Nankai Tracts in Mathematics Vol. 12

TOPOLOGY AND PHYSICS

Editors

Kevin Lin
Zhenghan Wang
Weiping Zhang

Proceedings of the
Nankai International Conference
in Memory of Xiao-Song Lin

World Scientific

TOPOLOGY
AND
PHYSICS

Proceedings of the Nankai International Conference
in Memory of Xiao-Song Lin

NANKAI TRACTS IN MATHEMATICS

Series Editors: Yiming Long and Weiping Zhang
Chern Institute of Mathematics

Published

Vol. 1 Scissors Congruences, Group Homology and Characteristic Classes
by J. L. Dupont

Vol. 2 The Index Theorem and the Heat Equation Method
by Y. L. Yu

Vol. 3 Least Action Principle of Crystal Formation of Dense Packing Type and Kepler's Conjecture
by W. Y. Hsiang

Vol. 4 Lectures on Chern–Weil Theory and Witten Deformations
by W. P. Zhang

Vol. 5 Contemporary Trends in Algebraic Geometry and Algebraic Topology
edited by Shiing-Shen Chern, Lei Fu & Richard Hain

Vol. 6 Riemann–Finsler Geometry
by Shiing-Shen Chern & Zhongmin Shen

Vol. 7 Iterated Integrals and Cycles on Algebraic Manifolds
by Bruno Harris

Vol. 8 Minimal Submanifolds and Related Topics
by Yuanlong Xin

Vol. 9 Introduction to Several Complex Variables: Three Methods of Several Complex Variables
by Yum-Tong Siu et al.

Vol. 10 Differential Geometry and Physics
edited by Mo-Lin Ge & Weiping Zhang

Vol. 11 Inspired by S S Chern
edited by Phillip A. Griffiths

Vol. 12 Topology and Physics
edited by Kevin Lin, Zhenghan Wang & Weiping Zhang

Nankai Tracts in Mathematics – Vol. 12

TOPOLOGY AND PHYSICS

Proceedings of the Nankai International Conference in Memory of Xiao-Song Lin

Tianjin, China 27 – 31 July 2007

Editors

Kevin Lin
University of California at Berkeley, USA

Zhenghan Wang
Microsoft Research Station Q, USA

Weiping Zhang
Chern Institute of Mathematics, China

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

TOPOLOGY AND PHYSICS
Proceedings of the Nankai International Conference in Memory of Xiao-Song Lin

ISBN-13 978-981-281-910-9
ISBN-10 981-281-910-X

Printed in Singapore by World Scientific Printers

Dedicated to the memory of Professor Xiao-Song Lin

July 27, 1957—Jan. 14, 2007

Independence,
confidence,
and dignity!

The most
cherished is the
light given off
when we cross
paths.

Xiao-Song Lin
Dec. 7, 2006

独立、自信、和尊严！
最可贵的是在这交会中
互发的光亮。

林晓松书
二〇〇六年
十二月七日

FOREWORD

This volume through its diversity reflects Lin's journal-editing side. In addition to the mathematics he did he always read widely and with good taste in and around all the lovely terrain which neighbors on topology. Topology, in this way is a good subject for an eclectic. There is almost no domain from condensed matter physics to p-adic analysis that one cannot find a topological excuse to study. Lin, the journal editor, did this. I hope he would find this volume interesting and thoughtful enough that he might have put it together himself.

Lin was, with Fred Hickling, one of my first two graduate students. So he and I figured out the thesis advising process together. I needed a lot of help. I had had an unusual education and had missed some of the steps. To my great relief Lin seemed to know how it would go. He would talk to me about what I was working on, focus in on something that I was not understanding properly and then dig in. As with most great students the advisor has little to do with the thesis. Lin had a great idea, which he developed with Habegger, and is further expanded in this volume. The idea was very reductionist, he thought the subject of knots and links had begun in the wrong place, as if physicists had tried to solve for H_2^+ without doing the hydrogen atom first. His idea was to study "string links" : arcs in a ball with fixed boundary conditions, rather than ordinary links in S^3. First thing first. Do the local problem before going global. It was a very sensible idea and amazing that knot theory had existed 100 years without this idea popping up. I think, 20 years later in 2008, this idea would be considered obvious. But this is a sign that Lin and a few like him exerted a systematic influence: localizing and systematizing geometric topology. If you like making topology more like quantum field theory.

Quantum field theory is another area where Lin was in the vanguard. The idea that some (but not all) classical link invariants are naturally described as coefficients of a perturbative expansion was sorted out between Bar-Natan and Lin in the early 1990s leading quickly to the theory of finite type invariants. It was a thrill for me to see how Lin took to the then new subject of quantum topology. At first I did not expect to learn this subject myself. However in the mid 90s I was thinking about building a computer based on the Chern-Simons Lagrangian and Lin became my tutor, explaining the Jones representations, and all the related algebras. He made the subject very friendly. For over a decade he would stay in touch what he was doing and thinking in the "Jones" world. His work in this area is among the finest and this volume serves, also, as a tribute to these contributions.

Finally, it is not well known but Lin made a serious effort to understand Perelman's proof of Thurston's geometrization conjecture and the many ancillary expositions. The article herein on generalized Ricci flow commemorates his efforts both as a researcher and journal editor here.

Lin was a great and generous spirit, well loved by our community. He was our dear friend. We miss him and dedicate our work to him.

Michael H. Freedman

PREFACE

On January 14, 2007, our beloved friend Xiao-Song Lin left us. On his 50th birthday—July 27, 2007, his friends, colleagues and family members gathered in the Chern Institute of Mathematics to celebrate his wonderful life. These proceedings resulted from this conference, and is a permanent tribute to a humble person, an excellent mathematician, a great friend, and a devoted family man.

Topology and physics are central themes in Xiao-Song's professional career. A central player in quantum topology is knot theory. Knot invariants such as the celebrated Jones polynomial and finite type invariants were constantly on Xiao-Song's mind. As one of the leading quantum knot theorists in the world, Xiao-Song made fundamental contributions to the development and popularization of knot theory. With his untimely death, the knot world lost a leader.

During the international conference from July 27 to July 31, 2007 at the Chern Institute of Mathematics, Xiao-Song's friends and colleagues covered a variety of topics in topology and physics. We are sure that Xiao-Song will smile in heaven when the topics dear to his heart continue to flourish.

We thank all the participants and speakers for making the conference a memorable one. Staff in the Chern Institute won the hearts of the participants for staying on top of everything. We thank all of them, especially Mrs. Hongqin Li.

Zhenghan Wang
Weiping Zhang

Short Biography of Lin

Xiao-Song Lin, a Professor of Mathematics at the University of California at Riverside, died on January 14, 2007 in Riverside, California, six months after being diagnosed with advanced stage liver cancer. He was 49.

Xiao-Song Lin was born in Songjiang, Shanghai, on July 27, 1957, and grew up in Suzhou, Jiangsu. In 1984, he received his M.S. in Mathematics from Beijing University under the direction of Professor Boju Jiang. That same year, he arrived in the United States to study at the University of California, San Diego under Professor Michael H. Freedman. After obtaining his Ph.D. in 1988, he began his career at Columbia University. In 1995, he joined the faculty at the University of California, Riverside, where he remained until the time of his death.

Xiao-Song Lin was a mathematician of exceptional ability and creativity. His areas of specialization were in low-dimensional topology and quantum topology. He was best known for his numerous contributions to knot theory. Throughout his entire career, Xiao-Song Lin maintained a passionate commitment to mathematical research and the mathematics community. He was co-founder and co-Editor-in-Chief of the research journal Communications in Contemporary Mathematics, and served on the editorial boards of several others. He advised five Ph.D. students, and he served as a mentor to many other graduate and post-doctoral students in topology. He will always be remembered by his students and his colleagues for his patience, his generosity, and his willingness to share mathematical ideas.

Xiao-Song Lin received many honors and awards, including the prestigious Sloan Fellowship (1992-1994); he was a member of the Institute for Advanced Study (Spring 1988 and 1993-1994); he was a Professor of Special Mathematics Lectures at Beijing University (1998-2000); and he was named Beijing University's Chang Jiang Scholar (2006-2008) by the Chinese Ministry of Education.

Despite his employment in the USA, Xiao-Song Lin was actively involved in the advancement of Chinese mathematics and kept in close contact with the topology research group at Beijing University. Beginning in the early 1990s, he spent most of his summers in China, primarily giving lectures and teaching classes at his alma mater Beijing University. Together with Professor Boju Jiang and Professor Shicheng Wang, he helped to plan and organize the annual Chinese Low-Dimensional Topology Summer School, the 2002 ICM Satellite Conference in Geometric Topology, as well as many other mathematical meetings and conferences in China.

His untimely death is a great loss to the international topology community, and to all who knew him. In his honor and in his memory, the Xiao-Song Lin Award was established by Xiao-Song Lin's family: Each year, a cash prize of at least 1000 USD will be awarded to a senior undergraduate at Beijing University who has demonstrated truly exceptional scholarship in mathematics.

Mathematics of Lin

Xiao-Song's first major work was a joint paper with Prof. M. Freedman on the A-B slice problem. In 1981, Freedman solved the 4-dimensional topological Poincaré conjecture. Actually, he achieved a complete classification of all closed simply-connected topological 4-manifolds. Later, Freedman proved that his method, in principle, works for a large class of fundamental groups including all finitely generated abelian groups. But he conjectured that his method will not cover the cases of non-abelian free groups. The A-B slice problem is a program to prove Freedman's conjecture. This difficult problem is still open. Prof. V. Krushkal's paper in this book gives an up-to-date account of this problem. The experience that Xiao-Song gained from this project strongly influenced his career and future research. In this paper, link homotopy was introduced into the study of 4-dimensional topology. Freedman recently wrote on this work: "Sometimes when I think of our work on that problem, I feel like an old time mountaineer stormed off a high peak just short of the summit."

After the A-B slice problem, it was natural for Xiao-Song to study link homotopy. In a joint work with Prof. N. Habegger, he solved a problem of Prof. J. Milnor from 1950s on the classification of links up to homotopy, where the notion "string link" was invented.

These two beautiful papers were essentially written during his graduate school years. After obtaining his Ph.D, he went to work at Columbia University. There in a joint paper with Prof. J. Birman, he axiomatized Prof. V. Vassiliev's knot invariants combinatorically, then expanded the Jones polynomial of knots into Vassiliev or finite type invariants. The Birman-Lin condition, which characterized finite type invariants, was discovered in this work. Soon afterwards, a second revolution in quantum knot theory after Prof. V. Jones' first one started.

Xiao-Song's personal favorite work was his paper *A knot invariant via representation spaces*. Following an idea of Prof. A. Casson, Xiao-Song defined a knot invariant, which turned out to be the knot signature. Recently, this work was generalized using symplectic Floer homology by Prof. W. Li.

The paper *Representations of knot groups and twisted Alexander polynomials* shows something that we are familiar with and grateful: Xiao-Song's generosity in mathematics. He introduced the twisted Alexander polynomials in this paper, but never rushed to publish the paper for the sake of credit. Actually this paper would never have been published if not solicited by an editor.

His unfinished work *Zeros of Jones polynomials* had captured his attention for a long time. Since the beginning of quantum knot invariants, Jones realized that his knot polynomial is related to statistical mechanics. In physics, the zeros of partition functions encode deep information about the corresponding physical systems. In his unfinished manuscript, Xiao-Song asked: how can one tell whether or not a Laurent polynomial with integer coefficients is the Jones polynomial of a knot? Then he wrote: maybe this was the wrong question. If so, what would be the right question? He believed that the right question would lead to some beautiful mathematics. At the end of the manuscript, he suggested to look for statistical laws for the norms and phases of the zeros of the Jones polynomials.

There are 38 publications by Xiao-Song listed in Mathematical Reviews so far, and 8 more papers are posted on the arXiv. Another 5 unpublished papers are collected in this volume. In addition, there are 4 more unfinished works on: the zeros of Jones polynomials, an L^2-approach to the volume conjecture, wood puzzle games, and finger loop braids, respectively, that can be found on Xiao-Song's webpage `http: //math.ucr.edu/~xl/`.

We cannot do justice to Xiao-Song's mathematics in a few pages. As Prof. D. Bar-Natan's talk title in Nankai said: following Lin. Then we are bound to discover beautiful gems in mathematics.

Quoting Freedman again: "There will ever after be string links and finite type invariants." We add: there will ever after be the name of Xiao-Song Lin in mathematics.

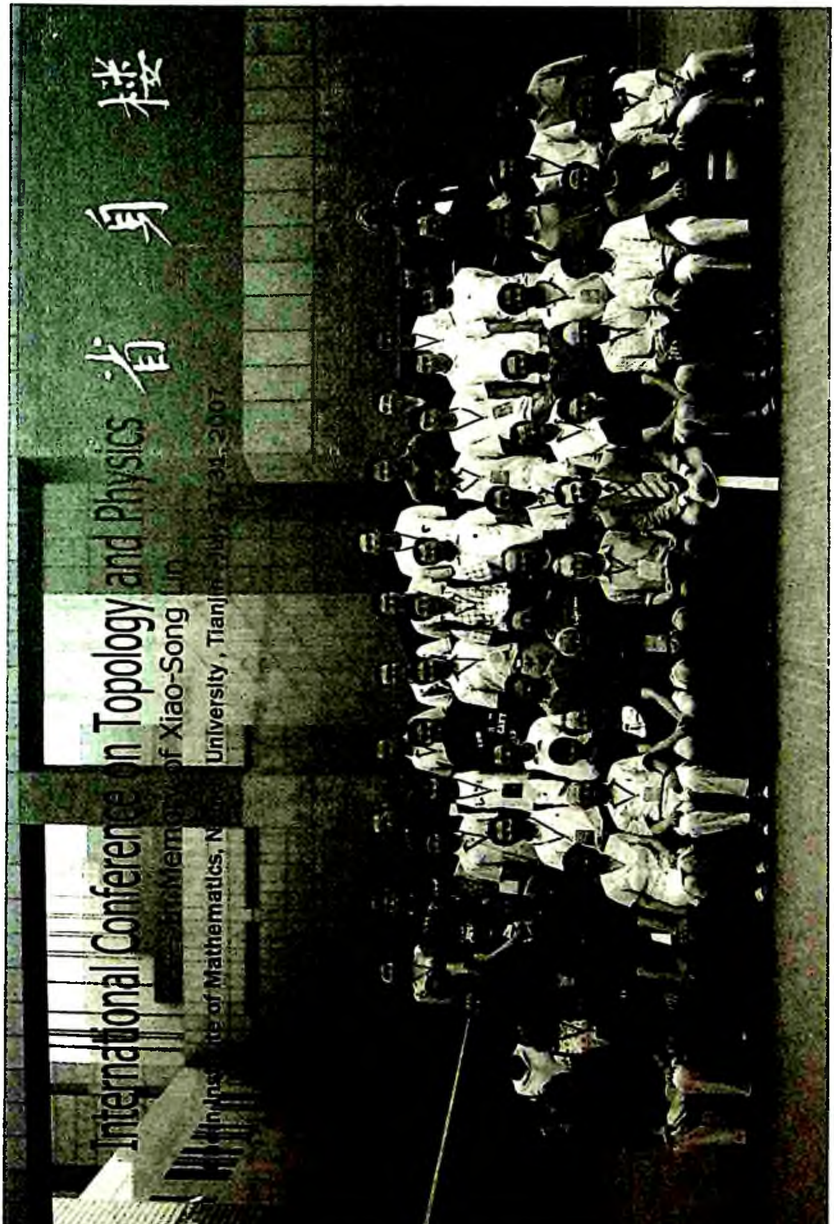
Link symbol in Xiao-Song's office

ADVISORY COMMITTEE

Michael Freedman	– Microsoft Research Station Q, USA
Mo-Lin Ge	– Chern Institute of Mathematics, China
Boju Jiang	– Beijing University, China
Vaughan Jones	– UC Berkeley, USA
Joan Birman	– Columbia University, USA

ORGANIZING COMMITTEE

Zhengxu He	– Chinese Academy of Sciences, China
Feng Luo	– Rutgers University, USA
Shicheng Wang	– Beijing University, China
Zhenghan Wang	– Microsoft Research Station Q, USA
Weiping Zhang	– Chern Institute of Mathematics, China


International Conference on Topology and Physics
省身楼

List of Participants

Speakers:

Boju Jiang	– Beijing University, China
Dror Bar-Natan	– University of Toronto, Canada
Dylan Thurston	– Columbia University, USA
Feng Luo	– Rutgers University, USA
Feng Xu	– UC Riverside, USA
Gyo Taek Jin	– KAIST, South Korea
Hao Zheng	– Zhongshan University, China
Huaidong Cao	– Lehigh University, USA
Jason Cantarella	– University of Georgia, USA
John Sullivan	– TU Berlin, Germany
Kefeng Liu	– UCLA, USA and Zhejiang University, China
Louis Kauffman	– University of Illinois at Chicago, USA
Mo-Lin Ge	– Chern Institute of Mathematics, China
Nathan Habegger	– Univerist de Nantes, France
Ruifeng Qiu	– Jilin University, China
Shicheng Wang	– Beijing University, China
Weiping Li	– Oklahoma State University, USA
Weiping Zhang	– Nankai University, China
William Jaco	– Oklahoma State University, USA
Xiaobo Liu	– Columbia University, USA
Yisong Yang	– Polytechnic University, USA
Yizhi Huang	– Rutgers University, USA
Zhenghan Wang	– Microsoft, USA
Zhengxu He	– Chinese Academy of Sciences, China
Pan Peng	– Harvard University, USA
Ying-Qing Wu	– University of Iowa, USA
Chengbo Yue	– Chinese Academy of Sciences, China
Vassily Manturov	– Moscow State Regional University, Russia

Other Participants:

Jian-Pin He	– UC Riverside, USA
Kevin Lin	– UC Berkeley, USA
Haibing Lin	– UC Berkeley, USA
Yiping Wang	– Rutgers University, USA
Jianyi He	– Suzhou, China
Xiao-Jiang Lin	– Suzhou, China
Sen Hu	– USTC, China
Thomas Au	– CUHK, Hong Kong
Guowu Meng	– University of Science and Technology, Hong Kong
Weidong Ruan	– KAIST, Korea
Helen Wong	– Yale University, USA
Yunping Jiang	– CUNY, USA and CAS, China
Huitao Feng	– Nankai University, China
Gengyu Zhang	– Tokyo University, Japan
Hwajeong Lee	– Kaist, Korea
Seojung Park	– Kaist, Korea
Zhongtao Wu	– Princeton University, USA
Haimiao Chen	– Zhejiang University, China
Kun Wang	– Zhejiang University, China
Yi Zhu	– Zhejiang University, China
Hongbin Sun	– Beijing University, China
Xiaoming Du	– Beijing University, China
Youling Li	– Beijing University, China
Jiming Ma	– Beijing University, China
Ping Li	– Beijing University, China
Yinhua Ai	– Beijing University, China
Qing Zhang	– Beijing University, China
Youquan Liang	– Beijing University, China
Yefeng Shen	– Beijing University, China
Zhengyi Liu	– Oklahoma State University, USA
Renyi Ma	– Tsinghua University, China
Su Yang	– University of Munster, Germany
Hao Yu	– University of Minnesota, USA

PROGRAM

July 27, 2007—50th Birthday of Xiao-Song Lin

Morning 9:00—10:00

1. Welcome remarks of Weiping Zhang
2. Opening remarks of Boju Jiang
3. DVD of X.-S. Lin's life
4. Kevin Lin on behalf of the family
5. Zhenghan Wang on Lin's mathematics
6. Presentation of Xiao-Song Lin award

Chair: Mo-Lin Ge

10:00—10:50 Huaidong Cao
Ricci flow on 3-manifolds
11:00—11:50 Shicheng Wang
Topology and dynamics around tame embeddings of solenoids

Afternoon

Chair: Zhengxu He
2:00—2:50 Chengpo Yue
Smooth rigidity at ∞ of Riemannian manifolds of negative curvature
3:00—3:50 Yizhi Huang
Representation theory of VOAs and knot and 3-manifold invariants

4:10—5:00 Dror Bar-Natan
Following Lin: Expansions for Groups
5:10—6:00 Boju Jiang
A trace formula for the forcing relation of braids

Dinner 6:30—8:30

1. In memory of X.-S. Lin by participants:
Led by Feng Luo and Thomas Au
2. Jian-Pin He on behalf of the family

July 28 Sightseeing for all day

July 29

Morning

Chair: William Jaco
9:00—9:50 Kefeng Liu
String Duality and Moduli Spaces
10:00—10:50 Louis Kauffman
q-deformed Spin Networks, Knot Invariants and TQC
11:10—12:00 Hao Zheng
A geometric categorification of representations of Uq(sl2)

Afternoon

Chair: Weiping Zhang
2:00—2:50 Mo-Lin Ge
Entangled States,Yang-Baxter approach and Berry's Phase
3:00—3:50 Gyo Taek Jin
Quadrsecants of Knots

4:10—5:00 Weiping Li
Knots, knot invariants and Xiao-Song's work
5:10—6:00 Dylan Thurston
Combinatorial Heegaard Floer Homology

July 30

Morning

Chair: Shicheng Wang
9:00—9:50 Weiping Zhang
Real embeddings and Riemann-Roch for Dirac operators
10:00—10:50 Ruifeng Qiu
Stabilizations of Heegaard splittings
11:10—12:00 Xiaobo Liu
Introduction to quantization of Teichmuller space

Afternoon

Chair: Kefeng Liu
2:00—2:50 Feng Luo
Milnor's conjecture on volume of simplices
3:00—3:50 William Jaco
Ideal triangulations and edge-slopes
4:10—5:00 Pan Peng

On a proof of the Labastida-Marino-Ooguri-Vafa conjecture
5:10—6:00 Vassily Manturov
Additional gradings in Khovanov homology.

July 31

Morning

Chair: Zhenghan Wang
9:00—9:50 John Sullivan
Ropelength and Distortion of Knots
10:00—10:50 Feng Xu
Mirror extensions of Local Nets
11:10—12:00 Jason Cantarella
Inscribed polygons on closed curves in Riemannian manifolds

Afternoon

Chair: Feng Luo
2:00—2:50 Yisong Yang
Static Knot Energy, Hopf Charge, and Universal Growth Law
3:00—3:50 Ying-Qing Wu
Toroidal Dehn surgery
4:00—4:50 Nathan Habegger
Recollections of La Jolla, 1985-87

Welcome Speech of Weiping Zhang

First of all, on behalf of the Chern Institute of Mathematics, I would like to welcome all of you to attend this International Conference on Topology and Physics, which is dedicated to the memory of our beloved friend Xiao-Song Lin.

As many of us might have already known, today is actually Xiao-Song's 50th birthday. Xiao-Song could not be with us. However, I am sure that his smile and friendship, along with his love of mathematics, will stay with us during the whole conference, as well as in the future to come.

Let me also say a few words about our Institute. As you may know, this Institute was founded in 1985 by Professor Shiing-shen Chern, and was called the Nankai Institute of Mathematics at that time. Since then, Professor Chern devoted almost all of his energy to the development of the Institute. Moreover, he established the unique style of the Institute among the mathematical centers in China. It was renamed as the Chern Institute of Mathematics in December of 2005, at the 20th Anniversary of the Institute. All our faculty and staff are now working hard to maintain the spirit of the Institute established by Professor Chern, and seeking various ways to make contributions to the development of mathematics in China.

I would like to thank all of you, in particular Xiao-Song's family members, for coming to the Chern Institute. My special thanks go to Zhenghan for all his endless effort to make this conference possible. It's really an honor for the Chern Institute to be able to host such a conference for Xiao-Song!

I wish all of you a pleasant stay, and I wish our conference success.

Thank you very much!

Speech of Boju Jiang

Our friend Xiao-Song Lin has left us, but he will always be remembered for many reasons. I'll speak from the perspective of a mathematician working in China. Xiao-Song knew the Chinese mathematical community very well. He kept in close contact with us, especially the topology group in Beijing University.

Around the year 1990, during a period of rapid social changes in China, Chinese mathematics was experiencing a kind of crisis. Funding was very low, but this was not new to us. What was special to that time was a sharp decline in mathematics enrollment. It never happened before and has never happened again so far. The recovery took almost a decade. In topology, the turning point was the decision by the Chinese National Science Foundation to support an annual low-dimensional topology summer school.

The idea came about in close discussion with Xiao-Song: the summer school would bring together faculty and students from different universities, along with some overseas scholars, to learn about recent advances and to discuss their own ideas. Xiao-Song played a key role in the planning of the summer schools, especially in inviting speakers and in shaping the themes. During almost every summer since 1994, Xiao-Song was busy working in China giving lecture series and organizing discussions. His lectures were always very popular, not only to graduate students, but undergraduate students also liked his lectures very much. Several collaborations between Xiao-Song and faculty and even graduate students grew out of these summer contacts. We have continued the annual school now for more than ten years.

Xiao-Song also played a significant role in China's international exchange. A remarkable example is the 2002 ICM Satellite Conference on Geometric Topology, held in Xi'an. During the two years of preparation, Xiao-Song dedicated a lot of his energy, and acted almost as the communication center for the conference because we had so many – more than 100 – foreign participants.

He cared deeply about the future of Chinese mathematics. Together with

Professor Gang Tian, Xiao-Song proposed to the Chinese National Science Foundation to pay more attention to gifted high school students, following the example of Russia and other countries. In the past, we all know that the Chinese International Mathematical Olympiad team has done quite well. But Xiao-Song thought that mathematical competitions are not the only way, nor the best way, to attract young students into mathematics. So Xiao-Song and Tian helped to start a summer camp for high school students, which started in 2000 and has continued every year since.

Xiao-Song was able to play a key role in the development of Chinese mathematics because he knew very clearly what Chinese mathematicians really need, not only in theory but also in practice. Xiao-Song was an excellent example of how an overseas Chinese mathematician can actively and effectively contribute to the mathematical community in China. Deep inside, he was a cultivated and thoughtful Chinese scholar.

It is only appropriate to have this conference in his memory in China. I want to express my thanks to Zhenghan Wang for proposing and organizing this conference, and to Weiping Zhang and the Chern Institute of Mathematics for hosting it.

Thank you.

Boju Jiang

CONTENTS

PART A

Invited Contributions

The Modified Calabi-Yau Problems for CR-manifolds*

Jianguo Cao

Mathematics Department, University of Notre Dame,
Notre Dame, IN 46556, USA;

Department of Mathematics, Nanjing University,
Nanjing 210093, China.
Email: cao.7@nd.edu

Shu-Cheng Chang

Department of Mathematics, National Tsing Hua University,
Hsinchu 30013, Taiwan, R. O. C.
Email: scchang@math.nthu.edu.tw

Dedicated to the memory of Xiao-Song Lin.

In this paper, we derive a partial result related to a question of Yau: "Does a simply-connected complete Kähler manifold M with negative sectional curvature admit a bounded non-constant holomorphic function?"

Main Theorem. *Let M^{2n} be a simply-connected complete Kähler manifold M with negative sectional curvature ≤ -1 and $S_\infty(M)$ be the sphere at infinity of M. Then there is an explicit bounded contact form β defined on the entire manifold M^{2n}.*

Consequently, if M^{2n} is a simply-connected Kähler manifold with negative sectional curvature $-a^2 \leq sec_M \leq -1$, then the sphere $S_\infty(M)$ at infinity of M admits a bounded contact structure and a bounded pseudo-Hermitian metric in the sense of Tanaka-Webster.

We also discuss several open modified problems of Calabi and Yau for Alexandrov spaces and CR-manifolds.

*2000 *Mathematics Subject Classification.* Primary 53C20; Secondary 53C23
Key words and phrases. Negative sectional curvature, bounded holomorphic functions, bounded cohomology, contact structures, Calabi problems, Yau's conjectures, CR-manifolds.

0. Introduction

In this paper, we will provide a detailed construction of *bounded* contact structures on a simply-connected complete Kähler manifold M with negative sectional curvature ≤ -1. Afterwards, we will discuss related open problems inspired by Calabi and Yau.

In 1979, Professor S. T. Yau [Y1] asked the following question.

Problem 0.1. *(Yau [Y1]) Let M^{2n} be a simply-connected complete Kähler manifold M with negative sectional curvature ≤ -1. Does M^{2n} admit a bounded non-constant holomorphic function?*

In fact, an even more attractive problem in complex analytic differential geometry is to characterize bounded domains in C^n within noncompact manifolds.

Problem 0.2. *(Yau [Y1]) Let M^{2n} be a simply-connected complete Kähler manifold M with negative sectional curvature ≤ -1. Is M bi-homeomorphic to a bounded domain in $\mathbb{C}^n$?*

Some partial progress has been made by Bland [Bl] and Nakano-Ohsawa [NO]. Under extra assumptions, they proved the existence of CR functions on the ideal boundary $S_\infty(M)$. In [Bl], two sufficient conditions were given for a complete Kähler manifold M of non-positive sectional curvature to admit nonconstant bounded holomorphic functions, which seems also to guarantee that M is a relatively compact domain with smooth boundary.

The precise definition of ideal boundary $S_\infty(M)$ can be found in [BGS].

Theorem 0.3. *Let M^{2n} be a simply-connected complete Kähler manifold M with negative sectional curvature ≤ -1 and $S_\infty(M)$ be the sphere at infinity of M. Then there is an explicit bounded contact form β defined on the entire manifold M^{2n}.*

Consequently, if M^{2n} is a simply-connected Kähler manifold with negative sectional curvature $-a^2 \leq sec_M \leq -1$, then the sphere $S_\infty(M)$ at infinity of M admits a bounded contact structure and a bounded pseudo-Hermitian metric in the sense of Tanaka-Webster.

Our proof of Theorem 0.3 was inspired by Gromov's bounded cohomology [Gro1-2] and calculations in [CaX].

Let ω be the Kähler metric on M^{2n}. It is clear that $d\omega = 0$. When M^{2n} is a simply-connected complete Kähler manifold with negative sectional curvature ≤ -1, Gromov observed that there must be a bounded 1-form β with

$$d\beta = \omega. \tag{0.1}$$

The proof of Gromov's assertion was outlined in [Pa] and [JZ]. In this paper, we provide a detailed proof of Gromov's assertion in §1. A similar sub-linear estimate for equation (0.1) on manifolds with non-positive curvature was given by the first author and Xavier in [CaX].

1. Bounded solutions to $d\beta = \alpha$ on manifolds with negative curvature

In this section, we prove Theorem 0.3. In addition, we present a new direct proof of Gromov's bounded cohomology theorem of negative curvature, see Theorem 1.4 and its proof below. Gromov's original approach to Theorem 1.4 below was based a volume estimate of k-dimensional cone over a $(k-1)$-dimensional chain, and then use a dual space argument to complete the proof. Our new method is to work on k-chains directly with a controlled Poincaré lemma for negative curvature. Our approach might have some potential independent applications.

Throughout this section (M^m, g) will be a complete simply-connected manifold of negative sectional curvature ≤ -1. Let also α be a bounded smooth closed k-form on M with $k \geq 1$. Since M^m is diffeomorphic to $\mathbb{R}^m$ there exists a form β such that $d\beta = \alpha$. The purpose of this section is to show that β can be chosen to be bounded. The proof will follow from the Poincaré lemma by a comparison argument.

Fix $p \in M$ and denote by $exp_p : T_pM \to M$ the exponential map based at p.

Lemma 1.1. *Consider the maps $\tau_t : M \to M$, given by $x \longmapsto exp_p(t\, exp_p^{-1}(x))$, where $0 \leq t \leq 1$. Then*

$$|(\tau_t)_*\xi| \leq \frac{\sinh tr}{\sinh r}|\xi| \tag{1.1}$$

for every tangent vector ξ, where $r = d(x,p)$.

Proof. Let $\sigma : [0,1] \to M^n$ be the geodesic segment joining p to x, $\xi \in T_xM^n$ and $y = (exp_p)^{-1}(x) \in T_pM^n$. By a straightforward computation one has

$$\begin{aligned}(\tau_t)_*\xi &= (d\, exp_p)_{t(exp_p)^{-1}(x)}\{td(exp_p^{-1})_{(x)}\xi]\\ &= (d\, exp_p)_{ty}\{t[d(exp_p)_y]^{-1}\xi\}.\end{aligned}$$

Recall that $\sigma(t) = exp_p(ty)$. It is now manifest from the above formula that

$$J(tr) := (\tau_t)_*\xi \tag{1.2}$$

is the Jacobi field along σ satisfying $J(0) = 0$, $J(r) = \xi$. On the other hand, since the sectional curvatures are ≤ -1, we estimate the function $f(s) := |J(s)|$ by a method inspired by Gromov. It is sufficient to verify

$$\frac{|J(s)|}{\sinh s} \leq \frac{|J(r)|}{\sinh r}, \tag{1.3}$$

for all $0 \leq s \leq r$.

We may assume that $r > 0$, otherwise the inequality (1.1) holds trivially. To do this, we consider the function

$$\eta(s) = \frac{f(s)}{\sinh s}.$$

It is sufficient to verify

$$\frac{f(s)}{\sinh s} \leq \frac{f(r)}{\sinh r} \text{ or } \eta'(s) \geq 0. \tag{1.4}$$

Since we have

$$\eta'(s) = \frac{f'(s)\sinh s - f(s)\cosh s}{[\sinh s]^2},$$

it remains to verify that

$$[f'(s)\sinh s - f(s)\cosh s]' = f''(s)\sinh s - f(s)\sinh s \geq 0. \tag{1.5}$$

Recall that the curvature tensor R is given by $R(X,Y)Z = -\nabla_X\nabla_Y Z + \nabla_Y\nabla_X Z + \nabla_{[X,Y]}Z$ where $[X,Y] = XY - YX$ is the Lie bracket of X and Y.

Following a calculation in [BGS], by our assumption of $sec_M \leq -1$ we have

$$\begin{aligned} f''(s) &= |J(s)|'' \\ &= [\frac{\langle J, J'\rangle}{|J|}]' \\ &= \frac{\langle J, J''\rangle|J|^2 + \langle J', J'\rangle|J|^2 - \langle J, J'\rangle^2}{|J|^3} \\ &\geq \frac{-\langle R(\sigma', J)\sigma', J\rangle|J|^2}{|J|^3} \\ &\geq f(s), \end{aligned} \tag{1.6}$$

where we used the assumption that $\langle J'', J\rangle = -\langle R(\sigma', J)\sigma', J\rangle \geq |J|^2$. It follows from (1.5)-(1.6) that (1.4) holds. This completes the proof of (1.3) as well as Lemma 1.1. $\square$

Recall that if α is a k-form and Z is a vector field, then $(\alpha\lfloor_Z)$ is the $(k-1)$-form given by

$$(\alpha\lfloor_Z)(\xi_1, \cdots, \xi_{k-1}) = \alpha(Z, \xi_1, \cdots, \xi_{k-1}).$$

For the sake of completeness we give a proof of the following elementary result.

Lemma 1.2. *Let Ψ be a closed k-form in $\mathbb{R}^m$. Then the $(k-1)$-form Φ defined by*

$$\Phi(x) = r\int_0^1 [(\tau_t)^*(\Psi\lfloor_{\frac{\partial}{\partial r}})](x)dt$$

satisfies $d\Phi = \Psi$; here $\frac{\partial}{\partial r} = \sum_{i=1}^m \frac{x_i}{r}\frac{\partial}{\partial x_i}$, $r = (\sum_{i=1}^m x_i^2)^{1/2}$ and $\tau_t(x) = tx$.

Proof. By the standard proof of the Poincaré lemma ([SiT], p.130), Φ can be taken to be $\Phi(x) =$

$$\sum_{i_1<\cdots<i_k}\sum_{j=1}^{k}(-1)^{j-1}x_{i_j}\Big(\int_0^1 t^{k-1}\Psi_{i_1\cdots i_k}(tx)dt\Big)dx_{i_1}\wedge\cdots\wedge\widehat{dx_{i_j}}\wedge\cdots\wedge dx_{i_k},$$

where $\Psi = \sum_{i_1<\cdots<i_k}\Psi_{i_1\cdots i_k}dx_{i_1}\wedge\cdots\wedge dx_{i_k}$.

In particular, one has

$$\begin{aligned}\Phi(x) &= \sum_{i_1<\cdots<i_k}\sum_{j=1}^{k}x_{i_j}\Big(\int_0^1 t^{k-1}\Psi_{i_1\cdots i_k}(tx)dt\Big)(dx_{i_1}\wedge\cdots\wedge dx_{i_k})\lfloor_{\frac{\partial}{\partial x_{i_j}}}\\ &= r\sum_{i_1<\cdots<i_k}\Big(\int_0^1 t^{k-1}\Psi_{i_1\cdots i_k}(tx)dt\Big)(dx_{i_1}\wedge\cdots\wedge dx_{i_k})\lfloor_{\frac{\partial}{\partial r}}\\ &= r\int_0^1 t^{k-1}(\Psi\lfloor_{\frac{\partial}{\partial r}})(tx)dt\\ &= r\int_0^1 [(\tau_t)^*(\Psi\lfloor_{\frac{\partial}{\partial r}})](x)dt,\end{aligned}$$

as desired. $\square$

We would also like to borrow another elementary but useful observation of Gromov, in order to prove our main theorem

Lemma 1.3. *(Gromov, [Cha, page 124]) Suppose that f and h are positive integrable functions, of real variable r, for which*

$$\frac{f}{g}$$

is an increasing with respect to r. Then the function

$$\frac{\int_0^r f}{\int_0^r g}$$

is also increasing with respect to $r \geq 0$.

Let us now provide a new detailed proof of a theorem of Gromov.

Theorem 1.4. *(Gromov) Let M^m be a simply-connected complete Riemannian manifold with negative sectional curvature ≤ -1. Suppose that α is bounded closed k-form with $k \geq 2$. There is a bounded $(k-1)$-form β with $d\beta = \alpha$ satisfying*

$$\|\beta\|_{L^\infty} \leq \frac{1}{k-1}\|\alpha\|_{L^\infty}. \tag{1.7}$$

Proof. Let $(x_1, ..., x_n)$ be Euclidean coordinates on T_pM and consider the pull-back metric h of the metric g under $exp_p : T_pM \to M$. Observe that there are now two ways to interpret the map τ_t. The first interpretation comes from Lemma 1.1 with (M, g) being replaced by (T_pM, h); alternatively, one can think of τ_t as the self-map of T_pM, $(x_1, ..., x_n) \longmapsto t(x_1, ..., x_n)$, that appears in the Poincaré lemma (Lemma 1.2). It is an easy and yet basic observation that these two ways of thinking about τ_t give rise to the same map.

We may also replace the form α that appears in the statement of Lemma 1.2 by a closed form Ψ on T_pM which is bounded in the induced metric h. Let Φ be given by Lemma 1.2 and observe that, by Lemma 1.1,

$$|(\tau_t)^*\varphi(x)|_h \leq \left(\frac{\sinh tr}{\sinh r}\right)^{k-1}|\varphi(\tau_t(x))|_h, \quad k \geq 2, \tag{1.8}$$

holds for any $(k-1)$-form φ on T_pM; here $|\cdot|_h$ is any one of the equivalent norms induced by h. Since $|\frac{\partial}{\partial r}| = 1$, it follows from (1.3) and Lemma 1.2

that

$$\begin{aligned}|\Phi(x)|_h &\leq r\int_0^1 |[(\tau_t)^*(\Psi\lfloor_{\frac{\partial}{\partial r}})](x)|_h dt \\ &\leq r\int_0^1 (\frac{\sinh tr}{\sinh r})^{k-1}|\Psi(tx)\lfloor_{\frac{\partial}{\partial r}}|_h dt \\ &= \int_0^r (\frac{\sinh s}{\sinh r})^{k-1}|\Psi(\frac{s}{r}x)\lfloor_{\frac{\partial}{\partial r}}|_h ds \\ &\leq \frac{\int_0^r(\sinh s)^{k-1}ds}{(\sinh r)^{k-1}} \sup_{0\leq s\leq r}|\Psi(\frac{s}{r}x)|_h\end{aligned} \tag{1.9}$$

Choosing $f(r) = (\sinh r)^{k-1}$ and $\hat{g}(r) = (k-1)(\sinh r)^{k-2}\cosh r$ in Lemma 1.3, we see that $[\frac{f}{\hat{g}}]' = \frac{1}{(k-1)(\sinh r)^2} > 0$ and

$$\frac{\int_0^r(\sinh s)^{k-1}ds}{(\sinh r)^{k-1}} \leq \frac{1}{k-1}. \tag{1.10}$$

It follows from (1.9)-(1.10) that

$$|\Phi(x)|_h \leq \frac{1}{k-1}\sup|\Psi|_h. \tag{1.11}$$

Hence Φ is a bounded solution of $d\Phi = \Psi$ and the proof of Theorem 1.4 is completed. □

Proof of Main Theorem:

Our main theorem Theorem 0.3 can be derived as follows. We fix a base point p as above. There is a differential structure Ξ_p imposed on $S_\infty(M)$ given by the map

$$\begin{aligned}F_p :&\overline{B_1(0)} \to M \cup S_\infty(M)\\ &\vec{v} \to Exp_p[\frac{\vec{v}}{1-|\vec{v}|}].\end{aligned}$$

For $p \neq q$, the transitive map $F_q^{-1} \circ F_p : \overline{B_1(0)} \to \overline{B_1(0)}$ is not necessarily smooth. However, we fix *one* differential structure Ξ_p on $S_\infty(M)$ via the map F_p.

Let J be the complex structure of our Kähler manifold M. Let $r(x) = d(x,p)$ and $\beta = J \circ dr$, i.e., $\beta(\vec{w}) = dr(J\vec{w})$ for all $\vec{w} \in T_x(M)$. When $-a^2 \leq sec_M - 1$, it is known that

$$|X|^2 \leq |(\nabla_X dr)(X)| = |Hess(r)(X,X)| \leq a|X|^2$$

for all $X \in T_x(\partial B_r(p))$ with $r >> 1$.

Since M is Kähler, we have $\nabla_X J = 0$. It follows that $|\nabla_X \beta| \leq a|X|$ for $X \in T_x(\partial B_r(p))$ with $r >> 1$.

Thus, $\{\beta|_{\partial B_r(p)}\}$ defines an equi-continuous family of contact forms on $S_\infty(M)$. By Ascoli lemma, there is a subsequence that converges to a bounded contact form β_∞ on $S_\infty(M)$. Since $sec_M \leq -1$, it is known that $d\beta(\tilde{X}, \bar{\tilde{X}}) = Hess(r)(X, X) + Hess(r)(JX, JX) \geq 2|X|^2$ for all $X \in T_x(\partial B_r(p))$ and $X \perp \nabla r$, where $\tilde{X} = \frac{1}{\sqrt{2}}[X - \sqrt{-1}JX]$. Therefore, β_∞ defines a non-trivial contact form on $S_\infty(M)$. Moreover, $\omega_\infty = d\beta_\infty$ gives rise to a pseudo-hermitian metric on $S_\infty(M)$.

Similarly, one can also choose β^* satisfying $d\beta^* = \omega$, where ω is the Kähler form of M and β^* in the proof of Theorem 1.4. With some extra effort, one can show that $|\nabla\beta^*| \leq c_1$ for some constant c_1. Thus, $\{\beta^*|_{\partial B_r(p)}\}$ defines an equi-continuous family of contact forms on $S_\infty(M)$ as well.

This completes the proof of our main theorem.

2. The modified Calabi-Yau problems for singular spaces and CR-manifolds

In this section, we will discuss the generalized Calabi problems on Kähler manifolds with boundaries. In addition, we will comment on the existence of positive sup-harmonic functions on (possibly singular) Alexandrov spaces with non-negative sectional curvature.

§A. Sup-harmonic functions on Alexandrov spaces with non-negative sectional curvature

Professor S. T. Yau also had earlier results on bounded harmonic functions on smooth complete Riemannian manifolds with non-negative Ricci curvature. We would like to extend this theorem of Yau to singular spaces.

In an important paper [Per1], Perelman provided an affirmative solution to the Cheeger-Gromoll soul conjecture. More precisely, he showed that *"if a smooth complete non-compact Riemannian manifold M^n of non-negative curvature has a point p_0 with strictly positive curvature $K|_{p_0} > 0$, then M^n must be diffeomorphic to $\mathbb{R}^n$*. In [Per1], Perelman also asked to what extent the conclusions of his paper [Per1] would hold for (possibly singular) Alexandrov spaces with non-negative curvature.

Recently, the first author, together with Dai and Mei, showed the following.

Theorem A.1. *(Cao-Dai-Mei, 2007, [CaMD1]) Let M^n be a n-dimensional complete, non-compact Alexandrov space with non-negative* *sectional curvature. Suppose that M^n has no boundary and M^n has positive sectional curvature on an non-empty open set. Then M^n is contractible.*

In 1976, Professor S. T. Yau proved the following Liouville type theorem.

Theorem A.2. *(Yau, 1976 [Y3]) Let M^n be a n-dimensional complete, non-compact smooth Riemannian space with non-negative Ricci curvature. Then any positive harmonic functions on M^n must be a constant function.*

On an (possibly singular) Alexandrov space, we introduce the following notion of sup-harmonic function.

Definition 0.1. Definition A.3 Let M^n be a n-dimensional complete, non-compact Alexandrov space with non-negative sectional curvature. Suppose that M^n has no boundary, $f : M^n \to \mathbb{R}$ is a Lipschitz continuous function and

$$f(x) \geq \frac{1}{Area(\partial B_\varepsilon(x))} \int_{\partial B_\varepsilon(x)} f dA \tag{A.1}$$

for any sufficiently small $\varepsilon > 0$. Then we say that f is a sup-harmonic function on M.

For example, $f(x) = -[d(x, x_0)]^2$ is a sup-harmonic function on M, whenever M has non-negative sectional curvature in generalized sense.

Problem A.4. *(Liouville-Yau type problem) Let M^n be a n-dimensional complete, non-compact Alexandrov space with non-negative sectional curvature. Suppose that M^n has no boundary. Is it true that any positive sup-harmonic functions on M^n must be a constant function?*

In [CaB], the first author and Benjamini studied a different Liouville-type problem of Schoen-Yau. One hopes to continue to work on Liouville-Yau type problem mentioned above.

§B. The generalized Calabi problems for Kähler domains with boundaries

The classical Calabi problems for Ricci curvatures on compact Kähler manifolds *without boundary* have been successfully solved by Professor S. T. Yau.

Theorem B.1. *(Yau [Y4]) Let M^{2n} be a compact smooth Kähler manifold without boundary. Then the following is true: (1) For any Kähler form $\omega_0 \in H^{(1,1)}(M^{2n})$ and any $(1,1)$-form β representing the first Chern class*

$c_1(M^{2n})$, there is a Kähler metric $\tilde{\omega} = \omega_0 + i\partial\bar{\partial}f$ such that its Ricci curvature tensor satisfies

$$Ric_{\tilde{\omega}} = \beta;$$

(2) If the first Chern class $c_1(M) \leq 0$, then M^{2n} admits a Kähler-Einstein metric.

For a Kähler manifold Ω with boundary $M^{2n-1} = b\Omega$, we consider a similar problem. This problem is closely related to the existence problem of CR-Einstein metrics, or partially Einstein metrics.

Definition B.2. *(CR-Einstein metrics or partially Einstein metrics, [Lee2]) Let Σ^{2n-1} be a CR-hypersurface with CR-distribution $\mathcal{H}_{\Sigma^{2n-1}} = \ker\theta$ for some contact 1-form θ and let $g_\theta(X, JY) = d\theta(X, JY)$ be a pseudo-hermitian metric as above. If the Ricci tensor of g_θ satisfies*

$$Ric_{g_\theta}(X, Y) = cg_\theta(X, Y)$$

for all $X, Y \in \mathcal{H}_{\Sigma^{2n-1}} = \ker\theta$ where c is a constant, then g_θ is called a CR-Einstein (partially Einstein) metric.

Inspired by Yau's result, Lee proposed to study the CR-version of the Calabi problem.

Problem B.3. *(CR-Calabi Problems, [Lee2]) Let M^{2n-1} be a CR-manifold, Φ be a closed form representing the first Chern class for the bundle $T^{(1,0)}(M^{2n-1})$ and $\Phi_b(X, Y) = \Phi(X, Y)$ for $X, Y \in \mathcal{H}_{\Sigma^{2n-1}} = \ker\theta$.*

(1) Can we find a pseudo-metric g_θ such that its Ricci tensor satisfies

$$Ric_{g_\theta}(X, Y) = \Phi_b(X, Y) \tag{B.1}$$

for all $X, Y \in \mathcal{H}_{\Sigma^{2n-1}} = \ker\theta$?

(2) Given a (1,1)-form $\beta_b \in [c_1(M^{2n-2}]_b$, can we find a pseudo-metric g_θ such that its Ricci tensor satisfies

$$Ric_{g_\theta}(X, Y) = \beta(X, Y) \tag{B.2}$$

for all $X, Y \in \mathcal{H}_{\Sigma^{2n-1}} = \ker\theta$?

The pseudo-Hermitian metric for general CR-manifolds was also discussed in [Ta1-2] and [Web]. Authors derived the following partial answer to Problem 3:

Problem B.4. *([CaCh]) Let M^{2n-1} be the smooth boundary of a bounded strongly pseudo-convex domain Ω in a complete Stein manifold V^{2n}. Then* *for $n \geq 3$, M^{2n-1} admits a CR-Einstein metric (or partially Einstein metric).*

One might be able to continue working on Problem B.3, using Kohn-Rossi's $\bar{\partial}_b$-theory described below.

§C. The Calabi-Escobar type problem for Kähler domains with boundaries

The first author and Mei-Chi Shaw studied the CR-version of the Poincaré-Lelong equation $i\partial_b\bar{\partial}_b u = \Psi_b$ in [CaS3]. The linearization equation for (B.2) is related to the CR-version of Poincaré-Lelong equation.

In fact, to solve the linear function

$$\bar{\partial}_b u = \beta_b \text{ on } b\Omega, \tag{C.1}$$

Kohn and Rossi [KoRo] used the solutions to the $\bar{\partial}$-Cauchy problem to solve $\bar{\partial}_b u = \beta_b$ extrinsically as follows. Let us first choose an arbitrary smooth extension $\hat{\beta}$ on Ω. If we can solve

$$\begin{cases} \bar{\partial} v = \bar{\partial}\hat{\beta} \text{ on } \Omega \\ v \llcorner_X = 0, \text{ for } X \in T_z^{(0,1)}(b\Omega) \end{cases} \tag{C.2}$$

Clearly $\tilde{\beta} = \hat{\beta} - v$ is a $\bar{\partial}$-closed extension on Ω of β. If we solve

$$\bar{\partial}\tilde{u} = \hat{\beta} - v \text{ on } (\Omega \cup b\Omega), \tag{C.3}$$

then the restriction $u = \tilde{u}\llcorner_{b\Omega}$ satisfies

$$\bar{\partial}_b[(\tilde{u})\lfloor_{b\Omega}] = \beta_b \text{ on } b\Omega.$$

The details for solving the $\bar{\partial}$-Cauchy problem (C.2) was given in Chapter 9 of [ChSh].

In 1992, Escobar [Esc] was able to solve the non-linear curvature equation on manifolds *with boundary.*

Theorem C.1. *(Escobar [Esc]) Let $\Omega \subset \mathbb{R}^n$ be a compact domain with smooth boundary $\partial\Omega$ and $n > 6$. Then there is a conformally flat metric g on Ω such that the scalar curvature $Scal_g$ of g is zero and the mean curvature H_g of $(\partial\Omega, g)$ is constant:*

$$\begin{cases} Scal_g = 0 \text{ on } \Omega \\ H_g = c \text{ on } \partial\Omega, \end{cases} \tag{C.4}$$

for some constant c.

Inspired by Theorem C.1 and the Kohn-Rossi's solution to $\bar{\partial}$-Cauchy problem, we are interested in the following type.

Problem C.2. *(Calabi-Escobar type problem) Let Ω be a compact domain in Stein manifold M with smooth strongly pseudo convex boundary $b\Omega$, and let H_g^{CR} be the partial sum of second fundamental form of $(b\Omega, g)$ over the CR-distribution $ker\theta$ of $b\Omega$. Is there is a Kähler-Einstein metric g on Ω with constant CR-mean curvature on the boundary $b\Omega$? In another words, we would like to find the existence of solution to the following non-linear boundary problem:*

$$\begin{cases} Ric_g = c_1 g \ \ on\ \Omega \\ H_g^{CR} = c_2 \ \ on\ b\Omega \end{cases} \tag{C.5}$$

for some constant numbers c_1 and c_2.

The linearalization of non-linear equation is the Poincare-Lelong equation with boundary conditions. The first author and Mei-Chi Shaw [CaS] were able to solve

$$i\partial_b\bar{\partial}_b u = \Theta_b \quad \text{on} \quad b\Omega \tag{C.6}$$

even for weakly pseudo-convex domains Ω in $\mathbb{C}P^n$.

One hopes to continue to work in direction, in order to investigate Problem C.2.

Acknowledgments

The first author is supported in part by an NSF grant Grant DMS0706513 and Changjiang Scholarship of China at Nanjing University. The first named author is also very grateful to National Center for Theoretical Sciences at National Tsinghua University for its warm hospitality. The second author is supported partially by the NSC of Taiwan.

We thank Professor Mei-Chi Shaw for pointing out a misquote in the earlier version of Section 1. She also proposed to use (not necessarily unique) one of smooth differential structures at infinity. The first author is indebted to Professor Rick Schoen for his advice on open manifolds of negative curvature.

References

Bah. Bahuaud, E. *Intrinsic characterization for Lipschitz asymptotically hyperbolic metrics.* Preprint 2007, posted as arXiv:math/0711.3371

BahM. Bahuaud, E. and Marsh, T. *Hölder Compactification for some manifolds with pinched negative curvature near infinity.* Preprint 2006, posted as arXiv:math/0601503

BGS. Ballmann, W, Gromov, M., and Schroeder, V. *Manifolds of nonpositive curvature.* 1985 Birkhäuser, Boston

Be. Berger, M. *Riemannian geometry during the second half of the twentieth century.* University Lect. Series, 17. Amer. Math. Society, Providence, RI, 2000. Reprint of the 1998 original (Jahresber. Deutsch. Math.-Verein. 100 (1998), no. 2, 45–208).

