λ picks out the coefficient of $a^{2(m-1)}e^{(m-1)}f^{(m-1)}$ in the algebra element to which it is applied.

Proof. This can be verified by a direct calculation. //

The Yang-Baxter element in A (See [KM] and [RA2].) has the formula

$$\rho = \sum_{i,j=0}^{n-1} \sum_{k=0}^{m-1} \{(t^{-jk+i(j-k)-k}(q - q^{-1})^k)/n[k]_q!\} f^k a^i \otimes e^k a^{-j}.$$

A has linear basis $\{a^i e^j f^k \mid 0 \leq i < n, 0 \leq j, k < m\}$.

The algebra $A = \cup_q(sl(2))'$ is a finite dimensional ribbon Hopf algebra. It is our primary object of study for elucidating the invariant $\text{INV}(K)$.

IX. Epilogue

It has been the purpose of this paper to elucidate connections between the diagrammatic theory of knots and links and the use of Hopf algebras to produce link invariants. In particular, we have shown how the classical Gauss code for a knot or link is intimately related both to abstract tensor algebra and to this approach via quantum groups. The code figures crucially in understanding the universality of these forms of invariant. We then continue this view point of a format of producing invariants directly in terms of the Hopf algebra (not necessarily depending upon a specific representation). Here the Gauss code becomes augmented into a (state) sum of words in the Hopf algebra, ready for the application of a trace. We have concluded the paper with a description of joint work of the author and David Radford: an application of these methods to a construction of Hennings' invariants of 3-manifolds, via finite dimensional Hopf algebras.

References

- [D] M. Dehn, *Über Kombinatorische Topologie*, Acta Math. **67** (1936), 1213–168.
- [DR] V. G. Drinfel'd, *Quantum groups*, Proceedings of the International Congress of Mathematicians, Berkeley, California, USA (1987), 798–820.
- [G] C. F. Gauss., Werke, Teubner, Leipzig 1900. pp. 272, 282–286.
- [H1] M. A. Hennings, *Hopf algebras and regular isotopy invariants for link diagrams*, Math. Proc. Camb. Phil. Soc. **109** (1991), 59–77.
- [H2] M. A. Hennings, *Invariants of links and 3-manifolds obtained from Hopf algebras*, preprint (1989).
- [K1] L. H. Kauffman, *Formal Knot Theory*, Princeton Univ. Press, Lecture Notes Series **30** (1983).
- [K2] L. H. Kauffman, *From knots to quantum groups and back*, Proceedings of the Argonne Workshop on Quantum Groups, edited by T. Curtright, D. Fairlie and C. Zachos, World Scientific (1990), pp. 1–32.
- [K3] L. H. Kauffman, *Knots abstract tensors and the Yang-Baxter equation*, Knots Topology and Quantum Field Theories - Proceedings of the Johns Hopkins Workshop on Current Problems in Particle Theory-13, Florence (1989), edited by L. Lussana, World Scientific (1989), pp. 179–334.
- [K4] L .H. Kauffman, *On Knots*, Annals Study **115d**, Princeton Univ. Press (1987).
- [K5] L. H. Kauffman, *State models and the Jones polynomial*, Topology **26** (1987), 395–407.
- [K6] L. H. Kauffman, *Knots and Physics*, World Scientific (1991).
- [K7] L. H. Kauffman, *Gauss Codes and Quantum Groups*, Quantum Field Theory, Statistical Mechanics, Quantum Groups and Topology, edited by T. Curtright, L. Mezincescu and R. Nepomechie, World Scientific (1992), pp. 123–154.
- [KR1] L. H. Kauffman, and D. E. Radford, *A necessary and sufficient condition for a finite-dimensional Drinfel'd double to be a ribbon Hopf algebra*, to appear in J. of Algebra.
- [KR2] L. H. Kauffman and D. E. Radford, *Hopf algebras and invariants of 3-Manifolds*, (Announcement 1992, paper to appear).
- [K] R. Kirby, *A calculus for framed links in S^3* , Invent. Math. **45** (1978), 35–56.
- [KM] R. Kirby and P. Melvin, *The 3-manifold invariants of Witten and Reshetikhin-Turaev for $sl(2, C)$* , Invent. Math. **105**, 473–545.
- [LS] R. G. Larson, and M. E. Sweedler, *A associative orthogonal bilinear form for Hopf algebras*, Amer. J. Math. **91** (1969), 75–94.
- [L] R. A. Lawrence, *A universal link invariant*, The Interface of Mathematics and Physics, edited by D. G. Quillen, G. B. Segal and S.T. Tsou, Oxford Univ. Press (1990), pp. 151–160.
- [LC] H. C. Lee, *Tangles, links and twisted quantum groups* (preprint 1989).

- [P] R. Penrose, *Applications of negative dimensional tensors*, Combinatorial Mathematics and its Applications. edited by D.J.A. Welsh, Academic Press (1971).
- [RA1] D. E. Radford, *On the antipode of a quasitriangular Hopf algebra* (preprint 1991).
- [RA2] D. E. Radford, *Generalized double crossproducts associated with quantized enveloping algebras*, preprint.
- [RA3] D. E. Radford, *The trace function and Hopf algebras*, preprint.
- [R] R. C. Read, and P. Rosenstiehl, *On the Gauss Crossing Problem*, Colloquia Mathematica Societatis Janos Bolyai 18, Combinatorics, Keszthely (Hungary) (1976), pp. 843–876.
- [RE] N. Y. Reshetikhin, *Quantized universal enveloping algebras, the Yang-Baxter equation and invariants of links I and II*, LOMI reprints E-4-87 and E-17-87, Steklov Institute, Leningrad USSR.
- [RT1] N. Y. Reshetikhin and V. Turaev, *Ribbon graphs and their invariants derived from quantum groups*, Comm. Math. Phys. **127** (1990), 1–96.
- [RT2] N. Yu. Reshetikhin and V. G. Turaev, *Invariants of 3-manifolds via link polynomials and quantum groups*, Invent. Math. **103** (1991), 547–597.
- [S] J. Stillwell, *Classical Topology and Combinatorial Group Theory*, Springer Verlag (1980).
- [SW] M. E. Sweedler, *Hopf Algebras*, Mathematics Lecture Notes Series, Benjamin, New York, 1969.
- [W] H. Whitney, *On regular closed curves in the plane*, Comp. Math. **4** (1937), 276–284.
- [Y] D. Yetter, *Quantum groups and representations of monoidal categories* (preprint 1990).

Spin Networks, Topology and Discrete Physics

This paper discusses combinatorial recoupling theory and generalisations of spin recoupling theory, first in relation to the vector cross product algebra and a reformulation of the Four Color Theorem, and secondly in relation to the Temperley Lieb algebra, Spin Networks, the Jones polynomial and the SU(2) 3-Manifold invariants of Witten, Reshetikhin and Turaev. We emphasize the roots of these ideas in the Penrose theory of spin networks.

I. Introduction

This paper discusses combinatorial recoupling theory, first in relation to the vector cross product algebra and a reformulation of the Four Color Theorem, and secondly in relation to the Temperley Lieb algebra, Spin Networks, the Jones polynomial and the SU(2) 3-Manifold invariants of Witten, Reshetikhin and Turaev.

Section 2 discusses a simple recoupling theory related to the vector cross product algebra that has implications for the coloring problem for plane maps. Section 3 discusses the combinatorial structure of the Temperley Lieb algebra. We investigate an algebra of capforms and boundaries (the boundary logic) that underlies the structure of the Temperley Lieb algebra. This capform algebra gives insight into the nature of the Jones-Wenzl projectors that are the basic construction for the recoupling theory for Temperley Lieb algebra discussed in Sec. 4. Section 5 discusses the relationship of this work with the spin networks of Roger Penrose. Section 6 discusses the definition of the Witten-Reshetikhin-Turaev invariant of 3-manifolds. Section 7 explains how to translate the definition in Sec. 6 into a partition function on a 2-cell complex by using a reformulation of

the Kirillov-Reshetikhin shadow world appropriate to the recoupling theory of Sec. 4. The work described in Secs. 4 and 7 is joint work of the author and S. Lins and will appear in [KL92]. Section 8 is a brief discussion of the recent invariant of 4-manifolds originated by Ooguri [OG92] and extended by Crane and Yetter [CR93].

The foundations of the recoupling theory presented here go back to the work of Roger Penrose on spin networks in the 1960's and 1970's. On the mathematical physics side this has led to a number of interrelations with the the 3-manifold invariants discussed here and theories of quantum gravity. There has not been room in this paper to go into these relationships. For the record, the reference list includes papers by Penrose and also more recent authors on this topic (Hasslacher and Perry, Crane, Williams and Archer, Ooguri, Ashtekar, Smolin, Rovelli, Gambini, Pullin). It is the author's belief that the approach to the recoupling theory discussed herein will illuminate questions about these matters of mathematical physics.

Work for this paper was partly supported by NSF Grant Number DMS 9205277 and the Program for Mathematics and Molecular Biology, University of California at Berkeley, Berkeley, California. It gives the author great pleasure to thank the Isaac Newton Institute of the University of Cambridge, Cambridge, England for its kind hospitality during the initial preparation of this work. The present paper is an expanded version of a talk given by the author at the NATO conference on Topological Quantum Field Theory held at the Newton Institute in September 1992.

II. Trees and Four Colors

It is amusing and instructive to begin with the subject of trees and the Four Color Theorem. This provides a miniature arena that illustrates many subtle issues in relation to recoupling theory.

Recall the main theorem of [Kauffm90]. There the Four Color Theorem is reformulated as a problem about the non-associativity of the vector cross product algebra in three dimensional space. Specifically, let $V = \{i, -i, j, -j, k, -k, 1, -1, 0\}$, closed under the vector cross product (Here the cross product of a and b is denoted in the usual fashion by $a \times b$) :

$$\begin{aligned} i \times j &= k , & j \times i &= -k \\ j \times k &= i , & k \times j &= -i \\ k \times i &= j , & i \times k &= -j \end{aligned}$$

$$i \times i = j \times j = k \times k = 0$$

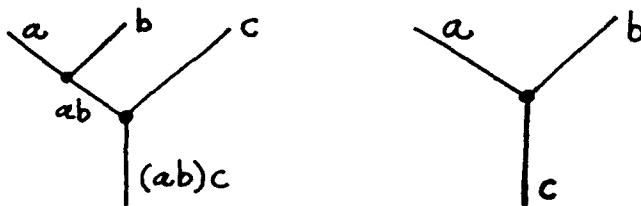
$$0 \times a = 0 \text{ for any } a$$

$$\sim 0 = 0$$

$$-1 \times a = -a \text{ for any } a$$

$$-(-a) = a \text{ for any } a .$$

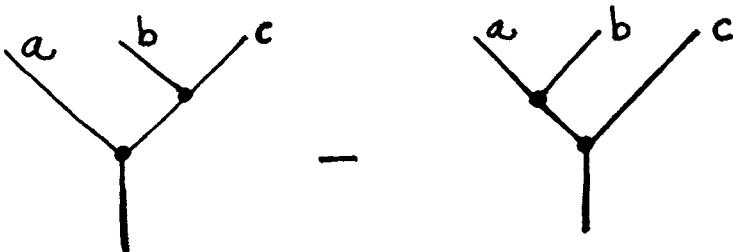
V is a multiplicatively closed subset of the usual vector cross product algebra. V is not associative, since (e.g.) $i \times (i \times j) = i \times k = -j$ while $(i \times i) \times j = 0 \times j = 0$. Given variables $a_1, a_2, a_3, \dots, a_n$, let L and R denote two parenthesizations of the product $a_1 \times a_2 \dots \times a_n$. Then we can ask for solutions to the equation $L = R$ with values for the a_i taken from the set $V^* = \{i, j, k\}$ and such that the resultant values of L and of R are non-zero. In [Kauffm90] such a solution is called a sharp solution to the equation $L = R$, and it is proved that *the existence of sharp solutions for all n and all L and R is equivalent to the Four Color Theorem*. The graph theory behind this algebraic reformulation of the Four Color Theorem comes from the fact that an associated product of a collection of variables corresponds to a rooted planar tree. This correspondence is illustrated below. Each binary product corresponds to a binary branching of the corresponding tree.



We tie the tree for L , $T(L)$, to the mirror image tree, $T(R)^*$, of the tree for R to form a planar graph $T(L)\#T(R)^*$. It is the region coloring of this graph that corresponds to a sharp solution to the equation $L = R$. (A region coloring of the graph corresponds to an edge coloring it with three colors (i, j, k) so that each vertex is incident to three distinct colors. See [Kauffm90] and [Kau92].)

There is, in this formulation of the coloring problem, an analogue to the sort of recoupling theory for networks that is common in theories of angular

momentum. In particular, we are interested in the difference between the following two branching situations:



In terms of multiplication, this asks the question about the difference generated by one local shift of associated variables:

$$a \times (b \times c) - (a \times b) \times c = ?$$

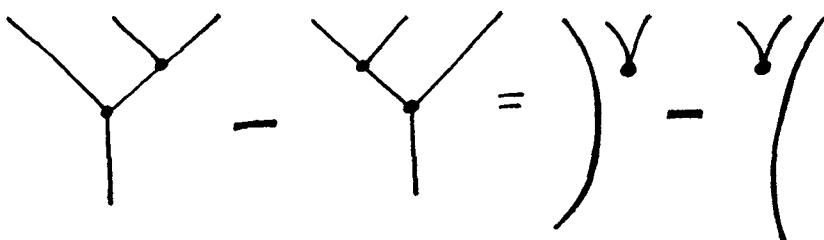
Since our algebra is the vector cross product algebra, we are well acquainted with the answer to this question. The answer is

$$a \times (b \times c) - (a \times b) \times c = a(b \cdot c) - (a \cdot b)c$$

where $a \cdot b$ denotes the dot product of vectors in three space. (Thus $i \cdot i = j \cdot j = k \cdot k = 1$, $i \cdot j = i \cdot k = j \cdot k = 0$ and $a \cdot b = b \cdot a$.)

(This formula is easily proved by using the fact that quaternion multiplication is associative, and that the formula for the product of two "pure" (i.e. of the form $ai + bj + ck$ for a, b, c real) quaternions u and v is given by the equation $uv = -u \cdot v + u \times v$.)

We can diagram this basic recoupling formula as shown below



where it is understood that a concurrence of two lines represents a dot product of the corresponding labels.

We can use this formulation to investigate the behaviour of tree evaluations under changes of parenthesization. It also provides a way to investigate purely algebraically the existence of sharp solutions, and hence the existence of map colorations.

Example. Find sharp solutions to the equation $a \times (b \times d) = (a \times b) \times d$. Since $a \times (b \times c) - (a \times b) \times c = a(b \cdot c) - (a \cdot b)c$, we see that the difference can be zero if $b \cdot c = 0$ and $a \cdot b = 0$. Thus we can try $a = i$, $b = j$ and $c = i$ or k . In this case both of these work.

Example. Find sharp solutions to the equation

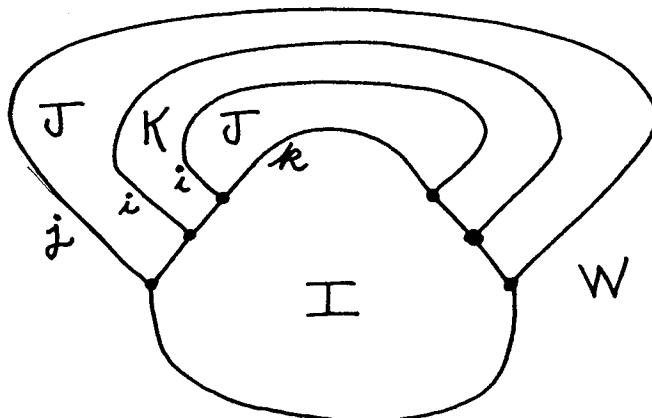
$$a \times (b \times (c \times d))) = ((a \times b) \times c) \times d .$$

Three applications of these recoupling formulas yield the equation

$$\begin{aligned} a \times (b \times (c \times d))) - ((a \times b) \times c) \times d = \\ -(a \cdot b)c \times d + (c \cdot d)a \times b - (a \cdot (b \times c))d + ((b \times c) \cdot d)a . \end{aligned}$$

From this we are led to try $a = j$, $b = i$, $c = i$, $d = k$, and this indeed works.

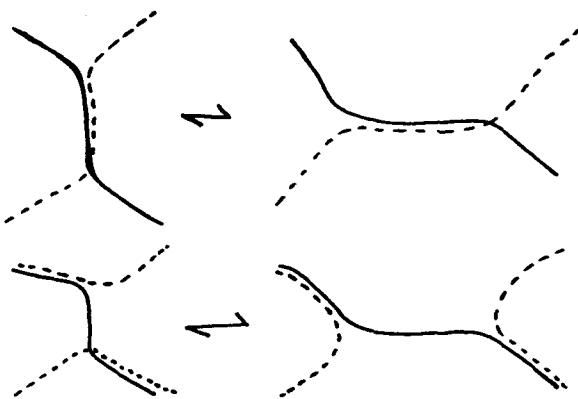
Remark. We hope to find a way to use the recoupling algebra associated with vector cross product to illuminate the original coloring problem. These are basically simple structures, but there arise gaps in language from one context to another. Thus the same problem as in the last example is very quickly and confidently solved by coloring the map shown below. This map represents the tied trees $T(L)\#T(R)^*$, and we color the regions of the map with colors W, I, J, K . The region coloring gives rise to an edge coloring by coloring an edge by the product of the colors for its adjacent regions with $IJ = k$, $JK = i$, $KI = j$, $WA = AW = a$ for all A , $AA = w$ for all A . (We use lower case letters to designate the colors of the edges.) Note that if the map is colored so that adjacent regions receive different colors, then the edges receive only the colors i, j, k .



The connection between the coloring approach and the recoupling theory of the vector cross product depends crucially upon our addition of signs to the products of colors (via the vector cross product algebra). That sharp solutions do derive from coloring including the sign also involves the quaternions. For an equation $L = R$ to have a sharp solution up to sign (as it does from bare map coloring) implies the agreement of the signs. This is because both sides of the equation (being non-zero) can be viewed as products in the quaternions. Since the quaternions are associative, this implies that the two sides are equal, hence the signs are the same [Kau92]. Both sides must be non-zero in order for this argument to work.

The Kryuchkov Conjecture: It is interesting to examine the conjecture of Kryuchkov [Kry92] in the light of these diagrammatics. He is concerned with the possibility that one could find a coloring of L and of R (i.e. a choice of values i, j, k for the variables) forming an amicable solution. By an amicable solution I mean a choice of values so that that it is possible to go from L to R by a series of elementary recombinations (such as $(xy)z \rightarrow x(yz)$) maintaining sharp solutions throughout the procedure. Kryuchkov conjectures that there is an amicable solution for every choice of L and R .

The notation of formations (see [Kauffm90]) allows a diagrammatic view of amicability. In this form we represent one color (red) by a solid line, one color (blue) by a dotted line, and the third color (purple) by a superposition of a dotted line and a solid line. The elementary forms of amicability are then as shown below:



Clearly, more work remains to be done in this version of the Four Color Theorem and its generalisations. We have begun this paper with a short review of this area both for its intrinsic interest, and for the sake of the possible analogies with other aspects of mathematics and mathematical physics.

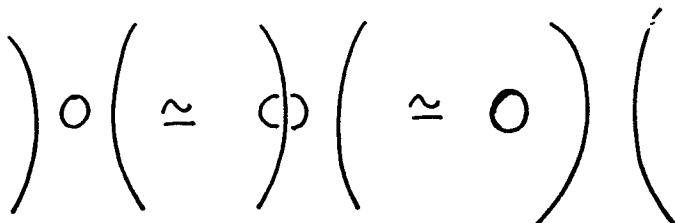
III. The Temperley Lieb Algebra

We now turn to the combinatorial underpinnings of the Temperley Lieb algebra.

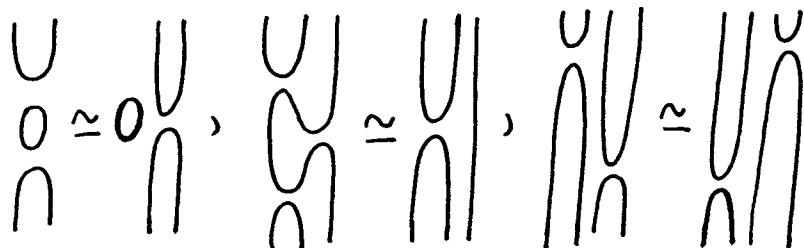
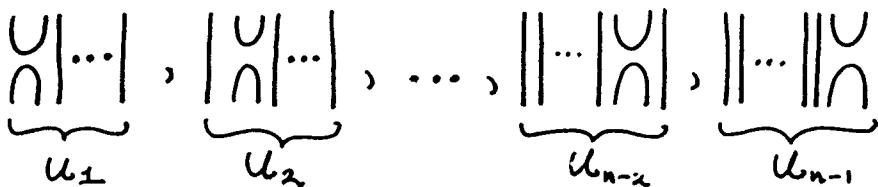
First recall the tangle-theoretic interpretation of the Temperley-Lieb Algebra [KA87]. In this interpretation, the additive generators of the algebra are flat tangles with equal numbers of inputs and outputs. We denote by T_n the (Temperley Lieb) algebra generated by flat tangles with n inputs and n outputs. A flat n -tangle is a *embedding* of disjoint curves and line segments into the plane so that the free ends of the segments are in one-to-one correspondence with the input and output lines of a rectangle in the plane that is denoted the *tangle box*. Except for these inputs and outputs, the disjoint curves and line segments are embedded in the interior of the rectangle.

Two such tangles are *equivalent* if there is a regular isotopy carrying one to the other occurring within the rectangle and keeping the endpoints fixed. Regular isotopy is generated by the Reidemeister Moves of type II and type III for link diagrams (see [KA87]). The reason we adhere to regular isotopy

at this point is that it is necessary to be able to freely move closed curves in such a tangle. Thus the two tangles illustrated below are equivalent via a regular isotopy in the tangle box that has intermediate stages that are out of the category of flat tangles.



The Temperley Lieb algebra T_n is freely additively generated by the flat n -tangles, over the ring $\mathbb{C}[A, A^{-1}]$ where \mathbb{C} denotes the complex numbers. A closed loop in a tangle is identified with a central element d in this algebra to be specified later. The familiar multiplicative generators of the Temperley Lieb algebra then appear as the following special flat tangles U_1, U_2, \dots, U_{n-1} in T_n :



These generators enjoy the relations shown above in diagrams and below in algebra.

$$U_i U_{i+1} U_i = U_i$$

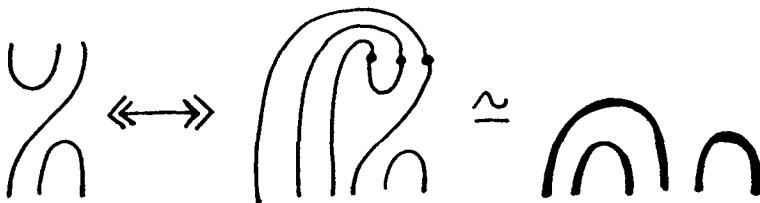
$$U_i U_{i-1} U_i = U_i$$

$$U_i U_i = dU_i$$

$$U_i U_j = U_j U_i \text{ for } |i - j| > 1$$

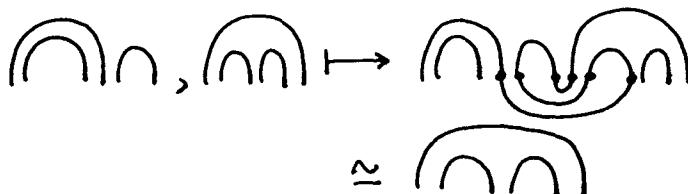
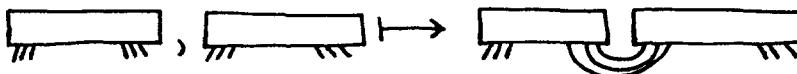
and these relations generate equivalence of flat tangles [KA90].

The purpose of this section is to point out a combinatorial algebra on parenthesis structures — *the boundary logic* — that forms a foundation for the Temperley Lieb algebra. First note that we can convert flat n -tangles to forms of parenthesization by bending the upper ends downward and to the left as illustrated below.

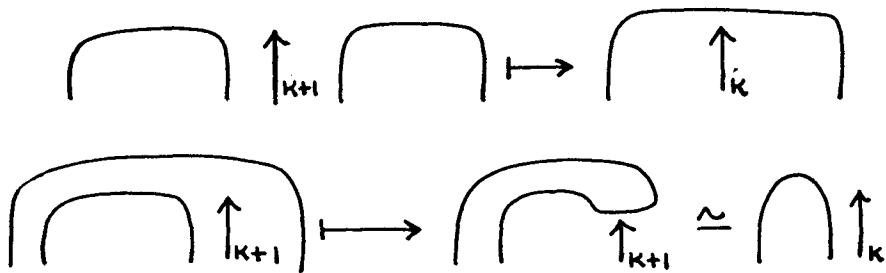


Call such a structure a capform.

In this way we convert the tangles to capforms with $2n$ strands restricted to the bottom of the form. These are capforms with n caps. The tangle multiplication then takes the form of tying the rightmost n strands of the left capform to the leftmost n strands of the right capform.



It is this operation that we shall convert into a series of more elementary operations: Regard the capform multiplication as done one pair of strands at a time. Then for a single pair of strands the pattern is to join them as shown below:



We denote this joining operation by a vertical arrow between adjacent strands. The resultant of the operation inherits an arrow in the place where the next joining can occur. An arrow with a subscript $k + 1$ has an arrow with subscript k as its immediate descendant. An arrow with the subscript 0 (zero) is equal to the empty arrow. In this language the multiplication in T_n becomes, for the corresponding capforms: $X, Y \rightarrow X \uparrow_n Y$.

It is sometimes convenient to omit the subscript on the arrow, in writing identities and also in specific calculations where the count of operations is being performed separately. We shall accordingly omit the subscripts in the text that follows.

The boundary logic of the Temperley Lieb algebra is based on the joining and breaking of adjacent boundaries in capforms. Note the following basic equations in this boundary logic:

$$\begin{aligned} A \uparrow B &= A \uparrow B \\ A(B \uparrow) C &= A B \uparrow C \\ C \uparrow (B \uparrow A) &= C \uparrow B \uparrow A \end{aligned}$$

If we wish, we can re-express this formal structure in terms of ordinary brackets rather than the arches that have been given us by the capforms. In this form the basic equations look like

$$\langle A \rangle | \langle B \rangle = \langle A | B \rangle$$

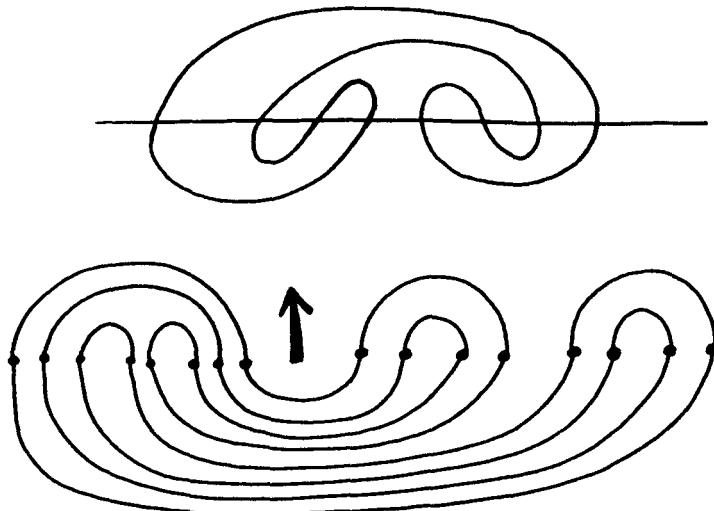
$$\langle A \langle B \rangle | \rangle = \langle A \rangle B |$$

$$\langle | \langle A \rangle B \rangle = | A \langle B \rangle$$

Here the vertical arrow indicating the join operation has been replaced by a bold face vertical line segment.

Boundary logic encapsulates the Temperley Lieb Algebra in a specific symbolic formalism that is suitable for machine computation. This formalism “knows” about the topology of Jordan curves in the plane! For example, take a Jordan curve, slice it by a line segment and regard the two halves as capforms. Then successive joining of these two halves will compute a single component.

Compare this symbolic computation with the complexity of the drawing that results from using joining arcs in the usual topological mode.



$$\begin{aligned}
 & \text{Diagram showing } \textcirclearrowleft \uparrow \textcirclearrowright \textcirclearrowleft = \textcirclearrowleft \uparrow \textcirclearrowright \textcirclearrowleft \\
 & = \textcirclearrowleft \uparrow \textcirclearrowright \textcirclearrowleft \\
 & = \textcirclearrowleft \uparrow \textcirclearrowleft \\
 & = \textcirclearrowleft \uparrow \textcirclearrowleft \\
 & = \textcirclearrowleft \uparrow \textcirclearrowleft = \textcirclearrowleft \uparrow \textcirclearrowleft \\
 & = \textcirclearrowleft \uparrow .
 \end{aligned}$$

Finally, here is the same verification of connectedness performed in the boundary logic.

$$\begin{aligned}
 & <<<><>>> | <<>><<>>= \\
 & <<<><>> | <>><<>>= \\
 & <<<><> | >><<>>= \\
 & <<<>> | ><<>>= \\
 & <><> | <<>>= \\
 & <><> | <>>= \\
 & <> | <>= \\
 & < | >
 \end{aligned}$$

We have left the vertical bar in the last entry, denoting a single loop. In this form it is interesting to note that with the extra rule

$$\| = \mid$$

the formalism contains an image of Dirac brackets:

$$P = | >< |$$

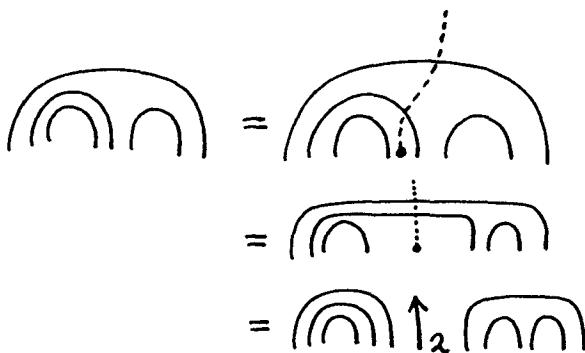
$$PP = | >< || >< | = | >< | >< | = < | > | >< | = dP .$$

Aside from the advantages of formalization, certain structural features of the Temperley Lieb algebra are easy to see from the point of view of the boundary logic. For example, consider the following natural map $T_n \times T_n \dashrightarrow T_{n+1}$ given by the formula $A, B \dashrightarrow A^*B = A \uparrow_{n-1} B$.

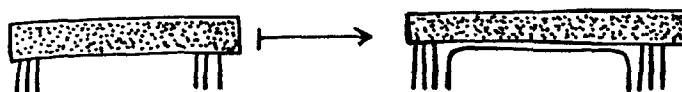
Theorem. Every element in T_{n+1} other than the identity element is of the form $A^*B = A \uparrow_{n-1} B$ for some elements A and B in T_n . This decomposition is not unique.

Proof. If C in T_{n+1} is not the identity element then it is possible to draw a curve from the midpoint of the base of C to the outer region crossing less than n arcs of C . This is easily modified in a non-unique way to cross exactly $n - 1$ arcs of C (possibly crossing some arcs twice). The curve so drawn then divides into the desired product. //

Example.



Another operation for going from T_n to T_{n+1} is $x \dashrightarrow x'$ where x' is obtained by adding an innermost cap as in



The operation $x \dashrightarrow x'$ takes the identity to the identity, and so together with $x \uparrow_{n-1} y$ encompasses all of T_{n+1} from T_n . This suggests combining these operations to produce inductive constructions in the Temperley Lieb algebra. An example that fits this idea is the well known ([JO83],[KA91],[LI91]) inductive construction of the Jones-Wenzl projectors. In tangle language these projectors are constructed by the recursion

$$\left| \begin{array}{c} \vdots \\ \vdots \\ n \\ \vdots \\ \vdots \end{array} \right| = \left| \begin{array}{c} \vdots \\ \vdots \\ n \\ \hline \vdots \end{array} \right| - (\Delta n / \Delta n_{\text{ref}}) \quad \text{Diagram: } \begin{array}{c} \text{A vertical stack of } n \text{ horizontal bars.} \\ \text{The bottom bar is highlighted with a thick black border.} \\ \text{A bracket below the stack indicates its height.} \end{array}$$

where Δ_n is a Chebyshev polynomial. These projectors are nontrivial idempotents in the Temperley Lieb algebra T_{n+1} , and they give zero when multiplied by the generators U_i for $i = 1, \dots, n+1$.

Now note the capform interpretation of the terms of this summation.

$$\begin{array}{c} \text{Diagram 1:} \\ \text{Left: } \begin{array}{c} \text{A vertical line with two horizontal branches. The top branch has a small loop labeled } n. \end{array} \\ \text{Center: } \begin{array}{c} \text{A vertical line with a curved loop labeled } n \text{ above it, and a horizontal branch below it.} \end{array} \\ \text{Right: } \begin{array}{c} \text{A vertical line with a curved loop labeled } n \text{ above it, and a horizontal branch below it.} \end{array} \end{array}$$

In the capform algebra the projectors are constructed via the recursion

$$g_{n+1} = g'_n - (\Delta_n / \Delta_{n+1}) g_n^* g_n .$$

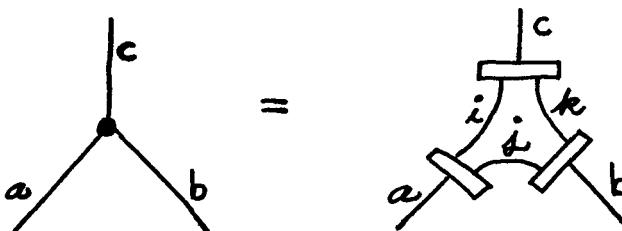
We shall return to these projectors in the next section.

A second use of the formalism is another reformulation of the Four Color Theorem. There is ([KS92], [Kau92]) a completely algebraic form of the Four Color Theory via the Temperley Lieb algebra. The boundary logic approach to the Temperley Lieb algebra provides a new way to look at the combinatorics of this version of the coloring problem. This relationship will be discussed elsewhere.

IV. Temperley Lieb Recoupling Theory

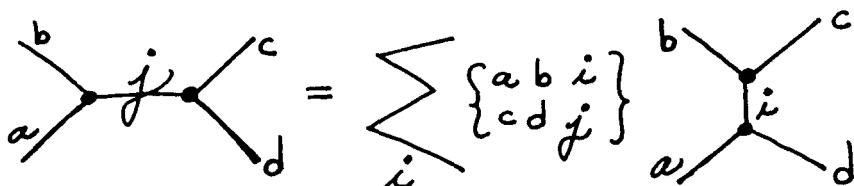
By using the Jones-Wenzl projectors, one builds a recoupling theory for the Temperley Lieb algebra that is essentially a version of the recoupling theory for the $SL(2)q$ quantum group (see [KR88],[KAU92],[KL90],[KL92]). From the vantage of this theory it is easy to construct the Witten-Reshetikhin-Turaev invariants of 3-manifolds.

We begin by recalling the basics of the recoupling theory. The 3-vertex in this theory is built from three interconnected projectors in the pattern indicated below.



The internal lines must add up correctly and this forces the sum of the external lines to be even and it also forces the sum of any two external line numbers to be greater than or equal to the third.

With these 3-vertices, we have a recoupling formula



Here the symbol

$$\left\{ \begin{matrix} a & b & i \\ c & d & j \end{matrix} \right\}$$

is a generalized $6j$ symbol.

A specific formula for the evaluation of this $6j$ symbol arises as the consequence of the following identity (see [KA91], [Kau92],[KL92]):

$$\text{Diagram: } \begin{array}{c} |a \\ \text{---} \\ b \text{---} \text{circle} \text{---} c \\ |a' \end{array} = \frac{\Theta(a,b,c)}{\Delta_a} \sum_{aa'} \begin{array}{c} |a \\ \text{---} \\ \text{box} \text{---} |a' \end{array}$$

$$\Delta_a = \begin{array}{c} |a \\ \text{---} \\ \text{box} \end{array}, \quad \Theta(a,b,c) = \begin{array}{c} |a \\ \text{---} \\ \text{circle} \text{---} b \\ |c \end{array}$$

From this identity it is easy to deduce that the $6j$ symbol is given by the network evaluation shown below:

$$\left\{ \begin{matrix} a & b & i \\ c & d & j \end{matrix} \right\} = \frac{\text{Diagram: } \begin{array}{c} |a \\ \text{---} \\ \text{triangle} \text{---} |b \\ |c \\ |d \\ |i \end{array}}{\Theta(a,d,i) \Theta(b,c,i)} \Delta_i$$

The key ingredients are the tetrahedral and theta nets. They, in turn, can be evaluated quite specifically (see [KL92],[LI91],[MV92]). There are a number of methods for obtaining these specific evaluations. For the gen-

eral case one can induct using the recursion formula for the Jones-Wenzl projectors. In the special case where $d = -2$ there is a method to obtain the results via counting loops and colorings of loops in the networks. See [PEN79], [MOU79], [KA91], [KL92]. It should be mentioned that the case $d = -2$ corresponds to the classical theory of $SU(2)$ recoupling.

V. Penrose Spin Networks

As mentioned at the end of the previous section, the Penrose spin networks are that special case of the recoupling theory where $d = -2$. In this case the projectors have a particularly simple expression as antisymmetrized sums of permutations, and there is an identity (the binor identity) that eliminates crossed lines as shown below:

$$\left(\begin{array}{c} X + U \\ +) (\end{array} \right) = \emptyset$$

The Binor Identity

The source of the binor identity in this special case is algebraic. It is a reformulation of the basic Fierz identity relating epsilons and deltas:

$$\epsilon^{ab} \epsilon_{cd} = \delta_c^a \delta_d^b - \delta_d^a \delta_c^b$$

Here the epsilon ϵ_{ab} denotes the alternating symbol where the indices a and b range over two values (say 0 and 1). Thus $\epsilon_{ab} = \epsilon^{ab} = 0$ if $a = b$, 1 if $a < b$, -1 if $a > b$. The delta is the Kronecker delta $\delta_a^b = 1$ if $a = b$, 0 if $a \neq b$.

Here we use the Einstein summation convention — *sum over repeated occurrences of upper and lower indices*.

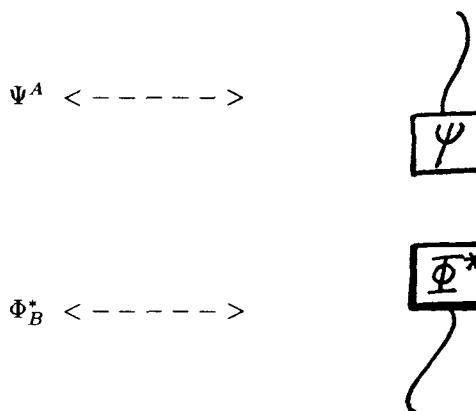
The story of how the Fierz identity becomes the binor identity, and how this translation effects a relationship among $SU(2)$, diagrams, spin and topology is the subject of this section. It is possible to make diagrammatic notations for tensor algebra. These diagrams can be allowed, in specific circumstances, to become spin networks, link diagrams or a combination of structures. In fact, in the case of spin networks alone, there are remarkable topological motivations for choosing certain diagrammatic conventions. These topological motivations provide a pathway into the beautiful combinatorics underlying the subject of spin angular momentum.

Let us begin by recalling that the groups $SU(2)$ and $SL(2)$ have the same Lie algebra structure. It is more convenient in this context to speak of $SL(2)$ because it has such a simple definition as the set of complex 2×2 matrices of determinant equal to one. If ε denotes the 2×2 matrix $\varepsilon = [\varepsilon_{ab}]$, then $SL(2)$ is the set of 2×2 complex matrices P such that $P\varepsilon P^T = \varepsilon$. This follows from the fact that for any 2×2 matrix P , $P\varepsilon P^T = \text{DET}(P)\varepsilon$ where $\text{DET}(P)$ denotes the determinant of the matrix P . Thus $SL(2)$ is the set of matrices for which the epsilon is invariant under conjugation.

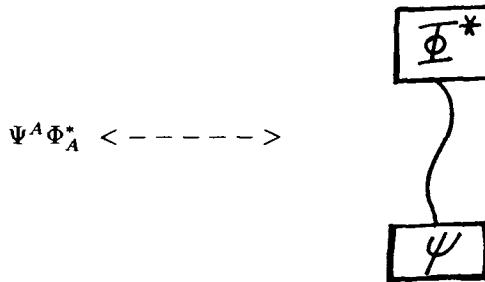
Recall also that a *spinor* is a complex vector of dimension two. That is a spinor is an element of the vector space $V = C \times C$, C the complex numbers, and V is equipped with the standard left action of the group $SL(2)$. If Ψ is a spinor and P belongs to $SL(2)$, then $P\Psi$ is the vector with coordinates $(P\Psi)_A = P_B^A \Psi^B$. Here we use the Einstein summation convention, summing (from 0 to 1) on repeated indices where one index is in the upper place, and one index is in the lower place.

With this action, there is an $SL(2)$ invariant inner product on spinors given by the formula $\Psi\Phi^* = \Psi^A \varepsilon_{AB} \Phi^B$. In other words, we define $(\Phi^*)_A = \varepsilon_{AB} \Phi^B$ and take the standard inner product of the (upper and lower indexed) vectors Ψ and Φ^* to form this inner product. It is easy to see that this inner product is invariant under the action of $SL(2)$.

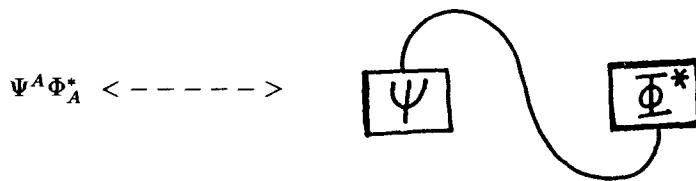
This basic algebra related to $SL(2)$ and to spinors is sufficient to use a case study in the art of diagramming tensor algebra. We let a spinor (which has one upper index or one lower index if it is a conjugate spinor) be denoted by a box with an arc emanating either from its top or its bottom. The arc will take the place of the index, and *common indices will be denoted by joined arcs*. Thus we represent Ψ^A and Φ_B^* as follows:



Consequently, we represent $\Psi\Phi^* = \Psi^A\Phi_A^*$ as the interconnection of the boxes for each term - tied by their common index lines.