Bl. Bland, J. *On the existence of bounded holomorphic functions on complete Kähler manifolds.* Invent. Math. vol **81** (1985), no. 3, 555–566

BGP. Burago, Yu., Gromov, M. and Perelman, G. 1-58 *A.D. Alexandrov spaces with curvature bounded below.* Russ. Math. Surv. vol. **47** 1992

Ca. Cao, J. *Cheeger isoperimetric constants of Gromov-hyperbolic spaces with quasi-poles.* Communications in Contemporary Mathematics vol. **2** No.4 2000 511-533

CaB. Cao, J. and Benjamini, I. *Examples of simply-connected Liouville manifolds with positive spectrum.* Differential Geom. Appl. vol. **6** No.1 1996 31-50

CCa. Cao, J. and Calabi, E. *Simple closed geodesics on convex surfaces.* J. Differential Geom. vol. **36** 1992 517-549

CaCh. Cao, J., Chang, Shu-Cheng *Pseudo-Einstein and Q-flat metrics with eigenvalue estimates on CR-hypersurfaces.* "Indiana Univ. Math. Journal", vol 56, No. 6 (2007), pages 2839-2858

CaDM1. Cao, J., Dai, B. and Mei, J. *An extension of Perelman's soul theorem for singular spaces.* Preprint 2007, 53 pages, "arxiv/math.dg/pdf/0706/0706.0565v3.pdf", submitted

CaDM2. Cao, J., Dai, B. and Mei, J. *An optimal extension of Perelman's comparison theorem for quadrangles and its applications.* Preprint 2007, 29 pages, "arXiv:0712.3221v1 [math.DG]", submitted

CaFL. Cao, J., Fan, Huijun and Ledrappier, F. *Martin points on open manifolds of non-positive curvature.* Transaction of Amer. Math. Soc., vol 359 (2007) page 5697-5723

CaS1. Cao, J., Shaw, Mei-Chi *The smoothness of Riemannian submersions with nonnegative sectional curvature.* Communication in Contemporary Math. vol. **7** 137-144 2005

CaS2. Cao, J., Shaw, Mei-Chi *A new proof of the Takeuchi theorem.* Lecture Notes of Seminario Interdisplinare di Mathematica vol. 4 65-72 2005

CaS3. Cao, J. and Shaw, Mei-Chi *The d-bar-Cauchy problem and nonexistence of Lipschitz Levi-flat hypersurfaces in CP^n with $n \geq 3$.* Math. Zeit, vol 256 (2007), No. 1, page 175-192

CaSW. Cao, J., Shaw, M. and Wang, L. *Estimates for the $\bar\partial$-Neumann problem and nonexistence of Levi-flat hypersurfaces.* in $\mathbf{C}P^n$. Mathematische Zeitschrift vol. **248** No.1 183-221 2004

CaT. Cao, J. and Tang, H. *An intrinsic proof of Gromoll-Grove diameter* *rigidity theorem.* "Communication in Contemporary Math.", vol 19 no.3 (2007) 401-419

CaW. Cao, J. and Wang, Youde *An Introduction to Modern Riemannian Geometry* . (in Chinese) 2006 Lectures in Contemporary Mathematics, Volume 1, Science Press, Beijing, China, ISBN 7-03-016435-0, 147 pages, English version, to appear

CaX. Cao, J. and Xavier, F. *Kahler parabolicity and the Euler number of compact manifolds of non-positive sectional curvature.* Mathematische Annalen, Volume 319 Issue 3 (2001) pp 483-491

Cha. Chavel, I. *Riemannian geometry—a modern introduction.* Cambridge Tracts in Mathematics, 108. Cambridge University Press, Cambridge, 1993. xii+386 pp. ISBN: 0-521-43201-4; 0-521-48578-9

ChSh. Chen, So-Chin; Shaw, Mei-Chi *Partial differential equations in several complex variables.* AMS/IP Studies in Advanced Mathematics vol. **19** American Mathematical Society, Providence, RI; International Press, Boston, MA, 2001. xii+380 pp. 2001

Ch. Chern, S. S. *On curvature and characteristic classes of a Riemannian manifold.* 1955 Abh. Math. Sem. Univ. Hamburg, **20** 117-126

Esc. Escobar, J. *Conformal deformation of a Riemannian metric to a scalar flat metric with constant mean curvature on the boundary.* Ann. of Math. vol. **136** No.1 1-50 1992

GW. Greene, R. and Wu, H. *Function theory on manifolds which posses a pole.* 1979 Springer Lect. Notes in Math., **699**

Gro1. Gromov, M. *Volume and bounded cohomology.* Inst. Hautes 07tudes Sci. Publ. Math. No. 56 (1982), 5–99 (1983)

Gro2. Gromov, M. *Kähler hyperbolicity and L^2-Hodge theory.* 1991 J. Differential Geom., **33** 263-292

JZ. Jost, J. and Zuo, K. *Vanishing theorems for L^2-cohomology on infinite coverings of compact Kähler manifolds and applications in algebraic geometry.* 2000 Comm. Anal. Geom. vol 8 1-30

HL. Harvey, F. R and Lawson, B. *On boundaries of complex analytic varieties. I.* 1975 Ann. of Math. vol. **102** 223-290

KoRo. Kohn, J. J. and Rossi, Hugo *On the extension of holomorphic functions from the boundary of a complex manifold.* 1965 Ann. of Math. vol. **81**, 451-472

Lee1. Lee, J. 411-429 *The Fefferman metric and pseudo-Hermitian invariants.* Trans. Amer. Math. Soc. vol. **296**, no. 1 1986

Lee2. Lee, J. 157-178 *Pseudo-Einstein structures on CR manifolds.* Amer. J. Math. vol. **110**, no. 1 1988

MY. Mok, N. and Yau, S. T.: 41-59 *Completeness of the Kähler-Einstein metric on bounded domains and the characterization of domains of holomorphy by curvature conditions.* The mathematical heritage of Henri Poincar, Part 1 (Bloomington, Ind., 1980), Proc. Sympos. Pure Math., Amer. Math. Soc., Providence, RI vol. **39** 1983.

No. Nakano, S. and Ohsawa, T. *Strongly pseudoconvex manifolds and strongly pseudoconvex domains.* Publ. Res. Inst. Math. Sci. 20 (1984), no. 4, 705–715.

Pa. Pansu, P. 53-86 *Introduction to L^2 Betti numbers.* In the book by G. Besson; J. Lohkamp; P. Pansu; P. Petersen, Riemannian geometry. Papers from the micro-program held in Waterloo, Ontario, August 3–13, 1993. Edited by L. Miroslav, M.-O. Maung and McKenzie Y.-K. Wang. Fields Institute Monographs, 4. American Mathematical Society, Providence, RI, 1996. xii+115 pp.

Per1. Perelman, G. 209-212 *Proof of the soul conjecture of Cheeger and Gromoll.* J. Diff. Geom. vol. **40** 1994

Per2. Perelman, G. 299-305 *Manifolds of positive Ricci curvature with almost maximal volume.* J. Amer. Math. Soc. vol. **7** no. 2 1994

Per3. Perelman, G. *The entropy formula for the Ricci flow and its geometric applications.* Preprint, math.DG/0211159.

Per4. Perelman, G. *Ricci flow with surgery on three-manifolds.* Preprint 2003, www.arXiv.org, math.DG/0303109

Per5. Perelman, G. *Alexandrov's spaces with curvatures bounded from below II.* Preprint 1991, see http: www.math.psu.edu/petrunin/paper.

Per6. Perelman, G. *Elements of Morse theory on Aleksandrov spaces.* St. Petersburg Math. J. vol. **5** no. 1 1994, 205-213

Per7. Perelman, G. *DC structure on Alexandrov spaces with curvatures bounded below.* Preprint 1994, see website in.[Per5]

Sch. Schoen, R. and Yau, S. T. *Lectures on differential geometry.* International Press, Cambridge, MA, 1994

ShW. Shaw, M.-C. and Wang, L *Hölder and L^p estimates for $\Box_b$ on CR manifolds of arbitrary codimension.* Math. Ann. vol. **331** no.2 2005 297-343

SiT. Singer, I and Thorpe, J. *Lecture Notes on Elementary Topology and Geometry.* Springer-Verlag, 1976

Ta1. Tanaka, N. 397-429 *On the pseudo-conformal geometry of hypersurfaces of the space of n complex variables.* J. Math Soc. Japan. vol. **14** 1962

Ta2. Tanaka, N. *A differential geometric study on strongly pseudo-convex manifolds.* Lectures in Mathematics, Department of Mathematics, Kyoto University Kinokunia Book-Store Co. Ltd addr Tokyo, Japan vol. **9** 2001

Web. Webster, S 25-41 *Pseudo-hermitian structures on a real hypersurface.* J. Diff. Geom. vol 13 1978

Wel. Wells, R. L., Jr. *Differential analysis on complex manifolds.* Springer-Verlag, New York, 1979

Y1. Yau, S. T. *Problem Section.* In Yau, S. T. (ed.) "Seminar on Differential Geometry", Princeton University Press 1982 Annals of Math Studies, vol. **102** 669-706

Y2. Yau, S. T. *Open problems in geometry* Differential Geometry: Partial Differential Equations on Manifolds, Greene, R. and Yau, S. T. (editors): Proceedings of Symposia in Pure Mathematics vol. **54**, Part I, American Math. Soc., Providence, Rhode Island 1993

Y3. Yau, S. T. *Some Function-Theoretic Properties of Complete Riemannian manifold and their applications to Geometry.* Indiana Univ. Math. J. vol. **25** 1976 659-670

Y4. Yau, S. T. *On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampére equation.* I. Comm. Pure Appl. Math. vol. **31** no. 3 1978 339-411

On Picture (2+1)-TQFTs

Michael Freedman, Chetan Nayak, Kevin Walker and Zhenghan Wang*

Microsoft Station Q
CNSI Bldg Rm 2243
University of California
Santa Barbara, CA 93106-6105
U.S.A.
Email: zhenghwa@microsoft.com

Dedicated to the memory of Xiao-Song Lin

1. Introduction

Topological quantum field theories (TQFTs) emerged into physics and mathematics in the 1980s from the study of three distinct enigmas: the infrared limit of $1+1$ dimensional conformal field theories, the fractional quantum Hall effect (FQHE), and the relation of the Jones polynomial to 3−manifold topology. Now 25 years on, about half the literature in 3−dimensional topology employs some "quantum" view point, yet it is still difficult for people to learn what a TQFT is and to manipulate the simplest examples. Roughly (axioms will follow later), a $(2+1)$−dimensional TQFT is a functor which associates a vector space $V(Y)$ called "modular functor" to a closed oriented surface Y (perhaps with some extra structures); sends disjoint union to tensor product, orientation reversal to dual, and is natural with respect to transformations (diffeomorphisms up to isotopy or perhaps a central extension of these) of Y. The empty set $\emptyset$ is considered to be a manifold of each dimension: $\{0, 1, \cdots\}$. As a closed surface, the associated vector space is $\mathbb{C}$, i.e., $V(\emptyset) = \mathbb{C}$. Also if $Y = \partial X$, X an oriented 3−manifold (also perhaps with some extra structure), then a vector $Z(X) \in V(Y)$ is determined (surfaces Y with boundary also play a role but we pass over this for now.) A closed 3−manifold X determines a vector $Z(X) \in V(\emptyset) = \mathbb{C}$, that is a number. In the case X is the 3−sphere with "extra structure"

*partially supported by NSF FRG grant DMS-034772.

a link L, then Witten's "$SU(2)$–family" of TQFTs yields a Jones polynomial evaluation $Z(S^3, L) = J_L(e^{2\pi i/r})$, $r = 3, 4, 5, \ldots$, as the "closed 3–manifold" invariants, which mathematically are the Reshetikhin-Turaev invariants based on quantum groups[Jo] [Witt]. [RT] This is the best known example. Note that physicists tend to index the same family by the levels $k = r - 2$. The shift 2 is the dual Coxeter number of $SU(2)$. We will use both indices. Most of the "quantum" literature in topology focuses on such closed 3–manifold invariants but there has been a growing awareness that a deeper understanding is locked up in the representation spaces $V(Y)$ and the "higher algebras" associated to boundary (Y) (circles) and points[FQ]. [Fd] Let us explain this last statement. While invariants of 3–manifolds may be fascinating in their interrelations there is something of a shortage of work for them within topology. Reidemeister was probably the last topologist to be seriously puzzled as to whether a certain pair of 3–manifolds were the same or different and, famously, solved his problem by the invention of "torsion". (In four dimensions the situation is quite the opposite, and the closed manifold information from (3+1) dimensional TQFTs would be most welcome. But in this dimension, we do not yet know interesting examples of TQFTs.) So while the subject in dimension 3 seems to be maturing away from the closed case it is running into a pedological difficulty. It is hard to develop a solid understanding of the vector spaces $V(Y)$ even for simple examples. Our goal in these notes is, in a few simple examples to provide an intuition and understanding on the same level of "admissible pictures" modulo relations, just as we understand homology as cycles modulo boundaries. This is the meaning of "picture" in the title. A picture TQFT is one where $V(Y)$ is the space of formal $\mathbb{C}$–linear combinations of "admissible" pictures drawn on Y modulo some <u>local</u> (i.e. on a disk) linear relations. We will use the terms formal links, or formal tangles, or formal pictures , etc. to mean $\mathbb{C}$-linear combinations of links, tangles, pictures, etc. Formal tangles in 3-manifolds are also commonly referred to as "skein"s. Equivalently, we can adopt a dual point of view: take the space of linear functionals on multicurves and impose linear constraints for functionals. This point of view is closer to the physical idea of "amplitude of an eigenstate": think of a functional f as a wavefunction and its value $f(\gamma)$ on a multicurve γ as as the amplitude of the eigenstate γ. Then quotient spaces of pictures become subspaces of wavefunctions.

Experts may note that central charge $c \neq 0$ is an obstruction to this "picture formulation": the mapping class group $\mathcal{M}(Y)$ acts directly on pictures and so induces an action on any $V(Y)$ defined by pictures. As c

determines a central extension $\widetilde{\mathcal{M}}^c(Y)$ which acts in place of $\mathcal{M}(Y)$, the feeling that all interesting theories must have $c \neq 0$ may have discouraged a pictorial approach. However this is not true: for any $V(Y)$ its endomorphism algebra End $V \cong V^* \bigotimes V$ has central charge $c = 0$ $(c(V^*) = -c(V))$ and remembers the original projective representation faithfully. In fact, all our examples are either of this form or slightly more subtle quantum doubles or Drinfeld centers in which the original theory V violates some axiom (the nonsingularity of the $S-$matrix) but this deficiency is "cured" by doubling[K].[Mu] Although those notes focus on picture TQFTs based on variations of the Jones-Wenzl projectors, the approach can be generalized to an arbitrary spherical tensor category. The Temperley-Lieb categories are generated by a single "fundamental" representation, and all fusion spaces are of dimension 0 or 1, so pictures are just 1-manifolds. In general, 1-manifolds need to be replaced by tri-valent graphs whose edges carry labels. But $c = 0$ is not sufficient for a TQFT to have a picture description. Given any two TQFTs with opposite central charges, their product has $c = 0$, e.g. TQFTs with $\mathbb{Z}_n$ fusion rules have $c = 1$, so the product of any theory with the mirror of a different one has $c = 0$, but such a product theory does not have a picture description in our sense.

While these notes describe the mathematical side of the story, we have avoided jargon which might throw off readers from physics. When different terminologies prevail within mathematics and physics we will try to note both. Within physics, TQFTs are referred to as "anyonic systems"[Wil].[DFNSS] These are 2-dimensional quantum mechanical systems with point like excitations (variously called "quas-particle" or just "particle", anyon, or perhaps "nonabelion") which under exchange exhibit exotic statistics: a nontrival representation of the braid groups acting on a finite dimensional Hilbert space V consisting of "internal degrees of freedom". Since these "internal degrees of freedom" sound mysterious, we note that this information is accessed by fusion: fuse pairs of anyons along a well defined trajectory and observe the outcome. Anyons are a feature of the fractional quantum Hall effect; Laughlin's 1998 Nobel prize was for the prediction of an anyon carrying change $e/3$ and with braiding statistics $e^{2\pi i/3}$. In the FQHE central charge $c \neq 0$ is enforced by a symmetry breaking magnetic field B. It is argued in[Fn] that solid state realizations of doubled or "picture" TQFTs may - if found - be more stable (larger spectral gap above the degenerate ground state manifold) because no symmetry breaking is required. The important electron - electron interactions would be at a lattice spacing scale $\sim 4\mathring{A}$ rather than at a "magnetic length" typically around $150\mathring{A}$. So it is hoped that the examples which are the subject of these notes will be the low energy limits of certain microscopic solid state models. Picture TQFTs have a Hamiltonian formulation, and describe string-net condensation in physics, which serve as a classification of non-chiral topological phases of matter. An interesting mathematical application is the proof of the asymptotic faithfulness of the representations of the mapping class groups.

As mentioned above, these notes are primarily about examples either of the form $V^* \bigotimes V$ or with a related but more general doubled structure $\mathcal{D}(V)$. In choosing a path through this material there seemed a basic choice: (1) present the picture (doubled) theories in a self contained way in two dimensions with no reference to their twisted ($c \neq 0$) and less tractable parent theories V or (2) weave the stories of $\mathcal{D}(V)$ and V together from the start and exploit the action of $\mathcal{D}(V)$ on V in analyzing the structure of $\mathcal{D}(V)$. In the end, the choice was made for us: we did not succeed in finding purely combinatorial "picture-proofs" for all the necessary lemmas — the action on V is indeed very useful so we follow course (2). We do recommend to some interested brave reader that she produce her own article hewing to course (1).

In the literature[BHMV] comes closest to the goals of the notes, and[Wal2] exploits deeply the picture theories in many directions. Actually, a large part of the notes will follow from a finished.[Wal2] If one applies the set up of [BHMV] to skeins in surface cross interval, $Y \times I$, and then resolves crossings to get a formal linear combination of $1-$submanifolds of $Y = Y \times \frac{1}{2} \subset Y \times I$ one arrives at (an example of) the "pictures" we study. In this doubled context there is no need for the p_1-structure (or "two-framing") intrinsic to the other approaches. To readers familiar with[BHMV] one should think of skeins in a handle body H, $\partial H = Y$, when an undoubled theory $V(Y)$ is being discussed, and skeins in $Y \times I$ when $\mathcal{D}V(Y)$ is under consideration.

By varying pictures and relations we produce many examples, and in the Temperley-Lieb-Jones context give a complete analysis of the possible local relations. Experts have long been troubled by certain sign discrepancies between the $S-$matrix arising from representations (or loop groups or quantum groups)[MS,Witt,KM] on the one hand and from the Kauffman bracket on the other[LiTu].[KL] The source of the discrepancy is that the fundamental representation of $SU(2)$ is anti-symmetrically self dual whereas there is no room in Kauffman's spin-network notation to record the antisymmetry. We rectify this by amplifying the pictures slightly, which yields exactly the modular functor V coming from representation theory of $SU(2)_q$.

The content of each section is as follows. In Sections 2, 3, we treat diagram TQFTs for closed manifolds. In Sections 4, 5, 7.1, we handle boundaries. In Sections 7, 9, 8, we cover the related Jones-Kauffman TQFTs, and the Witten-Reshetikhin-Turaev $SU(2)$-TQFTs which have anomaly, and non-trivial Frobenius-Schur indicators, respectively. In Section 10, we first prove the uniqueness of TQFTs based on Jones-Wenzl projectors, and then classify them according to the Kauffman variable A. A theory V or $\mathcal{D}(V)$ is unitary if the vector spaces V have natural positive definite Hermitian structures. Only unitary theories will have physical relevance so we decide for each theory if it is unitary.

2. Jones representations

2.1. *Braid statistics*

Statistics of elementary particles in 3-dimensional space is related to representations of the permutation groups S_n. Since the discovery of the fractional quantum Hall effect, the existence of anyons in 2-dimensional space becomes a real possibility. Statistics of anyons is described by unitary representations of the braid groups B_n. Therefore, it is important to understand unitary representations of the braid groups B_n. Statistics of n anyons is given by unitary representation of the n-strand braid group B_n. Since statistics of anyons of different numbers n is governed by the same local physics, unitary representations of B_n have to be compatible for different n's in order to become possible statistics of anyons. One such condition is that all representations of B_n come from the same unitary braided tensor category.

There is an exact sequence of groups: $1 \longrightarrow PB_n \longrightarrow B_n \longrightarrow S_n \longrightarrow 1$, where PB_n is the n-strand pure braid group. It follows that every representation of the permutation group S_n gives rise to a representation of the braid group B_n. An obvious fact for such representations of the braid groups is that the images are always finite. More interesting representations of B_n are those that do not factorize through S_n, in particular those with infinite images.

To construct representations of the braid groups B_n, we recall the construction of all finitely dimensional irreducible representations (irreps) of the permutation groups S_n: the group algebra $\mathbb{C}[S_n]$, as a representation of S_n, decomposes into irreps as $\mathbb{C}[S_n] \cong \bigoplus_i \mathbb{C}^{\dim V_i} \otimes V_i$, where the sum is over all irreps V_i of S_n. This construction cannot be generalized to B_n because B_n is an infinite group for $n \geq 2$. But by passing to various different finitely dimensional quotients of $\mathbb{C}[B_n]$, we obtain many interesting representations of the braid groups. This class of representations of B_n is Schur-Weyl dual to the the class of braid group representations from the quantum group approach and has the advantage of being manifestly unitary. This approach, pioneered by V. Jones,[Jo1] provides the best understood examples of unitary braid group representations besides the Burau representation, and leads to the discovery of the celebrated Jones polynomial of knots.[Jo2] The theories in this paper are related to the quantum $SU(2)_q$ theories.

2.2. *Generic Jones representation of the braid groups*

The n-strand braid group B_n has a standard presentation with generators $\{\sigma_i, i = 1, 2, \cdots, n-1\}$ and relations:

$$\sigma_i\sigma_j = \sigma_j\sigma_i, \quad \text{if} \quad |i-j| \geq 2, \tag{2.1}$$

$$\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}. \tag{2.2}$$

If we add the relations $\sigma_i^2 = 1$ for each i, we recover the standard presentation for S_n. In the group algebra $k[B_n]$, where k is a field (in this paper k will be either $\mathbb{C}$ or some rational functional field $\mathbb{C}(A)$ or $\mathbb{C}(q)$ over variables A or q), we may deform the relations $\sigma_i^2 = 1$ to linear combinations (superpositions in physical parlance) $\sigma_i^2 = a\sigma_i + b$ for some $a, b \in k$. By rescaling the relations, it is easy to show that there is only 1-parameter family of such deformations. The first interesting quotient algebras are the Hecke algebras of type A, denoted by $H_n(q)$, with generators $1, g_1, g_2, \cdots, g_{n-1}$ over $\mathbb{Q}(q)$ and relations:

$$g_ig_j = g_jg_i, \quad \text{if} \quad |i-j| \geq 2, \tag{2.3}$$

$$g_ig_{i+1}g_i = g_{i+1}g_ig_{i+1}. \tag{2.4}$$

and

$$g_i^2 = (q-1)g_i + q. \tag{2.5}$$

The Hecke relation 2.5 is normalized to have roots $\{-1, q\}$ when the corresponding quadratic equation is solved. The Hecke algebras $H_n(q)$ at $q = 1$ become $\mathbb{C}[S_n]$, hence they are deformations of $\mathbb{C}[S_n]$. When q is a variable, the irreps of $H_n(q)$ are in one-to-one correspondence with the irreps of $\mathbb{C}[S_n]$.

To obtain the Hecke algebras as quotients of $\mathbb{C}[B_n]$, we set $q = A^4$, and $g_i = A^3\sigma_i$, where A is a new variable, called the Kauffman variable since it is the conventional variable for the Kauffman bracket below. Note that

$q = A^2$ in.[KL] The prefactor A^3 is introduced to match the Hecke relation 2.5 exactly to a relation in the Temperley-Lieb algebras using the Kauffman bracket. In terms of the new variable A, and new generators σ_i's, the Hecke relation 2.5 becomes

$$\sigma_i^2 = (A - A^{-3})\sigma_i + A^{-2}. \tag{2.6}$$

The Kauffman bracket <> is defined by the resolution of a crossing in Figure 2.1

Fig. 2.1. Kauffman bracket

As a formula, $\sigma_i = A \cdot \mathrm{id} + A^{-1}U_i$, where U_i is a new generator. The Hecke algebra $H_n(q)$ in variable A and generators $1, U_1, U_2, \cdots, U_{n-1}$ is given by relations:

$$U_iU_j = U_jU_i, \quad \text{if} \quad |i-j| \geq 2, \tag{2.7}$$

$$U_iU_{i+1}U_i - U_i = U_{i+1}U_iU_{i+1} - U_{i+1}, \tag{2.8}$$

and

$$U_i^2 = dU_i, \tag{2.9}$$

where $d = -A^2 - A^{-2}$.

The relation 2.9 is the same as relation 2.6, which is the Hecke relation 2.5. The relation 2.8 is the braid relation 2.4.

The Temperley-Lieb (TL) algebras, denoted as $\mathrm{TL}_n(A)$, are further quotients of the Heck algebras. In the TL algebras, we replace the relations 2.8 by

$$U_iU_{i\pm1}U_i = U_i, \tag{2.10}$$

i.e., both sides of relation 2.8 are set to 0.

Prop 2.1.

The Kauffman bracket <>: $k[B_n] \longrightarrow \mathrm{TL}_n(A)$ is a surjective algebra homomorphism, where $k = \mathbb{C}(A)$.

The proof is a straightforward computation.

When A is generic, the TL algebras $\mathrm{TL}_n(A)$ are semi-simple, hence $\mathrm{TL}_n(A) \cong \oplus_i \mathrm{Mat}_{n_i}(\mathbb{C}(A))$, where Mat_{n_i} are $n_i \times n_i$ matrices over $\mathbb{C}(A)$ for some n_i's.

The generic Jones representation of the braid groups B_n is defined as follows:

Definition 2.1. By the decomposition $\mathrm{TL}_n(A) \cong \oplus_i \mathrm{Mat}_{n_i}(\mathbb{C}(A))$, each braid $\sigma \in B_n$ is mapped to a direct sum of matrices under the Kauffman bracket. It follows from Prop. 2.1 that the image matrix of any braid is invertible and the map is a group homomorphism when restricted to B_n.

It is an open question whether or not the generic Jones representation is faithful, i.e., are there non-trivial braids which are mapped to the identity matrix?

2.3. *Unitary Jones representations*

The TL algebras $\mathrm{TL}_n(A)$ have a beautiful picture description by L. Kauffman, inspired by R. Penrose's spin-networks, as follows: fix a rectangle $\mathcal{R}$ in the complex plane with n points at both the top and the bottom of $\mathcal{R}$ (see Fig.2.2), $\mathrm{TL}_n(A)$ is spanned formally as a vector space over $\mathbb{C}(A)$ by embedded curves in the interior of $\mathcal{R}$ consisting of n disjoint arcs connecting the $2n$ boundary points of $\mathcal{R}$ and any number of simple closed loops. Such an embedding will be called a diagram or a multi-curve in physical language, and a linear combination of diagrams will be called a formal diagram. Two diagrams that are isotopic relative to boundary points represent the same vector in $\mathrm{TL}_n(A)$. To define the algebra structure, we introduce a multiplication: vertical stacking from bottom to top of diagrams and extending bilinearly to formal diagrams; furthermore, deleting a closed loop must be compensated for by multiplication by $d = -A^2 - A^{-2}$. Isotopy and the deletion rule of a closed trivial loop together will be called "d-isotopy".

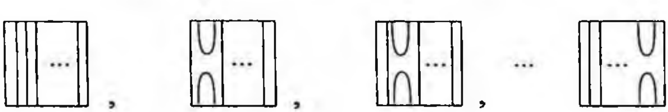


Fig. 2.2. Generators of TL

For our application, the variable A will be evaluated at a non-zero complex number. We will see later that when $d = -A^2 - A^{-2}$ is not a root of a Chebyshev polynomial Δ_i, $\mathrm{TL}_n(A)$ is semi-simple over $\mathbb{C}$, therefore, isomorphic to a matrix algebra. But when d is a root of some Chebyshev polynomial, $\mathrm{TL}_n(A)$ is in general not semi-simple. Jones discovered a semi-simple quotient by introducing local relations, called the Jones-Wenzl projectors[Jo4We].[KL] Jones-Wenzl projectors have certain rigidity. Represented by formal diagrams in TL algebras, Jones-Wenzl projectors make it possible to describe two families of TQFTs labelled by integers. Conventionally the integer is either $r \geq 3$ or $k = r-2 \geq 1$. The integer r is related to the order of A, and k is the level related to the $SU(2)$-Witten-Chern-Simons theory. One family is related to the $SU(2)_k$-Witten-Reshetikhin-Turaev (WRT) TQFTs, and will be called the Jones-Kauffman TQFTs. Although Jones-Kauffman TQFTs are commonly stated as the same as WRT TQFTs, they are really not. The other family is related to the quantum double of Jones-Kauffman TQFTs, which are of the Turaev-Viro type. Those doubled TQFTs, labelled by a level $k \geq 1$, are among the easiest in a sense, and will be called diagram TQFTs. The level $k = 1$ diagram TQFT for closed surfaces is the group algebras of $\mathbb{Z}_2$-homology of surfaces. Therefore, higher level diagram TQFTs can be thought as quantum generalizations of the $\mathbb{Z}_2$-homology, and the Jones-Wenzl projectors as the generalizations of the homologous relation of curves in Figure 2.3.

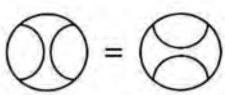


Fig. 2.3. Z_2 homology

The loop values $d = -A^2 - A^{-2}$ play fundamental roles in the study of Temperley-Lieb-Jones theories, in particular the picture version of $\mathrm{TL}_n(A)$ can be defined over $\mathbb{C}(d)$, so we will also use the notation $\mathrm{TL}_n(d)$. In the following, we focus the discussion on d, though for full TQFTs or the discussion of braids in $\mathrm{TL}_n(A)$, we need A's. Essential to the proof and to the understanding of the exceptional values of d is the trace tr: $\mathrm{TL}_n(d) \longrightarrow \mathbb{C}$ defined by Fig.2.4. This Markov trace is defined on diagrams by (and then extended linearly) connecting the endpoints at the top to the endpoints at the bottom of the rectangle by n non crossing arcs in the complement of the rectangle, counting the number # of closed loops (deleting the rectangle), and then forming $d^{\#}$.

The Markov trace $(x, y) \mapsto \mathrm{tr}(\overline{x}y)$ extends to a sesquilinear pairing on

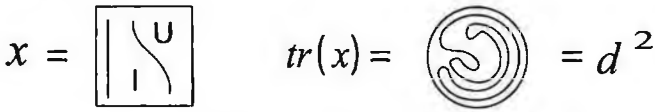


Fig. 2.4. Markov Trace

$\mathrm{TL}_n(d)$, where bar (diagram) is reflection in a horizontal middle-line and bar(coefficient) is complex conjugation.

Define the n^{th} Chebyshev polynomial $\Delta_n(x)$ inductively by $\Delta_0 = 1, \Delta_1 = x$, and $\Delta_{n+1}(x) = x\Delta_n(x) - \Delta_{n-1}(x)$. Let c_n be the Catalan number $c_n = \frac{1}{n+1}\binom{2n}{n}$. There are c_n different diagrams $\{D_i\}$ consisting of n disjoint arcs up to isotopy in the rectangle $\mathcal{R}$ to connect the $2n$ boundary points of $\mathcal{R}$. These c_n diagrams generate $\mathrm{TL}_n(d)$ as a vector space. Let $M_{c_n \times c_n} = (m_{ij})$ be the matrix of the Markov trace Hermitian pairing in a certain order of $\{D_i\}$, i.e. $m_{ij} = \mathrm{tr}(\overline{D_i}D_j)$, then we have:

$$\mathrm{Det}(M_{c_n \times c_n}) = \prod_{i=1}^{n} \Delta_i(d)^{a_{n,i}}, \tag{2.11}$$

where $a_{n,i} = \binom{2n}{n-i-2} + \binom{2n}{n-i} - 2\binom{2n}{n-i-1}$.
This is derived in.[DGG]

As a quick consequence of this formula, we have:

Lemma 2.1. *The dimension of $TL_n(d)$ as a vector space over $\mathbb{C}(d)$ is c_n if d is not a root of the Chebyshev polynomials $\Delta_i, 1 \leq i \leq n$, where $c_n = \frac{1}{n+1}\binom{2n}{n}$.*

Proof. By the formula 2.11, if d is not a root of $\Delta_i, 1 \leq i \leq n$, then $\{D_i\}$ are linearly independent. As a remark, since each D_i is a monomial of U_i's, it follows that $\{U_i\}$ generate $\mathrm{TL}_n(d)$ as an algebra. □

Next we show the existence and uniqueness of the Jones-Wenzl projectors.

Theorem 2.1. *For $d \in \mathbb{C}$ that is not a root of Δ_k for all $k < n$, then $\mathrm{TL}_n(d)$ contains a unique element p_n characterized by: $p_n^2 = p_n \neq 0$ and $U_i p_n = p_n U_i = 0$ for all $1 \leq i \leq n-1$. Furthermore p_n can be written as $p_n = 1 + U$ where $U = \sum c_j h_j$, h_j a product of $U_i^{'}s$, $1 \leq i \leq n-1$ and $c_j \in \mathbb{C}$.*

Proof. Suppose p_n exists and can be expanded as $p_n = a1 + U$, then $p_n^2 = p_n(a1 + U) = p_n(a1) = ap_n = a^2 1 + aU$, so $a = 1$. Now check uniqueness by supposing $p_n = 1+U$ and $p_n^{'} = 1+V$ both have the properties above and expand $p_n p_n^{'}$ from both sides:

$$p_n^{'} = 1 \cdot p_n^{'} = (1 + U)p_n^{'} = p_n p_n^{'} = p_n(1 + V) = p_n \cdot 1 = p_n.$$

The proof is completed by H. Wenzl's[We] inductive construction of p_{n+1} from p_n which also reveals the exact nature of the "generic" restriction on d. The induction is given in Figure 2.5, where $\mu_n = \frac{\Delta_{n-1}(d)}{\Delta_n(d)}$.

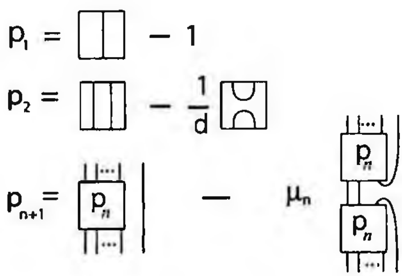


Fig. 2.5. Jones Wenzl projectors

Tracing the inductive definition of p_{n+1} yields $\mathrm{tr}(p_{n+1}) = d\ \mathrm{tr}(p_n) - \frac{\Delta_{n-1}}{\Delta_n}\mathrm{tr}(p_n)$ showing $\mathrm{tr}(p_n)$ satisfies the Chebyshev recursion (and the initial data). Thus $\mathrm{tr}(p_n) = \Delta_n$.

It is not difficult to check that $U_i p_n = p_n U_i = 0$, $i < n$. (The most interesting case is U_{n-1}.) Consult[KL] or[Tu] for details. □

The idempotent p_n is called the *Jones-Wenzl idempotent*, or the Jones-Wenzl projector, and plays an indispensable role in the pictorial approach to TQFTs.

Theorem 2.2. *(1): If $d \in \mathbb{C}$ is not a root of Chebyshev polynomials $\Delta_i, 1 \leq i \leq n$, then the TL algebra $\mathrm{TL}_n(d)$ is semisimple.*

(2): Fixing an integer $r \geq 3$, a non-zero number d is a root of Δ_i, $i < r$ if and only if $d = -A^2 - A^{-2}$ for some A such that $A^{4l} = 1, l \leq r$. If $d = -A^2 - A^{-2}$ for a primitive $4r$-th root of unity A for some $r \geq 3$ or a primitive $2r$th or rth for r odd, then the TL algebras $\{\mathrm{TL}_n(d)\}$ modulo the Jones-Wenzl idempotent p_{r-1} are semi-simple.

Proof.

(1): $\mathrm{TL}_n(d)$ is a $*$-algebra. By formula 2.11, the determinant of the Markov trace pairing is $\prod_{i=1}^{n} \Delta_i(d)^{a_{n,i}}$, hence the $*$-structure is non-degenerate. By Lemma B.2, $\mathrm{TL}_n(d)$ is semi-simple.

(2): The first part follows from $\Delta_n(d) = (-1)^n \frac{A^{2n+2}-A^{-2n-2}}{A^2-A^{-2}}$. In Section 5, we will show that the kernel of the Markov trace Hermitian pairing is generated by p_{r-1}, and the second part follows. □

The semi-simple quotients of $\mathrm{TL}_n(d)$ in the above theorem will be called the Temperley-Lieb-Jones (TLJ) algebras or just Jones algebras, denoted by $\mathrm{TLJ}_n(d)$. The TLJ algebras are semi-simple algebras over $\mathbb{C}$, therefore it is isomorphic to a direct sum of matrix algebras, i.e.,

$$\mathrm{TLJ}_n(d) \cong \oplus_i Mat_{n_i}(\mathbb{C}). \tag{2.12}$$

As in the generic Jones representation case, the Kauffman bracket followed by the decomposition yields a representation of the braid groups.

Prop 2.2.

(1): When the Markov trace Hermitian paring is $\pm$-definite, then Jones representations are unitary, but reducible. When $A = \pm ie^{\pm\frac{2\pi i}{4r}}$, the Markov trace Hermitian pairing is $+$-definite for all n's.

(2): Given a braid $\sigma \in B_n$, the Markov trace is a weighted trace on the matrix decomposition 2.12, and when multiplied by $(-A)^{-3\sigma}$ results in the Jones polynomial of the braid closure of σ evaluated at $q = A^4$.

Unitary will be established in Section 10, and reducibility follows from the decomposition 2.12. That the Markov trace, normalized by the framing-dependence factor, is the Jones polynomial follows from direct verification of invariance under Reidermeister moves or Markov's theorem (see e.g.[KL]).

2.4. *Uniqueness of Jones-Wenzl projectors*

Fix an $r \geq 3$ and a primitive $4r$th root of unity or a primitive $2r$th or rth root of unity for r odd, and $d = -A^2 - A^{-2}$. In this section, we prove that TL_d has a unique ideal generated by p_{r-1}. When A is a primitive $4r$th root of unity, this is proved in the Appendix of[Fn] by F. Goodman and H. Wenzl. Our elementary argument works for all A as above.

Notice that TL_d admits the structure of a (strict) monoidal category, with the tensor product given by horizontal "stacking", e.g., juxtaposition of diagrams. This tensor product (denoted $\otimes$) is clearly associative, and 1_0, the identity on 0 vertices or the empty object, serves as a unit. The tensor product and the original algebra product on TL_d satisfy the interchange law, $(f \otimes g) \cdot (f' \otimes g') = (f \cdot f') \otimes (g \cdot g')$, whenever the required vertical composites are defined.

We may use this notation to recursively define the projectors p_k: $p_{k+1} = p_k \otimes 1_1 - \mu_k(p_k \otimes 1_1)U_k^{k+1}(p_k \otimes 1_1)$. We define $p_0 = 1_0$, $p_1 = 1_1$ and $\mu_k = \frac{\Delta_{k-1}}{\Delta_k}$. Using this we can prove a sort of "decomposition theorem" for projectors:

Prop 2.3. $p_k = \left(\bigotimes_{i=1}^{\lfloor \frac{k}{r} \rfloor} p_r\right) \otimes p_{(k \bmod r)}$.

Proof. We proceed by induction, using the recursive definition of the Jones-Wenzl projectors. For p_1 the statement is trivial. Assuming the assertion holds for p_k, we then have (let $m = k \bmod r$):

$$p_{k+1} = p_k \otimes 1_1 - \mu_k(p_k \otimes 1_1)U_k^{k+1}(p_k \otimes 1_1)$$

$$= (\bigotimes_{i=1}^{\lfloor \frac{k}{r} \rfloor} p_r) \otimes p_m \otimes 1_1 - \mu_k((\bigotimes_{i=1}^{\lfloor \frac{k}{r} \rfloor} p_r) \otimes p_m \otimes 1_1)U_k^{k+1}((\bigotimes_{i=1}^{\lfloor \frac{k}{r} \rfloor} p_r) \otimes p_m \otimes 1_1)$$

Then, if $m \neq 0$,

$$= (\bigotimes_{i=1}^{\lfloor \frac{k}{r} \rfloor} p_r) \otimes p_m \otimes 1_1 -$$

$$\mu_k((\bigotimes_{i=1}^{\lfloor \frac{k}{r} \rfloor} p_r) \otimes p_m \otimes 1_1)(1_{k-m} \otimes U_m^{m+1})((\bigotimes_{i=1}^{\lfloor \frac{k}{r} \rfloor} p_r) \otimes p_m \otimes 1_1)$$

$$= (\bigotimes_{i=1}^{\lfloor \frac{k}{r} \rfloor} p_r) \otimes \left(p_m \otimes 1_1 - \mu_m(p_m \otimes 1_1)U_m^{m+1}(p_m \otimes 1_1)\right)$$

(The $\lfloor \frac{k}{r} \rfloor$ copies of p_r can be factored out of the second term by $p_r p_r = p_r$.)

$$= (\bigotimes_{i=1}^{\lfloor \frac{k}{r} \rfloor} p_r) \otimes p_{m+1}$$

If $m < r - 1$, then $m + 1 = (k + 1) \bmod r$; if $m = r - 1$, then we get one more copy of p_r, as needed. So it remains to consider the case above where $k \bmod r = 0$. But then $\mu_k = \mu_{k \bmod r} = \mu_0 = 0$, so that

$$p_{k+1} = (\bigotimes_{i=1}^{\frac{k}{r}} p_r) \otimes 1_1 = (\bigotimes_{i=1}^{\lfloor \frac{k+1}{r} \rfloor} p_r) \otimes p_1$$

as desired. $\square$

In analogy with the standard notion from ring theory, an ideal in TL is defined to be a class of morphisms which is internally closed under addition, and externally closed under both the vertical product (composition) $\cdot$ and the horizontal product $\otimes$. Given such an ideal I, we may form the quotient category TL$/I$, which has the same objects as TL, and hom-sets formed by taking the usual quotient of $\mathrm{Hom}(\mathbf{m},\mathbf{n})$ by those morphisms in $I \cap \mathrm{Hom}(\mathbf{m},\mathbf{n})$.

We can prove that $<p_{r-1}>$ is an ideal.

Lemma 2.2. *The ideal* $\mathcal{R}_d = <p_{r-1}>$ *is a proper ideal.*

Proof. It suffices to show that the $\otimes$-identity 1_0 is not in the ideal. In order for 1_0 to be in the ideal, it would have to be obtained from some closed network (e.g., element of $\mathrm{Hom}(\mathbf{0},\mathbf{0})$) which contains at least one copy of p_{r-1}. Fixing such a projector, we expand all other terms in the network (this includes getting rid of closed loops), so that we are left with a linear combination of closed networks, each having exactly one $r-1$ strand projector. Now, considering each term seperately, if there are any strands that leave and re-enter the projector on the same side, then the network is null (since $p_{r-1}U_i^{r-1} = 0$). So the only remaining terms will be strand closures of p_{r-1}; but by the above, these are null as well, so that every term in the expansion vanishes.

Since every closed network with p_{r-1} is null, it follows that $1_0 \notin \mathcal{R}_d$, and therefore $\mathcal{R}_d$ is a proper ideal of TL. □

In fact, this same ideal is generated by any p_k for $k \geq r-1$; this is established via a sequence of lemmas.

Lemma 2.3. $<p_r>=<p_{r-1}>$

Proof. It is clearly sufficient to show $p_{r-1} \in <p_r>$. Set $x = p_r \otimes 1_1$, and expand p_r in terms of p_{r-1} according to the recursive definition. Then connect the rightmost two strands in a loop: e.g., by pre- and post-multiplying by the appropriate elements of $\mathrm{Hom}(\mathbf{r-1},\mathbf{r+1})$ and $\mathrm{Hom}(\mathbf{r+1},\mathbf{r-1})$, respectively. Using the fact that $p_{r-1}p_{r-1} = p_{r-1}$, the resulting diagram simplifies to $(d-\mu_{r-1})p_{r-1}$; and since $\mu_{r-1} \neq d$, the coefficient is invertible, so that $p_{r-1} \in <p_r>$. □

Lemma 2.4. *For any integer* $k \geq 1$, $<p_{kr}>=<p_{r-1}>$.

Proof. By induction; the base case is established in the previous Lemma. For $k \geq 2$, we can write $p_{kr} = p_{(k-2)r} \otimes p_r \otimes p_r$, and then consider the tangle $p_{kr} \otimes 1_r$. By again pre- and post-multiplying by appropriate tangles, and using $p_r p_r = p_r$, we see that $p_{(k-1)r} \in <p_{kr}>$. □

Lemma 2.5. *For any $k \geq r-1$, $<p_k>=<p_{r-1}>$.*

Proof. This basically uses the same technique as the previous lemma, combined with the fact that $p_r(p_l \otimes 1_{r-l}) = p_r$ (see[KL]).

Let $m = k \bmod r$; if $m = 0$, then this falls under the case of the previous lemma, so $0 < m < r$. Now consider $x = p_k \otimes 1_{2r-m}$; we can use the technique of the previous lemma to merge the last three groups of r strands into one, so that the resulting element $x' = p_{\lfloor \frac{k}{r} \rfloor r} \otimes (p_r(p_l \otimes 1_{r-l})1_r)$. But $p_r(p_l \otimes 1_{r-l}) = p_r$, so that $x' = p_{(\lfloor \frac{k}{r} \rfloor + 1)r}$, whence, by the previous lemma, $<p_{r-1}>=<p_k>$. □

Thus, in the quotient category $\mathrm{TL}/\mathcal{R}_d$, all k-projectors, for $k \geq r-1$, are null.

We have shown that $\mathcal{R}_d$ is an ideal; our strategy in showing that $\mathcal{R}_d$ is unique will be to show that it has no proper ideals, and that the quotient $\mathrm{TL}/\mathcal{R}_d$ has no nontrivial ideals. To show the latter fact, we will show that the ideal (in the quotient) generated by any element is in fact all of $\mathrm{TL}/\mathcal{R}_d$.

We note also that $\mathrm{TL}/\mathcal{R}_d$ may be described succinctly as the subcategory of TL whose tangles have less than $r-1$ "through-passing" strands. This subcategory does not close under $\otimes$ as described above, but can be shown to be well-defined under the reduction $1_{r-1} \rightsquigarrow (1_{r-1} - p_{r-1})$. This view is not necessary in what follows, so we do not pursue it further; but it may be useful in thinking about the quotient category.

A preliminary observation is that $\mathrm{TL}/\mathcal{R}_d$ has no zero divisors:

Lemma 2.6. *Let $x, y \in TL/\mathcal{R}_d$. If $x \otimes y = 0$, then $x = 0$ or $y = 0$.*

Proof. The statement clearly holds in TL; so the only way it could fail in the quotient is if p_{r-1} had a tensor decomposition.

So, suppose, $x \otimes y = p_{r-1}$, where x is a tangle on $k > 0$ strands and y is a tangle on $l > 0$ strands, both nontrivial (that $\mathrm{dom}(x) = \mathrm{cod}(x)$ and $\mathrm{dom}(y) = \mathrm{cod}(y)$ follows from the fact that $p_{r-1}p_{r-1} = p_{r-1}$). Then the properties of projectors and the interchange law give:

$$x \otimes y = (x \otimes y)(x \otimes y) = xx \otimes yy \Longrightarrow xx = x, yy = y$$

Further $(x \otimes y)U_i^{k+l} = 0$ for all i, so that $xU_i^k \otimes y = 0 \Longrightarrow xU_i^k = 0$, and likewise $yU_i^l = 0$. Thus both x and y are projectors. But the strand closure of $p_k \otimes p_l$ is $\Delta_k \Delta_l$, which are both nonzero, and the strand closure of p_{r-1} is zero, so we have reached a contradiction. □

The next lemma introduces an algorithm that is the key to the rest of the proof:

Lemma 2.7. *Any nonzero ideal $I \subset TL/\mathcal{R}_d$ contains at least one element of* $\mathrm{Hom}(\mathbf{r-3}, \mathbf{r-3})$.

Proof. Let $x \neq 0 \in I$, say $x \in \mathrm{Hom}(\mathbf{m}, \mathbf{n})$. First, if $m \neq n$, then we can tensor with the unique basis element in either $\mathrm{Hom}(\mathbf{0}, \mathbf{2})$ or $\mathrm{Hom}(\mathbf{2}, \mathbf{0})$ the appropriate number of times so that we get an $x' \in \mathrm{Hom}(\mathbf{k}, \mathbf{k}) \in \mathbf{I}$, where $k = \max\{m, n\}$. (By the previous lemma, $x' \neq 0$.) If $k \leq r-3$, then $x' \otimes 1_{r-3-k} \in \mathrm{Hom}(\mathbf{r-3}, \mathbf{r-3})$ is an element of the ideal; so it remains to show the case where $k > r-3$.

First, assume k and $r-3$ have the same parity; if not, use $x' \otimes 1_1$ instead of x'. Then let $k_0 = k, x'_0 = x'$, and use the following algorithm (starting with $i = 0$):

(1) If $k_i = r-3$, then stop: $x'_i \in \mathrm{Hom}(\mathbf{r-3}, \mathbf{r-3})$ is in the ideal.
(2) Since $k_i \geq r-1$, and $x'_i \neq 0$, it follows that $x'_i \neq \alpha p_{k_i}$, since all $r-1$ and above projectors are null in $\mathrm{TL}/\mathcal{R}_d$. Recall that p_{k_i} is the unique element in $\mathrm{Hom}(\mathbf{k_i}, \mathbf{k_i})$ such that (i) $U_j^{k_i} p_{k_i} = p_{k_i} U_j^{k_i} = 0$ for $1 \leq j < k_i$; and (ii) $p_{k_i} p_{k_i} = p_{k_i}$. From this it follows that the only elements which satisfy (i) are αp_{k_i}, for some $\alpha \in \mathbb{C}$. Therefore, since $x'_i \neq \alpha p_{k_i}$, there exists some $U_i = U_{j_i}^{k_i}$ such that $U_i x' \neq 0$.
(3) Using an argument similar to the above, there exists some $U'_i = U_{j'_i}^{k_i}$ such that $(U_i x') U'_i \neq 0$.
(4) Set V_i to be the unique basis element in $\mathrm{Hom}(\mathbf{k_i - 2}, \mathbf{k_i})$ which connects the j_i and $j_i + 1$ vertices on the top (codomain) objects, and connects the remaining $k-2$ vertices on top and bottom to each other. Then $V_i U_i x' U'_i$ can be described as being exactly like $U_i x' U'_i$, except that the top half-loop of the U_i has been factored out as d, thus reducing the domain object by two vertices. It is thus clear that $V_i U_i x' U'_i \neq 0$.
(5) Similarly, choose V'_i to be the unique element in $\mathrm{Hom}(\mathbf{k_i}, \mathbf{k_{i-2}})$ connecting the j'_i and $j'_i + 1$ vertices of the domain object, thus closing the half loop of U'_i. Then $V_i U_i x' U'_i V'_i \neq 0$.
(6) Set $x'_{i+1} = V_i U_i x' U'_i V'_i$, $k_{i+1} = k_i - 2$, and return to step (1).

After $j = \frac{1}{2}(k_0 - (r-3))$ passes through the algorithm, the desired element $x'_j \in I$ is produced. □

The proof of the previous Lemma is useful in establishing that $\mathcal{R}_d$ has no proper sub-ideals.

Lemma 2.8. *For any $x \in \mathcal{R}_d, x \neq 0$, then $<x>= \mathcal{R}_d$.*

Proof. Use the techniques previous Lemma to get an element $x' \in <x>$ such that $x' \in \mathrm{Hom}(\mathbf{k}, \mathbf{k})$, and $k \equiv r-1 \bmod 2$. Then follow the algorithm, except for on steps (2) and (3): for, since $x'_i \in \mathcal{R}_d$, it is possible that $x'_i = \alpha p_{k_i}$. If this is not the case, proceed with the algorithm as it is stated. However, if $x'_i = \alpha p_{k_i}$, then it follows that $<x>= \mathcal{R}_d$, by Lemma 2.5. So it only remains to show that this does happen at some point before the algorithm terminates: e.g., that for some i, $x'_i = \alpha p_{k_i}$.

But, suppose this didn't happen; then, the algorithm goes through to completion, yielding an element $y \in <x>$ such that $y \in \mathrm{Hom}(\mathbf{r-3}, \mathbf{r-3})$, $y \neq 0$. But then $y \notin \mathcal{R}_d$, since every nonzero element of $\mathcal{R}_d$ must have at least $r-1$ strands. This contradicts the fact that $y \in <x> \subset \mathcal{R}_d$; therefore, there must be some i such that $x'_i = \alpha p_{k_i}$, and so the lemma follows. □

Now we can put all of this together to obtain our desired result:

Theorem 2.3. *TL_d has a unique proper nonzero ideal when A is as in Lemma 3.1.*

Proof. By Lemma 2.2, $\mathcal{R}_d = <p_{r-1}>$ is a proper ideal, which, by Lemma 2.8, has no proper sub-ideals. To prove the theorem, therefore, it suffices to show that the quotient category $\mathrm{TL}/\mathcal{R}_d$ has no proper nonzero ideals.

Consider $<x>$, for any $x \in \mathrm{TL}/\mathcal{R}_d$. By Lemma 2.7, there exists some $y \in <x>$ such that $x \in \mathrm{Hom}(\mathbf{r-3}, \mathbf{r-3})$. But now, instead of stopping at this point in the algorithm, we continue the loop, with the possibility that x'_i might actually be a projector. So we again modify steps (2) and (3), as below:

(1') If $k_i = 0$, stop; set $x' = x'_i$.

(2') If $x'_i = \alpha p_{k_i}$ for some constant α, then stop, with $x' = x'_i$. Otherwise, proceed with step (2) of the original algorithm.

(3') If $U_i x'_i = \alpha p_{k_i}$ for some constant α, then stop, with $x' = x'_i$. Otherwise, proceed with step (3) of the original algorithm.

So now, when the algorithm terminates, we are left with some $x' \in <x>$, with either: (a) $x' = \alpha 1_0$, $\alpha \neq 0$; or (b) $x' = \alpha p_k$ for some $1 \leq k < r-1$, $\alpha \neq 0$. In case (a), we have that $1_0 \in <x>$, so that $<x> = \mathrm{TL}/\mathcal{R}_d$. In case (b), consider the element $y = \frac{1}{\alpha}x' \otimes 1_k = p_k \otimes 1_k \in <x>$. We can then pre- and post-multiply by the elements of $\mathrm{Hom}(\mathbf{0}, \mathbf{2k})$ and $\mathrm{Hom}(\mathbf{2k}, \mathbf{0})$, respectively, which join the left group of k strands to the right group of k strands. In other words, the resulting element is simply Δ_k, the strand closure of p_k, times 1_0. Since $\Delta_k \neq 0$ for $1 \leq k \leq r-2$, it follows that $1_0 \in <x>$, so that we still have $<x> = \mathrm{TL}/\mathcal{R}_d$.

So we have shown that $\mathrm{TL}/\mathcal{R}_d$ has no proper nonzero ideals, and therefore, that TL has the unique ideal $\mathcal{R}_d$. □

As a corollary, we have the following:

Theorem 2.4. *1):If d is not a root of any Chebyshev polynomial $\Delta_k, k \geq 1$, then the Temperley-Lieb category TL_d is semisimple.*

2): Fixing an integer $r \geq 3$, a non-zero number d is a root of Δ_k, $k < r$ if and only if $d = -A^2 - A^{-2}$ for some A such that $A^{4l} = 1, l \leq r$. If $d = -A^2 - A^{-2}$ for a primitive $4r$-th root of unity A or $2r$-th r odd or r-th r odd for some $r \geq 3$, then the tensor category TLJ_d has a unique nontrivial ideal generated by the Jones-Wenzl idempotent p_{r-1}. The quotient categories TLJ_d are semi-simple.

3. Diagram TQFTs for closed manifolds

3.1. *"d-isotopy", local relation, and skein relation*

Let Y be an oriented compact surface, and $\gamma \subset Y$ be an imbedded unoriented 1-dimensional submanifold. If $\partial Y \neq \phi$ then fix a finite set F of points on ∂Y and require $\partial\gamma = F$ transversely. That is, γ a disjoint union of non-crossing loops and arcs, a "multi-curve". Let $\mathcal{S}$ be the set of such γ's. To "linearize" we consider the complex span $\mathbb{C}[\mathcal{S}]$ of $\mathcal{S}$, and then impose linear relations. We always impose the "isotopy" constraint $\gamma' = \gamma$, if γ' is isotopic to γ. We also always impose a constraint of the form $\gamma \cup \mathbf{O} = d \cdot \gamma$ for some $d \in \mathbb{C}\backslash\{0\}$, independent of γ (see an example below that we do not impose this relation). The notation $\gamma \cup \mathbf{O}$ means a multi curve made from γ by adding a disjoint loop $\mathbf{O}$ to γ where $\mathbf{O}$ is "trivial" in the sense that it is the boundary of a disk B^2 in the interior of Y. Taken together these two constraints are "d-isotopy" relation: $\gamma' - \frac{1}{d} \cdot (\gamma \cup \mathbf{O}) = 0$ if γ' is isotopic to γ.

A diagram *local relation* or just a local relation is a linear relation on multicurves $\gamma_1, \ldots, \gamma_m$ which are identical outside some disk B^2 in the interior of Y, and intersect ∂B^2 transversely. By a disk here, we mean a topological disk, i.e., any diffeomorphic image of the standard 2-disk in the plane. Local relations are usually drawn by illustrating how the γ_i differ on B^2. So the "isotopy" constraint has the form:

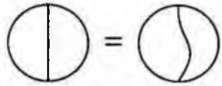


Fig. 3.1. Isotopy

and the "$d-$constraint" has the form:

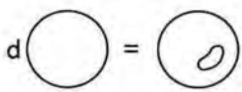


Fig. 3.2. d constraint

Local relations have been explored to a great generality in[Wal2] and encode information of topologically invariant partition functions of a ball. We may filter a local relation according to the number of points of $\gamma_i \cap \partial B^2$ which may be $0, 2, 4, 6, \ldots$ since we assume γ transverse to ∂B^2. "Isotopy" has degree $= 2$ and "d-constraint" degree $= 0$.

Formally, we define a local relation and a skein relation as follows:

Definition 3.1.

(1) Let $\{D_i\}$ be all the diagrams on a disk B^2 up to diffeomorphisms of the disk and without any loops. The diagrams $\{D_i\}$ are filtered into degrees $= 2n$ according to how many points of $D_i \cap \partial B^2$, and there are Catalan number c_n many diagrams of degree $2n$ ($c_0 = 1$ which is the empty diagram). A degree $= 2n$ diagram local relation is a formal linear equation of diagrams $\sum_i c_i D_i = 0$, where $c_i \in \mathbb{C}$, and $c_i = 0$ if D_i is not of degree $= 2n$.

(2) A skein relation is a resolution of over-/under-crossings into formal pictures on B^2. If the resolutions of crossings for a skein relation are all formal diagrams, then the skein relation induces a *set* map from $\mathbb{C}[B_n]$ to $TL_n(d)$.

The most interesting diagram local relations are the Jones-Wenzl projectors (the rectangle $\mathcal{R}$ is identified with a disk B^2 in an arbitrary way). When we impose a local relation on $\mathbb{C}[\mathcal{S}]$, we get a quotient vector space of $\mathbb{C}[\mathcal{S}]$ as follows: for any multi-curve γ and a disk B^2 in the interior of Σ, if γ intersects B^2 transversely and the part $\gamma \cap B^2$ of γ in B^2 matches one of the diagram D_j topologically in the local relation $\sum_i c_i D_i = 0$, and $c_j \neq 0$, then we set $\gamma = -\sum_{i \neq j} \frac{c_i}{c_j} \gamma_i'$ in $\mathbb{C}[\mathcal{S}]$ where γ_i' is obtained from γ by replacing the part $\gamma \cap B^2$ of γ in B^2 by the diagram D_i.

Kauffman bracket is the most interesting skein relation in this paper. More general skein relations can be obtained from minimal polynomials of R-matrices from a quantum group. Kauffman bracket is an unoriented version of the $SU(2)_q$ case.

As a digression we describe an unusual example where we impose "isotopy" but not the "d-constraint". It is motivated by the theory of finite type invariants. A singular crossing (outside $\mathcal{S}$) suggests the "type 1" relation in Figure 2.3.

This relation is closely related to $\mathbb{Z}_2-$homology and is compatible with the choice $d = 1$. We will revisit it again under the name $\mathbb{Z}_2-$gauge theory.

Now consider the "type 2" relation Figure 3.4 which comes by resolving the arc in Figure 3.3 using either arrow along the arc. (Reversing the arrow leaves the relation Figure 3.4 on unoriented diagrams unchanged.)

Fig. 3.3. Singular arc

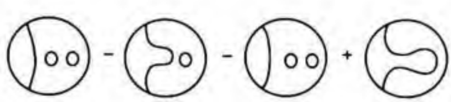


Fig. 3.4. Resolution relation

Formally we may write the resolution relation Figure 3.4 as the square of the 2 term relation drawn in Figure 3.5.

Interpreting "times" as "vertical stacking" makes the claim immediate as shown in Figure 3.6.

Since the two term relation Figure 3.5 does not appear to be a conse-

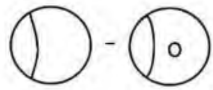


Fig. 3.5. Two term relation

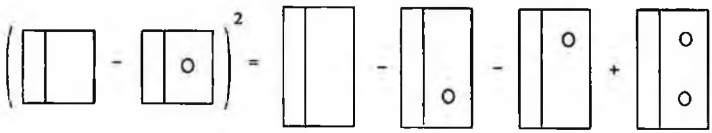


Fig. 3.6. Two term squared

quence of the resolution relation Figure 3.4, dividing by the resolution relation induces nilpotence in the algebra (of degree = 2 diagrams under vertical stacking). By imposing the "d" relation we find that only semi-simple algebras are encountered. This is closer to the physics (the simple pieces are symmetries of a fixed particle type or "super-selection sector") and easier mathematically so henceforth we always assume a "d–constraint" for some $d \in \mathbb{C}\backslash\{0\}$.

3.2. *Picture classes*

Fix a local relation $R = 0$. Given an oriented closed surface Y. The vector space $\mathbb{C}[\mathcal{S}]$ is infinitely dimensional. We define a finitely dimensional quotient of $\mathbb{C}[\mathcal{S}]$ by imposing the local relation R as in last section: $\mathbb{C}[\mathcal{S}]$ modulo the local relation. The resulting quotient vector space will be called the picture space, denoted as $\mathrm{Pic}^R(Y)$. Elements of $\mathrm{Pic}^R(Y)$ will be called picture classes. We will denote $\mathrm{Pic}^R(Y)$ as $\mathrm{Pic}(Y)$ when R is clear or irrelevant for the discussion.

Prop 3.1.

(1) $\mathrm{Pic}(Y)$ is independent of the orientation of Y.
(2) $\mathrm{Pic}(S^2) = \mathbb{C}\emptyset$, so it is either 0 or $\mathbb{C}$.
(3) $\mathrm{Pic}(Y_1 \amalg Y_2) \cong \mathrm{Pic}(Y_1) \otimes \mathrm{Pic}(Y_2)$.
(4) $\mathrm{Pic}(Y)$ is a representation of the mapping class group $\mathcal{M}(Y)$. Furthermore, the action of $\mathcal{M}(Y)$ is compatible with property (3).

Proof.

Properties (1) (3) and (4) are obvious from the definition. For (2), since every simple closed curve on S^2 bounds a disk, a multicurve with m loops is $d^m\emptyset$ by "d-isotopy". Therefore, if $\emptyset$ is not 0, it can be chosen as the canonical basis. □

For any choice of $A \neq 0$, we may impose the Jones-Wenzl projector as a local relation. The resulting finitely dimensional vector spaces Pic(Y) might be trivial. For example, if we choose a $d \neq \pm 1$ and impose the Jones-Wenzl projector $p_2 = 0$ as the local relation. To see that the resulted picture spaces $= 0$, we reconnect two adjacent loops in a disk into one using $p_2 = 0$; this gives the identity $(d^2 - 1)\emptyset = 0$. If $d \neq \pm 1$, then $\emptyset = 0$, hence Pic(S^2) $= 0$. Even if Pic(Y)'s are not 0, they do not necessarily form a TQFT in general. We do not know any examples. If exist, such non-trivial vector spaces might have interesting applications because they are representations of the mapping class groups. In the cases of Jones-Wenzl projectors, only certain special choices of A's lead to TQFTs.

3.3. *Skein classes*

Fix a $d \in \mathbb{C}\backslash\{0\}$, a skein relation $K = 0$ and a local relation $R = 0$. Given an oriented 3-manifold X (possibly with boundaries). Let $\mathcal{F}$ be all the non-crossing loops in X, i.e., all links l's in the interior of X, and $\mathbb{C}[\mathcal{F}]$ be their linear span. We impose the "d-isotopy" relation on $\mathbb{C}[\mathcal{F}]$, where a knot is trivial if it bounds a disk in X. For any 3-ball B^3 inside X and a link l, the part $l \cap B^3$ of l can be projected onto a proper rectangle $\mathcal{R}$ of B^3 using the orientation of X (isotopy l if necessary). Resolving all crossings with the given skein relation $K = 0$, we obtain a formal diagram in $\mathcal{R}$, where the local relation $R = 0$ can be applied. Such operations introduce linear relations onto $\mathbb{C}[\mathcal{F}]$. The resulting quotient vector space will be called the skein space, denoted by $S_{d,K,R}(X)$ or just $S(X)$, and elements of $S(X)$ will be called skein classes.

As mentioned in the introduction, the empty set $\emptyset$ has been regarded as a manifold of each dimension. It is also regarded as a multicurve in any manifold Y or a link in any X, and many other things. In the case of skein spaces, the empty multicurve represents an element of the skein space $S(X)$. For a closed manifold X, this would be the canonical basis if the skein space $S(X) \cong \mathbb{C}$. But the empty skein is the 0 vector for some closed 3-manifolds. In these cases, we do not have a canonical basis for the skein space $S(X)$ even if $S(X) \cong \mathbb{C}$.

Skein spaces behave naturally with respect to disjoint union, inclusion of spaces, orientation reversal, and self-diffeomorphisms: the skein space of a disjoint union is isomorphic to the tensor product; an orientation preserving embedding from $X_1 \to X_2$ induces a linear map from $S(X_1)$ to $S(X_2)$, orientation reversal induces a conjugate-linear map on $S(X)$, and diffeomorphisms of X act on $S(X)$ by moving pictures around, therefore $S(X)$ is a representation of the orientation preserving diffeomorphisms of X up to isotopy.

Prop 3.2.

(1) If Y is oriented, then $\mathrm{Pic}(Y)$ is an algebra.
(2) If $\partial X = Y$, then $\mathrm{Pic}(Y)$ acts on $S(X)$. If Y is oriented, then $S(X)$ is a representation of $\mathrm{Pic}(Y)$.

Proof.

(1): Given two multicurves x, y in Y, and consider $Y \times [-1, 1]$, draw x in $Y \times 1$ and y in $Y \times -1$. Push x into the interior of $Y \times [0, 1]$ and y into $Y \times [-1, 0]$. Isotope x, y so that their projections onto $Y \times 0$ are in general position. Resolutions of the crossings using the given skein relation result in a formal multicurve in Y, which is denoted by xy. We define $[x][y] = [xy]$, where $[\cdot]$ denotes the picture class. Suppose the local relation is $R = 0$, and let $\hat{R}$ be a multicurve obtained from the closure of R arbitrarily outside a rectangle $\mathcal{R}$ where the local relation resides. To show that this multiplication is well-defined, it suffices to show that $\hat{R}y = 0$. By general position, we may assume that y miss the rectangle $\mathcal{R}$. Then by definition, $\hat{R}y = 0$ no matter how we resolve the crossings away from the local relation R. It is easy to check that this multiplication yields an algebra structure on $\mathrm{Pic}(Y)$.

(2): The action is defined by gluing a collar of the boundary and then re-parameterizing the manifold to absorb the collar. Let Y_ϵ be $Y \times [0, \epsilon]$, which can be identified with a small collar neighborhood of Y in X. Given a multicurve x in X and γ in Y, draw γ on $Y \times 0$ and push it into Y_ϵ. Then the union $\gamma \cup x$ is a multicurve in $X_+ = Y_\epsilon \cup_Y X$. Absorbing Y_ϵ of X_+ into X yields a multicurve $\gamma \cup x$ in X, which is defined to be $\gamma.x$. □

3.4. *Recoupling theory*

In this section, we recall some results of the recoupling theory in,[KL] and deduce some needed results for later sections.

Fix a $A \in \mathbb{C}\backslash\{0\}$, two families of numbers are important for us: the Chebyshev polynomials $\Delta_n(d)$ and the quantum integers $[n]_A = \frac{A^{2n} - A^{-2n}}{A^2 - A^{-2}}$. When A is clear from the context, we will drop the A from $[n]_A$. The Chebyshev polynomials and quantum integers are related by the formula $\Delta_n(d) = (-1)^n [n+1]_A$.

Note that $[-n]_A = -[n]_A, [n]_{-A} = [n]_{\bar{A}} = [n]_A, [n]_{iA} = (-1)^{n+1}[n]_A$. Some other relations of quantum integers depend on the order of A.

Lemma 3.1. *Fix $r \geq 3$.*

(1) If A is a primitive $4r$th root of unity, then $[n+r] = -[n]$ and $[r-n] = [n]$. The Jones-Wenzl projectors $\{p_i\}$ exist for $0 \leq i \leq r-1$, and $\mathrm{Tr}(p_{r-1}) = \Delta_r = 0$.

(2) If r odd and A is a primitive $2r$th root of unity, then $[n+r] = [n]$ and $[r-n] = -[n]$. The Jones-Wenzl projectors $\{p_i\}$ exist for $0 \leq i \leq r-1$, and $\mathrm{Tr}(p_{r-1}) = \Delta_r = 0$.

(3) If r odd and A is a primitive rth root of unity, then $[n+r] = [n]$ and $[r-n] = -[n]$. The Jones-Wenzl projectors $\{p_i\}$ exist for $0 \leq i \leq r-1$, and $\mathrm{Tr}(p_{r-1}) = \Delta_r = 0$.

The proof is obvious using the induction formula for p_n in Lemma 2.1, and $[n] \neq 0$ for $0 \leq n \leq r-1$ for such A's.

Fix an r and A as in Lemma 3.1, and let I be the range that p_i exists and $\mathrm{Tr}(p_i) \neq 0$. Let $L_A = \{p_i\}_{i \in I}$, then $I = \{0, 1, \cdots, r-2\}$. Both L_A and I will be called the label set. Note that if A is a primitive $2r$th root of unity and r is even, then $\{p_i\}$ exist for $0 \leq i \leq \frac{r-2}{2}$, and $\mathrm{Tr}(p_{\frac{r-2}{2}}) = 0$.

Given a ribbon link l in S^3, i.e. each component is a thin annulus, also called a framed link, then the Kauffman bracket of l, i.e. the Kauffman bracket and "d-isotopy" skein class of l, is a framed version of the Jones polynomial of l, denoted by $< l >_A$. The Kauffman bracket can be generalized to colored ribbon links: ribbon links that each component carries a label from L_A; the Kauffman bracket of a colored ribbon link l is the Kauffman bracket of the formal ribbon link obtained by replacing each component a of l with its label p_i inside the ribbon a and thickening each component of p_i inside a into small ribbons. Since S^3 is simply-connected, the Kauffman bracket of any colored ribbon link is a Laurent polynomial in A, hence a complex number.

Let H_{ij} be the colored ribbon Hopf link in the plane labelled by Jones-Wenzl projectors p_i and p_j, then the Kauffman bracket of H_{ij} is

$$\tilde{s}_{ij} = (-1)^{i+j}[(i+1)(j+1)]_A. \tag{3.1}$$

The matrix $\tilde{s} = (\tilde{s}_{ij})_{i,j \in I}$ is called the modular $\tilde{s}$-matrix. Let $\tilde{s}_{\text{even}}$ be the restriction of $\tilde{s}$ to even labels. Define $\bar{i} = k - i = r - 2 - i$.

Lemma 3.2.

(1) *If A is a primitive 4rth root of unity, then the modular $\bar{s}$ matrix is non-singular.*
(2) *If r is odd and A is a primitive 2rth or rth root of unity, then $\bar{s}_{\bar{i}j} = \bar{s}_{ij}$.*
(3) *If r odd, and A is a primitive 2rth root of unity or rth root of unity, then the modular $\bar{s}$ has rank $= \frac{r-1}{2}$. Moreover, $\bar{s} = \bar{s}_{\text{even}} \otimes \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.*

Proof.

Since $\bar{s}$ is a symmetric real matrix, so the rank of $\bar{s}$ is the same as $\bar{s}^2$. By the formula 3.2, we have $(\bar{s}^2)_{ij}$

$$= \frac{(-1)^{i+j}}{(A^2 - A^{-2})^2} \sum_{l=0}^{r-2} [A^{2(i+1)(l+1)} - A^{-2(i+1)(l+1)}][A^{2(l+1)(j+1)} - A^{-2(l+1)(j+1)}]$$

$$= \frac{(-1)^{i+j}}{(A^2 - A^{-2})^2} \sum_{l=0}^{r-2} [A^{2(i+1)(l+1)+2(l+1)(j+1)} + A^{-2(i+1)(l+1)-2(l+1)(j+1)}$$

$$-A^{2(i+1)(l+1)-2(l+1)(j+1)} - A^{-2(i+1)(l+1)+2(l+1)(j+1)}].$$

The first sum $\sum_{l=0}^{r-2} A^{2(i+1)(l+1)+2(l+1)(j+1)}$ is a geometric series $= \frac{A^{2(i+j+2)} - (A^{2(i+j+2)})^r}{1 - A^{2(i+j+2)}}$ if $A^{2(i+j+2)} \neq 1$. The second sum $\sum_{l=0}^{r-2} A^{-2(i+1)(l+1)-2(l+1)(j+1)}$ is the complex conjugate of the first sum. The third sum $-\sum_{l=0}^{r-2} A^{2(i+1)(l+1)-2(l+1)(j+1)}$ is also a geometric series $= -\frac{A^{2(i-j)} - (A^{2(i-j)})^r}{1 - A^{2(i-j)}}$ if $A^{2(i-j)} \neq 1$. The 4th sum $-\sum_{l=0}^{r-2} A^{-2(i+1)(l+1)+2(l+1)(j+1)}$ is the complex conjugate of the third sum. If A is a 4rth primitive, since $0 \leq i, j \leq r-2$, we have $4 \leq 2(i + j + 2) \leq 4r - 4$ and $-(r-2) \leq i - j \leq r - 2$. Hence, $A^{2(i+j+2)} \neq 1$ and $A^{2(i-j)} \neq 1$ unless $i = j$. The first sum and the second sum add to $\frac{A^{2(i+j+2)} - (-1)^{i+j} - 1 + (-1)^{i+j} A^{2(i+j+2)}}{1 - A^{2(i+j+2)}}$. So if $i + j$ is even, then $= -2$; if $i + j$ is odd, then $= 0$. If $A^{2(i-j)} \neq 1$, then the third and 4th add to $-\frac{A^{2(i-j)} - (-1)^{i-j} - 1 + (-1)^{i-j} A^{2(i-j)}}{1 - A^{2(i-j)}}$. It is $= 2$ if $i - j$ is even and $= 0$ if $i - j$ odd. Therefore, if $i \neq j$, the four sums add to 0 and if $i = j$, then they add to $-2 - 2(r-1) = -2r$. It follows that $\bar{s}^2$ is a diagonal matrix with diagonal entries $= \frac{-2r}{(A^2 - A^{-2})^2}$.

If A is a 2rth or rth primitive, if $A^{2(i+j+2)} \neq 1$, then the first sum -1. The second sum is also -1 since it is the complex conjugate. If $A^{2(i-j)} \neq 1$, then the third sum is $= 1$ and so is the 4th sum. It follows that if neither $A^{2(i+j+2)} = 1$ nor $A^{2(i-j)} = 1$, then the (i, j)th entry of $\bar{s}^2$ is 0.

If $A^{2(i+j+2)} = 1$, then $2(i+j+2) = r, 2r, 3r$ as $0 \leq i, j \leq r-2$ and $4 \leq 2(i+j+2) \leq 4r-4$. When r is odd, $2(i+j+2) = 2r$, so $i = \bar{j}$. Therefore, if $i+j+2 \neq r$, the first and second sum is -1. If $i+j+2 = r$, then the first and second sum both are $r-1$. If $A^{2(i-j)} = 1$, then $2(i-j) = -r, 0, r$ as $-2(r-2) \leq 2(i-j) \leq 2(r-2)$. It follows that $i = j$ as r is odd. If $i = j$, the third and 4th sum both are $= -(r-1)$. Put everything together, we have if $i \neq j$ or $i+j \neq r-2$, then $(\bar{s}^2)_{ij} = 0$. If $i = j$, then $i+j \neq r-2$ because $r-2$ is odd, and $(\bar{s}^2)_{ij} = \frac{-2r}{(A^2-A^{-2})^2}$. If $i+j = r-2$, then $i \neq j$, and $(\bar{s}^2)_{ij} = \frac{-2r}{(A^2-A^{-2})^2}$. Hence $\bar{s}^2 = \frac{-2r}{(A^2-A^{-2})^2}(m_{ij})$, where $m_{ij} = 0$ unless $i = j$ or $i+j = k = r-2$. □

We define a colored tangle category Δ_A based on a label set L_A. Consider $\mathbb{C} \times I$, the product of the plane $\mathbb{C}$ with an interval I, the objects of Δ_A are finitely many labelled points on the real axis of $\mathbb{C}$ identified with $\mathbb{C} \times \{0\}$ or $\mathbb{C} \times \{1\}$. A morphism between two objects are formal tangles in $\mathbb{C} \times I$ whose arc components connect the objects in $\mathbb{C} \times \{0\}$ and $\mathbb{C} \times \{1\}$ transversely with same labels, modulo Kauffman bracket and Jones-Wenzl projector p_{r-1}. Horizontal juxtaposition as a tensor product makes Δ_A into a strict monoidal category.

The quantum dimension d_i of a label i is defined to the Kauffman bracket of the 0-framed unknot colored by the label i. So $d_i = \Delta_i(d)$. The total quantum order of Δ_A is $D = \sqrt{\sum_i d_i^2}$, so $D = \sqrt{\frac{-2r}{(A^2-A^{-2})^2}}$. The Kauffman bracket of the 1-framed unknot is of the form $\theta_i d_i$, where $\theta_i = A^{-i(i+2)}$ is called the twist of the label i. Define $p_\pm = \sum_{i \in I} \theta_i^{\pm 1} d_i^2$, then $D^2 = p_+ p_-$.

A triple (i, j, k) of labels is admissible if $\mathrm{Hom}(p_i \otimes p_j, p_k)$ is not 0. The theta symbol $\theta(i, j, k)$ is the Kauffman bracket of the theta network, see.[KL]

Lemma 3.3.

(1) $\mathrm{Hom}(p_i \otimes p_j, p_k)$ *is not* 0 *if and only if the theta symbol* $\theta(i, j, k)$ *is non-zero, then* $\mathrm{Hom}(p_i \otimes p_j, p_k) \cong \mathbb{C}$.

(2) $\theta(i, j, k) \neq 0$ *if and only if* $i+j+k \leq 2(r-2)$, $i+j+k$ *is even and* $i+j \geq k, j+k \geq i, k+i \geq j$.

3.5. *Handles and S-matrix*

There are various ways to present an n-manifold X: triangulation, surgery, handle decomposition, etc. The convenient ways for us are the surgery description and handle decompositions.

Handle decomposition of a manifold X comes from from a Morse function of X. Fix a dimension$= n$, a k-handle is a product structure $I^k \times I^{n-k}$ on the n-ball B^n, where the part of boundary $\partial I^k \times I^{n-k} \cong S^{k-1} \times I^{n-k}$ is specified as the attaching region. The basic operations in handlebody theory are handle attachment, handle slide, stabilization, and surgery. They correspond to how Morse functions pass through singularities in the space of smooth functions on X. Let us discuss handle attachment and surgery here. Given an n-manifold X with a sub-manifold $S^{k-1} \times I^{n-k}$ specified in its boundary, and an attach map $\phi : \partial I^k \times I^{n-k} \to S^{k-1} \times I^{n-k}$, we can attach a k-handle to X via ϕ to form a new manifold $X' = X \cup_\phi I^k \times I^{n-k}$. The new manifold X' depends on ϕ, but only on its isotopy class. It follows from Morse theory or triangulation that every smooth manifold X can be obtained from 0-handles by attaching handles successively, i.e., has a handle decomposition. Moreover, the handles can be arranged to be attached in the order of their indices, i.e., from 0-handles, first all 1-handles attached, then all 2-handles, etc.

Given an n-manifold X, a sub-manifold $S^k \times I^{n-k}$ and a map $\phi : \partial I^{k+1} \times S^{n-k-1} \to S^k \times S^{n-k-1}$, we can change X to a new manifold X' by doing index k surgery on $S^k \times I^{n-k}$ as follows: delete the interior of $S^k \times I^{n-k}$, and glue in $I^{k+1} \times S^{n-k-1}$ via ϕ along the common boundary $S^k \times S^{n-k-1}$. Of course the resulting manifold X' depends on the map ϕ, but only on its isotopy class. Handle decompositions of $n+1$-manifolds are related to surgery of n-manifolds as their boundaries.

It is fundamental theorem that every orientable closed 3-manifold can be obtained from surgery on a framed link in S^3; moreover, if two framed links give rise to the same 3-manifold, they are related by Kirby moves, which consist of stabilization and handle slides. This is extremely convenient for constructing 3-manifold invariants from link invariants: it suffices to write down a magic linear combination of invariants of the surgery link so that the combination is invariant under Kirby moves. The Reshetikhin-Turaev invariants were discovered in this way.

The magic combination is provided by the projector ω_0 from the first row of the S matrix: given a surgery link L of a 3-manifold, if every component of L is colored by ω_0, then the resulting link invariant is invariant under handle slides. Moreover, a certain normalization using the signature of the surgery link produces a 3-manifold invariant as in Theorem 3.1 below.

The projector ω_0 is a ribbon tensor category analogue of the regular representation of a finite group, and is related to surgery as below. In general, all projectors ω_i are related to surgery in a sense, which is responsible for the gluing formula for the partition function Z of a TQFT.

Lemma 3.4.

(1) Given a 3 manifold X with a knot K inside, if K is colored by ω_0, then the invariant of the pair (X, K) is the same as the invariant of X', which is obtained from X by 0-surgery on K.

(2) Let $S^2 \subset X$ be an embeded 2-sphere, then any labeled multicurve γ interests S^2 transversely must carry the trivial label. In other words, non-trivial particle type cannot cross an embeded S^2.

The colored tangle category Δ_A has natural braidings, and duality, hence is a ribbon tensor category. An object a is simple if $\mathrm{Hom}(a,a) = \mathbb{C}$. A point marked by a Jones-Wenzl projector p_i is a simple object of Δ_A. Therefore, the label set L_A can be identified with a complete set of simple object representatives of Δ_A. A ribbon category is premodular if the number of simple object classes is finite, and is called modular if furthermore, the modular S-matrix $S = \frac{1}{D}\tilde{s}$ is non-singular. A non-singular S-matrix $S = (s_{ij})$ can be used to define projectors $\omega_i = \frac{1}{D}\sum_{j\in I} s_{ij}p_j$, which projects out the ith label.

Given a ribbon link l, $< \omega_0 * l >$ denotes the Kauffman bracket of the colored ribbon link l that each component is colored by ω_0.

Theorem 3.1.

(1) The tangle category Δ_A is a premodular category, and is modular if and only if A is a primitive $4r$th root of unity.

(2) Given a premodular category Λ, and X an oriented closed 3-manifold with an m-component surgery link l, then $Z_{JK}(X) = \frac{1}{D^{m+1}}(\frac{p_-}{D})^{\sigma(l)} < \omega_0 * l >$ is a 3-manifold invariant, where $\sigma(l)$ is the signature of the framing matrix of l.

3.6. *Diagram TQFTs for closed manifolds*

In this section, fix an integer $r \geq 3$, A as in Lemma 3.1. For these special values, the picture spaces $\mathrm{Pic}^A(Y)$ form a modular functor which is part of a TQFT. These TQFTs will be called diagram TQFTs. In the following, we verify all the applicable axioms for diagram TQFTs for closed manifolds after defining the partition function Z.

The full axioms of TQFTs are given in Section 6.3. The applicable axioms for closed manifolds are:

(1) Empty surface axiom: $V(\emptyset) = \mathbb{C}$

(2) Sphere axiom: $V(S^2) = \mathbb{C}$. This is a consequence of the disk axiom and gluing formula.
(3) Disjoint union axiom for both V and Z:
(4) Duality axiom for V:
(5) Composition axiom for Z: This is a consequence of the gluing axiom.

These axioms together forms exactly a tensor functor as follows: the category $\mathfrak{X}^{2,cld}$ of oriented closed surfaces Y as objects and oriented bordisms up to diffeomorphisms between surfaces as morphisms is a strict rigid tensor category if we define disjoint union as the tensor product; $-Y$ as the dual object of Y; for birth/death, given an oriented closed surface Y, let $Y \times S^1_- : \emptyset \longrightarrow -Y \coprod Y$ be the birth operator, and $Y \times S^1_+ : -Y \coprod Y \longrightarrow \emptyset$ the death operator, here $S^1_{\mp}$ are the lower/up semi-circles.

Definition 3.2. A $(2+1)$-anomaly free TQFT for closed manifolds is a *nontrivial* tensor functor $V : \mathfrak{X}^{2,cld} \longrightarrow \mathcal{V}$, where $\mathcal{V}$ is the tensor category of finite dimensional vector spaces.

Non-triviality implies $V(\emptyset) = \mathbb{C}$ by the disjoint union axiom. Since $\emptyset \coprod \emptyset = \emptyset$, $V(\emptyset) = V(\emptyset) \otimes V(\emptyset)$. Hence $V(\emptyset) \cong \mathbb{C}$ because otherwise $V(\emptyset) = 0$ the theory is trivial. The empty set picture $\emptyset$ is the canonical basis, therefore, $V(\emptyset) = \mathbb{C}$.

The disjoint union axiom and the trace formula for Z in Prop. 6.1 fixes the normalization of 3-manifold invariants. Given an invariant of closed 3-manifolds, then multiplication of all invariants by scalars leads to another invariant. Hence Z on closed 3-manifolds can be changed by multiplying any scalar k. But this freedom is eliminated from TQFTs by the disjoint union axiom which implies $k = k^2$, hence $k = 1$ since otherwise the theory is trivial. The trace formula implies $Z(S^2 \times S^1) = 1$. We set $Z(S^3) = \frac{1}{D}$, and D is the total quantum order of the theory.

Recall that the picture space $\mathrm{Pic}^A(Y)$ is defined even for unorientable surfaces. When Y is oriented, $\mathrm{Pic}(Y)$ is isomorphic to $K_A(Y \times I)$. Given a bordism X from Y_1 to Y_2, we need to define $Z_D(X) \in \mathrm{Pic}^A(-Y_1 \coprod Y_2)$. It follows from the disjoint union axiom and the duality axiom, $Z_D(X)$ can be regarded as a linear map $\mathrm{Pic}^A(Y_1) \longrightarrow \mathrm{Pic}^A(Y_2)$.

Given a closed surface Y, let $V_D(Y) = \mathrm{Pic}^A(Y)$. For a diffeomorphism $f : Y \to Y$, the action of f on pictures is given by moving them in Y. This action on pictures descends to an action of mapping classes on $V(Y)$. To define $Z_D(X)$ for a bordism X from Y_1 to Y_2, fix a relative handle decomposition of X from Y_1 to Y_2.

Suppose that Y_2 is obtained from Y_1 by attaching a single handle of indices $= 0, 1, 2, 3$. For indices $= 0, 3$, the linear map is just a multiplication by $\frac{1}{D_{JK}}$, where $D_{JK} = \sqrt{\frac{-2r}{(A^2-A^{-2})^2}}$. Since S^3 is a 0-handled followed by a 3-handle, $Z_D(S^3) = \frac{1}{D_{JK}^2}$. This is not a coincidence, but a special case of a theorem of K. Walker and V. Turaev that $Z_D(X) = |Z_{JK}(X)|^2$ for any oriented closed 3-manifold.

Given a multicurve γ in Y_1,

1): If a 1-handle $I \times B^2$ is attached to Y_1, isotopy γ so that it is disjoint from the attaching regions $\partial I \times B^2$ of the 1-handle. Label the co-core circle $\frac{1}{2} \times \partial B^2$ of the 1-handle by ω_0 to get a formal multicurve in Y_2. This defines a map from $\mathrm{Pic}^A(Y_1)$ to $\mathrm{Pic}^A(Y_1)$ by linearly extending to pictures classes.

2): If a 2-handle $B^2 \times I$ is attached to Y_1, isotopy γ so that it intersects the attaching circle $\partial B^2 \times \frac{1}{2}$ of the 2-handle transversely. Expand this attaching circle slightly to become a circle s just outside the 2 handle and parallel to the attaching circle $\partial B^2 \times \frac{1}{2}$. Label s by ω_0. Fuse all strands of γ so that a single labeled curve intersects the attaching circle $\partial B^2 \times \frac{1}{2}$; only the 0-labeled curves survive the projector ω_0 on s. By drawing all remaining curves on the Y_1 outside the attaching region plus the two disks $B^2 \times \{0\}$ and $B^2 \times \{1\}$, we get a formal diagram in Y_2.

We need to prove that this definition is independent of handle-slides and cancellation pairs, which is left to the interested readers.

Now we are ready to verify all the axioms one by one:

The empty surface axiom: this is true as we have a non-trivial theory.

The sphere axiom: by the "d-isotopy" constraint, every multicurve with m loops $= d^m\emptyset$. If $\emptyset$ picture on S^2 is $= 0$ in $\mathrm{Pic}^A(S^2)$, then $Z_D(B^3) = 0$ which leads to $Z_D(S^3) = 0$. But $Z_D(S^3) \neq 0$, it follows that $\mathrm{Pic}^A(S^2) = \mathbb{C}$.

The disjoint union axioms for both V and Z are obvious since both are defined by pictures in each connected component.

$\mathrm{Pic}(-Y) = \mathrm{Pic}(Y)$ is the identification for the duality axiom. To define a functorial identification of $\mathrm{Pic}^A(-Y)$ with $\mathrm{Pic}^A(Y)^*$, we define a Hermitian paring: $\mathrm{Pic}^A(Y) \times \mathrm{Pic}^A(Y) \longrightarrow \mathbb{C}$. Since $\mathrm{Pic}^A(Y)$ is an algebra, and semi-simple, it is a matrix algebra. For any $x, y \in \mathrm{Pic}^A(Y)$, we identify them as matrices, and define $< x, y >= \mathrm{Tr}(x^\dagger y)$. This is a non-generate inner product. The conjugate linear map $x \to < x, \cdot >$ is the identification of $\mathrm{Pic}^A(-Y)$ with $\mathrm{Pic}^A(Y)^*$.

Summarizing, we have;

Theorem 3.2.
The pair (V_D, Z_D) is a $(2+1)$-anomaly free TQFT for closed manifolds.

3.7. *Boundary conditions for picture TQFTs*

In Section 3.1, we consider $\mathbb{C}[\mathbb{S}]$ for surfaces Y even with boundaries. Given a surface Y with m boundary circles with n_i fixed points on the ith boundary circle, by imposing Jones-Wenzl projector p_{r-1} away from the boundaries, we obtain some pictures spaces, denoted as $\mathrm{Pic}^A(Y; n_1, \cdots, n_m)$. To understand the deeper properties of the picture space $\mathrm{Pic}^A(Y)$, we need to consider the splitting and gluing of surfaces along circles. Given a simple closed curve (scc) s in the interior of Y, and a multicurve γ in Y, isotope s and γ to general position. If Y is cut along s, the resulted surface Y_{cut} has two more boundary circles with n points on each new boundary circle, where n is the number of intersection points of $s \cap \gamma$, and $n \in \{0, 1, 2, \cdots, \}$. In the gluing formula, we like to have an identification of $\oplus \mathrm{Pic}^A(Y_{\text{cut}})$ with all possible boundary conditions with $\mathrm{Pic}^A(Y)$, but this sum consists of infinitely many non-trivial vector space,which contradicts that $\mathrm{Pic}^A(Y)$ is finitely dimensional. Therefore, we need more refined boundary conditions. One problem about the crude boundary conditions of finitely many points is due to bigons resulted from the "d-isotopy" freedom: we may introduce a trivial scc intersecting s at many points, or isotope γ to have more intersection points with s. The most satisfactory solution is to define a picture category, then the picture spaces become modules over these categories. Picture category serves as crude boundary conditions. To refine the crude boundary conditions, we consider the representation category of the picture category as new boundary conditions. The representation category of a picture category is naturally Morita equivalent to the original picture category. The gluing formula can be then formulated as the Morita reduction of picture modules over the representation category of the picture category. The labels for the gluing formula are given by the irreps of the picture categories. This approach will be treated in the next two sections. In this section, we content ourselves with the description of the labels for the diagram TQFTs, and define the diagram modular functor for all surfaces. In Section 6.3, we will give the definition of a TQFT, and later verify all axioms for diagram TQFTs.

The irreps of the non-semi-simple TL annular categories at roots of unity were contained in,[GL] but we need the irreps of the semi-simple quotients of TL annular categories, i.e., the TLJ annular categories.

The irreps of the TLJ annular will be analyzed in Sections 5.5 5.6. In the following, we just state the result. By Theorem B.1 in Appendix B, each irrep can be represented by an idempotent in a morphism space of some object. Fix h $(0 \leq h \leq k)$ many points on S^1, and let $\omega_{i,j;h}$ be the following diagram in the annulus $\mathcal{A}$: the two circles in the annulus are labeled by ω_i, ω_j, and $h = 3$ in the diagram.

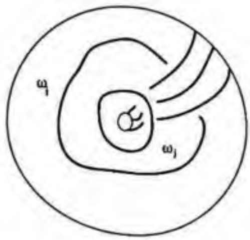


Fig. 3.7. Annular projector

The labels for diagram TQFTs are the idempotents $\omega_{i,j;h}$ above. Given a surface Y with boundary circles $\gamma_i, i = 1, .., m$. In the annular neighborhood $\mathcal{A}_i$ of γ_i, fix an idempotent $\omega_{i,j;h}$ inside $\mathcal{A}_i$. Let $\mathrm{Pic}_D^A(Y; \omega_{i,j;h})$ be the span of all multicurves that within $\mathcal{A}_i$ agree with $\omega_{i,j;h}$ modulo p_{r-1} .

Theorem 3.3.
If A is as in Lemma 3.1, then the pair $(Pic^A(Y), Z_D)$ is an anomaly-free TQFT.

3.8. *Jones-Kauffman skein spaces*

In this section, fix an integer $r \geq 3$, A as in Lemma 3.1, and $d = -A^2 - A^{-2}$.

Definition 3.3. Given any closed surface Y, let $\mathrm{Pic}^A(Y)$ be the picture space of pictures modulo p_{r-1}. Given an oriented 3-manifold X, the skein space of p_{r-1} and the Kauffman bracket is called the Jones-Kauffman skein space, denoted by $K_A(X)$.

The following theorem collects the most important properties of the Jones-Kauffman skein spaces. The proof of the theorem relies heavily on handlebody theory of manifolds.

Theorem 3.4.
a): Let A be a primitive $4r$th root of unity. Then

(1) $K_A(S^3) = \mathbb{C}$.
(2) There is a canonical isomorphism of $K_A(X_1 \amalg X_2) \cong K_A(X_1) \# K(X_2)$.
(3) If $\partial X_1 = \partial X_2$, *then* $K_A(X_1) \cong K_A(X_2)$, *but not canonically.*

(4) If the $\emptyset$ link is not 0 in $K_A(X)$ for a closed manifold X, then $K_A(X) \cong \mathbb{C}$ canonically. The $\emptyset$ link in $K_A(\#_{r=1}^m S^1 \times S^2)$ and $K_A(Y \times S^1)$ is not 0 for oriented closed surface Y.

(5) $K_A(-X) \times K_A(X) \longrightarrow K_A(DX)$ is non-degenerate. Therefore, $K_A(\overline{X})$ is isomorphic to $K_A(X)^*$, but not canonically.

(6) $K_A(Y \times I) \longrightarrow End(K_A(X))$ is an isomorphism if $\partial X = Y$.

(7) $Pic^A(Y)$ is canonically isomorphic to $K_A(Y \times I)$ if Y is orientable, hence also isomorphic to $End(K_A(X))$.

b): If A is a primitive $2r$th root of unity or rth root of unity, then (2) does not hold, and it follows that the rest fail for disconnected manifolds.

Proof.

(1) Obvious.

(2) The idea here in physical terms is that non-trivial particles cannot cross an S^2.

The skein space $K_A(X_1 \amalg X_2)$ is a subspace of $K_A(X_1)\#K(X_2)$ by inclusion. So it suffices to show this is onto. Given any skein class x in $K_A(X_1)\#K(X_2)$, by isotopy we may assume x intersects the connecting S^2 transversely. Put the projector ω_0 on S^2 disjoint from x, then ω_0 encircle x from outside. Apply ω_0 to x to project out the 0-label, we split x into two skein classes in $K_A(X_1 \amalg X_2)$, therefore the inclusion is onto.

(3) This is an important fact. For example, combining with (1), we see that the Jones-Kauffman skein space of any oriented 3-manifold is $\cong \mathbb{C}$.

We will show below that any bordism W^4 from X_1 to X_2 induces an isomorphism. Moreover, the isomorphism depends only on the signature of the 4-manifold W^4.

Pick a 4-manifold W such that $\partial W = -X_1 \cup_Y (Y \times I) \cup_Y X_2$ (W exists since every orientable 3 manifold bounds a 4 manifold), and fix a handle-decomposition of W. 0-handles, and dually 4-handles, induce a scalar multiplication. 1-handles, or dually 3-handles, also induce a scalar multiplication by (2). By (2), we may assume that $X_i, i = 1, 2$ are connected. Therefore, we will fix a relative handle decompositions of W with only 2-handles, and let $L_{X_i}, i = 1, 2$ be the attaching links for the 2-handles in X_i, respectively. Then $X_1 \backslash L_{X_1} \cong X_2 \backslash L_{X_2}$. The links $L_{X_i}, i = 1, 2$ are dual to each other in a sense: let $L_{X_i}^{\text{dual}}$ be the link consists of cocores of the 2-handles on X_i, then surgery on L_{X_1} in X_1 results X_2, while surgery on $L_{X_1}{}^{\text{dual}}$ in X_1 results X_1, and vice versa.

Define a map $h_1 : K_A(X_1) \to K_A(X_2)$ as follows: for any skein class representative γ_1, isotope γ_1 so that it misses L_{X_1}. Note that in the skein spaces, labeling a component L_1 of a link L by ω_0, denoted as $\omega_0 * L_1$, is equivalent to surgering the component; then $h(\gamma_1) = \gamma_1 \coprod \omega_0 * L_{X_2}$, where γ_1 is now considered as a link in X_2. Formally, we write this map as:

$$(X_1;\gamma_1) \to (X_1; L_{X_1} \coprod L_{X_1}{}^{\text{dual}} \coprod \gamma_1) \to (X_2; L_{X_2} \coprod \gamma_2),$$

where γ_2 is γ_1 regarded as a skein class in X_2. In this map, L_{X_1} is mapped to the empty skein as it has been surged out, while $L_{X_1}{}^{\text{dual}}$ is mapped to L_{X_2}. Then define h_2 similarly. The composition of h_1 and h_2 is the link invariant of the colored link L_{X_i} union a small linking circle for each component plus a parallel copy of $L_{X_i}{}^{\text{dual}}$ union its small linking circles as in the Fig. 3.8, which is clearly a scalar, hence an isomorphism.

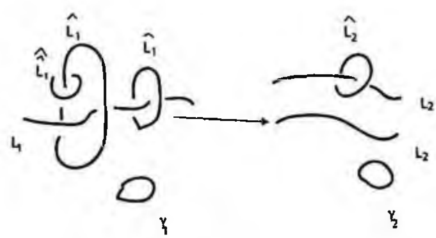


Fig. 3.8. Skein space maps

Now we see that a pair, (W, a handle decomposition), induces an isomorphism. Using Cerf theory, we can show that the isomorphism is first independent of the handle decomposition; secondly it is a bordism invariant: if there is a 5-manifold N which is a relative bordism from W to W', then W and W' induces the same map. Hence the isomorphism depends only on the signature of the 4-manifold W. The detail is a highly non-trivial exercise in Cerf theory.

(4) follows from (1)–(3) easily.

(5) The inner product is given by doubling. By (3) $K_A(X)$ is isomorphic to $K_A(H)$, where H is a handlebody with the same boundary. By (4), the same inner product is non-singular for $K_A(H)$. Chasing through the isomorphism in (3) shows that the inner product on $K_A(X)$ is also non-singular. Since $K_A(DX) \cong \mathbb{C}$, hence $K_A(\overline{X})$ is isomorphic to $K_A(X)^*$.

(6) $K_A(Y \times I)$ is isomorphic to $K_A(-X \coprod X)$ by (3). It follows that the action of $K_A(Y \times I)$ on $K_A(X)$: $K_A(X) \otimes K_A(Y \times I) \to K_A(X \cup_Y (Y \times I)$ becomes an action of $K_A(-X \coprod X)$ on $K_A(X)$: $K_A(X) \otimes K_A(-X \coprod X) \to K_A(DX \coprod X)$. By the paring in (5), we identify the action as the action of $\mathrm{End}(X) = K_A(-X) \otimes K_A(X)$ on $K_A(X)$.

(7) follows from (6) easily. □

The pairing $K_A(-X) \times K_A(X) \longrightarrow K_A(DX)$ allows us to define a Hermitian product on $K_A(X)$ as follows:

Definition 3.4.

Given an oriented closed 3-manifold X, and choose a basis e of $K_A(DX)$. Then $K_A(DX) = \mathbb{C}e$. For any multicurves x, y in X, consider x as a multicurve in $-X$, denoted as $\bar{x}$. Then define $\bar{x} \cup y =< x, y > e$, i.e. the ratio of the skien $\bar{x} \cup y$ with e. If $\emptyset$ is not 0 in $K_A(DX)$, then Hermitian pairing is canonical by choosing $e = \emptyset$.

Almost all notations are set up to define the Jones-Kauffman TQFTs. We see in Theorem 3.4 that if two 3-manifolds $X_i, i = 1, 2$ have the same boundary, then $K_A(X_1)$ and $K_A(X_2)$ are isomorphic, but not canonically. We like to define the modular functor space $V(Y)$ to be a Jones-Kauffman skein space. The dependence on X_i is due to a framing-anomaly, which also appears in Witten-Restikhin-Turaev $SU(2)$ TQFTs. To resolve this anomaly, we introduce an extension of surfaces. Recall by Poincare duality, the kernel λ_X of $H_1(\partial X; \mathbb{R}) \to H_1(X; \mathbb{R})$ is a Lagrangian subspace $\lambda \subset H_1(Y; \mathbb{R})$. This Lagrangian subspace contains sufficient information to resolve the framing dependence. Therefore, we define an extended surface as a pair $(Y; \lambda)$, where λ is a Lagrangian subspace of $H_1(Y; \mathbb{R})$. The orientation, homology and many other topological property of an extended surface $(Y; \lambda)$ mean that of the underlying surface Y.

The labels for the Jones-Kauffman TQFTs are the Jones-Wenzl projectors $\{p_i\}$. Given an extended surface $(Y; \lambda)$ with boundary circles $\gamma_i, i = 1, .., m$. Glue m disks B^2 to the boundaries to get a closed surface $\widehat{Y}$ and choose a handlebody H such that $\partial H = \widehat{Y}$, and the kernel λ_H of $H_1(Y; \mathbb{R}) \to H_1(H; \mathbb{R})$ is λ. In a small solid cylinder neighborhood $B_i^2 \times [0, \epsilon]$ of each boundary circle γ_i, fix a Jones-Wenzl projector p_{i_j} inside some $\mathrm{arc} \times [0, \epsilon]$, where the arc is any fixed diagonal of B_i^2. Let $V_{JK}^A(Y; \lambda, \{p_{i_j}\})$ be the Jones-Kauffman skein space of H of all pictures within the solid cylinders $B_i^2 \times [0, \epsilon]$ agree with $\{p_{i_j}\}$.

Lemma 3.5.

Let B^2 be a 2-disk, $\mathcal{A}$ an annulus and P a pair of pants, and (Y, λ) an extended surface with m punctures labelled by $p_{i_j}, j = 1, 2, \cdots, m$, then

(1) $V^A_{JK}(B^2; p_i) = 0$ unless $i = 0$, and $V^A_{JK}(B^2; p_0) = \mathbb{C}$.
(2) $V^A_{JK}(\mathcal{A}; p_i, p_j) = 0$ unless $i = j$, and $V^A_{JK}(\mathcal{A}; p_i, p_i) = \mathbb{C}$
(3) $V^A_{JK}(P; p_i, p_j, p_k) = 0$ unless i, j, k is admissible, and $V^A_{JK}(P; p_i, p_j, p_k) \cong \mathbb{C}$ if i, j, k is admissible.
(4) Given an extended surface $(Y; \lambda)$, and let H be a genus=g handlebody such that $\partial H = (\widehat{Y}; \lambda)$ as extended manifolds. Then admissible labelings of any framed trivalent spine dual to a pants decomposition of Y with all external edges labelled by the corresponding boundary label p_{i_j} is a basis of $V^A_{JK}(Y; \lambda, \{p_{i_j}\})$.
(5) $V^A_{JK}(Y)$ is generated by bordisms $\{X | \partial X = Y\}$ if Y is closed and oriented.

Given an extended surface $(Y; \lambda)$, to define the partition function $Z_{JK}(X)$ for any X such that $\partial X = Y$, let us first assume that $\partial X = (Y; \lambda)$ as an extended surface. Find a handlebody H such that $\lambda_H = \lambda$, and a link L in H such that surgery on L yields X. Then we define Z_X as the skein in $V_D(H)$ given by the L labeled by ω_0 on each component of L. If λ_X is not λ, then choose a 4-manifold W such that $\partial W = -X \coprod X$ and the Lagrange space λ and λ_X extended through W. W defines an isomorphism between $K_A(X)$ and itself. The image of the empty skein in $K_A(X)$ is $Z(X)$. Given $f : (Y_1; \lambda_1) \to (Y_2; \lambda_2)$, the mapping cylinder I_f defines an element in $V(Y_1 \coprod Y_2) \cong V(Y_1) \otimes V(Y_2)$ by the disjoint union axiom. This defines a representation of the mapping class group $\mathcal{M}(Y)$, which might be a projective representation.

Theorem 3.5.

If A is a primitive $4r$th root of unity for $r \geq 3$, then the pair (V_{JK}, Z_{JK}) is a TQFT.

This theorem will be proved in Section 7.

There is a second way to define the projective representation of $\mathcal{M}(Y)$. Given an oriented surface Y, the mapping class group $\mathcal{M}(Y)$ acts on $\text{Pic}(Y)$ by moving pictures in Y. This action preserves the algebra structure of $\text{Pic}(Y)$ in Prop. 3.2. The algebra $\text{Pic}(Y) \cong \text{End}(K_A(Y))$ is a simple matrix algebra, therefore any automorphism ρ is given by a conjugation of an invertible matrix M_ρ, where M_ρ is only defined up to a non-zero scalar. It follows that for each $f \in \mathcal{M}(Y)$, we have an invertible matrix $V_f = M_f$, which forms a projective representation of the mapping class group $\mathcal{M}(Y)$.