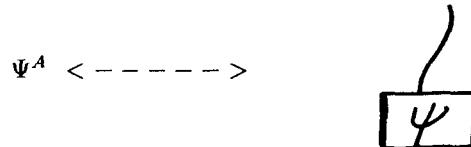
In translating into diagrammatic tensor algebra, it is convenient to have a convention for moving the boxes around so that, for example, we could also represent $\Psi^A\Phi_A^*$ by a picture with the two boxes next to one another. The connecting line must be topologically deformed as shown below:

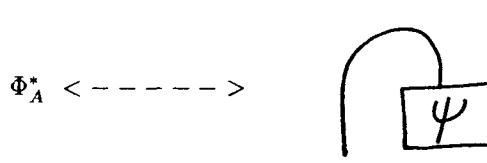


Since the lines in the tensor diagrams only serve to indicate the positions of common indices, *the expressions are invariant under topological deformations of the lines.*

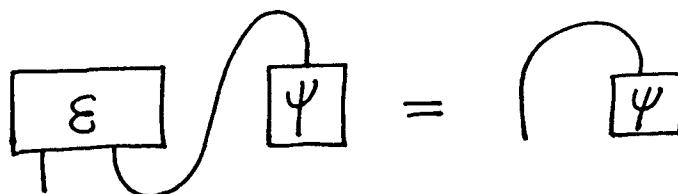
On the other hand, it sometimes happens that a very compelling convention suggests itself and appears to contradict such topological invariance. In the case of these spinors that convention is as follows:

Indicate the lowering of the index in the passage from Ψ^A to Ψ_A^ by bending the upper “antenna” downwards.*

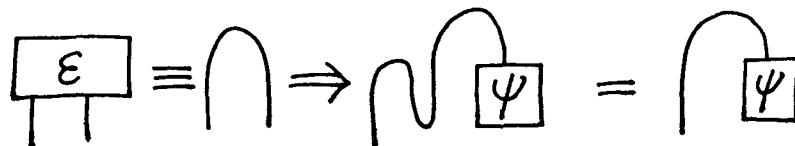




In our case we have $\Psi_A^* = \epsilon_{AB}\psi^B$. Therefore, if ϵ_{AB} is represented in diagrams as a box with two downward strokes, then we have the identity



and consequently we must identify a “cap” (a segment with vertical tangents and one maximum) as an epsilon:



However, this convention is *not* topologically invariant! The square of the epsilon matrix is minus the identity. This paradox is remedied by multiplying the epsilon by the square root of minus one. We let a cap or a cup denote $\sqrt{-1}\epsilon$.

$$\cap \underset{\text{def}}{=} \sqrt{-1} \begin{array}{|c|} \hline \epsilon \\ \hline \end{array}, \cup \underset{\text{def}}{=} \sqrt{-1} \begin{array}{c|} \hline \epsilon \\ \hline \end{array}$$

Then we have the identities

$$\text{curl} = \Big|, \quad \text{div} = \Big|, \quad \Big| \Leftrightarrow \Big|_A^B = \delta_A^B.$$

However it is still the case that a curl as shown below gives a minus sign.

$$\text{curl} = \sqrt{-1} \quad \boxed{\epsilon} = -\sqrt{-1} \quad \boxed{\epsilon}$$

Therefore associate a minus sign with each crossing.

$$\text{crossing} \stackrel{\text{def}}{=} (-1) \quad \text{crossing}, \quad \begin{array}{c} a \\ \times \\ c \end{array} \quad \begin{array}{c} b \\ \times \\ d \end{array} = \delta_d^a \delta_c^b$$

With this change of conventions, we obtain a topologically invariant calculus of modified epsilons in which the Fierz identity becomes the binor identity mentioned previously, and the value of a loop is equal to minus two (-2) rather than the customary 2 that would result from $\epsilon^{ab}\epsilon_{ab}$. Here the loop corresponds to

$$(\sqrt{-1})^2 \epsilon^{ab} \epsilon_{ab} = -2.$$

$$\text{loop} = (\sqrt{-1})^2 \epsilon^{ab} \epsilon_{ab} = -2.$$

The result of this reformulation is a topologically invariant diagrammatic calculus of tensors associated with the group $SL(2)$. We can now explain

an important special case of the Temperley-Lieb projectors in terms of this calculus.

Consider the Temperley Lieb algebra where the loop value is -2 . Form the *antisymmetrizer* F_n obtained by summing over all permutation diagrams for the permutations of $\{1, 2, \dots, n\}$ multiplied by their signs, and divide the whole sum by $n!$.

$$F_n = \frac{1}{n!} \sum_{\sigma \in S_n} \begin{array}{c} \text{---} \\ | \\ \text{---} \end{array} \begin{array}{c} \text{---} \\ | \\ \text{---} \end{array} \text{sgn}(\sigma)$$

Transform this sum into an element in the Temperley-Lieb algebra by using the binor identity at each crossing.

For example

$$\begin{aligned} F_2 &= \frac{1}{2!} [|| - X] \\ &= \frac{1}{2} [|| + \begin{smallmatrix} \cup \\ \cap \end{smallmatrix} + ||] \\ &= || + \frac{1}{2} \begin{smallmatrix} \cup \\ \cap \end{smallmatrix}. \end{aligned}$$

It is then not hard to see that $F_n F_n = F_n$ and that F_n annihilates each generator U_i of the Temperley Lieb algebra on n strands, for $i = 1, \dots, n-1$. Thus F_n is the unique projector in this case of loop value equal to -2 . These projectors serve as the basis for a recoupling theory as outlined in the last section. In this case, it is equivalent to the recoupling theory for standard angular momentum in quantum mechanics.

In the next section we shall use the bracket polynomial [KA87] to explain how the general recoupling theory is related to the topology of 3-manifolds. In that setting the binor identity becomes replaced by the bracket identity

$$\text{X} = A \text{X} + \bar{A}^l \text{D C}$$

and the loop value is $-A^2 - A^{-2}$. Note that when $A = -1$, the bracket identity reduces to the binor identity and the loop value equals -2 . At -2 there is no longer any distinction between an overcrossing and an undercrossing and the bracket calculus of knots and links degenerates to the planar binor calculus.

For the remainder of this section I wish to discuss some of the issues behind the binor calculus and spin network construction from the point of view of discrete physics. Penrose originally invented the spin networks in order to begin a program for constructing a substratum of spacetime. The substratum was to be a timeless domain of networks representing spin exchange processes. From this substrate one hopes to coax the three dimensional geometry of space and the Lorentzian geometry of spacetime (perhaps to emerge in the limit of large networks). The program was partially successful in that through the Spin Geometry Theorem [PEN79] the directions and angles of 3-space emerge in such a limit. But distances remain mysterious and it is not yet clear what is the nature of spacetime with respect to these nets. Furthermore, from the point of view of the classical spin nets, the move to the topologically invariant binor calculus is a delicate and precarious descent into combinatorics. One wants to understand why it works at all, and what the specific meaning of this combinatorics is in relation to three dimensional space.

There is a radically different path. That path begins with only the combinatorics and a hint of topology. Start with a generalization of the binor identity in the form

$$\begin{aligned} \text{X} &= A \text{X} + B \text{D C} \\ \text{O} &= \text{D} \end{aligned}$$

and ask that it be topologically invariant in the sense of regular isotopy generated by the knot diagrammatic moves II and III.

$$\text{II. } \text{D}\text{C} \rightarrow \text{DC}.$$

$$\text{III. } \text{X} \rightarrow \text{X}, \text{X} \rightarrow \text{X}.$$

We see at once that

$$\text{D}\text{C} = AB \text{DC} + [ABd + A^2 + B^2] \text{D}\text{C}$$

Proof.

$$\begin{aligned} \text{D}\text{C} &= \text{D} \begin{array}{c} B \\ \diagdown \\ A \end{array} \text{C} + \text{D} \begin{array}{c} B \\ \diagup \\ A \end{array} \text{C} \\ &+ \text{D} \begin{array}{c} A \\ \diagup \\ A \end{array} \text{C} + \text{D} \begin{array}{c} A \\ \diagdown \\ B \end{array} \text{C} // \end{aligned}$$

Hence we can achieve the invariance

$$\text{D}\text{C} = \text{DC}$$

by taking $B = A^{-1}$ and $d = -A^2 - A^{-2}$. Then, a miracle happens, and we are granted invariance under the triangle move with no extra restrictions:

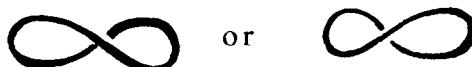
$$\begin{aligned} \cancel{\text{Diagram}} &= A \left(\text{Diagram} + \bar{A}^{-1} \right) \cancel{\text{Diagram}} \\ &= A \left(\text{Diagram} + \bar{A}^{-1} \right) \cancel{\text{Diagram}} = \cancel{\text{Diagram}}. \end{aligned}$$

The rest of the mathematical story is told in the sections of this paper that precede and follow the present section. Let us stay with the discussion of discrete physics. All of the basic combinatorics of spin nets is met and generalized by a strong demand for network topological invariance (in three dimensions). The *topology* of three dimensions can be built into the system from the beginning.

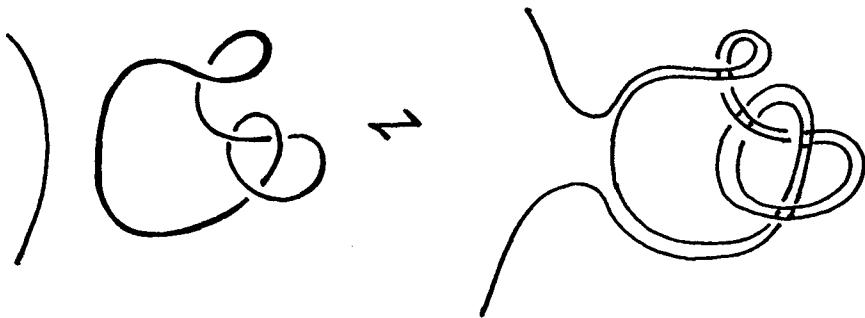
The groups $\mathrm{SL}(2)$ and $\mathrm{SU}(2)$ (and the corresponding quantum groups [KA91]) emerge not as symmetries of metric euclidean space, but as internal symmetries of the network structure of the topology. Furthermore, it is only through the well-known *interpretations* of the knot and link diagrams that the combinatorics becomes interpreted in terms of the topology of three dimensional space. The knotted spin network diagrams become webs of pattern in an abstract or formal plane where the only criterion of distinction is the fact that a simple closed curve divides the space in twain. The knot theoretic networks speak directly to the logic of this formal plane. The significant invariances of type II and type III (above) are moves that occur in the plane plus the very slightest of vertical dimensions. The vertical dimension for the formal plane is infinitesimal/imaginary. Angular momentum and the topology of knots and links are a fantasy and fugue on the theme of pattern in a formal plane. The plane sings its song of distinction, unfolding into complex topological and quantum mechanical structures.

A link diagram is a code for a specific three dimensional manifold. This is accomplished by regarding the diagram as instructions for doing surgery

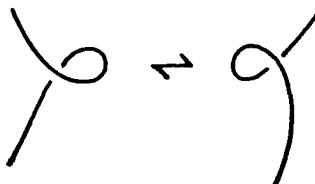
to build the manifold. Each component of the diagram is seen as an embedding in three space of a solid torus, and the curling of the diagram in the plane gives instructions for cutting out this torus and repasting a twisted version to produce the 3-manifold. (See [KA91] for a more detailed description of the process.) Extra moves on the diagrams give a set of equivalence classes that are in one-to-one correspondence with the topological types of three-manifolds. These moves consist in the addition or deletion of components with one curl as shown below



plus “handle sliding” as shown below and in the next section.



We also need the “ribbon” equivalence of curls shown below.



With these extra moves the links codify the topology of three dimensional manifolds. The recoupling theory and spin networks appear to be especially designed for handling this extra equivalence relation.

This is the real surprise. The topological generalization of the spin network theory actually handles the topology of three dimensional spaces. Each such space corresponds to a finite spin network, and there are no limiting approximations as in the original Spin Geometry Theorem. From the point of view of topology, each distinct three dimensional manifold is the direct correspondent of a finite spin network.

Where then is the geometry? Topology is neatly encoded at the level of the spin networks. Geometric structures on three manifolds, particularly metrics, are the life blood of relativity and other physics. It is interesting to speculate on how to allow the geometry to live in spin nets. One possibility is to use an appropriately fine spin network to encode the geometry. There are many spin networks corresponding to a given three manifold. We would like to have a mesh in the spin network that corresponds to the points of the three manifold. This could be accomplished by "engraving" the three manifold with many topologically extraneous handles, and then reexpressing the net as a coupling of these handles. This would make a big weaving pattern that could become the receptacle of the geometry. But really, this matter of putting in the geometry is an open problem and it is best stated as such. (But compare [ASH92].)

The moral of our story is that the spin networks most naturally articulate topology. How they implicate geometry is a quest of great worth.

VI. Knots and 3-Manifolds

This recoupling theory extends to trivalent graphs that are knotted and linked in three dimensional space. One way to delineate the connection is via the bracket model for the Jones polynomial [KA87]. In this model one obtains an invariant of regular isotopy of knots and links with the properties shown below.

$$\langle \diagup\diagdown \rangle = A \langle \diagup\diagup \rangle + A^{-1} \langle \circlearrowleft \circlearrowright \rangle$$

$$\langle \circlearrowright \circlearrowleft \rangle = -A^2 - A^{-2}$$

$$\langle \text{---} \rangle = \langle \text{---} \rangle$$

$$\langle \text{X} \rangle = \langle \text{X} \rangle$$

$$\langle \text{---} \rangle = -A^3 \langle \text{---} \rangle$$

$$\langle \text{---} \rangle = -A^{-3} \langle \text{---} \rangle$$

Braids expand under the bracket into sums of elements in the Temperley Lieb algebra. In this context the Jones-Wenzl projectors can be realized via sums of braids corresponding to all permutations on n -strands (see [KA91], [KL92]). In this context, the coefficient Δ_n is given by the formula $\Delta_{n-1} = (-1)^{n-1}(A^{2n} - A^{-2n})/(A^2 - A^{-2})$.

When $A = \exp(i\pi/2r)$ is a $4r$ -th root of unity, then the recoupling theory goes over to this context with an extra admissibility criterion for each 3-vertex with legs a, b, c . We require that $a + b + c \leq 2r - 4$. With indices ranging over the set $\{0, 1, 2, \dots, r-2\}$, everything we have said about recoupling goes over. In particular, the theta net evaluations

$$\Theta(a, b, c) = \begin{array}{c} c \\ \text{---} \\ b \\ \text{---} \\ a \end{array}$$

are non-zero in this range so that the network formulas for the recoupling coefficients still hold.

It is at the roots of unity that one can define invariants of 3-manifolds. There are many ways to make this definition. A particularly neat version using the Jones-Wenzl projectors is given by Lickorish [LIC92]. We reproduce his definition here:

Let a link component, K , labelled with ω (as in $\omega^* K$) denote the sum of i -cablings of this component over i belonging to the set $\{0, 1, 2, \dots, r-2\}$.

A projector is applied to each cabling. Thus

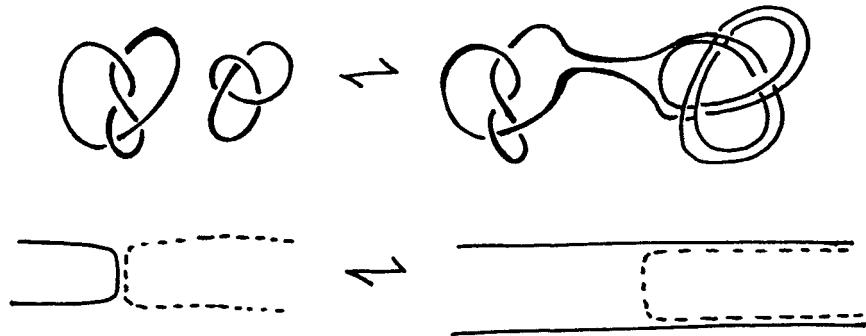
$$\omega = \sum \Delta_i$$

and

$$\omega \otimes \omega = \sum_{i,j} \Delta_i \Delta_j$$

Then $\langle \omega^* K \rangle$ denotes the sum of evaluations of the corresponding bracket polynomials. If K has more than one component, then $\omega^* K$ denotes the result of labelling each component, and taking the corresponding formal sum of products for the different cableings of individual components.

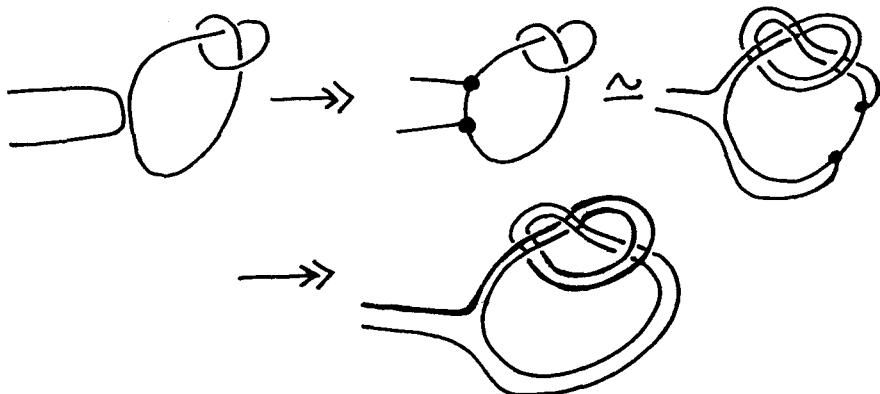
It is then quite an easy matter to prove that $\langle \omega^* K \rangle$ is invariant under handle sliding in the sense shown below.



This is the basic ingredient in producing an invariant of 3-manifolds as represented by links in the blackboard framing. (A normalization is needed for handling the fact the 3-sphere is returned after surgery on an unknot with framing plus or minus one.) A quick proof of handle sliding invariance using this definition and recoupling theory has been discovered by Justin Roberts [LIC92], and also by Oleg Viro. (An even quicker proof using less machinery is given by Lickorish in [LIC92].) Roberts proof is based on the formula (a special case of the general recoupling formula)

$$a) \quad b = \frac{\Delta_i}{\Theta(a, b, i)} \quad \text{Diagram: } \begin{array}{c} a \quad b \\ \diagdown \quad \diagup \\ i \end{array}$$

and the sequence of “events” shown below.



We turn these events into a proof of invariance under handle sliding by adding the algebra.

$$\begin{aligned}
 & \text{Diagram showing } a \text{ entering a loop } \omega \text{ and exiting} \\
 & = \sum_i \Delta_i \text{ Diagram showing } a \text{ entering a loop labeled } i \\
 & = \sum_i \Delta_i \sum_j \frac{\Delta_j}{\theta(a, i, j)} \text{ Diagram showing } a \text{ entering a loop labeled } i \\
 & = \sum_j \Delta_j \sum_i \frac{\Delta_i}{\theta(a, j, i)} \text{ Diagram showing } a \text{ entering a loop labeled } i \\
 & = \sum_j \Delta_j \text{ Diagram showing } a \text{ entering a loop labeled } j = \omega
 \end{aligned}$$

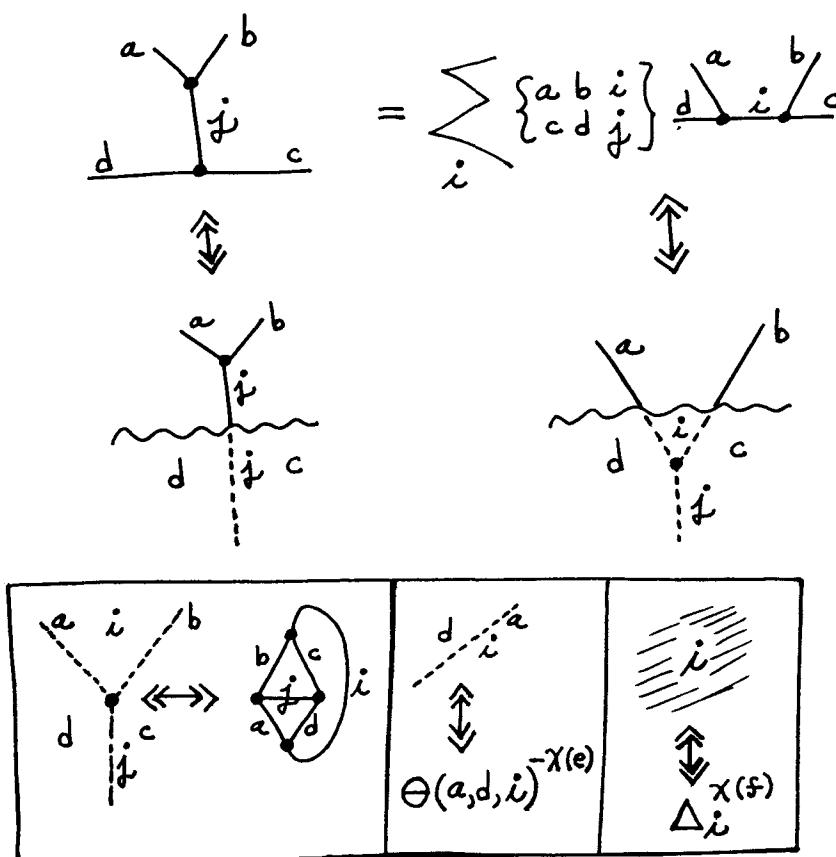
The logic of this construction of an invariant of 3-manifolds depends directly on formal properties of the recoupling theory, plus subtle properties of that theory at the roots of unity.

VII. The Shadow World

In this section we give a quick sketch of a reformulation of Kirillov-Reshetikhin Shadow World [KR88] from the point of view of the Temperley-Lieb Algebra recoupling theory. The payoff is an elegant expression for the Witten-Reshetikhin-Turaev Invariant as a partition function on a two-cell complex.

Shadow world formalism rewrites formulas in recoupling theory in terms of colorings of a two-cell complex. In the case of diagrams drawn in the

plane this means that we allow ourselves to color the regions of the plane as well as the lines of the diagram with indices from the set $\{0, 1, 2, \dots, r-2\}$ (working at $A = \exp(i\pi/2r)$, as in the last section). In this way a recoupling formula can be rewritten in terms of weights assigned to parts of the two-cell complex. The diagram below illustrates the process of translating between the daylight world (above the wavy line) and the shadow world (below the wavy line).

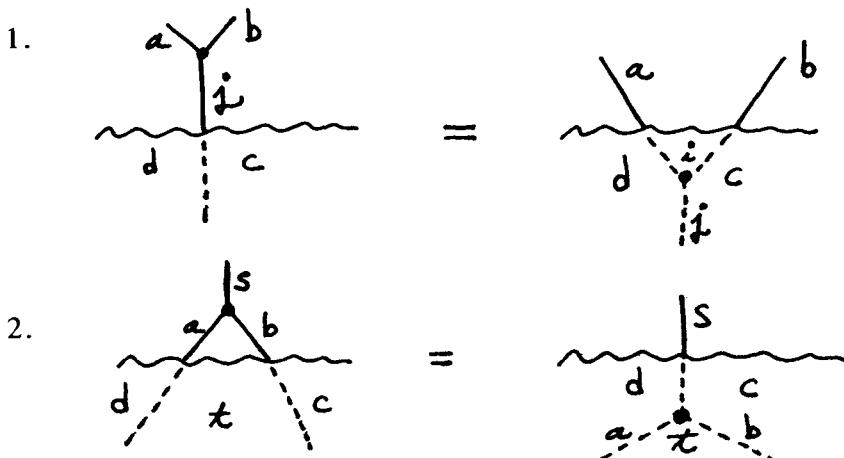


In this diagram we have indicated how if the tetrahedral evaluation is assigned to the six colors around a shadow world vertex (either 4-valent or 3-valent), the theta symbol is assigned to the three colors corresponding to an edge (the edge itself is colored as are the regions to either side of the edge) then we can take the shadow world picture as holding the information about 6-j symbols that are in the recoupling formula.

To complete this picture we assign Δ_i to a face labelled i (when this face is a disk, otherwise we assign Δ_i raised to the Euler characteristic of the face). There are phase factors corresponding to the crossings. The shadow world diagram is then interpreted as a sum of products of these weights over all colorings of its regions edges and faces.

The result is an expression for the handle sliding invariant $\langle \omega^* K \rangle$ (of Sec. 5) as a partition function on a two-cell complex. The extra Δ_i 's coming from the assignments of ω to the components of the link can be indexed by attaching a 2-cell to each component. The result is a 2-cell complex that is a "shadow" of the 4-dimensional handlebody whose boundary is the 3-manifold constructed by surgery on the link.

We shall spend the rest of this section giving more of the details of this shadow world translation. A complete description can be found in [KL93]. First, here are some elementary identities. The first is the one we have already mentioned above.



3.

$$\text{Diagram 3: } \text{wavy line } b \text{ under curve } a = \text{wavy line } b \text{ with a crossing over } a, \text{ with } i \text{ indicating orientation}$$

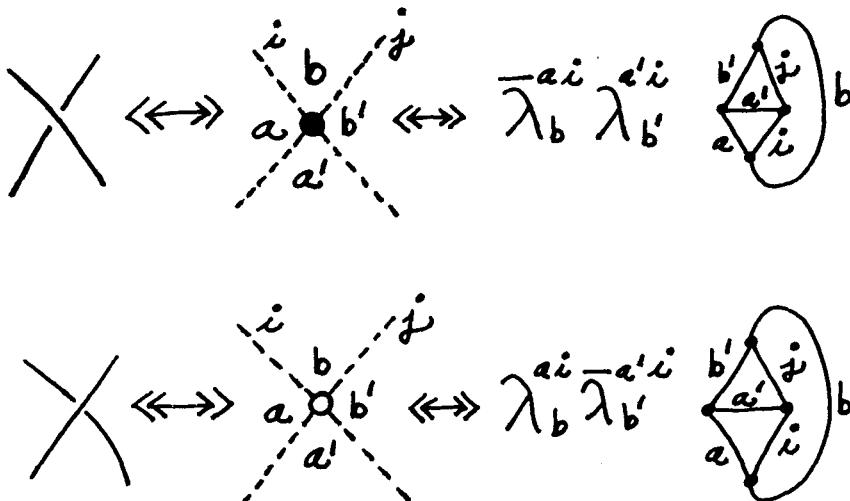
4.

$$\text{Diagram 4: } \text{wavy line } b \text{ under curve } a = \text{wavy line } b \text{ with a crossing over } a, \text{ with points } c, d \text{ mapped to } a, b \text{ and a factor } \delta_{cd}$$

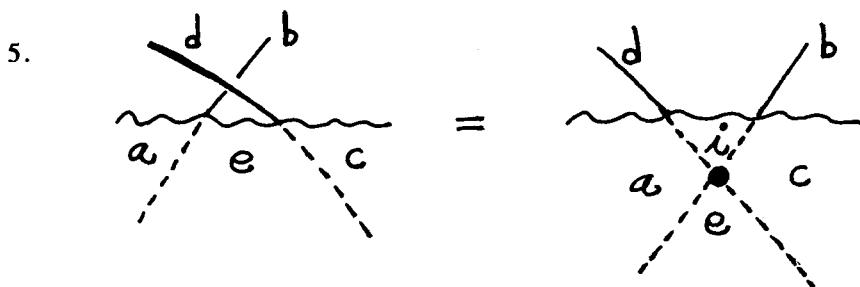
To illustrate this technique, we prove the next identity. This one shows how to translate a crossing across the shadow line. In order to explicate this identity, we need to recall a formula from the recoupling theory - namely that a 3-vertex with a twist in two of the legs is equal to an untwisted vertex multiplied by a factor of λ_c^{ab} . We refer to [KL93] for the precise value of lambda as a function of a , b and c .

$$\text{Diagram: } \text{vertex with } a, b \text{ crossed and } c = \lambda_c^{ab} \text{ vertex with } a, b \text{ untwisted and } c$$

We define two shadow vertices to correspond to the two types of unoriented crossings that occur with respect to a vertical direction in the plane. Each shadow crossing corresponds to a tetrahedral evaluation multiplied by a lambda factor. This lambda factor is the phase referred to above. This correspondence is illustrated below.



The crossing identity then becomes the following shadow equation.



Proof.

$$\begin{aligned}
 & \text{Diagram showing a horizontal line with points } a, e, c. \text{ A vertical line } b \text{ passes through } e. \text{ A diagonal line } d \text{ passes through } a \text{ and } e. \\
 = & \text{Diagram showing a horizontal line with points } a, e, c. \text{ A vertical line } b \text{ passes through } e. \text{ A wavy line } d \text{ connects } a \text{ to } e. \\
 = & \text{Diagram showing a horizontal line with points } a, e, c. \text{ A vertical line } b \text{ passes through } e. \text{ A wavy line } d \text{ connects } a \text{ to } e. \cdot \lambda_c^{de} \\
 = & \sum_i \text{Diagram showing a horizontal line with points } a, e, c. \text{ A vertical line } b \text{ passes through } e. \text{ A wavy line } d \text{ connects } a \text{ to } e. \text{ A point } i \text{ is marked on the wavy line.} \\
 & \quad \left\{ \begin{matrix} a & b & i \\ c & d & e \end{matrix} \right\} \lambda_c^{de} \\
 = & \sum_i \text{Diagram showing a horizontal line with points } a, e, c. \text{ A vertical line } b \text{ passes through } e. \text{ A wavy line } d \text{ connects } a \text{ to } e. \\
 & \quad \left\{ \begin{matrix} a & b & i \\ c & d & e \end{matrix} \right\} \bar{\lambda}_i^{ad} \bar{\lambda}_c^{ed} //
 \end{aligned}$$

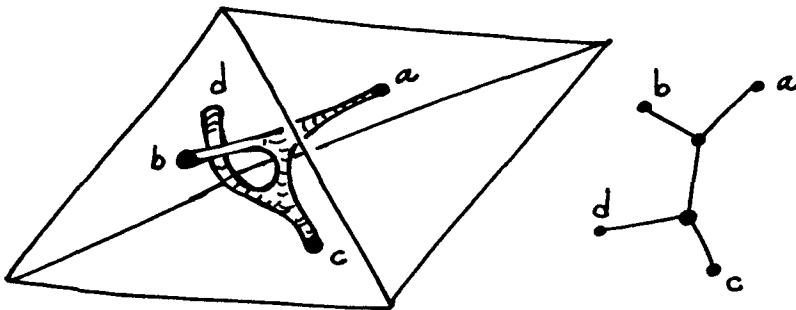
These formulas provide the means to rewrite the handle sliding invariant as a partition function on a two-cell complex. They also provide the means to rewrite the general bracket polynomial with Temperley Lieb projectors inserted into the lines as a similarly structured partition function. This gives an explicit (albeit complicated) formula for these invariants for arbitrary link diagrams.

Once the handle-sliding invariant is normalized to become an invariant of 3-manifolds, these techniques become very useful for studying the invariants. In particular this structure can be used to give a proof of the theorem of Turaev and Walker [TU92],[WA91] relating the Witten-Reshetikhin-Turaev invariant with the Turaev-Viro invariant. It remains to be seen what further applications and relationships will arise from this point of view.

VIII. The Invariants of Ooguri, Crane and Yetter

In [OG93] Ooguri discusses a method to obtain a piecewise linear invariant of 4-manifolds via a partition function defined on the triangulation. His partition function is a formal invariant for the classical group $SU(2)$, and he suggests that his approach will give a finitely defined invariant for the quantum group $SU(2)_q$ with q a root of unity. In [CR93] Crane and Yetter take up the theme of Ooguri's invariant, formulating it with respect to a modular tensor category and showing that the scheme does indeed lead to an invariant of 4-manifolds in the case of the $SU(2)_q$ quantum group at roots of unity. In this case, the technique is to apply the recoupling network theory for $SU(2)_q$ due to Kirillov and Reshetikhin. As we have remarked before, it is always possible to replace this particular network theory with the Temperley Lieb algebra recoupling theory. Thus in the case of this invariant there is a formulation that uses the Temperley Lieb recoupling theory, bringing it into the context of this paper.

We are given an oriented 4-manifold M with a specific triangulation. Each 4-simplex of M has a boundary consisting of five 3-simplices. To each 3-simplex is associated a network with four free ends and two trivalent vertices, as shown below.



The edges of this network are labelled in correspondence to a coloring of the faces of the 2-simplices in the triangulation. The five boundary simplices of a given 4-simplex yield a closed network whose evaluation is referred to as the $15 - j$ symbol associated with the given (colored) 4-simplex. The partition function for M is then the sum over all colorings of the products of these $15 - j$ symbols multiplied by extra factors from edges and faces.

The invariance is verified via the use of combinatorial moves that generate piecewise linear homeomorphisms of 4-manifolds. At this writing it

is too early to assess the significance of this invariant, but it is exceedingly remarkable that it can be constructed using only properties of the $SU(2)_q$ spin networks.

References

- [ASH92] Abhay Ashtekar, Carlo Rovelli, Lee Smolin. *Weaving a classical geometry with quantum threads*. (preprint 1992).
- [BL79] L. C. Biedenharn and J.D. Louck. Angular Momentum in Quantum Physics-Theory and Application. (*Encyclopedia of Mathematics and its Applications*). Cambridge University Press (1979).
- [BR92] Bernd Brugmann, Rodolfo Gambini, Jorge Pullin. *Knot invariants as nondegenerate quantum geometries*. Phys. Rev. Lett. **68**, No. 4 (27 Jan. 1992), 431–434.
- [CR91] L. Crane. *Conformal field theory, spin geometry and quantum gravity*. Phys. Lett. **B259**, No. 3 (1991), 243–248.
- [CR93] L. Crane and D. Yetter. *A categorical construction of 4D topological quantum field theories*. (preprint 1993).
- [FR79] R. A. Fenn and C. P. Rourke. *On Kirby's calculus of links*. Topology **18** (1979), 1–15.
- [HP81] B. Hasslacher and M. J. Perry. *Spin networks are simplicial quantum gravity*. Phys. Lett. **103B**, No. 1 (July 1981).
- [JO83] V. F. R. Jones. *Index for subfactors*. Invent. Math. **72** (1983), 1–25.
- [JO85] V. F. R. Jones. *A polynomial invariant for links via von Neumann algebras*. Bull. Amer. Math. Soc. **129** (1985), 103–112.
- [JO86] V. F. R. Jones. *A new knot polynomial and von Neumann algebras*. Notices of AMS **33** (1986), 219–225.
- [JO87] V. F. R. Jones. *Hecke algebra representations of braid groups and link polynomials*. Ann. of Math. **126** (1987), 335–38.
- [KA87] L. H. Kauffman. *State models and the Jones polynomial*. Topology **26** (1987), 395–407.
- [KAU87] L. H. Kauffman. *On knots*. Annals of Mathematics Studies No. 115, Princeton University Press (1987).
- [KA88] L. H. Kauffman. *New invariants in the theory of knots*. Amer. Math. Monthly **95**, No. 3 (March 1988), 195–242.
- [KA89] L. H. Kauffman. *Spin networks and the Jones polynomial*. Twistor Newsletter **29** (1989), 25–29.
- [KA90] L. H. Kauffman. *An invariant of regular isotopy*. Trans. Amer. Math. Soc. **318**, No. 2 (1990), 417–471.
- [Kauff90] L. H. Kauffman. *Spin networks and knot polynomials*. Int. J. Mod. Phys. **A5**, No. 1 (1990), 93–115.

- [Kauffm90] L. H. Kauffman. *Map coloring and the vector cross product.* J. Comb. Theo. Ser.B, **48**, No. 2 (April 1990), 145–154.
- [Kauff91] L. H. Kauffman. *$SU(2)q$ - Spin networks.* Twistor Newsletter **32** (1991), 10–14.
- [KA91] L. H. Kauffman. *Knots and Physics*, World Scientific (1991).
- [KA92] L. H. Kauffman. *Knots, Spin Networks and 3-Manifold Invariants.* Knots **90** (edited by A. Kawauchi) pp. 271-287 (1992).
- [KAU92] L. H. Kauffman. *Map coloring, q -deformed spin networks, and Turaev-Viro invariants for 3-manifolds.* In the Proceedings of the Conference on Quantum Groups (Como, Italy June 1991), edited by M. Rasetti, World Sci. Pub., Intl. J. Mod. Phys. **B6**, Nos. 11 & 12 (1992), 1765–1794.
- [KL90] L. H. Kauffman and S. Lins. *A 3-manifold invariant by state summation.* (announcement 1990).
- [KL91] L. H. Kauffman and S. Lins. *Computing Turaev-Viro invariants for 3-manifolds.* Manuscripta Math. **72** (1991), 81–94.
- [KL93] L. H. Kauffman and S. Lins. *Temperley Lieb recoupling theory and invariants of 3-manifolds.* (to appear)
- [KS92] L.H. Kauffman and H. Saleur. *Map coloring and the Temperley Lieb algebra.* (to appear in Comm. Math. Phys.)
- [KI78] R. Kirby. *A calculus for framed links in S^3 .* Invent. Math. **45** (1978), 35–56.
- [KM91] R. Kirby and P. Melvin. *On the 3-manifold invariants of Reshetikhin-Turaev for $sl(2, C)$.* Invent. Math. **105** (1991), 473–545.
- [KR88] A. N. Kirillov. and N. Y. Reshetikhin. *Representations of the algebra $U_q(sl_2)$, q -orthogonal polynomials and invariants of links.* In Infinite Dimensional Lie Algebras and Groups, ed. by V.G. Kac. Adv. Ser. in Math. Phys. **7** (1988), 285–338.
- [KRY92] S. I. Kryuchkov. *The four color theorem and trees.* I. V. Kurchatov Institute of Atomic Energy, Moscow (1992) - IAE-5537/1.
- [LI62] W. B. R. Lickorish. *A representation of orientable combinatorial 3-manifolds.* Ann. of Math. **76** (1962), 531–540.
- [LI90] W. B. R. Lickorish. *Calculations with the Temperley-Lieb algebra.* (preprint 1990).
- [LI91] W. B. R. Lickorish. *3-Manifolds and the Temperley Lieb algebra.* Math. Ann. **290** (1991), 657–670.
- [LI92] W. B. R. Lickorish. *Skeins and handlebodies.* (preprint 1992).
- [LIC92] W. B. R. Lickorish. *The Temperley Lieb Algebra and 3-manifold invariants.* (preprint 1992). (and private communication)
- [LD91] S. Lins and C. Durand. *Topological classification of small graph-encoded orientable 3-manifolds.* Notas Com. Mat. UFPE **177**, (1991).

- [MV92] G. Masbaum and P. Vogel. *3-valent graphs and the Kauffman bracket.* (preprint 1992).
- [MOU79] J. P. Moussouris. *The chromatic evaluation of strand networks.* In Advances in Twistor Theory (ed. by Huston and Ward) Research Notes in Mathematics. Pitman Pub. (1979), 308–312.
- [MOU83] J. P. Moussouris. *Quantum models of space-time based on recoupling theory.* (Mathematics Thesis, Oxford University - 1983).
- [OG91] H. Ooguri. *Discrete and continuum approaches to three-dimensional quantum gravity.* (preprint 1991).
- [OG92] H. Ooguri. *Topological lattice models in four dimensions.* Mod. Phys. Lett. **A7**, No. 30 (1992), 2799–2810.
- [PEN69] R. Penrose. *Angular momentum: an approach to combinatorial space-time.* In Quantum Theory and Beyond (ed. T.A. Bastin) Cambridge Univ. Press (1969).
- [PEN71] R. Penrose. *Applications of negative dimensional tensors.* Combinatorial Mathematics and its Applications (ed. by D. J. A. Welsh) Academic Press (1971).
- [PEN79] R. Penrose. *Combinatorial quantum theory and quantized directions.* Advances in Twistor Theory (ed by L. P. Hughston and R. S. Ward) Pitman (1979) pp. 301-307.
- [PIU92] Sergey Piunikhin. *Turaev-Viro and Kauffman-Lins invariants for 3-manifolds coincide.* J. Knot Theory and its Ramifications **1**, No. 2 (1992), pp. 105-135.
- [SM88] Lee Smolin. *Quantum gravity in the self-dual representation.* Contemp. Math. **71** (1988), pp. 55-97.
- [RE87] N. Y. Reshetikhin. *Quantized universal enveloping algebras, the Yang-Baxter equation and invariants of links, I and II.* LOMI reprints E-4-87 and E-17-87, Steklov Institute, Leningrad, USSR.
- [RT90] N. Y. Reshetikhin and V. Turaev. *Ribbon graphs and their invariants derived from quantum groups.* Comm. Math. Phys. **127** (1990), 1–26.
- [RT91] N. Y. Reshetikhin and V. Turaev. *Invariants of three manifolds via link polynomials and quantum groups.* Invent. Math. **103** (1991), 547–597.
- [TU87] V. G. Turaev. *The Yang-Baxter equations and invariants of links.* LOMI preprint E-3-87, Steklov Institute, Leningrad, USSR. Inventiones Math. **92** Fasc.3,527-553.
- [TV90] V. G. Turaev and O. Viro. *State sum invariants of 3-manifolds and quantum 6j symbols.* (preprint 1990).
- [TU90] V. G. Turaev. *Quantum invariants of links and 3-valent graphs in 3-manifolds.* (preprint 1990).
- [TUR90] V. G. Turaev. *Shadow links and face models of statistical mechanics.* Publ. Inst. Recherche Math. Avance. Strasbourg (1990).

- [TURA90] V. G. Turaev. *Quantum invariants of 3-manifolds and a glimpse of shadow topology.* (preprint 1990).
- [TU92] V. G. Turaev. *Topology of shadows.* (preprint 1992).
- [WA91] K. Walker. *On Witten's 3-manifold invariants.* (preprint 1991).
- [WA91] R. Williams and F. Archer. *The Turaev-Viro state sum model and 3-dimensional quantum gravity.* (preprint 1991).
- [WIT89] E. Witten. *Quantum field theory and the Jones polynomial.* Commun. Math. Phys. **121** (1989), 351–399.
- [YA92] S. Yamada. *A topological invariant of spatial regular graphs.* Knots 90 (edited by Akio Kawauchi). de Gruyter (1992), pp. 447-454.

LINK POLYNOMIALS AND A GRAPHICAL CALCULUS

LOUIS H. KAUFFMAN and PIERRE VOGEL

ABSTRACT

This paper constructs invariants of rigid vertex isotopy for graphs embedded in three dimensional space. For the Homfly and Dubrovnik polynomials, the skein formalism for these invariants is as shown below.

Homfly.

$$\left\{ \begin{array}{l} \left[\begin{array}{c} \nearrow \\ \times \\ \searrow \end{array} \right] = A \left[\begin{array}{c} \nearrow \\ \nearrow \\ \searrow \end{array} \right] + \left[\begin{array}{c} \nearrow \\ \times \\ \searrow \end{array} \right] \\ \left[\begin{array}{c} \nearrow \\ \times \\ \searrow \end{array} \right] = B \left[\begin{array}{c} \nearrow \\ \nearrow \\ \searrow \end{array} \right] + \left[\begin{array}{c} \nearrow \\ \times \\ \searrow \end{array} \right] \\ - \left[\begin{array}{c} \nearrow \\ \times \\ \searrow \end{array} \right] = (A - B) \left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right] \end{array} \right\} a \left[\begin{array}{c} \nearrow \\ \parallel \\ \searrow \end{array} \right]$$

Dubrovnik.

$$\left\{ \begin{array}{l} \left[\begin{array}{c} \times \\ \times \end{array} \right] = A \left[\begin{array}{c} () \\ () \end{array} \right] + B \left[\begin{array}{c}) \\ (\end{array} \right] + \left[\begin{array}{c} \times \\ \times \end{array} \right] \\ \left[\begin{array}{c} \times \\ \times \end{array} \right] = B \left[\begin{array}{c} () \\ () \end{array} \right] + A \left[\begin{array}{c}) \\ (\end{array} \right] + \left[\begin{array}{c} \times \\ \times \end{array} \right] \\ - \left[\begin{array}{c} \times \\ \times \end{array} \right] = (A - B) \left(\left[\begin{array}{c} () \\ () \end{array} \right] - \left[\begin{array}{c}) \\ (\end{array} \right] \right) \end{array} \right\} a \left[\begin{array}{c} \nearrow \\ \parallel \\ \searrow \end{array} \right]$$

0. Introduction

The purpose of this paper is to study models for both the Homfly and Kauffman (2-variable) polynomial invariants of links embedded in three-dimensional space. In both cases, the models extend these invariants to *three variable* invariants of 4-valent rigid-vertex embeddings of graphs in three-space. The models are based upon *graphical calculi* that assign well-defined polynomials to planar 4-valent graphs.

These graphical calculi determine values for each planar 4-valent graph by recursive formulas defined entirely in the category of graphs.

The approach to invariants in this paper is based on the combinatorial approach to knot theory. Thus we use standard knot and link diagrams, and define two diagrams to be *equivalent* (ambient isotopic) if they are related by a sequence of Reidemeister moves of type I, II or III. Figure 1 indicates representatives for each Reidemeister move. It is also useful to consider the equivalence generated by moves II and III alone. This is referred to as *regular isotopy* (see [5], [8], [10]).

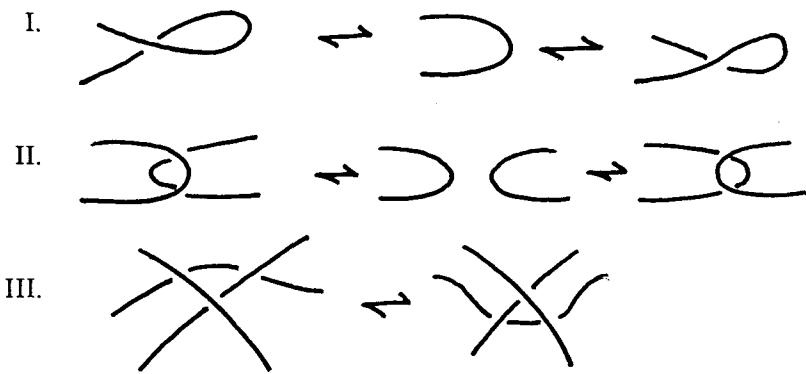


Figure 1

This paper is organized as follows. In section 1 we review the concept of rigid vertex isotopy for graphs embedded in three dimensional space. We also point out a general scheme for producing rigid vertex isotopy invariants of 4-valent graph embeddings from *any* regular isotopy invariant of knots and links. This method yields non-trivial results even for the writhe. The rest

of the paper uses this method in the context of the Homfly and Kauffman polynomials. In section 2 we review the Homfly polynomial in the context of regular isotopy, and explain how the state-model and graphical calculus works in that case. Proofs of the formulas in the graphical calculus are postponed to section 4. Section 3 shows that in the case of braids the identities in the graphical calculus are essentially intertwined with the generating relations for the Hecke algebra [4] associated with the n -strand braid group. This yields a new viewpoint on the Hecke algebra, and provides a new interpretation of the Jones-Ocneanu trace. Section 5 explains how our ideas work for the Dubrovnik version of the 2-variable Kauffman polynomial. Once again, in the case of braids and the braid-monoid ([5], [1]) *we obtain a new interpretation of the basic algebra relations for the Birman-Wenzl algebra in terms of a graphical calculus.* Proofs of formulas in the graphical calculus for the Dubrovnik polynomial are given in section 6. Section 7 discusses open problems and relationships with other state models.

We would like to take this opportunity to thank Ken Millett for sharing his ideas about state models with us. The approach taken in this paper is complementary to that of Millett-Jonish [12]. We deduce our graphical calculi only from the known existence of the two-variable polynomials. Finally, it gives us great pleasure to thank Jon Simon for many graphic conversations.

Work for this paper has been partially supported by NSF Grant Number DMS-8822602 and by the Program for Mathematics and Molecular Biology, University of California at Berkeley, Berkeley, California.

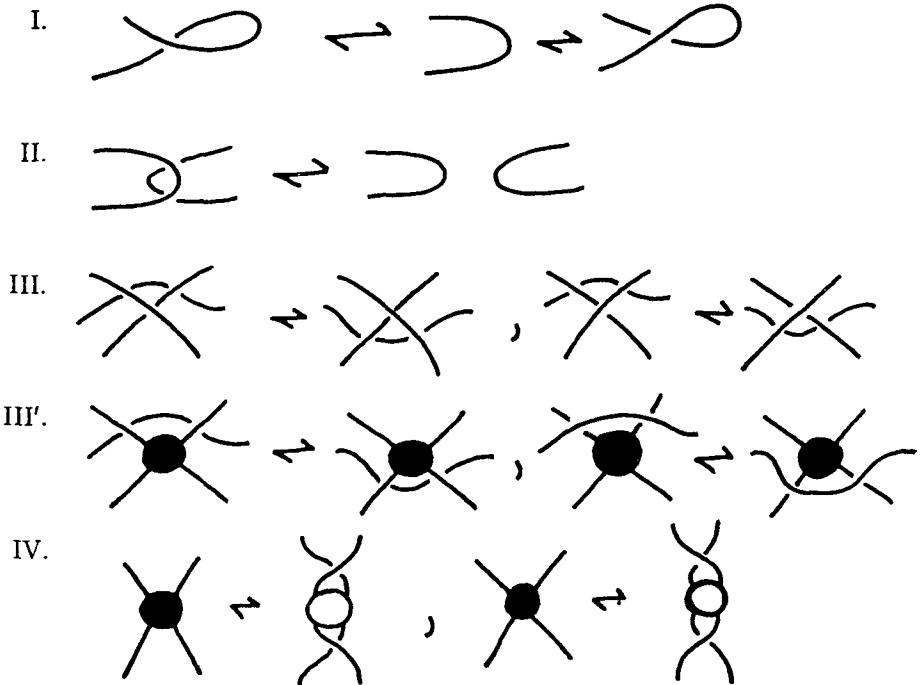
1. Rigid Vertex Isotopy

Before going into the main body of the paper, recall [10] some general facts about the properties of rigid-vertex isotopy for embeddings of 4-valent graphs. As explained in [10] a 4-valent graph with rigid vertices can be regarded as an embedding of a graph whose vertices have been replaced by rigid disks. Each disk has four strands attached to it, and the cyclic order of these strands is determined via the rigidity of the disk. An *RV*-isotopy (rigid vertex isotopy) of the embedding of such a graph G in \mathbf{R}^3 consists in affine motions of the disks, coupled with topological ambient isotopies of the strands (corresponding to the edges of G).

This notion of *RV* isotopy is a mixture of mechanical (Euclidean) and topological concepts. It arises naturally in the building of models for graph

embeddings, and it also arises naturally in regard to creating invariants of graph embeddings.

In [10] we have derived a collection of moves, analogous to the Reidemeister moves of Figure 1, that generate *RV*-isotopy for diagrams of 4-valent graph embeddings. We refer the reader to [10] for more information on this topic. The moves for *RV* – 4 isotopy are shown in Figure 1'.



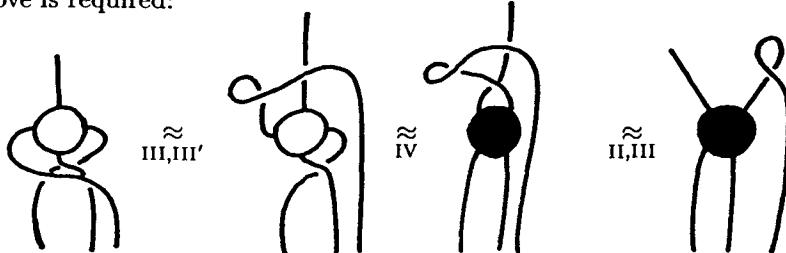
Graph Moves that Generate Rigid Isotopy
Figure 1'

One important combination of these move-types is discussed in [10] and must be mentioned again here. Consider the rigid isotopy shown below:



Call this move a *3-twist*. The 3-twist twists a triple of strands attached to

the rigid disk by π , while one strand is invariant - since it is the axis of the rotation. The three-twist can be accomplished by the rigid-isotopy moves I, II, III, III', and IV. However, if we look carefully, then we see that the type I move is required:



The equivalence relation generated by rigid vertex isotopy moves, II, III, III', and IV is called *regular vertex isotopy* (in analogy with regular isotopy for link diagrams). We see that if H_K is an invariant of regular (vertex) isotopy, then

$$H\left(\begin{array}{c} \text{rigid vertex} \\ \text{with strands} \end{array}\right) \text{ and } H\left(\begin{array}{c} \text{rigid vertex} \\ \text{with strands} \end{array}\right)$$

will differ by this invariant's behavior on the type I move. The invariants discussed in this paper will all have the property of being *multiplicative under type I moves*:

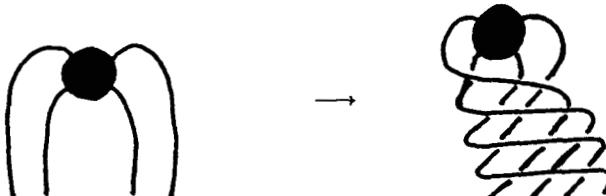
$$H_{\text{---}} = a H_{\text{---}}$$

$$H_{\text{---}} = a^{-1} H_{\text{---}}$$

for an appropriate scalar a . Thus when extending such an invariant to an invariant for rigid vertex isotopy classes of graphs we will find that it is well-defined *up to multiplication by powers of a* . This is sufficient for most of our purposes.

The same remarks that we have made for the 3-twist apply to the 4-

twist:

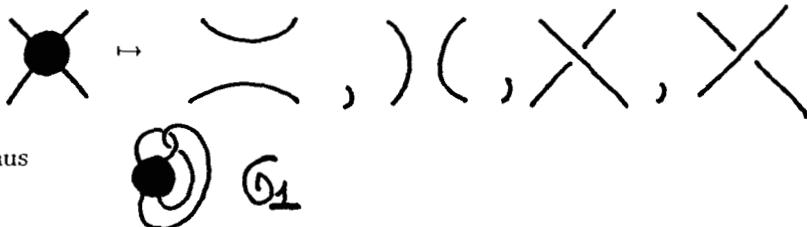


The 4-twist is the result of applying a full 360° twist to the four parallel strands. It is easy to see that the resulting diagram is regularly vertex isotopic to the original with some curls added as shown below:



With these facts in mind, let's turn to general comments about invariants of *RV4* isotopy.

In [10] it is pointed out that a collection $\mathcal{C}(G)$ of knots and links associated with the graph G is *as a collection of ambient isotopy classes of knots and links* an invariant of rigid isotopy of the graph G . This is explained for both oriented and unoriented graphs. In the case of unoriented graphs, we create $\mathcal{C}(G)$ by replacing each vertex by *crossings or smoothings*:



Thus



$$c(G_1) = \{\text{Trefoil}, \text{Trefoil}, \text{Trefoil}, \text{Trefoil}\}$$

$$= \{\text{Trefoil}, \text{O}, \text{Trefoil}\}.$$

(We take the associated *set* of knots and links up to ambient isotopy.) While



$$\begin{aligned} c(G_0) &= \left\{ \text{Trefoil}, \text{Figure-eight} \right\} \\ &= \left\{ \text{Trefoil}, \text{Figure-eight}^* \right\}. \end{aligned}$$

It follows that the graphs G_0 and G_1 are not rigid vertex isotopic since the knot and links types in their associated collections are distinct. It also follows that G_1 and G_1^* (the mirror image of G_1) are distinct, since $C(G_1^*)$ contains one knot (a trefoil) that is the mirror image of the corresponding knot in $C(G_1)$ - and the trefoil and its mirror image are well-known to be topologically distinct.

Given two graphs G, G' , one would like to discriminate the collections $C(G)$ and $C(G')$ if this is possible.

This can be done by applying invariants to the elements of the collection. In this paper, we do this with extra structure by creating (polynomial) invariants for the graphs from known (polynomial) invariants for knots and links. We shall explain the most general procedure of this kind in this introduction.

Thus, assume that $I(K)$ is a polynomial or scalar (integer, complex number, ...) valued regular isotopy invariant of knots and links K . Assume that $I(K)$ behaves multiplicatively under the type I move, so that

$$I(\text{Trefoil}) = a I(\text{Figure-eight})$$

$$I(\text{Trefoil}^*) = a^{-1} I(\text{Figure-eight}).$$

Unoriented Invariants of Graphs

Suppose that the graph G is unoriented and that $I(K)$ is an unoriented regular isotopy invariant for knots and links K . Then we define $I(G)$ recursively by the formula:

$$\begin{aligned} I\left(\text{X}\right) &= xI\left(\text{U}\right) + xI\left(\text{O}\right) + \\ &yI\left(\text{X}\right) + yI\left(\text{X}\right) \end{aligned}$$

where x and y are commuting algebraic variables.

Theorem 1.1. $I(G)$ is well-defined, and an invariant of rigid-vertex isotopy for the graphs G (up to \doteq where \doteq denotes equality up to powers of a).

Proof. This is a straightforward check via the graph-moves (Figure 1'). We will check move IV, and leave the rest for the reader.

$$\begin{aligned} I\left(\text{X}\right) &= xI\left(\text{U}\right) + xI\left(\text{O}\right) \\ &+ yI\left(\text{X}\right) + yI\left(\text{X}\right) \\ &= xaa^{-1}I\left(\text{U}\right) + xI\left(\text{O}\right) \\ &+ yI\left(\text{X}\right) + yI\left(\text{X}\right) = I\left(\text{X}\right). \end{aligned}$$

Well-definedness follows from the fact that x and y commute. Formally, we can define a *link-state* σ of G to be any link obtained from G by replacing each vertex in G by one of the smoothings or crossings in the expansion formula. Then

$$I(G) = \sum_{\sigma} x^{s(\sigma)} y^{c(\sigma)} I(\sigma)$$

where $s(\sigma)$ is the number of smoothings in σ and $c(\sigma)$ is the number of crossing replacements needed to obtain σ from G . This completes the proof. \square

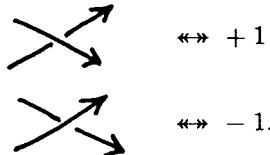
Remark. In principle, $I(G)$ for $I(\sigma) = \text{ambient isotopy class of } \sigma$, contains more information than the set $\mathcal{C}(G)$. The production of a specific invariant through such a summation is useful for handling information, both theoretically and computationally.

Example.



$$\begin{aligned}
 I(G) &= xI(\text{trefoil}) + xI(\text{circle}) + yI(\text{circle}) + yI(\text{circle}) \\
 &= xI(\text{trefoil}) + xa^{-2}I(\text{circle}) \\
 &\quad + yI(\text{circle}) + yaI(\text{circle}) \\
 &= (xa^{-2} + ya)I(\text{circle}) + xI(\text{trefoil}) + yI(\text{circle}).
 \end{aligned}$$

Example. Let $I(K) = a^{w(K)}$ where $w(K)$ is the *writhe of K* if K is a knot and $I(K) = 0$ if K is a link (> 1 component). The writhe is the sum of the crossing signs of the diagram K with



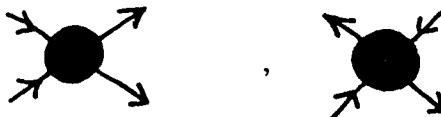
Unoriented knots receive a unique writhe, independent of the choice of orientation. In the example above, we then have

$$\begin{aligned}
 I(G) &= (xa^{-2} + ya) + ya^3 = xa^{-2} + y(a + a^3) \\
 &= a^{-2}[x + y(a^3 + a^5)] \\
 I(G^*) &= xa^2 + y(a^{-1} + a^{-3}) = a^2[x + y(a^{-3} + a^{-5})].
 \end{aligned}$$

Since $I(G) \neq I(G^*)$ we conclude that the graphs G and G^* are not equivalent in rigid vertex isotopy. The interest of this example is that it shows that the invariant $I(G)$ can detect chirality using very elementary means!

Oriented Graph Invariants

An oriented rigid vertex can take basic forms:



In general, it is assumed that the vertex has two incoming and two outgoing arrows. At a *given site* this means *six possible orientations*:



If $I(K)$ is an oriented, multiplicative regular isotopy invariant of knots and links, define $I(G)$ for oriented graphs (the edges are oriented so that the given local conditions of 2-in, 2-out are satisfied at each graphical vertex) by the expansions:

$$\begin{aligned} I\left(\begin{array}{c} \nearrow \\ \searrow \\ \text{---} \\ \swarrow \\ \nwarrow \end{array}\right) &= xI\left(\begin{array}{c} \nearrow \\ \text{---} \\ \text{---} \\ \swarrow \\ \nwarrow \end{array}\right) + yI\left(\begin{array}{c} \nearrow \\ \text{---} \\ \text{---} \\ \swarrow \\ \nearrow \end{array}\right) + zI\left(\begin{array}{c} \nearrow \\ \text{---} \\ \text{---} \\ \nearrow \\ \swarrow \end{array}\right). \\ I\left(\begin{array}{c} \nearrow \\ \text{---} \\ \text{---} \\ \swarrow \\ \nwarrow \end{array}\right) &= rI\left(\begin{array}{c} \downarrow \\ \uparrow \end{array}\right) + sI\left(\begin{array}{c} \nearrow \\ \text{---} \\ \text{---} \\ \swarrow \\ \nwarrow \end{array}\right). \end{aligned}$$

Once again we have the general

Theorem 1.2. Suppose that G and G' are $RV4$ isotopic oriented graphs, and that $I(K)$ is a regular isotopy invariant for oriented knots and links. Then $I(G) = I(G')$.

Proof. We check move IV, leaving the rest for the reader:

$$\begin{aligned}
 I\left(\begin{array}{c} \text{graph} \\ \text{with move IV} \end{array}\right) &= rI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) + rI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) \\
 &= ra^{-1}aI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) + rI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) \\
 &= I\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) \\
 I\left(\begin{array}{c} \text{graph} \\ \text{with move IV} \end{array}\right) &= xI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) + yI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) + zI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) \\
 &= xI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) + yI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) + zI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) \\
 &= I\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right). \\
 I\left(\begin{array}{c} \text{graph} \\ \text{with move IV} \end{array}\right) &= xI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) + yI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) + zI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) \\
 &= xaa^{-1}I\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) + yI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) + zI\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right) \\
 &= I\left(\begin{array}{c} \text{graph} \\ \text{without move IV} \end{array}\right).
 \end{aligned}$$

This completes the proof. □

This completes the description of the production of invariants of *RV4* graphs from invariants of knots and links. One good set of invariants of knots and links to use in this regard are the “classical” 2-variable knot polynomials (they are generalizations of the original Jones polynomial) - the Homfly and Kauffman polynomials. The rest of the paper is devoted to the rather remarkable interrelationship of these particular polynomials with graph invariants. In the case of these polynomials, it is not just that they can be used in the $I(G)$ general scheme. Rather, the relationship with graphs is actually fundamental to the nature of these polynomials.

Thus we shall define generalizations of the 2-variable knot polynomials to graphs and then show that the specializations of these invariants to planar graphs are sufficient to provide models for the 2-variable polynomials themselves. The invariants of planar graphs can be computed entirely through

a graphical calculus developed in the paper. This calculus shows that the planar graph evaluations form significant generalizations of the Hecke [4] and Birman-Wenzl [1] algebras related to these polynomials. Finally, in the second appendix to the paper we show how these graph polynomials are related to those that have been constructed previously in relation to skein models ([6], [9], [3]) for the knot polynomials.

While this paper does not consider $I(G)$ when I is any other invariant of knots and links than a knot polynomial, there is clearly a whole area to be investigated here. In particular, there are many invariants that arise directly from solutions to the Yang-Baxter Equation [13] and these will automatically give $RV4$ invariants by our scheme.

2. The Homfly Polynomial

The Homfly polynomial is a two-variable generalization of the Conway-Alexander polynomial and the original Jones polynomial. We shall denote the Homfly polynomial for a link L by $P_L = P_L(a, z)$. This (Laurent) polynomial is completely determined by the following axioms:

Axioms for the Homfly Polynomial

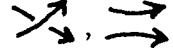
1. If links L and L' are ambient isotopic, then $P_L = P_{L'}$.
2. If three link diagrams differ only at the sites of the small diagrams
 ,  ,  , then the following identity is valid:

$$aP \begin{array}{c} \nearrow \\ \searrow \end{array} - a^{-1}P \begin{array}{c} \searrow \\ \nearrow \end{array} = zP \begin{array}{c} \nearrow \\ \nearrow \end{array} .$$

Remark. Some authors [11] use ℓ^{-1} for a and m for z in the above identity. We have chosen the letters a and z for our own convenience. The Conway polynomial is obtained for $a = 1$; it is denoted $\nabla_L = \nabla_L(z)$ and satisfies the identity $\nabla \begin{array}{c} \nearrow \\ \searrow \end{array} - \nabla \begin{array}{c} \searrow \\ \nearrow \end{array} = z\nabla \begin{array}{c} \nearrow \\ \nearrow \end{array} .$

For our purposes, it is useful to *reformulate the Homfly polynomial in the regular isotopy category*. Thus we define a polynomial $R_L(a, z) = R_L$ (See [8], Chapter 6) that is an invariant of regular isotopy (generated by Reidemeister moves II and III):

Axioms for the R -Polynomial

1. If L and L' are *regularly isotopic* then $R_L = R_{L'}$.
2. If  denote parts of larger link diagrams that differ only in these smaller diagrams, then

$$R \begin{array}{c} \nearrow \\ \searrow \end{array} - R \begin{array}{c} \nearrow \\ \swarrow \end{array} = z R \begin{array}{c} \nearrow \\ \rightarrow \end{array}$$

3. R_L behaves as indicated below under type I moves:

$$\begin{aligned} R \begin{array}{c} \nearrow \\ \overbrace{\searrow}^{\text{---}} \end{array} &= aR \\ R \begin{array}{c} \nearrow \\ \overbrace{\searrow}^{=} \end{array} &= a^{-1}R \end{aligned}$$

The R -polynomial is the regular isotopy version of the Homfly polynomial. In fact, we have the following result: Let $w(L)$ denote the *writhe* (or twist number) of the oriented link L . That is,

$$w(L) = \sum_p \epsilon(p)$$

where this sum is taken over all crossings in the link diagram L , and $\epsilon(p)$ denotes the sign (± 1) of a crossing. The convention for this sign is indicated in Figure 2.



Crossing Signs
Figure 2

Then

Proposition 2.1. The Homfly polynomial is a writhe-normalization of its regular isotopy counterpart:

$$P_L(a, z) = a^{-w(L)} R_L(a, z).$$

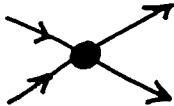
Proof. Note that $w(\text{X}) = 1 + w(\text{--})$ and $w(\text{X}) = -1 + w(\text{--})$. Then

$$R \text{X} - R \text{X} = zR \text{--}$$

$$\begin{aligned} & aa^{-w(\text{--})-1} R \text{X} - a^{-1} a^{-w(\text{--})+1} R \text{X} \\ &= za^{-w(\text{--})} R \text{--} \\ &\therefore a(a^{-w(\text{X})} R \text{X}) - a^{-1}(a^{-w(\text{X})} R \text{X}) \\ &= z(a^{-w(\text{--})} R \text{--}). \end{aligned}$$

This last identity shows that $a^{-w(L)} R_L$ satisfies the same identity as P_L . Hence they are equal. \square

We are now ready to describe the form of this state-model for the Homfly polynomial. Let $[L]$ denote R_L for any oriented link L . Then we shall extend the domain of definition for $[L]$ to diagrams with 4-valent graphical vertices such as



with orientations as shown above. Then R_L will satisfy the following

$$(*) \quad \left\{ \begin{array}{l} \left[\begin{array}{c} \text{X} \\ \text{X} \end{array} \right] = A \left[\begin{array}{c} \text{--} \\ \text{--} \end{array} \right] + \left[\begin{array}{c} \text{X} \\ \text{X} \end{array} \right] \\ \left[\begin{array}{c} \text{X} \\ \text{X} \end{array} \right] = B \left[\begin{array}{c} \text{--} \\ \text{--} \end{array} \right] + \left[\begin{array}{c} \text{X} \\ \text{X} \end{array} \right] \end{array} \right\}$$

where A and B are polynomial variables subject to the restriction that

$$z = A - B.$$

Note that it follows at once from this formalism that

$$\left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right] - \left[\begin{array}{c} \searrow \\ \nearrow \end{array} \right] = (A - B) \left[\begin{array}{c} \nearrow \\ \nearrow \end{array} \right],$$

hence

$$\left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right] - \left[\begin{array}{c} \searrow \\ \nearrow \end{array} \right] = z \left[\begin{array}{c} \nearrow \\ \nearrow \end{array} \right].$$

This is the desired exchange identity.

In order to create this model, we shall go about the business upside-down, defining the value of $[G]$ for G a 4-valent planar graph in terms of the known values of R_L on knots and links. In other words, we shall take *as definitions* the formulas

$$\begin{aligned} (**)\quad \left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right] &= \left[\begin{array}{c} \searrow \\ \nearrow \end{array} \right] - A \left[\begin{array}{c} \nearrow \\ \nearrow \end{array} \right] \\ \left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right] &= \left[\begin{array}{c} \searrow \\ \nearrow \end{array} \right] - B \left[\begin{array}{c} \nearrow \\ \nearrow \end{array} \right] \end{aligned}$$

(with $z = A - B$ and $\left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right] - \left[\begin{array}{c} \searrow \\ \nearrow \end{array} \right] = z \left[\begin{array}{c} \nearrow \\ \nearrow \end{array} \right]$ for link diagrams.)

For example, we have

$$\begin{aligned} \left[\begin{array}{c} \infty \\ \bullet \end{array} \right] &= \left[\begin{array}{c} \infty \\ \bullet \end{array} \right] - B \left[\begin{array}{c} \bullet \\ \bullet \end{array} \right] \\ &= a^{-1} - B \left(\frac{a - a^{-1}}{z} \right) \\ &= a^{-1} - B \left(\frac{a - a^{-1}}{A - B} \right) \\ &= (a^{-1}(A - B) - B(a - a^{-1}))/ (A - B) \\ \therefore \left[\begin{array}{c} \infty \\ \bullet \end{array} \right] &= (a^{-1}A - aB)/(A - B). \end{aligned}$$

Here we have used some facts about $R_L = [L]$. Letting $\delta = (a - a^{-1})/z$, it is easy to check that $[L \bullet] = \delta[L]$ where the left-hand side of this equation connotes disjoint union of L with a trivial component.

Note also that according to equations (*) (or (**)) if all crossings in a graph-diagram are switched, then the variables A, B and a, a^{-1} are interchanged (B with A , a^{-1} with a). Thus, for a planar graph G (without

crossings) we expect that $[G]$ is invariant under the simultaneous interchange of $A \leftrightarrow B$, $a \leftrightarrow a^{-1}$. This is indeed the case with our computed value

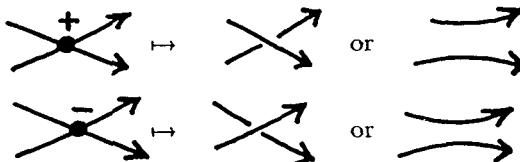
$$\left[\text{∞} \right] = (a^{-1}A - aB)/(A - B).$$

Definition of $[G]$.

In order to actually create the definition of $[G]$ for G a 4-valent graph, we proceed as follows.

- 1^o. Choose a *sign*, or , at each vertex of the graph G .
- 2^o. Define an *associated link* $\ell = \ell(G)$ to be any link diagram that is obtained from G by replacing each signed vertex by *either a crossing of that sign or by a smoothing of the vertex*.

Thus



- 3^o. Let $(G, \vec{\epsilon})$ denote the graph G where $\vec{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ is a list of the signs given to the vertices labelled $1, 2, \dots, n$. Let $\mathcal{L} = \mathcal{L}(G, \vec{\epsilon})$ denote the collection of links associated with $(G, \vec{\epsilon})$ as in 2^o above. For each $L \in \mathcal{L}$ let $i(L)$ denote the number of smoothings () corresponding to negative vertices. Define $[G, \vec{\epsilon}]$ by the formula

$$[G, \vec{\epsilon}] = \sum_{L \in \mathcal{L}} (-A)^{i(L)} (-B)^{j(L)} R_L$$

where R_L is the regular isotopy form of the Homfly polynomial as described in this section.

This definition applies to any graph-diagram. Such a diagram may contain 4-valent vertices and/or ordinary crossings. A simple example of a diagram of mixed type is



Lemma 2.2. The value of $[G, \vec{\epsilon}]$ is independent of the choice of signs $\vec{\epsilon}$.

Proof. Consider the effect of choosing a sign at a given vertex. According to our definitions, we can write

$$\left[\begin{array}{c} \nearrow \\ \nwarrow \\ \times \end{array} \right] = \left[\begin{array}{c} \nearrow \\ \nwarrow \\ \times \end{array} \right] - A \left[\begin{array}{c} \nearrow \\ \nearrow \\ \parallel \end{array} \right]$$

and

$$\left[\begin{array}{c} \nearrow \\ \nwarrow \\ \times \end{array} \right] = \left[\begin{array}{c} \nearrow \\ \nwarrow \\ \times \end{array} \right] - B \left[\begin{array}{c} \nearrow \\ \nearrow \\ \parallel \end{array} \right].$$

Here it is assumed that all other choices of sign are retained.

We can also state that for any fixed choice of signs at the vertices, $\left[\begin{array}{c} \nearrow \\ \nwarrow \end{array} \right] - \left[\begin{array}{c} \nearrow \\ \nwarrow \end{array} \right] = z \left[\begin{array}{c} \nearrow \\ \nearrow \end{array} \right]$ where $z = A - B$, since this is true for each triple of L_+, L_-, L_0 in the sums over associated links by the identity for the R -polynomial. Thus

$$\left[\begin{array}{c} \nearrow \\ \nwarrow \\ \times \end{array} \right] - \left[\begin{array}{c} \nearrow \\ \nwarrow \\ \times \end{array} \right] = \left[\begin{array}{c} \nearrow \\ \nwarrow \\ \times \end{array} \right] - \left[\begin{array}{c} \nearrow \\ \nwarrow \\ \times \end{array} \right] + (B - A) \left[\begin{array}{c} \nearrow \\ \nearrow \\ \parallel \end{array} \right] = 0.$$

This completes the proof. \square

Example.

$$\begin{aligned} \left[\begin{array}{c} \textcirclearrowleft \\ \textcirclearrowright \\ \times \end{array} \right] &= \left[\begin{array}{c} \textcirclearrowleft \\ \textcirclearrowright \\ \times \end{array} \right] - A \left[\begin{array}{c} \textcirclearrowleft \\ \textcirclearrowright \\ \textcirclearrowright \end{array} \right] \\ &= \delta - Aa^{-1} \\ &= \frac{a - a^{-1}}{A - B} - Aa^{-1} \\ &= \frac{a - a^{-1} - Aa^{-1}(A - B)}{A - B} \\ \left[\begin{array}{c} \textcirclearrowleft \\ \textcirclearrowright \\ \times \end{array} \right] &= \frac{a - a^{-1} - A^2a^{-1} + ABa^{-1}}{A - B}. \end{aligned}$$

We are now in possession of a well-defined 3-variable (A, B, a) rational function for graph-diagrams, generalizing the Homfly polynomial and satisfying the identities

$$\begin{aligned} (*) \quad \left[\begin{array}{c} \nearrow \\ \nwarrow \end{array} \right] &= A \left[\begin{array}{c} \nearrow \\ \nearrow \\ \parallel \end{array} \right] + \left[\begin{array}{c} \nearrow \\ \nwarrow \\ \times \end{array} \right] \\ \left[\begin{array}{c} \nearrow \\ \nwarrow \end{array} \right] &= B \left[\begin{array}{c} \nearrow \\ \nearrow \\ \parallel \end{array} \right] + \left[\begin{array}{c} \nearrow \\ \nwarrow \\ \times \end{array} \right]. \end{aligned}$$

Repeated applications of these identities will evaluate the Homfly polynomial for any link in terms of the value of this bracket on planar graphs. Before working on this graphical calculus, we point out some invariance properties of the graph-polynomial.

Proposition 2.3.

$$(i) \quad \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] = \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right]$$

$$\left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] = \left[\begin{array}{c} \curvearrowleft \\ \curvearrowleft \end{array} \right]$$

$$(ii) \quad \left[\begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array} \right] = \left[\begin{array}{c} \diagdown \diagup \\ \diagup \diagdown \end{array} \right]$$

Type III invariance holds for any of the other choices of crossings and orientations.

$$(iii) \quad \left[\begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array} \right] = \left[\begin{array}{c} \diagdown \diagup \\ \diagup \diagdown \end{array} \right]$$

This graphical version of type III invariance also holds for all other displays of orientation (and allowing undercrossings of the free line).

$$(iv) \quad \left[\begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array} \right] = \left[\begin{array}{c} \star \star \star \\ \diagdown \diagup \end{array} \right] = \left[\begin{array}{c} \star \star \star \\ \diagup \diagdown \end{array} \right] \text{ (and the mirror images).}$$

Proof. (i) and (ii) follow from the regular isotopy invariance of R_L . For (iii), we use the expansion formula:

$$\begin{aligned} \left[\begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array} \right] &= \left[\begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array} \right] - A \left[\begin{array}{c} \diagup \diagdown \\ \diagdown \diagdown \end{array} \right] \\ &= \left[\begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array} \right] - A \left[\begin{array}{c} \diagdown \diagup \\ \diagdown \diagup \end{array} \right] \\ &= \left[\begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array} \right]. \end{aligned}$$

For (iv) again expand:

$$\begin{aligned}
 & \left[\begin{array}{c} \text{Diagram with a rigid vertex} \\ \text{and two crossing arrows} \end{array} \right] = \left[\begin{array}{c} \text{Diagram with a rigid vertex} \\ \text{and one crossing arrow} \end{array} \right] - B \left[\begin{array}{c} \text{Diagram with a rigid vertex} \\ \text{and no crossing arrows} \end{array} \right] \\
 &= \left[\begin{array}{c} \text{Diagram with a rigid vertex} \\ \text{and two crossing arrows} \end{array} \right] - B(a^{-1}a) \left[\begin{array}{c} \text{Diagram with a rigid vertex} \\ \text{and no crossing arrows} \end{array} \right] \\
 &= \left[\begin{array}{c} \text{Diagram with a rigid vertex} \\ \text{and two crossing arrows} \end{array} \right]. \\
 \left[\begin{array}{c} \text{Diagram with a rigid vertex} \\ \text{and three crossing arrows} \end{array} \right] &= \left[\begin{array}{c} \text{Diagram with a rigid vertex} \\ \text{and two crossing arrows} \end{array} \right] - B \left[\begin{array}{c} \text{Diagram with a rigid vertex} \\ \text{and one crossing arrow} \end{array} \right] \\
 &= \left[\begin{array}{c} \text{Diagram with a rigid vertex} \\ \text{and three crossing arrows} \end{array} \right] - B \left[\begin{array}{c} \text{Diagram with a rigid vertex} \\ \text{and no crossing arrows} \end{array} \right] \\
 &= \left[\begin{array}{c} \text{Diagram with a rigid vertex} \\ \text{and three crossing arrows} \end{array} \right].
 \end{aligned}$$

This completes the proof. \square

We summarize this statement (2.3) by saying that $[G]$ is an invariant of *rigid-vertex regular isotopy* for oriented graphs G . The moves indicated in this proposition, when augmented by the standard type I move correspond to ambient isotopy of rigid vertex graphs embedded in space. One way to conceptualize the rigid vertex is to regard it as a disk that must be moved rigidly (isometrically) throughout the ambient isotopy. The strings attached to the disk must remain attached, but are otherwise free to move topologically. (See [10]).

We define $w(G)$ just as we did for the usual writhe - as the sum of the crossing signs. Then $w(G)$ is an invariant of rigid-vertex regular isotopy, and we define

$$P_G = a^{-w(G)}[G],$$

an invariant of rigid-vertex ambient isotopy for 4-valent graphs in three-space. This is the promised three-variable generalization of the Homfly polynomial. Note that for links P_L is exactly the usual Homfly polynomial since $z = A - B$.

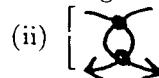
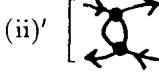
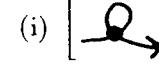
Example. P  $= a^{-1} \left[\begin{array}{c} \text{Diagram with a rigid vertex} \\ \text{and two crossing arrows} \end{array} \right] = \frac{1 - a^{-2} - A^2 a^{-2} + AB a^{-2}}{A - B}$

Since this rational function is not symmetric (invariant) under the interchange of A and B , a and a^{-1} , we conclude that this graph is not rigid-vertex ambient isotopic to its mirror image.

Graphical Calculus

We now turn to the properties of $[G]$ for G a planar graph. Here we shall see that *all computations can be performed recursively in the category of planar graphs.*

Proposition 2.4. Let G be an arbitrary 4-valent graph-diagram. Then the following identities hold:

- (ii)  $= (1 - AB) \left[\begin{array}{c} \downarrow \\ \downarrow \end{array} \right] - (A + B) \left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right]$
- (ii)'  $= \left[\begin{array}{c} \downarrow \\ \uparrow \end{array} \right] + \left[\frac{B^2 a - A^2 a^{-1}}{A - B} \right] \left[\begin{array}{c} \leftarrow \\ \rightarrow \end{array} \right]$
- (i)  $= \left(\frac{A a^{-1} - B a}{A - B} \right) \left[\begin{array}{c} \rightarrow \end{array} \right]$
- (iii)  $- \left[\begin{array}{c} \downarrow \\ \downarrow \\ \downarrow \end{array} \right] = AB \left(\left[\begin{array}{c} \nearrow \\ \searrow \\ \downarrow \end{array} \right] - \left[\begin{array}{c} \downarrow \\ \searrow \\ \downarrow \end{array} \right] \right)$
- (iii')  $- \left[\begin{array}{c} \uparrow \\ \downarrow \\ \downarrow \end{array} \right] = \left(\frac{B^3 a - A^3 a^{-1}}{A - B} \right) \left(\left[\begin{array}{c} \downarrow \\ \uparrow \\ \downarrow \end{array} \right] - \left[\begin{array}{c} \uparrow \\ \downarrow \\ \downarrow \end{array} \right] \right)$

The proof of this proposition is postponed to section 4. Since any 4-valent planar graph can be “undone” by a series of moves of the type - shadow of a Reidemeister move, it follows at once that the identities of 2.4 coupled with the loop identity $[G \text{ } \bullet] = \delta[G]$ will suffice to compute the value of $[G]$ within the category of planar graphs. *This is the graphical calculus.*

Example. 1. $\left[\text{ } \right] = (a^{-1} A - a B) / (A - B)$ (previously computed)

$$\begin{aligned} 2. \left[\text{ } \right] &= \left[\text{ } \right] + \left(\frac{B^2 a - A^2 a^{-1}}{A - B} \right) \left[\text{ } \right] \\ &= \frac{a - a^{-1} + B^2 a - A^2 a^{-1}}{A - B} \end{aligned}$$

3. $\left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right] = (1 - AB) \left[\begin{array}{c} \text{O} \\ \text{O} \end{array} \right] - (A + B) \left[\begin{array}{c} \text{Q} \\ \text{Q} \end{array} \right]$

$$= \frac{(1 - AB)(a - a^{-1})}{A - B} - \frac{(A + B)(a^{-1}A - aB)}{A - B}$$

$$= \frac{a - ABA - a^{-1} + ABa^{-1} - a^{-1}A^2 + aAB - a^{-1}AB + aBB}{A - B}$$

$$= (a - a^{-1} - a^{-1}A^2 + aB^2)/(A - B)$$

$$\therefore \left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right] = \left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right].$$

4. $\left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right] = (1 - AB) \left[\begin{array}{c} \text{O} \\ \text{O} \end{array} \right] - (A + B) \left[\begin{array}{c} \text{Q} \\ \text{Q} \end{array} \right]$

$$= \frac{(1 - AB)(a^{-1}A - aB) - (A + B)(a - a^{-1} - a^{-1}A^2 + aB^2)}{A - B}$$

$$= (a^{-1}A - aB - a^{-1}A^2B + aAB^2 - aA + a^{-1}A + a^{-1}A^3$$

$$- aAB^2 - aB + a^{-1}B + a^{-1}A^2B - aB^3)/(A - B)$$

$$= \frac{2(a^{-1}A - aB) + (a^{-1}B - aA) + (a^{-1}A^3 - aB^3)}{A - B}.$$

5. Here is an example of a calculation using the state-expansion via graphical calculus:

$$\left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right] = B \left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right] + \left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right]$$

$$= B \left\{ B \left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right] + \left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right] \right\}$$

$$+ \left\{ B \left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right] + B \left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right] \right\}$$

$$\left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right] = B^2 \left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right] + 2B \left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right] + \left[\begin{array}{c} \text{S} \\ \text{S} \end{array} \right].$$

Thus

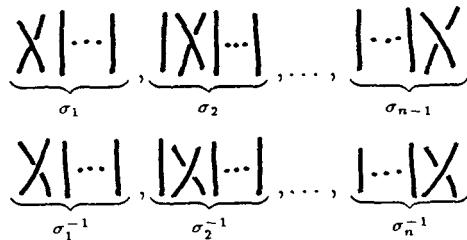
$$\begin{aligned}
 \left[\text{Diagram} \right] &= a^{-1}AB^2 - aB^3 \\
 &\quad + 2aB - 2a^{-1}B - 2a^{-1}A^2B + 2aB^3 \\
 &\quad + \left(\frac{2(a^{-1}A - AB) + (a^{-1}B - aA) + (a^{-1}A^3 - aB^3)}{A - B} \right) \\
 &= \frac{\left\{ \begin{array}{l} 2aB - 2a^{-1}B - 2a^{-1}A^2B + a^{-1}AB^2 + aB^3 \\ + 2(a^{-1}A - aB) + (a^{-1}B - aA) + (a^{-1}A^3 - aB^3) \end{array} \right\}}{(A - B)}
 \end{aligned}$$

It follows at once from $P_L = a^{-w(L)}[L] = a^2[L]$ that L is not rigid-vertex ambient isotopic to its mirror image.

Remark. These results about rigid vertex isotopy can be obtained more cheaply as in [10]. We bring them up here to illustrate the state expansion into values of planar graphs, and the corresponding applications of the graphical calculus. More intensive computations and tables will be given in a sequel to this paper.

3. Braids and the Hecke Algebra

Recall that the Artin Braid Group on n -strands, denoted B_n , is generated by elements $\sigma_1, \sigma_2, \dots, \sigma_{n-1}$ and their inverses, as shown in Figure 3.



Generators of the n -strand Braid Group
Figure 3

Braids are multiplied by attaching the bottom strands of the left element to the top strands of the right element (when reading a product bb' from left to right, b is the left element.). The *closure* \bar{b} of a braid b is the link obtained by attaching top strands to corresponding bottom strands, as shown in Figure 4.