4. Morita equivalence and cut-paste topology

Temperley-Lieb-Jones algebras can be generalized naturally to categories by allowing different numbers of boundary points at the top and bottom of the rectangle $\mathcal{R}$. Another interesting generalization is to replace the rectangle by an annulus $\mathcal{A}$. Those categories provide crude boundary conditions for $V(Y)$ when Y has boundary, and serve as "scalars" for a "higher" tensor product structure which provides the formal framework to discuss relations among $V(Y)$'s under cut-paste of surfaces. The vector spaces $V(Y)$ of a modular functor V can be formulated as bimodules over those picture categories. An important axiom of a modular functor is the gluing formula which encodes locality of a TQFT, and describes how a modular functor $V(Y)$ behaves under splitting and gluing of surfaces along boundaries. The gluing formula is best understood as a Morita reduction of the crude picture categories to their representation categories, which provides refined boundary conditions for surfaces with boundaries. Therefore, the Morita reduction of a picture category amounts to the computation of all its irreps. The use of bimodules and their tensor products over linear categories to realize gluing formulas appeared in [BHMV]. In this section, we will set up the formalism. The irreducible representations of our examples will be computed in the next section.

We work with the complex numbers $\mathbb{C}$ as the ground ring. Let Λ denote a linear category over $\mathbb{C}$ meaning that the morphisms set of Λ are vector spaces over $\mathbb{C}$ and composition is bilinear. We consider two kinds of examples: "rectangular" and "annular" Λ's. (The adjectives refer to methods for building examples rather than additional axioms.) We think of rectangles($\mathcal{R}$) as oriented vertically with a "top" and "bottom" and annuli($\mathcal{A}$) has an "inside" and an "outside". Sometimes, we draw an annulus as a rectangle, and interpret the rectangles as having their left and right sides glued. The objects in our examples are finite collections of points, or perhaps points marked by signs, arrows, colors, etc., on "top" or "bottom" in the rectangular case, and on "inside" or "outside" in the annular case. The morphisms are formal linear combinations of "pictures" in $\mathcal{R}$ or $\mathcal{A}$ satisfying some linear relations. The most important examples are the Jones-Wenzl projectors. Pictures will variously be unoriented submanifolds (i.e. multicurves), 1-submanifolds with various decorations such as orienting arrows, reversal points, transverse flags, etc., and trivalent graphs. Even though the pictures are drawn in two dimensions they may in some theories be allowed to indicate over-crossings in a formal way. A morphism is sometimes called an "element" as if Λ had a single object and were an algebra.

Our $\mathcal{R}$'s and $\mathcal{A}$'s are parameterized, i.e. not treated merely up to diffeomorphism. One crucial part of the parameterization is that a base point arc $* \times I \subset S^1 \times I = \mathcal{A}$ be marked. The $*$ marks the base point on S^1 and brings us to a technical point. Are the objects of Λ the continuously many collections of finitely many points in I (or S^1) or are they to be simply one representative example for each non-negative integer m. The second approach makes the category feel bit more like an algebra (which has only one object) and the linear representations have a simpler object grading. One problem with this approach is that if an annulus $\mathcal{A}$ factored as a composition of two by drawing a degree=1 scc $\gamma \subset \mathcal{A}$ (and parameterizing both halves), even if γ is transverse to an element x in $\mathcal{A}$ $\gamma \cap x$ may not be the representative set of its cardinality. This problem can be overcome by picking a base point preserving re-parameterization of γ. This amounts to "skeletonizing" the larger category and replacing some "strict" associations by "weak" ones. Apparently a theorem of S. MacLane guarantees that no harm follows, so either viewpoint can be adopted.[Ma] We will work with the continuously many objects version.

Recall the following definition from Appendix B:

Definition 4.1. A representation of a linear category Λ is a functor $\rho : \Lambda \to \mathcal{V}$, where $\mathcal{V}$ is the category of finite dimensional vector spaces. The action is written on the right: $\rho(a) = V_a$ and given $m \in \mathrm{Mor}(a, b), \rho(m) : V_a \to V_b$. We write on the left to denote a representation of Λ^{op}.

Let us track the definitions with the simplest pair of examples, temporarily denote $\Lambda^{\mathcal{R}}$ and $\Lambda^{\mathcal{A}}$ the $\mathcal{R}$ and $\mathcal{A}$-categories with objects finite collections of points and morphisms transversely embedded un-oriented 1-submanifolds with the marked points as boundary data. Let us say that 1) we may vary multicurves by "d-isotopy" for $d = 1$, and 2) to place ourselves in the simplest case let us enforce the skein relation: $p_2 = 0$ for $d = 1$. This means that we allow arbitrary recoupling of curves. This is the Kauffman bracket relation associated to $A = e^{\frac{2\pi i}{6}}, d = -A^2 - A^{-2} = 1$. The admissible pictures may be extended to over-crosssings by the local Kauffman bracket rule in Figure 2.1 in Section 2.2.

In these theories, which we call the rectangular and annular Temperley-Lieb-Jones categories TLJ at level $= 1, d = 1$, over-crossings are quite trivial, but at higher roots of unity they are more interesting. In schematic

Figures 4.1, 4.2 give examples of $\Lambda^{\mathcal{R}}$ and $\Lambda^{\mathcal{A}}$ representations.

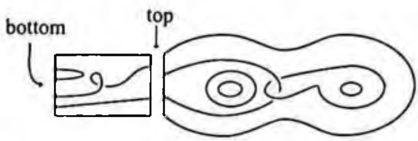


Fig. 4.1. $\Lambda^{\mathcal{R}}$ acts (represents) on vector space of pictures in twice punctured disk (or genus=2 handlebody)

It is actually the morphism between objects index by 4 to 2 points, resp. ${}_4\Lambda_2^R$, which is acting in Figure 4.1, Figure 4.2.

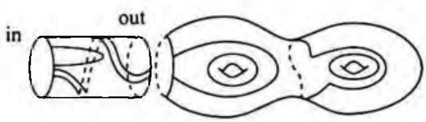


Fig. 4.2. $\Lambda^{\mathcal{A}}$ acts on vector space of pictures in punctured genus=2 surface

The actions above are by regluing and then re-parameterizing to absorb the collar. For each object a in $\mathrm{TLJ}_{d=1}$, the functor assigns the vector space of pictures lying in a given fixed space with boundary data equal the object a. Given a fixed picture in $\mathcal{R}$ and $\mathcal{A}$, i.e. an element e of Λ, gluing and absorbing the collar defines a restriction map: $f(e) : V_4 \to V_2$ (in the case illustrated) between the vector spaces with the bottom (in) and the top (out) boundary conditions. To summarize the annular categories act on vector spaces which are pictures on a surface by gluing on an annulus. The rectangular categories, in practice, act on handlebodies or other 3-manifolds with boundary by gluing on a solid cylinder; Figure 4.1 is intentionally ambiguous and may be seen as a diagram or 3−manifolds. Because we can use framing and overcrossing notations in the rectangle we are free to think of $\mathcal{R}$ either as 2-dimensional, $I \times I$, or 3−dimensional $I \times B^2$.

4.1. *Bimodules over picture category*

Because a rectangle or annulus can be glued along two sides, we need to consider $\Lambda^{\mathrm{op}} \times \Lambda$ actions: $\Lambda^{\mathrm{op}} \times \Lambda \xrightarrow{\rho} \mathcal{V}$. The composition $\Lambda \xrightarrow[m \mapsto (m^{\mathrm{op}}, m)]{\Delta} \Lambda^{\mathrm{op}} \times \Lambda \xrightarrow{\rho} \mathcal{V}$ describes the action of gluing an $\mathcal{R}$ or $\mathcal{A}$ on two sides (Figure 4.3, Figure 4.4).

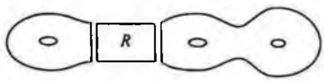


Fig. 4.3. $\Lambda^{\mathcal{R}}$ acts on both sides

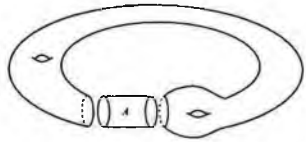


Fig. 4.4. Λ^{A} acts on both sides

We refer to the action of Λ^{op} as "left" and the action of Λ as "right".

Definition 4.2.

1) Let M be a right Λ-representation (or "module") and N a left Λ-module. Denote by $M \bigotimes_\Lambda N$ the $\mathbb{C}$-module quotient of $\bigoplus_a M_a \bigotimes_{\mathbb{C}} {}_aN$ by all relations of the form $u\alpha \bigotimes v = u \bigotimes \alpha v$, where "a" and "b" are general objects of Λ, $u \in M_a$, $v \in {}_bN$ and $\alpha \in {}_a\Lambda_b = \mathrm{Mor}(a,b)$.

2) A $\mathcal{C} \times \mathcal{D}$ bimodule is a functor $\rho : \mathcal{C}^{\mathrm{op}} \times \mathcal{D} \longrightarrow \mathcal{V}$. Note that $\mathcal{C}$ is naturally a $\mathcal{C} \times \mathcal{C}$ bimodule, which will be called the regular representation of $\mathcal{C}$.

Suppose now that Λ is semi-simple. This means that there is a set I of isomorphism classes of (finite dimensional) irreducible representations $\rho_i, i \in I$ of Λ and every (finite dimensional) representation of ρ may be decomposed $\rho \cong \bigoplus V_i \bigotimes \rho_i$, where V_i is a $\mathbb{C}$-vector space with no Λ action; $\dim(V_i)$ is the multiplicity of ρ_i. (If $\rho_1(a) = M_a^1$ and $\rho_2(a) = M_a^2$, then $\rho_1 \bigoplus \rho_2 = M_a^1 \bigoplus M_a^2$, and similarly for morphisms.)

The following example is contained in the section in the general discussion, but it is instructive to see how things work in $\mathrm{TLJ}^{\mathcal{R}}_{d=1}$ and $\mathrm{TLJ}^{A}_{d=1}$, the TLJ rectangular and picture categories for $d = 1$. These simple examples include the celebrated toric codes TQFT in[Ki1] or $\mathbb{Z}_2$ gauge theory, and illustrate the general techniques. Since it is almost no extra work, we will include the corresponding calculation for $\mathrm{TLJ}^{\mathcal{R}}_{d=-1}$ and $\mathrm{TLJ}^{A}_{d=-1}$ where $A = e^{\frac{2\pi i}{12}}, d = -1$ and $p_2 = 0$ for $d = -1$.

A general element $x \in {}_a\Lambda^{\mathcal{R}}_{\pm 1 b}$ is determined by its coefficients of "squeezed" diagrams where only 0 and 1 arcs cross the midlevel of the rectangle such diagrams look like:

Similarly $x \in {}_a\Lambda^{A}_{\pm 1, b}$ are determined by the coefficients of the diagrams

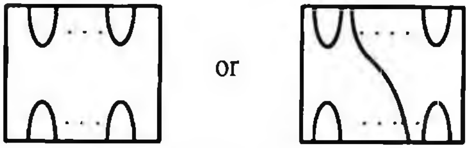


Fig. 4.5. Squeezed morphisms

in an annulus made by gluing the left and right sides of FIGS??. To each $a \in \Lambda^0, \Lambda = \Lambda_{\pm 1}^{\mathcal{J} \text{ or } \mathcal{A}}$, let V_a be the vector space spanned by diagrams, with a end points on the top (outside) and zero (a even) or one (a odd) end point on the bottom (inside), thus $V_a = {}_{0 \text{ or } 1}\Lambda_a$. The gluing map ${}_{0 \text{ or } 1}\Lambda_a^{\mathcal{J}} \bigotimes {}_a\Lambda_b^{\mathcal{J}} \xrightarrow{\rho^{\mathcal{J}}} {}_{0 \text{ or } 1}\Lambda_b^{\mathcal{J}}$ provides the two representations $\rho_0^{\mathcal{J}}$ and $\rho_1^{\mathcal{J}}$ of $\Lambda_{\pm 1}^{\mathcal{J}}$ ($\rho_{0 \text{ (or } 1)}$ sends a odd (or even) to the 0-dimensional vector space.)

Lemma 4.1. *The representation $\rho_0^{\mathcal{J}}$ and $\rho_1^{\mathcal{J}}$ are irreducible.*

Proof. Consider $\rho_0^{\mathcal{J}}$, the morphism vector space ${}_{2k}\Lambda_{2k}$ has dimension=1 (spanned by the empty diagram in a rectangle) so that in "grade", $2k$, $\rho_0^{\mathcal{J}}$ is automatically irreducible. There is a morphism $m \in {}_{2k}\Lambda_{2n}$, $m^\dagger \in {}_{2n}\Lambda_{2k}$:

Fig. 4.6. Factored morphisms

and $m^\dagger m = \text{id} \in {}_{2k}\Lambda_{2k}$. Thus any representation $\{V\}$ on the even grades of the categories must have equal dimension in all (even) grades since $\rho_0^{\mathcal{J}}(m)$ and $\rho_0^{\mathcal{J}}(m^\dagger)$ are inverse to each other. It follows that any proper subrepresentation of $\rho_0^{\mathcal{J}}$ must be zero dimensional in all grades. Thus $\rho_0^{\mathcal{J}}$ is irreducible.

The argument for $\rho_1^{\mathcal{J}}$ is similar, simply add a vertical line near the right margin of the rectangles in Fig. 4.6 to obtain the corresponding $m, m^\dagger$ in the odd grades. □

Lemma 4.2. *Any irreducible representation of $\Lambda_{\pm 1}^{\mathcal{J}}$ is isomorphic to $\rho_0^{\mathcal{J}}$ or $\rho_1^{\mathcal{J}}$.*

Proof. The proof is based on "resolutions of the identity". In this case that means:

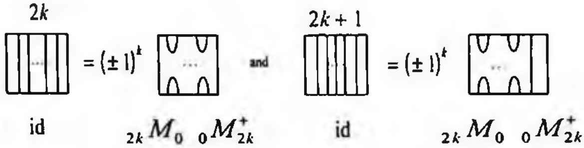


Fig. 4.7. Resolution of identity

Acting by ρ on $\{V\}$ may be factored schematically as shown in Figure 4.8.

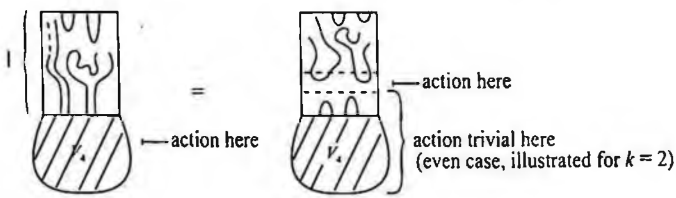


Fig. 4.8. Picture action

By Theorem B.1, for every $a \in \Lambda^0$, V_a is a subspace of ${}_a\Lambda_b$ for sone $b \in \Lambda^0$. In formulas, let $l \in {}_{2n}\Lambda_{2k}$ (for the even case) $\rho(l) = \rho\left(l \cdot {}_{2k}m_0^\dagger \cdot {}_0m_{2k}\right)$, so the action factors through ${}_0\Lambda_{2k}$. On the even (odd) grades the action is isomorphic to $\rho_0^{\mathcal{T}}$ $(\rho_1^{\mathcal{T}})$ tensor the subspace of ${}_{2n}\Lambda_0$ generated by elements of the form $l \cdot {}_{2k}m_0^\dagger$ with trivial action. So the general representation is isomorphic to a direct sum of irreducibles. In this simple case it was not necessary (as it will be in other cases) to construct the Hermitian structure on Λ to derive semi-simplicity. $\square$

Now consider representations of $\Lambda^{\mathcal{A}}_{\pm 1}$. Again $x \in {}_a\Lambda^{\mathcal{T}}_{\pm 1,b}$ is determined by diagrams with a "weight" of 0 or 1.

In the special ("principle graph") cases: ${}_0\Lambda_0$ and ${}_1\Lambda_1$ there are four diagrams (Figure 4.9) up to isotopy in the presence of relations $p_2 = 0$ for $d = \pm 1$.

The reader should observe that if pictures are glued to be outside of $\emptyset$, ring R, straight arc I, or twist T they may be transformed to another

Fig. 4.9. Idempotents for annular $d = \pm 1$

picture:

$$\emptyset \bigotimes R = R,\ R \bigotimes R = \pm\emptyset,\ I \bigotimes T = T, \text{ and } T \bigotimes T = \pm T.$$

(The signs are for $d = \pm 1$). Let us call the object (i.e. number of end points) a "crude label". We have two crude labels "0" and "1" in this example. For each crude label the symmetric $\left(\frac{\emptyset+R}{2}, \text{and } \frac{I+T}{2}\right)$ and anti-symmetric $(\frac{\emptyset-R}{2}, \text{and } \frac{I-T}{2})$ averages are in fact $(+1,-1)$ eigenvectors under gluing on a ring R in ${}_0\Lambda_{1,0}$ and gluing on T in ${}_1\Lambda_{1,1}$. The combinations $(\emptyset - iR)$ and $(\emptyset + iR)$ are ± 1-eigenvectors for the action of R in ${}_0\Lambda_{-1,0}$ and $(T - iT)$ and $(T + iT)$ are ± 1-eigenvectors for the action of T in ${}_0\Lambda_{-1,0}$. In all cases these vectors span a 1-dimensional representation of four algebras ${}_0\Lambda_{1,0}, {}_1\Lambda_{-1,0}, {}_1\Lambda_{1,1}, {}_1\Lambda_{-1,1}$ in which they lie. That is, ${}_0\Lambda_0$ and ${}_1\Lambda_1$ have the structure of commutative rings under gluing $(\cdot)$ and formal sum $(+)$. They satisfy ${}_0\Lambda_{\pm 1,0} \cong \mathbb{C}[R]/(R^2 = \pm\emptyset)$ and ${}_1\Lambda_{\pm 1,1} = \mathbb{C}(T)/(T^2 = \pm I)$ with $\emptyset$ and I serving as respective identities.

What is more important than the representations of these algebras is the representations of the entire category ${}_a\Lambda_{\pm 1,b}$ in which they lie. Similar to the rectangular case, these four representatives together form the "principle graph" from which the rest of the "Bratteli diagram" for full category representations follows formally.

Lemma 4.3. *These 4 representations of $\Lambda^{\mathcal{A}}_{\pm 1}$ are a complete set of irreducibles.*

The Bratteli diagram in Figure 4.10 explains how to extend the algebra representations to the linear category ("algebroid") in the rectangle $\mathcal{R}$ case.

All vector spaces above are 1-dimensional and spanned by the indicated picture in $\mathcal{R}$ and the $\nearrow$ is "add line on right", the $\searrow$ "bend right". The annular case is similar and is shown in Figures 4.11, 4.12.

Note that we interpret the rectangles as having their left and right sides glued.

In the case of annular categories there is no tensor structure (horizontal stacking) so in general the arrows present in the $\mathcal{R}$-case seems more difficult

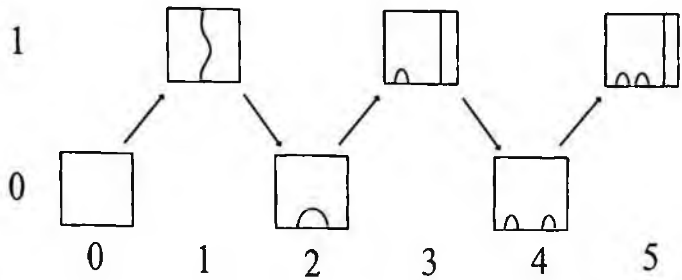


Fig. 4.10. Bratteli diagram

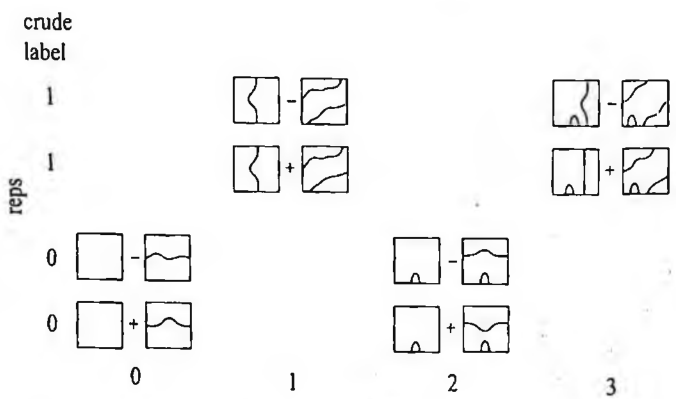


Fig. 4.11. Λ^A_{+1}

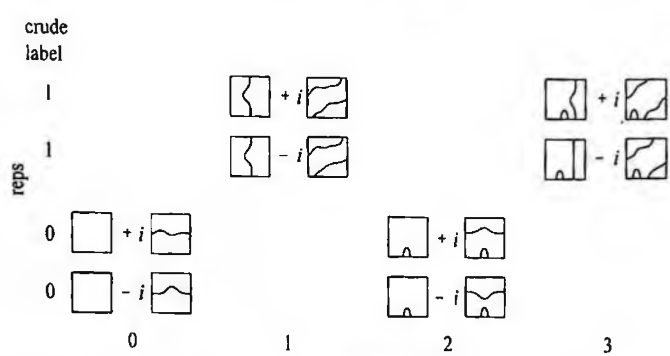


Fig. 4.12. Λ^A_{-1}

to define in generality, but should be clear in these examples. In the annular diagrams above all vector spaces of morphisms ${}_a\Lambda_b$ have dimension=2 if $a = b \ mod \ 2$ and zero otherwise. As in the rectangular case, resolutions of the identity morphisms of $\Lambda^A_{\pm 1}$ into morphism which factor through 0 or 1-strand show that all representations are sums of the four. One dimensionality and the existence of invertible morphisms between grades (exactly those shown in Fig. 4.6, but now with the convention that the vertical sides of rectangles are glued to form an annulus) again show that the four are irreducible.

By the corollary B.1 to Schur's lemma, the above decompositions into irreducibles are all unique. There are direct generalizations of the categories so far considered to Temperley-Lieb-Jones categories in the next section.

4.2. *Cutting and paste as Morita equivalence*

Crude labels for picture categories are given as finitely many points of the boundary. In the gluing formulas for TQFTs, labels are irreps of the picture categories. The passage from the crude labels of points to the refined labels of irreps is Morita equivalence.

Definition 4.3. Two linear categories $\mathcal{C}$ and $\mathcal{D}$ are Morita equivalent if there are $\mathcal{C} \times \mathcal{D}$ bimodule M and $\mathcal{D} \times \mathcal{C}$ bimodule N such that $M \otimes N \cong \mathcal{C}$ and $N \otimes M \cong \mathcal{D}$ as bimodules.

Let Λ be a linear category, $\{a_i\}_{i \in I}$ a family of objects of Λ. For each $i \in I$, let e_i be an idempotent in the algebra ${}_{a_i}\Lambda_{a_i}$. Define a new linear category Δ as follows: the objects of Δ is the index set I, and the morphism set ${}_i\Delta_j = e_{i a_i}\Lambda_{a_j}$.

Given an object a in Λ, define the $\Delta \times \Lambda$ bimodule M as ${}_iM_a = e_{i a_i}\Lambda_a$, and the $\Lambda \times \Delta$ bimodule N as ${}_aN_i = {}_a\Lambda_{a_i} e_i$.

A key lemma is the following theorem in Appendix A of:[BHMV]

Theorem 4.1.

Suppose the idempotents e_i generate Λ as a two-sided ideal. Then the bimodule $M \otimes_\Lambda N \cong \Delta$ and $N \otimes_\Delta M \cong \Lambda$, i.e., Λ and Δ are Morita equivalence.

Consequently, tensoring (on the left or right), by the modules N and M, gives rise to the Morita equivalence of Λ and Δ. Moreover, these equivalences preserves tensor product of bimodules.

Given two surfaces Y_1, Y_2 such that $\partial Y_1 = \bar{\gamma}_1 \coprod \gamma, \partial Y_2 = \bar{\gamma} \coprod \gamma_2$, and the picture spaces $\mathrm{Pic}(Y_1), \mathrm{Pic}(Y_2)$ are bimodules over the picture category

Λ, then the picture space $\mathrm{Pic}(Y_1 \cup_\gamma Y_2)$ is the tensor product of $\mathrm{Pic}(Y_1)$ and $\mathrm{Pic}(Y_2)$ over Λ. Morita equivalence, applied to the picture category Λ, sends the bimodule $_-\mathrm{Pic}(Y_1 \cup_\gamma Y_2)_-$ to the bimodule $_-\mathrm{Pic}(Y_1) \otimes \mathrm{Pic}(Y_2)_-$ (over the representation category Δ of the picture category Λ) because tensor products are preserved under Morita equivalence. Now the general gluing formula can be stated as a consequence of Morita equivalence:

Theorem 4.2.
Let Y_1, Y_2 are two oriented surfaces such that $\partial Y_1 = \bar{\gamma}_1 \coprod \gamma$ and $\partial Y_2 = \bar{\gamma} \coprod \gamma_2$. Then the picture bimodule $_-V(Y_1 \cup_\gamma Y_2)_-$ is isomorphic to $_-V(Y_1) \otimes_\Delta V(Y_2)_-$ as bimodules.

As explained in Appendix B, the idempotents e_i label a complete set of irreps of the linear category Λ. Therefore, gluing formulas for picture TQFTs need the representation categories of the picture categories. In the axioms of TQFTs the label set was a mysterious feature, now we will see its origins in picture TQFTs.

Now let us write the Morita equivalence more explicitly. Let Λ be some $\Lambda^{\mathcal{R}}$ (or $\Lambda^{\mathcal{A}}$) and suppose Λ is semi-simple with index set I. The pictures in a fixed 3-manifold (surface) with a "left" and "right" gluing region provide a bimodule $_aB_b$ on for Λ. If the gluing region is not connected within the 3-manifold (surface) then $B \cong B^{\text{left}} \bigotimes_\Lambda B^{\text{right}}$. We treat this case first.

Lemma 4.4. $B^l \bigotimes_\Lambda B^r \cong_{i\in I} (V_i \bigotimes W_i)$, where $_aB^{\text{right}} \cong \bigoplus_i V_i \otimes \rho_i$ and $_aB^{\text{left}} \cong \bigoplus W_j \otimes \rho_j^{\text{op}}$, $(\rho_j^{\text{op}}(m) = (\rho_j(m))^\dagger)$.

Proof. Note that $\rho_i^{\text{op}} \bigotimes \rho_j \cong \mathrm{Hom}_\Lambda(\rho_i, \rho_j) \cong \begin{cases} \mathbb{C} \text{ if } i=j \\ 0 \text{ if } i\neq j \end{cases}$. As usual $\bigoplus_\Lambda$ distributes over $\bigotimes_\Lambda$, the unusual feature is that the coefficients are vector spaces V_i and W_j, not complex numbers. They are "multiplied" by (ordinary) tensor product $\bigotimes$. $\square$

The manipulations above are standard in the context of 2-vector spaces[Fd, Wal2] and in fact a representation is a 2-vector in the 2-vector space of all formal representations.

Now suppose the regions to be glued to the opposite ends of $\mathcal{R}$ ($\mathcal{A}$) are part of a connected component of a 3-manifold (surface), then write the bi-module $_aB_b \cong \bigoplus_{(i,j)\in I^{\text{op}}\times I} W_{ij} \otimes (\rho_i^{\text{op}} \bigotimes_\Lambda \rho_j)$ as a bimodule. Define a 2-trace,

$$\mathrm{tr}B = \bigoplus_{a \in \mathrm{obj}(\Lambda)} {}_aB_a/u\alpha \cong \alpha^{\mathrm{op}}u,$$

where $\alpha \in {}_a\Lambda_b, u \in {}_bB_a$ are arbitrary. Again, "linear algebra" yields:

Lemma 4.5. *$tr(B) \cong \bigoplus_{i \in I} W_{ii}$.*

Proof. Schur's lemma implies $\rho_i^{\mathrm{opp}} \bigotimes_{\mathbb{C}} \rho_j \cong \mathbb{C}$ iff $i = j$. □

Note that disjoint union of the spaces carries over to tensor product, $\bigotimes_{\mathbb{C}}$, of the modules of pictures on the space. This makes lemma 4.4 a special case of lemma 4.5. And further observe that both lemmas match the form of the "gluing formula" as expected, with $I = \mathcal{L}$, the label set, and adjoint ($\dagger$) is the involution$\widehat{\ }$: $\mathcal{L} \to \mathcal{L}$.

4.3. *Annualization and quantum double*

Annular categories are closely related to the corresponding rectangle categories. In particular, there is an interesting general principle:

Conjecture: *If $\Lambda^{\mathcal{R}}$ and $\Lambda^{\mathcal{A}}$ are rectangular and annular versions of locally defined picture/relation categories, then $(\mathcal{D}(Rep(\Lambda^{\mathcal{R}})) \cong Rep(\Lambda^{\mathcal{A}})$, the Drinfeld center or quantum double of the representation category of the rectangular picture category is isomorphic to the representation category of the corresponding annular category.*

The conjecture and its higher category generalizations are proved in.[Wal2]

5. Temperley-Lieb-Jones categories

To obtain the full strucure of the picture TQFTs, we need to consider surfaces with boundaries, and boundary conditions for the corresponding vector spaces $V(Y)$. The crude boundary conditions using objects in TLJ categories are not suitable for the gluing formulas. As shown in Section 3.7, Section 4, we need to find the irrpes of the TLJ categories. Two important properties of boundary condition categories needed for TQFTs are semisimplicity and the finiteness of irreps. For TLJ categories, both properties follow from a resolution of the identity in the Jones-Wenzl projectors.

Let X be a compact *parameterized* n-manifold. The interesting cases in this paper are the unit interval $I = [0, 1]$ or the unit circle S^1. Define a category $\mathcal{C}(X)$ as follows: an object a of $\mathcal{C}(X)$ consists of finitely many points in the interior of X, and given two objects a, b, a morphism in $\mathrm{Mor}(a, b)$ is an $(n+1)$-manifold, not necessarily connected, in the interior of $X \times [0,1]$ whose boundaries are $a \times 0, b \times 1$, and intersects the boundary of $X \times [0,1]$ transversely. Given two morphisms $f \in {}_a\mathcal{C}_b, g \in {}_b\mathcal{C}_c$, the composition of f, g will be just the vertical stacking from bottom to top followed by the rescaling of the height to unit length 1. When X is a circle, we will also draw the vertical stacking of two cylinders as the gluing of two annuli in the plane from inside to outside. More often, we will draw the stacking of cylinders as vertical stacking of rectangles one on top of the other with periodic boundary conditions horizontally. Note the two boundary circles of a cylinder are parameterized, so they have base points and are oriented. The gluing respects both the base-point and orientation.

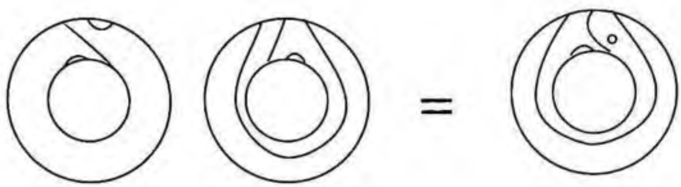


Fig. 5.1. Composition of annular morphisms

Given a non-zero number $d \in \mathbb{C}$, the *Temperley-Lieb category* TL_d is the linear category obtained from $\mathcal{C}([0,1])$ by first imposing d-isotopy in each morphism set, and then taking formal finite sums of morphisms as follows: the objects of TL_d are the same as that of $\mathcal{C}([0,1])$, and for any two objects a, b, the vector space $\mathrm{Mor}_{\mathrm{TL}}(a, b)$ is spanned by the set $\mathrm{Mor}(a, b)$ modulo d-isotopy.

The structure of the Temperley-Lieb categories TL_d depends strongly on the values of d as we have seen in the *Temperley-Lieb algebras* $\mathrm{TL}_n(d) = \mathrm{Mor}(a, a)$ for any object $a \in \mathrm{TL}_d^0$ consisting of n points. When A is as in Lemma 3.1, the semi-simple quotient of the Temperley-Lieb category TL_d by the Jones-Wenzl idempotent p_{r-1} is a semi-simple category. The associated semi-simple algebras $\mathrm{TL}_n(d)$ were first discovered by Jones in.[Jo4] Therefore the semi-simple quotient categories of TL_d for a particular d will be called the *rectangular Temperley-Lieb-Jones category* $\mathrm{TLJ}_d^{\mathcal{R}}$, where $d = -A^2 - A^{-2}$. Note that there will be several different A's which result in the same TLJ category as the coefficients of the Jones-Wenzl idempotents are rational functions of d. If we replace the interval $[0,1]$ in the definition of the Temperley-Lieb categories by the unit circle S^1, we get the *annular*

Temperley-Lieb categories $\mathrm{TL}_d^{\mathcal{A}}$, and their semi-simple quotients *the annular Temperley-Lieb-Jones categories* $\mathrm{TLJ}_d^{\mathcal{A}}$.

5.1. *Annular Markov trace*

In the analysis of the structure of the TL algebras, the Markov trace defined by Figure 2.4 in Section 2 plays an important rule. In order to analyze the annular TLJ categories, we introduce an annular version of the Markov trace and 2-category generalizations.

Recall that $\Delta_n(x)$ is the Chebyshev polynomial. Let $C_n(x)$ be the algebra $\mathbb{C}[x]/(\Delta_n(x))$. Inductively, we can check that the constant term of Δ_n is not 0 if n is odd, and is 0 if n is even. For n even, the coefficient of x is $(-1)^{\frac{n}{2}}n$. Let $n = 2m$ and $q_{2m}(x)$ be the element of $C_n(x)$ represented by $\frac{\Delta_{2m}(x)}{x}$.

Define the annular Markov trace $Tr^{\mathcal{A}}$ as follows: $Tr^{\mathcal{A}} : \mathrm{TL}_{n,d} \to C_n(x)$ is defined exactly the same as in Figure 2.4 in Section 2 except instead of counting the number of simple loops in the plane, the image becoming elements in the annular algebra, where x is represented by the center circle(=called a ring sometimes).

Prop 5.1. $Tr^{\mathcal{A}}(p_n) = \Delta_n(x)$.

It follows that the algebra $C_n(x)$ can be identified as the annular algebra when d is a simple root of $\Delta_n(x)$.

If the inside and outside of the annulus $\mathcal{A}$ are identified, we have a torus T^2. The annular Markov trace followed by this identification leads to a 2-trace from $\mathrm{TL}_{n,d}$ to the vector space of pictures in T^2.

5.2. *Representation of Temperley-Lieb-Jones categories*

Our goal is to find the representations of a TLJ category $\mathrm{TLJ}_d^{\mathcal{R}}$ or $\mathrm{TLJ}_d^{\mathcal{A}}$. The objects consisting of the same number of points in such categories are isomorphic, therefore the set of natural numbers $\{0, 1, 2, \cdots\}$ can be identified with a skeleton of the category (a complete set of representatives of the isomorphism classes of objects). Each morphism set $\mathrm{Mor}(i, j)$ is spanned by pictures in a rectangle or an annulus.

To find all the irreps of a TLJ category, we use Theorem B.1 in Appendix B to introduce a table notation as follows: we list a skeleton $\{0, 1, \cdots,\}$ in the bottom row. Each *isomorphism* class ρ_j of irreps of the category is represented by a row of vector spaces $\{V_{j,i}\} = \{\rho_j(i)\}$. Each column of vector spaces $\{V_{i,j}\}$ determines an isomorphism class of objects of the category. The graded morphism linear maps of any two columns will be ${}_i\mathrm{TLJ}_j$, in particular the graded linear maps of any column to itself give rise to the decomposition of ${}_i\mathrm{TLJ}_i$ into matrix algebras, i.e., ${}_i\mathrm{TLJ}_i = \oplus_j \mathrm{Hom}(V_{j,i}, V_{j,i})$. To find all irreps of TLJ, we look for minimal idempotents of ${}_i\mathrm{TLJ}_i$ starting from $i = 0$. Suppose there exists an m_0 such that the irreps $\{e_{j_i}\}$ of ${}_j\mathrm{TLJ}_j, j \leq m_0$ are sufficient to decompose every ${}_m\mathrm{TLJ}_m$ as $\oplus_{j_i}\mathrm{Hom}(V_{j_i,m}, V_{j_i,m})$ for all $m \geq m_0$, then it follows that all irreps of TLJ are found; otherwise, a non-zero new representation space $V_{k,a}$ from some new irrep ρ_k and $a \in \mathrm{TLJ}^0$ implying $\mathrm{Hom}(V_{k,a}, V_{k,a}) \subset {}_a\mathrm{TLJ}_a$ will contradict the fact that ${}_a\mathrm{TLJ}_a \cong \oplus_{j\neq k}\mathrm{Hom}(V_{j,a}, V_{j,a})$.

Remark: For the annulus categories, we can identify one irrep as the trivial label using the disk axiom of a TQFT. Given a particular formal picture x in an annulus, we define the disk consequences of x as all the formal pictures obtained by gluing x to a collar of the disk: given a picture y on the disk, composition x and y is a new picture in the disk. By convention, pictures with mismatched boundary conditions are 0. Then the trivial label is the one whose disk consequences form the vector space $\mathbb{C}$, while all others would result in 0.

For an object $m \in \mathrm{TLJ}^0$ if $\mathrm{id}_m = \oplus_{j<m}(\oplus_l f^l_{m,j} \cdot g^l_{j,m})$ for $f^l_{m,j} \in {}_m\mathrm{TLJ}_j, g^l_{j,m} \in {}_j\mathrm{TLJ}_m$, where l is a finite number depending on j, then we have a *resolution of the identity of* m into lower orders.

Lemma 5.1. *If for some object m of a TLJ category, we have a resolution of its identity id_m into lower orders, then every irrep of the category TLJ is given by a minimal idempotent in ${}_j TLJ_j$ for some $j < m$.*

Given a TLJ category and two objects $a, c \in \mathrm{TLJ}^0$, there is a subalgebra, denoted by A^a_{cc}, of the algebra $A_{cc} = {}_c\mathrm{TLJ}_c$ consisting all morphisms generated by those factoring through the object a: $f \cdot g, f \in {}_c\mathrm{TLJ}_a, g \in {}_a\mathrm{TLJ}_c$. If e_a is an idempotent of ${}_a\mathrm{TLJ}_a$, then $A^{e_a}_{cc}$ denotes the subalgebra of A^a_{cc} consisting all morphisms generated by those factoring through e_a, i.e., those of the form $f \cdot e_a \cdot g$.

Lemma 5.2. *Given two objects a, b of a TLJ category, and two minimal idempotents $e_a \in {}_a TLJ_a, e_b \in {}_b TLJ_b$, then*

1): $A^{e_a}_{cc}$ is the simple matrix algebra over the vector space ${}_c TLJ_a e_a$.

2): If the two representations $e_a TLJ, e_b TLJ$ are isomorphic, then for any $c \in TLJ^0$, which is neither a nor b, the subalgebras $A^{e_a}_{cc}, A^{e_b}_{cc}$ of A_{cc} are equal.

We will use these lemmas to analyze representations of TLJ categories, but first we consider only the low levels.

5.3. *Rectangular Tempeley-Lieb-Jones categories for low levels*

Denote $A_{ij} = {}_i\Lambda_j$. Note that A_{ii} is an algebra, and $A_{ij} = 0$ if $i \neq j \bmod 2$. The Markov trace induces an inner product $<,>: A_{ij} \times A_{ij} \to \mathbb{C}$ on all A_{ij} given by $< x, y >= \mathrm{Tr}(\bar{x}y)$.

5.3.1. *Level=1,* $d^2 = 1$

Using $p_2 = 0$, we can "squeeze" a general element $x \in A_{ij}$ so that there are only 0 or 1 arcs cross the mid-level of the rectangle. Such diagrams in Figure 4.5) in Section 4.1.

The algebra $A_{00} = \mathbb{C}$, and the empty diagram is the generator. The first irrep ρ_0 of $\mathrm{TLJ}^{\mathfrak{R}}_{d=\pm1}$ is given the idempotent p_0, which is just the identity $\mathrm{id}_\emptyset$ on the empty diagram: if j is odd, $\rho_0(j) = 0$; if j is even, $\rho_0(j) = A_{0j} \cong \mathbb{C}$.

The algebra $A_{11} = \mathbb{C}$, generated by a single vertical line. The identity does not factor through the 0-object, so we have a new idempotent p_1 (=idenity on the vertical line). The resulting irrep ρ_1 sends even j to 0, and odd j to $A_{1j} \cong \mathbb{C}$.

Continuing to A_{22}, we see that the identity on two strands does factor through p_0 given by the Jones-Wenzl idempotent p_2. By Lemma 5.1, we have found all the irreps of $\mathrm{TLJ}^{\mathfrak{R}}_{\pm1}$, which are summarized into Table 1.

ρ_1	0	1
ρ_0	1	0
	0	1

$\mathrm{TLJ}^{\mathfrak{R}}_{d=1}$ does not lead to a TQFT since the resulting S-matrix $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ is singular. Although $\mathrm{TLJ}^{\mathfrak{R}}_{d=-1}$ does give rise to a TQFT, the resulting theory with S-matrix $= \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}$ is not unitary. The semion theory with S-matrix $= \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$ can be realized only by the representation category of the quantum group $SU(2)$ at level=1. This subtlety comes from the Frobenius-Schur indicator of the non-trivial label, which is 1 for TLJ and -1 for quantum group.

5.3.2. *Level=2,* $d^2 = 2$

Since p_3 is a resolution of the identity of id_3 into lower orders, it suffices to analyze A_{ii} for $i \le 2$. The cases of A_{00}, A_{11} are the same as level=1. Since $\dim A_{20}$=1, $\dim A_{21}$=0 and $\dim A_{22}$=2, id_2 does not factor through lower orders, so there is a new idempotent in A_{22}. The 1-dimensional subalgebra A_{22}^0 is generated by e_2, which is the following diagram:

$\frac{1}{d}$ 

It is easy to check e_2 is the identity of A_{22}^0. Since the identity of A_{22} is the sum of the two central idempotents (the two identities of each 1-dimensional subalgebra), the new idempotent p_2 is $\mathrm{id}_2 - e_2$. The irrep corresponding to p_2 sends each odd j to 0, and each even j to $p_2 A_{2j}$.

Therefore, the irreps of the level=2 $\mathrm{TLJ}^{\mathfrak{R}}$ are given by p_0TLJ, p_1TLJ, p_2TLJ, which are summarized into Table 2.

ρ_2	0	0	1
ρ_1	0	1	0
ρ_0	1	0	1
	0	1	2

5.3.3. *Level=3,* $d^2 = 1 + d$ *or* $d^2 = 1 - d$

The same analysis for objects $0, 1, 2$ yields three idempotents p_0, p_1, p_2. Direct computation shows $\mathrm{Hom}(3,3) \cong \mathbb{C}^5$, $\mathrm{Hom}(3,0) = \mathrm{Hom}(3,2) = 0$ and $\mathrm{Hom}(3,1) \cong \mathbb{C}^2$. By Lemma 5.2, $A_{33}^{p_1} = A_{33}^1$ is the 4-dimensional algebra of 2×2 matrices over the vector space A_{13}. Let v_1, v_2 be the two vectors of A_{31} represented by diagrams such that $< v_1, v_2 >=< v_2, v_1 >= d^2$, and $< v_1, v_1 >=< v_2, v_2 >= d$. Using Gram-Schmidt on the vectors v_1, v_2, we get an orthonormal basis $e_1 = \frac{v_1}{d}, e_2 = \frac{v_2 - e_1}{d^2 - 1}$ of A_{31}. Hence the identity of the algebra A_{33}^1 is $|e_1 >< e_1| + |e_2 >< e_2|$. Therefore, the remaining idempotent of A_{33} is $\mathrm{id}_3 - |e_1 >< e_1| - |e_2 >< e_2|$, which is just p_3. It follows that the irreps of TLJ are given by p_0TLJ, p_1TLJ, p_2TLJ, p_3TLJ.

5.4. *Annular Temperley-Lieb-Jones theories for low levels*

First we have the following notations for the pictures in the annular morphism sets $A_{00}, A_{11}, A_{02}, A_{22}$, where $\mathbf{1_0}, R, B, \mathbf{1_1}, T_1, \mathbf{1_2}, T_2$ are annular diagrams: $\mathbf{1_0}, \mathbf{1_1}, \mathbf{1_2}$ are identities with $0, 1, 2$ strands, R is the ring, B is the birth, and T_1 is the Dehn twisted curve, and T_2 is the fractional Dehn twisted curve. We also use B' to denote the diagram of RB after Z_2 homology surgery. A diagram with a $\bar{}$ is the one obtained from a reflection through a horizontal line.

5.4.1. *Level=1,* $d^2 = 1$

The Jones-Wenzl idempotent p_2 is a resolution of id_2 into the lower orders, so we need only to find the minimal idempotents of $\mathrm{Hom}(0,0)$ and $\mathrm{Hom}(1,1)$. Since any two parallel lines can be replaced by a turn-back, the algebra A_{00} is generated by the empty picture $\emptyset$ and the ring circle R. Stacking two rings R together and resolving the two parallel lines give $R^2 = 1$, hence A_{00} is the algebra $\mathbb{C}[R]/(R^2-1)$. By Lemma 2.1, the two minimal idempotents of A_{00} are $e_1 = \frac{\emptyset+R}{2}, e_2 = \frac{\emptyset-R}{2}$. To test which idempotent is of the trivial type, we apply e_1, e_2 to the empty diagram on the disk and obtain $e_1\emptyset = (\frac{d+1}{2})\emptyset, e_2\emptyset = (\frac{1-d}{2})\emptyset$. Hence if $d=1$, then e_1 is of the trivial type, and if $d=-1$, then e_2 is of the trivial type.

The algebra A_{11} is generated by the straight arc I and the tiwst T. By stacking two rings R together and resolving the two parallel lines, we see that A_{11} is the algebra $\mathbb{C}[T]/(T^2-dI)$. By Lemma 2.1, for $d=1$, we have two minimal idempotents $e_{3,1} = \frac{I+T}{2}, e_{4,1} = \frac{I-T}{2}$. For $d=-1$, we have two minimal idempotents $e_{3,-1} = \frac{1-iT}{2}, e_{4,-1} = \frac{1+iT}{2}$. Note that $\mathrm{Hom}(0,1) = \mathrm{Hom}(1,0) = 0$. Therefore, the annular TLJ categories for $d = \pm 1$ have 4 irreps $e_i, i = 1,2,3,4$.

5.4.2. *Level=2,* $d^2 = 2$

For the TLJ categories at level=2, $d^2 = 2$, p_3 is a resolution of the identity of id_3 into lower orders, so we need to analyze the algebras A_{00}, A_{11}, A_{22}. The algebra A_{00} is generated by the empty picture $\emptyset$ and the ring R. Since $R^3 = 2R$, $A_{00} = \mathbb{C}[R]/(R^3-2R)$. By Lemma 2.1, the three minimal idempotents are $e_1 = \emptyset - \frac{R^2}{2}, e_2 = \frac{R^2+dR}{4}, e_3 = \frac{R^2-dR}{4}$. Testing on the disk, we know that e_2 is of the trivial type.

For A_{11}, we apply the Jones-Wenzl idempotent p_3 to the stacking of two twists T^2. After simplifying, we get $T^4 - dT^2 + 1 = 0$. Again by Lemma 2.1, we have 4 minimal idempotents: $\frac{1}{2d}(\alpha^2 I + \alpha T - \alpha^4 T^2 - \alpha^3 T^3)$, where

$\alpha^4 - d\alpha^2 + 1 = 0$.

A new phenomenon arises in the algebra A_{22}, which is generated by 8 diagrams: $1_2, \bar{B}B, T_2, \bar{B}'B, \bar{B}B', \bar{B}'B', \bar{B}RB, \bar{B}'RB$. Computing their inner products shows that $A_{22} \cong \mathbb{C}^8$. A_{02} is spanned by B, B', RB, RB'. Using the three minimal idempotents in A_{00}, we see that $e_0 A_{02}$ is spanned by $RB + RB' = f_0$, $e_1 A_{02}$ is spanned by $B - \frac{d}{2}RB' = f_1, B' - \frac{d}{2}RB = f_1'$, and $e_1 A_{02}$ is spanned by $RB - RB' = f_2$. Hence $A_{22}^0 \cong \mathbb{C}^6$ as the direct sum of 2 1×1 matrix algebras generated by f_0, f_2 and a 2×2 algebra generated by f_1, f_1'. Therefore there are two more idempotents in A_{22}. Applying p_3 to the action of the 1/2-Dehn twist F on A_{22}, we get $F^2 = 1$ modulo lower order terms, hence the last two idempotents are of the form $\frac{1}{2}(I_2 \pm iF)$ plus lower order terms in A_{22}^0. Since $A_{22}^0 = BA_{02} + B'A_{20}$, we need to find an x such that $e = -\frac{1}{2}1_2 \pm \frac{i}{2}T_2 + x$ is a projector and $eB = eB' = 0$. Solve the equations, we find

$$e_\pm = \frac{1}{2}1_2 \pm \frac{i}{2}T \mp \frac{i}{2d}\bar{B}'B - \frac{1}{2d}\bar{B}B \mp \frac{i}{2d}BB' - \frac{1}{2d}\bar{B}'B' \pm \frac{i}{2d^2}\bar{B}RB + \frac{1}{2d^2}\bar{B}'RB.$$

5.4.3. *Level=3, $d^2 = 1 + d$ or $d^2 = 1 - d$*

The algebra A_{00} is the algebra $\mathbb{C}[R]/(R^4 - 3R^2 + 1)$, so we have 4 minimal idempotents.

The algebra A_{11} is generated by the twist T, so A_{11} is the algebra $\mathbb{C}[T]/(T^6 - dT^4 - dT^2 + 1)$, so we have 6 minimal idempotents.

Let F be the fractional Dehn twist on A_{22}, then p_4 results in a dependence among $F^{-2}, F^{-1}, I_2, F, F^2$: $F^4 - dF^2 + 1$ modulo lower order terms. So we have 4 minimal idempotents.

Let F be the fractional Dehn twist on A_{33}, then p_4 results in a relation between F^{-1}, I_3, F. So we have 2 minimal idempotents.

We leave the exact formula for the idempotents to interested readers. Note that the number of irreps of the annular TLJ categories is the square of the corresponding TLJ rectangular categories.

5.5. *Temperley-Lieb-Jones categories for primitive 4rth roots of unity*

Let A be a primitive $4r$-th root of unity, and $d = -A^2 - A^{-2}$. $\mathrm{TLJ}^{\mathfrak{R},k,A}$ is just the TL_d modulo its annihilator p_{r-1}. We found that it has minimal idempotents $p_0, p_1, p_2, \cdots, p_k$, $k = r - 2$ and with image (p_{k+1}) being the annihilator of the Hermitian paring $\langle\, ,\rangle$.

The case A a primitive $2r$-th or rth root of unity, r odd, e.g. $A =$

$e^{2\pi i/6}, k = 1, d = 1$; is identical as far as the rectangle categories go, but for the annular categories is more complicated; it is analyzed in the next section.

Theorem 5.1. Rectangle diagrams with $p_i, 0 \leq i \leq k$, near the bottom and object t at top span spaces $\{W^t_{A,i}\} := W_{A,i}$ on which $\Lambda := \Lambda^{\mathcal{R},k,A}$ acts from above. The families $\{W_{A,i}\}$ (as i varies) are the $k+1$ (isomorphism classes of) irreducible representations of Λ. The involution $\hat{}$ is the identity.

Proof. Most of the argument is by now familiar. Resolving the identity shows that any representation is a direct sum of $\{W_{A,i}\}, 0 \leq i \leq k$.

For the first time $\dim(W^t_{A,i})$ may be > 1 and there will not be invertible morphism $t \to t'$ but irreducibility can still be proved as follows: for all $m = p_i \cdot m_0$ and $m' = p_i \cdot m'_0$ one may construct morphism x and y so that $m' = mx$ and $m = m'y$, where $p_i \in {}_i\Lambda_i, m, m_0 \in {}_i\Lambda_a, m', m'_0 \in {}_i\Lambda_b, x \in {}_a\Lambda_b, y \in {}_a\Lambda_b$. □

It is a bit harder to find the irreps of $\Lambda^{\mathcal{A},k,A} := \Lambda$, but we will do this now. Similar irreps for $\mathrm{TL}^{\mathcal{A}}$ categories were previously found by Graham-Lehner,[GL] in a different context.

We do not know how to proceed in a purely combinatorial fashion but must invoke the action of the doubled theory on the undoubled. Topologically this amounts to the action on pictures in the solid cylinder $(B^2 \times I, B^2 \times \partial I)$ under the addition of additional strands in a shell $(B^2_2 \backslash B^2_1 \times I; B^2_2 \backslash B^2_1 \times \partial I)$. Logically our calculation should be done until we have established† the undoubled TQFT based on $\Lambda^{\mathcal{R},k,A}$ where the hypothesis A a primitive $4r$th root is used. This can be done in Section 3 already or from here by going directly to Section 7 which does not depend on this section. Therefore, we will freely invoke this material.

In the low level cases we found that $\#\mathrm{irreps}\Lambda^{\mathcal{A}} = (\#\mathrm{irreps}\Lambda^{\mathcal{R}})^2$. This is not an accident but comes from identifying $\Lambda^{\mathcal{A}}$ with $\mathrm{End}(\Lambda^{\mathcal{R}})$. $\Lambda^{\mathcal{A},k,A}$ is too complicated to "guess" the irreps so we compute them from the endomorphism view point.

Recall from Section 3.4 the projectors $\omega_a = \sum_{c=0}^{k} \frac{\Delta_{(a+1)(c+1)}[c]}{D}$ onto the

†It has been shown in Lemma 3.2 that the Temperley-Lieb-Jones theories $\Lambda^{\mathcal{R},k,A}$ violate an important TQFT-axiom when $A^{2r} = 1$. The S-matrix is singular, half the expected rank, so the action of the mapping class group is not completely defined. In this case irreps of $\Lambda^{\mathcal{A},k,A}$ have a more complicated structure.

a-label, and $D^2 = \sum_{c=0}^{k} \Delta_{c+1}^2$.

Also recall from Section 7 if $Y = \partial X$ and $\gamma \subset \text{interior}X$ is a family of sccs labelled by $\bar{\omega_a}$ and γ cobounds a family of imbedded annuli $A \subset X$ with $\gamma' \subset Y$,, i.e. $\partial A = \gamma \cup \gamma'$, then $Z(X, \gamma_{\omega_a}) \in V(Y\backslash\gamma; a, \hat{a}) \subset \bigoplus_{\text{admissible}} V(Y\backslash\gamma; l, \hat{l}) = V(Y)$.