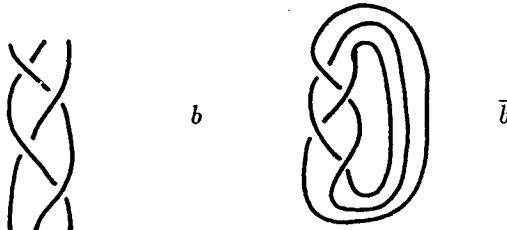


Figure 4

Upon closing a braid b to form the link \bar{b} , we orient \bar{b} in a standard way by assuming that each braid strand is oriented from top to bottom. For this reason we can dispense with indications of orientations in our discussion of braids.

Within the braid group B_n , two braids are equivalent if and only if they are related by a series of moves of the types:

1. ~ $\Leftrightarrow \sigma_i \sigma_i^{-1} = \sigma_i^{-1} \sigma_i = 1$
2. ~ $\Leftrightarrow \sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}$ (same for inverses)

These special versions of the type II and type III moves suffice for braid equivalence, giving generators and relations for the braid group.

In extending knots and links to graphs and the oriented graphical calculus it is natural to extend braids by adding the graphical vertex elements C_1, C_2, \dots, C_{n-1} as shown below:

$$\underbrace{X| \cdots |}_{C_1}, \underbrace{| X | \cdots |}_{C_2}, \dots, \underbrace{|| \cdots | X}_{C_{n-1}}.$$

Since

$$C_1^2 \leftrightarrow \begin{array}{c} \text{Diagram of } C_1 \\ | \\ \dots \end{array}$$

we have no license to make the extended structure into a group. Hence this becomes a *nodal braid monoid* (to distinguish it from the braid monoid of [1] and [5]). (Remark: The monoid structure will be extended even further in section 5 of this paper.)

Should the C_i 's satisfy further relations? *The graphical calculus provides a strong clue.* For note that in the case of braids the recursive formulas in the graphical calculus take the form: (See Proposition 2.4)

- (ii) $\left[\begin{array}{c} \text{Diagram of } C_1 \\ | \end{array} \right] = (1 - AB) \left[\begin{array}{c} \text{Diagram of } C_1 \\ | \end{array} \right] - (A + B) \left[\begin{array}{c} \text{Diagram of } C_1 \\ | \end{array} \right]$
- (i) $\left[\begin{array}{c} \text{Diagram of } C_1 \\ | \end{array} \right] = \left(\frac{Aa^{-1} - Ba}{A - B} \right) \left[\begin{array}{c} \text{Diagram of } C_1 \\ | \end{array} \right]$
- (iii) $\left[\begin{array}{c} \text{Diagram of } C_1 \\ | \end{array} \right] - \left[\begin{array}{c} \text{Diagram of } C_1 \\ | \end{array} \right] = AB \left(\left[\begin{array}{c} \text{Diagram of } C_1 \\ | \end{array} \right] - \left[\begin{array}{c} \text{Diagram of } C_1 \\ | \end{array} \right] \right)$

These three rules are sufficient to compute the value of $[\bar{b}]$ for the closure of any braid-graph (product of C_i 's) inductively. In B_n it suffices to compute explicit values for any $n!$ representations of permutations as products of the C_i 's.

For example, in B_2 we need

$$\left[\begin{array}{c} \text{Diagram of } C_1 \\ | \end{array} \right] = \delta = \frac{a - a^{-1}}{A - B}$$

and

$$\left[\begin{array}{c} \text{Diagram of } C_1 \\ | \end{array} \right] = \left(\frac{Aa^{-1} - Ba}{A - B} \right) \stackrel{\text{def}}{=} \epsilon.$$

In general, define $[b]$ for any b in the *nodal braid monoid* NB_n by the formula

$$[b] = [\bar{b}]$$

where \bar{b} is the closure of b .

Computing in NB_3 is entirely facilitated by the values for

$$[\text{|||}], [\text{X|}], [|\text{X}|], [\text{XX}], [\text{X}\text{X}], [\text{X}\text{X}\text{X}].$$

Then

$$\begin{aligned} [\text{|||}] &= \delta^2, [\text{X|}] = \delta\epsilon = [|\text{X}|] \\ [\text{XX}] &= [\text{O}\text{O}] = \epsilon^2 = [\text{X}\text{X}] \\ [\text{X}\text{X}] &= [\text{O}\text{O}] = \epsilon [\text{O}] \\ &= \epsilon \left((1 - AB) [\text{O}] - (A + B) [\text{O}] \right) \\ &= \epsilon((a - AB)\delta - (A + B)\epsilon). \end{aligned}$$

These specific values on representations of permutations then provide the basis for the computation of any product of C_i 's and thence, by our relations to any braid or mixture of braids and nodal elements.

Note that via our interpretations in terms of rigid vertex isotopies all of these graphical evaluations have specific topological meaning (i.e. invariance).

Thus the graphical calculus suggests that we adopt the following relations on the C_i :

$$(ii) \Leftrightarrow (ii)^* \quad C_i^2 = (a - AB) - (A + B)C_i$$

$$(iii) \Leftrightarrow (iii)^* \quad C_i C_{i+1} C_i - C_{i+1} C_i C_{i+1} = AB(C_i - C_{i+1})$$

Since these relations are satisfied by the graphical evaluation we see that $[]$ gives a well-defined mapping on the resulting quotient structure. Here we are working in the free algebra over $\mathbb{Z}[a, a^{-1}, A, A^{-1}, B, B^{-1}, (A - B)^{\pm 1}]$ with generators $1, \sigma_1, \dots, \sigma_{n-1}, \sigma_a^{-1}, \dots, \sigma_{n-1}^{-1}, C_1, \dots, C_{n-1}$. We already have the multiplicative braiding relations and we shall also add (ii)*, (iii)* above plus mixed relations as follows:

$$\begin{aligned} 1^0. \quad \sigma_i^{-1} &= A + C_i \Leftrightarrow [\text{X}\text{X}] = A[\text{V}\text{V}] + [\text{X}\text{X}] \\ \sigma_i &= B + C_i \Leftrightarrow [\text{X}\text{X}] = B[\text{V}\text{V}] + [\text{X}\text{X}] \end{aligned}$$

$$(\text{Note: } \sigma_i^{-1} - \sigma_i = (A - B) = z)$$

$$2^0. \sigma_i C_{i+1} \sigma_i^{-1} = \sigma_{i+1}^{-1} C_i \sigma_{i+1}$$

$$\leftrightarrow \quad \begin{array}{c} \diagup \\ \diagdown \end{array} \sim \quad \begin{array}{c} \diagdown \\ \diagup \end{array}$$

It may be desirable to omit the first relation 1^0 . using it only as a definition of $[]$ on braids. If 1^0 is included, then the braiding relations and 2^0 . are consequences. In fact we see that the algebra of the C_i 's alone is a version of the Hecke algebra. This is seen most clearly by taking $AB = 1$. Then

$$C_i^2 = -(A + B)C_i$$

and

$$C_i C_{i+1} C_i - C_i = C_{i+1} C_i C_{i+1} - C_{i+1}.$$

The relation $[\sigma_i] = B[1] + [C_i]$ then corresponds to a representation $\rho : B_n \rightarrow \mathcal{C}_n$ where \mathcal{C}_n is this algebra of the C_i 's.

$$\begin{aligned}\rho(\sigma_i^{-1}) &= A + C_i \\ \rho(\sigma_i) &= B + C_i.\end{aligned}$$

Thus we have given an extension of the original Hecke algebra in terms of the relations in the graphical calculus.

In this sense the full graphical calculus for planar graphs may be regarded as a non-trivial generalization of the Hecke algebra, and its combinatorial context.

4. Demonstration of Identities in Oriented Graphical Calculus

First recall the set up. We have

$$\begin{aligned}z &= A - B \\ \delta &= (a - a^{-1})/(A - B).\end{aligned}$$

The basic identities for $[K]$ are:

$$\begin{aligned}\left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right] &= A \left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right] + \left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right] \\ \left[\begin{array}{c} \nearrow \\ \swarrow \end{array} \right] &= B \left[\begin{array}{c} \nearrow \\ \swarrow \end{array} \right] + \left[\begin{array}{c} \nearrow \\ \swarrow \end{array} \right].\end{aligned}$$

$$1^0. \quad \left[\begin{array}{c} \nearrow \\ \text{loop} \\ \searrow \end{array} \right] = (1 - AB) \left[\begin{array}{c} \nearrow \\ \searrow \end{array} \right] - (A + B) \left[\begin{array}{c} \nearrow \\ \swarrow \end{array} \right].$$

Dem.

$$\begin{aligned}
 \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right] &= \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right] - B \left[\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right] \\
 &= \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right] - A \left[\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right] - B \left[\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right] \\
 &= \left[\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right] - A \left(B \left[\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right] + \left[\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right] \right) \\
 &\quad - B \left[\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right] \\
 &= (1 - AB) \left[\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right] - (A + B) \left[\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right] \quad //
 \end{aligned}$$

$$2^0. \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right] = \epsilon \left[\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right] \text{ with } \epsilon = \frac{Aa^{-1} - Ba}{A - B}.$$

Dem.

$$\begin{aligned}
 \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right] &= \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right] - B \left[\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right] \\
 &= (a^{-1} - B\delta) \left[\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right]. \\
 a^{-1} - B\delta &= a^{-1} - B \left(\frac{a - a^{-1}}{A - B} \right) \\
 &= ((A - B)a^{-1} - B(a - a^{-1})) / (A - B) \\
 &= \left(\frac{Aa^{-1} - Ba}{A - B} \right) \quad //
 \end{aligned}$$

$$3^0. \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right] = \left[\begin{array}{c} \downarrow \uparrow \end{array} \right] + \left(\frac{B^2a - A^2a^{-1}}{A - B} \right) \left[\begin{array}{c} \curvearrowright \\ \curvearrowright \end{array} \right]$$

Dem.

$$\begin{aligned}
 \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] &= \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] - A \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right] \\
 &= \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] - B \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right] - A \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right] \\
 &= \left[\begin{array}{c} \downarrow \uparrow \\ \downarrow \uparrow \end{array} \right] - Ba \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right] - A\epsilon \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right]. \\
 -Ba - A\epsilon &= -Ba - \frac{A(Aa^{-1} - Ba)}{A - B} \\
 &= (-Ba(A - B) - A^2a^{-1} + ABa)/(A - B) \\
 &= (B^2a - A^2a^{-1})/(A - B) \quad //
 \end{aligned}$$

$$4^0. \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] - \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] = AB \left(\left[\begin{array}{c} \curvearrowleft \\ \curvearrowleft \end{array} \right] - \left[\begin{array}{c} \downarrow \curvearrowleft \\ \downarrow \curvearrowright \end{array} \right] \right)$$

Dem.

$$\begin{aligned}
 \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] &= \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] - A \left[\begin{array}{c} \curvearrowleft \\ \curvearrowleft \end{array} \right] \\
 &= \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] - B \left[\begin{array}{c} \curvearrowleft \\ \curvearrowleft \end{array} \right] - A \left[\begin{array}{c} \curvearrowleft \\ \curvearrowleft \end{array} \right] \\
 &= \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] - B \left(A \left[\begin{array}{c} \downarrow \curvearrowleft \\ \downarrow \curvearrowright \end{array} \right] + \left[\begin{array}{c} \curvearrowleft \\ \curvearrowleft \end{array} \right] \right) - A \left[\begin{array}{c} \curvearrowleft \\ \curvearrowleft \end{array} \right] \\
 \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] &= \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] - B \left[\begin{array}{c} \curvearrowleft \\ \curvearrowleft \end{array} \right] \\
 &= \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] - A \left[\begin{array}{c} \curvearrowleft \\ \curvearrowleft \end{array} \right] - B \left[\begin{array}{c} \curvearrowleft \\ \curvearrowleft \end{array} \right] \\
 &= \left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] - A \left(B \left[\begin{array}{c} \downarrow \curvearrowleft \\ \downarrow \curvearrowright \end{array} \right] + \left[\begin{array}{c} \curvearrowleft \\ \curvearrowleft \end{array} \right] \right) - B \left[\begin{array}{c} \curvearrowleft \\ \curvearrowleft \end{array} \right].
 \end{aligned}$$

The identity now follows from the fact that

$$\left[\begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \right] = \left[\begin{array}{c} \curvearrowleft \\ \curvearrowright \end{array} \right] \quad //$$

$$5^0. \quad \left[\begin{array}{cc} \nearrow & \searrow \\ \swarrow & \downarrow \end{array} \right] - \left[\begin{array}{cc} \nearrow & \searrow \\ \swarrow & \downarrow \end{array} \right] = \mu \left(\left[\begin{array}{cc} \downarrow & \curvearrowleft \\ \curvearrowright & \downarrow \end{array} \right] - \left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right] \right)$$

$$\mu = (B^3 a - A^3 a^{-1})/(A - B).$$

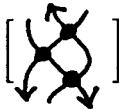
Dem.

$$\begin{aligned}
\left[\begin{array}{cc} \nearrow & \searrow \\ \swarrow & \downarrow \end{array} \right] &= \left[\begin{array}{cc} \nearrow & \searrow \\ \swarrow & \downarrow \end{array} \right] - B \left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right] \\
&= \left[\begin{array}{cc} \nearrow & \searrow \\ \swarrow & \downarrow \end{array} \right] - A \left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right] - B \left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right] \\
&= \left[\begin{array}{cc} \nearrow & \searrow \\ \swarrow & \downarrow \end{array} \right] - A \left(B \left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right] + \left[\begin{array}{cc} \nearrow & \searrow \\ \swarrow & \downarrow \end{array} \right] \right) \\
&\quad - B \left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right] \\
&= \left[\begin{array}{cc} \nearrow & \searrow \\ \swarrow & \downarrow \end{array} \right] - AB\epsilon \left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right] \\
&\quad - A \left(\left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right] + \eta \left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right] \right) \\
&\quad - B \left(\left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right] + \eta \left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right] \right) \text{ (by } 3^0.) \\
&= \left[\begin{array}{cc} \nearrow & \searrow \\ \swarrow & \downarrow \end{array} \right] - (AB\epsilon + A\eta + B\eta) \left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right] \\
&\quad - A \left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right] - B \left[\begin{array}{cc} \curvearrowleft & \downarrow \\ \curvearrowright & \downarrow \end{array} \right].
\end{aligned}$$

Here $\eta = (B^2 a - A^2 a^{-1})/(A - B)$ (from 3^0). Thus

$$\begin{aligned}
&AB\epsilon + A\eta + B\eta \\
&= AB \left(\frac{Aa^{-1} - BA}{A - B} \right) + (A + B) \left(\frac{B^2 a - A^2 a^{-1}}{A - B} \right) \\
&= \frac{A^2 Ba^{-1} - AB^2 a + AB^2 a - A^3 a^{-1} + B^3 a - A^2 Ba^{-1}}{A - B} \\
&= (B^3 a - A^3 a^{-1})/(A - B) \\
\therefore AB\epsilon + A\eta + B\eta &= \mu.
\end{aligned}$$

The identity now follows by expanding



and subtracting the two formulas. //

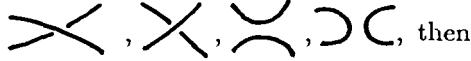
This completes the verification of the basic identities in the oriented graphical calculus.

5. The Dubrovnik Polynomial

We shall work with the Dubrovnik version of the Kauffman polynomial. This is based on a regular isotopy invariant for *unoriented* knots and links that we denote by D_L . This polynomial satisfies the axioms:

Dubrovnik Polynomial Axioms

- To each unoriented knot or link diagram L is associated a Laurent polynomial $D_L = D_L(a, z)$ in the variables a, z . If L is regularly isotopic to L' , then $D_L = D_{L'}$.
- If four links differ only at the sites of the four small diagrams



$$D \text{ } \cancel{\times} - D \cancel{\times} = z(D \cancel{\times} - D \cancel{\circ} \text{ } \cancel{\circ}).$$

$$3. D \text{ } \circ = 1$$

$$4. D \text{ } \cancel{\circ} = a D \text{ } \cancel{\circ} \quad D \text{ } \cancel{\circ} \text{ } \cancel{\circ} = a^{-1} D \text{ } \cancel{\circ} .$$

There is a corresponding *ambient isotopy* invariant to D_L for oriented L given by the formula

$$Y_L = a^{-w(L)} D_L.$$

We shall refer to Y_L as the *oriented* Dubrovnik polynomial.

Remark. The Kauffman polynomial refers to a regular isotopy invariant L_K for unoriented links K (and its companion $F_K = a^{-w(K)}L_K$ for K oriented). L_K satisfies the recursion

$$\begin{aligned} L \times + L \times &= z(L \equiv + L \circ \bullet) \\ L \circ &= 1 \\ L \overline{\circ} &= aL \\ L \overline{\bullet} &= a^{-1}L \end{aligned}$$

The L and D polynomials are interrelated by the following change of variables (an observation of W.B.R. Lickorish):

$$D_K(a, z) = -i^{-w(K)}(-1)^{c(K)}L_K(ia, -iz)$$

where $i = \sqrt{-1}$ and $c(K)$ is the number of components of K .

For example, we have

$$\begin{aligned} L \circ \circ + L \circ \circ &= z(L \circ \circ + L \bullet \bullet) \\ \Rightarrow a + a^{-1} &= z(L \circ \circ + 1) \\ \Rightarrow L \circ \circ &= \left(\frac{a + a^{-1}}{z}\right) - 1 \end{aligned}$$

while $D \circ \circ = \frac{a - a^{-1}}{z} + 1$.

In the calculations to follow we shall use

$$\mu = D \circ \circ = \frac{a - a^{-1}}{z} + 1.$$

Note that $D_{KO} = \mu D_K$, is the formula for adding a disjoint component.

Everything that we have done for the extension of the Homfly polynomial to a 3-variable polynomial for rigid-vertex graphs has its counterpart with the Dubrovnik polynomial. We now sketch how this is done.

In the extension we use variables A, B, a with $z = A - B$. Then D is extended to 4-valent nodal diagrams via

$$D \times = D \times - AD \equiv - BD \circ \bullet$$

$$D \times = D \times - BD \equiv - AD \circ \bullet.$$

The technical summation showing that this extension is well-defined proceeds as in Lemma 2.2. The definition is as follows:

- (1) Choose a *marker*  or  at each graphical vertex.
(2) Let \mathcal{L} be the set of links obtained by replacing:

$$\begin{array}{ccc} \text{---} & \mapsto & \text{---} \text{---} \\ \text{---} & \mapsto & \text{---} \text{---} \end{array} \text{ or } \begin{array}{c} \text{---} \\ \text{---} \end{array}.$$

3. Let D denote the standard regular isotopy Dubrovnik satisfying

$$D \text{---} D \text{---} = z(D \text{---} - D \text{---})$$

on unoriented *link diagrams*. Let $z = A - B$.

4. Define

$$[G] = \sum_{L \in \mathcal{L}} (-A)^{i(L)} (-B)^{j(L)} D_L$$

where $i(L)$ is the number of replacements of type  \mapsto 

and $j(L)$ is the number of replacements of type  \mapsto 

Use (3) to verify that this definition is independent of the marker choice.

We then rewrite this definition as an expansion for D :

$$\begin{aligned} D \text{---} &= AD \text{---} + BD \text{---} + D \text{---} \\ D \text{---} &= BD \text{---} + AD \text{---} + D \text{---} . \end{aligned}$$

We shall write $[K] = D_K$. There should be no confusion here, since the diagrams inside a bracket for the Dubrovnik polynomial are unoriented. Thus

$$\begin{aligned} \mu &= \left[\text{---} \text{---} \right] = \left(\frac{a - a^{-1}}{A - B} \right) + 1. \\ \mathcal{O} &= \left[\text{---} \text{---} \right] = \left[\text{---} \text{---} \right] - A \left[\text{---} \text{---} \right] - B \left[\text{---} \text{---} \right] \\ &= a^{-1} - A - B\mu \\ &= a^{-1} - A - B \left(\frac{a - a^{-1}}{A - B} + 1 \right) \\ \therefore \mathcal{O} &= a^{-1} - \frac{B(a - a^{-1})}{A - B} - (A + B) \\ &= \frac{a^{-1}A - a^{-1}B - aB + a^{-1}B}{A - B} - (A + B) \\ \boxed{\mathcal{O} = \frac{a^{-1}A - aB}{A - B} - (A + B)} \end{aligned}$$

Once again, planar graphs receive values that are invariant under simultaneous interchange of $A \leftrightarrow B$, $a \leftrightarrow a^{-1}$.

It is easy to verify that $[L]$ is an invariant of rigid-vertex regular isotopy, and that for oriented graphs G (with $w(G) = \sum \pm 1 = \sum_p \text{sign}(p)$ where p is a crossing) we can define $Y_G = a^{-w(G)}[G]$ to obtain an invariant of rigid vertex ambient isotopy. ($[G]$ forgets the orientation on G .)

Example.

$$\begin{aligned} [\textcirclearrowleft] &= [\textcirclearrowleft] - A[\textcirclearrowright] - B[\textcirclearrowright] \\ &= [\textcirclearrowright\textcirclearrowleft] - Aa - Ba^{-1} \\ &= (\delta + z(a - a^{-1})) - Aa - Ba^{-1}. \end{aligned}$$

The basic result in the *unoriented graphical calculus* is

Theorem 5.1. The following identities hold, giving a calculus for unoriented planar 4-valent graphs:

$$(a) \quad [\text{Y}] = \mathcal{O}[\text{)}]$$

$$\begin{aligned} (b) \quad [\text{X}] &= (1 - AB)[\text{)}(\text{)}] + \gamma[\text{XX}] \\ &\quad - (A + B)[\text{X}\text{X}] \end{aligned}$$

$$\begin{aligned} (c) \quad &[\text{X}\text{X}] - [\text{X}\text{X}\text{X}] = \\ &AB([\text{IX}] - [\text{XI}] + [\text{X}\text{X}] + [\text{X}\text{X}]) \\ &\quad - ([\text{X}\text{X}\text{X}] - [\text{X}\text{X}\text{X}]) \\ &\quad + \xi([\text{X}\text{X}\text{I}] - [\text{I}\text{X}\text{X}]) \end{aligned}$$

where

$$\mathcal{O} = (Aa^{-1} - Ba)/(A - B) - (A + B)$$

$$\gamma = (B^2a - A^2a^{-1})/(A - B) + AB$$

$$\xi = (B^3a - A^3a^{-1})/(A - B).$$

Proof of 5.1. See Appendix 1. □

In the case of braids we are working in an *extended braid monoid* with generators $\sigma_1^{\pm 1}, \dots, \sigma_{n-1}^{\pm 1}$ (the usual braiding generators) plus

$$\underbrace{X| \cdots |}_{c_1}, \underbrace{| X \cdots |}_{c_2}, \dots, \underbrace{|\cdots| X}_{c_{n-1}}$$

and

$$\underbrace{\cup| \cdots |}_{h_1}, \underbrace{| \cup \cdots |}_{h_2}, \dots, \underbrace{|\cdots| \cup}_{h_n}$$

Each of the graphical identities of Theorem 5.1 can then be re-written to give:

- (a) $[c_i h_i] = [h_i c_i] = \mathcal{O}[h_i]$
- (b) $[c_i^2] = (1 - AB)[1] + \gamma[h_i] - (A + B)[c_i]$
- (c) $[c_{i+1} c_i c_{i+1}] - [c_i c_{i+1} c_i] =$
 $AB([c_{i+1}] - [c_i] + [h_{i+1} c_i] + [c_i h_{i+1}]$
 $- [h_i c_{i+1}] - [c_{i+1} h_i])$
 $+ \xi([h_i] - [h_{i+1}])$

Thus we can form a free additive algebra with generators c_1, \dots, c_{n-1} , h_1, \dots, h_{n-1} and relations

$$(I) \quad \left\{ \begin{array}{rcl} c_i c_j & = & c_j c_i & |i - j| > 1 \\ h_i h_j & = & h_j h_i & |i - j| > 1 \\ h_i^2 & = & \mu h_i \\ h_i h_{i \pm 1} h_i & = & h_i \\ c_i h_i & = & h_i c_i = \mathcal{O} h_i \\ c_i h_{i+1} & = & c_{i+1} h_i h_{i+1} \\ h_i c_{i+1} & = & h_i h_{i+1} c_i \end{array} \right\}$$

$$(II) \quad \left\{ \begin{array}{l} c_i^2 = (1 - AB) + \gamma h_i - (A + B)c_i \\ c_{i+1}c_i c_{i+1} - c_i c_{i+1} c_i \\ = AB(c_{i+1} - c_i + h_{i+1}c_i + c_i h_{i+1}) \\ - h_i c_{i+1} - c_{i+1} h_i \\ + \xi(h_i - h_{i+1}) \end{array} \right\}$$

The relations in group I describe the underlying connection monoid. Group II constitutes the deformation from

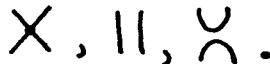
$$\{c_i^2 = 1, c_{i+1}c_i c_{i+1} = c_i c_{i+1} c_i\}.$$

Call this algebra \mathcal{W}_n .

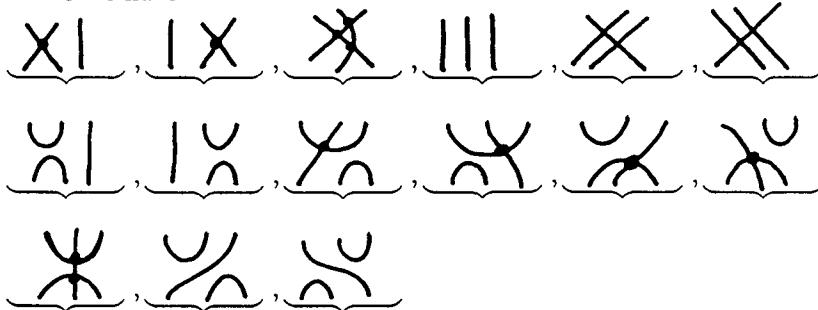
Our results imply that \mathcal{W}_n is isomorphic to the Birman-Wenzl algebra [1] associated to the L and D polynomials. The existence of the un-oriented graphical calculus proves the existence of a “trace” []: $\mathcal{W}_N \rightarrow \mathbb{Z}[A^\pm, B^\pm, (A_B)^\pm, a^\pm]$ and gives explicit rules for calculating this trace.

It is sufficient to know the values $[G]$ for the closures of elements of the Brauer monoid [2] or *full connection monoid on n points*. In other words, if we know $[G]$ on a set of graphs that include all connections top-top, top-bottom, bottom-bottom, between two rows of n -points, then the graphical identities suffice to calculate for all remaining diagrams of elements of \mathcal{W}_n .

For example, if $n = 2$ then the Brauer monoid is generated by



For $n = 3$ we have



These 15 elements of the Brauer monoid BR_3 represent all possible pairwise connections among 6 points. Up to free loops, the product of any two is a third. Thus

$$\begin{array}{c} \text{Diagram A} \\ \text{Diagram B} \end{array} \rightleftharpoons \begin{array}{c} \text{Diagram C} \end{array}$$

In BR_n itself we assume that

$$\text{Diagram A} \equiv \text{Diagram B}, \quad \text{Diagram B} \equiv \text{Diagram C}, \quad \text{Diagram C} \equiv \text{Diagram A}$$

so that two elements are equivalent if and only if they have the same connection structure (i.e. the same set of paired points).

Thus the connection monoid BR_n is described by the relations:

$$c_i c_j = c_j c_i \quad |i - j| > 1$$

$$\begin{array}{c} \text{Diagram A} \\ \text{Diagram B} \end{array}$$

$$h_i h_j = h_j h_i \quad |i - j| > 1$$

$$\begin{array}{c} \text{Diagram A} \\ \text{Diagram B} \\ \text{Diagram C} \\ \text{Diagram D} \\ \text{Diagram E} \\ \text{Diagram F} \end{array}$$

$$h_i^2 = \mu h_i$$

$$\leftarrow \mu$$

$$h_i h_{i \pm 1} h_i = h_i$$

$$\sim \begin{array}{c} \text{Diagram A} \\ \text{Diagram B} \end{array}$$

$$c_i h_i = h_i c_i = \mathcal{O} h_i$$

$$= \mathcal{O} \begin{array}{c} \text{Diagram A} \\ \text{Diagram B} \end{array}$$

$$c_i h_{i+1} = c_{i+1} h_i h_{i+1}$$

$$\sim \begin{array}{c} \text{Diagram A} \\ \text{Diagram B} \end{array}$$

$$h_i c_{i+1} = h_i h_{i+1} c_i$$

$$\sim \begin{array}{c} \text{Diagram A} \\ \text{Diagram B} \end{array}$$

$$c_i^2 = 1$$

$$\sim \begin{array}{c} \text{Diagram A} \\ \text{Diagram B} \end{array}$$

$$c_i c_{i+1} c_i = c_{i+1} c_i c_{i+1}$$

$$\sim \begin{array}{c} \text{Diagram A} \\ \text{Diagram B} \end{array}$$

Our version of the Birman-Wenzl Algebra deforms these last two relations.

The algebra \mathcal{W}_n is to BR_n as the Hecke algebra H_n is to the symmetric group S_n . In both cases (Hecke algebra and Birman-Wenzl algebra) we

have shown how a graphical calculus provides the appropriate trace. This interpretation gives a direct geometric meaning to the algebra trace - for $[b]$ is a rigid-vertex invariant of the corresponding graph.

The braid group is represented to \mathcal{W}_n via $\rho : B_n \rightarrow \mathcal{W}_n$

$$\begin{aligned}\rho(\sigma_i) &= Ah_i + B + c_i \\ \rho(\sigma_i^{-1}) &= Bh_i + A + c_i\end{aligned}$$

whence $\rho(\sigma_i) - \rho(\sigma_i^{-1}) = z(h_i - 1)$ where $z = A - B$.

Appendix 1. Unoriented Graphical Identities

Recall the set up:

$$\mu = \left[\textcircled{O} \textcircled{O} \right]$$

$$\boxed{\mu = \frac{a - a^{-1}}{A - B} + 1}.$$

$$\begin{aligned}\left[\begin{array}{c} \diagup \diagdown \\ \diagdown \diagup \end{array} \right] &= A \left[\begin{array}{c} \diagup \diagup \\ \diagup \diagup \end{array} \right] + B \left[\begin{array}{c} \curvearrowleft \curvearrowright \\ \curvearrowright \curvearrowleft \end{array} \right] + \left[\begin{array}{c} \times \\ \times \end{array} \right] \\ \left[\begin{array}{c} \diagdown \diagup \\ \diagup \diagdown \end{array} \right] &= B \left[\begin{array}{c} \diagup \diagup \\ \diagup \diagup \end{array} \right] + A \left[\begin{array}{c} \curvearrowleft \curvearrowright \\ \curvearrowright \curvearrowleft \end{array} \right] + \left[\begin{array}{c} \times \\ \times \end{array} \right]\end{aligned}$$

$$\begin{aligned}\mathcal{O} &= \left[\textcircled{O} \textcircled{\infty} \right] = \left[\textcircled{\infty} \right] - A \left[\textcircled{O} \textcircled{\text{wavy}} \right] - B \left[\textcircled{O} \textcircled{O} \right] \\ &\quad = a^{-1} - A - B\mu \\ \therefore \mathcal{O} &= (a^{-1} - aB)/(A - B) - (A + B).\end{aligned}$$

$$1^0. \quad \left[\begin{array}{c} \text{X} \\ \text{O} \end{array} \right] = \mathcal{O} \left[\begin{array}{c} \text{O} \\ \text{O} \end{array} \right]$$

Dem.

$$\begin{aligned} \left[\begin{array}{c} \text{X} \\ \text{O} \end{array} \right] &= \left[\begin{array}{c} \text{X} \\ \text{O} \end{array} \right] - A \left[\begin{array}{c} \text{O} \\ \text{O} \end{array} \right] - B \left[\begin{array}{c} \text{O} \\ \text{O} \end{array} \right] \\ &= (a^{-1} - A - B\mu) \left[\begin{array}{c} \text{O} \\ \text{O} \end{array} \right] \end{aligned}$$

$$\therefore \left[\begin{array}{c} \text{X} \\ \text{O} \end{array} \right] = \mu \left[\begin{array}{c} \text{O} \\ \text{O} \end{array} \right]. \quad //$$

$$2^0. \quad \left[\begin{array}{c} \text{X} \\ \text{O} \end{array} \right] = (1 - AB) \left[\begin{array}{c} \text{O} \\ \text{O} \end{array} \right] + \gamma \left[\begin{array}{c} \text{O} \\ \text{O} \end{array} \right] - (A + B) \left[\begin{array}{c} \text{X} \\ \text{X} \end{array} \right]$$

where $\gamma = (B^2a - A^2a^{-1})/(A - B) + AB$

Dem.

$$\begin{aligned} \left[\begin{array}{c} \text{X} \\ \text{O} \end{array} \right] &= \left[\begin{array}{c} \text{X} \\ \text{O} \end{array} \right] - A \left[\begin{array}{c} \text{O} \\ \text{O} \end{array} \right] - B \left[\begin{array}{c} \text{X} \\ \text{X} \end{array} \right] \\ &= \left[\begin{array}{c} \text{X} \\ \text{O} \end{array} \right] - A \left[\begin{array}{c} \text{X} \\ \text{X} \end{array} \right] - B \left[\begin{array}{c} \text{X} \\ \text{O} \end{array} \right] \\ &\quad - A\mathcal{O} \left[\begin{array}{c} \text{X} \\ \text{O} \end{array} \right] - B \left[\begin{array}{c} \text{X} \\ \text{X} \end{array} \right] \\ &= \left[\begin{array}{c} \text{O} \\ \text{O} \end{array} \right] - (Ba + A\mathcal{O}) \left[\begin{array}{c} \text{X} \\ \text{O} \end{array} \right] - B \left[\begin{array}{c} \text{X} \\ \text{X} \end{array} \right] \\ &\quad - A(A \left[\begin{array}{c} \text{X} \\ \text{O} \end{array} \right] + B \left[\begin{array}{c} \text{O} \\ \text{O} \end{array} \right] + \left[\begin{array}{c} \text{X} \\ \text{X} \end{array} \right]) \\ &= (1 - AB) \left[\begin{array}{c} \text{O} \\ \text{O} \end{array} \right] - (A^2 + Ba + A\mathcal{O}) \left[\begin{array}{c} \text{X} \\ \text{O} \end{array} \right] \\ &\quad - (A + B) \left[\begin{array}{c} \text{X} \\ \text{X} \end{array} \right]. \end{aligned}$$

$$\begin{aligned}
A^2 + Ba + A\mathcal{O} &= A^2 + Ba + A \left(\frac{a^{-1}A - aB}{A - B} - (A + B) \right) \\
&= \frac{(A^2 + Ba)(A - B) + a^{-1}A^2 - aAB - A(A^2 - B^2)}{A - B} \\
&= \left(\begin{array}{c} A^3 + ABa - A^2B - B^2a + a^{-1}A^2 \\ -A^3 - aAB + AB^2 \end{array} \right) / (A - B) \\
&= (AB(-A + B) + a^{-1}A^2 - aB^2) / (A - B) \\
&= \left(\frac{a^{-1}A^2 - aB^2}{A - B} \right) - AB.
\end{aligned}$$

$3^0.$

$$\begin{aligned}
[\text{X}] - [\text{X}] &= \xi([\text{U}] - [\text{U}]) \\
&\quad + AB([\text{X}] - [\text{X}]) \\
&\quad + [\text{X}] + [\text{X}] - [\text{X}] - [\text{X}] \\
&\quad \xi(B^3a - A^3a^{-1}) / (A - B)
\end{aligned}$$

Dem.

$$\begin{aligned}
 [\text{X} \text{X}] &= [\text{X} \text{X}] - A[\text{X} \text{X}] - B[\text{X} \text{X}] \\
 &= [\text{X} \text{X}] - A[\text{X} \text{X}] - B[\text{X} \text{X}] \\
 &\quad - A[\text{X} \text{X}] - B[\text{X} \text{X}] \\
 &= [\text{X} \text{X}] - A(A[\text{X} \text{X}] + B[\text{X} \text{X}] + [\text{X} \text{X}]) \\
 &\quad - B(A[\text{X} \text{X}] + B[\text{X} \text{X}] + [\text{X} \text{X}]) \\
 &\quad - A[\text{X} \text{X}] - B[\text{X} \text{X}].
 \end{aligned}$$

Hence

$$\begin{aligned}
 [\text{X} \text{X}] &= [\text{X} \text{X}] - B(B[\text{X} \text{X}] + A[\text{X} \text{X}] + [\text{X} \text{X}]) \\
 &\quad - A(B[\text{X} \text{X}] + A[\text{X} \text{X}] + [\text{X} \text{X}]) \\
 &\quad - B[\text{X} \text{X}] - A[\text{X} \text{X}].
 \end{aligned}$$

Thus

$$\begin{aligned}
 & [\text{X} \text{X}] - [\text{X} \text{X}] = A^2 ([\text{X} \text{X}] - [\text{X} \text{X}]) \\
 & + AB ([|\text{X}|] - [\text{X}|]) \\
 & + AB\gamma ([\text{X}|] - [|\text{X}|]) \\
 & + B^2 ([\text{X} \text{X}] - [\text{X} \text{X}]) \\
 & + A ([\text{X} \text{X}] - [\text{X} \text{X}]) \\
 & + B ([\text{X} \text{X}] - [\text{X} \text{X}]).
 \end{aligned}$$

Now apply 2^o to the last two pairs of terms

$$\begin{aligned}
 [\text{X} \text{X}] &= (1 - AB) [\text{X} \text{X}] + \gamma [\text{X}|] - (A + B) [\text{X} \text{X}] \\
 [\text{X} \text{X}] &= (1 - AB) [\text{X} \text{X}] + \gamma [|\text{X}|] - (A + B) [\text{X} \text{X}] \\
 [\text{X} \text{X}] &= (1 - AB) [\text{X} \text{X}] + \gamma [\text{X}|] - (A + B) [\text{X} \text{X}] \\
 [\text{X} \text{X}] &= (1 - AB) [\text{X} \text{X}] + \gamma [|\text{X}|] - (A + B) [\text{X} \text{X}]
 \end{aligned}$$

$$\begin{aligned}
 \therefore [\text{X} \text{X}] - [\text{X} \text{X}] &= \gamma ([\text{X}|] - [|\text{X}|]) + (A + B) ([\text{X} \text{X}] - [\text{X} \text{X}]) \\
 [\text{X} \text{X}] - [\text{X} \text{X}] &= \gamma ([\text{X}|] - [|\text{X}|]) + (A + B) ([\text{X} \text{X}] - [\text{X} \text{X}])
 \end{aligned}$$

Hence

$$[\text{X}] - [\text{X}]$$

$$= AB \left([|\text{X}|] - [\text{X}|] + [\text{X}] - [|\text{X}|] + [\text{X}] - [|\text{X}|] \right) \\ + (AB\mathcal{O} + (A+B)\gamma)([|\text{X}|] - [|\text{X}|])$$

$$AB\mathcal{O} + (A+B)\gamma = \frac{AB(Aa^{-1} - Ba)}{A - B} - AB(A + B)$$

$$+ \frac{(A + B)(B^2a - A^2a^{-1})}{A - B} + (A + B)AB$$

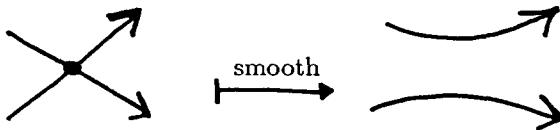
$$= \frac{A^2Ba^{-1} - AB^2a + AB^2a - A^3a^{-1} + B^3a - A^2Ba^{-1}}{A - b}$$

$$= (B^3a - A^3a^{-1})/(A - B).$$

This completes the demonstration of the third identity.

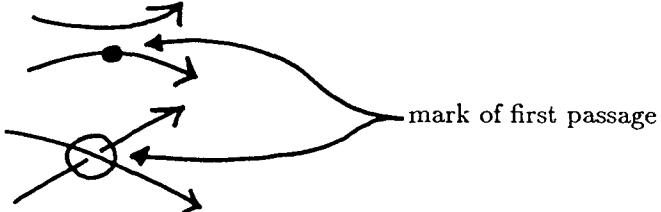
Appendix 2. Skein Models and Graph Polynomials

The purpose of this appendix is to give a rapid sketch of the combinatorics that underlies the graphical calculi presented in this paper. In particular, we shall give specific state summation formulas for the graphical evaluations corresponding to the Homfly and Dubrovnik polynomials. First we consider the case of the Homfly polynomial. For this, we need to define a state sum that has the properties indicated in Proposition 2.4. Thus, a planar 4-valent oriented graph \mathcal{G} is given, and we must define $[\mathcal{G}]$. A *state* of \mathcal{G} is defined relative to a *template* T . The template is a labelling of the edges of \mathcal{G} from a given ordered index set. A template is chosen at the outset – it is a generalized basepoint for the model. To obtain a state σ of \mathcal{G} , first replace some subset of the crossings of \mathcal{G} by oriented smoothings, as shown below.



Call $\tilde{\sigma}(\mathcal{G})$ this plane graph with smoothings.

Now choose the least label on the template T and begin walking along $\tilde{\sigma}(\mathcal{G})$ in the direction of the orientation. Whenever the walker meets a smoothing or a crossing for the *first time* this is indicated by marking the site as shown below:



In the case of the crossing, the overcrossing line indicates the first passage.

Once a circuit is completed, choose the least template label that has not been so far used and repeat the process. Once a crossing is decorated with a mark of first passage, it is left alone. When all sites are decorated we obtain the state $\sigma = \sigma(\mathcal{G})$. Each site v of σ is assigned a *vertex weight* $[v]$ according

to the rules:

$$\begin{aligned} \left[\begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} \right] &= a \\ \left[\begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array} \right] &= a^{-1} \\ \left[\begin{array}{c} \text{Diagram 5} \\ \text{Diagram 6} \end{array} \right] &= -B \\ \left[\begin{array}{c} \text{Diagram 7} \\ \text{Diagram 8} \end{array} \right] &= -A \end{aligned}$$

and we take $[\sigma]$ to be the product of these vertex weights. This will be multiplied by $\delta^{||\sigma||-1}$ where $||\sigma||$ denotes the number of closed oriented loops in σ (A loop is a projection of a link component) and $\delta = (a - a^{-1})/z$ where $z = A - B$. Then

$$[\mathcal{G}] = \sum_{\sigma} [\sigma] \delta^{||\sigma||-1}.$$

For example,  has template  and states , .

Thus

$$\begin{aligned} \left[\begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} \right] &= \left[\begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array} \right] \delta^{1-1} + \left[\begin{array}{c} \text{Diagram 5} \\ \text{Diagram 6} \end{array} \right] \delta^{2-1} \\ &= a\delta^0 + (-A)\delta^1 \\ &= \left(a + \frac{(-A)(a - a^{-1})}{A - B} \right) \\ &= \left(\frac{aA - Ba - Aa + Aa^{-1}}{A - B} \right) \\ &= \left(\frac{Aa^{-1} - Ba}{A - B} \right). \end{aligned}$$

Note that this agrees with the evaluation given in example 1 after Proposition 2.4.

The proof that this algorithm gives our graphical evaluations can be found in the discussion of skein models in [9] and in [3]. The apparently difficult point of proving independence of the choice of template is handled by

interpreting these models in terms of the link theory of the Homfly polynomial. Nevertheless, these graphical evaluations form the combinatorial basis for the Homfly polynomial and its associated graph invariants.

A similar state model gives the graphical evaluations corresponding to the Dubrovnik polynomial. Here the graph \mathcal{G} is unoriented, and we will give a state summation for $[\mathcal{G}]$ corresponding to the calculus of Theorem 5.1. Choose a template \mathcal{T} as before, but *orient each edge in \mathcal{T}* (This is a choice, but will be fixed for the given template.) Given \mathcal{G} , we form $\tilde{\sigma}$ by smoothing some subset of the crossings. A given crossing can be smoothed in two ways:



Traverse $\tilde{\sigma}$ using the template \mathcal{T} . Start from the edge with least label and *orient the journey via the direction on this labelled edge*. Thus the starting edge determines the direction of the journey. *Label the passages as before but include orientation corresponding to the direction of the journey*. The vertex weights are then given as shown below:

$$\begin{aligned} \left[\begin{array}{c} \curvearrowright \\ \bullet \end{array} \right] &= -B \\ \left[\begin{array}{c} \curvearrowright \\ \bullet \end{array} \right] &= -A \\ \left[\begin{array}{c} \curvearrowright \\ \text{double line} \end{array} \right] &= a \\ \left[\begin{array}{c} \text{double line} \\ \curvearrowright \end{array} \right] &= a^{-1}. \end{aligned}$$

These apply to any orientation of the site. Thus

$$\left[\begin{array}{c} \curvearrowright \\ \bullet \end{array} \right] = \left[\begin{array}{c}) \\ \bullet \end{array} \right] = \left[\begin{array}{c} \curvearrowleft \\ \bullet \end{array} \right] = \left[\begin{array}{c} \downarrow \\ (\end{array} \right]$$

and only the orientation of the edge of first passage is relevant.

We then have $[\sigma]$ equal to the product of the vertex weights, and

$$[\mathcal{G}] = \sum_{\sigma} [\sigma] \mu^{||\sigma||-1}$$

where $\mu = \left(\frac{a - a^{-1}}{A - B} \right) + 1$, and $||\sigma||$ is the number of loops in σ .

These state summations for the graphical evaluation deserve further investigation. They are inspired by the combinatorial models in [3]. A deeper investigation could turn the approach upside down so that the combinatorics of these graphical evaluations would be seen as the fundamental basis of the skein polynomials.

REFERENCES

1. BIRMAN, J.S. AND WENZL, H. Braids, link polynomials and a new algebra. *Trans. Amer. Math. Soc.*, **313**(1989), pp. 249-273.
2. BRAUER, R. On algebras which are connected with the semisimple continuous groups. *Annals of Math.*, Vol. **58**, No. 4, October 1937.
3. JAEGER, F. Composition products and models for the Homfly polynomial. *L'Enseignement Math.*, t**35**(1989), pp. 323-361.
4. JONES, V.F.R. Hecke algebra representations of braid groups and link polynomials. *Ann. of Math.*, **126**(1987), pp. 335-388.
5. KAUFFMAN, L.H. An invariant of regular isotopy. *Trans. Amer. Math. Soc.*, Vol. **318**, No. 2 (1990), pp. 417-471.
6. KAUFFMAN, L.H. **Knots and Physics**. World Sci. Pub. (1991).
7. KAUFFMAN, L.H. Knots, abstract tensors, and the Yang-Baxter equation. In **Knots, Topology and Quantum Field Theories - Proceedings of the Johns Hopkins Workshop on Current Problems in Particle Theory 13**. Florence (1989). ed. by L. Lussana. World Sci. Pub. (1989), pp. 179-334.
8. KAUFFMAN, L.H. **On Knots**. *Annals of Mathematics Study*, **115**, Princeton University Press (1987).
9. KAUFFMAN, L.H. State models for link polynomials. *L'Enseignement Math.*, t**36**(1990), pp. 1-37.
10. KAUFFMAN, L.H. Invariants of graphs in three-space. *Trans. Amer. Math. Soc.* Vol. **311**, 2 (Feb. 1989), 697-710.
11. LICKORISH, W.B.R. AND MILLETT, K.C. A polynomial invariant for oriented links. *Topology*, **26**(1987), 107-141.
12. MILLETT, K. AND JONISH, D. Isotopy invariants of graphs. (To appear in *Trans. A.M.S.*)
13. TURAEV, V.G. The Yang-Baxter equations and invariants of links. LOMI preprint E-3-87, Steklov Institute, Leningrad, USSR. *Inventiones Math.*, **92** Fasc. 3, 527-553.

Knots, Tangles, and Electrical Networks

JAY R. GOLDMAN*

School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

AND

LOUIS H. KAUFFMAN†

*Department of Mathematics, Statistics and Computer Science,
University of Illinois at Chicago, Chicago, Illinois 60680*

The signed graphs of tangles or of tunnel links (special links in $\{R^3\}$ -two parallel lines) are two terminal signed networks. The latter contain the two terminal passive electrical networks. The conductance across two terminals of a network is defined, generalizing the classical electrical notion. For a signed graph, the conductance is an ambient isotopy invariant of the corresponding tangle or tunnel link. Series, parallel, and star triangle methods from electrical networks yield techniques for computing conductance, as well as giving the first natural interpretation of the graphical Reidemeister moves. The conductance is sensitive to detecting mirror images and linking. The continued fraction of a rational tangle is a conductance. Algebraic tangles correspond to two terminal series parallel networks. For tangles, the conductance can be computed from a special evaluation of quotients of Conway polynomials and there is a similar evaluation using the original Jones polynomial. © 1993 Academic Press, Inc.

CONTENTS

1. *Introduction.*
 2. *Knots, tangles, and graphs.*
 3. *Classical electricity.*
 4. *Modern electricity—The conductance invariant.*
 5. *Topology: Mirror images, tangles, and continued fractions.*
 6. *Classical topology.*
- Appendix: From electricity to trees.*

*Partially supported by NSA Grant No. MDA 904 67-H-2016.

†Partially supported by NSF Grant No. DMS-8822602 and the Program for Mathematics and Molecular Biology, University of California at Berkeley, Berkeley, California.

1. INTRODUCTION

In this paper, we show how a generalization of the conductance of a classical electrical network gives rise to topological invariants of knots and tangles in three-dimensional space. The invariants that we define are chirality sensitive and are related, in the case of tangles, to the Alexander–Conway polynomial at a special value. The most general class of invariants defined here are the conductance invariants for special *tunnel links*. A tunnel link is an embedding of a disjoint collection of closed curves into the complement (in Euclidean three-space) of two disjoint straight lines or tubes.

We were led to these invariants by an analogy between the Reidemeister moves in the theory of knots and certain transformations of electrical networks. A knot or link projection is encoded by a (planar) signed graph (a graph with a $+1$ or a -1 assigned to each edge). The signed graphs are a subset of the signed networks (graphs with a nonzero real number assigned to each edge). These networks can be viewed as generalizations of passive resistive electrical networks, where the number on an edge is a generalized conductance. The series, parallel, and so-called star-triangle (or delta-wye) transformations of an electrical network, leave the conductance across *two terminals* unchanged. These notions generalize to invariance of a generalized conductance across two terminals of a signed network under corresponding transformations. When specialized to planar signed graphs, these transformations are the graph-theoretic translations of the Reidemeister moves for link projections (see Fig. 2.6 for a glimpse of this correspondence). Hence the invariance of conductance across two terminals of our graph yields an ambient isotopy invariant for the corresponding special tunnel link or tangle. Note that this electrical setting yields the first “natural” interpretation of the graph-theoretical version of the Reidemeister moves.

The paper is organized as follows. Section 2 gives background on knots, tangles, and graphs, and the translation of Reidemeister moves to graphical Reidemeister moves on signed graphs. The notion of a tunnel link is introduced and we discuss the relation with tangles. Section 3 reviews the background in classical electricity which motivates the connection between graphical Reidemeister moves and conductance-preserving transformations on networks. In Section 4, we define conductance across two terminals of a signed network in terms of weighted spanning trees. This coincides with the classical notion in the special case of electrical networks. We prove the invariance of conductance under generalized series, parallel, and star-delta transformations which then gives us the topological invariant for the corresponding tangles and special tunnel links. In Section 5, we show how the conductance invariant behaves under mirror images

and we provide numerous examples. We also treat the relation between continued fractions, conductance, and Conway's rational tangles. Section 6 shows how, for tangles, the conductance invariant is related to the Alexander-Conway polynomial. The proof uses a combination of state models for the Conway polynomial and the Jones polynomial. Finally, a short appendix traces the flow of ideas from classical electrical calculations to spanning trees of networks.

2. KNOTS, TANGLES, AND GRAPHS

The purpose of this section is to recall basic notions from knot and tangle theory, introduce the new notion of a tunnel link and to show how the theory is reformulated in terms of signed planar graphs. This will enable us to pursue the analogy between knot theory and electrical networks via the common language of the graphs.

A *knot* is an embedding of a circle in Euclidean three-space \mathbf{R}^3 . (The embedding is assumed to be smooth or, equivalently, piecewise linear [6].) Thus a knot is a simple closed curve in three-dimensional space. The standard simplest embedding (Fig. 2.1) will be denoted by the capital letter U and is called the unknot. Thus in this terminology the unknot is a knot! Two knots are said to be *equivalent* or *ambient isotopic* if there is a continuous deformation through embeddings from one to the other. A knot is said to be *knotted* if it is not ambient isotopic to the unknot. The simplest nontrivial (knotted) knot is the trefoil as shown in Fig. 2.1. As Fig. 2.1 illustrates, there are two trefoil knots, a right-handed version K and its left-handed mirror image K^* . K is not equivalent to K^* and both are knotted [16].

A *link* is an embedding of one or more circles (a knot is a link). Equivalence of links follows the same form as for knots. In Fig. 2.2 we illustrate an *unlink* of two components, the *Hopf link*, consisting of two unknotted components linked once with each other, and the *Borromean rings*, consisting of three linked rings such that each pair of rings in unlinked.

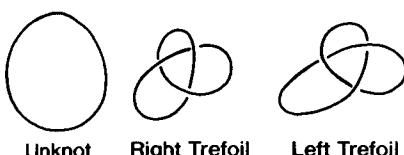


FIGURE 2.1

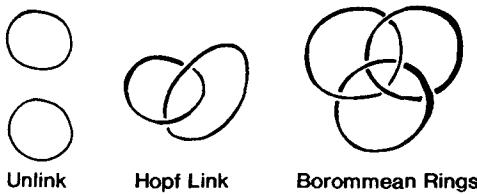


FIGURE 2.2

As these figures show, it is possible to represent knots and links by diagrammatic figures (*projections*). These projections are obtained by a parallel projection of the knot to a plane so that the strands cross transversely (Figs. 2.1, 2.2). The projections can also be thought of as planar graphs where each vertex (crossing) is 4-valent and where the diagram also shows which strand crosses over the other at a crossing.

Not only can one represent knots and links by diagrams, but the equivalence relation defined by ambient isotopy can be generated by a set of transformations of these diagrams known as the *Reidemeister moves*. Reidemeister [21] proved that two knots or links are equivalent (ambient isotopic) if and only if any projection of one can be transformed into any projection of the other by a sequence of the three Reidemeister moves (shown in Fig. 2.3) coupled with ordinary planar equivalence of diagrams given by homeomorphisms (i.e., continuous deformations) of the plane. We indicate the latter as a type-zero move in Fig. 2.3 (zero is indicated in

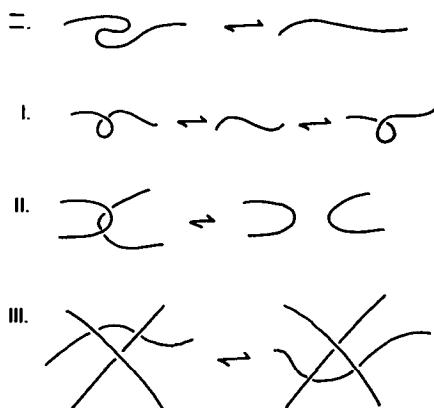


FIG. 2.3. Reidemeister moves.

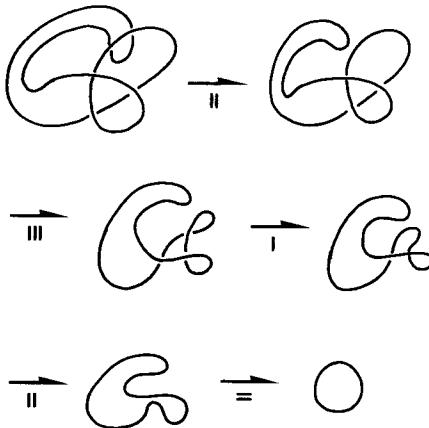


FIGURE 2.4

the Roman style by two parallel lines), showing the straightening of a winding arc. Each move, as shown in Fig. 2.3, is to be performed locally in a disc-shaped region without disturbing the rest of the diagram and without moving the endpoints of the arcs of the local diagrams that indicate the moves. In Fig. 2.4 we illustrate the use of the Reidemeister moves in unknotting a sample diagram.

We now translate this diagrammatic theory into another diagrammatic theory that associates to every knot or link diagram an arbitrary *signed planar graph* (a planar graph with a +1 or -1 assigned to each edge) as follows. Recall that a planar graph is a graph, together with a given embedding in the plane, where we usually do not distinguish between two such graphs that are equivalent under a continuous deformation of the plane.

Let K be a connected knot or link *projection* (i.e., it is connected as a 4-valent planar graph). It is easy to see (via the Jordan curve theorem [15] or by simple graph theory [22, Theorem 2-3]) that one can color the regions of K with two colors so that regions sharing an edge of the diagram receive different colors. Call such a coloring a *shading* of the diagram. Letting the colors be black and white, we refer to the black regions as the *shaded regions* and the white regions as the *unshaded* ones. To standardize the shading, let the unbounded region be unshaded. We can now refer to the shading of a connected diagram (Fig. 2.5).

Now there is a connected planar graph $G(K)$ associated to the shading of K . The nodes (or vertices) of $G(K)$ correspond to the shaded regions of K . If two shaded regions have a crossing in common, the corresponding

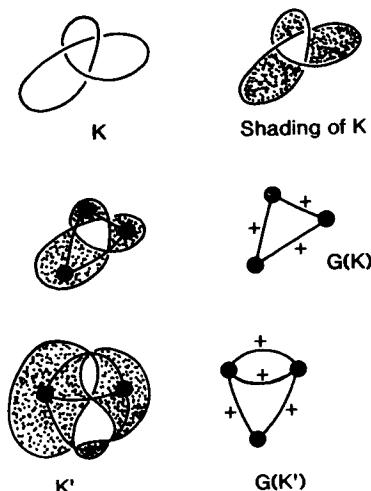


FIG. 2.5. Knots and their graphs.

nodes in $G(K)$ are connected by an edge (one edge for each shared crossing). Each edge of $G(K)$ is assigned a number, $+1$ or -1 , according to the relationship of the corresponding crossing in K with its two shaded regions. If by turning the over crossing strand at a crossing counterclockwise, this line sweeps over the shaded region, then this crossing and the corresponding edge in $G(K)$ is assigned a $+1$; otherwise a -1 . In Fig. 2.5, the construction of the signed graph is shown for the right-hand trefoil and for the figure eight knot. We shall refer to $G(K)$ as the (*signed*) graph of the knot or link diagram K . Note, by the construction, that since K is connected, $G(K)$ is connected.

Conversely, given a connected planar signed graph H , there is a link diagram $L(H)$ such that $G(L(H)) = H$ (see Fig. 2.5 and [31, 17]). Note that H connected implies that K is connected. Moreover, $L(G(K)) = K$ and thus $K \leftrightarrow G(K)$ is a bijection between connected link projections and connected signed planar graphs. (More precisely, the bijection is between equivalence classes of each under continuous deformation of the plane.) This correspondence is called the *medial* construction; K and H are called the *medials of each other*. The correspondence can be extended to nonconnected link diagrams and graphs by applying the constructions separately to each component of the link diagram or graph. This extension, which we also call the medial construction, is a bijection only if we ignore the relative positions of the components in the plane.

One can transfer the Reidemeister moves (of Fig. 2.3) to a corresponding set of moves on signed graphs [31]. These graphical Reidemeister moves

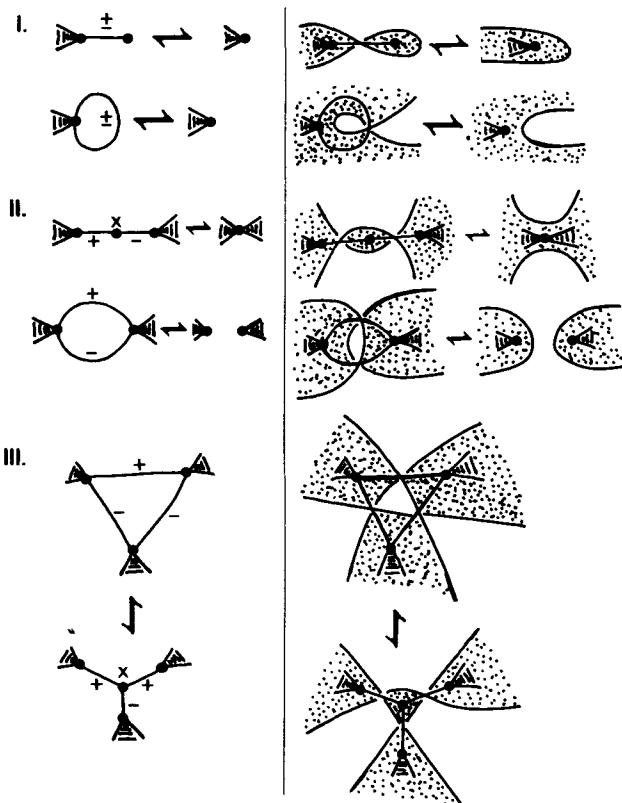


FIG. 2.6. Graphical Reidemeister moves.

are shown in Fig. 2.6, where, as for link diagrams, they are local moves, restricted to the signed configurations enclosed by the black triangles (which represent connections to the rest of the graph) and not affecting the other part of the graph. In Fig. 2.6, we also indicate informally how the graphical moves come about from the link diagram moves. The vertices incident to the triangles can have any valency, but the nodes marked with an x are *only* incident to the edges shown in the figures. Note that there are two graphical Reidemeister moves corresponding to the first Reidemeister move on link diagrams and two more corresponding to the second. This multiplicity is a result of the two possible local shadings. Move zero corresponds to allowing two-dimensional isotopies of the planar graph. The different possibilities for over and undercrossings in move three require exactly two of the three numbers on the triangle to have the same sign and the same for the star.

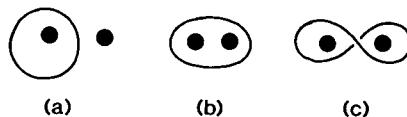


FIGURE 2.7

We now turn to the notions of tunnel knots and links, tangles, and their associated graphs—the topological concepts that are our main objects of study.

A *tunnel link* is a link that is embedded in Euclidean three-space with two infinite straight tunnels removed from it. More precisely, let D_1 and D_2 be two disjoint closed disks in the plane \mathbf{R}^2 . Regard \mathbf{R}^3 as the Cartesian product $\mathbf{R}^3 = \mathbf{R}^2 \times \mathbf{R}^1$. Then let the *tunnel space* $\tau(\mathbf{R}^3)$ be defined by $\tau(\mathbf{R}^3) = \mathbf{R}^3 - [(D_1 \times \mathbf{R}^1) \cup (D_2 \times \mathbf{R}^1)]$. The theory of knots and links in the tunnel space is equivalent to the theory of link diagrams (up to Reidemeister moves) in the punctured plane $\mathbf{R}^2 - (D_1 \cup D_2)$. In other words, tunnel links are represented by diagrams with two special disks such that no Reidemeister moves can pass through those disks. For example, the three “knots” in Fig. 2.7 are distinct from each other in this theory, where the disks D_1, D_2 are indicated by the black disks in these diagrams. The relevance of tunnel links will become apparent shortly.

A (two input–two output) *tangle* is represented by a link diagram, inside a planar rectangle, cut at two points of the diagram. Two adjacent lines of one cut are regarded as the top of the tangle and emanate from the top of the rectangle, and the other two lines from the bottom. Hence, inside the rectangle, no lines with endpoints occur. Examples of tangles are shown in Fig. 2.8.

The two simplest tangles, shown in Fig. 2.8, are “ ∞ ” and “ 0 .” In “ ∞ ” each top line is connected directly to its corresponding bottom line. In “ 0 ” the two top lines are connected directly to each other, as are the two bottom ones. The third tangle shown in Fig. 2.8 is the *Borromean tangle* B . Note the two general tangle constructions shown in Fig. 2.8. The *numerator* of a tangle, denoted $n(T)$, is the link obtained from T by joining the top strands to each other and the bottom strands to each other, on the outside of the tangle box. The *denominator*, $d(T)$, is the link obtained by joining top strands to bottom strands by parallel arcs—as shown in Fig. 2.8. The figure illustrates that the numerator, $n(B)$, of the Borromean tangle is the Borromean rings (see Fig. 2.2) and the denominator, $d(B)$, is the Whitehead link (see [6]).

Two tangles are *equivalent* if there is an ambient isotopy between them that leaves the endpoints of the top and bottom lines fixed and is

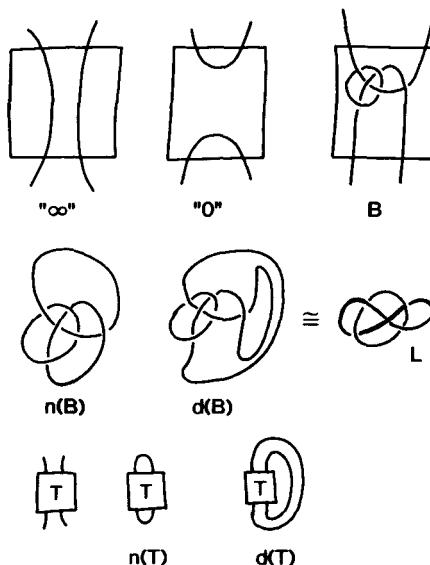


FIG. 2.8. Tangles: Numerators and denominators.

restricted to the space inside the tangle box. For more details on tangles, see [11, 26, 16, 6, 7].

Tangles are related to tunnel links as follows. If T is a tangle, let $\bar{n}(T)$ be the tunnel link obtained by placing disks in the (“top and bottom”) regions of the numerator $n(T)$ (see Fig. 2.9a for an example). A tunnel link K gives rise to a tangle if its disks are in regions adjacent to the unbounded region of the diagram (or to a common region if the diagrams

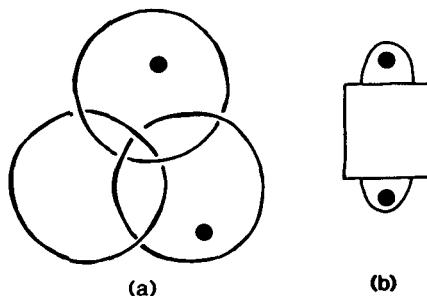


FIG. 2.9. Borromean tunnel link.

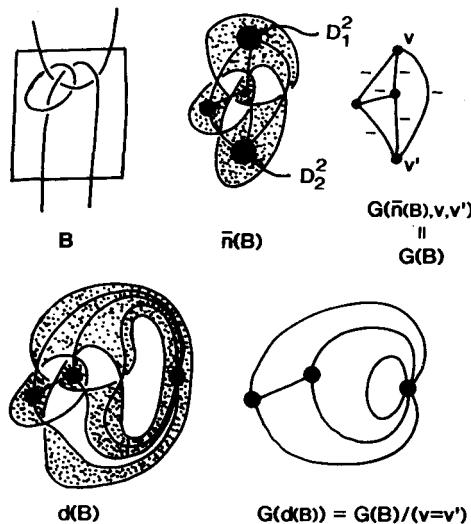


FIGURE 2.10

are drawn on the sphere). If this is the case the diagram will have the form of Fig. 2.9b, and hence $K = \bar{n}(T)$, for a tangle T . In this sense, tunnel links are generalizations of tangles.

A tunnel link K is said to be *special* if its disks lie in the shaded regions of K . This is the case for tunnel links of the form $\bar{n}(T)$. Note that the disks of a special tunnel link remain in shaded regions throughout isotopies generated by Reidemeister moves. *From now on, when we say tunnel link, we mean a special tunnel link.*

The *graph of a tunnel link* \mathbf{K} has two distinguished nodes corresponding to the shaded regions occupied by the disks D_1 and D_2 . Let these nodes by v and v' , respectively. The graph of the special tunnel link (K, D_1, D_2) will also be called the *tunnel graph of* \mathbf{K} , denoted by $(G(\mathbf{K}), v, v')$ (Fig. 2.10). A signed planar graph with designated nodes v and v' will be called a *two-terminal (signed) graph with terminals v and v'*. By the medial construction, every such two-terminal graph is a tunnel graph for some tunnel link.

In the case of a tangle T , where $\bar{n}(T)$ defines a special tunnel link (disks in the top and bottom regions), the two terminal graph $(G(\bar{n}(T)), v, v')$, also denoted by $(G(T), v, v')$, is the *tangle graph of* T (Fig. 2.10). We also use $G(T)$ for short, when no confusion can arise. Again, by the medial construction, a two-terminal graph is a tangle graph if and only if its terminals are both incident to the unbounded region in the plane. Note

that $G(d(T))$, the graph of the link $d(T)$, is obtained from the tangle graph $(G(T), v, v')$ by identifying the terminals v and v' (Fig. 2.10). Thus, as graphs, $G(n(T)) = G(T)$, while $G(d(T)) = G(T)/(v = v')$.

In order to be faithful to the restrictions on Reidemeister moves for tangle equivalence (moves only in the tangle box), we must add a corresponding restriction on the graphical moves allowed on the two terminal graphs corresponding to tangles. These restrictions are as follows:

1. If a node labeled x in a signed graph in Fig. 2.6 is a terminal, then the corresponding Reidemeister move is not allowed.
2. In the graph of a tangle, extend lines from each terminal to infinity, as shown in Fig. 2.11. No isotopy of the plane (graphical move zero) can carry parts of the graph across these lines.

Restriction 1 is general for the two terminal graphs corresponding to any tunnel link. Restriction 2 is necessary for tangles as shown in Fig. 2.11; otherwise, we could move a locally knotted piece on one strand over to another strand, which does not correspond to an ambient isotopy of tangles.

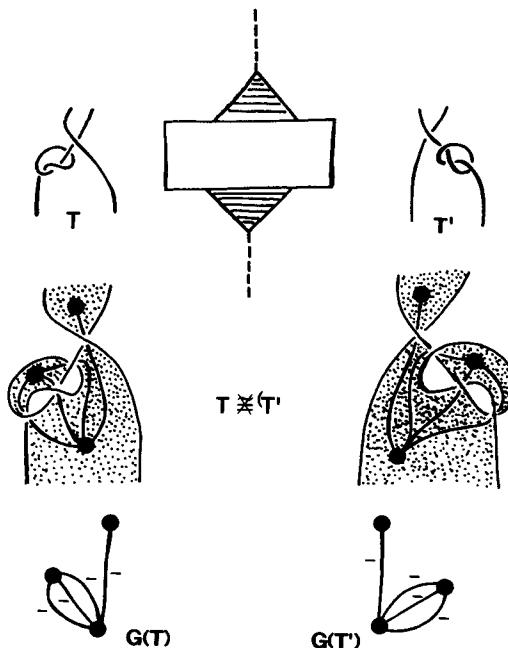


FIGURE 2.11

3. CLASSICAL ELECTRICITY

A classical (passive) electrical network can be modeled by a graph (with multiple edges allowed) such that each edge e is assigned a positive real number $r(e)$, its *resistance*. The basic theory of electrical networks is governed by Ohm's law and Kirchoff's laws (see [3, 23]).

Ohm's law, $V = ir(e)$, relates the current flowing in an edge e to the resistance $r(e)$ and the potential difference (voltage) measured across the endpoints. Kirchoff's *current law* states that the sum of the currents at a node (taking account of the direction of flow) is zero, while his *voltage law* says that the potential $V(v, v')$, measured across an arbitrary pair of nodes v and v' , equals the sum of the changes of potential across the edges of any path connecting v and v' .

If a current i is allowed to flow into the network N , only at the node v , and leaves it, only at v' , then the *resistance* $r(v, v')$ across v and v' is given by $r(v, v') = V(v, v')/i$. We find it more convenient to deal with the dual concept of the *conductance* $c(v, v')$ (or $c(N, v, v')$) across the nodes v and v' , which is defined to be $1/r(v, v')$. We also define $c(e)$, the conductance of an edge to be $1/r(e)$.

It is an interesting fact that $c(v, v')$ can be computed strictly in terms of the $c(e)$, for all edges e , independent of any currents or voltages (a formal definition, in a more general context, is given in the next section). Moreover, there are two types of "local" simplifications that can be made to a network, the series and parallel addition of edges, *without changing* $c(v, v')$:

(i) If two edges with conductances r and s are *connected in parallel* (they are incident at two distinct nodes), then $c(v, v')$ does not change when the network is changed by replacing these edges by a single edge, with conductance $t = r + s$, connecting the same endpoints (Fig. 3.1).

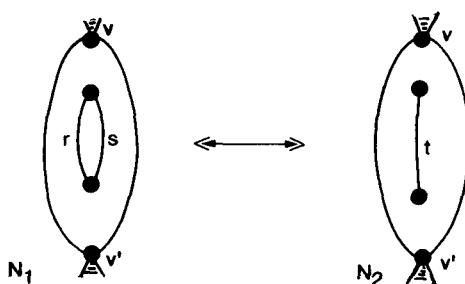


FIGURE 3.1

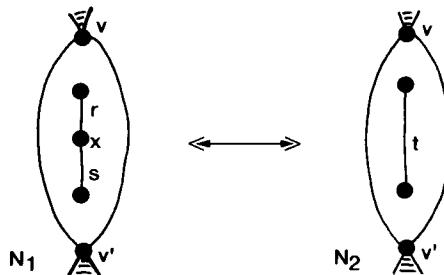


FIGURE 3.2

(ii) If two edges with conductances r and s are *connected in series* (they have exactly one vertex x in common and no other edges are incident to x) and $x \neq v, v'$, then $c(v, v')$ does not change when the network is changed by replacing these edges by a single edge of conductance t , where $1/t = 1/r + 1/s$ or $t = rs/(r+s)$ (Fig. 3.2).

There are two more subtle transformations of networks allowed, the so-called star to triangle and triangle to star substitutions, which leave the conductance invariant. We present it in a more symmetric form, the *star–triangle* relation [3, 19]:

(iii) Let the network N_1 contain the triangle with vertices u, w, y , with conductances a', b', c' on the edges (Fig. 3.3—other edges may also connect u, w , and y). Let the network N_2 be identical to N_1 , except that the triangle is replaced by the star with endpoints u, w, y and a new center point x and edge conductances a, b, c (Fig. 3.3—no other edges are incident to x). If the conductances satisfy the relations

$$\begin{aligned} a' &= S/a, & b' &= S/b, & c' &= S/c, \\ a &= D/a', & b &= D/b', & c &= D/c', \end{aligned} \tag{1}$$

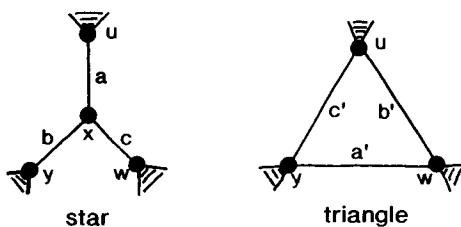


FIGURE 3.3

where $D = a'b' + a'c' + b'c'$ and $S = abc/(a + b + c)$ (of course, by (1), $S = D$), then $c(N_1, v, v') = c(N_2, v, v')$.

Note an absolutely remarkable “coincidence.” The operation of replacing a star by a triangle or a triangle by a star corresponds, purely graph-theoretically, to the third graphical Reidemeister move (Fig. 2.6). Moreover, if we let $a = -1$, $b = c = +1$, $a' = +1$, $b' = c' = -1$ in Fig. 3.3, then these values correspond to the signs on graphical Reidemeister move three in Fig. 2.6 and *they also satisfy the relation in (1)*. But the relations in (1) refer to *positive* conductances. The observation that it holds for the signed graphs describing the Reidemeister move is amazing and the goal of this paper is to understand it.

There is a generalization of electrical networks, which includes a generalization of the notion of conductance across two nodes. The two terminal signed graphs of tunnel links and tangles are a subclass of these networks. The graphical Reidemeister moves are special cases of series and parallel addition (for move two), star-triangle transformations (for move three), and one other transformation (for move one), and the conductance across two nodes is invariant under these operations. This allows us to define a new ambient isotopy invariant for special tunnel links and tangles. We develop these ideas in the next two sections.

4. MODERN ELECTRICITY—THE CONDUCTANCE INVARIANT

Up to now we have discussed (two terminal) signed graphs and classical electrical networks, as well as motivating the need for allowing negative conductances. Now we think of these graphs as embedded in a broader class of “generalized electrical networks” and generalize the notion of conductance across the terminals.

We allow graphs with loops and multiple edges. A (signed) *network* is a graph with a nonzero real number $c(e)$ assigned to each edge e , called the *conductance of e*. If T is a spanning tree of the network N , the *weight of T*, $w(T)$, is the product $\prod c(e)$, taken over all edges e of T . $w(N)$, the *weight of N*, is the sum $\sum w(T)$, taken over all all spanning trees T of N . If $c(e) > 0$, for all e , then N is an electrical network.

A *two-terminal* network (N, v, v') is a signed network N with two distinguished nodes v and v' , called *terminals* (we allow $v = v'$). Let $N + e'$ (really $(N + e', v, v')$) be the net obtained from N by adding a new edge e' , connecting v and v' , with $c(e') = 1$ (Fig. 4.1). $w'(N + e')$ is the sum $\sum w(T)$, taken over all all spanning trees T of $N + e'$ which contain e' .

We define $c(N, v, v')$, the *conductance across the terminals v and v'* in (N, v, v') . For now, let N be understood and just write $c(v, v')$. $c(v, v')$

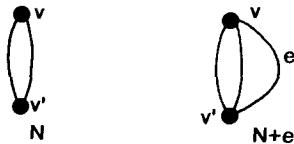


FIGURE 4.1

will take values in the real numbers with the symbols $0/0$ and ∞ adjoined, with the conventions $r/0 = \infty$, for r a nonzero real number, and $r/\infty = 0$, for any real r .

DEFINITION. (1) If N is connected, then

$$c(v, v') = w(N)/w'(N + e') \quad (1)$$

(if $v = v'$, we let $w'(N + e') = 0$ —there are no spanning trees containing the loop e' —and if $N = v$, i.e., there are no edges, we take $w(N) = 1$ —there is only the empty spanning tree of weight 1—and, in this case, $c = 1/0 = \infty$).

(2) If N has two connected components, one containing v and the other containing v' , then $c(v, v')$ is also given by Eq. (1), where $w(N) = 0$ (N has no spanning trees). In this case, $c(v, v')$ is either 0 or $0/0$.

(3) In all other cases, $c(v, v') = 0/0$.

Remarks. (i) If all edges have positive conductance, then our definition does yield the classical notion of conductance across two terminals in an electrical network (see [5, Theorem 3.4] and our appendix). We also define $r(N, v, v')$, the *resistance* between v and v' to be $1/c(N, v, v')$, which also reduces to classical resistance when all $c(e)$ are positive.

(ii) If e is an edge of the network N , let $N - e$ denote the network obtained from N by deleting e , and let N/e denote the network obtained from N by *contracting along the edge e* (deleting e and identifying its endpoints—Fig. 4.2).

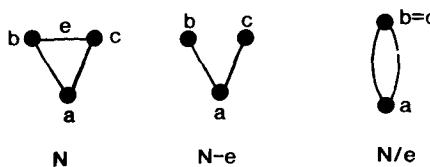


FIGURE 4.2

Now Eq. (1) can also be written as

$$c(v, v') = w(N' - e')/w(N'/e'), \quad (1')$$

where $N' = N + e'$. Since $N = N' - e'$, the numerators in (1) and (1') are equal. If e' is not a loop, then $T \leftrightarrow T/e'$ is a bijection between those trees T of N which contain e' and all spanning trees of N/e' . Since $w(T) = c(e')w(T/e') = w(T/e')$, we have $w'(N') = w(N'/e')$, and the denominators of (1) and (1') are also equal. When e' is a loop or in other cases, the equality of the denominators is handled by the definition of $w(N'/e')$.

In the case of a tangle graph (G, v, v') of a tangle K , the somewhat more symmetric form (1') is important because $(G + e') - e'$ and $(G + e')/e'$ are the graphs of the numerator and denominator of K , respectively.

(iii) Both tangle and tunnel graphs are planar. However, conductance is a function of the underlying abstract graph and independent of any special planar embedding.

The following theorems show that the operations discussed in the last section, together with the hanging edge result in Theorem 1 which are routinely used to calculate resistance and conductance in electrical networks, are true for our more general networks. Special cases of these operations correspond to the invariance of conductance under the graphical Reidemeister moves and therefore the conductance of the graph of a tunnel knot or a tangle is an invariant of the corresponding tunnel knot or tangle.

Unless otherwise stated, *tree* will mean *spanning tree*. Recall that an edge e with endpoints a and b is a *loop* if $a = b$ and a *link* if $a \neq b$. When no confusion can arise, we use the same symbol to denote an edge and its conductance.

An edge e of the network N is a *hanging edge* if it is a loop or if it is a link connected to the rest of the graph at exactly one endpoint (Figs. 4.3a, b). Hanging edges play no role in computing conductance, i.e.,

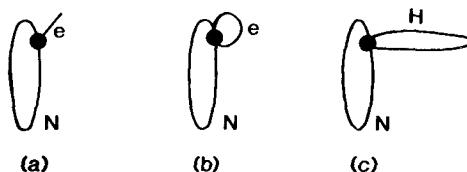


FIGURE 4.3

THEOREM 1. *If (N, v, v') , $v \neq v'$, has hanging edge e , then $c(N, v, v') = c(N - e, v, v')$.*

Proof. If e is a loop, it is not in any tree of N . Therefore, the trees of N and $N - e$ are identical, as are the trees of $N + e'$ and $N - e + e'$ containing e' . So the corresponding weights are the same and we are done.

If e is a link, then every tree of N (resp. $N + e'$) contains e , and the correspondence $T \leftrightarrow T - e$ defines a bijection between the trees of N and the trees of $N - e$ (resp. between the trees of $N + e'$ containing e' and the trees of $N - e + e'$ containing e'). Since $w(T) = ew(T - e)$, in all cases, e cancels from the numerator and denominator of the formula for $c(N, v, v')$ and $c(N, v, v') = c(N - e, v, v')$.

When the conductance of the hanging edge is ± 1 , Theorem 1 implies that the conductance of a tangle graph or a tunnel graph across the terminals is invariant under Reidemeister move I. More generally, it is easy to prove that we can delete a “hanging” subnetwork H from a network N without changing the conductance (Fig. 4.3c), if $w(H) \neq 0, 0/0, \infty$. Figure 4.10 at the end of this section illustrates the problem when $w(H) = 0$.

THEOREM 2. *Let (N_1, v, v') be a two-terminal network which contains two links r and s connected in series, with their common node x not incident to any other edges, and assume $x \neq v, v'$.*

(i) *If $r + s \neq 0$, let N_2 be the network obtained from N_1 by replacing the edges r and s by the single edge t (Fig. 3.2). If $1/t = 1/r + 1/s$ or $t = (rs)/(r + s)$, then*

$$w(N_1) = (r + s)w(N_2), \quad w'(N_1 + e') = (r + s)w'(N_2 + e'), \quad (2)$$

and

$$c(N_1, v, v') = c(N_2, v, v'). \quad (3)$$

(ii) *If $r + s = 0$, let N_2 be the network obtained from N_1 by contacting the edges r and s to a point (Fig. 4.4). Then*

$$w(N_1) = (-r^2)w(N_2), \quad w'(N_1 + e') = (-r^2)w'(N_2 + e'), \quad (4)$$

and

$$c(N_1, v, v') = c(N_2, v, v'). \quad (5)$$

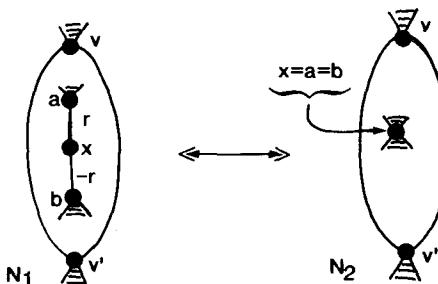


FIGURE 4.4

Proof. We need only prove (2) and (4), since taking quotients yields (3) and (5).

(i) First we work with $w(N_1)$ and $w(N_2)$. Since a tree in N_1 must contain the node x , it either contains both r and s or exactly one of them. In N_2 , a tree either contains t or it does not.

If a tree T_1 in N_1 contains r , but not s , it has weight of the form rw . We pair T_1 with the tree $T_2 = T_1 - r + s$ containing s , but not r , of weight sw . Hence the two trees $\{T_1, T_2\}$ contribute $(r+s)w$ to $w(N_1)$. Let $\{T_1, T_2\}$ correspond to $T' = T_1 - r (= T_2 - s)$, a tree in N_2 of weight w . Hence $w(T_1) + w(T_2) = (r+s)w(T')$.

A tree T in N_1 containing r and s has weight of the form wrs . Let T correspond to $T' = T - r - s + t$, a tree in N_2 of weight $tw = (rs/r + s)w$. Hence $w(T) = (r+s)w(T')$.

Combining both cases, which include all trees in N_1 and N_2 , we see that $w(N_1) = (r+s)w(N_2)$. The same argument, restricted to trees containing e' proves that $w'(N_1 + e') = (r+s)w'(N_2 + e')$ and we are done.

(ii) We have two edges r and $-r$ in series. The proof requires a small modification of the proof in (i). Again we work with $w(N_1)$ and $w(N_2)$.

The sum of the weights of a tree T_1 in N_1 containing r but not $-r$, and the tree $T_2 = T_1 - r + (-r)$ is zero; so they contribute zero to $w(N_1)$.

A tree T in N_1 containing r and $-r$ has weight of the form $-r^2w$. Let T correspond to $T' = T - r - (-r)$, a tree in N_2 of weight w ; so $w(T) = -r^2w(T')$. Conversely, a spanning tree T' in N_2 must contain x . Thus we can split the node x and reinsert the two edges r and $-r$ in series (going from N_2 back to N_1) obtaining the tree $T = T' + r + (-r)$. Therefore $T \leftrightarrow T'$ is a bijection between trees in N_1 containing r and $-r$ and all trees in N_2 . Since these are the only types of trees contributing nonzero terms to $w(N_1)$ and $w(N_2)$, we have $w(N_1) = (-r^2)w(N_2)$. Similarly, $w'(N_1 + e') = (-r^2)w'(N_2 + e')$ and we are done.