Consider the 4−component formal tangle in annulus cross interval, $-A\times I$, where $h = |i-j|$:

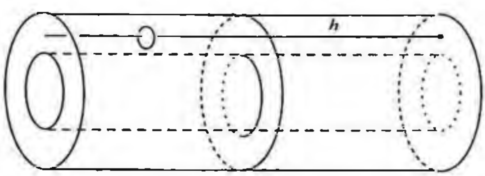


Fig. 5.2. 4-component formal tangle

Let X be the 3-manifold made by removing small tubular neighborhoods of the h-labeled arc, and write $X \cong Y\times I$, where Y is the annulus with a new puncture, and $\partial X = DY$ the double of Y. Let (Y, l_0) be Y with ∂Y labeled as follows: outer boundary $\to j$, inner boundary $\to i$, new boundary $\to h$. From Lemma 3.5 we know $V(Y, l_0) \cong V_{i,j,h} \cong \mathbb{C}$.

Another useful decomposition of ∂X results from expanding the inner and outer boundary components of Y to annuli, A_i and A_0: $\partial X = -Y \cup +Y \cup -A_i \cup +A_0 \cup A_h$. Applying V we have: $V(\partial X) =$

$$\bigoplus_{\text{admissible labels}} V^*(Y,l) \bigotimes V(Y,l) \bigotimes V^*(A_i,l) \bigotimes V(A_0,l) \bigotimes V(A_h,l). (*)$$

Let us restrict to label: l_0. By lemma, $\dim V(Y, l_0) = 1$ and let x be the unit normalized vector $\kappa \in V(Y, l_0)$,

The Jones-Wenzl projectors p_i, p_j and p_h are inserted as shown.

The arc diagram should be pushed into a ball B^+ bounding the 2-sphere S^2 made by capping ∂Y, to define an element of $V(Y, l_0)$. The root θ−symbol normalizes $||x'||^2$ to the invariant of $D(B^3) = S^3$ and the $s_{00}^{\frac{1}{2}}$ kills this factor so as defined $||x'||^2 = 1$.

Let $V_{l_0}(\partial X)$ denote the l_0 summand of the rhs (*). Fixing x', and therefore its dual $\hat{x}'$, give an isomorphism $\epsilon'_{ij} : V_{l_0} \to \text{Hom}(V(A_i; i, \hat{i}), V(A_0; j, \hat{j}) =: \text{Hom}(V_i, V_j)$.

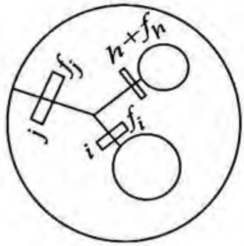


Fig. 5.3. Insert Jones-Wenzl

Consider the partition function $Z(X,L)$ of (X,L), where L is the 3-component link in X labelled by ω_i, ω_j and ω_h in FIG??, and $Z(X,L) \subset V_{l_0}(X)$.

We now check that $\epsilon'_{i,j}(Z(X,L))$ is a non-zero vector in $\mathrm{Hom}(V_i, V_j)$ whose definition is independent of phase(X).

The pairing axiom can be used to analyze the result of gluing X to the genus two handlebody $(H, \overline{\theta}_{i,j,h})$ which is a thickening of the i,j,h labelled θ-graph (with the graph inside), we get:

$$S_{00}\theta_{ijh} = S_{0i}^2 S_{0j}^2 S_{0h} \langle x, \hat{x}\rangle \langle \beta_{ii}^*, \beta_{ii}\rangle \langle \beta_{jj}, \beta_{jj}^*\rangle. \qquad (**)$$

The two factors of S_{0i} and S_{0j} come from gluing along the seems separating off the inner and outer annuli (respectively); the factor S_{0h} derives from gluing across the "new component" of ∂Y, S_{00} is the 3-sphere normalization constant, making the lhs (**) a "spherical θ-symbol". Previously we arranged $\langle x', \widehat{x'}\rangle = 1$ and $\langle \beta_{aa}^*, \beta_{a0}\rangle = S_{0,a}^{-1}$ so if we define $x = \frac{\sqrt{S_{0i}}\sqrt{S_{0j}}\sqrt{S_{0h}}}{\sqrt{S_{00}}\sqrt{\theta_{ijk}}} \cdot x' = \frac{\sqrt{S_{0i}}\sqrt{S_{0j}}\sqrt{S_{0h}}}{S_{00}}\kappa$, and redefine ϵ'_{ij} to ϵ_{ij} by replacing x' with x in its definition, we obtain:

$$\epsilon_{ij}(Z(X,L)) = \beta_{ii}^* \beta_{jj},$$

i.e. the canonical element of $\mathrm{Hom}(V_i, V_j)$.

Now attach a 2-handle to X along the "new" component of ∂Y to reverse our original construction: $X \cup$2-handle $= A \times I$. The co-core of the 2-handle should now be labeled by h and the ω_h-labeled component can be dispensed with (it is now irrelevant). Call this new idempotent 3−component formal tangle $\bar{L}_{ij}$. Fix x, as above, a map $\overline{\epsilon_{ij}}$ closely related to ϵ_{ij} is now defined: $\bar{\epsilon_{ij}} : V(A \times I, \bar{L}) \to \mathrm{Hom}(V_i, V_j)$ and as before we have

Lemma 5.3. $\bar{\epsilon_{ij}}(Z(A \times I, \bar{L})) = \beta_{ii}^* \beta_{jj}$.

Using the geometric interpretation of links in a product as operators, and using the product structure from the middle factor in $A \times I = S^1 \times I \times I$, we see $\beta^*_{i,i}\beta_{jj}$ realized by a formal knot projection $(\overline{L}) \subset S^1 \times \frac{1}{2} \times I$. This projection can be Kauffman-resolved to a formal 1−submanifold $=: L_{ij} \subset S^1 \times I$ (ignoring the constant $\frac{1}{2}$). This is the minimal, in fact 1-dimensional idempotent of $\Lambda^{A,k,A}$. In fact by counting we see that we have achieved a complete resolution of the identity, and in the annular algebras ${}_i\Lambda^A_i$ a complete list of isomorphisms classes of irreducible representations of the full annular category $\Lambda^{A,k,A}$.

Our assumption in Section 7 is that A is a primitive $4r$-th root of unity, $r = k + 2$. Section 7 constructs a TQFT with $\{\text{labels}\} = \{\text{irreps}\Lambda^{\mathcal{R},k,A}\}$. Using the s-matrix of this TQFT, we have just constructed a basis $\{\beta^*_{ii}\beta_{jj}\}$ of $(k+1)^2$ operators for $\mathrm{Hom}\left(\bigoplus_{i=0}^{k} V_i, \bigoplus_{j=0}^{k} V_j\right)$ which are geometrically represented as formal submanifolds $\{L_{ij} \subset A\}$, also an idempotent in ${}_h\Lambda^{A,k,A}_h$.

The counting argument below holds for A a primitive $4r$th or $2r$th root of unity r odd or rth root of unity r odd, and so applies in the next section as well.

By a direct count of classical (not formal) pictures up to the projector relation $p_{k+1} = p_{r-1}$ we find:

$$\dim\left({}_h\Lambda^{A,k,A}_h\right) \leq \begin{cases} k, & \text{for} \quad i = 0 \\ 2k+2-2h, & \text{for } 1 \leq h \leq k \end{cases} \qquad (***).$$

Summing over h, $\dim\left(\bigoplus_{h=0}^{k} {}_h\Lambda^{A,k,A}_h\right) \leq (k+1)^2$.

Since $\{L_{ij}, 0 \leq i,j \leq k\}$ represent as $(k+1)^2$ linearly independent operators, the above inequalities must, in fact, be equalities.

${}_0\Lambda^{A,k,A}_0$ is spanned by the empty picture $\emptyset$, the ring circle R, and its powers up to R^k. The projector decomposes R^{k+1} into a linear combination of lower terms.

For $h = 1$, let I denote the straight arc picture and T the counter clockwise Dehn twist. The pictures: $\bar{T}^{(k-1)}, \bar{T}^{k-2)}, \cdots, I, T, \cdots T^{k-1}$ appear (and are) independent but there is an obvious dependency if the list is expanded to $T^{-k} \cdots, T^k$. This dependency leads quickly to the claimed bound for $h = 1$.

For $h > 1$, the argument is similar to the above, except a fractional Dehn twist F replaces T.

Lemma 5.4. $\{L_{ij}, 0 \leq i,j \leq k\}$ is a complete set of minimal idempotents for $\{{}_h\Lambda_h^{A,k,A}\}, 0 \le h \le k$.

Proof: From Fig. 5.2, each $\bar{L}$ is a minimal idempotent and L_{ij} represents the same operator.

Fixing $h > 0$ now consider the action of "fractional Dehn twist", F on L_{ij}.

Lemma 5.5. For $i < j$, $F(L_{ij}) = -A^{i+j+2}L_{ij}$, and for $i > j$, $F(L_{ij}) = -A^{i+j+2}L_{ij}$.

Proof: Use the Kauffman relation to resolve the diagram below, noting the left kink is equal to a factor of $-A^3$ and that only the resolution indicated by arrows gives a term not killed by the projectors; its coefficient is A^{i+j-1}.

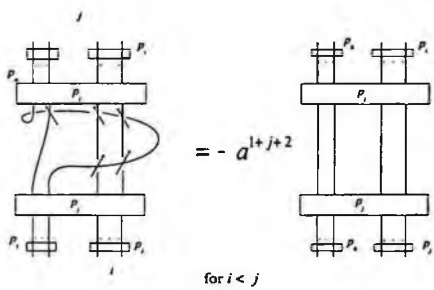


Fig. 5.4. Annular idempotent

For $i > j$ one considers the mirror image of the above, interchanging A and A^{-1}.

If $h = 0, i = j$, then $L_{ij} = \tilde{\omega_i}$ and we may consider the action R of ring addition to ω_i. Since ω_i is the projector to the i-th label, we have $R(\tilde{\omega_i}) = -(A^{2i+2} + A^{-2i-2})\tilde{\omega}_{i,}$.

This establishes:

Lemma 5.6. $R(L_{ii}) = -(A^{2i+2} + A^{-2i-2})L_{ii}$.

Let V_{ij} be the vector space spanned by formal 1-submanifolds in the annulus which near the inner boundary agree with L_{ij}.

Essentially the same argument employed for the rectangle categories: resolution of the identities but now for id $\in {}_i\Lambda_i^{A,k,A}, 0 \le i \le k$ shows:

Theorem 5.2. *The spaces $\{V_{ij}\}$ form a complete set of irreps for Rep$(\Lambda^{A,k,A})$. Direct sum decompositions into these irreducibles are unique.*

The irreps V_{ij} have another "diagonal" indexing by $(h = |i - j|, \text{eigenvalue}(i,j))$. We think of h as the "crude label" it specifies the boundary condition (object); it is refined into a true label by the additional information of eigenvalue under fractional Dehn twist.

Note that for ring addition: eigenvalue(i,j) $= -A^{i+j+2}$ for $i \neq j$.

5.6. *Temperley-Lieb-Jones categories for primitive 2rth root or rth root of unity, r odd*

In Lemma 3.2 we compute the S matrix associated to the rectangle category $\Lambda^{\mathcal{R},k,A}$, A a primitive $2r^{\text{th}}$ root or rth root of unity, r odd and find it is singular (Note that the last theorem of Ch XI[Tu] holds only for even r because the S matrix is singular for odd r). We find there is an involution on the label set $^-$: $\{0, \cdots, k\} \to \{0, \cdots, k\}$ defined by $\bar{a} = k - a$ so that $S = S_{\text{even}} \bigotimes \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$.) We use the notation i_{even} or just $i_e, 0 \leq i \leq k$, to denote the even number i or $\bar{i}$. (Note that $^-$ is not the usual duality $\hat{}$ on labels which is trivial in the TLJ theory. Also note that since k is odd exactly one of i and $\bar{i}$ is even,) The $\frac{k+1}{2}$ by $\frac{k+1}{2}$ matrix S_{even} is nonsingular and defines an $SU(2)_k^{\text{even}}$-TQFT‡ on the even labels at level k, explicitly:

$$S_{i_e j_e} = \sqrt{\frac{2}{r}}(-1)^{i+j}\left([i+1][j+1]\right). \tag{5c.1}$$

The formal 1−submanifolds $L_{i_e j_e}, 0 \leq i, j \leq k$ can be defined just as in last section. As operators on the $SU(2)_k^{\text{even}}$ TQFT they are $\beta^*_{i_e i_e}\beta_{j_e j_e}$. Also each $L_{i,j}$ has an interpretation as a formal 1-manifold in the category $\Lambda^{A,k,e^{2\pi i/6}}$. (This is the "$d = 1$" category ("$\mathbb{Z}_2$"-gauge theory) that we have been developing as a simple example.)

Letting i_O denote the odd index, i or $\bar{i}$, the tensor decomposition of the S-matrix implies $\omega_{i_e} = \omega_{i_O}$. It follows that $L_{i_e,j_e} = L_{i_O,j_O}$ and $L_{i_e,j_O} = L_{i_O,j_e}$ so we have found only half of the expected number of minimal idempotents. Let R_{i_e,j_e} be "reverse"(L_{i_e,j_e}), L_{i_e,j_e} with certain (-1) phase factors. That is, if $L_{i_e,j_e} = \sum_n a_n\alpha_n$ where α_i is a classical tangle then $R_{i_e,j_e} = \sum(-1)^{k_n} a_n\alpha_n$, where $k_n \stackrel{\text{def}}{=} k(\alpha_n))$ is the transverse intersec-

‡These TQFTs are called $SO(3)$-TQFTs by many authors. As noted in,[RSW] there is some mystery about those TQFTs as $SO(3)$-Witten-Chern-Simons TQFTs. Since they are the same TQFTs as $SU(2)$-Witten-Reshetikhin-Turaev TQFTs restricted to integral spins, therefore we adopt this notation. Their corresponding MTCs are denoted by $(A_1, k)_{\frac{1}{2}}$ in.[RSW]

tion number with a radial segment, $s \times I \subset S^1 \times I = \mathcal{A}$ in the annulus. Similarly define R_{i_e,j_O}.

Recall the four irreps of $\mathbb{Z}_2$-gauge theory, $0 = \emptyset + R$, $e = \emptyset - R$, $m = I + T$, and $em = I - T$, and consider the following bijection:

$$\{L_{i_e,j_e}, R_{i_e,j_e}, L_{i_e,j_O}, R_{i_e,j_O}\} \xrightarrow{\beta}$$
$$\{\beta^*_{i_e i_e} \bigotimes \beta_{j_e j_e} \bigotimes 0, \beta^*_{i_e i_e} \bigotimes \beta_{j_e j_e} \bigotimes e, \beta^*_{i_e i_e} \bigotimes \beta_{j_e j_e} \bigotimes m, \beta^*_{i_e i_e}$$
$$\bigotimes \beta_{j_e j_e} \bigotimes em\} \qquad (5c.2)$$

Theorem 5.3. *β is a bijection between the minimal idempotents of two graded algebras:*

$$_h\Lambda_h^{A,k,A} \xrightarrow{\beta} {}_h\Lambda_h^{A,k,e^{2\pi i/6}} \bigotimes End\left(\bigoplus_{j=even} {}_j\Lambda_j^{\mathcal{R},k,A}\right),$$

where A is a primitive $2r^{th}$ root or rth root of unity, r odd. The bijection β induces a bijection between the isomorphism classes of irreps of categories:

$$irreps.\left(\Lambda^{A,k,A}\right) \xrightarrow{\bar{\beta}} irreps.\left(\Lambda^{A,k,e^{2\pi i/6}} \bigotimes End\left(\Lambda_{even}^{\mathcal{R},k,A}\right)\right).$$

Proof: The second statement is by now the familiar consequence of the first and a "resolution of the identity."

The dimension count (upper bound) of last section applies equally for primitive $2r^{\text{th}}$ roots or rth roots, so it suffices to check that R_{i_e,j_e} and R_{i_e,j_o} are idempotents. Writing either as reverse$(L) = \sum_n (-1)^{k_n} a_n \alpha_n$ we square:

$$(\text{reverse}(L))^2 = (\sum_n (-1)^{k_n} a_n \alpha_n)^2 = \sum_{n,m} (-1)^{k_n + k_m} a_n a_m \alpha_n \alpha_m =$$
$$\text{reverse}\left(\sum_{n,m} a_n a_m \alpha_n \alpha_m\right) = \text{reverse}(L^2) = \text{reverse}(L). \quad (5c.3)$$

In the third equality holds since intersection number with a product ray in $\mathcal{A}$ is additive under stacking annuli:

$$k_n + k_m = k(\alpha_n) + k(\alpha_m) = k(\alpha_n \alpha_m) = k_{n,m}.$$

6. The definition of a TQFT

There are two subtle ingredients in the definition of a TQFT: the framing anomaly and the Frobenius-Schur (FS) indicator. For the TQFTs in this paper, the diagram and black-white TQFTs have neither anomaly nor non-trivial FS indicators, therefore, they are the easiest in this sense. The Jones-Kauffman TQFTs have anomaly, but no non-trivial FS indicators. Our version of the Turaev-Viro $SU(2)$-TQFTs have non-trivial FS indicators, but no anomaly; while the WRT TQFTs have both anomaly, and non-trivial FS indicators.

Our treatment essentially follows[Wal1] with two variations: first the axioms in[Wal1] apply only to TQFTs with trivial FS indicators, so we extend the label set to cover the non-trivial FS indicators; secondly we choose to resolve the anomaly for 3-manifolds only half way in the sense that we endow every 3-manifold with its canonical extension, so the modular functors lead to only projective representations of the mapping class groups. One reason for our choices is to minimize the topological prerequisite, and the other is that for application to quantum physics projective representations are adequate.

6.1. *Refined labels for TQFTs*

A TQFT assigns a vector space $V(Y)$ to a surface Y. If Y has boundaries, then certain conditions for ∂Y have to be specified for the vector space $V(Y)$ to satisfy desired properties for a TQFT. In Section 4, we see that crude boundary conditions need to be refined to the irreps of the picture categories, which are the labels. But for more complicated theories such as Witten-Reshetikhin-Turaev TQFTs, labels are not sufficient to encode the FS indicators. Therefore, we will introduce a boundary condition category to formalize boundary conditions. More precisely, boundary conditions are for small annular neighborhoods of the boundary circles. Our boundary condition category will be a strict weak fusion category $\mathcal{C}$, which enables us to encode the FS indicator for a label by marking boundaries with $\pm U$, where $U \in \mathcal{C}^0$. In our examples, the strict weak fusion categories are the representation categories of the TLJ categories. Then the labels are irreps of TLJ categories. In anyonic theory, labels are called superselection sectors, topological charges, or anyon types, etc. Boundary conditions which are labels are preferred because anyonic systems with such boundary conditions are more stable, while general boundary conditions such as superpositions of labels are difficult to maintain.

A fusion category is a finitely dominated semi-simple rigid linear monoidal category with finite dimensional morphism spaces and simple unit. A weak fusion category is like a fusion category except that rigidity is relaxed to weak rigidity as follows. A monoidal category $\mathcal{C}$ is weakly rigid if every object U has a weak dual: an object U^* such that $Hom(1, U \otimes W) \cong Hom(U^*, W)$ for any object W of $\mathcal{C}$.

A refined label set for a TQFT with a boundary condition category $\mathcal{C}$ is a finite set $L^e = \{\pm V_i\}_{i \in I}$, where the label set $L = \{V_i\}_{i \in I}$ is a set of representatives of isomorphism classes of simple objects of $\mathcal{C}$, and I a finite index set with a distinguished element 0 and $V_0 = 1$. An involution $\hat{}: L^e \longrightarrow L^e$ is defined on refined labels $l = \pm V_i$ by $\hat{l} = -l$ formally. There is also an involution on the index set I of the label set: $\hat{i} = j$ if $V_j \cong V_i^*$. A label $V_i \in L$ is self-dual if $\hat{i} = i$, and a refined label is self dual if the corresponding label is self dual. A (refined) label set is self-dual if every (refined) label is self-dual. Each label V_i has an FS indicator ν_l: 0 if not self-dual, and ± 1 if self-dual. A self-dual label V_i is symmetrically self-dual or real in conformal field theory language if $V_i^* = V_i$ in $\mathcal{C}$, then we say $\nu_i = 1$, and anti-symmetrically self-dual or pseudo-real if otherwise, then we say $\nu_i = -1$, i.e., V_i^* is not the same object V_i in $\mathcal{C}$, though they are isomorphic. Secretly $-V_i$ is V_i^*, and we will identify the label $-V_i$ with V_i if the label is symmetrically self-dual; but we cannot do so if the label is anti-symmetrically self-dual, e.g., in the Witten-Reshetikhin-Turaev $SU(2)$ TQFTs. Frobenius-Schur indicators are determined by the modular S and T matrices.[RSW] Note that the trivial label 1 is always symmetrically self-dual.

6.2. *Anomaly of TQFTs and extended manifolds*

In diagram TQFTs in Section 3, we see that $Z(X_1 \cup_{Y_2} X_2) = Z(X_1) \cdot Z(X_2)$ as composition of linear maps. For general TQFTs, this identity only holds up to a phase factor depending on X_1, X_2 and the gluing map. Moreover, for general TQFTs, the vector spaces $V(Y)$ for oriented surface Y are not defined canonically, but depend on extra structures under the names of 2-framing, Lagrange subspace, or p_1 structure, etc. A Lagrangian subspace of a surface Y is a maximal isotropic subspace of $H_1(Y; \mathbb{R})$ with respect to the intersection pairing of $H_1(Y; \mathbb{R})$. We choose to work with Lagrange subspaces to resolve the anomaly of a TQFT.

An extended surface Y is a pair (Y, λ), where λ is a Lagrangian subspace of $H_1(Y; \mathbb{R})$. Note that if $\partial X = Y$, then Y has a canonical Lagrange subspace $\lambda_X = ker(H_1(Y; \mathbb{R}) \longrightarrow H_1(X; \mathbb{R}))$. In the following, the boundary Y of a 3-manifold X is always extended by the canonical Lagrangian subspace λ_X unless stated otherwise. For any planar surface Y, $H_1(Y; \mathbb{R}) = 0$, so the extension is unique. Therefore, extended planar surfaces are just regular surfaces.

To resolve the anomaly for surfaces, we define a category of labeled extended surfaces. Given a boundary condition category $\mathcal{C}$, and a surface Y, a labeled extended surface is a triple $(Y; \lambda, l)$, where λ is a Lagrnagian subspace of $H_1(Y; \mathbb{R})$, and l is an assignment of a signed object $\pm U \in \mathcal{C}^0$ to each boundary circle. Moreover each boundary circle is oriented by the induced orientation from Y, and parameterized by an orientation preserving map from the standard circle S^1 in the plane.

Given two labeled extended surfaces $(Y_i; \lambda_i, l_i), i = 1, 2$, their disjoint union is the labeled extended surface $(Y_1 \coprod Y_2; \lambda_1 \oplus \lambda_2, l_1 \cup l_2)$. Gluing of surfaces has to be carefully defined to be compatible with the boundary structures and Lagrangian subspaces. Given two components γ_1 and γ_2 of ∂Y parameterized by ϕ_i and labeled by signed objects $\pm U$, and let gl be a diffeomorphism $\phi_2 \cdot r \cdot \phi_1^{-1}$, where r is the standard involution of the circle S^1. Then the glued surface Y_{gl} is the quotient space of $q : Y \to Y_{gl}$ given by $x \sim x'$ if $\mathrm{gl}(x) = \mathrm{gl}(x')$. If Y is extended by λ, then Y_{gl} is extended by $q_*(\lambda)$. The boundary surface ∂M_f of the mapping cylinder M_f of a diffeomorphism $f : Y \to Y$ of an extended surface $(Y; \lambda)$ has a canonical extension by the inclusions of λ.

Labeled diffeomorphisms between two labeled extended surfaces are orientation, boundary parameterization, and label preserving diffeomorphisms between the underlying surfaces. Note that we do not require the diffeomorphisms to preserve the Lagrangian subspaces.

6.3. *Axioms for TQFTs*

The category $\mathcal{X}^{2,e,l}$ of labeled extended surfaces is the category whose objects are labeled extended surfaces, and the morphism set of two labeled extended surfaces (Y_1, λ_1, l_1) and (Y_2, λ_2, l_2) are labeled diffeomorphisms.

The anomaly of a TQFT is a root of unity κ, and to match physical convention, we write $\kappa = e^{\pi i c/4}$, and $c \in \mathbb{Q}$ is well-defined mod 8, and called the central charge of a TQFT. Therefore, a TQFT is anomaly free if and only if the central charge c is 0 mod 8.

Definition 6.1.

A $(2+1)$-TQFT with a boundary condition category $\mathcal{C}$, a refined label set L^e, and anomaly κ consists of a pair (V, Z), where V is a functor from the category $\mathcal{X}^{2,e,l}$ of oriented labeled extended surfaces to the category $\mathcal{V}$ of finitely dimensional vector spaces and linear isomorphisms composed up to powers of κ, and Z is an assignment for each oriented 3-manifold X with extended boundary, $Z(X, \lambda) \in V(\partial X; \lambda)$, where ∂X is extended by a Lagrangian subspace λ. We will use the notation $Z(X), V(\partial X)$ if ∂X is extended by the canonical Lagrangian subspace λ_X. V is called a modular functor. Z is the partition function if X is closed in physical language, and we will call Z the partition function even when X is not closed.

Furthermore, V and Z satisfy the following axioms.

Axioms for V:

(1) Empty surface axiom:

$V(\emptyset) = \mathbb{C}$

(2) Disk axiom:

$V(B^2; l) \cong \begin{cases} \mathbb{C} & \text{if } l \text{ is the trivial label} \\ 0 & \text{otherwise} \end{cases}$, where B^2 is a 2-disk.

(3) Annular axiom:

$V(\mathcal{A}; a, b) \cong \begin{cases} \mathbb{C} & \text{if } a = \hat{b} \\ 0 & \text{otherwise} \end{cases}$, where $\mathcal{A}$ is an annulus, and $a, b \in L^e$ are refined labels.

(4) Disjoint union axiom:

$V(Y_1 \amalg Y_2; \lambda_1 \oplus \lambda_2, l_1 \coprod l_2) \cong V(Y_1; \lambda_1, l_1) \otimes V(Y_2; \lambda_2, l_2)$. The isomorphisms are associative, and compatible with the mapping class group actions.

(5) Duality axiom:

$V(-Y; l) \cong V(Y; \hat{l})^*$.

The isomorphisms are compatible with mapping class group actions, with orientation reversal and disjoint union axiom as follows:
a) The isomorphisms $V(Y) \to V(-Y)^*$ and $V(-Y) \to V(Y)^*$ are mutually adjoint.
b) Given $f : (Y_1; l_1) \to (Y_2; l_2)$ and let $\bar{f} : (-Y_1; \hat{l_1}) \to (-Y_2; \hat{l_2})$, then $< x, y >=< V(f)x, V(\bar{f})y >$, where $x \in V(Y_1; l_1), y \in V(-Y_1; \hat{l_1})$.
c) Let $\alpha_1 \otimes \alpha_2 \in V(Y_1 \coprod Y_2) = V(Y_1) \otimes V(Y_2)$, and $\beta_1 \otimes \beta_2 \in V(-Y_1 \coprod -Y_2) = V(-Y_1) \otimes V(-Y_2)$, then

$$< \alpha_1 \otimes \alpha_2, \beta_1 \otimes \beta_2 >=< \alpha_1, \beta_1 >< \alpha_2, \beta_2 > .$$

(6) Gluing Axiom:

Let Y_{gl} be the surface obtained from gluing two boundary components of Y, then $V(Y_{gl}) \cong \oplus_{l \in L} V(Y; (l, \hat{l}))$, where $l, \hat{l}$ label the two glued boundary components. The isomorphism is associative and compatible with mapping class group actions.

Moreover, the isomorphism is compatible with duality as follows: let $\oplus_{i \in L} \alpha_i \in V(Y_{gl}; l) = \oplus_{i \in L} V(Y; l, (i, \hat{i}))$ and $\oplus_i \beta_i \in V(-Y_{gl}; \hat{l}) = \oplus_{i \in L} V(-Y; \hat{l}, (i, \hat{i}))$, then there are non-zero real numbers s_i for each label V_i such that

$$< \oplus_i \alpha_i, \oplus_i \beta_i >= \sum_i s_i < \alpha_i, \beta_i > .$$

Axioms for Z:

(1) Disjoint axiom:

If $X = X_1 \amalg X_2$, then $Z(X) = Z(X_1) \otimes Z(X_2)$.

(2) Naturality axiom:

If $f : (X_1, (\partial X_1, \lambda_1)) \longrightarrow (X_2, (\partial X_2, \lambda_2))$ is a diffeomorphism, then $V(f) : V(\partial X_1) \longrightarrow V(\partial X_2)$ sends $Z(X_1, \lambda_1)$ to $Z(X_2, \lambda_2)$.

(3) Gluing axiom:

If $\partial X_1 = -Y_1 \amalg Y_2, \partial X_2 = -Y_2 \amalg Y_3$, then $Z(X_1 \cup_{Y_2} X_2) = \kappa^n Z(X_1) Z(X_2)$, where $n = \mu((\lambda_- X_1), \lambda_2, (\lambda_+ X_2))$ is the Maslov index (see Appendix C).

More generally, if X is an oriented 3-manifold and let $Y_i, i = 1, 2$ be disjoint surfaces in ∂X, extended by $\lambda_i \subset \lambda_X, i = 1, 2$, and $f : Y_1 \to Y_2$ be an orientation reversing dffeomorphism sending λ_1 to λ_2.

Then $V(\partial X)$ is isomorphic to $\sum_{l_1, l_2} V(Y_1; l_1) \otimes V(Y_2; l_2) \otimes V(\partial X \backslash (Y_1 \cup Y_2); (\hat{l_1}, \hat{l_2}))$ by multiplying κ^m, where l_i runs through all labelings of Y_i, and $m = \mu(K, \lambda_1 \oplus \lambda_2, \Delta)$ (see Appendix C). Hence $Z(X) = \oplus_{l_1, l_2} \kappa^m \sum_j \alpha^j_{l_1} \otimes \beta^j_{l_2} \otimes \gamma^j_{\hat{l_1}, \hat{l_2}}$.

If gluing Y_1 to Y_2 by f results in the manifold X_f, then

$$Z(X_f) = \kappa^m \sum_{j,l} < V(f) \alpha^j_l, \beta^j_l > \gamma^j_{\hat{l}, l}.$$

(4) Mapping cylinder axiom:

If Y is closed and extended by λ, and $Y \times I$ is extended canonically by $\lambda \oplus (-\lambda)$. Then $Z(Y \times I, \lambda \oplus (-\lambda)) = \mathrm{id}_{V(Y)}$.

More generally, let I_{id} be the mapping cylinder of $\mathrm{id} : Y \to Y$, and id_l be the identity in $V(Y;l) \otimes V(Y;l)^*$, then

$$Z(\mathrm{I}_{\mathrm{id}}, \lambda \oplus (-\lambda)) = \oplus_{l \in L(Y)} \mathrm{id}_l.$$

□

First we derive some easy consequences of the axioms:

Prop 6.1.

(1) $V(S^2) \cong \mathbb{C}$
(2) $Z(X_1 \sharp X_2) = \frac{Z(X_1) \otimes Z(X_2)}{Z(S^3)}$.
(3) *Trace formula:* Let X be a bordism from closed surfaces Y, extended by λ, to itself, and X_f be the closed 3-manifold obtained by gluing Y to itself with a diffeomorphism f.
Then $Z(X_f) = \kappa^m Tr_{V(Y)}(V(f))$, where $m = \mu(\lambda(f), \lambda_Y \oplus f_*(\lambda), \Delta_Y)$ and $\lambda(f)$ is the graph of f_*, Δ_Y is the diagonal of $H_1(-Y;\mathbb{R}) \oplus H_1(Y;\mathbb{R})$. In particular, $Z(Y \times S^1) = \dim(V(Y))$.
(4) The dimension of $V(T^2)$ is the number of particle types.

For a TQFT with anomaly, the representations of the mapping class groups are projective in a very special way. From the axioms, we deduce:

Prop 6.2.
The representations of the mapping class groups are given by the mapping cylinder construction: given a diffeomorphsim $f : Y \longrightarrow Y$ and Y extended by λ, the mapping cylinder Y_f induces a map $V(f) = Z(Y_f) : V(Y) \longrightarrow V(Y)$. We have $V(fg) = \kappa^{\mu(g_*(\lambda), \lambda, f_*^{-1}(\lambda))} V(f)V(g)$.

It follows from this proposition that the anomaly can be incorporated by an extension of the bordisms X, in particular, modular functors yield linear representations of certain central extensions of the mapping class groups.

6.4. *More consequences of the axioms*

For refined labels a, b, c, we have vector spaces $V_a = V(B^2; a), V_{a,b} = V(A_{ab}), V_{a,b,c} = V(P_{abc})$, where P is a pair of pants or three-punctured sphere. Denote the standard orientation reversing maps on $B^2, \mathcal{A}_{ab}, P_{abc}$ by ψ. Then $\psi^2 = id$, therefore ψ induces identifications $V_{abc} = V^*_{\hat{a}\hat{b}\hat{c}}, V_{a\hat{a}} = V^*_{a\hat{a}}$, and $V_1 = V_1^*$. Choose basis $\beta_1 \in V_1, \beta_{a\hat{a}} \in V_{a\hat{a}}$ such that $< \beta_a, \beta_a >= \frac{1}{d_a}$.

Prop 6.3.

(1) $Z(B^2 \times I) = \beta_1 \otimes \beta_1$
(2) $Z(S^1 \times B^2) = \beta_{11}$
(3) $Z(X \backslash B^3) = \frac{1}{D} Z(X) \otimes \beta_1 \otimes \beta_1$.

Proof.

Let B^3 be a 3-ball regarded as the mapping cylinder as the identity map $id : B^2 \longrightarrow B^2$. By the mapping cylinder axiom, $Z(B^3) = \beta_1 \otimes \beta_1$. Gluing two copies of B^3 together yields S^3. By the gluing axiom $Z(S^3) = s_{00} = \frac{1}{D}$. It follows that $Z(X \backslash B^3) = \frac{1}{D} Z(X) \otimes \beta_1 \otimes \beta_1$. □

Prop 6.4.

The action of the left-handed Dehn twist along a boundary component labeled by a of $B^2, \mathcal{A}_{ab}, P_{abc}$ on $V_1, V_{a,\hat{a}}$ or V_{abc} is a multiplication by a scalar θ_a. Furthermore, $\theta_1 = 1, \theta_a = \theta_{\hat{a}}$, and θ_a is a root of unity for each refined label a.

6.5. *Framed link invariants and modular representation*

Let K be a framed link in a 3-manifold X. The framing of K determines a decomposition of the boundary tori of the link compliment $X \backslash \text{nbd}(K)$ into annuli. With respect to this decomposition,

$$Z(X \backslash \text{nbd}(K)) = \oplus_l J(K; l) \beta_{a_1 \hat{a_1}} \otimes \cdots \otimes \beta_{a_n \hat{a_n}},$$

where $J(k; l) \in \mathbb{C}$ and $l = (a_1, \cdots, a_n)$ ranges over all labelings of the components of K. $J(K; l)$ is an invariant of the framed, labeled link $(K; l)$. When (V, Z) is a Jones-Kauffman or WRT TQFT, and $X = S^3$, the resulting link invariant is a version of the celebrated colored Jones polynomial evaluated at a root of unity. This invariant can be extended to an invariant of labeled, framed graphs.

A framed link K represents a 3-manifold $\chi(K)$ via surgery. Using the gluing formula for Z, we can express $Z(\chi(K))$ as a linear combination of $J(K; l)$:

$$Z(\chi(K)) = \sum_l c_l J(K; l).$$

Consider the Hopf link H_{ij} labeled by $i, j \in L$. Let $\tilde{s}_{ij}$ be the link invariant of H_{ij}. Note that when a component is labeled by the trivial label, then we may drop the component from the link when we compute link invariant. Therefore, the first row of $\tilde{s}$ consists of invariants of the unknot labeled by $i \in L$. Denote $\tilde{s}_{i0}$ as d_i, and d_i is called the quantum dimension of label i. In Prop. 6.4, each label is associated with a root of unity θ_i, which will be called the twist of label i. Define $D = \sqrt{\sum_{i \in L} d_i^2}$, and $S = \frac{1}{D}\tilde{s}, T = (\delta_{ij}\theta_i)$, then S, T give rise to a representation of $SL(2, \mathbb{Z})$, the mapping class group of T^2.

6.6. *Verlinde algebras and Verlinde formulas*

Let $T^2 = S^1 \times S^1 = \partial D^2 \times S^1$ be the standard torus. Define the meridian to be the curve $\mu = S^1 \times 1$ and the longitude to be the curve $\lambda = 1 \times S^1$.

Let (V, Z) be a TQFT, then the Verlinde algebra of (V, Z) is the vector space $V(T^2)$ with a multiplication defined as follows: consider the two decompositions of T^2 into annuli by splitting along μ and λ, respectively. These two decompositions determine two bases of $V(T^2)$ denoted as $m_a = \beta_{a\hat{a}}$, and $l_a = \beta_{\hat{a}a}$. These two bases are related by the modular S-matrix as follows:

$$l_a = \sum_b s_{ab} m_b, m_a = \sum_b s_{\hat{a}b} l_b. \tag{6.1}$$

Define $N_{abc} = \dim V(P_{abc})$, then we have

$$m_b m_c = \sum_a N_{a\hat{b}\hat{c}} m_a. \tag{6.2}$$

The multiplication makes $V(T^2)$ into an algebra, which is called the Verlinde algebra of (V, Z).

In the longitude bases l_a, the multiplication becomes

$$l_a l_b = \delta_{ab} s_{0a}^{-1} l_a. \tag{6.3}$$

This multiplication also has an intrinsic topological definition: $Z(P \times S^1)$ gives rise to a linear map from $V(T^2) \times V(T^2) \to V(T^2)$ by regarding $P \times S^1$ as a bordism from $T^2 \coprod T^2$ to T^2.

The fusion coefficient N_{abc} can be expressed in terms of s_{ab}, we have

$$N_{abc} = \sum_{x \in L} \frac{s_{ax} s_{bx} s_{cx}}{s_{0x}}. \tag{6.4}$$

More generally, for a genus=g surface Y with m boundaries labeled by $l = (a_1 \cdots a_m)$,

$$\dim V(Y) = \sum_{x \in L} s_{0x}^{2-2g-n} (\prod s_{a_i x}). \tag{6.5}$$

7. Diagram and Jones-Kauffman TQFTs

For the remaining part of the paper, we will construct picture TQFTs and verify the axioms for those TQFTs. Our approach is as follows: start with a local relation and a skein relation, we first define a picture category Λ whose objects are points with decorations in a 1-manifold X which is either an interval I or a circle S^1, and morphisms are unoriented sub-1-manifolds in $X \times I$ with certain structures connecting objects (=points in $X \times \{0\}$ or $X \times \{1\}$). More generally, the morphisms can be labeled trivalent graphs with coupons. Those picture categories serve as crude boundary conditions for defining picture spaces for surfaces with boundaries. Secondly, we find the representation category $\mathcal{C}$ of Λ, which is a spherical tensor category. The irreps will be the labels. In the cases that we are interested, the resulting spherical categories are all ribbon tensor categories. Thirdly, we define colored framed link invariants with the resulting ribbon tensor category in the second step. Invariants of the colored Hopf links with labels form the so-called modular S-matrix. Each row of the S-matrix can be used to define a projector ω_i which projects out the i-th label if a labeled strand goes through a trivial circle labeled by ω_i.

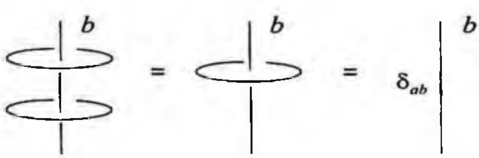


Fig. 7.1. Projectors

The projector ω_0 is used to construct the resulting 3-manifold invariant. Finally, we define the partition function Z for a bordism X using a handle decomposition. This construction will yield a TQFT if the S-matrix is nonsingular, which is always true for the annular TLJ cases. If the S-matrix is singular, we still have a 3-manifold invariant, but we cannot define the representations of the mapping class groups for high genus surfaces, though representations of the braid groups are still well defined.

7.1. *Diagram TQFTs*

In this section, we outline the proof that for some $r \geq 3$, A a primitive $4r$th root of unity, or a primitive $2r$th root of unity and r odd, or a primitive rth root of unity and r odd, the diagram theories $\mathrm{Pic}^A(Y), Z_D$ defined in Section 3 indeed satisfy the axioms of TQFTs.

The diagram TQFTs are constructed based on the TLJ annular categories. The boundary condition categories $\mathcal{C}$ are the representation categories of the TLJ annular categories Λ. A nice feature of those TQFTs is that we can identify the objects of the TLJ annular categories Λ as boundary conditions using Theorem B.1: each object in TLJ gives rise to a representation of Λ and therefore becomes an object of $\mathcal{C}$, which is in general not simple, i.e., not a label. Hence picture vector spaces are naturally vector spaces for the diagram TQFTs.

For the diagram TQFTs, all labels are self-dual with trivial FS indicators. Therefore, it suffices to use only the label set. The label sets of the diagram TQFTs are given by the idempotents $L = \{\omega_{i,j,h}\}$ in Fig. 3.7. Given a surface Y with $\partial Y = \gamma_1, \cdots, \gamma_m$, and each boundary circle γ_i labeled by an idempotent $e_i \in L$. Then the picture space $\mathrm{Pic}_D^A(Y; e_1, \cdots, e_m)$ consists of all formal pictures that agree with e_i inside a small annular neighborhood $\mathcal{A}_i$ of the boundary γ_i modulo the Jones-Wenzl projector p_{r-1} outside all $\mathcal{A}_i$'s in Y. Given a bordism X from Y_1 to Y_2, the partition function $Z_D(X)$ is defined in Section 3.6. Now we verify that (Pic^A, Z_D) is indeed a TQFT.

For the axioms for modular functor V:

(1) is obvious.

(2) Since Jones-Wenzl projectors kill any turn-backs, then $\mathrm{Pic}^A(B^2; \omega_{i,j,h}) = 0$ unless $h = 0$. For $h = 0$, all pictures are multiples of the empty diagram.

(3) Since $\mathrm{Hom}(p_i, p_j) = 0$ unless $i = j$, so $\mathrm{Pic}^A(\mathcal{A}; \omega_{i',j',h'}, \omega_{i,j,h}) = 0$ unless $h = h'$. If $h = h'$, then we have $\omega_i \cdot \omega_{i'}$ and $\omega_j \cdot \omega_{j'}$, respectively in the annulus. Recall that $\omega_a \omega_b = \delta_{ab} \omega_a$, it follows that unless $i = i', j = j'$, $\mathrm{Pic}^A(\mathcal{A}; \omega_{i',j',h'}, \omega_{i,j,h}) = 0$.

(4) Obvious

(5) $\mathrm{Pic}^A(-Y) = \mathrm{Pic}^A(Y)$, hence duality is obvious.

(6) Gluing follows from Morita equivalence.

The axioms of partition function Z follow from handle-body theory and properties of the S matrix.

The action of the mapping class groups is easy to see: a diffeomorphism maps one multicurve to another. Since a diffeomorphism preserves the local relation and skein relation, this action sends skein classes to skein classes.

The compatibility of the action with the axioms for vectors spaces is easy to check.

7.2. *Jones-Kauffman TQFTs*

In this section, we outline the proof that for $r \geq 3$, A a primitive 4rth root of unity, the Jones-Kauffman skein theories $V_{JK}^{A}(Y), Z_{JK}$ defined in Section 3 indeed satisfy the axioms of TQFTs.

The boundary condition category for a Jones-Kauffman TQFT is the representation category of a TLJ rectangular category. The label set is $L = \{p_i\}_{i \in I}$, and $I = \{0, 1, \cdots, r-2\}$. Same reason as for the diagram TQFTs, we need only the label set.

The new feature of the Jones-Kauffman TQFTs is the framing anomaly. If A and r as in Lemma 3.1, then the central charge is $\frac{3(r-2)}{r}$.

Given an extended surface $(Y; \lambda)$, the modular functor $V(Y; \lambda)$ is defined in Section 3.4. If $\partial X = Y$, then we define $Z(X)$ as the skein class in $K_A(\partial X)$ represented by the empty skein. TQFT axioms for V and Z follow from theorems in Section 3.4. The non-trivial part is the mapping class group action. This is explained at the end of Section 3.4.

8. WRT and Turaev-Viro $SU(2)$-TQFTs

The pictorial approach to the Witten-Reshetikhin-Turaev $SU(2)$ TQFTs was based on.[KM] The paper[KM] finished with 3-manifold invariants, just as[KL] for the Jones-Kauffman theories. The paper[BHMV] took the picture approach in[KL] one step further to TQFTs, but the same for WRT TQFTs has not been done using a pictorial approach. The reasons might be either people believe that this has been done by[BHMV] or realize that the Frobennius-Scur indicators make a picture approach more involved. It is also widely believed that the two approaches resulted in the same theories. But they are different. The spin 1/2 representation of quantum group $SU(2)_q$ for $q = e^{\pm 2\pi i/r}$ has a Frobenius-Schur indicator$=-1$, whereas the corresponding label 1 in Temperley-Lieb-Jones theories has Frobenius-Schur indicator $= 1$. The Frobenus-Schur indicators -1 in the Witten-Reshetikhin-Turaev theories introduce some -1's into the S-matrix, hence for the odd levels k, these -1's change the S-matrix from singular in the Jones-Kauffman theories when $A = \pm i e^{\pm \frac{2\pi i}{4r}}$ to non-singular. For even levels k, the S-matrices are the same as those of the Jones-Kauffman TQFTs, even though the TQFTs are different theories (see[RSW] for the level=2 case).

In the pictorial TLJ approach to TQFTs, there is no room to encode the Frobenius-Schur indicators -1. In this section, we introduce "flag" decorations on each component of a framed multicurve which can point to either side of the component. These flags allow us to encode the FS indicator -1, hence reproduce the Witten-Reshetikhin-Turaev $SU(2)$ TQFTs exactly. The doubled theories of WRT TQFTs are not the diagram TQFTs, and will be called the Turaev-Viro $SU(2)$-TQFTs. They are direct products of WRT theories with their mirror theories.

8.1. *Flagged TLJ categories*

In flagged TLJ categories, the local relation is still the Jones-Wenzl projectors, but the skein relation is not the Kauffman bracket exactly, but a slight variation discovered by R. Kirby and P. Melvin in.[KM]

The skein relation for resolving a crossing p is given in[KM] is as follows: if the two strands of the crossing belongs to two different components of the link, then the resolution is the Kauffman bracket in Figure 2.1; but if the two strands of the crossing p are from the same component, then a sign $\epsilon(p) = \pm 1$ is well-defined, and the skein relation is:

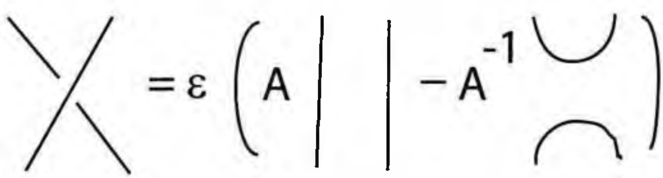


Fig. 8.1. Kirby-Melvin skein relation

The flagged TLJ categories have objects signed points in the interval and morphism flagged multicurves as follows: given an oriented surface Y, and a multicurve γ in the interior of Y, and no critical points of γ are within small neighborhoods of ∂Y. Let $\gamma \times [-\epsilon, \epsilon]$ be a small annulus neighborhood of γ. A flag of γ at $p \in \gamma$ is an arc $p \times [0, \epsilon]$ or $p \times [-\epsilon, 0]$. A flag is admissible if p is not a critical point of γ. A multicurve γ is flagged if all flags on γ are admissible and the number of flags has the same parity as the number of critical points of γ. An admissible flag on γ can be parallel transported on γ so that when the flag passes through a critical point, it flips to the other side. In the plane, this is the same as parallel transport by keeping the flag parallel at all times in the plane. A multicurve is flagged if all its components are flagged.

Given a surface Y with signed points on the boundary. Each signed point is flagged so that if the sign is $+$, the flag agrees with the induced orientation of the boundary; if the sign is $-$, the flag is opposite to the induced orientation. Let $\mathbb{C}[\mathcal{S}]$ be the space of all formal flagged multicurves in $\mathcal{R}$ with signed points at the bottom and top, then the morphism set between the bottom signed point and the top signed point of $\mathrm{TL}^{\mathrm{flag}}$ is the quotient space of $\mathbb{C}[\mathcal{S}]$ such that

(1) Flags can be parallel transported
(2) Flipping a flag to the other side results in a minus sign
(3) Two neighboring flags can be cancelled if there are no critical points between them and they are on the opposite sides.
(4) Apply Jones-Wenzl projector to any part of an multicurve with no flags.

Then all discussions for TL apply to $\mathrm{TL}^{\mathrm{flag}}$. The representation category is similarly given by the same Jones-Wenzl projectors. The biggest difference from TL is the resulting framed link invariant.

Lemma 8.1.
Given a framed link diagram D, then the WRT invariant $< D >_{KM}$ of D using Kirby-Melvin skein relation and the Jones-Kauffman invariant $< D >_K$ using Kauffman bracket is related by:

$$< D >_{KM} (A) = (-i)^{D \cdot D} < D >_K (iA). \tag{8.1}$$

8.2. *Turaev-Viro Unitary TQFTs*

Fix $A = \pm e^{\pm \frac{2\pi i}{4r}}$ for some $r \geq 3$.

The label set is the same as that of the corresponding diagram TQFT, but for the first time we need to work with the refined label set.

Given a surface Y with boundaries labeled by refined labels $\epsilon_i V_i$. If $\epsilon_i = 1$, we flag the point to the orientation of ∂Y; if $\epsilon_i = -1$, we flag the points opposite to the orientation of ∂Y. Then define the modular functor space analogous to the skein space replacing multicurves with flagged multicurves. The theories are similar enough so we will leave the details to interested readers. The difference is that when the level=k is odd, our version of the Turaev-Viro theory is a direct product-a trivial quantum double, while the corresponding diagram TQFT is a non-trivial quantum double.

8.3. *WRT Unitary TQFTs*

Fix $A = \pm e^{\pm \frac{2\pi i}{4r}}$ for some $r \geq 3$.

The label set of a WRT TQFT is the same as that of the corresponding Jones-Kauffman TQFT, but it needs to be extended to the refined label set. The central charge of a level=k theory is $\frac{3k}{k+2}$. The discussion together with the Turaev-Viro theories is completely parallel to the Jones-Kauffman TQFTs with diagram TQFTs.

9. Black-White TQFTs

Interesting variations of the TLJ categories can also be obtained by 2-colorings: the black-white annular categories $\mathrm{TLJ}_d^{\mathrm{BW}}$. The objects of the category are the objects of the corresponding annular TLJ category enhanced by two colorings of the complements of the points. In particular there are two circles: black and white. Morphisms between two objects are enhanced by colorings of the regions. A priori there are two enhancements of each Jones-Wenzl idempotent, but it has been proved in[Fn] that the two versions are equivalent.

9.1. *Black-white TLJ categories*

Fix some $r \geq 3$ and A, where A is a primitve $4r$th root of unity, or a primitive $2r$th root of unity and r odd, or a primitive rth root of unity and r odd

The objects of black-white TLJ categories are points in the interval or S^1 with a particular 2-coloring of the complementary intervals so that adjacent intervals having different colors. Given two objects, morphisms are multicurves from the bottom to top whose complement regions have black-white colors that are compatible with the objects, and any two neighboring regions receive different colors. The local relation is the 2-color enhanced Jones-Wenzl projector and the skein relation is the 2-color enhancement of the Kauffman bracket. The representation theories of the black-white categories are considerably harder to analyze.

The object with no points in the circle has two versions $0_B, 0_W$, which might be isomorphic. Indeed sometimes they are isomorphic and sometimes not. Therefore a skeleton of a black-white TLJ category can be identified with $\{0_B, 0_W, 2, 4, \cdots\}$ with the possibility that $0_B = 0_W$. We will draw the black object 0_b as a bold solid circle, and the white object as a dotted circle. Interface circles between black and white regions will be drawn as regular solid circles. Morphisms will be drawn inside annuli, directed and composed from inside-out. There are two color changing morphisms $r_{bw} \in \mathrm{Hom}(0_b, 0_w), r_{wb} \in \mathrm{Hom}(0_w, 0_b)$.

Let us denote the two compositions $r_{bw} \cdot r_{wb} = x_b \in \mathrm{Hom}(0_b, 0_b), r_{wb} \cdot r_{wb} = x_w \in \mathrm{Hom}(0_w, 0_w)$, which are just rings in the annulus.

Given an oriented closed surface Y, a 2-colored multicurve in Y is a pair (γ, c), where γ is a multicurve, and c is an assignment of black or white to all regions of $Y \backslash \gamma$ so that any two neighboring regions have opposite colors. Let $\mathbb{C}[\mathcal{S}]$ be the vector space of formal 2-colored multicurves, and $\mathrm{Pic}^{\mathrm{BW}}(Y)$ be the quotient space of $\mathbb{C}[\mathcal{S}]$ modulo the BW-enhancement of JW projectors.

9.2. *Labels for black-white theories*

Recall in Section 5.1, we define the element $q_{2m} \in C_{2m}(x)$.

Lemma 9.1.

The element q_{2m} is a minimal idempotent of $C_{2m}(x)$.

Prop 9.1.

(1): If r is even, then x_b, x_w are not invertible, hence 0_b is not isomorphic to 0_w.

(2): If r is odd, then x_b, x_w are invertible, hence 0_b and 0_w are isomorphic.

(3): The color swap involution is the identity on the TQFT vector spaces.

9.2.1. *Level=2, $d^2 = 2$*

The algebra $\mathrm{A}_{0_b 0_b} \cong \mathbb{C}^2$, and so is the algebra $\mathrm{A}_{0_w 0_w}$. Both are generated by x, so $\cong \mathbb{C}[x]/(x^2 = 2x)$.

$\mathrm{Hom}(0_b, 0_w) \cong \mathbb{C}$ is generated by r_{bw}. Similarly, $\mathrm{Hom}(0_w, 0_b) \cong \mathbb{C}$ is generated by r_{wb}.

$A_{2,2} \cong \mathbb{C}^4$. Following the same analysis as in Section 5, we get the irreps denoted by the following:

ρ_3	0	0	1
ρ_2	1	0	1
ρ_1	0	1	1
ρ_0	1	1	1
	0_b	0_w	2

9.2.2. *Level=3*

The algebra $A_{0_B, 0_B} \cong \mathbb{C}^2$ is generated by x so $\cong \mathbb{C}[x]/(x^2 - 3x + 1)$.