Note that if the numerator or denominator in the conductance are zero, our arguments are still okay; so we have included the cases where the conductance may be infinite or $0/0$.

In electrical circuit terminology, the theorem just says that for two edges connected in series, the reciprocals of their conductances, viz., their resistances, can be added without changing $c(v, v')$. If $r + s = 0$, then $t = \infty$ (a “short circuit”), which is the same as contracting t to a point.

When $r = 1$, $s = -1$, Theorem 2 says that the conductance of a tangle graph or tunnel graph across the terminals is invariant under the series version of Reidemeister move II.

THEOREM 3. *Let (N_1, v, v') be a two-terminal network which contains two links r and s connected in parallel (the two edges are incident at two distinct vertices).*

(i) If $r + s \neq 0$, let N_2 be the network obtained from N_1 by replacing the edges r and s by the single edge t . If $t = r + s$ (Fig. 3.1), then

$$w(N_1) = w(N_2), \quad w'(N_1 + e') = w'(N_2 + e'), \quad (6)$$

and

$$c(N_1, v, v') = c(N_2, v, v'). \quad (7)$$

(ii) If $r + s = 0$, let N_2 be the network obtained from N_1 by deleting the edges r and s (Fig. 4.5). Then

$$w(N_1) = w(N_2), \quad w'(N_1 + e') = w'(N_2 + e'), \quad (8)$$

and

$$c(N_1, v, v') = c(N_2, v, v'). \quad (9)$$

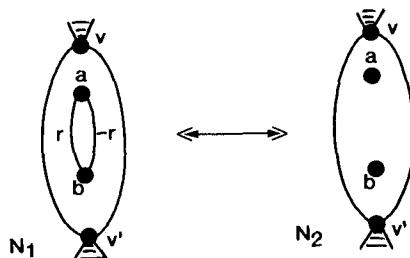


FIGURE 4.5

Proof. Equations (7) and (9) follow from (6) and (8), so we prove the latter. The proofs are very similar to those of Theorem 2 for series connections.

(i) Since r and s are in parallel, every tree in N_1 contains exactly one of the two edges or neither edge. If a tree T_1 in N_1 contains r , but not s , it has weight of the form rw . We pair T_1 with the tree $T_2 = T_1 - r + s$, containing s , but not r , of weight sw . Hence the two trees $\{T_1, T_2\}$ contribute $(r+s)w$ to $w(N_1)$. Let $\{T_1, T_2\}$ correspond to $T_3 = T_1 - r + t$, a tree in N_2 of weight tw . Therefore, $\{T_1, T_2\}$ and T_3 contribute the same weight to $w(N_1)$ and $w(N_2)$, respectively.

A tree T in N_1 which does not contain r or s is also a tree in N_2 , not containing t . Hence it contributes the same weight to $w(N_1)$ and $w(N_2)$.

Since we have considered all trees in N_1 and N_2 , we have $w(N_1) = w(N_2)$. Similarly, $w'(N_1 + e') = w'(N_2 + e')$.

(ii) When $r + s = 0$, the only modification we need in (i) is to note that $w(T_1) + w(T_2) = rw + (-r)w = 0$ and there is no corresponding T_3 . Hence only the trees T not containing r or s may contribute nonzero weights to $w(N_1)$ and $w(N_2)$, and they are the same trees in both networks.

In electrical circuit terminology, the theorem just says that for two edges connected in parallel, their conductances can be added without changing $c(v, v')$. If $r + s = 0$, then $t = 0$ (no current can flow), which is the same as deleting the edge.

When $r = 1, s = -1$, Theorem 2 says that the conductance of a tangle graph or a tunnel graph across the terminals is invariant under the parallel version of Reidemeister move II.

The last class of operations we need to generalize from electrical theory are the star–triangle (or wye–delta) transformations, which are also used in a variety of other network theories (switching, flow, etc.) as well as in statistical mechanics (see [1, 20, 2, 18, 24]). A *triangle* (or *delta*) in a network is, of course, a subset of three edges uy, yw, wu , with distinct nodes u, y, w (Fig. 3.3). A *star* (or *wye*) in a network is a subset of three edges xu, xy, xw , on the distinct nodes u, y, w, x , where x , the *center* of the star, has degree three (Fig. 3.3).

Given a star in a network N , with conductances a, b, c and end-nodes u, y and center x , such that $Y = a + b + c \neq 0$, a *star–triangle* (or *wye–delta*) transformation N consists of replacing the center and edges of the star by the triangle with nodes u, y, w and conductances a', b', c' (Fig. 3.3), given by

$$a' = S/a, \quad b' = S/b, \quad c' = S/c, \quad (10)$$

where $S = abc/Y$. Equivalently,

$$a' = bc/Y, \quad b' = ac/Y, \quad c' = ab/Y. \quad (10')$$

Given a triangle in a network N , with conductances a', b', c' and nodes u, v, w , such that $D = a'b' + b'c' + c'a' \neq 0$, a *triangle-star* (or *delta-wye*) transformation N consists of replacing the triangle by the center and edges of the star with end-nodes u, v, w , center x , and conductances a', b', c' (Fig. 3.3), where

$$a = D/a', \quad b = D/b', \quad c = D/c'. \quad (11)$$

These operations are inverse to one another and Eqs. (10) and (11) have a pleasing symmetry. To prove this and obtain some useful relations, we start with a triangle, but state the results as a purely arithmetic lemma.

LEMMA. *Let a', b', c' be nonzero real numbers, such that $D = a'b' + b'c' + c'a' \neq 0$. Define a, b, c , by Eqs. (11), and let $Y = a + b + c$ and $S = abc/Y$. Then*

- (i) $D^2 = Ya'b'c'$,
- (ii) $Y, S \neq 0$,
- (iii) $DY = abc$ and $S = D$,
- (iv) Eqs. (10) are true.

Proof. (i)

$$\begin{aligned} Y &= a + b + c = D/a' + D/b' + D/c', \quad \text{by (11)} \\ &= D(1/a' + 1/b' + 1/c') = D((a'b' + b'c' + c'a')/a'b'c') = D^2/a'b'c'. \end{aligned}$$

(ii) By assumption, $D, a', b', c' \neq 0$, so $Y \neq 0$, by (i), and $S = abc/Y \neq 0$.

(iii) Multiplying the three equations in (11) yields $abc = D^2D/a'b'c'$, which by (i) equals $(Ya'b'c')D/a'b'c' = YD$, and, by the definition of S , $S = D$.

(iv) By Eq. (11), $a' = D/a$, which equals S/a , by (iii). Similarly for b' and c' .

Of course, (iv) yields the invertibility; i.e., beginning with a triangle ($D \neq 0$), replacing it with a star ($S \neq 0$, by (ii)), and replacing this star with a triangle, yields the original triangle. Conversely, we could start with a star (with $S \neq 0$, and a', b', c' defined by (10)) and prove that $D \neq 0$ and Eqs. (11) are satisfied. Thus $S \neq 0 \Leftrightarrow D \neq 0$.

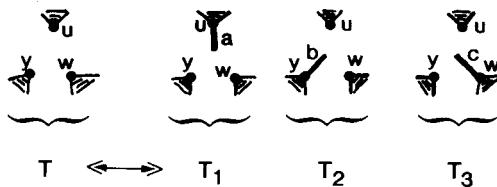


FIGURE 4.6

THEOREM 4. *If the network (N_2, v, v') is obtained from (N_1, v, v') by a star–triangle or a triangle–star transformation, where the center of the star is not a terminal, then*

$$w(N_2) = Yw(N_1), \quad w'(N_2 + e') = Yw'(N_1 + e'),$$

and

$$c(N_1, v, v') = c(N_2, v, v').$$

Proof. Assume we go from N_1 to N_2 by a triangle–star as labeled in Figure 3.3. We first consider $w(N_1)$ and $w(N_2)$. A tree in N_1 contains 0, 1, or 2 edges of the triangle, while a tree in N_2 contains 1, 2, or 3 edges of the star (it must contain the center x). Let T be a tree in N_1 .

(1) If T does not contain any edges of the triangle, let it correspond to the three trees $\{T_1 = T + a, T_2 = T + b, T_3 = T + c\}$ in N_2 (Fig. 4.6). Then $\sum w(T_i) = (a + b + c)w(T) = Yw(T)$.

(2) If T contains exactly one edge of the triangle, say a' , let it correspond to the tree $T' = T - a' + b + c$ in N_2 , where we replaced a' by the two edges b, c of the star which are incident to it (Fig. 4.7). Then $w(T') = w(T)bc/a'$ which, by the first equation in (10'), equals $Yw(T)$, and similarly if T contains b' or c' .

(3) If T contains two edges of the triangle, say a' and b' , then replacing these edges by the pair b', c' or c', a' yields two more spanning

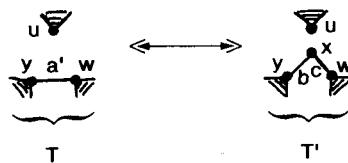


FIGURE 4.7

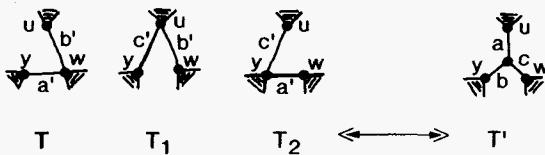


FIGURE 4.8

trees T_1 and T_2 . Associate these three trees with the tree $T' = T - a' - b' + a + b + c$ in N_2 ; i.e., replace the triangle edges by the complete star (Fig. 4.8). Writing $w(T)$ in the form $a'b'w$, we then have $w(T_1) = b'c'w$, $w(T_2) = a'c'w$, and $w(T') = abcw$. By part (ii) of the lemma, $abc = DY$; so $w(T') = DYw$. But $w(T) + w(T_1) + w(T_2) = (a'b' + b'c' + c'a')w = Dw$ and, therefore, $w(T') = Y(w(T) + w(T_1) + w(T_2))$.

Combining all three cases, which include all trees contributing to $w(N_1)$ and $w(N_2)$, we have proved that $w(N_2) = Yw(N_1)$. The same reasoning applies to prove that $w'(N_2 + e') = Yw'(N_1 + e')$ (note that case 3 cannot occur if both terminals are nodes of the triangle). Hence the conductances are equal.

When the conductances on a star (or triangle) are ± 1 and not all of the same sign, then Theorem 4 says that the conductance of a tangle graph is invariant under the graphical Reidemeister move III.

We define the *conductance* $c(T)$ of a special tunnel link T or of a tangle T to be the conductance across the terminals of its two terminal signed graph, i.e., $c(T) = c(G(T), v, v')$. For the case of a tangle we have, by Remark (ii) at the beginning of this section,

$$c(T) = w(G(n(T))) / w(G(d(T))). \quad (12)$$

Since we have proved that two terminal signed graphs are invariant under all the graphical Reidemeister moves, we have our main theorem.

THEOREM 5. *The conductance of a special tunnel link or of a tangle is an ambient isotopy invariant.*

Not only do we have an invariant, but also some very powerful techniques for computing it. Using series, parallel, and star–triangle transformations, which usually transform the two-terminal graph to a more general two-terminal network with the same conductance, provides the general approach. For example we compute $c(v, v')$ for the Borromean tangle (Fig. 2.8). The series of transformations shown in Fig. 4.9, consisting of a star–triangle transformation followed by three parallel additions and

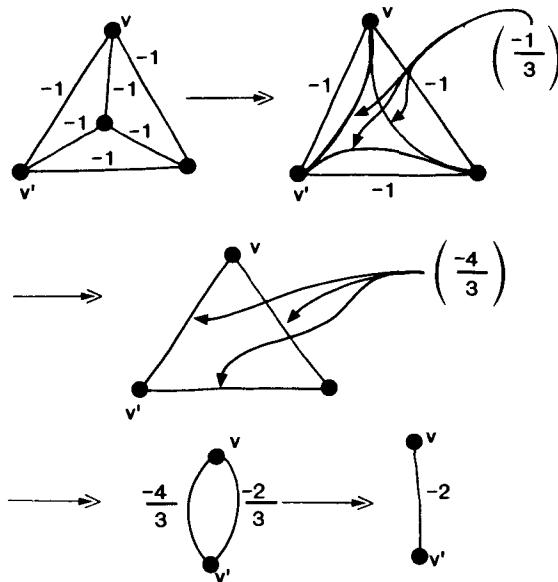


FIGURE 4.9

finally a series addition, convert the two-terminal graph to a two-terminal network with the same conductance. The final network consists of two terminals connected by one edge with conductance -2 , and, therefore, $c(v, v') = -2$ by definition (Eq. (1)).

It is natural to ask whether we can always compute the conductance of a two-terminal signed network (or more special, a tangle or tunnel graph) using only the operations of series and parallel addition, star and triangle swaps, and deleting allowable hanging subnetworks. For classical planar electrical networks (positive conductance on all edges) a conjecture that this could always be done was made by A. Lehman in 1953 [20]. It was first proved true by Epifanov [10], with subsequent proofs by Truemper [28] and Feo and Proven [12]. In this case it is equivalent to using our operations to transform any connected planar electrical network to a network consisting of the two terminals connected by one edge.

When negative conductances are allowed on the edges, the question is more subtle and difficult. As we see in Fig. 4.10, a connected planar graph can be transformed to a disconnected one; hence the question is whether we can use our operations to transform any network to one where each connected component consists of one edge or one node. One difficulty is that star and triangle swaps can only be done when $S, D \neq 0$, whereas in the electrical case they can always be done. Series and parallel addition

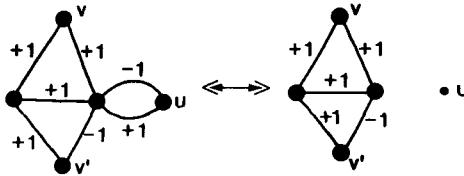


FIGURE 4.10

can be reduced to the classical case even when deletion and contraction occur (by allowing $c(e) = 0$ or ∞).

5. TOPOLOGY: MIRROR IMAGES, TANGLES AND CONTINUED FRACTIONS

5.1. Mirror Images

The main purpose of this section is to present a set of examples which give some idea of the range of applicability of the conductance invariant. An important property of conductance is its ability to detect mirror images. The *mirror image* of a tunnel link or tangle with diagram K is the tunnel link or tangle whose diagram K^* is obtained from K by reversing all the crossings. If G is a network, G^* is the graph obtained from G by multiplying $c(e)$ by -1 , for all e . Hence the tunnel or tangle graph of K^* is $(G(K)^*, v, v')$. A tunnel link or tangle is *achiral* if it is equivalent to its mirror image and *chiral* if not.

THEOREM. *If $c(G, v, v') \neq 0, 0/0, \infty$, then $c(G^*, v, v') = -c(G, v, v')$. If we also have $G = G(K)$, for some tangle or tunnel link K , then $c(K^*) = -c(K) \neq c(K)$ and K is chiral.*

Proof. By the definition of conductance (Eq. (4.1)), $c(G^*, v, v') = w(G^*)/w'(G^* + e')$. A spanning tree T of G corresponds to a spanning tree T^* of G^* (same edges—conductance multiplied by -1). If G and G^* have $n + 1$ vertices, then T and T^* have n edges. Therefore $w(T^*) = (-1)^n w(T)$ and $w(G^*) = (-1)^n w(G)$.

We have a similar correspondence between the spanning trees T of $G + e'$, containing e' , and the spanning trees T^* of $G^* + e'$, containing e' . Since $c(e') = 1$, $w(T^*) = (-1)^{n-1} w(T)$ and $w'(G^* + e') = (-1)^{n-1} w'(G + e')$.

Hence, since $c(G, v, v') \neq 0, 0/0, \infty$ guarantees $w(G^*)$ and $w'(G^* + e')$ are not zero,

$$c(G^*, v, v') = \frac{(-1)^n w(G)}{(-1)^{n-1} w'(G + e')} = -c(G, v, v').$$

COROLLARY. *If the tunnel link or tangle K is alternating (the conductances of all the edges of $G(K)$ have the same sign), then K is chiral.*

Proof. By the tree definition of conductance, $c(G, v, v') \neq 0, 0/0, \infty$, and we apply the theorem.

EXAMPLE 1. A tunnel link or tangle formed from the *trefoil* knot of Fig. 2.5 is alternating (with any choice of terminals) and the corollary applies. In particular, conductance detects the simplest type of knottedness.

EXAMPLE 2. The *figure eight* knot of Fig. 2.5 is achiral [16]. However, by the theorem, any way of turning this into a tunnel link or tangle (i.e., any choice of terminals in the corresponding graph) destroys this property since it has positive conductance (Fig. 5.1 shows one such example).

EXAMPLE 3. The *Hopf* link H (Fig. 2.2) has a graph consisting of two -1 edges in parallel; thus there is only one choice of terminals, $G(H)$ is alternating and the corollary applies. In particular, we see that the conductance can detect the simple linking of two strands in tunnel links or tangles,

EXAMPLE 4. The *Borromean* tunnel link and tangle (Fig. 2.10) is alternating and thus chiral. Hence, we see that the conductance can detect the more subtle linking of this example.

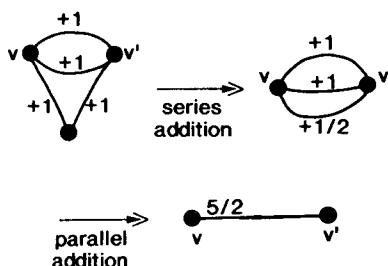


FIGURE 5.1

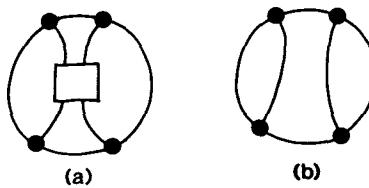


FIGURE 5.2

5.2 Rational Tangles and Continued Fractions

Let T be a *two-strand tangle* (only two strands are used to construct it) enclosed in the interior of a two-sphere, except for the endpoints of the four lines emanating from it which are on the boundary of the sphere (Fig. 5.2a). If there is an ambient isotopy taking T into the trivial ∞ tangle (Fig. 5.2b), with the restriction that the endpoints can only move on the sphere and the rest of the tangle remains inside the sphere during the deformation, then T is called a *rational tangle*. Intuitively, a rational tangle is a tangle that can “untwist” by moving the endpoints on the sphere. Every rational tangle is equivalent to a *canonical* rational tangle [6, 7, 11, 27], i.e., a tangle constructed as follows:

Start with two vertical strands (the ∞ tangle). Holding the top two endpoints fixed, twist the bottom two endpoints around each other some number of times in the positive or negative direction (the sign is the sign

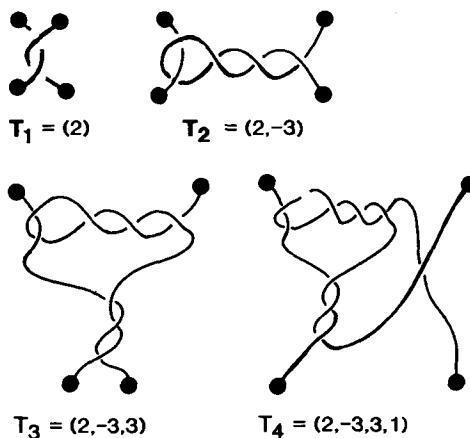


FIGURE 5.3

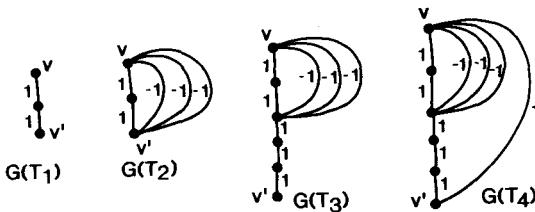


FIGURE 5.4

of the edges of the graph of this twist considered as a tangle T_1 —see Figs. 5.3 and 5.4). Then hold the two left endpoints fixed and twist the two right endpoints some number of times in the positive or negative direction to obtain a new tangle T_2 (again, the sign is determined by the sign on the edges of $G(T_2)$ corresponding to this twist—Figs. 5.3 and 5.4). Continue by alternately twisting the two bottom endpoints and the two right endpoints, stopping either at the bottom or on the right after a finite number of twists.

For example, the sequence of twists $(2, -3, 3, 1)$ leads to the sequence of simple tangles shown in Fig. 5.3 and the corresponding tangle graphs of Fig. 5.4.

More generally, let T be the tangle given by the sequence of twists (t_1, t_2, \dots, t_r) , $t_i \in \mathbf{Z}$. The graph of T is obtained by t_1 horizontal edges in series, connecting the top and bottom by t_2 edges in parallel, adding t_3 horizontal edges in series to the bottom, connecting the top and bottom by t_4 edges in parallel, and so on. The signs ($\neq 1$) on the edges are determined by the signs on the t_i 's and the top and bottom points of the final graph are the terminals.

We compute the conductance of $G(T_4)$ by an alternating sequence of series and parallel additions which correspond to the twists (Fig. 5.5). So $c(G(T_4))$ is given by a continued fraction. In fact, using the sequence of twists to denote the corresponding simple tangle, we have, by the intermediate steps in Fig. 5.5,

$$c(T_1) = c((2)) = \frac{1}{2}, \quad c(T_2) = c((2, -3)) = -3 + \frac{1}{2},$$

$$c(T_3) = c((2, -3, 3)) = \frac{1}{3 + 1/(-3 + 1/2)},$$

$$c(T_4) = c((2, -3, 3, 1)) = 1 + \frac{1}{3 + 1/(-3 + 1/2)}.$$

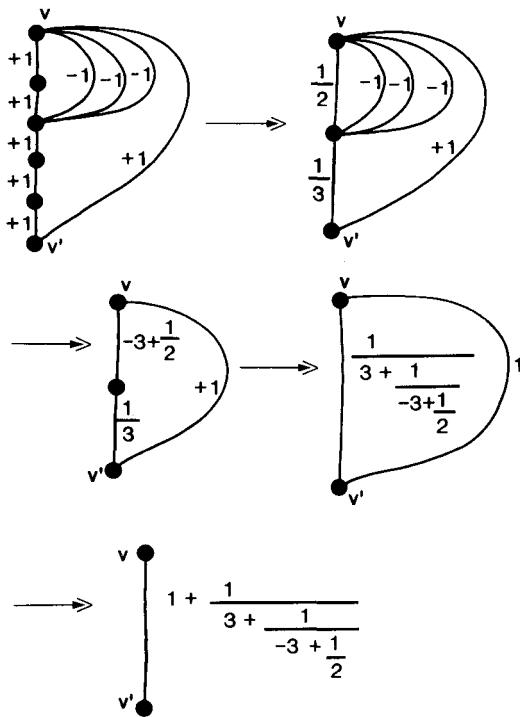


FIGURE 5.5

In general, we associate with the simple rational tangle (t_1, \dots, t_r) the continued fraction

$$t_r + \frac{1}{t_{r-1}} +$$

$$\dots + \frac{1}{t_1}$$

Then, the calculations of our examples immediately generalize and we have, for r even, that the continued fraction equals the conductance of the tangle, and, for r odd, it equals the resistance (the reciprocal of the conductance).

Conway taught us that every rational tangle is equivalent to a (not necessarily unique) simple tangle and that the rational number represented by the continued fraction of a simple tangle is a complete invariant

for ambient isotopy of simple tangles [7, 6], so we have the remarkable fact that *Conway's continued fraction is either the conductance or resistance of the tangle*.

It would be interesting to have a proof of Conway's classification using our electrical ideas. The graphs of simple tangles (with arbitrary positive conductances on their edges) have long been known to electrical engineers as *ladder networks*, and they have used continued fractions to study them [30].

Our electrical concept can be extended to study Conway's notion of an *algebraic tangle* [7, 27]. We only indicate some ideas. The graphs of algebraic tangles correspond to series-parallel networks, well known in electrical theory [9]. Brylawski [4], in his studies of the matroids of series-parallel networks introduced an algebra for describing them. With some minor additions to this notation and some conventions on canonical embeddings of these networks in the plane, Brylawski's algebra can be extended to describe the graphs of algebraic tangles, with "monomials" in this algebra corresponding to simple rational tangles.

6. CLASSICAL TOPOLOGY

In this section we show how our conductance invariant is related to the Alexander-Conway polynomial in the case of tangles. The conductance invariant for the more general class of special tunnel links appears to require further analysis in order to be related to classical topology (if indeed it is related).

Recall [14] that the Conway (Alexander) polynomial is an ambient isotopy invariant of oriented links. It is denoted by $\nabla_K(z) \in Z[z]$ and enjoys the following properties:

- (i) If K is ambient isotopic to K' , then $\nabla_K(z) = \nabla_{K'}(z)$.
- (ii) $\nabla_U(z) = 1$ for the unknot U .
- (iii) If K_+ , K_- , and K_0 are three links differing at the site of one crossing, as shown in Fig. 6.1, then $\nabla_{K_+}(z) - \nabla_{K_-}(z) = z\nabla_{K_0}(z)$.

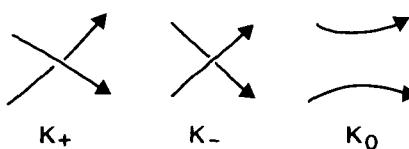


FIGURE 6.1



FIGURE 6.2

Recall also that the *writhe*, $wr(K)$, of an *oriented* link diagram (K can be a tangle as well) is equal to the sum of the *signs of the crossings* of K (defined in Fig. 6.2).

THEOREM. *Let F be a tangle, and $c(F)$ the conductance invariant, as defined in Section 4. Let $n(F)$ and $d(F)$ denote the numerator and denominator of F , as defined in Section 2. Then*

$$c(F) = -i i^{\left[\frac{wr(d(F)) - wr(n(F))}{2} \right]} \frac{\nabla_{n(F)}(2i)}{\nabla_{d(F)}(2i)} .$$

Remark. It is assumed here that the numerator and denominator of F are oriented separately. If F itself can be oriented as in Fig. 6.3a, then both $n(F)$ and $d(F)$ can inherit orientation directly from F (Figs. 6.3b, c). Thus, in this case the theorem simply reads

$$c(F) = -i \frac{\nabla_{n(F)}(2i)}{\nabla_{d(F)}(2i)} .$$

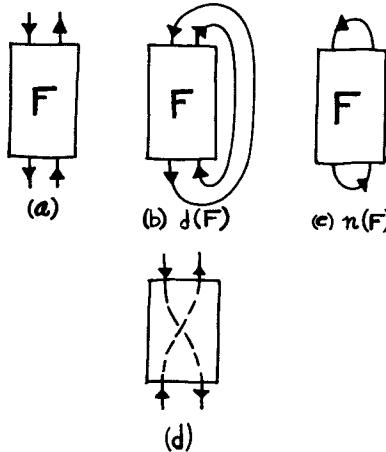


FIGURE 6.3

In general F can be oriented as in Fig. 6.3a or as in Fig. 6.3d, and then it is easy to reorient $d(F)$ so that $(\text{wr}(n(F)) - \text{wr}(d(F)))/2$ is the linking number between two components of $d(F)$. With this convention, the ratio of Conway polynomials is itself an invariant of the tangle.

The properties of $c(F)$ with respect to mirror images of tangles are then seen to correspond to known properties of the Conway polynomial (see [14]), but it is remarkable that they can be elucidated with the much more elementary means of the conductance invariant.

The proof of this theorem will be based on the state model for the Conway polynomial given in [15], plus some properties of the state model for the Jones polynomial in bracket form [16–18]. We shall assume this background for the proof.

Proof of the theorem. We must identify the tree sums of Section 4 with factors in the Conway polynomials of the numerator and denominator of the tangle F . Accordingly, let K be any (oriented) link diagram that is connected. Recall the state sum for $\nabla_K(z)$ given in [15]. In this state sum we have $\nabla_K(z) = \sum_S \langle K|S \rangle$, where S runs over the Alexander states of the diagram K and $\langle K|S \rangle$ is the product of the vertex weights for the state S . These vertex weights are obtained from the crossings in the link diagram by the conventions shown in Fig. 6.4. Each state S designates one of the four quadrants of each crossing. The designated quadrant has a vertex weight of 1, $\pm t$, $\pm t^{-1}$ and $\langle K|S \rangle$ equals the product of these vertex weights. The states are in one-one correspondence with Jordan–Euler trails on the link diagram, and they are also in one-one correspondence with maximal trees in the graph of the diagram.

Now focus on the case $\nabla_K(2i)$. Here $z = t - t^{-1} = i - i^{-1}$, whence $t = i$. We write this symbolically in Fig. 6.5. Familiarity with the bracket polynomial state model [16–18] then shows that this reformulation of the vertex weights implies that, as state sums

$$\nabla_K(2i) = i^{\text{wr}(K)/2} \langle K \rangle (i^{1/2}). \quad (1)$$

Now recall that the clock theorem [15] implies that any two Alexander states can be connected by a series of “clocking” moves. In terms of $G(K)$, the graph of K , a clocking move always removes one edge e from a



FIGURE 6.4

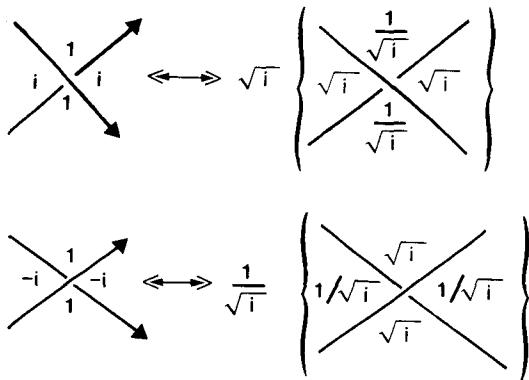


FIGURE 6.5

maximal tree T of $G(T)$ and replaces it with another edge e' to form a new tree T' . It is easy to verify for $\nabla_K(2i)$ that we have the equality $\langle K|S \rangle \text{wt}(e) = \langle K|S' \rangle \text{wt}(e')$, when S' is obtained from S by a single clocking move and $\text{wt}(e), \text{wt}(e')$ are the weights (± 1) assigned to the edges of $G(K)$ by our standard procedure in Section 2 (we used the notation $c(e), c(e')$ in Section 4). We leave the verification of this equation to the reader. It follows at once that for any state S , the product

$$p(K) = \langle K|S \rangle \prod_{e \in T(S)} \text{wt}(e)$$

is *invariant* under clocking (here $T(S)$ is the tree in $G(K)$ determined by the state S); hence it is a constant depending only on the diagram K . We can rewrite this equation as

$$\langle K|S \rangle = p(K) \prod_{e \in T(S)} \text{wt}(e), \quad (2)$$

since $(\text{wt}(e))^2 = 1$. Therefore,

$$\nabla_K(2i) = \sum_S \langle K|S \rangle = p(K) \sum_S \prod_{e \in T(S)} \text{wt}(e),$$

$$\nabla_K(2i) = p(K)w(G(K)),$$

where $w(G(K))$ is the tree sum (weight) defined at the beginning of Section 4.

Returning to our tangle F , we recall that the conductance invariant is given by Eq. 4.12, viz.,

$$c(F) = w(G(n(F)))/w(G(d(F))).$$

Therefore, the theorem will follow from a verification of the formula

$$p(n(F))/p(d(F)) = i \cdot i^{(wr(n(F))-wr(d(F)))/2}.$$

We omit the details of this verification, but point out that it is an application of Eqs. (1) and (2). These equations give a specific expression for $p(K)$, namely,

$$p(K) = i^{wr(K)/2} \left[\prod_{e \in T} wt(e) i^{wt(e)/2} \prod_{e \in T^*} i^{wt(e)/2} \right].$$

Here T is any maximal tree in $G(K)$ and T^* is the corresponding maximal tree in $G(K)^*$, where $G(K)^*$ denotes the planar graph dual to $G(K)$. By $e \in T$ (or $e \in T^*$) we mean that e is an edge of the tree T (or T^*). The formula for $p(n(F))/p(d(F))$ follows from this expression for $p(K)$. This completes the proof.

Remark. It is remarkable to note that, for tangles, our conductance invariant is related to both the Alexander-Conway and the Jones polynomials. This follows here because of the relationship (proved in the course of the theorem above) $\nabla_K(2i) = i^{wr(K)/2} \langle K \rangle(\sqrt{i})$; $\langle K \rangle$ is an unnormalized form of the Jones polynomial. In the usual form $V_K(t)$ for the Jones polynomial [13, 17], this identity becomes $\nabla_K(2i) = V_K(-1)$. This point of coincidence of the Alexander-Conway and the Jones polynomials is of independent interest.

Since the Jones polynomial can be expressed in terms of the signed Tutte polynomial [17], the conductance is also expressible as a special evaluation of the quotient of two signed Tutte polynomials.

APPENDIX: FROM ELECTRICITY TO TREES

It is worthwhile recalling how the calculation of conductance occurs in electrical network theory, based on Ohm's and Kirchoff's laws. Ohm's law states that the electrical potential (voltage drop) between two nodes in a network is equal to the products of the current between these two nodes and the resistance between the two points. This law is expressed as $E = IR$, where E denotes the potential, I is the current, and R is the resistance. Kirchoff's law states that the total sum of currents into and out of any node is zero.

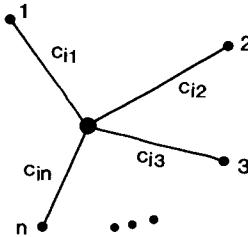


FIGURE A.1

These laws give the local rules for combining conductance. Conductance C is defined to be the reciprocal of resistance, viz., $C = 1/R$. Thus $E = I/C$ and $C = I/E$. Note that one also uses the principle that the voltage drop or potential along a given path in the network is equal to the sum of the voltage drops from node to node along the path.

With these ideas in mind, consider a node i with conductances $c_{i1}, c_{i2}, \dots, c_{in}$ on the edges incident to this node (Fig. A.1). Here c_{ij} labels an edge joining nodes i and j . Then let E_i denote the voltage at node i with respect to some fixed reference node (the *ground*) in the network (the voltages are only determined up to a fixed constant). Then the current in the edge labeled c_{ik} is $(E_i - E_k)c_{ik}$. ($EC = I$, where C is the conductance.) Hence we have by Kirchoff's law,

$$\sum_{k=1}^n (E_i - E_k)c_{ik} = 0.$$

This equation will be true without exception at all nodes but *two*. These special nodes v, v' are the ones where we have set up a current source on a special edge between them as shown in Fig. A.2. We can assume that the

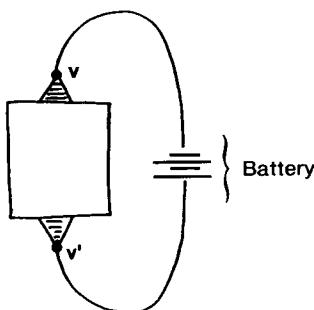


FIGURE A.2

battery edge delivers a fixed current I_0 and voltage E_0 . Then the set of equations for voltages and currents takes the form

$$\begin{pmatrix} -I_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = M \begin{pmatrix} E_0 \\ E_1 \\ E_2 \\ \vdots \\ E_m \end{pmatrix},$$

where the nodes in the graph are labeled $0, 1, \dots, m, 0'$, with 0 labeling v and $0'$ labeling v' . We take $0'$ to be the ground, whence $E_{0'} = 0$. We take M to be the matrix for the system of equations for nodes $0, 1, 2, \dots, m$. Then, if M is invertible, we have

$$\begin{pmatrix} E_0 \\ E_1 \\ E_2 \\ \vdots \\ E_m \end{pmatrix} = M^{-1} \begin{pmatrix} -I_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

M is the famous Kirchoff matrix. It is a remarkable fact that the determinant of M enumerates the spanning trees in G (no battery edge) in the sense that, up to sign, $\det(M)$ is the sum over these trees of the products of the conductances along the edges of the tree. This is the matrix tree theorem (see [29, 25]).

EXAMPLE. Consider the equations for the network of Fig. A.3:

$$(\text{node } 0), -I_0 = (E_1 - E_0)a + (E_2 - E_0)b$$

$$(\text{node } 1), 0 = (E_0 - E_1)a + (E_2 - E_1)c + (E_{0'} - E_1)d$$

$$(\text{node } 2), 0 = (E_0 - E_2)b + (E_1 - E_2)c + (E_{0'} - E_2)e.$$

Thus

$$\begin{pmatrix} -I_0 \\ 0 \\ 0 \end{pmatrix} = M \begin{pmatrix} E_0 \\ E_1 \\ E_2 \end{pmatrix},$$

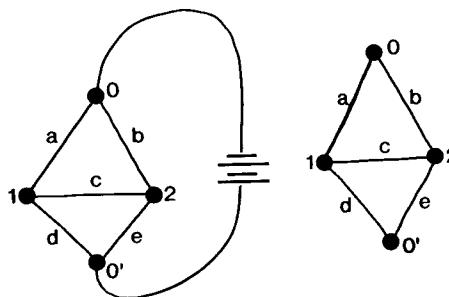


FIGURE A.3

where

$$M = \begin{pmatrix} -a - b & a & b \\ a & -a - c - d & c \\ b & c & -b - c - e \end{pmatrix}.$$

The matrix M is obtained from a matrix K defined for the graph G without the special edge

$$0 \cdot \text{---} \parallel \text{---} \cdot 0'.$$

The matrix K is a node-node matrix with nondiagonal entry k_{ij} equal to the sum of the labels (conductances) on the edges connecting nodes i and j , if such edges exist, and 0 if there is no edge. The diagonal entries k_{ii} are the negative sum of the labels of edges incident to node i . Hence

$$K = \begin{pmatrix} 0 & 1 & 2 & 0' \\ 0 & -a - b & a & b \\ 1 & a & -a - c - d & c \\ 2 & b & c & -b - c - e \\ 0' & 0 & d & e \end{pmatrix}.$$

Thus we see that M is obtained from K by removing the $0'$ row and the $0'$ column.

Now go back to our system of equations and solve for E_0 :

$$\begin{aligned} E_0 &= \frac{\text{Det} \begin{bmatrix} -I_0 & a & b \\ 0 & -a - c - d & c \\ 0 & c & -b - c - e \end{bmatrix}}{\text{Det } M} \\ &= -I_0 \frac{\text{Det} \begin{bmatrix} -a - c - d & c \\ c & -b - c - e \end{bmatrix}}{\text{Det } M}. \end{aligned}$$

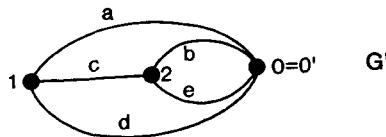


FIGURE A.4

Since $I_0 = CE$, where C is the conductance from v to v' (0 to $0'$), we have

$$C = \frac{-\text{Det } M}{\text{Det} \begin{bmatrix} -a - c - d & c \\ c & -b - c - e \end{bmatrix}}.$$

Let G' be the graph obtained by identifying v with v' (Fig. A.4). The Kirchoff matrix K' is given by

$$K' = \begin{pmatrix} 0 & 1 & 2 \\ 0 & -a - b - d - e & a + d & b + e \\ 1 & a + d & -a - c - d & c \\ 2 & b + e & c & -b - c - e \end{pmatrix}$$

and we take

$$M' = \begin{pmatrix} -a - c - d & c \\ c & -b - c - e \end{pmatrix}.$$

Thus $\text{Det}(M')$ enumerates the trees in G' .

Therefore, we see that the conductance is given by the ratio of tree summations

$$C(G, v, v') = |\text{Det}(M)/\text{Det}(M')| = w(G)/w(G').$$

This is the full electrical background to our combinatorics and topology.

Remark. That the determinants of minors of the Kirchoff matrix enumerate spanning trees in the graph has been the subject of much study. In the electrical context, it is worth mentioning the Wang algebra [8] and the work of Bott and Duffin [9] that clarified some of these issues in terms of Grassmann algebra.

REFERENCES

1. H. AU-YANG AND J. H. H. PERK, Onsager's star-triangle equation: Master key to integrability, in "Integrable Systems in Quantum Field Theory and Statistical Mechanics," Advanced Studies in Pure Mathematics, Vol. 19 pp. 57–94, Academic Press, Boston, 1989.
2. R. J. BAXTER, "Exactly Solved Models in Statistical Mechanics," Academic Press, New York/London, 1982.
3. B. BOLLOBAS, "Graph Theory: An Introductory Course," Graduate Texts in Math, Vol. 63, Springer-Verlag, New York, 1979.
4. T. H. BRYLAWSKI, A combinatorial model for series-parallel networks, *Trans. Am. Math. Soc.* **154** (1971), 1–22.
5. T. H. BRYLAWSKI, A determinantal identity for resistive networks, *SIAM J. Appl. Math.* **32**, No. 2 (1977), 443–449.
6. G. BURDE AND H. ZIESCHANG, "Knots," deGruyter Berlin, 1986.
7. J. H. CONWAY, An enumeration of knots and links, and some of their algebraic properties, in "Computational Problems in Abstract Algebra, Proc. Conference in Oxford 1967," pp. 329–358, Pergamon, Elmsford, NY, 1970.
8. R. J. DUFFIN, An analysis of the Wang algebra of networks, *Trans. Am. Math. Soc.* **93** (1959), 114–131.
9. R. J. DUFFIN, Topology of series parallel networks, *J. Math. Anal. Appl.* **10** (1965), 303–318.
10. G. V. EPIFANOV, Reduction of a plane graph to an edge by a star-triangle transformation, *Soviet. Math. Dokl.* **166** (1966), 13–17.
11. C. ERNST AND D. W. SUMMERS, A calculus for rational tangles—Applications to DNA recombination, *Math. Proc. Cambridge Philos. Soc.* **108** (1990), 489–515.
12. T. A. FEO AND J. S. PROVEN, Delta-wye transformations and the efficient reduction of two-terminal planar graphs, Univ. of North Carolina, Chapel Hill, 1988, preprint.
13. V. F. R. JONES, A polynomial invariant for links via von Neumann algebras, *Bull. Amer. Math. Soc.* **129** (1985), 103–112.
14. L. H. KAUFFMAN, The Conway polynomial, *Topology* **26** (1980), 101–108.
15. L. H. KAUFFMAN, "Formal Knot Theory," Princeton Univ. Press, Mathematical Notes #30, (1983).
16. L. H. KAUFFMAN, "On Knots," Annals of Math. Studies, Vol. 115, Princeton Univ. Press, Princeton, New Jersey, 1987.
17. L. H. KAUFFMAN, New invariants in the theory of knots, *Amer. Math. Monthly* **95**, No. 3 (1988), 195–242.
18. L. H. KAUFFMAN, "Knots and Physics," World Scientific, Singapore, 1991.
19. A. E. KENNELLY, The equivalence of triangles and three pointed stars in conducting networks, *Elec. World Eng.* **34** (1889), 413–414.
20. A. LEHMAN, Wye-delta transformations in probabilistic networks, *J. SIAM* **11** (1963), 773–805.
21. K. REIDEMEISTER, "Knot Theory," BSC Associates, Moscow, ID, 1983; transl. of "Knotentheorie," Springer, Berlin, 1932.
22. T. L. SAATY AND P. C. KAINEN, "The Four Color Theorem: Assaults and Conquest," McGraw-Hill, NY, 1977.
23. S. SESHU AND M. B. REED, "Linear Graphs and Electrical Networks," Addison-Wesley, Reading, MA, 1961.
24. C. E. SHANNON, A symbolic analysis of relay and switching circuits, *Trans. Am. Inst. Elec. Eng.* **57** (1938), 713–723.