The algebra $A_{22} \cong \mathbb{C}^7$. Similar analysis as above leads to:

ρ_3	0	0	1
ρ_2	0	0	1
ρ_1	1	1	2
ρ_0	1	1	1
	0_b	0_w	2

9.3. *BW TQFTs*

Theorem 9.1.

(1): If $r \geq 3$, and A a primitive $4r$th root of unity, or a primitive $2r$th root of unity and r odd, or a primitive rth root of unity and r odd, then (V^A_{BW}, Z^A_{BW}) is a TQFT.

(2): If r odd, then (V^A_{BW}, Z^A_{BW}) is isomorphic to the doubled even TLJ sub-category TQFT, i.e., the TQFT from the quantum double of the even TLJ subcategory at the corresponding A.

The proof of this theorem and the irreps for all r are left to a future publication.

We have not be able to identify the BW TQFTs with known ones when r is even, and A is a primitive $4r$th root of unity. If $r = 4$, then (V^A_{BW}, Z^A_{BW}) is isomorphic to the toric code TQFT. We conjecture Theorem 9.1 (2) still holds for these cases. Furthermore, each (V^A_{BW}, Z^A_{BW}) decomposes into a direct product of the toric code TQFT with another TQFT.

10. Classification and Unitarity

In this section, we classify all TQFTs based on Jones-Wenzl projectors and Kauffman brackets. Then we decide when the resulting TQFT is unitary. In literature A has been chosen to be either as a primitive $4r$-th root of unity or as a primitive $2r$th root of unity. We notice that for r odd, when A is a primitive rth root of unity, the resulting TLJ rectangular categories give rise to ribbon tensor categories with singular S-matrices, but their annular versions lead to TQFTs which are potentially new. Also when A is a primitive $4r$th root of unity and r even, the BW TQFTs seem to contain new theories.

10.1. *Classification of diagram local relations*

By d generic we mean that d is not a root of some Chebyshev polynomial Δ_i. Equivalently $d \neq B + \bar{B}$ for some B such that $B^\ell = 1$.

Let us consider d-isotopy classes of multicurves on a closed surface Y. Call this vector space $\mathrm{TL}_d(Y)$. This vector space has the subtle structure of gluing formula associated to cutting into subsurfaces (and then regluing); there is a product analogous to both times and tensor products in TLJ_d. Also for special values of d $\mathrm{TL}_d(Y)$ has a natural singular Hermitian structure.

Theorem 10.1. *If d has the form: $d = -A^2 - A^{-2}$, A a root of unity. Then there is a (single) local relation $R(d)$ so that $TL_d(Y)$ modulo $R(d)$, denoted by $V_d(Y)$, have finite nonzero dimension. If d is not of the above form then $V_d(Y) = 0$ or $= TL_d(Y)$ for any given $R(d)$. Furthermore the quotient space $V_d(Y)$ of $TL_d(Y)$ when it is neither $\{0\}$ nor $TL_d(Y)$ is uniquely determined, and when A is a primitive 4rth root of unity, then $V_d(Y)$ is the "Drinfeld double" of a Jones-Kauffman TQFT at level $k = r - 2$.*

Proof. Consider a local relation $R_0(d)$ of smallest degree, say $2n$, which holds in TL_d (i.e. is a consequence of $R(d)$). Arbitrarily draw $R_0(d)$ in a rectangle with n endpoints assigned to the top and n endpoints assigned to the bottom, to place $R_0(d)$ in the algebra $\mathrm{TL}_n(d)$. Adding any cup or cap to R_0 gives a consequent relation of degree $= 2n - 2$; this relation, by minimality, must be zero. This implies that $e_i R_0(d) = R_0(d) e_i = 0$ for $1 \leq i \leq n-1$. So by Proposition 2.1, $R_0(d) = c p_{n,d}$, c a nonzero scalar.

The trace, $\mathrm{tr}(p_{n,d}) \in \mathbb{C}$ is a degree $= 0$ consequence of $p_{n,d}$ so unless d is a root of Δ_n, $\mathrm{tr}(p_{n,d}) \neq 0$ and so generates all relations: $p_{n,d}(Y) = 0$.

Now suppose d is the root of two Chebyshev polynomials Δ_m and Δ_ℓ, $m < \ell$. This happens exactly when $(m+1)$ divides $(\ell + 1)$. In fact to understand the roots of $\Delta_n(d)$ introduces a change of variables $d = B + B^{-1}$, then: $\Delta_n(d) = (B^{n+1} - B^{-n-1})/(B - B^{-1})$. The r.h.s. vanishes (simply) when (and only when) B is a $2n+2-$ root of unity $\neq \pm 1$. In particular if d is a root of Δ_m and Δ_ℓ then $p_{m,d}$ is a consequence of $p_{m,\ell}$ by "partial trace" as shown in Figure 10.1.

Trace both sides to verify the coefficient $f = \frac{\Delta_m}{\Delta_\ell}$ and note that since both numerator and denominator have simple roots at d the coefficient at d is well defined and nonzero.

Thus for a diagram local relation (or set thereof) to yield a nontrivial set of quotient space $\neq 0$, d must be a root of lowest degree Δ_n and the

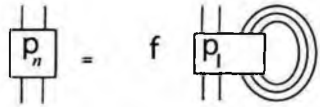


Fig. 10.1. Partial trace

relation(s) are equivalent to the single relation $p_{n,d}$.

Geometrically, $0 = p_n = 1 + U$ implies $1 = -U$ means that the multicurves whose multiplicity is less than n along the $1-$ cells, $SK^1(Y)$, ("bonds" in physical language) of any fixed triangulation of Y determine $p_n(Y)$. Here multiplicity $\leq n$ for a multicurve γ means that γ runs near $SK^1(Y)$ and with fewer than n parallel copies of a $1-$ cell (bond) of $SK^1(Y)$. Finite dimensionality of $p_n(Y)$ is an immediate consequence. (SK^1 stands for "1−skeleton.")

The quotient space $\neq 0$, for $Y = S^2$ this follows from the nonvanishing of certain θ-symbols; for Y of higher genus the Verlinde formulas. □

10.2. *Unitary TQFTs*

A unitary modular functor is a modular functor such that each $V(Y)$ is endowed with a non-degenerate Hermitian pairing:

$$<,>: \overline{V(Y)} \times V(Y) \longrightarrow \mathbb{C},$$

and each morphism is unitary. The Hermitian structures are required to satisfy compatibility conditions as in the naturality axiom of a modular functor. In particular,

$$< \oplus_i v_i, \oplus_j w_j >= \sum_i s_{i0} < v_i, w_j > .$$

Note that this implies that all quantum dimensions of particles are positive reals. It might be true that any theory with all quantum dimensions positive is actually unitary. Moreover, the following diagram commutes for all Y:

$$\begin{tikzcd} V(Y) \arrow[r, "\cong"] \arrow[d, "\cong"'] & V(-Y)^* \arrow[d, "\cong"] \\ \overline{V(Y)^*} \arrow[r, "\cong"] & \overline{V(-Y)} \end{tikzcd}$$

A unitary TQFT is a TQFT whose modular functor is unitary and whose partition function satisfies $Z(-M) = \overline{Z(M)}$.

10.3. *Classification and unitarity*

There are two kinds of TQFTs that we studied in this paper: undoubled and doubled, which are indexed by the Kauffman variable A.

When A is a primitive $4r$th root of unity for $r \geq 3$, we have the Jones-Kauffman TQFTs. The even sub-categories of TLJs yield TQFTs for r odd, but have singular S-matrix if r even. If r is even, and $A = \pm ie^{\pm\frac{2\pi i}{4r}}$, then the Jones-Kauffman TQFTs are unitary.

We also have the WRT $SU(2)$-TQFTs for $q = e^{\pm\frac{2\pi i}{r}}$, which are unitary. WRT TQFTs were believed to be the same as the Jones-Kauffman TQFTs with $q = A^{\pm 4}$, but they are actually different. Jones-Kauffman TQFTs and WRT TQFTs are related by a version of Schur-Weyl duality as alluded in Section 2 for the braid group representations.

All above theories can be doubled to get picture TQFTs: the doubled Jones-Kauffman TQFTs are the diagram TQFTs, while the doubled WRT TQFTs are the Turaev-Viro TQFTs.[TV] The doubles of even sub-categories for r odd form part of the Black-White TQFTs in Theorem 9.1, while for r even this is still a conjecture.

When A is a primitive $2r$th or rth root of unity and r odd, the TLJ categories do not yield TQFTs. But the restrictions to the even labels lead to TQFTs. When $A = \pm ie^{\pm\frac{2\pi i}{4r}}$, the resulting TQFTs are unitary. Those unitary TQFTs are the same as those obtained from the restrictions of WRT TQFTs to integral spins. All can be doubled to picture TQFTs. Note that for these cases when A is a primitive rth root of unity, then $-A$ is a primitive $2r$th root of unity. For the even sub-categories, they lead to the same TQFTs, which form part of the Black-White TQFTs in Theorem 9.1. $\square$

Appendix A. Topological phases of matter

Fractional quantum Hall liquids are new phases of matter exhibiting topological orders, and Chern-Simons theories are proposed as effective theories to describe the universal properties of such quantum liquids. Quantum Chern-Simons theories are (2+1)-dimensional topological quantum field theories (TQFTs), so we define a topological phase of matter as a quantum system with a TQFT effective theory.

Ground states manifolds as modular functors

While in real experiments, we will prefer to work with quantum systems in the plane, it is useful in theory to consider quantum systems on 2-dimensional surfaces such as the torus. Given a quantum system on a 2-dimensional oriented closed surface Σ, there associates a Hilbert space $\mathbb{H}$ consisting of all states of the system. The lowest energy states form the ground states manifold $V(\Sigma)$, which is a subspace of $\mathbb{H}$. For a given theory, the local physics of the quantum systems on different surfaces are the same, so there are relations among the ground states manifolds $V(\Sigma)$ for different Σ's dictated by the local physics. In a topological quantum system, the ground states manifolds form a modular functor—the 2-dimensional part of a TQFT. In particular, $V(\Sigma)$ depends only on the topological type of Σ.

Elementary excitations as particles

A topological quantum system has many salient features including an energy gap in the thermodynamic limit, ground states degeneracy and the lack of continuous evolutions for the ground states manifolds. The energy gap implies that elementary excitations are particle-like and particle statistics is well-defined. These quasi-particles are anyons, whose statistics are described by representations of the braid groups rather than representations of the permutation groups.

The mathematical model for an anyonic system is a ribbon category. In this model an anyon is pictured as a framed point in the plane: a small interval. Given a collection of n anyons in the plane, we will arbitrarily order them and place them onto the real axis, so we can represent them by intervals $[i-\epsilon, i+\epsilon], i = 1, 2, \cdots, n$ on the real axis for some small ϵ. The worldlines of any n anyons from time $t = 0$ to the same set of n anyons at time $t = 1$ form a *framed* braid in $\mathbb{R}^2 \times [0, 1]$. We will represent worldlines of anyons by diagrams of ribbons in the plane which are projections from $\mathbb{R}^2 \times [0, 1]$ to the real axis $\times [0, 1]$ with crossings. (Technically we need to perturb the worldlines in order to avoid more singular projections.) A further convention is the so-called blackboard framing: we will draw only single lines to represent ribbons with the understanding the ribbon is the parallel thickening of the lines in the plane.

Suppose n elementary excitations of a topological quantum system on a surface Σ are localized at points $p_1, p_2, \cdots, p_n$, by excising the particles from Σ, we have a topological quantum system on a punctured surface Σ' obtained from Σ by deleting a small disk around each point p_i. Then the ground states manifold of the quantum system on Σ' form a Hilbert space $V(\Sigma')$. The resulting Hilbert space should depend only on the topological properties of the particles—particle types that will be referred to also as labels. In this way we assign Hilbert spaces $V(\Sigma, a_1, a_2, \cdots, a_n)$ to surfaces with boundary components $\{1, 2, \cdots, n\}$ labelled by $\{a_1, a_2, \cdots, a_n\}$.

Braid statistics

The energy gap protects the ground states manifold, and when two particles are exchanged adiabatically within the ground states manifolds, the wave-functions are changed by a unitary transformation. Hence particle statistics can be defined as the resulting unitary representations of the braid groups.

Appendix B. Representation of linear category

Category theory is one of the most abstract branch of mathematics. It is extremely convenient to use category language to describe topological phases of matters. It remains to be seen whether or not this attempt will lead to useful physics. But tensor category theory might prove to be the right generalization of group theory for physics. On a superficial level, the two layers of structures in a category fit well with physics: objects in a category represent states, and morphisms between objects possible "physical processes" from one state to another. For quantum physics, the category will be linear so each morphism set is a vector space. Functors might be useful for the description of topological phase transitions and condensations of particles or string-nets. For more detailed introduction, consult the book.[Ma]

A *category* $\mathcal{C}$ consists of a collection of objects, denoted by $a, b, c, \cdots$, and a morphism set ${}_a\mathcal{C}_b$ (also denoted by $\mathrm{Mor}(a, b)$) for each ordered pair (a, b) of objects which satisfy the following axioms:

Given $f \in {}_a\mathcal{C}_b$ and $g \in {}_b\mathcal{C}_c$, then there is a morphism $f \cdot g \in {}_a\mathcal{C}_c$ such that

1)(Associativity):

If $f \in {}_a\mathcal{C}_b, g \in {}_b\mathcal{C}_c, h \in {}_c\mathcal{C}_d$, then $(f \cdot g) \cdot h = f \cdot (g \cdot h)$.

2)(Identity):

For each object a, there is a morphism $\mathrm{id}_a \in {}_a\mathcal{C}_a$ such that for any $f \in {}_a\mathcal{C}_b$ and $g \in {}_c\mathcal{C}_a$, $\mathrm{id}_a \cdot f = f$ and $g \cdot \mathrm{id}_a = g$.

We denote the objects of $\mathcal{C}$ by $\mathcal{C}^0$ and write $a \in \mathcal{C}^0$ for an object of $\mathcal{C}$. We use $\mathcal{C}^1$ to denote the disjoint union of all the sets ${}_a\mathcal{C}_b$. The morphism $f \cdot g \in {}_a\mathcal{C}_c$ is usually called the *composition* of $f \in {}_a\mathcal{C}_b$ and $g \in {}_b\mathcal{C}_c$, but our notation $f \cdot g$ is different from the usual convention $g \cdot f$ as we imagine the composition as the join of two consecutive arrows rather than the composition of two functions. This convention is convenient when the composition in ${}_a\mathcal{C}_a$ of a linear category is regarded as a multiplication to turn ${}_a\mathcal{C}_a$ into an algebra.

A category $\mathcal{C}$ is a *linear category* if each morphism set ${}_a\mathcal{C}_b$ is a finitely dimensional vector space, and the composition of morphisms is a bilinear map of vector spaces. It follows that for each object a, ${}_a\mathcal{C}_a$ is a finitely dimensional unital algebra. It follows that a finitely dimensional unital algebra can be regarded as a linear category with a single object. Another important linear category is the category of finitely dimensional vectors spaces $\mathcal{V}$. An object of $\mathcal{V}$ is a finitely dimensional vector space V. The morphism set $\mathrm{Mor}(V, W)$ between two objects V, W is $\mathrm{Hom}(V, W)$. More generally, given any finite set I, consider the linear category $\mathcal{V}[I]$ of I-graded vector spaces, which is a categorification of the group algebra $\mathbb{C}[G]$ if I is a finite group G. An object of $\mathcal{V}[I]$ is a collect of finitely dimensional vector spaces $\{V_i\}_{i\in I}$ labelled by elements of I, and the morphism set $\mathrm{Mor}(\{V_i\}_{i\in I}, \{W_j\}_{j\in I})$ is the (graded) vector space of linear maps $\oplus_{i\in I}\mathrm{Hom}(V_i, W_i)$. In the following all categories will be linear categories, and we will see that any semisimple linear category with finitely many irreducible representations is isomorphic to a category of a finite set graded vector spaces.

General representation theory

Definition B.1. A (right) representation of a linear category $\mathcal{C}$ is a functor $\rho : \mathcal{C} \to \mathcal{V}$, where $\mathcal{V}$ is the category of finitely dimensional vector spaces. The action is written on the right: $\rho(a) = V_a$ and given an $f \in {}_a\mathcal{C}_b, v.\rho(f) = v.f = v \cdot \rho(f) : V_a \to V_b$ for any $v \in V_a$.

The 0-representation of a category is the representation which sends every object to the 0-vector space. Fix an object $a \in \mathcal{C}^0$, we have a representation of the category $\mathcal{C}$, denoted by $a\mathcal{C}$: the representation sends a to the vector space ${}_a\mathcal{C}_a$, and any other $b \in \mathcal{C}^0$ to ${}_a\mathcal{C}_b$. An important construct which gives rise to all the representations of a semi-simple linear category is as follows: fix an object $a \in \mathcal{C}^0$ and a right ideal J_a of the algebra ${}_a\mathcal{C}_a$, then the map which sends each object $b \in \mathcal{C}^0$ to $J_a \cdot {}_a\mathcal{C}_b \subseteq {}_a\mathcal{C}_b$ affords $\mathcal{C}$ a representation, where $J_a \cdot {}_a\mathcal{C}_b$ is the subspace of ${}_a\mathcal{C}_b$ generated by all elements $f \cdot g, f \in J_a \subseteq {}_a\mathcal{C}_a, g \in {}_a\mathcal{C}_b$. If the right ideal J_a is generated by an element $p_a \in {}_a\mathcal{C}_a$, then the resulting representation of $\mathcal{C}$ will be denoted by $p_a\mathcal{C}$. In particular if $J_a = {}_a\mathcal{C}_a$, we will have the regular representation $a\mathcal{C}$.

The technical part of the paper will be the analysis of the representations of certain picture categories. In order to do this, we first recall the representation theory for an algebra—a linear category with a single object.

Definition B.2. Let A be an algebra, an element $e \in A$ is an idempotent if $e^2 = e \neq 0$. Two idempotents e_1, e_2 are orthogonal if $e_1e_2 = e_2e_1 = 0$. An idempotent is minimal if it is not the sum of two orthogonal idempotents.

Given an idempotent e of a finitely dimensional semi-simple algebra A, the right ideal eA is an irreducible representation of A if and only if the idempotent e is minimal. Since every irreducible right representation of A is isomorphic to a right ideal eA for some idempotent of A, the representations of A are completely known once we find a collection of pairwise orthogonal minimal idempotents e_i of A such that $1 = \oplus_i e_i$. It follows that $A = \oplus_{i=1}^n e_iA$.

Let $p(x)$ be a polynomial of degree=n with n distinct roots $a_1, a_2, \cdots, a_n$ and A be the quotient algebra $\mathbb{C}[x]/(p(x))$ of the polynomial algebra $\mathbb{C}[x]$. Let $u_j = \prod_{i=1, i\neq j}^n (x - a_j), \lambda_j = \prod_{i=1,i\neq j}^n (\lambda_j - \lambda_i)$ and $e_j = \frac{u_j}{\lambda_j}$. Then we have the following lemma.

Lemma B.1. *The idempotents $\{e_j\}_{j=1}^n$ of A are pair-wise orthogonal and $\oplus_{j=1}^n e_j = 1$. It follows that A is semi-simple and a direct sums of $\mathbb{C}$'s. Note that e_j is an eigenvector of the element $x \in A$ associated with the eigenvalue a_j.*

Proof. Since $u_j(x - a_j) = p(x) = 0$, so $u_j \cdot x = u_j \cdot a_j = a_ju_j$. It follows that $u_j^2 = u_j \prod_{i=1,i\neq j}^n (x - a_i) = \lambda_j u_j$, therefore $e_j^2 = e_j \neq 0$. Now consider $u_i \cdot u_j$, in u_j there is the factor $(x - a_i)$ if $i \neq j$, but $u_i(x - a_i) = p(x) = 0$, hence $u_iu_j = 0$.

The polynomial $g(x) = (\sum_{j=1}^n e_j) - 1$ is a polynomial of degree $n - 1$, but it has n distinct roots $a_1, a_2, \cdots, a_n$, so $g(x)$ is identically 0. □

A representation ρ of $\mathcal{C}$ is *reducible* if ρ is the direct sum of two nonzero representations of $\mathcal{C}$. Otherwise ρ is *irreducible*. A linear category $\mathcal{C}$ is *semi-simple* if every representation $\{\rho, V\}$ of $\mathcal{C}$ is a direct sum of irreducible representations.

Definition B.3. Λ has a positive definite Hermitian inner product (pdhi) iff each morphism set ${}_a\Lambda_b$ has a finite dimensional pdhi and composition ${}_a\Lambda_b \bigotimes {}_b\Lambda_c \xrightarrow{P} {}_a\Lambda_c$ satisfies the compatibility $< P({}_am_b \bigotimes {}_bm_c), {}_cm_d > = < {}_am_b, P({}_bm_c \bigotimes {}_cm_d) >, {}_im_j \in {}_i\Lambda_j$, and for all $i, j \in \Lambda^0, {}_i\Lambda_j$ is identified with $\overline{{}_j\Lambda_i}$.

Lemma B.2. *Suppose Λ has positive definite Hermitian inner product, then A is semi-simple.*

Proof. If Λ has a pdhi, then any (finite dimensional) representation $\{\rho, V\}$ of Λ may also be given a pdhi structure. This means that the V_i are individually pdhi-spaces and that for all morphisms $m, (\rho(m))^\dagger = \rho(\bar{m})$. One may check that any collections of pdhi-structures on $\{V\}$ which are averaged under the invertible morphisms (and therefore invariant) satisfies this condition. □

Most of the usual machinery of linear algebra, including Schur's lemma, holds for $\mathbb{C}$-linear categories.

Lemma B.3. *(Schur's Lemma for $\mathbb{C}$-linear categories) Suppose $\{\rho_m, V_i\}$ and $\{\chi_m, W_i\}, i \in obj(\Lambda), m \in Morph(i,j)$, are irreducible representations of a $\mathbb{C}$-linear category Λ (called an algebroid by some authors e.g. [BHMV]). Irreducibility means no ρ_m invariant class of proper subspaces $V_i' \subset V_i$ exists. Suppose that $\phi : \{V\} \to \{W\}$ is a Λ-map commuting with the action Λ. That is for $m \in Morph(i,j)$ and $v_i \in V_i$ we have $\chi_m(\phi_i(v_i)) = \phi_j \cdot \rho_m(v_i)$. Then either ϕ is identically zero for all $i, \phi_i : V_i \to W_i$, or ϕ is an isomorphism. If $\{V\} = \{W\}$ then $\phi = \lambda \cdot id$ for some $\lambda \in \mathbb{C}$.*

Proof. As in the algebra case $\ker(\phi)$ (and image(ϕ)) are both invariant families of subspaces (indexed by $i \in$ obj(Λ).) So if either is a nontrivial proper subspace for any i irreducibility of $\{V\}$ (or $\{W\}$) fails. For the second assertion, since $\mathbb{C}$ is algebraically closed for any $i \in$ obj(Λ), the characteristic equation $\det(\phi_i - x\mathrm{I}) = 0$, has roots, call one λ_i. The Λ-map $\phi - \lambda(I)$ has non-zero kernel (at least at object i) so by part one, $\phi - \lambda I = 0$ identically or $\phi = \lambda I$. □

Corollary B.1. *Suppose the representation $\{\rho, V\}$ of Λ has decomposition $\{V\} = V_{a_1} \bigotimes\{V_1\} \bigoplus \cdots \bigoplus V_{a_k} \bigotimes\{V_k\}$, where the V_l are distinct (up to isomorphism) irreducible representations, l is finite index, $1 \leq l \leq k$, and the V_{a_l} are ordinary $\mathbb{C}$-vector spaces with no Λ-action, (Dimension $(V_{a_l}) =: d(a_l)$ is the multiplicity of V_l.) The decomposition is unique up to permutation and of course isomorphism of V_{a_l} and scalars acting on $\{V_l\}$.*

Proof. Suppose $\{V\} = \bigoplus_m W_{a_m} \bigotimes\{W_j\}$. Apply Schur's lemma to compositions:

$$v_{a_l} \bigotimes\{V_l\} \to \{V\} \to w_{a_m} \bigotimes\{W_m\}.$$

for all $v_{a_l} \in V_{a_l}$ and $w_{a_l} \in W_{a_l}$ to conclude that given l, $V_l \cong W_m$ for some m and $d(a_l) = d(a_m)$. This established uniqueness. □

Now we state a structure theorem from[Wal2] for the representation theory of semisimple linear categories. Both the statement and the proof are analogous to those for the semi-simple algebras. A *right ideal* of $\mathcal{C}$ is a subset J of $\mathcal{C}^1$ such that for each object $a \in \mathcal{C}^0$, $J \cap {}_a\mathcal{C}_a$ is a right ideal of ${}_a\mathcal{C}_a$. Note that each right ideal of $\mathcal{C}$ affords $\mathcal{C}$ a representation.

Theorem B.1.

1): Let $\mathcal{C}$ be a semi-simple linear category, and $\{X_i\}_{i\in I}$ be a complete set of representatives for the simple right ideals of $\mathcal{C}$. Then $\mathcal{C}$ is naturally isomorphic to the category of the finite set I-graded vector spaces with each object $a \in \mathcal{C}^0$ corresponding to the graded vector space X_{ia}, where X_{ia} is $X_i \cap {}_a\mathcal{C}_a$.

2): Each irreducible representation ρ of $\mathcal{C}$ is given by a right ideal of the form $e_a\mathcal{C}$ for some object $a \in \mathcal{C}^0$, where e_a is a minimal idempotent of ${}_a\mathcal{C}_a$. If for some $b \in \mathcal{C}^0$, which may be a, and e_b is a minimal idempotent of ${}_b\mathcal{C}_b$, then the irrep $e_b\mathcal{C}$ of $\mathcal{C}$ is isomorphic to $e_a\mathcal{C}$ if and only if there exist $f \in {}_a\mathcal{C}_b$ and $g \in {}_b\mathcal{C}_a$ such that $f \cdot g = e_a$, $g \cdot f = e_b$.

Appendix C. Gluing and Maslov index

Gluing

Gluing of 3-manifolds needs to be addressed carefully due to anomaly. The basic problem is when X is a bordism, the canonical Lagrangian subspace $\lambda_X \in H_1(\partial X; \mathbb{R})$ is in general not a direct sum. λ_X is determined by the intrinsic topology as it is the kernel of the inclusion homomorphism:

$H_1(\partial X;\mathbb{R}) \to H_1(X;\mathbb{R})$. But the anomaly is related to the parameterizations of the bordisms, which are extrinsic.

Suppose $X_i, i = 1,2$ are bordisms from $-Y_i$ to Y_{i+1} extended by $\lambda_j, j = 1,2,3$. The canonical Lagrangian subspace λ_{X_i} defines a Lagrangian subspace of $H_1(Y_2;\mathbb{R})$ as follows: let $\lambda_- X_1 = \{b \in H_1(Y_2;\mathbb{R}) | \text{for some } a \in \lambda_1, (a,b) \in \lambda_{X_1}\}$, and $\lambda_+ X_2 = \{c \in H_1(Y_2;\mathbb{R}) | \text{for some } d \in \lambda_3, (c,d) \in \lambda_{X_2}\}$. Then we have three Lagrangian subspaces in $H_1(Y_2;\mathbb{R})$ together with λ_2.

More generally, let (Y_i, λ_i) be extended sub-surfaces of $(\partial X; \lambda_X)$, and $f : (Y_1; \lambda_1) \to (-Y_2; \lambda_2)$ be a gluing map. Then we have three Lagrangian subspaces in $H_1(Y_1;\mathbb{R}) \oplus H_1(Y_2;\mathbb{R})$: the direct sum $\lambda_1 \oplus \lambda_2$, the anti-diagonal $\Delta = \{(x \oplus -f_*(x))\}$, and K—the complement of λ_i in λ_X mapped here.

Maslov index

Given three isotropic subspaces $\lambda_i, i = 1,2,3$ of a symplectic vector space (H,ω), we can define a symmetric bilinear form $<,>$ on $(\lambda_1 + \lambda_2) \cap \lambda_3$ as follows: for any $v, w \in (\lambda_1 + \lambda_2) \cap \lambda_3$, write $v = v_1 + v_2, v_i \in \lambda_i$, then set $< v, w >= \omega(v_2, w)$. The Maslov index $\mu(\lambda_1, \lambda_2, \lambda_3)$ is the signature of the symmetric bilinear form $<,>$ on $(\lambda_1 + \lambda_2) \cap \lambda_3$.

References

BHMV. C. Blanchet; N. Habegger; G. Masbaum; P. Vogel, *Topological quantum field theories derived from the Kauffman bracket.* Topology 34 (1995), no. 4, 883–927.

DFNSS. S. Das Sarma; M. Freedman; C. Nayak; S. H. Simon,.; A. Stern, *Non-Abelian Anyons and Topological Quantum Computation,* arXiv:0707.1889.

DGG. P. Di Francesco; O. Golinelli; E. Guitter, *Meanders and the Temperley-Lieb algebra.* Comm. Math. Phys. 186 (1997), no. 1, 1—59.

Fd. D. S. Freed, *Higher algebraic structures and quantization.* Comm. Math. Phys. 159 (1994), no. 2, 343–398.

Fn. M. H. Freedman, *A magnetic model with a possible Chern-Simons phase.* With an appendix by F. Goodman and H. Wenzl. Comm. Math. Phys. 234 (2003), no. 1, 129–183.

FKLW. M. Freedman;A. Kitaev; M. Larsen; Z. Wang, *Topological quantum computation.* Bull. Amer. Math. Soc. (N.S.) 40 (2003), no. 1, 31–38.

FLW. M. H. Freedman, M. J. Larsen, and Z. Wang, *The two-eigenvalue problem and density of Jones representation of braid groups.* Comm. Math. Phys. **228** (2002), 177-199.

FQ. D. Freed; F. Quinn, *Chern-Simons theory with finite gauge group.* Comm. Math. Phys. 156 (1993), no. 3, 435–472.

GL. J. Graham; G. Lehrer, *The representation theory of affine Temperley-Lieb algebras.* Enseign. Math. (2) 44 (1998), no. 3-4, 173–218.

Jo1. V. Jones, *A polynomial invariant for knots via von Neumann algebras.* Bull. Amer. Math. Soc. (N.S.) 12 (1985), no. 1, 103–111.

Jo2. V. Jones, *Hecke algebra representations of braid groups and link polynomials.* Ann. of Math. (2) 126 (1987), no. 2, 335–388.

Jo3. V. Jones, *Index for subfactors.* Invent. Math. 72 (1983), no. 1, 1–25.

Jo4. V. F. R. Jones, *Braid groups, Hecke algebras and type* II_1 *factors,* Geometric methods in operator algebras (Kyoto, 1983), 242–273, Pitman Res. Notes Math. Ser. **123**, Longman Sci. Tech., Harlow, 1986.

K. C. Kassel, Quantum Groups. Graduate Texts in Mathematics, 155. Springer-Verlag, New York, 1995.

Ki1. A. Kitaev, *Fault-tolerant quantum computation by anyons.* Ann. Physics 303 (2003), no. 1, 2–30.

Ki2. A. Kitaev, *Anyons in an exactly solved model and beyond.* Ann. Physics 321 (2006), no. 1, 2–111.

KL. L. Kauffman; S. Lins, Temperley-Lieb recoupling theory and invariants of 3-manifolds. Annals of Mathematics Studies, 134. Princeton University Press, Princeton, NJ, 1994. x+296 pp.

KM. R. Kirby; P. Melvin, *The 3-manifold invariants of Witten and Reshetikhin-Turaev for* $\mathrm{sl}(2, C)$. Invent. Math. 105 (1991), no. 3, 473–545.

Li. W. Lickorish, *Three-manifolds and the Temperley-Lieb algebra.* Math. Ann. 290 (1991), no. 4, 657–670.

LRW. M. J. Larsen; E. C. Rowell: Z. Wang, *The N-eigenvalue problem and two applications,* Int. Math. Res. Not. **2005** (2005), no. 64, 3987–4018.

Ma. S. MacLane, Categories for the working mathematician. Graduate Texts in Mathematics, Vol. 5. Springer-Verlag, New York-Berlin, 1971. ix+262 pp.

Mu. M. Müger, *From subfactor to categories and topology, II* J. Pure Appl. Algebra 180 (2003), no. 1-2, 159–219.

MS. G. Moore; N. Seiberg, *Classical and quantum conformal field theory.* Comm. Math. Phys. 123 (1989), no. 2, 177–254.

RSW. E. Rowell; R. Stong; Z. Wang, *On classification of modular tensor categories*, arXiv: 0712.1377.

RT. N. Reshetikhin; V. Turaev, *Invariants of 3-manifolds via link polynomials and quantum groups.* Invent. Math. 103 (1991), no. 3, 547–597.

Tu. V. Turaev, Quantum Invariants of Knots and 3-Manifolds, De Gruyter Studies in Mathematics, Walter de Gruyter (July 1994).

TV. V. Turaev and O. Viro, *State sum invariants of 3-manifolds and quantum 6j-symbols.* Topology 31 (1992), no. 4, 865–902.

Wal1. K. Walker, On Witten's 3-manifold Invariants, 1991 notes at http://canyon23.net/math/.

Wal2. K. Walker, TQFTs, 2006 notes at http://canyon23.net/math/.

We. H. Wenzl, *On sequences of projections.* C. R. Math. Rep. Acad. Sci. Canada 9 (1987), no. 1, 5–9.

Wil. F. Wilczek, Fractional Statistics and Anyon Superconductivity, World Scientific Pub Co Inc (December 1990).

Witt. E. Witten, *Quantum field theory and the Jones polynomial.* Comm. Math. Phys. 121 (1989), no. 3, 351–399.

Berry Phase in Yang-Baxter Approach Related to Entangled States and the Yang-Baxterization of Braid Relation for Two-state Anyon Model

Mo-Lin Ge

Chern Institute of Mathematics
Nankai University
Tianjin, 300071, China

Louis H. Kauffman

Department of Mathematics, Statistics and Computer Science
University of Illinois at Chicago
IL 60607-7045, USA

Kang Xue

Department of Physics
Northeast Normal University
Changchun, Jilin 130024, China

Some entangled states can be generated based on Temperley-Lieb algebra. The dynamics in the entangling degree (θ in this paper) space is set up through the Yang-Baxterization and the Berry phase is found in the Yang-Baxter approach. The Yang-Baxterization has been presented for $A(u)$ and $B(u)$, i.e. the 2-dimensional braid matrices related to 2-state anyon model. We also show the equivalence between the usual 4×4 $\breve{R}(u)$-matrix and the set of $A(u)$ and $B(u)$.

1. Introduction

The Yang-Baxter equation (YBE) originates in solving δ-function interaction model[1] by C.N. Yang and statistical models[2] by R. Baxter and introduced to solve many quantum integrable models[3] by Faddeev and Leningrad Scholar. Through RTT relation[4] the new algebraic (quantum groups) structure was established[5] by V. Drinfeld. The usual YBE takes the form

$$\breve{R}_{12}(u)\breve{R}_{23}(u+v)\breve{R}_{12}(v) = \breve{R}_{23}(v)\breve{R}_{12}(u+v)\breve{R}_{23}(u) \qquad (1.1)$$

that is valid for three types of $\check{R}$-matrices, i.e. rational, trigonometric and elliptic solutions of YBE. The spectral parameter u plays important role that is 1-dimensional momentum (rapidity) in some typical models. The asymptotic behavior of $\check{R}_{12}(u)$ is u-independent.

$$\lim \check{R}_{i\ i+1}(u) = b_i \tag{1.2}$$

which satisfy braid relation

$$\begin{aligned} b_i b_{i+1} b_i &= b_{i+1} b_i b_{i+1} \\ b_i b_j &= b_j b_i, \qquad |i-j| > 2 \end{aligned} \tag{1.3}$$

where

$$b_i = 1 \times 1 \times 1 \cdots b_{i\ i+1} \times \cdots \times 1$$

For a statistical model all the elements of $\check{R}(u)$-matrix should be positive because they are related to the Boltzmann weights. The relationship between $\check{R}(u)$ and b was set up by M.Jimbo,[6] V.Jones[7] and others.[8] We call the process obtaining $\check{R}(u)$ for a given braid matrix b "Yang-Baxterization" that depends on the number of independent eigenvalues in matrix b.

If B has two independent eigenvalues we have simply

$$\check{R}(u) = \rho(xB - x^{-1}B^{-1}) \tag{1.4}$$

where $x = e^u$(or e^{iu}) and ρ is a normalization factor.

As was pointed out by Kauffman and Lomonaco[9,10] the braid matrix $B^{\frac{1}{2}\frac{1}{2}}$ transforms the "natural basis" $(|\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle\ |\downarrow\uparrow\rangle, |\downarrow\downarrow\rangle)$ to the Bell state $(\frac{1}{\sqrt{2}}(|\downarrow\downarrow\rangle \pm |\uparrow\uparrow\rangle), \frac{1}{\sqrt{2}}(|\downarrow\uparrow\rangle \pm |\uparrow\downarrow\rangle))$. It is emphased that here the elements of $B^{\frac{1}{2}\frac{1}{2}}$ is no longer positive. However, a braid matrix b is nothing with dynamics. We should Yang-Baxterize b to be $\check{R}(x)$-matrix and look for the physical consequence of the extension.

Generally a solution of $\check{R}(u)$ depends on two parameters, say, ϕ (the q-deformation parameter with $q = e^{i\phi}$, or it may be originated from other parameter such as $\eta = e^{i\phi}$, see below Eq(1)) and θ (the spectral parameter or the one dimensional momentum). In physics the parameter ϕ is flux that can be dependent on time t. If we define a quantum state

$$|\Phi(\theta, \phi(t))\rangle = \check{R}(\theta, \phi(t))|\Phi(0)\rangle, \tag{1.5}$$

where $|\Phi(0)\rangle$ is the initial state independent of t. The normalization condition of the quantum states $\langle\Phi(\theta,\phi(t))|\Phi(\theta,\phi(t))\rangle = \langle\Phi(0)|\Phi(0)\rangle = 1$ requires the unitary condition $\breve{R}^\dagger(\theta,\phi(t)) = \breve{R}^{-1}(\theta,\phi(t))$. It follows from Eq(5) that

$$\begin{aligned} & i\hbar\frac{\partial|\Phi(\theta,\phi(t))\rangle}{\partial t} \\ & = i\hbar\left[\frac{\partial\breve{R}(\theta,\phi(t))}{\partial t}\breve{R}^\dagger(\theta,\phi(t))\right]\breve{R}(\theta,\phi(t))|\Phi(0)\rangle \\ & = H(t)|\Phi(\theta,\phi(t))\rangle, \end{aligned} \tag{1.6}$$

where the Hamiltonian reads

$$H(t) = i\hbar\frac{\partial\breve{R}(\theta,\phi(t))}{\partial t}\breve{R}^\dagger(\theta,\phi(t)). \tag{1.7}$$

The $\breve{R}(\theta,\phi)$-matrix is related to the braiding matrix through Eq(3.4). Suppose $|\Phi(0)\rangle$ represents a direct-product state without entangling, and $B(\phi)|\Phi(0)\rangle$ yields an entangled state (in general, a maximally entangled state). For the case with two distinct eigenvalues, through the Yang-Baxterization $B(\phi) \to \breve{R}(\theta,\phi)$, Eq(1.7) defines the Hamiltonian for a Yang-Baxter systems. It is interesting to note that as was shown in[11] the meaning of $|2\cos\theta|$ is the entangling degree for the Bell state.

The purpose of this paper is to investigate the Berry phase (BP)[12] in Yang-Baxter systems, quantum criticality (QC) phenomenon[13–17] is also discussed. Special attention will be paid to set up the connection between the Yang-Baxter approach and topologic quantum field theory through 4×4 in YB and 2-dimensional in TQF model.[18]

2. Berry phase in Yang-Baxter approach

The braiding operators satisfy the following braid relations:

$$\begin{aligned} b_ib_{i+1}b_i &= b_{i+1}b_ib_{i+1}, \qquad & 1 \le i \le n-1, \\ b_ib_j &= b_jb_i, \qquad & |i-j| \ge 2, \end{aligned} \tag{2.1}$$

where the notation $b_i \equiv b_{i,i+1}$ has been used. Let us consider the following type of braiding matrix for two spin-1/2 particle[9,11]

$$B^{\frac{1}{2}\frac{1}{2}} = \frac{1}{\sqrt{2}}(I + M^{\frac{1}{2}\frac{1}{2}}), \tag{2.2}$$

where I is the 4×4 identity matrix and

$$M^{\frac{1}{2}\frac{1}{2}} = \begin{pmatrix} & & & e^{i\phi} \\ & & \epsilon & \\ & -\epsilon & & \\ -e^{-i\phi} & & & \end{pmatrix}, \tag{2.3}$$

with $\epsilon = \pm 1$, and $\phi = \phi(t)$ represents the arbitrary flux. The matrix $M^{\frac{1}{2}\frac{1}{2}}$ satisfies the algebraic relations of the extra-special two-group[10].[18]

The trigonometric Yang-Baxterization approach[6-8] gives $\breve{R}(x) = \rho(xB - x^{-1}B^{-1})$ (here ρ is a normalization factor) gives

$$\begin{aligned} \breve{R}(x) &= [2(x^2 + x^{-2})]^{-1/2}[(x + x^{-1})I^{\frac{1}{2}\frac{1}{2}} \\ &\quad +(x - x^{-1})M^{\frac{1}{2}\frac{1}{2}}], \end{aligned}$$

$$\begin{aligned} [\breve{R}(x)]^{-1} &= [2(x^2 + x^{-2})]^{-1/2}[(x + x^{-1})I^{\frac{1}{2}\frac{1}{2}} \\ &\quad -(x - x^{-1})M^{\frac{1}{2}\frac{1}{2}}]. \end{aligned} \tag{2.4}$$

The unitary condition $[\breve{R}(x)]^{-1} = \breve{R}(x^{-1})$ leads to $\phi(t)$ being real.

Equation (3.11) can be rewritten as

$$M^{\frac{1}{2}\frac{1}{2}} = e^{i\phi(t)}S_1^+S_2^+ - e^{-i\phi}S_1^-S_2^- + \epsilon(S_1^+S_2^- - S_1^-S_2^+), \tag{2.5}$$

where $S_i^{\pm} = S_i^1 \pm iS_i^2$ are raising and lowering operators of spin-1/2 angular momentum for the i-th lattice. We then have from Eq(1.7) that

$$\begin{aligned} H_1(x, \phi(t)) &= -\hbar\dot{\phi}[2(x^2 + x^{-2})]^{-1}(x - x^{-1}) \times \\ &\quad \{(x - x^{-1})(S_1^3 + S_2^3) + \\ &\quad (x + x^{-1})(e^{i\phi}S_1^+S_2^+ + e^{-i\phi}S_1^-S_2^-)\}. \end{aligned} \tag{2.6}$$

By using

$$\begin{aligned} x &= [-\cos 2\theta]^{-1/2}(\cos\theta + \sin\theta) \\ x^{-1} &= [-\cos 2\theta]^{-1/2}(\sin\theta - \cos\theta), \end{aligned} \tag{2.7}$$

Eq(5) can be recast to

$$\begin{aligned} H_1(\theta, \phi(t)) &= -\hbar\dot{\phi}\cos\theta[\cos\theta(S_1^3 + S_2^3) + \sin\theta(e^{i\phi}S_1^+S_2^+ \\ &\quad +e^{-i\phi}S_1^-S_2^-)]. \end{aligned} \tag{2.8}$$

The eigen-problem of Eq(10) under adiabatic approximation is

$$H_1(\theta, \phi(t))|\Phi_{\pm}(\theta, \phi(t))\rangle_1 = E_{\pm}^1(t)|\Phi_{\pm}(\theta, \phi(t))\rangle_1, \tag{2.9}$$

where the two non-zero eigenvalues are

$$\begin{aligned} E_{\pm}^{1} &= \mp\hbar\dot{\phi}\cos\theta \\ &= \mp\hbar\omega\cos\theta \quad (\text{for } \phi=\omega t), \end{aligned} \tag{2.10}$$

and the corresponding eigenstates are

$$\begin{aligned} |\Phi_{+}(\theta,\phi)\rangle &= \cos\tfrac{\theta}{2}|\uparrow\uparrow\rangle + \sin\tfrac{\theta}{2}e^{-i\phi}|\downarrow\downarrow\rangle, \\ |\Phi_{-}(\theta,\phi)\rangle &= -\sin\tfrac{\theta}{2}e^{i\phi}|\uparrow\uparrow\rangle + \cos\tfrac{\theta}{2}|\downarrow\downarrow\rangle. \end{aligned} \tag{2.11}$$

The physical consequence of Berry phase for the above Yang–Baxter Hamiltonian system, i.e., for $H_1(\theta,\phi(t))$, has been discussed in.[19] Namely, from the definition of Berry phase

$$\begin{aligned} \gamma(c) &= i\int_0^T dt\langle n(\vec{R})|\frac{\partial}{\partial t}|n(\vec{R})\rangle = i\int_0^T dt A(t) \\ &= i\int_0^{2\pi} dt\dot{\phi}^{-1}\langle n(\vec{R})|\frac{\partial}{\partial\phi}|n(\vec{R})\rangle, \end{aligned} \tag{2.12}$$

here $\vec{R}=(\sin\theta\cos\phi,\sin\theta\sin\phi,\cos\theta)$ and $|n(\vec{R})\rangle = |\Phi_{\pm}(\theta,\phi)\rangle$, one then obtains the Berry phases for the Yang–Baxter system as

$$\gamma_{\pm}^{1} = (\pm\int_0^{2\pi} d\phi)\sin^2\frac{\theta}{2} = \pm\pi(1-\cos\theta) = \pm\frac{\Omega}{2}, \tag{2.13}$$

where $\Omega = 2\pi(1-\cos\theta)$ is the familiar solid angle enclosed by the loop on the Bloch sphere in θ-space.

In terms of $S_i^+ = f_i^+, S_i^- = f_i$ and $n_i = f_i^+ f_i$ the Hamiltonian (10) can be recast to the form

$$\begin{aligned} &H_1(\theta,\phi(t)) = -\hbar\omega\cos\theta[\cos\theta\cdot(\hat{n}_1+\hat{n}_2-1) \\ &+\sin\theta(e^{i\phi(t)}S^+ + e^{-i\phi(t)}S^-)], \end{aligned} \tag{2.14}$$

or

$$H_1(\theta,\phi(t)) = -\hbar\omega\eta(\theta)H_0(\theta,\phi(t)) \tag{2.15}$$

where

$$H_0(\theta,\phi(t)) = 2\eta(\theta)S^3 + \Delta(t)S^+ + \Delta(t)^*S^- \tag{2.16}$$

$$\eta(\theta) = \cos\theta, \quad \Delta(t) = \sin\theta e^{i\phi(t)} \tag{2.17}$$

The standard procedure making $H_0(\theta,\phi(t))$ diagonal is

$$W^\dagger H_0 W = 2\mathcal{E}S_3, \quad \mathcal{E} = \sqrt{(\eta(\theta))^2 + |\Delta(t)|^2} = 1 \tag{2.18}$$

and the eigenstate is

$$\begin{aligned} &|\xi(\theta)\rangle = W|\text{vacuum}\rangle = \exp(\xi S_+ - \xi^* S_-)|\text{vacuum}\rangle, \\ &S_-|\text{vacuum}\rangle = 0, \end{aligned} \tag{2.19}$$

with

$$\xi = re^{i\phi(t)}, \quad \cot(2r) = -\frac{\eta(\theta)}{|\Delta(t)|} \tag{2.20}$$

Substituting Eq(2.17) into Eq(17) and Eq(19) we obtain

$$E = 1, \quad r = -\frac{\theta}{2},$$

in other words, we have

$$\begin{aligned} W^\dagger H W|\xi(\theta)\rangle &= -\hbar\omega \cdot 2\cos\theta S_3|\xi(\theta)\rangle \\ &= -\hbar\omega\cos\theta(\hat{n}_1 + \hat{n}_2 - 1)|\xi(\theta)\rangle. \end{aligned} \tag{2.21}$$

It is nothing but an oscillator Hamiltonian formed by two fermions with the frequency $\omega\cos\theta$. When $\theta = 0$ Eq (16) reduces to the standard oscillator for $\Delta(t) = 0$. However, When $\theta \neq 0$, $\Delta(t)$ plays a role of the "energy gap" and the wave function takes the form of spin coherent state.[20]

3. Berry phase for hamiltonian $H_2(\theta, \phi(t))$

In this section, we come to study the Berry phase for a kind of Yang–Baxter Hamiltonian related to the well-known six-vertex model[3] and the Temperley-Lieb algebra.

For the well-known six-vertex model, the braiding matrix reads

$$\begin{aligned} B = S^{\frac{1}{2}\frac{1}{2}} &= \begin{bmatrix} q & 0 & 0 & 0 \\ 0 & 0 & -\eta & 0 \\ 0 & -\eta & q - q^{-1} & 0 \\ 0 & 0 & 0 & q \end{bmatrix} \\ &= q(I - q^{-1}U^{\frac{1}{2}\frac{1}{2}}), \end{aligned} \tag{3.1}$$

where

$$U^{\frac{1}{2}\frac{1}{2}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & q & \eta & 0 \\ 0 & \eta^{-1} & q^{-1} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.2}$$

The matrix $U^{\frac{1}{2}\frac{1}{2}}$ satisfies the Temperley-Lieb algebra, i.e., $U_iU_{i\pm1}U_i = U_i$, $U_i^2 = d\ U_i$ (for the above matrix $U^{\frac{1}{2}\frac{1}{2}}$, $d = q + q^{-1}$). Similarly, the trigonometric Yang-Baxterization approach gives

$$\begin{aligned}\breve{R}(x) = [q^2 + q^{-2} - (x^2 + x^{-2})]^{-1/2}[(qx - q^{-1}x^{-1})I \\ -(x - x^{-1})U^{\frac{1}{2}\frac{1}{2}}],\end{aligned} \tag{3.3}$$

$$\begin{aligned}[\breve{R}(x)]^{-1} = [q^2 + q^{-2} - (x^2 + x^{-2})]^{-1/2}[(qx^{-1} - q^{-1}x)I \\ +(x - x^{-1})U^{\frac{1}{2}\frac{1}{2}}].\end{aligned} \tag{3.4}$$

It is easy to check that $[\breve{R}(x)]^\dagger = [\breve{R}(x)]^{-1} = \breve{R}(-x)$ for $x = e^{i\vartheta}$, $\eta = e^{i\varphi(t)}$, and $\theta, \varphi(t), q \in$ real.

One may symmetrize the matrix $\breve{R}(x)$ given by Eq(3.3) (i.e., to make the matrix elements $\left[\breve{R}(x)\right]^{1/2,-1/2}_{1/2,-1/2} = \left[\breve{R}(x)\right]^{-1/2,1/2}_{-1/2,1/2}$) through the following unitary transformation

$$\breve{R}_{i\ i+1}(V(x)) = V(x)\breve{R}_{i\ i+1}(x)V(x)^\dagger, \tag{3.5}$$

where $V(x) = V_i(x) \otimes [V_{i+1}(x)]^{-1}$ and

$$V_i(x) = \begin{pmatrix} 0 & x^{-\frac{1}{4}} \\ x^{\frac{1}{4}} & 0 \end{pmatrix}. \tag{3.6}$$

The resultant $\breve{R}_{i\ i+1}(V(x))$ is still a solution of YBE. Let only the parameter $\eta = e^{i\varphi(t)}$ be time-dependent, it yields from Eq(1.7) and Eq(3.5) that

$$\begin{aligned}H_2(x, \varphi(t)) = \hbar\dot{\varphi}\left[q^2 + q^{-2} - (x^2 + x^{-2})\right]^{-1}(x - x^{-1}) \\ \times[(x - x^{-1})(S_1^3 - S_2^3) + \\ (q - q^{-1})(e^{i\varphi}S_1^+S_2^- - e^{-i\varphi}S_1^-S_2^+)].\end{aligned} \tag{3.7}$$

Putting $x = e^{i\vartheta}$, $\vartheta = \pi/2 - \theta$ and $\varphi(t) = \phi(t) - \pi/2 = \omega t$, we have

$$\begin{aligned}H_2(\theta, \phi(t)) = -4\hbar\omega\left[q^2 + q^{-2} + 2\cos 2\theta\right]^{-1}\cos\theta \\ \times[\cos\theta(S_1^3 - S_2^3) + \\ \tfrac{1}{2}(q - q^{-1})(e^{i\phi}S_1^+S_2^- + e^{-i\phi}S_1^-S_2^+)],\end{aligned} \tag{3.8}$$

whose two nonzero eigenvalues are

$$\begin{aligned}E_\pm^2 &= -4\hbar\dot{\phi}(q^2 + q^{-2} + 2\cos 2\theta)^{-1}\cos\theta\lambda_\pm \\ &= -\frac{4\hbar\dot{\phi}\cos\theta}{\lambda_\pm},\end{aligned} \tag{3.9}$$

with

$$\lambda_\pm = \pm\sqrt{\cos^2\theta + (q - q^{-1})^2/4}. \tag{3.10}$$

Under the adiabatic approximation the corresponding eigenstates are

$$\begin{aligned} |\Phi_+(\theta,\phi)\rangle &= \tfrac{1}{\sqrt{2\lambda_+}}[(\lambda_+ - \cos\theta)^{-1/2}\left(\tfrac{q-q^{-1}}{2}\right)|\uparrow\downarrow\rangle \\ &\quad + i(\lambda_+ - \cos\theta)^{1/2}e^{-i\phi}|\downarrow\uparrow\rangle], \\ |\Phi_-(\theta,\phi)\rangle &= \tfrac{1}{\sqrt{2\lambda_-}}[i(\lambda_- - \cos\theta)^{-1/2}\left(\tfrac{q-q^{-1}}{2}\right)e^{i\phi}|\uparrow\downarrow\rangle \\ &\quad -(\lambda_- - \cos\theta)^{1/2}e^{-i\phi}|\downarrow\uparrow\rangle]. \end{aligned} \tag{3.11}$$

The corresponding Berry phases for the Yang–Baxter system are

$$\begin{aligned} \gamma^2_\pm &= \pm\pi\left(1 - \frac{1}{\lambda_+}\cos\theta\right) \\ &= \pm\pi\left[1 - \frac{\cos\theta}{[\cos^2\theta + (q-q^{-1})^2/4]^{1/2}}\right]. \end{aligned} \tag{3.12}$$

The above Berry phases have been "q-deformed", when $\lambda_+ = 1$, or $q = \sqrt{1+\sin^2\theta} \pm \sin\theta$, Eq(3.12) reduces to Eq(15). Remarkably the Berry phases in Eq(3.12) can still be expressed in terms of the concurrence of the states $|\Phi_\pm(\theta,\varphi)\rangle$ in Eq(3.11) as $\gamma^2_\pm = \mp\pi(1-\sqrt{1-\mathcal{C}^2})$, where $\mathcal{C} = (q-q^{-1})/(2\lambda_+)$.