25. D. STANTON AND D. WHITE, "Constructive Combinatorics," Undergraduate Texts Math., Springer-Verlag, New York, 1986.
26. D. W. SUMNERS, Untangling DNA, *Math. Intelligencer* **12**, No. 3 (1990), 71–80.
27. M. THISTLETHWAITE, Knot tabulations and related topics, in "Aspects of Topology in Memory of Hugh Dowker" (I. M. James and E. H. Kronheimer, Eds.), London Math. Soc. Lecture Notes Series, No. 93, pp. 1–76, Cambridge Univ. Press, Cambridge, 1984.
28. K. TRUEMPER, On the delta-wye reduction for planar graphs, *J. Graph Theory* **13** (1989), 141–148.
29. W. TUTTE, "Graph Theory," Encyclopedia of Mathematics and Its Applications, Vol. 21, Cambridge Univ. Press, Cambridge, 1984.
30. M. E. VAN VALKENBURG, "Introduction to Modern Network Synthesis," Wiley, New York, 1960.
31. T. YAJIMA AND S. KINOSHITA, On the graphs of knots, *Osaka Math. J.* **9** (1957), 155–163.

Knot Theory and Functional Integration

Louis H. Kauffman

Department of Mathematics, Statistics and Computer Science
University of Illinois at Chicago
851 South Morgan Street
Chicago, IL, 60607-7045

Abstract

This paper is an exposition of the relationship between the heuristics of Witten's functional integral and the theory of knots and links in three dimensional space.

1 Introduction

This paper shows how the Kontsevich Integrals, giving Vassiliev invariants in knot theory, arise naturally in the perturbative expansion of Witten's functional integral. The relationship between Vassiliev invariants and Witten's integral has been known since Bar-Natan's thesis [9] where he discovered, through this connection, how to define Lie algebraic weight systems for these invariants.

The paper is a sequel to [32] and [31] and an expanded version of a talk given at The Fifth Taiwan International Symposium on Statistical Physics (August 1999). See also the work of Labastida and Pérez [37] on this same subject. Their work comes to an identical conclusion, interpreting the Kontsevich integrals in terms of the light-cone gauge and thereby extending the original work of Fröhlich and King [14]. The purpose of this paper is to give an exposition of these relationships and to introduce diagrammatic techniques that illuminate the connections. In particular, we use a diagrammatic

operator method that is useful both for Vassiliev invariants and for relations of this subject with the quantum gravity formalism of Ashtekar, Smolin and Rovelli [3]. This paper also treats the perturbation expansion via three-space integrals leading to Vassiliev invariants as in [9] and [4]. See also [11]. We do not deal with the combinatorial reformulation of Vassiliev invariants that proceeds from the Kontsevich integrals as in [12].

The paper is divided into five sections. Section 2 discusses Vassiliev invariants and invariants of rigid vertex graphs. Section 3 introduces the basic formalism and shows how the functional integral is related directly to Vassiliev invariants. In this section we also show how our formalism works for the loop transform of Ashtekar, Smolin and Rovelli. Section 4 discusses, without gauge fixing, the integral heuristics for the three dimensional perturbative expansion. Section 5 shows how the Kontsevich integral arises in the perturbative expansion of Witten's integral in the light-cone gauge. One feature of section 5 is a new and simplified calculation of the necessary correlation functions by using complex numbers and the two-dimensional Laplacian. We show how the Kontsevich integrals are the Feynman integrals for this theory. In a final section we discuss some of the possibilities of justifying functional integration on formal grounds.

Acknowledgement. It gives the author pleasure to thank Louis Licht, Chris King and Jurg Fröhlich for helpful conversations and to thank the National Science Foundation for support of this research under NSF Grant DMS-9205277 and the NSA for partial support under grant number MSPF-96G-179.

2 Vassiliev Invariants and Invariants of Rigid Vertex Graphs

If $V(K)$ is a (Laurent polynomial valued, or more generally - commutative ring valued) invariant of knots, then it can be naturally extended to an invariant of rigid vertex graphs [26] by defining the invariant of graphs in terms of the knot invariant via an unfolding of the vertex. That is, we can regard the vertex as a "black box" and replace it by any tangle of our choice. Rigid vertex motions of the graph preserve the contents of the black box,

and hence implicate ambient isotopies of the link obtained by replacing the black box by its contents. Invariants of knots and links that are evaluated on these replacements are then automatically rigid vertex invariants of the corresponding graphs. If we set up a collection of multiple replacements at the vertices with standard conventions for the insertions of the tangles, then a summation over all possible replacements can lead to a graph invariant with new coefficients corresponding to the different replacements. In this way each invariant of knots and links implicates a large collection of graph invariants. See [26], [27].

The simplest tangle replacements for a 4-valent vertex are the two crossings, positive and negative, and the oriented smoothing. Let $V(K)$ be any invariant of knots and links. Extend V to the category of rigid vertex embeddings of 4-valent graphs by the formula

$$V(K_*) = aV(K_+) + bV(K_-) + cV(K_0)$$

where K_+ denotes a knot diagram K with a specific choice of positive crossing, K_- denotes a diagram identical to the first with the positive crossing replaced by a negative crossing and K_* denotes a diagram identical to the first with the positive crossing replaced by a graphical node.

This formula means that we define $V(G)$ for an embedded 4-valent graph G by taking the sum

$$V(G) = \sum_S a^{i_+(S)} b^{i_-(S)} c^{i_0(S)} V(S)$$

with the summation over all knots and links S obtained from G by replacing a node of G with either a crossing of positive or negative type, or with a smoothing of the crossing that replaces it by a planar embedding of non-touching segments (denoted 0). It is not hard to see that if $V(K)$ is an ambient isotopy invariant of knots, then, this extension is an rigid vertex isotopy invariant of graphs. In rigid vertex isotopy the cyclic order at the vertex is preserved, so that the vertex behaves like a rigid disk with flexible strings attached to it at specific points.

There is a rich class of graph invariants that can be studied in this manner. The Vassiliev Invariants [52],[10],[8] constitute the important special case of

these graph invariants where $a = +1$, $b = -1$ and $c = 0$. Thus $V(G)$ is a Vassiliev invariant if

$$V(K_*) = V(K_+) - V(K_-).$$

Call this formula the *exchange identity* for the Vassiliev invariant V . See Figure 1

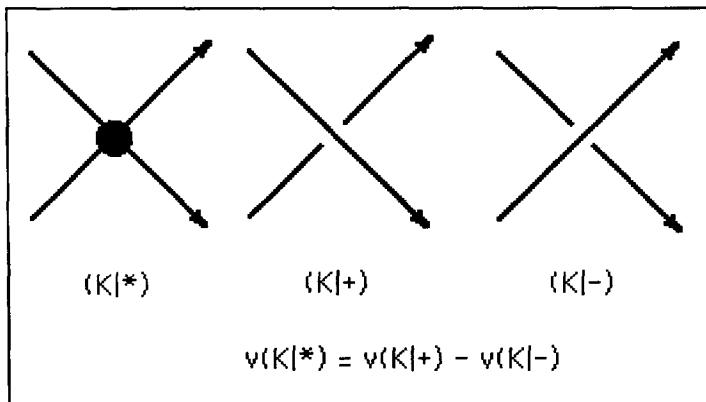


Figure 1 — Exchange Identity for Vassiliev Invariants

V is said to be of *finite type* k if $V(G) = 0$ whenever $|G| > k$ where $|G|$ denotes the number of (4-valent) nodes in the graph G . The notion of finite type is of extraordinary significance in studying these invariants. One reason for this is the following basic Lemma.

Lemma. If a graph G has exactly k nodes, then the value of a Vassiliev invariant v_k of type k on G , $v_k(G)$, is independent of the embedding of G .

Proof. The different embeddings of G can be represented by link diagrams with some of the 4-valent vertices in the diagram corresponding to the nodes of G . It suffices to show that the value of $v_k(G)$ is unchanged under switching of a crossing. However, the exchange identity for v_k shows that this difference

is equal to the evaluation of v_k on a graph with $k + 1$ nodes and hence is equal to zero. This completes the proof.//

The upshot of this Lemma is that Vassiliev invariants of type k are intimately involved with certain abstract evaluations of graphs with k nodes. In fact, there are restrictions (the four-term relations) on these evaluations demanded by the topology and it follows from results of Kontsevich [8] that such abstract evaluations actually determine the invariants. The knot invariants derived from classical Lie algebras are all built from Vassiliev invariants of finite type. All this is directly related to Witten's functional integral [55].

In the next few figures we illustrate some of these main points. In Figure 2 we show how one associates a so-called chord diagram to represent the abstract graph associated with an embedded graph. The chord diagram is a circle with arcs connecting those points on the circle that are welded to form the corresponding graph. In Figure 3 we illustrate how the four-term relation is a consequence of topological invariance. In Figure 4 we show how the four term relation is a consequence of the abstract pattern of the commutator identity for a matrix Lie algebra. This shows that the four term relation is directly related to a categorical generalisation of Lie algebras. Figure 5 illustrates how the weights are assigned to the chord diagrams in the Lie algebra case - by inserting Lie algebra matrices into the circle and taking a trace of a sum of matrix products.

3 Vassiliev Invariants and Witten's Functional Integral

In [55] Edward Witten proposed a formulation of a class of 3-manifold invariants as generalized Feynman integrals taking the form $Z(M)$ where

$$Z(M) = \int DA e^{(ik/4\pi)S(M,A)}.$$

Here M denotes a 3-manifold without boundary and A is a gauge field (also

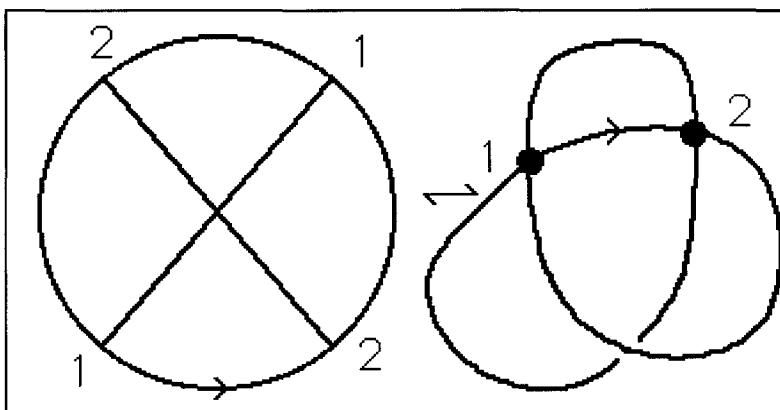


Figure 2 — Chord Diagrams

called a gauge potential or gauge connection) defined on M . The gauge field is a one-form on a trivial G -bundle over M with values in a representation of the Lie algebra of G . The group G corresponding to this Lie algebra is said to be the gauge group. In this integral the action $S(M, A)$ is taken to be the integral over M of the trace of the Chern-Simons three-form $A \wedge dA + (2/3)A \wedge A \wedge A$. (The product is the wedge product of differential forms.)

$Z(M)$ integrates over all gauge fields modulo gauge equivalence (See [5] for a discussion of the definition and meaning of gauge equivalence.)

The formalism and internal logic of Witten's integral supports the existence of a large class of topological invariants of 3-manifolds and associated invariants of knots and links in these manifolds.

The invariants associated with this integral have been given rigorous combinatorial descriptions [46],[51],[34],[40], [53],[29], but questions and conjectures arising from the integral formulation are still outstanding. (See for example [6], [15],[17], [19],[47], [1].) Specific conjectures about this integral take the form of just how it implicates invariants of links and 3-manifolds, and how these invariants behave in certain limits of the coupling constant k in the integral. Many conjectures of this sort can be verified through the

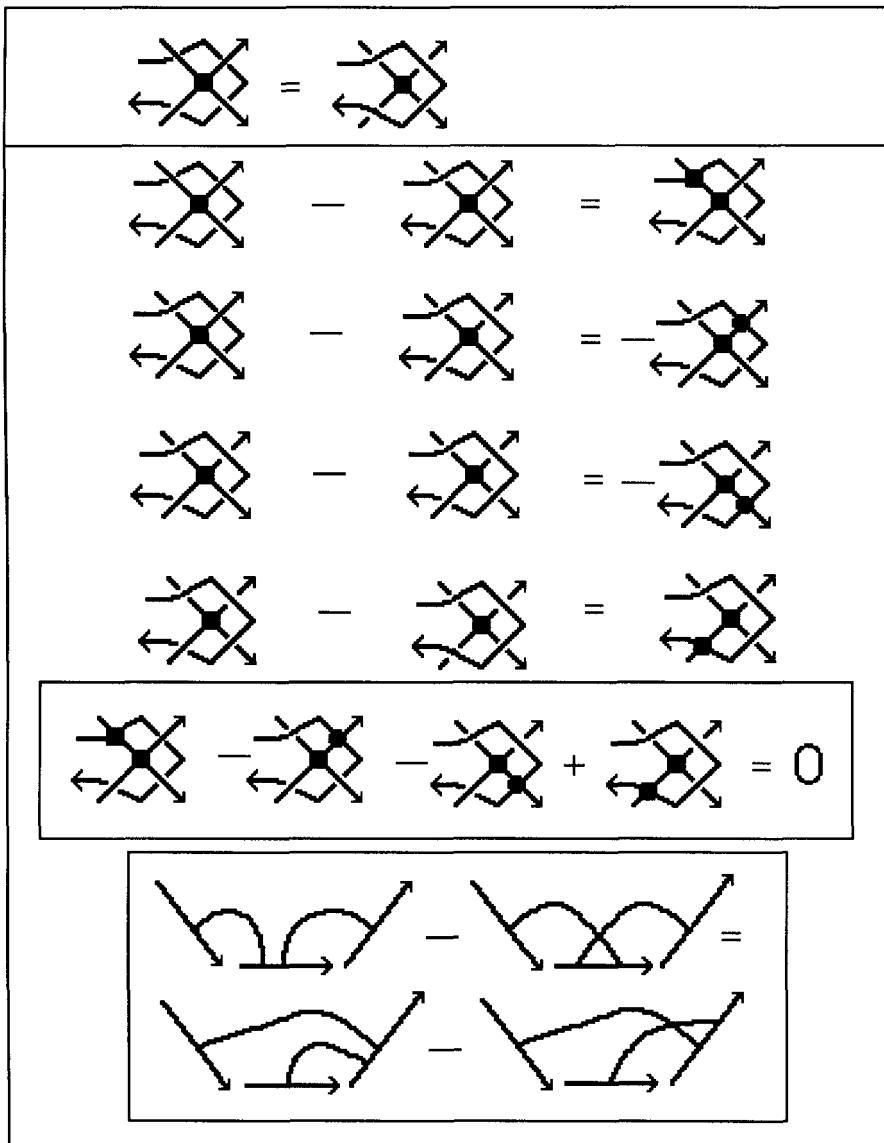


Figure 3 — The Four Term Relation from Topology

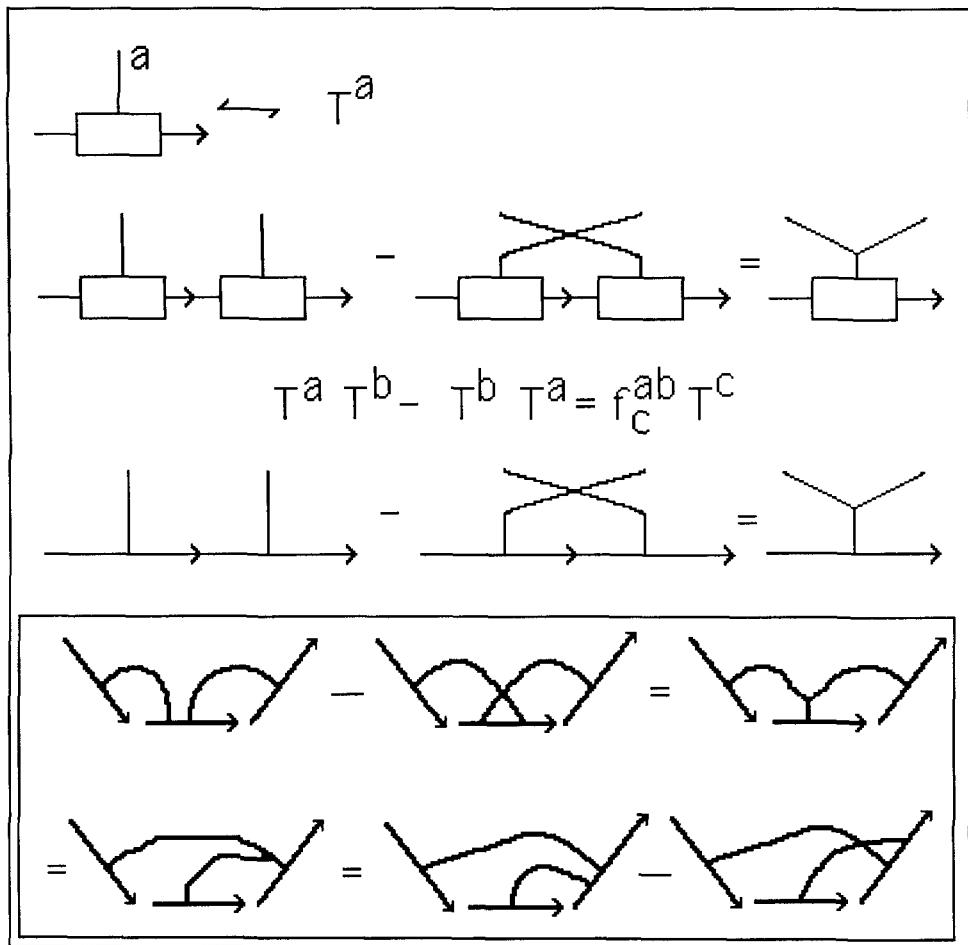


Figure 4 — The Four Term Relation from Categorical Lie Algebra

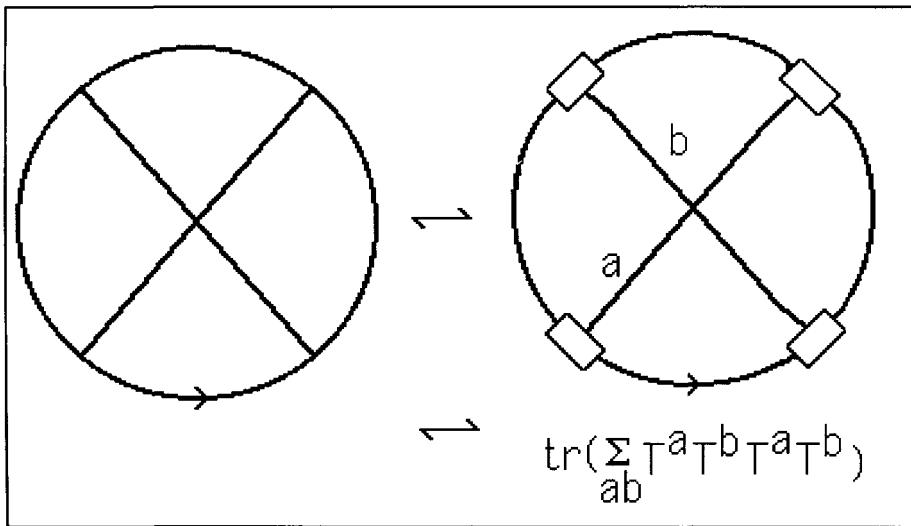


Figure 5 — Calculating Lie Algebra Weights

combinatorial models. On the other hand, the really outstanding conjecture about the integral is that it exists! At the present time there is no measure theory or generalization of measure theory that supports it. Here is a formal structure of great beauty. It is also a structure whose consequences can be verified by a remarkable variety of alternative means.

We now look at the formalism of the Witten integral in more detail and see how it implicates invariants of knots and links corresponding to each classical Lie algebra. In order to accomplish this task, we need to introduce the Wilson loop. The Wilson loop is an exponentiated version of integrating the gauge field along a loop K in three space that we take to be an embedding (knot) or a curve with transversal self-intersections. For this discussion, the Wilson loop will be denoted by the notation $W_K(A) = \langle K | A \rangle$ to denote the dependence on the loop K and the field A . It is usually indicated by the symbolism $\text{tr}(Pe^{\oint_K A})$. Thus

$$W_K(A) = \langle K | A \rangle = \text{tr}(Pe^{\oint_K A}).$$

Here the P denotes path ordered integration - we are integrating and exponentiating matrix valued functions, and so must keep track of the order of the operations. The symbol tr denotes the trace of the resulting matrix.

With the help of the Wilson loop functional on knots and links, Witten writes down a functional integral for link invariants in a 3-manifold M :

$$\begin{aligned} Z(M, K) &= \int DA e^{(ik/4\pi)S(M,A)} \text{tr}(Pe^{\oint_K A}) \\ &= \int DA e^{(ik/4\pi)S} \langle K|A \rangle . \end{aligned}$$

Here $S(M, A)$ is the Chern-Simons Lagrangian, as in the previous discussion. We abbreviate $S(M, A)$ as S and write $\langle K|A \rangle$ for the Wilson loop. Unless otherwise mentioned, the manifold M will be the three-dimensional sphere S^3 .

An analysis of the formalism of this functional integral reveals quite a bit about its role in knot theory. This analysis depends upon key facts relating the curvature of the gauge field to both the Wilson loop and the Chern-Simons Lagrangian. The idea for using the curvature in this way is due to Lee Smolin [48] (See also [43]). To this end, let us recall the local coordinate structure of the gauge field $A(x)$, where x is a point in three-space. We can write $A(x) = A_k^a(x)T_a dx^k$ where the index a ranges from 1 to m with the Lie algebra basis $\{T_1, T_2, T_3, \dots, T_m\}$. The index k goes from 1 to 3. For each choice of a and k , $A_k^a(x)$ is a smooth function defined on three-space. In $A(x)$ we sum over the values of repeated indices. The Lie algebra generators T_a are matrices corresponding to a given representation of the Lie algebra of the gauge group G . We assume some properties of these matrices as follows:

1. $[T_a, T_b] = if^{abc}T_c$ where $[x, y] = xy - yx$, and f^{abc} (the matrix of structure constants) is totally antisymmetric. There is summation over repeated indices.
2. $\text{tr}(T_a T_b) = \delta_{ab}/2$ where δ_{ab} is the Kronecker delta ($\delta_{ab} = 1$ if $a = b$ and zero otherwise).

We also assume some facts about curvature. (The reader may enjoy comparing with the exposition in [28]. But note the difference of conventions on the use of i in the Wilson loops and curvature definitions.) The first fact is the relation of Wilson loops and curvature for small loops:

Fact 1. The result of evaluating a Wilson loop about a very small planar circle around a point x is proportional to the area enclosed by this circle times the corresponding value of the curvature tensor of the gauge field evaluated at x . The curvature tensor is written

$$F_{rs}^a(x)T_a dx^r dy^s.$$

It is the local coordinate expression of $F = dA + A \wedge A$.

Application of Fact 1. Consider a given Wilson line $\langle K|S \rangle$. Ask how its value will change if it is deformed infinitesimally in the neighborhood of a point x on the line. Approximate the change according to Fact 1, and regard the point x as the place of curvature evaluation. Let $\delta \langle K|A \rangle$ denote the change in the value of the line. $\delta \langle K|A \rangle$ is given by the formula

$$\delta \langle K|A \rangle = dx^r dx^s F_a^{rs}(x) T_a \langle K|A \rangle.$$

This is the first order approximation to the change in the Wilson line.

In this formula it is understood that the Lie algebra matrices T_a are to be inserted into the Wilson line at the point x , and that we are summing over repeated indices. This means that each $T_a \langle K|A \rangle$ is a new Wilson line obtained from the original line $\langle K|A \rangle$ by leaving the form of the loop unchanged, but inserting the matrix T_a into that loop at the point x . In Figure 6 we have illustrated this mode of insertion of Lie algebra into the Wilson loop. Here and in further illustrations in this section we use $W_K(A)$ to denote the Wilson loop. Note that in the diagrammatic version shown in Figure 6 we have let small triangles with legs indicate dx^i . The legs correspond to indices just as in our work in the last section with Lie algebras and chord diagrams. The curvature tensor is indicated as a circle with three legs corresponding to the indices of F_a^{rs} .

Notation. In the diagrams in this section we have dropped mention of the factor of $(1/4\pi)$ that occurs in the integral. This convention saves space in the figures. In these figures L denotes the Chern–Simons Lagrangian.

Remark. In thinking about the Wilson line $\langle K|A \rangle = \text{tr}(P e^{\oint_K A})$, it is helpful to recall Euler's formula for the exponential:

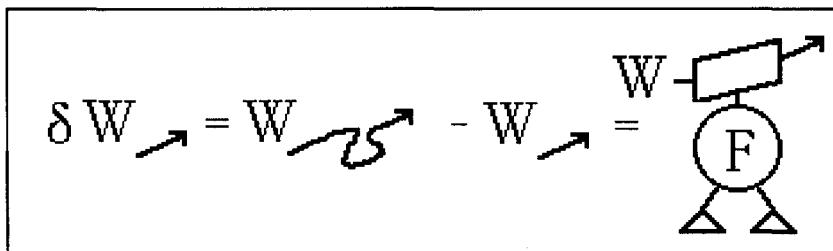


Figure 6 — Lie algebra and Curvature Tensor insertion into the Wilson Loop

$$e^x = \lim_{n \rightarrow \infty} (1 + x/n)^n.$$

The Wilson line is the limit, over partitions of the loop K , of products of the matrices $(1 + A(x))$ where x runs over the partition. Thus we can write symbolically,

$$\begin{aligned} < K | A > &= \prod_{x \in K} (1 + A(x)) \\ &= \prod_{x \in K} (1 + A_k^a(x) T_a dx^k). \end{aligned}$$

It is understood that a product of matrices around a closed loop connotes the trace of the product. The ordering is forced by the one dimensional nature of the loop. Insertion of a given matrix into this product at a point on the loop is then a well-defined concept. If T is a given matrix then it is understood that $T < K | A >$ denotes the insertion of T into some point of the loop. In the case above, it is understood from context in the formula that the insertion is to be performed at the point x indicated in the argument of the curvature.

Remark. The previous remark implies the following formula for the variation of the Wilson loop with respect to the gauge field:

$$\delta \langle K|A \rangle / \delta(A_k^a(x)) = dx^k T_a \langle K|A \rangle .$$

Varying the Wilson loop with respect to the gauge field results in the insertion of an infinitesimal Lie algebra element into the loop. Figure 7 gives a diagrammatic form for this formula. In that Figure we use a capital D with up and down legs to denote the derivative $\delta/\delta(A_k^a(x))$. Insertions in the Wilson line are indicated directly by matrix boxes placed in a representative bit of line.

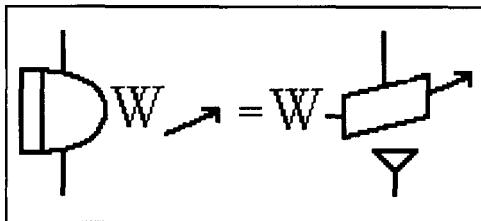


Figure 7 — Differentiating the Wilson Line

Proof.

$$\begin{aligned}
 \delta \langle K|A \rangle / \delta(A_k^a(x)) &= \delta \prod_{y \in K} (1 + A_k^a(y) T_a dy^k) / \delta(A_k^a(x)) \\
 &= \prod_{y < x \in K} (1 + A_k^a(y) T_a dy^k) [T_a dx^k] \prod_{y > x \in K} (1 + A_k^a(y) T_a dy^k) \\
 &= dx^k T_a \langle K|A \rangle .
 \end{aligned}$$

Fact 2. The variation of the Chern-Simons Lagrangian S with respect to the gauge potential at a given point in three-space is related to the values of the curvature tensor at that point by the following formula:

$$F_{rs}^a(x) = \epsilon_{rst} \delta S / \delta(A_t^a(x)).$$

Here ϵ_{abc} is the epsilon symbol for three indices, i.e. it is +1 for positive permutations of 123 and -1 for negative permutations of 123 and zero if any two indices are repeated. A diagrammatic for this formula is shown in Figure 8.

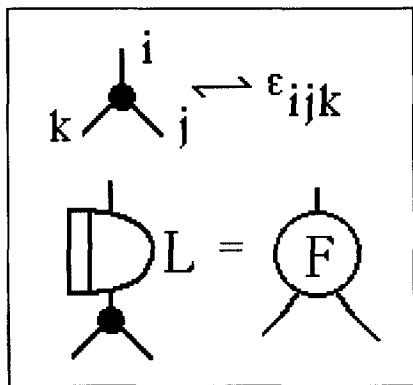


Figure 8 — Variational Formula for Curvature

With these facts at hand we are prepared to determine how the Witten integral behaves under a small deformation of the loop K .

Theorem. 1. Let $Z(K) = Z(S^3, K)$ and let $\delta Z(K)$ denote the change of $Z(K)$ under an infinitesimal change in the loop K . Then

$$\delta Z(K) = (4\pi i/k) \int dA e^{(ik/4\pi)S} [Vol] T_a T_a < K | A >$$

where $Vol = \epsilon_{rst} dx^r dx^s dx^t$.

The sum is taken over repeated indices, and the insertion is taken of the matrices $T_a T_a$ at the chosen point x on the loop K that is regarded as the center of the deformation. The volume element $Vol = \epsilon_{rst} dx_r dx_s dx_t$ is taken with regard to the infinitesimal directions of the loop deformation from this point on the original loop.

2. The same formula applies, with a different interpretation, to the case where x is a double point of transversal self intersection of a loop K , and the deformation consists in shifting one of the crossing segments perpendicularly to the plane of intersection so that the self-intersection point disappears. In this case, one T_a is inserted into each of the transversal crossing segments so that $T_a T_a < K|A >$ denotes a Wilson loop with a self intersection at x and insertions of T_a at $x + \epsilon_1$ and $x + \epsilon_2$ where ϵ_1 and ϵ_2 denote small displacements along the two arcs of K that intersect at x . In this case, the volume form is nonzero, with two directions coming from the plane of movement of one arc, and the perpendicular direction is the direction of the other arc.

Proof.

$$\begin{aligned}
\delta Z(K) &= \int DA e^{(ik/4\pi)S} \delta < K|A > \\
&= \int DA e^{(ik/4\pi)S} dx^r dy^s F_{rs}^a(x) T_a < K|A > \\
&= \int DA e^{(ik/4\pi)S} dx^r dy^s \epsilon_{rst} (\delta S / \delta (A_t^a(x))) T_a < K|A > \\
&= (-4\pi i/k) \int DA (\delta e^{(ik/4\pi)S} / \delta (A_t^a(x))) \epsilon_{rst} dx^r dy^s T_a < K|A > \\
&= (4\pi i/k) \int DA e^{(ik/4\pi)S} \epsilon_{rst} dx^r dy^s (\delta T_a < K|A > / \delta (A_t^a(x))) \\
&\quad (\text{integration by parts and the boundary terms vanish}) \\
&= (4\pi i/k) \int DA e^{(ik/4\pi)S} [Vol] T_a T_a < K|A > .
\end{aligned}$$

This completes the formalism of the proof. In the case of part 2., a change of interpretation occurs at the point in the argument when the Wilson line

is differentiated. Differentiating a self intersecting Wilson line at a point of self intersection is equivalent to differentiating the corresponding product of matrices with respect to a variable that occurs at two points in the product (corresponding to the two places where the loop passes through the point). One of these derivatives gives rise to a term with volume form equal to zero, the other term is the one that is described in part 2. This completes the proof of the Theorem. //

The formalism of this proof is illustrated in Figure 9.

In the case of switching a crossing the key point is to write the crossing switch as a composition of first moving a segment to obtain a transversal intersection of the diagram with itself, and then to continue the motion to complete the switch. One then analyses separately the case where x is a double point of transversal self intersection of a loop K , and the deformation consists in shifting one of the crossing segments perpendicularly to the plane of intersection so that the self-intersection point disappears. In this case, one T_a is inserted into each of the transversal crossing segments so that $T^a T^a < K|A >$ denotes a Wilson loop with a self intersection at x and insertions of T^a at $x + \epsilon_1$ and $x + \epsilon_2$ as in part 2. of the Theorem above. The first insertion is in the moving line, due to curvature. The second insertion is the consequence of differentiating the self-touching Wilson line. Since this line can be regarded as a product, the differentiation occurs twice at the point of intersection, and it is the second direction that produces the non-vanishing volume form.

Up to the choice of our conventions for constants, the switching formula is, as shown below (See Figure 10).

$$\begin{aligned} Z(K_+) - Z(K_-) &= (4\pi i/k) \int DA e^{(ik/4\pi)S} T_a T_a < K_{**}|A > \\ &= (4\pi i/k) Z(T^a T^a K_{**}), \end{aligned}$$

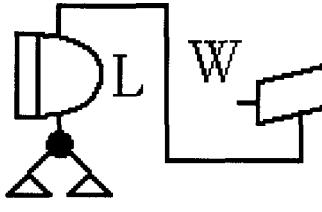
where K_{**} denotes the result of replacing the crossing by a self-touching crossing. We distinguish this from adding a graphical node at this crossing by using the double star notation.

A key point is to notice that the Lie algebra insertion for this difference

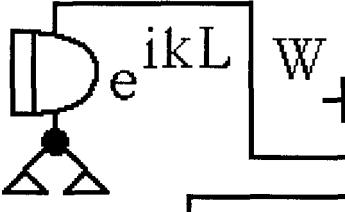
$$\delta Z = \int D A e^{ikL} \delta W$$



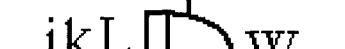
$$= \int D A e^{ikL} \quad \begin{array}{c} \text{F} \\ \triangle \end{array} \quad \begin{array}{c} W \\ \square \end{array}$$



$$= \int D A e^{ikL} \quad \begin{array}{c} L \\ \square \end{array} \quad \begin{array}{c} W \\ \square \end{array}$$



$$= (-i/k) \int D A \quad \begin{array}{c} \square \\ \square \end{array} \quad e^{ikL} \quad \begin{array}{c} W \\ \square \end{array}$$



$$= (i/k) \int D A e^{ikL} \quad \begin{array}{c} \square \\ \square \end{array} \quad \begin{array}{c} W \\ \square \end{array}$$



$$= (i/k) \int D A e^{ikL} \quad \begin{array}{c} \square \\ \square \end{array} \quad \begin{array}{c} W \\ \square \end{array}$$



Figure 9 — Varying the Functional Integral by Varying the Line

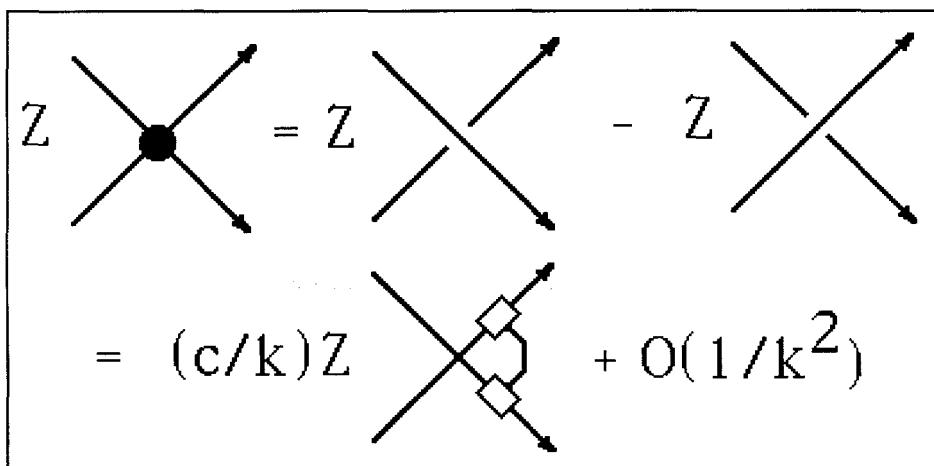


Figure 10 — The Difference Formula

is exactly what is done (in chord diagrams) to make the weight systems for Vassiliev invariants (without the framing compensation). Here we take formally the perturbative expansion of the Witten integral to obtain Vassiliev invariants as coefficients of the powers of $(1/k^n)$. Thus the formalism of the Witten functional integral takes one directly to these weight systems in the case of the classical Lie algebras. In this way the functional integral is central to the structure of the Vassiliev invariants.

3.1 The Loop Transform

Suppose that $\psi(A)$ is a (complex valued) function defined on gauge fields. Then we define formally the *loop transform* $\hat{\psi}(K)$, a function on embedded loops in three dimensional space, by the formula

$$\hat{\psi}(K) = \int DA \psi(A) W_K(A).$$

If Δ is a differential operator defined on $\psi(A)$, then we can use this integral transform to shift the effect of Δ to an operator on loops via integration by

parts:

$$\begin{aligned}\widehat{\Delta\psi}(K) &= \int DA \Delta\psi(A) W_K(A) \\ &= - \int DA \psi(A) \Delta W_K(A).\end{aligned}$$

When Δ is applied to the Wilson loop the result can be an understandable geometric or topological operation. In Figures 11, 12 and 13 we illustrate this situation with diagrammatically defined operators G and H .

We see from Figure 12 that

$$\widehat{G\psi}(K) = \delta\widehat{\psi}(K)$$

where this variation refers to the effect of varying K by a small loop. As we saw in this section, this means that if $\widehat{\psi}(K)$ is a topological invariant of knots and links, then $\widehat{G\psi}(K) = 0$ for all embedded loops K . This condition is a transform analogue of the equation $G\psi(A) = 0$. This equation is the differential analogue of an invariant of knots and links. It may happen that $\delta\widehat{\psi}(K)$ is not strictly zero, as in the case of our framed knot invariants. For example with

$$\psi(A) = e^{(ik/4\pi) \int \text{tr}(A \wedge dA + (2/3)A \wedge A \wedge A)}$$

we conclude that $\widehat{G\psi}(K)$ is zero for flat deformations (in the sense of this section) of the loop K , but can be non-zero in the presence of a twist or curl. In this sense the loop transform provides a subtle variation on the strict condition $G\psi(A) = 0$.

In [3] and earlier publications by these authors, the loop transform is used to study a reformulation and quantization of Einstein gravity. The differential geometric gravity theory is reformulated in terms of a background gauge connection and in the quantization, the Hilbert space consists in functions $\psi(A)$ that are required to satisfy the constraints

$$G\psi = 0$$

and

$$\hat{\Psi}(K) = \int dA \Psi(A) W_K$$

$$\begin{aligned}\widehat{\Delta\Psi}(K) &= \int dA \Delta\Psi(A) W_K \\ &= - \int dA \Psi(A) \Delta W_K\end{aligned}$$

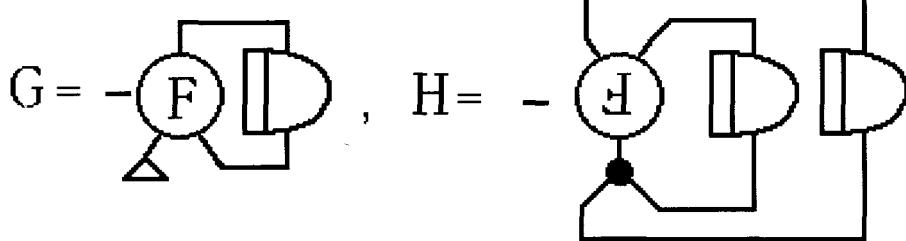
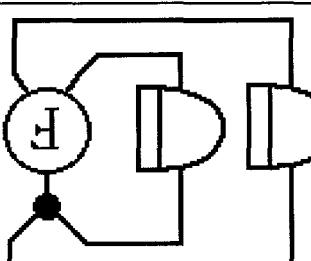
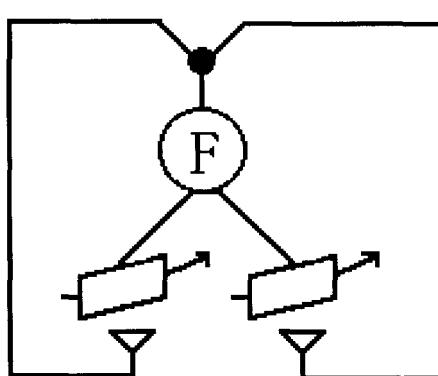


Figure 11—The Loop Transform and Operators G and H

$$\widehat{G}\Psi(\rightarrow) = \int DA G\Psi W_{\rightarrow} = - \int DA \Psi GW_{\rightarrow}$$
$$= \int DA \Psi \begin{array}{c} \text{F} \\ \text{---} \\ \text{W} \end{array}_{\rightarrow}$$
$$= \int DA \Psi \begin{array}{c} \text{F} \\ \text{---} \\ \text{W} \end{array} \begin{array}{c} \rightarrow \\ \text{---} \\ \nabla \end{array} = \int DA \Psi \delta W_{\rightarrow}$$

Figure 12 — The Diffeomorphism Constraint

$$\widehat{H}\Psi(\rightarrow) = \int DA \Psi$$


$$= \int DA \Psi W$$


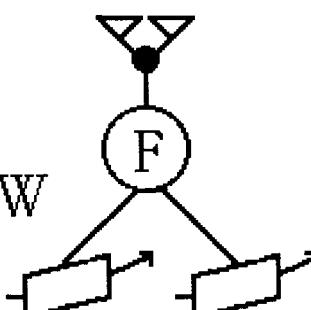
$$= \int DA \Psi W$$


Figure 13 — The Hamiltonian Constraint

$$H\psi = 0$$

where H is the operator shown in Figure 13. Thus we see that $\widehat{G}(K)$ can be partially zero in the sense of producing a framed knot invariant, and (from Figure 13 and the antisymmetry of the epsilon) that $\widehat{H}(K)$ is zero for non-self intersecting loops. This means that the loop transforms of G and H can be used to investigate a subtle variation of the original scheme for the quantization of gravity. This program is being actively pursued by a number of researchers. The Vassiliev invariants arising from a topologically invariant loop transform should be of significance to this theory. This theme will be explored in a subsequent paper.

4 Gaussian Integrals

The purpose of this section is to review well-known facts about Gaussian integration in finite dimensions and to indicate the basic hypothesis that extends these results to functional integrals in infinite dimensions. It is these results on Gaussian integration that form the basis for the next level of heuristics that will be explored in subsequent sections of this paper.