Similarly, the Hamiltonian $H_2(\theta,\phi(t))$ can be rewritten in terms of $SU(2)$ generators $J^+ = S_1^+S_2^- = \hat{f}_1^\dagger\hat{f}_2$, $J^- = S_1^-S_2^+ = \hat{f}_1\hat{f}_2^\dagger$, $J^3 = (S_1^3 - S_2^3) = (\hat{n}_1 - \hat{n}_2)/2$ as

$$H_2(\theta,\phi(t)) = -4\hbar\omega\frac{\cos\theta}{q^2+q^{-2}+2\cos 2\theta}H_0'(\theta,\phi(t)), \tag{3.13}$$

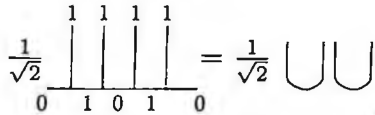
where

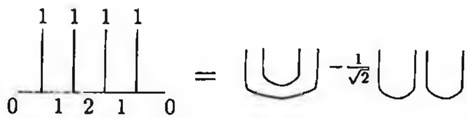
$$\begin{aligned} &H_0'(\theta,\phi(t)) = 2\varepsilon(\theta)J^3 + \Delta(t)J^+ + \Delta(t)^*J^-, \\ &\varepsilon(\theta) = \cos\theta, \quad \Delta(t) = e^{i\phi(t)}(q-q^{-1})/2. \end{aligned} \tag{3.14}$$

When $q - q^{-1} = 0$, or $q = \pm 1$, the Hamiltonian $H_0'(\theta,\phi(t))$ contracts to $H_0'(\theta,\phi(t)) = \varepsilon(\theta)(\hat{n}_1 - \hat{n}_2)$, thus the quantum criticality occurs. Correspondingly, one may easily see that the Berry phases in Eq(3.12) vanish.

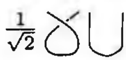
4. Yang-Baxterization of a simple model in 2-dimensional braid relation.

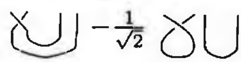
We have seen that the Yang-Baxterization procedure in the section I is used to yield the Berry phase. However, in connection with the FQHE there are two-dimensional braid matrices as was shown by Z. Wang in his lectures at Nankai Institute based on the works of M. Freedman, Z. Wang and others.[21] For the model with two types of particles the basis is taken as $|e_1\rangle$ and $|e_2\rangle$ that in terms of Kauffman's graphic method[22] are shown by

$|e_1\rangle =$  ,

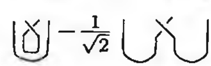
$|e_2\rangle =$  , (4.1)

while the corresponding operators behave as

$A|e_1\rangle =$  $= |e_1\rangle$

$A|e_2\rangle =$  $= i|e_1\rangle$

$B|e_1\rangle =$  $= (\frac{1+i}{2})\,|e_1\rangle + (\frac{1-i}{2})\,|e_2\rangle$

$B|e_2\rangle =$  $= (\frac{1-i}{2})\,|e_1\rangle + (\frac{1+i}{2})\,|e_2\rangle$, (4.2)

where the operator A makes crossing in the spaces 1 and 2, whereas B for 2 and 3. Thus their matrix representations in the basis $|\Phi\rangle = \begin{pmatrix} |e_1\rangle \\ |e_2\rangle \end{pmatrix}$ are given by

$$A = \rho \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}, \quad B = \frac{\rho}{2} \begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix}, \quad \rho = e^{-i\pi/8}, \tag{4.3}$$

and they satisfy braid relation

$$ABA = BAB. \tag{4.4}$$

We emphasize that Eq(4.4) should act on the combined basis $|\Phi\rangle$. It is worthy noting that the "crossing" in Eq(4) means the usual 4×4 braid matrix.

In the following we shall show the Yang-Baxterization of Eq(4.4) is given by

$$\breve{R}_{12}(u)\breve{R}_{23}(\frac{u+v}{1+\beta^2 uv})\breve{R}_{12}(v) = \breve{R}_{23}(v)\breve{R}_{12}(\frac{u+v}{1+\beta^2 uv})\breve{R}_{23}(u), \tag{4.5}$$

where $\breve{R}_{12}(u) = \breve{R}(u) \otimes \mathbf{1}_2, \breve{R}_{23}(v) = \mathbf{1}_2 \otimes \breve{R}(v)$ and the YBE Eq(7) admits the celebrated Temperley-Lieb algebra (TLA) for

$$\begin{aligned} \breve{R}_{12}(u) &= a_1(u)\mathbf{1}_4 + b_1(u)U_{12}, \\ \breve{R}_{23}(u) &= a_2(u)\mathbf{1}_4 + b_2(u)U_{23}, \end{aligned} \tag{4.6}$$

with U satisfying TLA

$$U^2 = dU,\ U_{12}U_{23}U_{12} = U_{12},\ U_{23}U_{12}U_{23} = U_{23}. \tag{4.7}$$

Actually substituting Eq(8) and Eq(9) into Eq(1) we obtain the following independent relations

$$\begin{aligned}
&a_1(u)a_2(u+v)a_1(v) = a_2(v)a_1(u+v)a_2(u)\\
&a_1(u)b_2(u+v)b_1(v) = b_2(v)b_1(u+v)a_2(u)\\
&a_1(u)b_1(v) + b_1(u)a_1(v) + d\,b_1(u)b_1(v)a_2(u+v) + b_1(u)b_2(u+v)b_1(v)\\
&\qquad = a_2(v)b_1(u+v)a_2(u)\\
&a_1(u)b_2(u+v)a_2(v) = [a_2(v)b_2(u) + b_2(v)a_2(u) + d\,b_2(v)b_2(u)]a_1(u+v)\\
&\qquad + b_2(v)b_1(u+v)b_2(u)
\end{aligned} \tag{4.8}$$

where Eq(9) has been used.

On the other hand by directly acting the 4-dimensional $\check{R}(u)$-matrix on the base $|e_1\rangle$ and $|e_2\rangle$ one obtains

$$\begin{aligned}
\check{R}_{12}(u)\,|e_1\rangle &= [a_1(u) + d\ b_1(u)]\,|e_1\rangle\,,\\
\check{R}_{12}(u)\,|e_2\rangle &= a_1(u)\,|e_2\rangle\,,\\
\check{R}_{23}(u)\,|e_1\rangle &= [a_2(u) + \frac{b_2(u)}{d}]\,|e_1\rangle + \frac{\sqrt{d^2-1}}{d}b_2(u)\,|e_2\rangle\,,\\
\check{R}_{23}(u)\,|e_2\rangle &= \frac{\sqrt{d^2-1}}{d}b_2(u)\,|e_1\rangle + [a_2(u) + \frac{d^2-1}{d}b_2(u)]\,|e_2\rangle\,.
\end{aligned} \tag{4.9}$$

Introducing matrix elements $A(u)_{ij} = \langle e_i|\check{R}_{12}(u)|e_j\rangle$ and $B(u)_{ij} = \langle e_i|\check{R}_{23}(u)|e_j\rangle$ $(i,j = 1,2)$ we have

$$\mathrm{A}(u) = \begin{bmatrix} a_1(u) + d\,b_1(u) & 0\\ 0 & a_1(u)\end{bmatrix}, \mathrm{B}(u) = \begin{bmatrix} a_2(u) + \frac{b_2(u)}{d} & \frac{\sqrt{d^2-1}}{d}b_2(u)\\ \frac{\sqrt{d^2-1}}{d}b_2(u) & a_2(u) + \frac{d^2-1}{d}b_2(u)\end{bmatrix} \tag{4.10}$$

that in terms of Pauli matrices can be recast to

$$A(u) = f_1(u)I + g_1(u)\sigma_3,\ B(u) = f_2(u)I + g_2(u)\sigma_1 + h_2(u)\sigma_3 \tag{4.11}$$

where

$$\begin{aligned}
f_1(u) &= a_1(u) + \frac{d}{2}b_1(u), \qquad g_1(u) = \tfrac{d}{2}b_1(u);\\
f_2(u) &= a_2(u) + \frac{d}{2}b_2(u),\ g_2(u) = \frac{\sqrt{d^2-1}}{2}b_2(u),\\
h_2(u) &= \frac{2-d^2}{2d}b_2(u).
\end{aligned} \tag{4.12}$$

We first discuss a system other than Eq(4.3). It has $d = 2$ in which case we should let A(u) and B(u) satisfy the spectral parameter dependent braid relation

$$A(u)B(u+v)A(u) = B(v)A(u+v)B(u) \tag{4.13}$$

We substitute Eq(13) into Eq(15) and find the independent relations

$$\begin{aligned}
&[f_1(u)f_1(v) + g_1(u)g_1(v)]f_2(u+v) + \alpha[f_1(u)g_1(v) + g_1(u)f_1(v)]g_2(u+v) \\
&\quad = [f_2(v)f_2(u) + (\alpha^2+1)g_2(v)g_2(u)]f_1(u+v) \\
&\quad\quad + \alpha[f_2(v)g_2(u) + g_2(v)f_2(u)]g_1(u+v)
\end{aligned} \tag{4.14}$$

$$\begin{aligned}
&[f_1(u)g_1(v) + g_1(u)f_1(v)]f_2(u+v) + 2\alpha g_1(u)g_1(v)g_2(u+v) \\
&\quad = [f_2(v)f_2(u) - (\alpha^2+1)g_2(v)g_2(u)]g_1(u+v)
\end{aligned} \tag{4.15}$$

$$\begin{aligned}
&[f_1(u)g_1(v) - g_1(u)f_1(v)]g_2(u+v) \\
&\quad = [g_2(v)f_2(u) - f_2(v)g_2(u)]g_1(u+v)
\end{aligned} \tag{4.16}$$

For the particular case

$$\begin{aligned}
a_1(u) = a_2(u) = a(u) \\
b_1(u) = b_2(u) = b(u)
\end{aligned} \tag{4.17}$$

We then have

$$\check{R}_{i,i+1}(u) = a(u)I_{i,i+1} + b(u)U_{i,i+1} \tag{4.18}$$

and the corresponding $A(u)$ and $B(u)$ read

$$\begin{aligned}
A(u) &= f(u)\,I + g(u)\sigma_3 \\
B(u) &= f(u)\,I + \tfrac{2\sqrt{d^2-1}}{d^2}g(u)\sigma_1 + \tfrac{2-d^2}{d^2}g(u)\sigma_3
\end{aligned} \tag{4.19}$$

where $f(u) = a(u) + \frac{d}{2}b(u), g(u) = \frac{d}{2}b(u)$ and satisfy

$$\begin{aligned}
&[f(u)g(v) + g(u)f(v)]f(u+v) = [f(v)f(u) + \\
&\quad (1 - \frac{4}{d^2})g(v)g(u)]g(u+v)
\end{aligned} \tag{4.20}$$

or equivalently

$$\begin{aligned}
&[a(u)b(v) + b(u)a(v) + d\,b(u)b(v)]a(u+v) = \\
&\quad [a(v)a(u) - b(v)b(u)]b(u+v)
\end{aligned} \tag{4.21}$$

When $a(u)$ and $b(u)$ are independent of u we put $a = \rho, b = \rho\xi$, then

$$A = \rho\begin{bmatrix} 1+\xi d & 0 \\ 0 & 1 \end{bmatrix}, \qquad B = \rho\begin{bmatrix} 1+\frac{\xi}{d} & \frac{\sqrt{d^2-1}}{d}\xi \\ \frac{\sqrt{d^2-1}}{d}\xi & 1+\frac{d^2-1}{d}\xi \end{bmatrix} \tag{4.22}$$

and Eq(23) leads to

$$\xi^2 + \xi d + 1 = 0 \tag{4.23}$$

i.e.

$$\xi = \frac{1}{2}(-d \pm \sqrt{d^2 - 4}) \tag{4.24}$$

Eq(24) satisfies the braid relation

$$ABA = BAB \tag{4.25}$$

For $\check{R}(u)$ Eq(23) and Eq(20) it follows

$$\begin{aligned} a(u) &= \rho(u), \quad b(u) = \rho(u)G(u) \\ G(u) &= \frac{u}{\gamma - u} \quad (d = 2) \quad (\gamma \text{ arbitrary}) \end{aligned} \tag{4.26}$$

The corresponding U-matrix takes the form

$$U = \begin{bmatrix} 0 & & & 0 \\ & 1 & e^{i\phi} & \\ & e^{-i\phi} & 1 & \\ 0 & & & 0 \end{bmatrix} \tag{4.27}$$

that had been given by Eq(2).

However this type of solution does not compatible with Eq(4.3). In order to make the equivalence between the 4-dimensional $\check{R}(x)$ -matrix and Yang-Baxterized braid relation of Eq(4.3) we should go in another way.

5. Yang-Baxterization of Eq(4.3)

We shall show that the consistent Yang-Baxterization for both $\check{R}(u)$ and $A(u), B(u)$ is given by

$$\check{R}_{12}(u)\check{R}_{23}(\frac{u+v}{1+\beta^2 uv})\check{R}_{12}(v) = \check{R}_{23}(v)\check{R}_{12}(\frac{u+v}{1+\beta^2 uv})\check{R}_{23}(u), \tag{5.1}$$

$$A(u)B(\frac{u+v}{1+\beta^2 uv})A(v) = B(v)A(\frac{u+v}{1+\beta^2 uv})B(u), \tag{5.2}$$

where β is an arbitrary constant. It looks like Lorentz transformation for $\beta = iv$ and the constraint equation for $a(u)$ and $b(u)$ are

$$\begin{aligned} &[a(u)b(v) + b(u)a(v) + db(u)b(v)]a(\frac{u+v}{1+\beta^2 uv}) = \\ &[a(v)b(u) + b(v)a(u)]b(\frac{u+v}{1+\beta^2 uv}) \end{aligned} \tag{5.3}$$

In fact the $\check{R}(\alpha)$-matrix can be understood to be dependent on a new spectral parameter u through $\alpha = i\beta^{-1}\tan u$.

Putting $a(u) = \rho(u)$ and $b(u) = \rho(u)G(u)$ again, the relation satisfied by $G(u)$ is given by

$$G(u) + G(v) + d\,G(u)G(v) = [1 - G(u)G(v)]G\left(\frac{u+v}{1+\beta^2 uv}\right) \tag{5.4}$$

whose solution can be found

$$G(u) = \frac{i\,4\epsilon\beta u}{d\,[1+\beta^2u^2 - i\,2\epsilon\beta u]} \qquad \text{for} \quad d = \sqrt{2} \tag{5.5}$$

where $\epsilon = \pm 1$ and β is arbitrary. It is emphasized that Eq(5.5) is the solution of YBE for $d = \sqrt{2}$ only.

Substituting Eq(5.5) into Eq(12) with $a_1(u) = a_2(u) = a(u)$, $b_1(u) = b_2(u) = b(u)$, we obtain

$$\begin{aligned} A(u) &= \rho(u)\begin{bmatrix} \frac{1+\beta^2u^2+i\,2\epsilon\beta u}{1+\beta^2u^2-i\,2\epsilon\beta u} & 0 \\ 0 & 1 \end{bmatrix} \\ B(u) &= \frac{a(u)}{1+\beta^2u^2-i\,2\epsilon\beta u}\begin{bmatrix} 1+\beta^2u^2 & i\,2\epsilon\beta u \\ i\,2\epsilon\beta u & 1+\beta^2u^2 \end{bmatrix} \end{aligned} \tag{5.6}$$

or for real β by letting $\dfrac{1+\beta^2u^2+i\,2\epsilon\beta u}{1+\beta^2u^2-i\,2\epsilon\beta u} = e^{-2i\theta}$ and $\rho(u) = e^{i\theta}$

$$A(u) = \begin{bmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{bmatrix}, \qquad B(u) = \begin{bmatrix} \cos\theta & -i\sin\theta \\ -i\sin\theta & \cos\theta \end{bmatrix} \tag{5.7}$$

On the other hand we have had the U-matrix with $d = \sqrt{2}$

$$U = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 & 0 & e^{i\phi} \\ 0 & 1 & i\epsilon & 0 \\ 0 & -i\epsilon & 1 & 0 \\ -e^{-i\phi} & 0 & 0 & 1 \end{bmatrix} \qquad \epsilon = \pm 1 \tag{5.8}$$

It follows from Eq(26) for $d = \sqrt{2}$ $\quad \xi = -\exp(\pm i\pi/4)$ the corresponding matrices A and B read

$$A = \mp i\rho\begin{bmatrix} 1 & 0 \\ 0 & \pm i \end{bmatrix}, \quad B = \frac{e^{\mp i\pi/4}}{\sqrt{2}}\rho\begin{bmatrix} 1 & \mp i \\ \mp i & 1 \end{bmatrix} \tag{5.9}$$

The relation Eq(5.1) tells when $\check{R}(u)$ reduces to braid matrix we have to set $u = v = \dfrac{u+v}{1+\beta^2 uv}$. There are only two possibilities: $u = v = 0$ and $u = v = \beta^{-1}$. When $u = v = 0$, $A(0) = B(0) = \rho(0)I$ whereas for the later case it leads to the familiar braid matrices

$$A(\beta^{-1}) = i\epsilon\rho \begin{bmatrix} 1 & 0 \\ 0 & -i\epsilon \end{bmatrix}, \quad B(\beta^{-1}) = \frac{e^{i\epsilon\pi/4}}{\sqrt{2}} \begin{bmatrix} 1 & i\epsilon \\ i\epsilon & 1 \end{bmatrix} \qquad (\epsilon = \pm 1) \tag{5.10}$$

Hence, the A and B given by Eq(5.9) is nothing but the light velocity limit of Eq(5.6).

In conclusion we have bridged the 4-dimensional $\check{R}(u)$-matrix satisfying TLA with $d = \sqrt{2}$ and the 2-dimensional braid matrices related to the anyon model with two states.

We are grateful to Prof. Z. Wang for enlighten discussions. This work is in part supported by NSFC (Grant No.10575053).

References

1. C.N. Yang, Phys. Rev. Lett. **19**, 1312 (1967), Phys. Rev. **168**, 1920 (1968). C.N. Yang, "Some Exactly soluble problem in statistical mechanics" Lectures given at the Karpacz Winter School of Theoretical Physics, Feb.1970, University of Wroclaw, Poland.
2. R.J. Baxter, Exactly solved Models in statistical Mechanics (Academic Press, London, 1982).
3. M. Jimbo (ed), Yang-Baxter Equation in Integrable Systems, Wrold Scientific Pub., Singapore, 1989.
4. E.K. Sklyanin, Zapiski Nauchnykh Seminarov Leningradskogo Otdeleniya Matematicheskogo. Instituta im. V. A. Stekiova AN SSSR, **95**, pp.55-128, 1980
   L.D. Faddeev, Integrable models in 1 + 1 dimensional QFT, Les Houches Lectures, pp.536-608, Elsevier, Amsterdam, 1984
   P.P. Kulish and E.K. Sklyanin, Lecture Nots in Phys. 151, pp.61-119.
5. V.G. Drinfeld, Soviet Math. Dokl **32**, pp.254-258, (1985). Dokl **35**, pp.212-216, (1988). Proceedings of ICM, pp.269-291 (1986), Berkeley.
6. M.Jimbo, Lett. Math. Phys. **10**, 63-69 (1985).
7. V.F.R. Jones, Commun. Math. Phys. **125**, 459 (1987).
8. M.L. Ge, Y.S. Wu and K. Xue, Int. J. Mod. Phys. **6A**, 3735 (1991).
9. L.H. Kauffman and S.J. Lomonaco, New J. Phys. **6**, 134 (2004).
10. J. M. Franko, E. C. Rowell and Z. Wang, J. Knot Theory Ramifications **15**, 413 (2006).
11. Y. Zhang, L.H. Kauffman and M.L. Ge, J.Quant. Inform. **3**, 669 (2005).
12. A. Shapere and F. Wilczek (ed), Geometric Phases in Physics, World Scientific, Singapore, 1986.
13. A. Osterloh, L. Amico, G. Falci and R. Fazio, Nature **416**, 608 (2002).
14. G. Vidal, J. I. Latorre, E. Rico and A. Kitaev, Phys. Rev. Lett. **90**, 227902 (2003); J. I. Latorre, E. Rico, and G. Vidal, Quant. Inform. Comput. **4**, 48 (2004).
15. L. A. Wu, S. Bandyopadhyay, M. S. Sarandy and D. A. Lidar, Phys. Rev. A **72**, 032309 (2005).

16. J. Anders and V. Vedral, e-print quant-ph/0610268.
17. J. L. Birman and S. Q Zhou, Quantum phase transitions and contraction of dynamical symmetry, to be published.
18. Y. Zhang, E. C. Rowell, Y. S. Wu, Z. Wang and M. L. Ge, e-print arXiv:0706.1761.
19. J. L. Chen, K. Xue and M. L. Ge, Phys. Rev. A **76**, 042324 (2007).
20. W. Zhang, D. Feng and R. Gilmore, Rev. Mod. Phys. **62**, 867 (1990).
    S. Chaturvedi, M.S. Sriram and V. Srinivasan, J. Phys. A**70** L1091 (1987).
21. S.D. Sarma, M. Freedman, C. Nayak, S.H. Simon and A. Stern, Arxiv:0707.1899.
    J. Preskill, http://www.theory.caltech.edu/~preskill/ph219/topological.pdf.
22. L.H. Kauffman, Knots and Physics, World Scientific, Singapore, 1991.

On a Conjecture of Milnor about Volume of Simplexes*

Ren Guo
Department of Mathematics, Rutgers University,
Piscataway, NJ, 08854
Email: renguo@math.rutgers.edu

Feng Luo
The Center of Mathematical Science, Zhejiang University,
Hangzhou, China, 310027

Department of Mathematics, Rutgers University,
Piscataway, NJ, 08854
Email: fluo@math.rutgers.edu

Dedicated to the memory of Xiao-Song Lin.

We establish the second part of Milnor's conjecture on the volume of simplexes in hyperbolic and spherical spaces. A characterization of the closure of the space of the angle Gram matrices of simplexes is also obtained.

1. Introduction

Milnor's conjecture

In,[5] John Milnor conjectured that the volume of a hyperbolic or spherical $n-$simplex, considered as a function of the dihedral angles, can be extended continuously to the degenerated simplexes. Furthermore, he conjectured that the extended volume function is non-zero except in the closure of the space of Euclidean simplexes. The first part of the conjecture on the continuous extension was established in[4] ([7] has a new proof of it which

*2000 *Mathematics Subject Classification.* 53C65, 53A35
Key words and phrases. constant curvature space, simplex, volume, angle Gram matrix.
This paper has been accepted for publication in the Journal of Differential Geometry and will appear in the journal. The authors thank the editor Professor Zhenghan Wang for inviting them to contribute it to the proceedings. This paper is published by permission from the Journal of Differential Geometry and the International Press.

generalizes to many polytopes). The purpose of the paper is to establish the second part of Milnor's conjecture.

To state the result, let us begin with some notations and definitions. Given an n−simplex in a spherical, hyperbolic or Euclidean space with vertices $u_1, ..., u_{n+1}$, the i-th codimension-1 face is defined to be the $(n-1)$-simplex with vertices $u_1, ..., u_{i-1}, u_{i+1}, ..., u_{n+1}$. The dihedral angle between the i-th and j-th codimension-1 faces is denoted by θ_{ij}. As a convention, we define $\theta_{ii} = \pi$ and call the symmetric matrix $A = [-\cos(\theta_{ij})]_{(n+1)\times(n+1)}$ the *angle Gram matrix* of the simplex. It is well known that the angle Gram matrix determines a hyperbolic or spherical n-simplex up to isometry and Euclidean n-simplex up to similarity. Let $\mathfrak{X}_{n+1}, \mathfrak{Y}_{n+1}, \mathfrak{Z}_{n+1}$ in $\mathbf{R}^{(n+1)\times(n+1)}$ be the subsets of $(n+1)\times(n+1)$ symmetric matrices corresponding to the angle Gram matrices of spherical, hyperbolic, or Euclidean n−simplexes respectively.

The volume of an n-simplex can be expressed in terms of the angle Gram matrix by the work of Aomoto,[1] Kneser[2] and Vinberg.[8] Namely, for a spherical or hyperbolic n-simplex σ^n with angle Gram matrix A, the volume V is

$$V(A) = \mu_n^{-1}\sqrt{|\det(ad(A))|}\int_{\mathbf{R}^{n+1}_{\geq 0}} e^{-x^t ad(A) x}dx, \tag{1}$$

where $\mathbf{R}^{n+1}_{\geq 0} = \{(x_1, ..., x_{n+1})|\ x_i \geq 0\}$, the constant $\mu_k = \int_0^\infty x^k e^{-x^2}dx$ and $ad(A)$ is the adjoint matrix of A. In,[4] it is proved that the volume function $V : \mathfrak{X}_{n+1} \cup \mathfrak{Y}_{n+1} \to \mathbf{R}$ can be extended continuously to the closure $\overline{\mathfrak{X}}_{n+1} \cup \overline{\mathfrak{Y}}_{n+1}$ in $\mathbf{R}^{(n+1)\times(n+1)}$. The main result of this paper, which verifies the second part of Milnor's conjecture, is the following theorem.

Theorem 1.1. *The extended volume function V on the closure $\overline{\mathfrak{X}}_{n+1} \cup \overline{\mathfrak{Y}}_{n+1}$ in $\mathbf{R}^{(n+1)\times(n+1)}$ vanishes at a point A if and only if A is in the closure $\overline{\mathfrak{Z}}_{n+1}$.*

A characterization of angle Gram matrices

We will use the following conventions. Given a real matrix $A = [a_{ij}]$, we use $A \geq 0$ to denote $a_{ij} \geq 0$ for all i, j and $A > 0$ to denote $a_{ij} > 0$ for all i, j. A^t is the transpose of A. We use $ad(A)$ to denote the adjoint matrix of A. The diagonal $k \times k$ matrix with diagonal entries $(x_1, ...x_k)$ is denoted by $diag(x_1, ...x_k)$. A characterization of the angle Gram matrices in $\mathfrak{X}_{n+1}, \mathfrak{Y}_{n+1}$ or $\mathfrak{Z}_{n+1}$ is known by the work of[3] and.[5]

Proposition 1.1 (,[35]). *Given an $(n+1)\times(n+1)$ symmetric matrix $A = [a_{ij}]$ with $a_{ii} = 1$ for all i, then*

(a) *$A \in \mathcal{Z}_{n+1}$ if and only if $det(A) = 0$, $ad(A) > 0$ and all principal $n\times n$ submatrices of A are positive definite,*
(b) *$A \in \mathcal{X}_{n+1}$ if and only if A is positive definite,*
(c) *$A \in \mathcal{Y}_{n+1}$ if and only if $det(A) < 0$, $ad(A) > 0$ and all principal $n\times n$ submatrices of A are positive definite.*

In particular, all off-diagonal entries a_{ij} have absolute values less than 1, i.e.,$|a_{ij}| < 1$ for $i \neq j$.

The following gives a characterization of matrices in $\overline{\mathcal{X}}_{n+1}, \overline{\mathcal{Y}}_{n+1}$ and $\overline{\mathcal{Z}}_{n+1}$ in $\mathbf{R}^{(n+1)\times(n+1)}$.

Theorem 1.2. *Given an $(n+1)\times(n+1)$ symmetric matrix $A = [a_{ij}]$ with $a_{ii} = 1$ for all i, then*

(a) *$A \in \overline{\mathcal{Z}}_{n+1}$ if and only if $det(A) = 0$, A is positive semi-definite, and there exists a principal $(k+1)\times(k+1)$ submatrix B of A so that $det(B) = 0, ad(B) \geq 0$ and $ad(B) \neq 0$,*
(b) *$A \in \overline{\mathcal{X}}_{n+1}$ if and only if either A is in $\mathcal{X}_{n+1}$ or there exists a diagonal matrix $D = diag(\varepsilon_1, ..., \varepsilon_{n+1})$ where $\varepsilon_i = 1$ or -1 for each $i = 1, ..., n+1$, such that $DAD \in \overline{\mathcal{Z}}_{n+1}$,*
(c) *$A \in \overline{\mathcal{Y}}_{n+1}$ if and only if either $A \in \overline{\mathcal{Z}}_{n+1}$ or $det(A) < 0$, $ad(A) \geq 0$ and all principal $n\times n$ submatrices of A are positive semi-definite.*

The paper is organized as follows. In section 2, we characterize normal vectors of degenerated Euclidean simplexes. In section 3, we characterize angle Gram matrices of degenerated hyperbolic simplexes. Theorem 1 is proved in section 4 and Theorem 3 is proved in section 5.

Acknowledgment

We would like to thank the referee for a very careful reading of the manuscript and for his/her nice suggestions on improving the exposition.

2. Normal vectors of Euclidean simplexes

As a convention, all vectors in $\mathbf{R}^m$ are column vectors and the standard inner product in $\mathbf{R}^m$ is denoted by $u \cdot v$. In the sequel, for a non-zero vector $w \in \mathbf{R}^n$, we call the set $\{x \in \mathbf{R}^n | w \cdot x \geq 0\}$ a *closed half space*, and the set $\{x \in \mathbf{R}^n | w \cdot x > 0\}$ an *open half space*. Define

$\mathcal{E}_{n+1} = \{(v_1, ..., v_{n+1}) \in (\mathbf{R}^n)^{n+1} | v_1, ..., v_{n+1}$ form unit outward normal vectors to the codimension-1 faces of a Euclidean n-simplex $\}$. Following Milnor,[5] a matrix is called *unidiagonal* if its diagonal entries are 1. An $(n+1) \times (n+1)$ symmetric unidiagonal matrix A is in $\mathcal{Z}_{n+1}$ if and only if $A = [v_i \cdot v_j]$ for some point $(v_1, ..., v_{n+1}) \in \mathcal{E}_{n+1}$ (this is proved in,[35]). We claim that an $(n+1) \times (n+1)$ symmetric unidiagonal matrix A is in $\overline{\mathcal{Z}}_{n+1}$ if and only if $A = [v_i \cdot v_j]$ for some point $(v_1, ..., v_{n+1})$ in the closure $\overline{\mathcal{E}}_{n+1}$ in $(\mathbf{R}^n)^{n+1}$. Indeed, if $A = [v_i \cdot v_j]$ for some point $(v_1, ..., v_{n+1}) \in \overline{\mathcal{E}}_{n+1}$, then there is a sequence $(v_1^{(m)}, ..., v_{n+1}^{(m)}) \in \mathcal{E}_{n+1}$ converging to $(v_1, ..., v_{n+1})$. We have a sequence of matrices $A^{(m)} = [v_i^{(m)} \cdot v_j^{(m)}] \in \mathcal{Z}_{n+1}$ converging to A. Conversely if $A \in \overline{\mathcal{Z}}_{n+1}$, then there is a sequence of matrices $A^{(m)} \in \mathcal{Z}_{n+1}$ converging to A. Write $A^{(m)} = [v_i^{(m)} \cdot v_j^{(m)}]$, where $(v_1^{(m)}, ..., v_{n+1}^{(m)}) \in \mathcal{E}_{n+1}$. Since $v_i^{(m)}$ has norm 1 for all i, m, by taking subsequence, we may assume $\lim_{m\to\infty}(v_1^{(m)}, ..., v_{n+1}^{(m)}) = (v_1, ..., v_{n+1}) \in \overline{\mathcal{E}}_{n+1}$ so that $A = [v_i \cdot v_j]$.

A geometric characterization of elements in $\mathcal{E}_{n+1}$ was obtained in.[3] For completeness, we include a proof here.

Lemma 2.1. *A collection of unit vectors $(v_1, ..., v_{n+1}) \in (\mathbf{R}^n)^{n+1}$ is in $\mathcal{E}_{n+1}$ if and only if one of the following conditions is satisfied.*

(4.1) The vectors $v_1, ..., v_{n+1}$ are not in any closed half-space.
(4.2) Any n vectors of $v_1, ..., v_{n+1}$ are linear independent and the linear system $\sum_{i=1}^{n+1} a_i v_i = 0$ has a solution $(a_1, ..., a_n)$ so that $a_i > 0$ for all $i = 1, ..., n+1$.

Proof. (4.2) $\Rightarrow$ (4.1). Suppose otherwise, $v_1, ..., v_{n+1}$ are in a closed half-space, i.e., there is a non-zero vector $w \in \mathbf{R}^n$ so that $w \cdot v_i \geq 0, i = 1, ..., n+1$. Let $a_1, ..., a_{n+1}$ be the positive numbers given by (4.2) so that $\sum_{i=1}^{n+1} a_i v_i = 0$. Then

$$0 = w \cdot (\sum_{i=1}^{n+1} a_i v_i) = \sum_{i=1}^{n+1} a_i (w \cdot v_i).$$

But by the assumption $a_i > 0$, $w \cdot v_i \geq 0$ for all i. Thus $w \cdot v_i = 0$ for all i. This means that $v_1, ..., v_{n+1}$ lie in the $(n-1)$-dimensional subspace perpendicular to w. It contradicts the assumption in (4.2) that any n vectors of $v_1, ..., v_{n+1}$ are linear independent.

(4.1) $\Rightarrow$ (4.2). To see that any n vectors of $v_1, ..., v_{n+1}$ are linear independent, suppose otherwise, some n vectors of $v_1, ..., v_{n+1}$ are linear dependent. Therefore there is an $(n-1)$-dimensional hyperplane containing these n vectors. Then $v_1, ..., v_{n+1}$ are contained in one of the two closed half spaces bounded by the hyperplane. It contradicts to the assumption of (4.1).

Since $v_1, ..., v_{n+1}$ are linear dependent, and any n of them are linear independent, we can find real numbers $a_i \neq 0$ for all i such that $\sum_{i=1}^{n+1} a_i v_i = 0$. For any $i \neq j$, let H_{ij} be the $(n-1)$-dimensional hyperplane spanned by the $n-1$ vectors $\{v_1, ..., v_{n+1}\} \setminus \{v_i, v_j\}$ and $u \in \mathbf{R}^n - \{0\}$ be a vector perpendicular to H_{ij}. We have

$$0 = u \cdot (\sum_{i=1}^{n+1} a_i v_i) = a_i(u \cdot v_i) + a_j(u \cdot v_j).$$

By the assumption of (4.1), v_i and v_j must lie in the different sides of H_{ij}. Thus $u \cdot v_i$ and $u \cdot v_j$ have different sign. This implies that a_i and a_j have the same sign. Hence we can make $a_i > 0$ for all i.

$\mathcal{E}_{n+1} \Leftrightarrow$ (4.1). We will show $(v_1, ..., v_{n+1}) \in \mathcal{E}_{n+1}$ if and only if the condition (4.1) holds. In fact, given an n-dimensional Euclidean simplex σ, let S^{n-1} be the sphere inscribed to σ. We may assume after a translation and a scaling that S^{n-1} is the unit sphere centered at the origin. Then the unit vectors $v_1, ..., v_{n+1}$ are the tangent points of S^{n-1} to the codimension-1 faces of σ. The tangent planes to S^{n-1} at v_i's bound a compact region (the Euclidean simplex σ) containing the origin if and only if the tangent points $v_1, ..., v_{n+1}$ are not in any closed hemisphere of S^{n-1}. □

Lemma 2.2. *A collection of unit vectors $(v_1, ..., v_{n+1}) \in (\mathbf{R}^n)^{n+1}$ is in $\overline{\mathcal{E}}_{n+1}$ if and only if one of the following conditions is satisfied:*

(5.1) The vectors $v_1, ..., v_{n+1}$ are not in any open half-space.
(5.2) The linear system $\sum_{i=1}^{n+1} a_i v_i = 0$ has a nonzero solution $(a_1, ..., a_{n+1})$ so that $a_i \geq 0$ for all $i = 1, ..., n+1$.

Proof. $\overline{\mathcal{E}}_{n+1} \Rightarrow$(5.1). To see that elements in $\overline{\mathcal{E}}_{n+1}$ satisfy (5.1), if $(v_1, ..., v_{n+1}) \in \overline{\mathcal{E}}_{n+1}$, there is a family of $(v_1^{(m)}, ..., v_{n+1}^{(m)}) \in \mathcal{E}_{n+1}$ converging to $(v_1, ..., v_{n+1})$. Since vectors $v_1^{(m)}, ..., v_{n+1}^{(m)}$ are not in any closed half-space for any m, by continuity, vectors $v_1, ..., v_{n+1}$ are not in any open half-space.

(5.1)$\Rightarrow$(5.2). Consider the linear map

$$f : \mathbf{R}^n \longrightarrow \mathbf{R}^{n+1}$$

$$w \longmapsto f(w) = [v_1, v_2, ..., v_{n+1}]^t w = \begin{pmatrix} v_1 \cdot w \\ v_2 \cdot w \\ \vdots \\ v_{n+1} \cdot w \end{pmatrix}.$$

Statement (5.1) says that

$$\begin{aligned}\emptyset &= \{w \in \mathbf{R}^n | v_i \cdot w > 0, i = 1, ..., n+1\}\\ &= \{w \in \mathbf{R}^n | f(w) > 0\}\\ &= f(\mathbf{R}^n) \cap \mathbf{R}^{n+1}_{>0}.\end{aligned}$$

Since $f(\mathbf{R}^n)$ and $\mathbf{R}^{n+1}_{>0}$ are convex and disjoint, by the separation theorem for convex sets, there is a vector $a = (a_1, ..., a_{n+1})^t$ satisfying the conditions (i) and (ii) below.

(i) For all $u \in \mathbf{R}^{n+1}_{>0}$,

$$a \cdot u > 0.$$

and

(ii) For all $w \in \mathbf{R}^n$,

$$0 \geq a \cdot f(w) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n+1} \end{pmatrix} \cdot \begin{pmatrix} v_1 \cdot w \\ v_2 \cdot w \\ \vdots \\ v_{n+1} \cdot w \end{pmatrix} = (\sum_{i=1}^{n+1} a_i v_i) \cdot w.$$

Condition (i) implies that $a_i \geq 0$, for $i = 1, ..., n+1$ and $a \neq 0$. Condition (ii) implies $\sum_{i=1}^{n+1} a_i v_i = 0$. Thus (5.2) holds.

(5.2)$\Rightarrow$ $\overline{\mathcal{E}}_{n+1}$. To see that a point $(v_1, ..., v_{n+1})$ satisfying (5.2) is in $\overline{\mathcal{E}}_{n+1}$, we show that in any ε-neighborhood of $(v_1, ..., v_{n+1})$ in $(\mathbf{R}^n)^{n+1}$, there is a point $(v_1^\varepsilon, ..., v_{n+1}^\varepsilon) \in \mathcal{E}_{n+1}$.

Let $\mathcal{N}_k$ be the set of $(v_1, ..., v_k)$ such that $v_i \in \mathbf{R}^{k-1}, |v_i| = 1$ for all i and $\sum_{i=1}^k a_i v_i = 0$ has a nonzero solution $(a_1, ..., a_k)$ with $a_i \geq 0$ for all i. The goal is to prove $\mathcal{N}_{n+1} \subset \overline{\mathcal{E}}_{n+1}$. We achieve this by induction on n. It is obvious that $\mathcal{N}_2 \subset \overline{\mathcal{E}}_2$. Assume that $\mathcal{N}_n \subset \overline{\mathcal{E}}_n$ holds.

For a point $(v_1, ..., v_{n+1}) \in \mathcal{N}_{n+1}$, if any n vectors of $v_1, ..., v_{n+1}$ are linear independent, then each entry a_i of the non-zero solution of the linear system $\sum_{i=1}^{n+1} a_i v_i = 0, a_i \geq 0, i = 1, ..., n+1$ must be nonzero. Thus $a_i > 0$ for all i and $(v_1, ..., v_{n+1})$ satisfies (4.2), therefore it is in $\mathcal{E}_{n+1}$.

In the remain case, without loss of generality, we assume that $v_1, ..., v_n$ are linear dependent. We may assume after a change of coordinates that $v_i \in \mathbf{R}^{n-1} = \mathbf{R}^{n-1} \times \{0\} \subset \mathbf{R}^n$, for $i = 1, ..., n$, and $v_{n+1} = (u_{n+1}\cos(\theta), \sin(\theta))^t$, where $0 \leq \theta \leq \frac{\pi}{2}$ and $|u_{n+1}| = 1$.

We claim that there exists some $1 \leq i \leq n+1$ such that $(v_1, ..., \widehat{v_i}, ..., v_{n+1}) \in \mathcal{N}_n$, where $\widehat{x}$ means deleting the element x.

Case 1. If $\theta > 0$ i.e., v_{n+1} is not in $\mathbf{R}^{n-1}$, consider the nonzero solution of the linear system $\sum_{i=1}^{n+1} a_i v_i = 0, a_i \geq 0, i = 1, ..., n+1$. The last coordinate gives $a_1 0 + ... + a_n 0 + a_{n+1}\sin(\theta) = 0$, which implies $a_{n+1} = 0$. This means $(a_1, ..., a_n) \neq 0$, i.e., $(v_1, ..., v_n) \in \mathcal{N}_n$.

Case 2. If $\theta = 0$ i.e., $v_{n+1} \in \mathbf{R}^{n-1}$, then the dimension of the solution space $W = \{(a_1, ..., a_{n+1})^t \in \mathbf{R}^{n+1} | \sum_{i=1}^{n+1} a_i v_i = 0\}$ is at least 2. Since $(v_1, ..., v_{n+1}) \in \mathcal{N}_{n+1}$, the intersection $W \cap \mathbf{R}^{n+1}_{\geq 0} - \{(0, ..., 0)\}$ is nonempty. The vector space W must intersect the boundary of the cone $\mathbf{R}^{n+1}_{\geq 0} - \{(0, ..., 0)\}$. Let $(a_1, ..., a_{n+1})$ be a point in both W and the boundary of the cone $\mathbf{R}^{n+1}_{\geq 0} - \{(0, ..., 0)\}$. Then there is some $a_i = 0$. Then $\{v_1(t), ..., \widehat{v_i}(t), ..., v_{n+1}(t)\} \in \mathcal{N}_n$.

By the above discussion, without lose of generality, we may assume that $(v_1, ..., v_n) \in \mathcal{N}_n$. By the induction assumption $\mathcal{N}_n \subset \overline{\mathcal{E}}_n$, i.e., in any $\frac{\epsilon}{2}$-neighborhood of $(v_1, ..., v_n)$, we can find a point $(u_1, ..., u_n) \in \mathcal{E}_n$, where $u_i \in \mathbf{R}^{n-1}$ for all i. Recall we write $v_{n+1} = (u_{n+1}\cos(\theta), \sin(\theta))^t$. Let us define a continuous family of $n+1$ unit vectors $v_1(t), ..., v_{n+1}(t)$ by setting

$$v_i(t) = (u_i \cos(t^2), -\sin(t^2))^t, 1 \leq i \leq n,$$

$$v_{n+1}(t) = (u_{n+1}\cos(\theta + t), \sin(\theta + t))^t.$$

We claim that there is a point $(v_1(t), ..., v_{n+1}(t)) \in \mathcal{E}_{n+1}$ for small $t > 0$ within $\frac{\epsilon}{2}$-neighborhood of $((u_1, 0)^t, ..., (u_n, 0)^t, v_{n+1})$. By triangle inequality, this point is within ε-neighborhood of $(v_1, ..., v_{n+1})$. We only need to check that $(v_1(t), ..., v_{n+1}(t)) \in \mathcal{E}_{n+1}$ for sufficiently small $t > 0$ by verifying the condition (4.2).

To show any n vectors of $v_1(t), ..., v_{n+1}(t)$ are linear independent, it is equivalent to show that

$$\det[v_1(t), ..., \widehat{v_i}(t), ..., v_{n+1}(t)] \neq 0$$

for each $i = 1, ..., n+1$.

First,

$$\det[v_1(t), ..., v_n(t)] = \det \begin{bmatrix} u_1 \cos(t^2) & u_2 \cos(t^2) & ... & u_n \cos(t^2) \\ -\sin(t^2) & -\sin(t^2) & ... & -\sin(t^2) \end{bmatrix}_{n \times n}$$
$$= -\sin(t^2)\cos(t^2)^{n-1} \det \begin{bmatrix} u_1 & u_2 & ... & u_n \\ 1 & 1 & ... & 1 \end{bmatrix}.$$

To see the determinant is nonzero, suppose there are real numbers $a_1, ..., a_n$ such that $\sum_{i=1}^n a_i (u_1, 1)^t = 0$. Then we have $\sum_{i=1}^n a_i u_i = 0$ and $\sum_{i=1}^n a_i = 0$. By assumption $(u_1, ..., u_n) \in \mathcal{E}_n$, we know either $a_i = 0$ for all i or $a_i \neq 0$ and have the same sign for all i. Hence $\sum_{i=1}^n a_i = 0$ implies $a_i = 0$ for all i. Thus the vectors $(u_1, 1)^t, ..., (u_n, 1)^t$ are linear independent. Hence $\det[v_1(t), ..., v_n(t)] \neq 0$ for $t \in (0, \sqrt{\frac{\pi}{2}}]$.

Second, we calculate the determinant of the matrix whose columns are $v_{n+1}(t)$ and some $n-1$ vectors of $v_1(t), ..., v_n(t)$. Without loss of generality, consider

$$f(t) = \det[v_2(t), ..., v_n(t), v_{n+1}(t)]$$

$$= \det \begin{bmatrix} u_2 \cos(t^2) & \dots & u_n \cos(t^2) & u_{n+1} \cos(\theta + t) \\ -\sin(t^2) & \dots & -\sin(t^2) & \sin(\theta + t) \end{bmatrix}$$

If $\theta > 0$, by the assumption $(u_1, ..., u_n) \in \mathcal{E}_n$, then

$$f(0) = \det \begin{bmatrix} u_2 & \dots & u_n & u_{n+1} \cos(\theta) \\ 0 & \dots & 0 & \sin(\theta) \end{bmatrix} = \sin(\theta) \det[u_2, ..., u_n] \neq 0.$$

It implies $f(t) \neq 0$ holds for sufficiently small $t > 0$.

If $\theta = 0$, then $f(0) = 0$. By expanding the determinant,

$$f(t) = -\sin(t^2)g(t) + \sin(t) \det[u_2 \cos(t^2), ..., u_n \cos(t^2)],$$

for some function $g(t)$, therefore $f'(0) = \det(u_2, ..., u_n) \neq 0$. Hence $f(t) \neq 0$ holds for sufficiently small $t > 0$.

Next, let

$$a_i(t) = (-1)^{i-1} \det[v_1(t), ..., \widehat{v}_i(t), ..., v_{n+1}(t)], 1 \leq i \leq n+1,$$

then $\sum_{i=1}^{n+1} a_i(t)v_i(t) = 0$. Since $\det[v_1(0), ..., v_n(0)] = 0$, we have $a_{n+1}(0) = 0$. This shows that $\sum_{i=1}^{n} a_i(0)v_i(0) = 0$, therefore $\sum_{i=1}^{n} a_i(0)u_i = 0$. By the assumption $(u_1, ..., u_n) \in \mathcal{E}_n$, we obtain $a_i(0) \cdot a_j(0) > 0$ for $0 \leq i, j \leq n$. By the continuity we obtain $a_i(t) \cdot a_j(t) > 0$ for $0 \leq i, j \leq n$, for sufficient small $t > 0$. Consider the last coordinate of $\sum_{i=1}^{n+1} a_i(t)v_i(t) = 0$ we obtain

$$-\sin(t^2) \sum_{i=1}^{n} a_i(t) + \sin(\theta + t)a_{n+1}(t) = 0.$$

Thus $a_{n+1}(t)$ has the same sign as that of $a_i(t)$. Thus $(a_1(t), ..., a_{n+1}(t))$ or $(-a_1(t), ..., -a_{n+1}(t))$ is a solution required in condition (4.2). □

3. Degenerate hyperbolic simplexes

Let $\mathbf{R}^{n,1}$ be the Minkowski space which is $\mathbf{R}^{n+1}$ with an inner product $\langle , \rangle$ where

$$\langle (x_1, ..., x_n, x_{n+1})^t, (y_1, ..., y_n, y_{n+1})^t \rangle = x_1 y_1 + ... + x_n y_n - x_{n+1} y_{n+1}.$$

Let $H^n = \{x = (x_1, ..., x_{n+1})^t \in \mathbf{R}^{n,1} | \langle x, x \rangle = -1, x_{n+1} > 0\}$ be the hyperboloid model of the hyperbolic space. The de Sitter space is

$\{x \in \mathbf{R}^{n,1} | \langle x, x \rangle = 1\}$. For a hyperbolic simplex σ in H^n, the *center* and the *radius* of the simplex σ are defined to be the center and radius of its inscribed ball.

Lemma 3.1. *For an n-dimensional hyperbolic simplex $\sigma \in H^n$ with center $e_{n+1} = (0, ..., 0, 1)^t$, its unit outward normal vectors in the de Sitter space are in a compact set independent of σ.*

Proof. Let $v_1, ..., v_{n+1}$ be the unit outward normal vectors of σ, i.e.,

$$\sigma = \{x \in H^n | \langle x, v_i \rangle \leq 0 \text{ and } \langle v_i, v_i \rangle = 1 \text{ for all } i\}.$$

Let $v_i^\perp$ be the totally geodesic hyperplane in H^n containing the $(n-1)-$dimensional face of σ perpendicular to v_i for each $i = 1, ..., n+1$. The radius of σ is the distance from the center e_{n+1} to $v_i^\perp$ for any $i = 1, ..., n+1$ which is equal to $\sinh^{-1}(|\langle e_{n+1}, v_i \rangle|)$ (see for instance[9] p26). It is well known that the volume of an $n-$dimensional hyperbolic simplex is bounded by the volume of the $n-$dimensional regular ideal hyperbolic simplex which is finite (see for instance[6] p539). It implies that the radius of a hyperbolic simplex σ is bounded from above by a constant independent of σ. Hence $\langle e_{n+1}, v_i \rangle^2$ is bounded from above by a constant c_n independent of σ for any $i = 1, ..., n+1$. It follows that $v_1, ..., v_{n+1}$ are in the compact set

$$\begin{aligned} X_n &= \{x = (x_1, ..., x_{n+1})^t | \langle x, x \rangle = 1, \langle e_{n+1}, x \rangle^2 \leq c_n\} \\ &= \{x = (x_1, ..., x_{n+1})^t | x_1^2 + ... + x_n^2 = x_{n+1}^2 + 1, x_{n+1}^2 \leq c_n\} \end{aligned}$$

independent of σ. □

Lemma 3.2. *If $A \in \overline{\mathcal{Y}}_{n+1}$ and* $\det(A) = 0$, *then $A \in \overline{\mathcal{Z}}_{n+1}$.*

Proof. Let $A^{(m)}$ be a sequence of angle Gram matrixes in $\mathcal{Y}_{n+1}$ converging to A. By Proposition 2 (c), for any m, all principal $n \times n$ submatrices of $A^{(m)}$ are positive definite. Thus all principal $n \times n$ submatrices of A are positive semi-definite. Since $\det(A) = 0$, we see that A is positive semi-definite.

Let $\sigma^{(m)}$ be the $n-$dimensional hyperbolic simplex in the hyperboloid model H^n whose angle Gram matrix is $A^{(m)}$ and whose center is $e_{n+1} = (0, ..., 0, 1)^t$. By Lemma 6, its unit outward normal vector $v_i^{(m)}$ is in a compact set. Thus by taking a subsequence, we may assume $(v_1^{(m)}, ..., v_{n+1}^{(m)})$ converges to $(v_1, ..., v_{n+1})$ with $\langle v_i, v_i \rangle = 1$. Since $A^{(m)} = [\langle v_i^{(m)}, v_j^{(m)} \rangle]$ and $A^{(m)}$ converges to A, we obtain

$$A = [\langle v_i, v_j \rangle] = [v_1, ..., v_{n+1}]^t S [v_1, ..., v_{n+1}],$$

where S is the diagonal matrix $diag(1, ..., 1, -1)$.

Since $\det(A) = 0$, the vectors $v_1, ..., v_{n+1}$ are linear dependent. Assume that the vectors $v_1, ..., v_{n+1}$ span a k-dimensional subspace W of $\mathbf{R}^{n,1}$, where $k \leq n$.

For any vector $x \in W$, write $x = \sum_{i=1}^{n+1} x_i v_i$. Then

$$\begin{aligned}\langle x, x\rangle &= (x_1, ..., x_{n+1})[v_1, ..., v_{n+1}]^t S[v_1, ..., v_{n+1}](x_1, ..., x_{n+1})^t \\ &= (x_1, ..., x_{n+1})A(x_1, ..., x_{n+1})^t \\ &\geq 0\end{aligned}$$

due to the fact that A is positive semi-definite.

Now for any $x, y \in W$, the inequality $\langle x + ty, x + ty\rangle \geq 0$ for any $t \in \mathbf{R}$ implies the Schwartz inequality

$$\langle x, y\rangle^2 \leq \langle x, x\rangle\langle y, y\rangle.$$

To prove that $A \in \overline{Z}_{n+1}$, we consider the following two possibilities.

Case 1. If $\langle x, x\rangle > 0$ holds for any non-zero $x \in W$, then the Minkowski inner product restricted on W is positive definite. Since the Minkowski inner product restricted on $\mathbf{R}^k = \mathbf{R}^k \times 0 \subset \mathbf{R}^{n,1}$ is positive definite, by Witt's theorem, there is an isometry γ of $\mathbf{R}^{n,1}$ sending W to $\mathbf{R}^k$ (see[9] p14-p15). By replacing $v_i^{(m)}$ by $\gamma(v_i^{(m)})$ for each i and m, we may assume that $v_1, ..., v_{n+1}$ are contained in $\mathbf{R}^k$. Thus $\langle v_i, v_i\rangle = v_i \cdot v_j$ for all i, j. Therefore

$$A = [v_1, ..., v_{n+1}]^t S[v_1, ..., v_{n+1}] = [v_1, ..., v_{n+1}]^t[v_1, ..., v_{n+1}].$$

To show $A \in \overline{Z}_{n+1}$, by Lemma 5, we only need to show that $v_1, ..., v_{n+1}$ are not contained in any open half space of $\mathbf{R}^k$. This is the same as that $v_1, ..., v_{n+1}$ are not contained in any open half space of $\mathbf{R}^n$.