Lets begin by recalling the proof of the following integral formula:

$$\int_{-\infty}^{\infty} e^{-Ax^2/2} dx = \sqrt{2\pi/A}.$$

We prove this formula as follows: Let

$$\Delta = \left(\int_{-\infty}^{\infty} e^{-Ax^2/2} dx \right)^2.$$

Then

$$\begin{aligned} \Delta &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-A(x^2+y^2)/2} dx dy \\ &= \int_0^{2\pi} \int_0^{\infty} e^{-Ar^2/2} r dr d\theta \\ &= 2\pi \int_0^{\infty} e^{-Ar^2/2} d(r^2/2) \end{aligned}$$

$$= 2\pi \int_0^\infty e^{-As} ds \\ = 2\pi/A.$$

Now let

$$Z(J) = \int_{-\infty}^\infty e^{-Ax^2/2+Jx} dx$$

where J is a constant. Then, by completing the square,

$$Z(J) = \int_{-\infty}^\infty e^{-A/2(x-J/A)^2+J^2/2A} dx \\ = \sqrt{2\pi/A} e^{J^2/2A}.$$

Note that

$$\int_{-\infty}^\infty e^{-Ax^2/2} x^n dx = d^n Z(J)/dJ^n|_{J=0} \\ = \sqrt{2\pi/A} d^n (e^{J^2/2A})/dJ^n|_{J=0}.$$

This last derivative can be performed and thus gives a specific algorithm for evaluating the moment integral

$$\int_{-\infty}^\infty e^{-Ax^2/2} x^n dx.$$

Even in this one-variable case, the evaluation of

$$d^N (e^{J^2/2A})/dJ^N|_{J=0}$$

involves non-trivial combinatorics. The reader may enjoy taking on the exercise of proving that

$$d^{2n} (e^{J^2/2A})/dJ^{2n}|_{J=0} = ((2n)!/n!2^n)/A^n.$$

Many ways to do this exercise will naturally lead one to make a few diagrams that help in seeing how to count those terms in

$$d^N (e^{J^2/2A})/dJ^N$$

that contribute to the evaluation at $J = 0$. Diagrams of this kind are precursors to the Feynman diagrams that aid in the evaluation of the descendants and generalizations of these one-dimensional integrals.

The first generalization of the calculations we just made is to go to n variables where n is a positive integer, and to replace A by an $n \times n$ non-singular symmetric matrix of positive determinant. We then want to integrate

$$\int_{R^n} e^{-x A x^t / 2} dx$$

where $x = (x_1, x_2, \dots, x_n)$, x^t denotes the column vector corresponding to the row vector x , and $dx = dx_1 dx_2 \dots dx_n$. The same methods that we have just detailed extend in the context of the linear algebra to give the formula

$$\int_{R^n} e^{-x A x^t / 2} dx = \sqrt{(2\pi)^n / \text{Det}(A)}$$

where $\text{Det}(A)$ denotes the determinant of the matrix A .

Letting $J = (J_1, J_2, \dots, J_n)$ we define

$$Z(J) = \int_{R^n} e^{-x A x^t / 2 + J x^t} dx.$$

Letting

$$\langle x, y \rangle = x A y^t$$

and $J/A = JA^{-1}$,
we have

$$\begin{aligned} Z(J) &= \int_{R^n} e^{-\langle x, x \rangle / 2 + \langle J/A, x \rangle} dx \\ &= \int_{R^n} e^{-\langle x - J/A x, x - J/A x \rangle / 2 + \langle J/A, J/A x \rangle / 2} dx \\ &= \int_{R^n} e^{-\langle x, x \rangle / 2 + \langle J/A, J/A x \rangle / 2} dx \\ &= e^{\langle J/A, J/A \rangle / 2} \int_{R^n} e^{-\langle x, x \rangle / 2} dx \\ &= e^{\langle J/A, J/A \rangle / 2} Z(0) \end{aligned}$$

$$= e^{<J/A, J/A>/2} \sqrt{(2\pi)^n / \text{Det}(A)}.$$

Noting that $< J/A, J/A > = JA^{-1}J^t$ and letting $[x, y] = xA^{-1}y^t$, we see that this formula is the exact analog in n dimensions of the above result in the one dimensional case and we have that correlation functions such as

$$< x_{i1}^{\alpha_1} \dots x_{ik}^{\alpha_k} > = \int_{R^n} e^{-<x, x>/2} x_{i1}^{\alpha_1} \dots x_{ik}^{\alpha_k} dx / Z(0),$$

can be computed by taking successive partial derivatives of $e^{[J, J]/2}$:

$$< x_{i1}^{\alpha_1} \dots x_{ik}^{\alpha_k} > = \partial^{\alpha_1} / \partial x_{i1} \dots \partial^{\alpha_k} / \partial x_{ik} e^{[J, J]/2} |_{J=0}.$$

The next stage is to consider functional integrals in Gaussian form with A replaced by an infinite dimensional operator. The hypothesis for evaluating such integrals is that they can be treated in direct analogy to the finite dimensional case so that the formula

$$Z(J) = e^{[J, J]} Z(0)$$

continues to hold. With this hypothesis in place we continue the specifics of our integral heuristics.

5 The Three-Dimensional Perturbative Expansion

In this section we will first show how linking and self-linking numbers arise from the functional integral, and then how this formalism generalizes to produce combinations of space integrals that express Vassiliev invariants. See [9] for a treatment of the gauge fixing that is relevant to this approach. Here we will follow a heuristic that avoids gauge-fixing altogether. This makes the derivation in this section wholly heuristic even from the physical point of view, but it is worth the experiment since it actually does come up with a correct topological answer as the work of Bott, Taubes [11] and Altschuler-Friedel [4] shows.

Lets begin with a simple vector field $(A_1(x), A_2(x), A_3(x)) = A(x)$ where x is a point in three dimensional Euclidean space R^3 . This is equivalent to

taking the circle ($U(1)$) as gauge group. Now consider fields of the form $A + \phi$ where ϕ is a scalar field (regarded as a fourth coordinate if you like). Define an operator L on $A + \phi$ by the formula

$$L(A + \phi) = -\nabla \times A - \nabla \cdot A + \nabla \phi.$$

then it is easy to see that

$$L^2(A + \phi) = -\nabla^2(A) - \nabla^2(\phi).$$

If we identify A with the one-form $A_1 dx^1 + A_2 dx^2 + A_3 dx^3$, then

$$\int_{R^3} \text{tr}(A \wedge dA) = \int_{R^3} A \cdot (\nabla \times A) dvol$$

where $dvol$ denotes the volume form on R^3 and the A appearing in the right-hand part of the above formula is the vector field version of A . Since we are integrating with respect to volume, the scalar part of $L(A)$ does not matter and we can write

$$\int_{R^3} \text{tr}(A \wedge dA) = \int_{R^3} A \cdot L(A) dvol$$

and take as a definition of a quadratic form on vector fields

$$\langle A, B \rangle = \int_{R^3} A \cdot B dvol$$

so that

$$\langle A, LA \rangle = \int_{R^3} \text{tr}(A \wedge dA).$$

With this in mind, consider the following functional integral defined for a link of two components K and K' :

$$\Lambda(K, K') = \int DA e^{-(1/2)\langle A, LA \rangle} \int \int_{(x,y) \in K \times K'} A(x) A(y).$$

Regarding this as a Gaussian integral, we need to determine the inverse of the operator L . Within the context of the quadratic form we have $L^2 = -\nabla^2$. We know that the Green's function

$$G = (-1/4\pi)(1/|x - y|)$$

is the inverse of the three dimensional Laplacian.

$$\nabla^2 G(x) = \delta(x - y)$$

where $\delta(x - y)$ denotes the Dirac delta function.

Thus if $J(y)$ stands for an arbitrary vector field, and we define

$$G * J = \int_{R^3} dy G(x - y) J(y),$$

then

$$\begin{aligned} (\nabla^2 G) * J(x) &= \int_{R^3} dy \nabla^2 G(x - y) J(y) \\ &= \int_{R^3} dy \delta(x - y) J(y) = J(x). \end{aligned}$$

Thus

$$(L^2 G) * J = J.$$

Hence

$$L((LG) * J) = J.$$

Hence

$$L^{-1} J = (LG) * J.$$

It is then easy to calculate that

$$\langle J, L^{-1} J \rangle = (1/4\pi) \int \int_{K \times K'} (J(x) \times J(y)) \cdot (x - y) / |x - y|^3.$$

Now returning to our functional integral we have

$$\begin{aligned} \Lambda(K, K') &= \int DA e^{-(1/2)\langle A, LA \rangle} \int \int_{(x,y) \in K \times K'} A(x) A(y) \\ &= \int \int_{K \times K'} \int DA e^{(-1/2)\langle A, LA \rangle} A(x) A(y) \\ &= \int \int_{K \times K'} \partial/\partial J(x) \partial/\partial J(y)|_{J=0} e^{(1/2)\langle J, L^{-1} J \rangle}. \end{aligned}$$

The last equality, being the basic hypothesis for evaluating the functional integral.

Using our results on L^{-1} and a few calculations suppressed here, we conclude that

$$\begin{aligned}\Lambda(K, K') &= (1/4\pi) \int \int_{K \times K'} (dx \times dy) \cdot (x - y) / |x - y|^3 \\ &= Lk(K, K').\end{aligned}$$

Here $L(K, K')$ denotes the linking number of the curves K and K' . The last equality is the statement of Gauss's integral representation for the linking number of two space curves.

Now consider the Witten functional integral for arbitrary gauge group. Then one can check that a formula of the following type expresses the trace of the Chern-Simons differential form. Here f_{abc} denotes the structure constants for the Lie algebra.

$$tr(AdA + (2/3)A^3) = (\epsilon^{ijk}/2) [\sum_a A_i^a \partial_j A_k^a + (1/3) \sum_{a,b,c} A_i^a A_j^b A_k^c f_{abc}] dvol$$

This means that we can think of this trace as a sum of a Gaussian term directly analogous to the term for the linking number (but involving the Lie algebra indices) and a separate term corresponding to the A^3 term in the Chern-Simons form. We want the coefficient of the Gaussian term in the integral to be constant (independent of the coupling constant k). To accomplish this, we replace A by A/\sqrt{k} and obtain

$$Z_K = \int DA e^{(i/4\pi) \int tr(AdA)} e^{(i/4\pi\sqrt{k}) \int tr((2/3)A^3)} \sum_n (1/\sqrt{k})^n \int \int \dots \int_{K_1 < K_2 < \dots < K_n} A$$

Here we have used an expansion of the Wilson loop as a sum of integrated integrals:

$$W_K(A) = tr \prod_{x \in K} (1 + A/\sqrt{k})$$

$$= \sum_n (1/\sqrt{k})^n \int \int \dots \int_{K_1 < K_2 < \dots < K_n} A^n$$

The notation $K_1 < K_2 < \dots < K_n$ means that in the integration process we choose points x_i from K so that $x_1 < x_2 < \dots < x_n$ and the term A^n in the integration formula denotes $A(x_1)A(x_2)\dots A(x_n)$. With this reformulation of Z_K , we see that the full integral can be considered as a Gaussian integral with correlations corresponding to the products of copies of A in the iterated integrals and other products involving the A^3 term. We will explicate this integration heuristically, without doing any gauge fixing.

Using

$$e^{(i/4\pi\sqrt{k}) \int \text{tr}((2/3)A^3)} = \sum_m (1/m!)(i/6\pi\sqrt{k})^m (\int \text{tr}(A^3))^m,$$

substituting and rewriting, we have

$$Z_K = \sum_m \sum_n (1/m!)(i/6\pi)^m (1/\sqrt{k})^{m+n} C_{mn}$$

where

$$C_{mn} = \int DA e^{(i/4\pi) \int \text{tr}(AdA)} (\int \text{tr}(A^3)^m \int \int \dots \int_{K_1 < K_2 < \dots < K_n} A^n$$

The term C_{mn} shows that we will have to consider correlations combined from the Wilson loop and from the A^3 term. The quadratic form $\int \text{tr}(AdA)$ is a direct generalization of the quadratic form that we explicated at the beginning of the section and the inverse operator is obtained by applying the Laplacian to the same Green's function. As a result, the Feynman diagrams for these correlations have triple vertices in space (corresponding to the A^3 term) and single vertices on the knot or link (corresponding to an A in the iterated integral). Each triple vertex contributes $1/\sqrt{k}$ and each vertex on the link also contributes $1/\sqrt{k}$. At a triple vertex there is a factor proportional to

$$\epsilon^{ijk} f_{abc}$$

(coming from the expression we gave earlier in this section for $\text{tr}(AdA + (2/3)A^3)$). At a single vertex (corresponding to one A), there is an insertion of Lie algebra in the line (i.e. the knot or link), and along each edge in the Feynman diagram there is a Gauss kernel (the propagator) $i/4\pi\delta_{ab}\epsilon^{ijk}(x-y)_k/|x-y|^3$. This is tied into the relevant indices which are the indices of tangent directions along the lines of the link and the factors at the triple vertices. For a given order of diagram one integrates the positions of the vertices along the link and the positions of the triple vertices in three dimensional space to obtain the contribution of the diagram to the correlator. In this way, a given C_{mn} contributes to the order $(1/\sqrt{k})^{m+n}$ of the perturbative expansion of the integral. By taking an appropriate sum of the C_{mn} for $m+n=2N$ one obtains an explicit formula for the N -th order Vassiliev invariants arising from the functional integral.

6 Wilson Lines, Light-Cone Gauge and the Kontsevich Integrals

In this section we follow the gauge fixing method used by Fröhlich and King [14]. Their paper was written before the advent of Vassiliev invariants, but contains, as we shall see, nearly the whole story about the Kontsevich integral. A similar approach to ours can be found in [37]. In our case we have simplified the determination of the inverse operator for this formalism and we have given a few more details about the calculation of the correlation functions than is customary in physics literature. I hope that this approach makes this subject more accessible to mathematicians. A heuristic argument of this kind contains a great deal of valuable mathematics. It is clear that these matters will eventually be given a fully rigorous treatment. In fact, in the present case there is a rigorous treatment, due to Albevario and Sen-Gupta [2] of the functional integral *after* the light-cone gauge has been imposed.

Let (x^0, x^1, x^2) denote a point in three dimensional space. Change to light-cone coordinates

$$x^+ = x^1 + x^2$$

and

$$x^- = x^1 - x^2.$$

Let t denote x^0 .

Then the gauge connection can be written in the form

$$A(x) = A_+(x)dx^+ + A_-(x)dx^- + A_0(x)dt.$$

Let $CS(A)$ denote the Chern-Simons integral (over the three dimensional sphere)

$$CS(A) = (1/4\pi) \int tr(A \wedge dA + (2/3)A \wedge A \wedge A).$$

We define *axial gauge* (light-cone gauge) to be the condition that $A_- = 0$. We shall now work with the functional integral of the previous section under the axial gauge restriction. In axial gauge we have that $A \wedge A \wedge A = 0$ and so

$$CS(A) = (1/4\pi) \int tr(A \wedge dA).$$

Letting ∂_{\pm} denote partial differentiation with respect to x^{\pm} , we get the following formula in axial gauge

$$A \wedge dA = (A_+ \partial_- A_0 - A_0 \partial_- A_+) dx^+ \wedge dx^- \wedge dt.$$

Thus, after integration by parts, we obtain the following formula for the Chern-Simons integral:

$$CS(A) = (1/2\pi) \int tr(A_+ \partial_- A_0) dx^+ \wedge dx^- \wedge dt.$$

Letting ∂_i denote the partial derivative with respect to x_i , we have that

$$\partial_+ \partial_- = \partial_1^2 - \partial_2^2.$$

If we replace x^2 with ix^2 where $i^2 = -1$, then $\partial_+ \partial_-$ is replaced by

$$\partial_1^2 + \partial_2^2 = \nabla^2.$$

We now make this replacement so that the analysis can be expressed over the complex numbers.

Letting

$$z = x^1 + ix^2,$$

it is well known that

$$\nabla^2 \ln(z) = 2\pi\delta(z)$$

where $\delta(z)$ denotes the Dirac delta function and $\ln(z)$ is the natural logarithm of z . Thus we can write

$$(\partial_+ \partial_-)^{-1} = (1/2\pi)\ln(z).$$

Note that $\partial_+ = \partial_z = \partial/\partial z$ after the replacement of x^2 by ix^2 . As a result we have that

$$(\partial_-)^{-1} = \partial_+(\partial_+ \partial_-)^{-1} = \partial_+(1/2\pi)\ln(z) = 1/2\pi z.$$

Now that we know the inverse of the operator ∂_- we are in a position to treat the Chern-Simons integral as a quadratic form in the pattern

$$(-1/2) \langle A, LA \rangle = -iCS(A)$$

where the operator

$$L = \partial_-.$$

Since we know L^{-1} , we can express the functional integral as a Gaussian integral:

We replace

$$Z(K) = \int DA e^{ikCS(A)} \text{tr}(Pe^{\oint_K A})$$

by

$$Z(K) = \int DA e^{iCS(A)} \text{tr}(Pe^{\oint_K A/\sqrt{k}})$$

by sending A to $(1/\sqrt{k})A$. We then replace this version by

$$Z(K) = \int DA e^{(-1/2)\langle A, LA \rangle} \text{tr}(Pe^{\oint_K A/\sqrt{k}}).$$

In this last formulation we can use our knowledge of L^{-1} to determine the correlation functions and express $Z(K)$ perturbatively in powers of $(1/\sqrt{k})$.

Proposition. Letting

$$\langle \phi(A) \rangle = \int DA e^{(-1/2)\langle A, LA \rangle} \phi(A) / \int DA e^{(-1/2)\langle A, LA \rangle}$$

for any functional $\phi(A)$, we find that

$$\begin{aligned}\langle A_+^a(z, t) A_+^b(w, s) \rangle &= 0, \\ \langle A_0^a(z, t) A_0^b(w, s) \rangle &= 0, \\ \langle A_+^a(z, t) A_0^b(w, s) \rangle &= \kappa \delta^{ab} \delta(t-s)/(z-w)\end{aligned}$$

where κ is a constant.

Proof Sketch. Let's recall how these correlation functions are obtained. The basic formalism for the Gaussian integration is in the pattern

$$\begin{aligned}\langle A(z) A(w) \rangle &= \int DA e^{(-1/2)\langle A, LA \rangle} A(z) A(w) / \int DA e^{(-1/2)\langle A, LA \rangle} \\ &= ((\partial/\partial J(z))(\partial/\partial J(w))|_{J=0}) e^{(1/2)\langle J, L^{-1}J \rangle}\end{aligned}$$

Letting $G * J(z) = \int dw G(z-w) J(w)$, we have that when

$$LG(z) = \delta(z)$$

($\delta(z)$ is a Dirac delta function of z .) then

$$LG * J(z) = \int dw LG(z-w) J(w) = \int dw \delta(z-w) J(w) = J(z)$$

Thus $G * J(z)$ can be identified with $L^{-1}J(z)$.

In our case

$$G(z) = 1/2\pi z$$

and

$$L^{-1}J(z) = G * J(z) = \int dw J(w)/(z-w).$$

Thus

$$\begin{aligned} < J(z), L^{-1}J(z) > &= < J(z), G \ast J(z) > = (1/2\pi) \int tr(J(z)(\int dw J(w)/(z-w))d \\ &= (1/2\pi) \int \int dz dw tr(J(z)J(w))/(z-w). \end{aligned}$$

The results on the correlation functions then follow directly from differentiating this expression. Note that the Kronecker delta on Lie algebra indices is a result of the corresponding Kronecker delta in the trace formula $tr(T_a T_b) = \delta_{ab}/2$ for products of Lie algebra generators. The Kronecker delta for the $x^0 = t, s$ coordinates is a consequence of the evaluation at J equal to zero. //

We are now prepared to give an explicit form to the perturbative expansion for

$$\begin{aligned} < K > &= Z(K)/ \int DA e^{(-1/2)< A, LA >} \\ &= \int DA e^{(-1/2)< A, LA >} tr(P e^{\oint_K A/\sqrt{k}})/ \int DA e^{(-1/2)< A, LA >} \\ &= \int DA e^{(-1/2)< A, LA >} tr(\prod_{x \in K} (1 + (A/\sqrt{k}))) / \int DA e^{(-1/2)< A, LA >} \\ &= \sum_n (1/k^{n/2}) \oint_{K_1 < \dots < K_n} < A(x_1) \dots A(x_n) >. \end{aligned}$$

The latter summation can be rewritten (Wick expansion) into a sum over products of pair correlations, and we have already worked out the values of these. In the formula above we have written $K_1 < \dots < K_n$ to denote the integration over variables x_1, \dots, x_n on K so that $x_1 < \dots < x_n$ in the ordering induced on the loop K by choosing a basepoint on the loop. After the Wick expansion, we get

$$< K > = \sum_m (1/k^m) \oint_{K_1 < \dots < K_n} \sum_{P = \{x_i < x'_i | i=1, \dots, m\}} \prod_i < A(x_i) A(x'_i) >.$$

Now we know that

$$\langle A(x_i)A(x'_i) \rangle = \langle A_k^a(x_i)A_l^b(x'_i) \rangle T_a T_b dx^k dx^l.$$

Rewriting this in the complexified axial gauge coordinates, the only contribution is

$$\langle A_+^a(z, t)A_0^b(s, w) \rangle = \kappa \delta^{ab} \delta(t - s)/(z - w).$$

Thus

$$\begin{aligned} & \langle A(x_i)A(x'_i) \rangle \\ &= \langle A_+^a(x_i)A_0^a(x'_i) \rangle T_a T_a dx^+ \wedge dt + \langle A_0^a(x_i)A_+^a(x'_i) \rangle T_a T_a dx^+ \wedge dt \\ &= (dz - dz')/(z - z')[i/i'] \end{aligned}$$

where $[i/i']$ denotes the insertion of the Lie algebra elements $T_a T_a$ into the Wilson loop.

As a result, for each partition of the loop and choice of pairings $P = \{x_i < x'_i | i = 1, \dots, m\}$ we get an evaluation D_P of the trace of these insertions into the loop. This is the value of the corresponding chord diagram in the weight systems for Vassiliev invariants. These chord diagram evaluations then figure in our formula as shown below:

$$\langle K \rangle = \sum_m (1/k^m) \sum_P D_P \oint_{K_1 < \dots < K_n} \bigwedge_{i=1}^m (dz_i - dz'_i)/((z_i - z'_i))$$

This is a Wilson loop ordering version of the Kontsevich integral. To see the usual form of the integral appear, we change from the time variable (parametrization) associated with the loop itself to time variables associated with a specific global direction of time in three dimensional space that is perpendicular to the complex plane defined by the axial gauge coordinates. It is easy to see that this results in one change of sign for each segment of the knot diagram supporting a pair correlation where the segment is oriented (Wilson loop parameter) downward with respect to the global time direction. This results in the rewrite of our formula to

$$\langle K \rangle = \sum_m (1/k^m) \sum_P (-1)^{|P|} D_P \int_{t_1 < \dots < t_n} \bigwedge_{i=1}^m (dz_i - dz'_i)/((z_i - z'_i))$$

where $|P \downarrow|$ denotes the number of points (z_i, t_i) or (z'_i, t_i) in the pairings where the knot diagram is oriented downward with respect to global time. The integration around the Wilson loop has been replaced by integration in the vertical time direction and is so indicated by the replacement of $\{K_1 < \dots < K_n\}$ with $\{t_1 < \dots < t_n\}$

The coefficients of $1/k^m$ in this expansion are exactly the Kontsevich integrals for the weight systems D_P . See Figure 14.

It was Kontsevich's insight to see (by different means) that these integrals could be used to construct Vassiliev invariants from arbitrary weight systems satisfying the four-term relations. Here we have seen how these integrals arise naturally in the axial gauge fixing of the Witten functional integral.

Remark. The careful reader will note that we have not made a discussion of the role of the maxima and minima of the space curve of the knot with respect to the height direction (t). In fact one has to take these maxima and minima very carefully into account and to divide by the corresponding evaluated loop pattern (with these maxima and minima) to make the Kontsevich integral well-defined and actually invariant under ambient isotopy (with appropriate framing correction as well). The corresponding difficulty appears here in the fact that because of the gauge choice the Wilson lines are actually only defined in the complement of the maxima and minima and one needs to analyse a limiting procedure to take care of the inclusion of these points in the Wilson line. This points to one of the places where this correspondence with the Kontsevich integrals as Feynman integrals for Witten's functional integral could stand closer mathematical scrutiny. One purpose of this paper has been to outline the correspondences that exist and to put enough light on the situation to allow a full story to eventually appear.

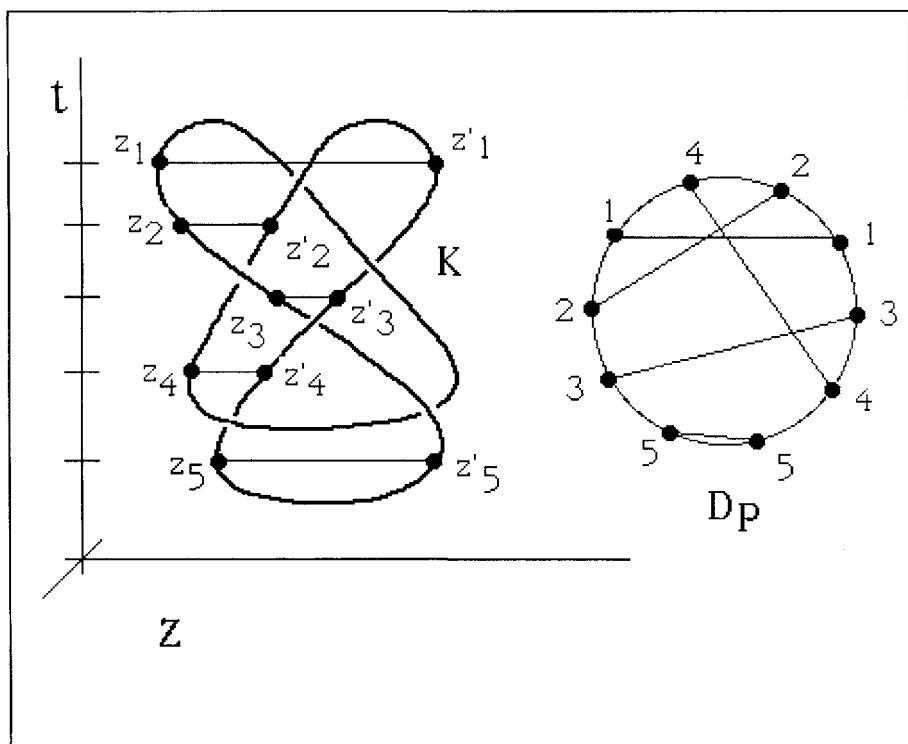


Figure 14 — Applying The Kontsevich Integral

7 Formal Integration

In light of the fact that many beautiful and rigorous results come forth from the formal manipulation of functional integrals, it is of interest to attempt to see whether one can create a formal (essentially combinatorial) category in which these manipulations can exist. This section is devoted to some very elementary remarks along this line.

Suppose that we start with an extended real line in the sense of Robinson's non-standard analysis [45], [18]. (For a delightful introduction to a similar construction of an extended real line see also Conway's work on Surreal Numbers [13].) In this line there are infinitesimals δ each less than any standard real and the extended line is still a field. Can we make formal integrals that make sense of Leibniz's $\int_a^b f(x)dx$ as a "sum" over infinitely many small quantities?

The method for the Robinson approach is to define a Riemann sum $S(\Delta x) = \sum_{i=0}^n f(x_i)\Delta x$ where $\{x_0, \dots, x_n\}$ is a partition of the interval $[a, b]$ and $n\Delta x = b - a$. The Riemann sum is regarded as a function of Δx , and then by the Robinson logic of transferring statements about reals to statements about hyperreals, one can let Δx be infinitesimal to define the integral.

Here, I want to note that there is a very formal counterpart to this process that lets one define an indefinite "micro-integral" by a straightforward formula where δ is an infinitesimal:

$$F(x) = f(x)\delta + f(x - \delta)\delta + f(x - 2\delta)\delta + \dots$$

$F(x)$ is a formal representative for $\int^x f(t)dt$. In particular, let's compute the derivative of F :

$$F(x + \delta) = f(x + \delta)\delta + f(x)\delta + f(x - \delta)\delta + f(x - 2\delta)\delta + \dots$$

Thus

$$(F(x + \delta) - F(x))/\delta = f(x + \delta).$$

Every finite hyperreal is a unique sum of a real and an infinitesimal. The derivative of a function is defined to the real part of the difference quotient using δ . It follows that

$$F'(x) = f(x).$$

In this sense, we can directly write down a formal antiderivative for a real-valued function $f(x)$.

By using this same construction, we can write down a micro-integral that represents the indefinite integral of a real-value function:

$$\int f(x)dx := \Sigma_{n=-\infty}^{n=+\infty} f(n\delta)\delta.$$

Here $\int f(x)dx$ is the micro analog of the integral of $f(x)$ from negative to positive infinity, but it does not have a specific real value and may be an infinitesimal itself. We assume that if δ' is any infinitesimal such that $\delta' = a\delta$ where a is a standard real number, then

$$\Sigma_{n=-\infty}^{n=+\infty} f(n\delta')\delta' = \Sigma_{n=-\infty}^{n=+\infty} f(n\delta)\delta.$$

This should be called *the basic principle of micointegration*. It defines an equivalence relation on the (non-numerical) values of these integrals. But we need more to actually model an integral from minus infinity to infinity. The best we can do with a micointegral is model the intgegration between two finite points. For suppose that δ has the property that $\omega\delta = a$ for some finite real number a where ω is the first infinite number (the limit as n goes to infinity, if you will). Then the above micointegral represents the integral

$$\int_{-a}^a f(x)dx = \Sigma_{n=-\omega}^{n=+\omega} f(n\delta)\delta.$$

In order to get the integral from minus infinity to infinity, we will have to add up a countable number of these micointegrals, one for each interval in the real line. With this in place, we can make a formal model for that integration and it will follow from the basic principle of micointegration that

$$\int_b^a f(cx)dx = \int_{b/c}^{a/c} f(x)dx.$$

It is very interesting in the light of these constructions to note how the Leibniz notation makes everything about the formalities of integration quite automatic. In Leibniz notation we take dx to denote the infinitesimal δ in the summation, and we assume the formal rules

$$d(x + c) = dx$$

when c is a constant, and

$$d(ax) = ad(x)$$

when a is a constant. (Constants are fixed standard reals.) Consequently

$$\int f(x + c)dx = \int f(x)dx.$$

The combinatorics of microintegrals or the usual work with limits make these facts into properties to be verified.

How far are we on the road toward writing down a formal version of a functional integral in gauge field theory? Is there a combinatorial formula for

$$Z(M, K) = \int DA e^{(ik/4\pi)S(M, A)} \text{tr}(Pe^{\oint_K A})$$

utilizing an orchestration of *countable* sums of infinitesimals from the hyperreals? If we do make such a construction it may, like our toy example $F(x)$, be itself infinitesimal, or even infinite. The problem is to make it well-defined in the hyperreals and so that the typical results about integrating Gaussians are logical consequences of its construction. That is the question!

Since I do not have an answer to this question at this writing, let me add a few more ideas. There is another approach to the infinitesimal calculus due to Lawvere and Bell [7] that uses *square zero infinitesimals*. These are infinitesimals ϵ such that $\epsilon^2 = 0$. Of course the extension of the reals to these new hyperreals is no longer a field. Furthermore these structures are usually wrapped in a cloak of intuitionistic mathematics and category theory (topos theory – see [7] and [41]). The intuitionism comes from the desire that ϵ be indistinguishable from zero and yet not zero. Such subtlety is possible in a world where the double negative is distinct from an affirmation.

The square zero infinitesimals are charming. In this system, the formula

$$f(x + \epsilon) = f(x) + f'(x)\epsilon$$

defines the derivative of $f(x)$, and all the familiar properties of derivatives follow easily.

Now square zero infinitesimals remind the differential geometer immediately of Grassmann algebra and exterior differential forms. Actually there are two significant choices that one can make at this fundamental level. One can choose to assume that distinct square zero infinitesimals commute with one another, and do not necessarily annihilate each other when multiplied. Or one can assume that they do not necessarily commute, but that the squares of any finite sum of them is itself of square zero. Let us consider these two choices one at a time.

In the case of commuting square zero infinitesimals α and β , we have

$$\begin{aligned} (\alpha + \beta)^2 &= \alpha^2 + 2\alpha\beta + \beta^2 \\ &= 2\alpha\beta \end{aligned}$$

while

$$(\alpha + \beta)^3 = 0.$$

In general, the commutative situation forces higher and higher orders of nilpotencey for sums of independent infinitesimals. An *infinite sum* of square zero infinitesimals can stand for a standard Robinson infinitesimal. This indicates that the Robinson theory should be seen as the limit of square zero theories. This is the analog of thinking of a Taylor series in terms of its truncations.

In the case of all sums having square zero and non-commutation, we recover an infinitesimal version of Grassmann algebra:

$$\begin{aligned} 0 &= (\alpha + \beta)^2 = \alpha^2 + \alpha\beta + \beta\alpha + \beta^2 \\ &= \alpha\beta + \beta\alpha. \end{aligned}$$

Thus

$$0 = \alpha\beta + \beta\alpha.$$

I take these elementary observations as hints that the correct version of calculus for our purposes will contain infinitesimals that partake of all these options, and that we should press ahead and create the theory that contains them.

References

- [1] D. Adams, The semiclassical approximation for the Chern-Simons partition function, *Phys. Lett. B* 417 (1998), pp. 53-60.
- [2] S. Albevario and A. Sen-Gupta, A Mathematical Construction of the Non-Abelian Chern-Simons Functional Integral, *Commun. Math. Phys.*, Vol. 186 (1997), pp. 563-579.
- [3] Ashtekar, Abhay, Rovelli, Carlo and Smolin, Lee [1992], "Weaving a Classical Geometry with Quantum Threads", *Phys. Rev. Lett.*, vol. 69, p. 237.
- [4] Daniel Altschuler and Laurent Freidel, Vassiliev knot invariants and Chern-Simons perturbation theory to all orders, *Commun. Math. Phys.* 187 (1997), 261-287.
- [5] M.F. Atiyah, *Geometry of Yang-Mills Fields*, Accademia Nazionale dei Lincei Scuola Superiore Lezioni Fermiare, Pisa ,1979.
- [6] M.F. Atiyah, *The Geometry and Physics of Knots*, Cambridge University Press, 1990.
- [7] J. L. Bell, *A Primer of Infinitesimal Calculus*, Cambridge University Press, 1998.
- [8] D. Bar-Natan, On the Vassiliev knot invariants, *Topology* 34 (1995), 423-472.

- [9] Dror Bar-Natan, *Perturbative Aspects of the Chern-Simons Topological Quantum field Theory*, Ph. D. Thesis, Princeton University, June 1991.
- [10] J. Birman and X.S.Lin, Knot polynomials and Vassilievs invariants, *Invent. Math.* **111** No. 2 (1993), 225-270.
- [11] R. Bott and C. Taubes, On the self-linking of knots, *Jour. Math. Phys.* **35** (1994), pp. 5247-5287.
- [12] P. Cartier, Construction combinatoire des invariants de Vassiliev - Kontsevich des noeuds, *C. R. Acad. Sci. Paris*, **316**, Serie I, (1993), pp. 1205-1210.
- [13] J. H. Conway, *On Numbers and Games*, Academic Press, 1976.
- [14] J. Fröhlich and C. King, The Chern Simons Theory and Knot Polynomials, *Commun. Math. Phys.* **126** (1989), 167-199.
- [15] S. Garoufalidis, Applications of $TQFT$ to invariants in low dimensional topology, (preprint 1993).
- [16] E. Guadagnini, M. Martellini and M. Mintchev, Chern-Simons model and new relations between the Homfly coefficients, *Physics Letters B*, Vol. 238, No. 4, Sept. 28 (1989), pp. 489-494.
- [17] D. S. Freed and R. E. Gompf, Computer calculation of Witten's three-manifold invariant, *Commun. Math. Phys.*, No. 141, (1991), pp. 79-117.
- [18] J. M. Henle and E. W. Kleinberg, *Infinitesimal Calculus*, MIT Press, 1980.
- [19] L.C. Jeffrey, Chern-Simons-Witten invariants of lens spaces and torus bundles, and the semi-classical approximation, *Commun. Math. Phys.*, No. 147, (1992), pp. 563-604.
- [20] V.F.R. Jones, A polynomial invariant of links via von Neumann algebras, *Bull. Amer. Math. Soc.*, 1985, No. 129, pp. 103-112.
- [21] V.F.R.Jones, Hecke algebra representations of braid groups and link polynomials, *Ann. of Math.*, Vol.126, 1987, pp. 335-338.

- [22] V.F.R.Jones, On knot invariants related to some statistical mechanics models, *Pacific J. Math.*, Vol. 137, no. 2,1989, pp. 311-334.
- [23] L.H. Kauffman, *On Knots*, Annals Study No. 115, Princeton University Press (1987)
- [24] L.H.Kauffman, State Models and the Jones Polynomial, *Topology*,Vol. 26, 1987,pp. 395-407.
- [25] L.H.Kauffman, Statistical mechanics and the Jones polynomial, *AMS Contemp. Math. Series*, Vol. 78,1989, pp. 263-297.
- [26] L.H.Kauffman, New invariants in the theory of knots, *Amer. Math. Monthly*, Vol.95,No.3,March 1988. pp 195-242.
- [27] L.H.Kauffman and P.Vogel, Link polynomials and a graphical calculus, *Journal of Knot Theory and Its Ramifications*, Vol. 1, No. 1,March 1992, pp. 59- 104.
- [28] L.H.Kauffman, *Knots and Physics*, World Scientific Pub.,1991 and 1993
- [29] L. H. Kauffman and S. L. Lins, *Temperley-Lieb Recoupling Theory and Invariants of 3- Manifolds*, Annals of Mathematics Study 114, Princeton Univ. Press,1994.
- [30] L. H. Kauffman, Functional Integration and the theory of knots, *J. Math. Physics*, Vol. 36 (5), May 1995, pp. 2402 - 2429.
- [31] L. H. Kauffman, Witten's Integral and the Kontsevich Integrals, in *Particles, Fields, and Gravitation*, Proceedings of the Lodz, Poland (April 1998) Conference on Mathematical Physics edited by Jakub Remblinski, AIP Conference Proceedings 453 (1998), pp. 368 -381.
- [32] L. H. Kauffman Knot Theory and the heuristics of functional integration, *Physica A* 281 (2000), 173-200.
- [33] R. Kirby, A calculus for framed links in S^3 , *Invent. Math.*, **45**, (1978), pp. 35-56.
- [34] R. Kirby and P. Melvin, On the 3-manifold invariants of Reshetikhin-Turaev for $sl(2, C)$, *Invent. Math.* 105, pp. 473-545,1991.

- [35] M. Kontsevich, Graphs, homotopical algebra and low dimensional topology, (preprint 1992).
- [36] T. Kohno, Linear representations of braid groups and classical Yang-Baxter equations, *Contemporary Mathematics*, Vol. 78, Amer. Math. Soc., 1988, pp. 339-364.
- [37] J. M. F. Labastida and E. Pérez, Kontsevich Integral for Vassiliev Invariants from Chern-Simons Perturbation Theory in the Light-Cone Gauge, *J. Math. Phys.*, Vol. 39 (1998), pp. 5183-5198.
- [38] T.Q.T. Le and J. Murakami, Universal finite type invariants of 3-manifolds, (preprint 1995.)
- [39] R. Lawrence, Asymptotic expansions of Witten-Reshetikhin-Turaev invariants for some simple 3- manifolds. *J. Math. Phys.*, No. 36(11), November 1995, pp. 6106-6129.
- [40] W.B.R.Lickorish, The Temperley Lieb Algebra and 3-manifold invariants, *Journal of Knot Theory and Its Ramifications*, Vol. 2,1993, pp. 171-194.
- [41] S. MacLane and L. Moerdijk, *Sheaves in Geometry and Logic*, Springer-Verlag, 1992.
- [42] Michael Polyak and Oleg Viro, Gauss diagram formulas for Vassiliev invariants, *Intl. Math. Res. Notices*, No. 11, (1994) pp. 445-453.
- [43] P. Cotta-Ramusino,E.Guadagnini,M.Martellini,M.Mintchev, Quantum field theory and link invariants, *Nucl. Phys. B* **330**, Nos. 2-3 (1990), pp. 557-574
- [44] N.Y.Reshetikhin, Quantized universal enveloping algebras, the Yang-Baxter equation and invariants of links, I and II, *LOMI reprints E-4-87 and E-17-87*, Steklov Institute, Leningrad, USSR.
- [45] Abraham Robinson, *Non-Standard Analysis*, Princeton University Press, 1996, 1965.

- [46] N.Y. Reshetikhin and V. Turaev, Invariants of Three Manifolds via link polynomials and quantum groups, *Invent. Math.*, Vol.103,1991, pp. 547-597.
- [47] L. Rozansky, Witten's invariant of 3-dimensional manifolds: loop expansion and surgery calculus, In *Knots and Applications*, edited by L. Kauffman, (1995), World Scientific Pub. Co.
- [48] Lee Smolin, Link polynomials and critical points of the Chern-Simons path integrals, *Mod. Phys. Lett. A*, Vol. 4,No. 12, 1989, pp. 1091-1112.
- [49] V.G.Turaev, The Yang-Baxter equations and invariants of links, LOMI preprint E-3-87, Steklov Institute, Leningrad, USSR., *Inventiones Math.*,Vol. 92, Fasc.3, pp. 527-553.
- [50] V.G.Turaev and O. Y. Viro, State sum invariants of 3-manifolds and quantum $6j$ symbols, *Topology* 31, (1992), pp. 865-902.
- [51] V.G. Turaev and H. Wenzl, Quantum invariants of 3-manifolds associated with classical simple Lie algebras, *International J. of Math.*, Vol. 4, No. 2,1993, pp. 323-358.
- [52] V. Vassiliev, Cohomology of knot spaces, In *Theory of Singularities and Its Applications*, V.I.Arnold, ed., Amer. Math. Soc.,1990, pp. 23-69.
- [53] K. Walker, On Witten's 3-Manifold Invariants, (preprint 1991).
- [54] J. H. White, Self-linking and the Gauss integral in higher dimensions. *Amer. J. Math.*, **91** (1969), pp. 693-728.
- [55] E. Witten, Quantum field Theory and the Jones Polynomial, *Commun. Math. Phys.*,vol. 121, 1989, pp. 351-399.

Series on Knots and Everything – Vol. 1

KNOTS AND PHYSICS

(3rd Edition)

This invaluable book is an introduction to knot and link invariants as generalised amplitudes for a quasi-physical process. The demands of knot theory, coupled with a quantum-statistical framework, create a context that naturally and powerfully includes an extraordinary range of interrelated topics in topology and mathematical physics. The author takes a primarily combinatorial stance toward knot theory and its relations with these subjects. This stance has the advantage of providing direct access to the algebra and to the combinatorial topology, as well as physical ideas.

The book is divided into two parts: Part I is a systematic course on knots and physics starting from the ground up, and Part II is a set of lectures on various topics related to Part I. Part II includes topics such as frictional properties of knots, relations with combinatorics, and knots in dynamical systems.

In this third edition, a paper by the author entitled "Knot Theory and Functional Integration" has been added. This paper shows how the Kontsevich integral approach to the Vassiliev invariants is directly related to the perturbative expansion of Witten's functional integral. While the book supplies the background, this paper can be read independently as an introduction to quantum field theory and knot invariants and their relation to quantum gravity. As in the second edition, there is a selection of papers by the author at the end of the book. Numerous clarifying remarks have been added to the text.

ISBN 981-02-4112-7 (pbk)

9 789810 241124