Suppose otherwise, there exists a vector $w \in \mathbf{R}^k$ such that $v_i \cdot w > 0$ for all i. Thus $\langle v_i, w\rangle = v_i \cdot w > 0$. By taking m large enough, we obtain $\langle v_i^{(m)}, w\rangle > 0$ for all i.

It is well known that for the unit normal vectors $v_i^{(m)}$ of a compact hyperbolic simplex in H^n, the conditions $\langle v_i^{(m)}, w\rangle > 0$ for all i implies $\langle w, w\rangle < 0$. But this contradicts the assumption that $w \in \mathbf{R}^k$ which implies $\langle w, w\rangle \geq 0$.

Case 2. If there exists some non-zero vector $x_0 \in W$ such that $\langle x_0, x_0\rangle = 0$, then by the Schwartz inequality we have

$$\langle x_0, y\rangle^2 \leq \langle x_0, x_0\rangle\langle y, y\rangle = 0$$

for any $y \in W$. Thus $\langle x_0, y\rangle = 0$ for any $y \in W$. This implies that the subspace W is contained in $x_0^{\perp}$, the orthogonal complement of x_0.

Since the vector $u = (0, ..., 0, 1, 1)^t \in \mathbf{R}^{n,1}$ satisfies $\langle u, u \rangle = 0$, there is an isometry γ of $\mathbf{R}^{n,1}$ sending x_0 to u. Thus γ sends $x_0^\perp$ to $u^\perp$. By replacing $v_i^{(m)}$ by $\gamma(v_i^{(m)})$ for each i and m, we may assume that $v_1, ..., v_{n+1}$ are contained in $u^\perp$.

For any i, since $\langle v_i, u \rangle = \langle v_i, (0, ..., 0, 1, 1)^t \rangle = 0$, we can write v_i as

$$v_i = w_i + a_i u$$

for some $w_i \in \mathbf{R}^{n-1}$ and $a_i \in \mathbf{R}$. Since $\langle w_i, u \rangle = 0$, thus $\langle v_i, v_j \rangle = w_i \cdot w_j$ for all i, j. Therefore

$$\begin{aligned} A &= [v_1, ..., v_{n+1}]^t S [v_1, ..., v_{n+1}] \\ &= [w_1, ..., w_{n+1}]^t [w_1, ..., w_{n+1}]. \end{aligned}$$

To show $A \in \overline{\mathfrak{Z}}_{n+1}$, by Lemma 5, we only need to show that $w_1, ..., w_{n+1}$ are not contained in any open half space of $\mathbf{R}^{n-1}$ which is equivalent to that $w_1, ..., w_{n+1}$ are not contained in any open half space of $\mathbf{R}^n$.

Suppose otherwise, there exists a vector $w \in \mathbf{R}^{n-1}$ such that $w_i \cdot w > 0$ for all i. Then

$$\langle v_i, w \rangle = \langle w_i, w \rangle + \langle (0, ..., 0, a_i, a_i)^t, w \rangle = w_i \cdot w + 0 > 0$$

for all i. By taking m large enough, we obtain $\langle v_i^{(m)}, w \rangle > 0$ for all i. By the same argument above, it is a contradiction. $\square$

4. Proof of theorem 1

Spherical case

We begin with a brief review of the relevant result in.[4] For any positive semi-definite symmetric matrix A, there exists a unique positive semi-definite symmetric matrix $\sqrt{A}$ so that $(\sqrt{A})^2 = A$. It is well know that the map $A \longmapsto \sqrt{A}$ is continuous on the space of all positive semi-definite symmetric matrices.

Suppose $A \in \mathfrak{X}_{n+1} = \{A = [a_{ij}] \in \mathbf{R}^{(n+1)\times(n+1)} |\ A^t = A$, all $a_{ii} = 1$, A is positive definite$\}$, the space of the angle Gram matrices of spherical simplexes (by the Proposition 2). By making a change of variables, the Aomoto-Kneser-Vinberg formula (1.1) is equivalent to

$$V(A) = \mu_n^{-1} \int_{\mathbf{R}^{n+1}} \chi(\sqrt{A}x) e^{-x^t x} dx, \tag{1}$$

where χ is the characteristic function of the set $\mathbf{R}^{n+1}_{\geq 0}$ in $\mathbf{R}^{n+1}$. It is proved in[4] that volume formula (2) still holds for any matrix in $\overline{\mathfrak{X}}_{n+1}$=$\{A = [a_{ij}] \in \mathbf{R}^{(n+1)\times(n+1)} |\ A^t = A$, all $a_{ii} = 1$, A is positive semi-definite$\}$.

Suppose $V(A) = 0$, by formula (2), we see the function $\chi \circ h : \mathbf{R}^{n+1} \to \mathbf{R}$ is zero almost everywhere, where $h : \mathbf{R}^{n+1} \to \mathbf{R}^{n+1}$ is the linear map sending x to $\sqrt{A}x$. Equivalently, the $(n+1)$-dimensional Lebesque measure of $h^{-1}(\mathbf{R}^{n+1}_{>0})$ is zero. We claim $h(\mathbf{R}^{n+1}) \cap \mathbf{R}^{n+1}_{>0} = \emptyset$. For otherwise, $h^{-1}(\mathbf{R}^{n+1}_{>0})$ is a nonempty open subset in $\mathbf{R}^{n+1}$ with positive $(n+1)$-dimensional Lebesque measure. This is a contradiction.

Now let $\sqrt{A} = [v_1, ..., v_{n+1}]^t_{(n+1)\times(n+1)}$, where $v_i \in \mathbf{R}^{n+1}$ is a column vector for each i. First $h(\mathbf{R}^{n+1}) \cap \mathbf{R}^{n+1}_{>0} = \emptyset$ implies that $\det \sqrt{A} = 0$. Therefore $\{v_1, ..., v_{n+1}\}$ are linear dependent. We may assume, after a rotation $r \in O(n+1)$, the vectors $v_1, ..., v_{n+1}$ lie in $\mathbf{R}^n \times \{0\}$. Now

$$\begin{aligned}\emptyset &= h(\mathbf{R}^{n+1}) \cap \mathbf{R}^{n+1}_{>0} \\ &= \{\sqrt{A}w | w \in \mathbf{R}^{n+1}\} \cap \mathbf{R}^{n+1}_{>0} \\ &= \{(v_1 \cdot w, ..., v_{n+1} \cdot w)^t | w \in \mathbf{R}^{n+1}\} \cap \mathbf{R}^{n+1}_{>0}\end{aligned}$$

This shows that there is no $w \in \mathbf{R}^{n+1}$ such that $v_i \cdot w > 0$ for $i = 1, ..., n+1$, i.e., the vectors $v_1, ..., v_{n+1}$ are not in any open half space. By lemma 5, we have $(v_1, ..., v_{n+1}) \in \overline{\mathcal{E}}_{n+1}$, therefore $A = [v_i \cdot v_j] \in \overline{\mathcal{Z}}_{n+1}$.

Hyperbolic case

Let $A \in \overline{\mathcal{Y}}_{n+1}$. If $\det(A) \neq 0$, it is proved in[4] that the volume formula (1.1) still holds for A. In formula (1.1), since $-x^t ad(A)x$ is finite, the integrant $e^{-x^t ad(A)x} > 0$. Therefore the integral $\int_{\mathbf{R}^{n+1}_{\geq 0}} e^{-x^t ad(A)x} dx > 0$. Hence $V(A) > 0$.

It follows that if the extended volume function vanishes at A, then $\det A = 0$. By Lemma 7 , we have $A \in \overline{\mathcal{Z}}_{n+1}$.

5. Proof of theorem 3

Proof of (a)

If $A \in \overline{\mathcal{Z}}_{n+1}$, then $A = [v_i \cdot v_j]$ for some point $(v_1, ..., v_{n+1}) \in \overline{\mathcal{E}}_{n+1}$. By Lemma 5, the linear system $\sum_{i=1}^{n+1} a_i v_i = 0, a_i \geq 0, i = 1, ..., n+1$ has a nonzero solution. Let $(a_1, ..., a_{n+1})$ be a solution with the least number of nonzero entries among all solutions. By rearrange the index, we may assume $a_1 > 0, ..., a_{k+1} > 0, a_{k+2} = ... = a_{n+1} = 0$. We claim $rank[v_1, ..., v_{k+1}] = k$. Otherwise $rank[v_1, ..., v_{k+1}] \leq k-1$, then the dimension of the solution space $W = \{(x_1, ..., x_{k+1})^t \in \mathbf{R}^{k+1} | \sum_{i=1}^{k+1} x_i v_i = 0\}$ is at least 2. Thus $\Omega = W \cap \mathbf{R}^{k+1}_{>0}$ is a nonempty open convex set in W whose dimension is at least 2. Hence Ω contains a boundary point

$(b_1, ..., b_{k+1}) \in \Omega - \{(0, ..., 0)\}$ with some $b_j = 0$, due to $dimW \geq 2$. Now we obtain a solution $(b_1, ...b_{j-1}, 0, b_{j+1}, ..., b_{k+1}, 0, ..., 0)$ which has lesser number of nonzero entries than $(a_1, ..., a_{n+1})$. This is a contradiction.

Let $B = [v_i \cdot v_j]_{(k+1)\times(k+1)}$. Since $rank[v_1, ..., v_{k+1}] = k$, we have $\det(B) = 0$. We claim that $ad(B) \geq 0$ and $ad(B) \neq 0$. This will verify the condition (a) in Theorem 3 for A. Let $ad(B) = [b_{ij}]_{(k+1)\times(k+1)}$. Evidently, due to $rank(B) = k, ad(B) \neq 0$. It remains to prove that $ad(B) \geq 0$. By the construction of B, we see $b_{jj} \geq 0$, for all j. Since $rank[v_1, ..., v_{k+1}] = k$, it follows the dimension of the solution space of $\sum_{i=1}^{k+1} a_i v_i = 0$ is 1. Since $\sum_{i=1}^{k+1} b_{ij} v_j = 0$, $(b_{i1}, ..., b_{ik+1})$ is proportional to $(a_1, ..., a_{k+1})$, where $a_i > 0$ for $1 \leq i \leq k+1$. This shows that if $b_{jj} > 0$, then $b_{ij} > 0$ for all i, if $b_{jj} = 0$, then $b_{ij} = 0$ for all i. This shows $ad(B) \geq 0$.

Conversely, if A is positive semi-definite so that $\det(A) = 0$ and there exists a principal $(k+1)\times(k+1)$ submatrix B so that $\det(B) = 0, ad(B) \geq 0$ and $ad(B) \neq 0$, we will show that $A \in \overline{\mathfrak{Z}}_{n+1}$. Since A is positive semi-definite and unidiagonal, there exist unit vectors $v_1, ..., v_{n+1}$ in $\mathbf{R}^n$ such that $A = [v_i \cdot v_j]$. We may assume $B = [v_i \cdot v_j]_{(k+1)\times(k+1)}, 1 \leq i, j \leq k+1$ and $ad(B) = [b_{ij}]$. Due to $\det(B) = 0, ad(B) \neq 0$, we have $rank(v_1, ..., v_{k+1}) = k$. We may assume $v_2, ..., v_{k+1}$ are independent. Thus the cofactor $b_{11} > 0$. By the assumption $ad(B) \geq 0$, we have $b_{1s} \geq 0$ for $s = 1, ...k+1$. Since $\sum_{s=1}^{k+1} b_{1s}(v_s \cdot v_j) = 0$ for all $j = 2, ..., k+1$ and $v_2, ..., v_{k+1}$ are independent, we get $\sum_{s=1}^{k+1} b_{1s} v_s = 0$. Thus we get a nonzero solution for the linear system $\sum_{i=1}^{n+1} a_i v_i = 0, a_i \geq 0, i = 1, ..., n+1$.

Proof of (b)

If $A \in \overline{\mathfrak{X}}_{n+1} - \mathfrak{X}_{n+1}$, then $A = [v_i \cdot v_j]$ where $v_1, ..., v_{n+1}$ are linear dependent. We can assume $v_1, ..., v_{n+1}$ lie in $\mathbf{R}^n \times \{0\}$. By change subindex, we may assume $\sum_{i=1}^{n+1} a_i v_i = 0$ has a non-zero solution with $a_i \geq 0$ if $i = 1, ..., k$ while $a_i < 0$ if $i = k+1, ..., n+1$. Thus vectors $v_1, ..., v_k, -v_{k+1}, ..., -v_{n+1}$ satisfy the condition (5.2) in Lemma 5. Let D be the diagonal matrix $diag(1, ..., 1, -1, ..., -1)$ with k diagonal entries being 1 and $n-k+1$ diagonal entries being -1. Thus by Lemma 5, $DAD \in \overline{\mathfrak{Z}}_{n+1}$.

On the other hand, if for some diagonal matrix D in Theorem 3 (b), we have $DAD \in \overline{\mathfrak{Z}}_{n+1}$, then by Theorem 3 (a), DAD is positive semi-definite. Therefore A is positive semi-definite. Take $B \in \mathfrak{X}_{n+1}$ and consider the family $A(t) = (1-t)A + tB$ for $t \in [0, 1]$. Then $\lim_{t\to 0} A(t) = A$ and $A(t) \in \mathfrak{X}_{n+1}$ for $t > 0$. Thus $A \in \overline{\mathfrak{X}}_{n+1}$.

Proof of (c)

First we show that the conditions are sufficient. Suppose $A = [a_{ij}]_{(n+1)\times(n+1)}$ is a symmetric unidiagonal matrix with all principal $n \times n$ submatrices positive semi-definite so that either $A \in \overline{\mathcal{Z}}_{n+1}$ or $\det(A) < 0$ and $ad(A) \geq 0$. We will show $A \in \overline{\mathcal{Y}}_{n+1}$. If $A \in \overline{\mathcal{Z}}_{n+1}$, it is sufficient to show that $\mathcal{Z}_{n+1} \subset \overline{\mathcal{Y}}_{n+1}$, i.e., we may assume $A \in \mathcal{Z}_{n+1}$. In this case, let $J = [c_{ij}]_{(n+1)\times(n+1)}$ so that $c_{ii} = 1$ and $c_{ij} = -1$ for $i \neq j$. Consider the family $A(t) = (1-t)A + tJ$, for $0 \leq t \leq 1$. Evidently $\lim_{t\to 0} A(t) = A$. We claim that $A(t) \in \mathcal{Y}_{n+1}$ for small $t > 0$. Since all principal $n \times n$ submatrices of A are positive definite, by continuity, all principal $n \times n$ submatrices of $A(t)$ are positive definite for small $t > 0$. It remains to check $\det(A(t)) < 0$ for small $t > 0$. To this end, let us consider $\frac{d}{dt}|_{t=0} \det(A(t))$. We have

$$\frac{d}{dt}|_{t=0} \det(A(t)) = \sum_{i \neq j} (-a_{ij} - 1) cof(A)_{ij} < 0,$$

due to $ad(A) = [cof(A)_{ij}] > 0$ and $-a_{ij} - 1 < 0$ for all $i \neq j$. Since $\det(A) = 0$ it follows that $det(A(t)) < 0$ for small $t > 0$.

In the second case that $\det(A) < 0$ and $ad(A) \geq 0$ and all principal $n \times n$ submatrices of A are positive semi-definite. Then A has a unique negative eigenvalue $-\lambda$, where $\lambda > 0$. Consider the family $A(t) = A + t\lambda I$, for $0 \leq t \leq 1$, where I is the identity matrix, so that

$$\lim_{t \to 0} \frac{1}{1+\lambda t} A(t) = A.$$

We claim there is a diagonal matrix D whose diagonal entries are ± 1 so that

(1) $DAD = A$,
(2) $\frac{1}{1+\lambda t} DA(t)D \in \mathcal{Y}_{n+1}$ for $0 < t < 1$.

As a consequence, it follows

$$\begin{aligned} A &= DAD \\ &= \lim_{t \to 0} \frac{1}{1+\lambda t} DA(t)D \in \overline{\mathcal{Y}}_{n+1}. \end{aligned}$$

To find this diagonal matrix D, by the continuity, $\det(A(t)) < 0$ for $0 < t < 1$ and $\det(A(1)) = 0$. Furthermore, all principal $n \times n$ submatrices of A are positive definite for $t > 0$ due to positive definiteness of $t\lambda I$. Let us recall the Lemma 3.4 in[4] which says that if B is a symmetric $(n+1) \times (n+1)$ matrix so that all $n \times n$ principal submatrices in B are positive definite and

$\det(B) \leq 0$, then no entry in the adjacent matrix $ad(B)$ is zero. It follows that every entry of $ad(A(t))$ is nonzero for $0 < t \leq 1$.

Let $ad(A(1)) = [b_{ij}]_{(n+1)\times(n+1)}$ and D to be the diagonal matrix with diagonal entries being

$$\frac{b_{1i}}{|b_{1i}|} = \pm 1$$

for $i = 1, ..., n+1$. Then the entries of the first row and the first column of $Dad(A(1))D$ are positive. Since $\det(A(1)) = 0$ and $ad(A(1)) \neq 0$, we see the rank of $ad(A(1))$ is 1. Thus any other column is propositional to the first column. But $b_{ii} > 0$ for all i, hence $ad(A(1)) > 0$. Now since every entry of $Dad(A(t))D$ is nonzero for $t > 0$, by continuity $Dad(A(t))D > 0$ for $t > 0$ and $Dad(A)D = Dad(A(0))D \geq 0$. By the assumption $ad(A) \geq 0$, it follows $Dad(A)D = ad(A)$. On the other hand $Dad(A)D = ad(D^{-1}AD^{-1})$, and $det(A) \neq 0$. Thus $D^{-1}AD^{-1} = A$ or the same $A = DAD$. This shows

$$\begin{aligned} A &= DAD \\ &= \lim_{t\to 0} DA(t)D \\ &= \lim_{t\to 0} \frac{1}{1+\lambda t} DA(t)D. \end{aligned}$$

By the construction above $\frac{1}{1+\lambda t}DA(t)D \in \mathcal{Y}_{n+1}$ for $0 < t < 1$, this shows $A \in \overline{\mathcal{Y}}_{n+1}$.

Finally, we show the condition in (c) is necessary. Suppose $A = \lim_{m\to\infty} A^{(m)}$ where $A^{(m)} \in \mathcal{Y}_{n+1}$. By Proposition 2, $\det(A^{(m)}) < 0$, $ad(A^{(m)}) > 0$ and all principal $n \times n$ submatrices of $A^{(m)}$ are positive definite. We want to show that A satisfies the conditions stated in (c). Evidently, all principal $n \times n$ submatrices of A are positive semi-definite, $ad(A) \geq 0$ and $det(A) \leq 0$. If $\det(A) < 0$, then we are done. If $\det(A) = 0$, by Lemma 7, we see that $A \in \overline{\mathcal{Z}}_{n+1}$.

References

1. K. Aomoto, *Analytic structure of Schläfli function.* Nagoya Math J. 68 (1977), 1–16.
2. H. Kneser, *Der Simplexinhalt in der nichteuklidischen Geometrie.* Deutsche Math. 1(1936), 337-340.
3. F. Luo, *On a problem of Fenchel.* Geom. Dedicata 64 (1997), no. 3, 277–282.
4. F. Luo, *Continuity of the Volume of Simplices in Classical Geometry.* Commun. Contemp. Math. 8 (2006), no. 3, 411–431.
5. J. Milnor, *The Schläfli differential equality.* Collected papers, vol. 1. Publish or Perish, Inc., Houston, TX, 1994.

6. J. G. Ratcliffe, *Foundations of Hyperbolic Manifolds*. Second Edition. GTM 149, Springer, 2006.
7. I. Rivin, *Continuity of volumes – on a generalization of a conjecture of J. W. Milnor*. arXiv:math.GT/0502543.
8. E. B. Vinberg, *Volumes of non-Euclidean polyhedra.* (Russian) Uspekhi Mat. Nauk 48 (1993), no. 2(290), 17–46; translation in Russian Math. Surveys 48 (1993), no. 2, 15–45.
9. E. B. Vinberg (Ed.), *Geometry II, Spaces of Constant Curvature*. Encyclopaedia of Mathematical Sciences, Vol. 29, Springer-Verlag Berlin Heidelberg 1993.

On the Classification of Links up to Finite Type

N. Habegger (Ju Li Wu Xiao-Song)
UMR 6629 du CNRS, Université de Nantes
Département de Mathématiques
2 rue de la Houssinière
44072 NANTES Cedex 03, France
Email: habegger@math.univ-nantes.fr

Jean-Baptiste Meilhan
CTQM - Department of Mathematical Sciences
University of Aarhus
Ny Munkegade, bldg. 1530
8000 Aarhus C, Denmark
Email: meilhan@imf.au.dk

We use an action, of $2l$-component string links on l-component string links, defined by Habegger and Lin, to lift the indeterminacy of finite type link invariants. The set of links up to this new indeterminacy is in bijection with the orbit space of the restriction of this action to the stabilizer of the identity. Structure theorems for the sets of links up to C_n-equivalence and Self-C_n-equivalence are also given.

(In fond rememberance of Xiao-Song Lin, 1957-2007)

1. Introduction

In,[M1,M2] Milnor defined invariants of links, known as the Milnor $\overline{\mu}$ invariants. In fact, these invariants are not universally defined, *i.e.*, if the lower order invariants do not vanish, they are either not defined, or at best, they have indeterminacies.

In,[HL1] the notion of string link was introduced, together with the philosophy that Milnor's invariants are actually invariants of string links. Indeterminacies are determined precisely by the indeterminacy of representing a link as the closure of a string link. This philosophy led to the classification of links up to homotopy, and to an algorithm constructed by Xiao-Song. (Here and throughout, we will often refer to Xiao-Song Lin by his first name.) More precisely, Xiao-Song and the first author constructed an orbit space structure for the set of links up to homotopy. The group action was 'unipotent', meaning it acted trivially on the successive layers of the nilpotent homotopy string link group. This was the determining structural feature which underlay the successful construction of Xiao-Song's algorithm.

In,[HL2] an analogous orbit space structure for link concordance was obtained and a study of the algebraic part of link concordance, corresponding to the Milnor concordance invariants, was made. The theory also applies to more general 'concordance-type' equivalence relations, in particular to those studied by Kent Orr[O] and developed in Xiao-Song's thesis.

With the advent of the physical interpretation of the Jones Polynomial,[J] predicted by Atiyah[A] and established by Witten,[W] a whole new area, known as Quantum Topology, emerged. Its perturbative aspects are succinctly summarized in the Universal Finite Type Invariant known as the Kontsevich Integral.[K]

Recall that, in the seminal paper,[L] Xiao-Song had shown that Milnor Invariants are finite type invariants of string links. We refer the reader to the paper of the first author and Gregor Masbaum,[HM] where a formula is given which computes the Milnor Invariants directly from the Kontsevich Integral.

No successful attempt has been made at applying the methods of[HL1,HL2] to the Vassiliev invariants.[V] (The Vassiliev invariants were shown by Xiao-Song and Joan Birman[BL] to be those invariants which satisfy the properties of finite type invariants. Subsequently, Bar-Natan[B] adopted those properties as axioms for finite type invariants.) This is so, because, as we show, the classification scheme does *not* hold. Thus, in the philosophy of,[HL1,HL2] the finite type invariants of links ought to be refined.

In this paper, we make such a refinement and show that after refinement, the classification scheme applies. We also show it applies to C_n-equivalence and to Self-C_n-equivalence.

Acknowledgments

The authors extend thanks to Jørgen Andersen and Bob Penner for organizing the stimulating conference *Finite Type Invariants, Fat Graphs and Torelli-Johnson-Morita Theory* at the CTQM in Aarhus, March 2008, during which this work was done. They also thank Jørgen Andersen and Gwénaël Massuyeau for useful conversations.

The first author thanks the University of Nantes for releasing his time for research. He also wishes to thank Zhenghan Wang for the invitation to speak at the conference in honor of Xiao-Song at the Chern Institute in

Tianjin, July 2007. It was there that he was baptized by Xiao-Song's wife, Jean, during a short ceremony in which he received his Chinese name, Ju Li Wu Xiao-Song.

2. Preliminaries

Let D^2 be the standard two-dimensional disk, and let I denote the unit interval. Recall from[HL1] the notion of string link.

Definition 2.1. Let $l \geq 1$. An l-component string link is a proper embedding,

$$\sigma : \bigsqcup_{i=1}^{l} I_i \to D^2 \times I,$$

of the disjoint union $\coprod_{i=1}^{l} I_i$ of l copies of I in $D^2 \times I$, such that the $j = 0, 1$ levels are preserved and $\partial_j \sigma \subset D^2 \times \{j\}$ is the standard inclusion of l points in D^2. By an abuse of notation, we will also denote by $\sigma \subset D^2 \times I$, the image of the map σ.

Note that we do not require that the t levels, for $t \in I$, be preserved. A string link is a pure braid precisely when it preserves the t levels for all $t \in I$. Note also that each string of an l-component string link is equipped with an (upward) orientation induced by the natural orientation of I.

The set $SL(l)$ of isotopy classes of l-component string links (fixing the boundary) has a monoidal structure, with composition given by the *stacking product* and with the trivial l-component string link 1_l as unit element. See Figure 1.

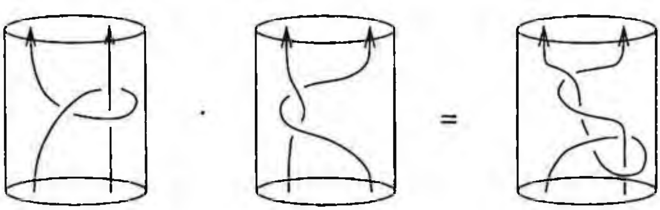


Fig. 1. Multiplying two 2-component string links.

Remark 2.1. In the above, one may replace the disk D^2 with any surface S to get the notion of a string link in $S \times I$. The l-component string links in S, up to isotopy, again has a monoidal structure.

We denote by $L(l)$ the set of isotopy classes of l-component links. By a link, we mean an embedding $\coprod_{i=1}^{l} \mathbf{S}_i^1 \to \mathbf{R}^3$. Thus the components are ordered and oriented. There is an obvious surjective *closure* map

$$\hat{} : SL(l) \longrightarrow L(l)$$

which closes an l-component string link σ into an l-component link $\hat{\sigma}$.

In,[HL1] Xiao-Song and the first author introduced a certain left (resp. right) action of the monoid of isotopy classes of $2l$-component string links on l-component string links. See Figure 2 for an illustration of these actions. Thus given two l-component string links σ, σ', and a $2l$-component string link Σ, one has l-component string links $\Sigma\sigma$, $\sigma\Sigma$, and a closed link $\sigma\Sigma\sigma'$.

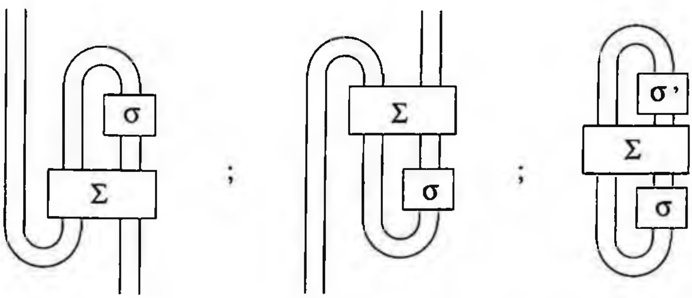


Fig. 2. Schematical representations of the left and right actions of Σ on σ, $\Sigma\sigma$ and $\sigma\Sigma$, and of the closed link $\sigma\Sigma\sigma'$.

One may represent the closure $\hat{\sigma}$ of a string link σ as $1_l 1_{2l}\sigma$, as well as $\sigma 1_{2l} 1_l$, and also as $1_l(1_l \otimes \sigma)1_l$, where $\sigma_1 \otimes \sigma_2$ denotes the $2l$-component string link obtained by horizontal juxtaposition. (One orients all strands appropriately in the above, e.g., in Σ, one must reverse the parametrization of the first l strands.)

The following result on basing links was proven in.[HL2]

Prop 2.1. Let σ_1, σ_2 be two l-component string links whose closures are isotopic. Then there is a $2l$-component string link Σ, with $1_l\Sigma$ isotopic to σ_1, and $\Sigma 1_l$ isotopic to σ_2.

3. The Habegger-Lin Classification Scheme

In[HL2] a structure theorem was proven for certain 'concordance-type' equivalence relations on the set of links. Given here for the convenience of the reader, though stated slightly differently, the result is in fact implicit in the proof in.[HL2]

Consider an equivalence relation E on string links and on links (for all l), which is implied by isotopy. We will denote by $E(x)$, the E equivalence class of x. We denote by $ESL(l)$, resp. $EL(l)$, the set of E equivalence classes of l-component string links, resp. links. We will also denote by E the map which sends a link or string link to its equivalence class.

Consider the following set of Axioms for an equivalence relation E:

For $i = 1, 2$, let σ_i be l-component string links with $E(\sigma_1) = E(\sigma_2)$, and let Σ_i be $2l$-component string links with $E(\Sigma_1) = E(\Sigma_2)$.

(1) $E(\hat{\sigma}_1) = E(\hat{\sigma}_2)$
(2) $E(1_l \otimes \sigma_1) = E(1_l \otimes \sigma_2)$
(3) $E(\sigma_1\Sigma_1) = E(\sigma_2\Sigma_2)$
(4) $E(\Sigma_1\sigma_1) = E(\Sigma_2\sigma_2)$.
(5) For all string links σ, there is a string link σ_1, such that $E(\sigma\sigma_1) = E(1_l)$.
(6) If $E(L) = E(L')$, then there is an m and a sequence of string links σ_i, for $i = 1, \ldots, m$, such that L is isotopic to $\hat{\sigma}_1$, and L' is isotopic to $\hat{\sigma}_m$, and for all i, $1 \leq i < m$, either $E(\sigma_i) = E(\sigma_{i+1})$, or $\hat{\sigma}_i$ is isotopic to $\hat{\sigma}_{i+1}$ (i.e., the equivalence relation E on links is generated by the equivalence relation of isotopy on links and the equivalence relation E on string links).
(5′) For all string links σ, $E(\sigma\overline{\sigma}) = E(1_l)$. Here the string link $\overline{\sigma}$ is defined by, $\overline{\sigma} = R_t \circ \sigma \circ R_s$, where R_s and R_t are the reflection mappings at the source and target.

Definition 3.1. An equivalence relation satisfying Axioms (1) – (4) is called local.

We have the following result.

Prop 3.1. Let E be a local equivalence relation. For $i = 1, 2$, let σ_i and σ_i' be l-component string links with $E(\sigma_1) = E(\sigma_2)$ and $E(\sigma_1') = E(\sigma_2')$, and let Σ_i be $2l$-component string links with $E(\Sigma_1) = E(\Sigma_2)$. Then $E(\sigma_1\sigma_1') = E(\sigma_2\sigma_2')$ and $E(\sigma_1\Sigma_1\sigma_1') = E(\sigma_2\Sigma_2\sigma_2')$.

The monoidal structures, the left (resp. right) action and the closure mapping all pass to maps of equivalence classes. Let $ES^R(l)$ (resp. $ES^L(l)$), denote the right (resp. left) stabilizer of the unit element of $ESL(l)$. Then $ES^R(l)$ (resp. $ES^L(l)$) is a submonoid of $ESL(2l)$. Furthermore, the closure mapping of $ESL(l)$ to $EL(l)$, passes to the set of orbits of the $ES^R(l)$ (resp. $ES^L(l)$) action, i.e., we have a map

$$\frac{ESL(l)}{S^R(l)} \longrightarrow EL(l).$$

If in addition, Axiom (5) holds, then the monoid $ESL(l)$ is a group, and $ES^R(l)$ (resp. $ES^L(l)$) is a subgroup of $ESL(2l)$. If Axiom (5′) holds, then $ES^R(l) = ES^L(l)$.

Proof. By Axiom (2), $E(1_l \otimes \sigma'_1) = E(1_l \otimes \sigma'_2)$. Using Axiom (3), we have that $E(\sigma_1\sigma'_1) = E(\sigma_1(1_l \otimes \sigma'_1)) = E(\sigma_2(1_l \otimes \sigma'_2)) = E(\sigma_2\sigma'_2)$.

One defines $E(\sigma_1)E(\sigma_2) = E(\sigma_1\sigma_2)$. This is well defined by the above, and the element $E(1_l)$ is a unit. One also defines $E(\sigma)E(\Sigma) = E(\sigma\Sigma)$ and $E(\Sigma)E(\sigma) = E(\Sigma\sigma)$. These are well defined, by Axioms (3) and (4), and are monoidal actions (of sets).

Suppose σ and σ' define the same element of $\frac{ESL(l)}{S^R(l)}$, i.e., there is $E(\Sigma) \in S^R(l)$ such that $E(\Sigma)E(\sigma) = E(\sigma')$. One has $E(\hat{\sigma}') = E(1_l\Sigma\sigma) = E(1_l 1_{2l}\sigma) = E(\hat{\sigma})$.

If Axiom (5) holds, each element $E(\sigma)$ has a right inverse. Hence $ESL(l)$ is a group, for all l.

If $E(\Sigma)$ belongs to $ES^R(l)$, then $E(\overline{\Sigma})$ belongs to $ES^L(l)$. But if Axiom (5′) holds, then $E(\Sigma)$ is the inverse of $E(\overline{\Sigma})$, so also belongs to $ES^L(l)$. This proves one inclusion and the other is proven similarly. □

Theorem 3.1 (Structure Theorem for E-equivalence).

(1) Let E be a local equivalence relation satisfying Axiom (5)*. Then the quotient map*

$$SL(l) \longrightarrow \frac{ESL(l)}{ES^R(l)}$$

factors through the closure mapping, i.e., we have a link invariant

$$\widetilde{E} : L(l) \longrightarrow \frac{ESL(l)}{ES^R(l)}.$$

such that the composite map to $EL(l)$ is E.

(2) Furthermore, if Axiom (6) *also holds, then we have a bijection*

$$\frac{ESL(l)}{S^R(l)} = EL(l).$$

Proof.

Suppose $\hat{\sigma}$ is isotopic to $\hat{\sigma}'$. By Proposition 3.9, one has that, for some Σ_0, σ' is isotopic to $\Sigma_0 1_l$ and σ is isotopic to $1_l\Sigma_0$. Set $\Sigma = \Sigma_0(1_l \otimes \sigma_1)$, where $E(\sigma_1)$ satisfies $E(\sigma\sigma_1) = E(1_l)$ (and hence also $E(\sigma_1\sigma) = E(1_l)$). One has that $E(1_l)E(\Sigma) = E(1_l\Sigma) = E(\sigma\sigma_1) = E(1_l)$, so $E(\Sigma) \in S^R(l)$. See Figure 1.

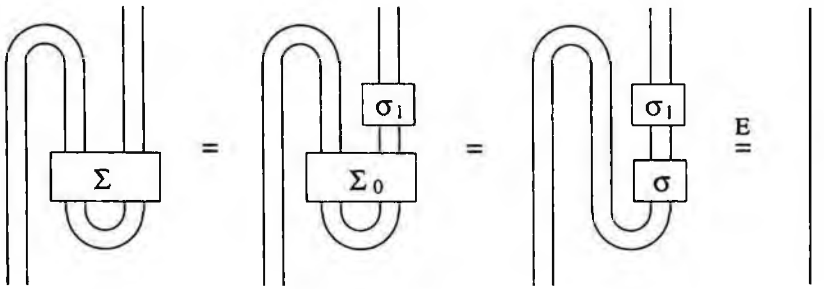


Fig. 1. Proof that Σ lies in $S^R(l)$.

Finally, one has that $E(\Sigma)E(\sigma) = E(\Sigma\sigma) = E(\Sigma_0)E(\sigma_1\sigma) = E(\Sigma_0)E(1_l) = E(\Sigma_0 1_l) = E(\sigma')$. See Figure 2. This completes the proof of (1).

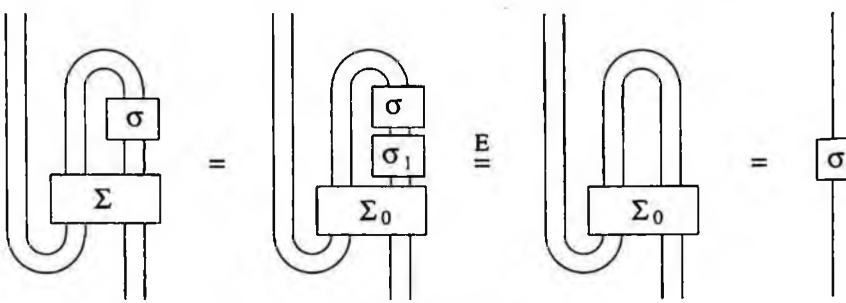


Fig. 2. Proof that $\Sigma\sigma$ is equivalent to σ'.

To see (2), note that we have already shown that if the closures of two string links are isotopic, then they define the same element of $\frac{ESL(l)}{S^R(l)}$. Thus we have that for all i in Axiom (6), $E(\sigma_i)$ and $E(\sigma_{i+1})$ both agree in $\frac{ESL(l)}{S^R(l)}$. Thus the surjective map from $\frac{ESL(l)}{S^R(l)}$ to $EL(l)$ is injective. □

4. Structure Theorems for C_n-equivalence and for Self-C_n-equivalence.

We will denote by FT_n, the equivalence relation on tangles determined by finite type equivalence up to degree n, i.e., FT_n-equivalent tangles differ by an element in the $n+1$st term of the Vassiliev filtration. In,[H] K. Habiro showed that, *for knots*, FT_n-equivalence agrees with another equivalence relation, called C_{n+1}-equivalence. Habiro conjectured in[H] that for string links, FT_n equivalence is equivalent to C_{n+1}-equivalence.

Habiro also showed that for links, the result does not hold. Note that, since the structure theorem holds for C_{n+1}-equivalence, if the equivalence relations were the same both for string links and for links, it would also hold for FT_n equivalence. However, for FT_n equivalence, the structure theorem does not hold (see Theorem 5.1 and the Borromean ring example of Section 5).

By definition, two tangles are said to be C_n-equivalent, if there is a finite sequence of tree clasper surgeries, of degree greater than or equal to n, taking one tangle to the other, up to isotopy. See[H] for the definition. (Note that in,[H] a tree clasper is called an admissible, strict tree clasper.) Here the leaves of the tree can be assumed to be trivial and intersect the tangle in a single point. It is known that C_{n+1}-equivalent tangles are FT_n-equivalent (see [H, Σ6]).

By definition, two tangles are said to be Self-C_n-equivalent, if there is a finite sequence of tree clasper surgeries, of degree greater than or equal to n, taking one tangle to the other, up to isotopy, such that the leaves of each tree are restricted to all intersect the same tangle component.

Remark 4.1. Self-C_n-Equivalence, for $n=1$, is link-homotopy. For $n=2$ it is also known as Self-Delta equivalence.

C_n-equivalence and Self-C_n-equivalence are obviously local, i.e., they satisfy Axioms (1) – (4) of Section 3. Axiom 5 was shown in [H, Theorem 5.4] for C_n-equivalence.

Prop 4.1. Self-C_n-equivalence satisfies Axiom (5) of Section 3.

Prop 4.2. C_n-equivalence and Self-C_n-equivalence satisfy Axiom (6) of Section 3.

Applying Theorem 3.2, one has the following result.

Theorem 4.1 (Structure Theorem and Self-C_n-equivalence).

$$\frac{C_n SL(l)}{C_n S^R(l)} = C_n L(l).$$

$$\frac{\textit{Self-}C_n SL(l)}{\textit{Self-}C_n S^R(l)} = \textit{Self-}C_n L(l).$$

Proof of Proposition 5.2. Suppose that L' is obtained from L by surgery on a disjoint union F of tree claspers of degree $\geq n$. Let L be the closure of σ. Since the disk base for L retracts onto a 1-complex, we may assume it is disjoint from F. Thus L' is the closure of a string link σ', obtained from σ by surgery on a union of tree claspers of degree $\geq n$. This shows that C_n-equivalence (resp. Self-C_n-equivalence) for links is implied by C_n-equivalence (resp. Self-C_n-equivalence) for string links and isotopy. □

Proposition 5.1 is a special case of the following result.

Prop 4.3. Let S be any surface. Self-C_n-equivalence (and consequently C_n-equivalence) classes of string links in $S \times I$ form a group.

Proof of Proposition 5.3. The proof is by induction on the number l of components. For $l = 1$, Self-C_n-equivalence is C_n-equivalence, so we may invoke [H, Theorem 5.4].

Suppose the result is true for $l-1$. Removing the first component from σ, we have an $l-1$-component string link σ_0. By the induction hypothesis, σ_0 has an inverse σ_1, up to Self-C_n-equivalence. Let $\sigma' = \sigma(1_1 \otimes \sigma_1)$. It suffices to find a right inverse for σ', up to Self-C_n-equivalence. Note that the string link σ'_0, obtained from σ' by removing the first component, is Self-C_n-equivalent to the trivial string link 1_{l-1}. Thus 1_{l-1} is obtainable from σ'_0 by surgery on a disjoint union F of trees of degree n such that the leaves of each tree are restricted to intersect a single component.

We may assume that F is disjoint from the first component of σ'. Perform surgery on F to obtain from σ' a string link σ''. As σ' is Self-C_n-equivalent to σ'', it remains to find a right inverse for σ''. Note that after removing the first component of σ'', we obtain the trivial $l-1$-component string link. Thus if we remove from σ'' the last $l-1$ components, we have a one-component string link σ''_0 in $S' \times I$, where S' is the surface obtained from S by removing $l-1$ points. Since, by the result for $l = 1$, the string link σ''_0 has a right inverse, up to Self-C_n-equivalence, so does σ''. □

5. The Indeterminacy of Finite Type Invariants.

In this section we assume the reader is familiar with the notion of finite type invariant as well as the Kontsevich Integral, which is the Universal Finite Type Invariant. Recall from the last section that we have denoted by FT_n, the equivalence relation on tangles determined by finite type equivalence up to degree n, i.e., FT_n-equivalent tangles differ by an element in the $n+1$st term of the Vassiliev filtration.

Let us begin with a disturbing fact about finite type invariants of links. The Borromean Rings are distinguished from the unlink by the triple Milnor Invariant. Unfortunately, this invariant, which is really only defined as an integer when the linking numbers of the 2-component sublinks vanish, dies in the space of trivalent Feynman diagrams (also known as Jacobi diagrams) on 3 circles. This is because, when passing from 3 intervals to 3 circles, invariants of linear combinations of string links, which die upon closure, must also die upon closure for other linear combinations which are equivalent.

This can be seen using the Kontsevich Integral. Specifically, in the space of Jacobi diagrams on 3 intervals we have

where the right-hand side is obviously mapped to zero when closing. (Recall that the coefficient of the Y-shaped diagram on the left-hand side corresponds to the triple Milnor Invariant.)

To see how this comes about more geometrically, consider the free group on 2 generators as a subgroup of the 3 component pure braid group. The word $xyx^{-1}y^{-1}$ represents the Borromean rings, after closure. Since xyx^{-1} and y are conjugate, and thus agree after closure, we see that the quantity $xyx^{-1}y^{-1}-1$, which (say after applying the Magnus expansion) is in degree 2 before closure, lies in degree 3 after closure, since it agrees after closure with the quantity $(xyx^{-1}y^{-1}-1)(y-1)$, which is in degree 3. (The degree considerations here are valid in the Vassiliev filtration as well). Thus we see that we can no longer distinguish the Borromean rings from the unlink!

In summary, the indeterminacies of higher order invariants due to the non-vanishing of lower order ones, propagate to destroy what should be invariants of links whose lower order invariants vanish. We are thus led to a problem of refining the indeterminacies in a less algebraic way. We are guided by the structure theorem of the last section.

Rationally, it is known, see,[HM] that the set rational finite type equivalence classes of l-component string links is a finitely generated torsion free nilpotent group. Over the integers, it follows from the last section, since $C_{n+1}SL(l)$ is a group and surjects to $FT_nSL(l)$, that $FT_nSL(l)$ is also a group.

The set $FT_nSL(2l)$ acts on $FT_nSL(l)$ on the left and right. Let $FT_nS^R(l)$ denote the stabilizer of $FT_n(1_l)$ under the right action. $FT_nL(l)$ denotes the set of FT_n equivalence classes of l-component links.

The main result of this paper is the following.

Theorem 5.1 (Structure Theorem for Finite Type Equivalence).

(1) The projection mapping, of $SL(l)$ to the set $\frac{FT_nSL(l)}{FT_nS^R(l)}$ of left $FT_nS^R(l)$ orbits, factors through $L(l)$ and thus gives a well defined invariant of links

$$\widetilde{FT}_n : L(l) \longrightarrow \frac{FT_nSL(l)}{FT_nS^R(l)}.$$

(2) The above link invariant lifts the indeterminacies given by finite type invariants of links, i.e., if two links determine the same element of $\frac{FT_nSL(l)}{FT_nS^R(l)}$, then they have the same finite type invariants up to degree n. That is, the above map, $\widetilde{FT}_n$, factors through a (surjective, but not generally injective) map,

$$\frac{FT_nSL(l)}{FT_nS^R(l)} \longrightarrow FT_nL(l),$$

and the composite mapping is

$$FT_n : L(l) \longrightarrow FT_nL(l).$$

Proof. Axioms (1) – (4) follow from the local definition of the Vassiliev filtration. Axiom (5) follows from the remark above that $FT_nSL(l)$ is a group. □

Remark 5.1. The analogous theorem also holds if one restricts to the equivalence relation FT_n^Q, given by rational invariants of finite type of degree up to n. One can use the local property of the Kontsevich Integral (and the result cited above from[HM]) to give an alternative proof of the Axioms (1) – (5) in this case.

Let $A_{\leq n}(l)$ denote the algebra of Jacobi diagrams on l strands of degree up to n. The action of $FT_n^QSL(2l)$ on the set $FT_n^QSL(l)$ is induced, via the Kontsevich Integral, by an analogously defined action of $A_{\leq n}(2l)$ on $A_{\leq n}(l)$, given purely diagrammatically. (In the definition of the action of string links, just replace the string links with diagrams.) Let $A_{\leq n}(2l)_1$ be the stabilizer of the unit element in $A_{\leq n}(l)$. The stabilizer $A_{\leq n}(2l)_1$ contains $FT_n^QS^R(l)$. It is easily seen that there are surjective maps of the space of covariants $A_{\leq n}(l)/FT_n^QS^R(l)$ to the space of covariants $A_{\leq n}(l)/A_{\leq n}(2l)_1$, and from $A_{\leq n}(l)/A_{\leq n}(2l)_1$ to the space $A_{\leq n}(\coprod_{i=1}^l S_i^1)$ of diagrams on l circles, up to degree n. Using the link invariance of our theorem, and the universal property of the Kontsevich Integral, one can check that these maps are both isomorphisms. (We do not have a diagrammatical proof of this fact.) It follows that one should not pass to covariants to try to refine finite type invariants of links!

We conclude this section with several problems.

Problem 5.1:

Use the 'unipotent' action to write an algorithm, analogous to Xiao-Song's link-homotopy algorithm, 'calculating' whether or not two (string) links determine the same element in the orbit space.

Problem 5.2:

Does the full Kontsevich Integral for links (or integrally, modulo the intersection of the Vassiliev filtration) 'recapture' the information lost at each finite level? (For example, the triple Milnor Invariant dies, but its cube does not. But of course the degree is now 6 and not 2.)

References

A. M. F. Atiyah, *New Invariants of Three and Four Dimensional Manifolds*, The Mathematical Heritage of Hermann Weyl, Proc. Symp. Pure Math. 48, ed. R. Wells, AMS, 1988.

B. D. Bar-Natan, *On the Vassiliev knot invariants*, Topology **34** (1995), no. 2, 423–472.

BL. J. Birman, X.S. Lin, *Knot polynomials and Vassiliev's invariants*, Invent. Math. **111** (1993) 225–270.

HL1. N. Habegger, X.S. Lin, *The classification of links up to link-homotopy*, J. Amer. Math. Soc. **3** (1990), no. 2, 389-419.

HL2. N. Habegger, X.S. Lin, *On Link Concordance and Milnor's $\overline{\mu}$ Invariants*, Bull. London Math. Soc. **30** (1998), 419-428.

HM. N. Habegger and G. Masbaum, *The Kontsevich integral and Milnor's invariants*, Topology **39** (2000), no. 6, 1253–1289.

J. V. F. R. Jones, *Hecke algebra representations of braid groups and link polynomials*, Ann. Math. **126** (1987) 335–388.

K. M. Kontsevich, *Vassiliev's knot invariants*, "I. M. Gel'fand Seminar", 137–150, Adv. Soviet Math., 16, Part 2, Amer. Math. Soc., Providence, RI, 1993.

H. K. Habiro, *Claspers and finite type invariants of links*, Geom. Topol. **4** (2000), 1–83.

L. X.S. Lin, *Power series expansions and invariants of links*, in "Geometric topology" (W. Kazez, Ed.), Proc. Georgia Int. Topology Conf.1993, AMS/IP Studies in Adv. Math. (1997).

M1. J. Milnor, *Link groups*, Annals of Math. **59** (1954), 177-195.

M2. J. Milnor, *Isotopy of links*, Algebraic geometry and topology, A symposium in honor of S. Lefschetz, pp. 280–306, Princeton University Press, Princeton, N. J., 1957.

O. K. Orr, *Homotopy invariants of links*, Invent. Math. **95**, No. 2, 1989, 379-394.

V. V.A. Vassiliev, *Cohomology of knot spaces*, in "Theory of singularities and its Applications" (V.I. Arnold ed.) Amer. Math. Soc., Providence, 1990.

W. E. Witten, *Quantum field theory and the Jones polynomial*, Comm. Math. Phys. **121** (1989), 351-399.

Generalized Ricci Flow I: Local Existence and Uniqueness

Chun-Lei He[*], Sen Hu[†], De-Xing Kong[‡] and Kefeng Liu[§]

In this paper we investigate a kind of generalized Ricci flow which possesses a gradient form. We study the monotonicity of the given function under the generalized Ricci flow and prove that the related system of partial differential equations are strictly and uniformly parabolic. Based on this, we show that the generalized Ricci flow defined on an n-dimensional compact Riemannian manifold admits a unique short-time smooth solution. Moreover, we also derive the evolution equations for the curvatures, which play an important role in our future study.

Keywords: Generalized Ricci flow, uniformly parabolic system, short-time existence, Thurston's eight geometries.

1. Introduction

In the early eighties R. Hamilton introduced the Ricci flow to construct canonical metrics for some manifolds. Since then many mathematicians, including Hamilton, Yau, Perelman and others, developed many tools and techniques to study the Ricci flow. The latest developments confirmed that the Ricci flow approach is very powerful in the study of three-manifolds. In fact, a complete proof of Poincare's conjecture and Thurston's geometrization conjecture has been offered in Cao-Zhu's paper[3] and others after Perelman's breakthrough.

It is useful to observe that, in Perelman's work,[10] a key step is to introduce a functional for a metric g and a function f on a manifold M

$$W(g,f)=\int_{M^3} d^3x\sqrt{g}e^{-f}(R+|\nabla f|^2).$$

The variation of this functional generates a gradient flow which is a system

[*]Department of Mathematics, Shanghai Jiao Tong University, Shanghai 200240, China;
[†]Department of Mathematics, University of Science and Technology of China, Hefei 230026, China;
[‡]Center of Mathematical Sciences, Zhejiang University, Hangzhou 310027, China;
[§]Department of Mathematics, University of California at Los Angeles, CA 90095, USA.

of partial differential equations

$$\dot{g}_{ij} = -2(R_{ij} + \nabla_i \nabla_j f),$$

$$\dot{f} = -(R + \triangle f).$$

If we fix a measure for the conformal class of metrics $e^f ds^2$ of a metric, i.e., let $dm = e^{-f} dV$ be fixed, then we get back to the original Ricci flow after we apply a transformation of diffeomorphism generated by the vector field $\nabla_i f$ to the metric. In this way, we express the Ricci flow as a gradient flow. Dynamics of a gradient flow is much easier to handle. The functional generating the flow gives a monotone functional along the orbit of the flow automatically. If the flow exists for all time, then it shall flow to a critical point which leads to the existence of a canonical metric. Even for a flow which does not exist for all time, the generating functional helps very much in the analysis of singularities.

Perelman's above idea came from physics. Ricci flow arises as the first order approximation of the renormalization flow of a sigma model. Since there are many kinds of sigma models, it would be interesting to try some other models. Indeed such a generalization was made by physicists in.[11] For a three-manifold M^3, they proposed to add a $U(1)$ gauge field with potential 1-form A and field strength F which are coupled as a *Maxwell-Chern-Simons theory*. The corresponding action given by[6] or[5] reads

$$S = \int_M d^3x\sqrt{g}e^{-f}(-\chi + R + |\nabla f|^2) - \frac{1}{2}e^{-f}H \wedge *H - e^{-f}F \wedge *F.$$

The $U(1)$ gauge field A is a one-form potential whose field strength $F = dA$. The Wess-Zumino field B is a two-form potential whose field strength $H = dB$, f is a dilaton. In their paper, they find that Thurston's eight geometries appear as critical points of the above functional. Furthermore they show that there are no other critical points. So basically critical points of the above functional are eight geometries of Thurston. They also propose to study the gradient flow of the functional S as a generalization of the Ricci flow. Unfortunately, they modify the gradient flow in a way to change sign for the variable of gauge fields. Although the modified flow shares the same set of critical points they lost the important monotone property (along an orbit).

In addition, we are also able to consider a flow for a similar functional for a four-dimension manifold

$$S_1 = \int_M d^4x\sqrt{g}e^{-f}(\chi + R + |\nabla f|^2) - \frac{1}{2}e^{-f}H \wedge *H - e^{-f}F \wedge *F + \frac{e}{2}F \wedge F$$

where e is the Euler number $e(\eta)$ of the bundle η. The corresponding flow is given by

$$\begin{cases} \dfrac{\partial g_{ij}}{\partial t} = -2[R_{ij} + \nabla_i \nabla_j f - \dfrac{1}{4} H_{ikl} H_j{}^{kl} - F_i{}^k F_{jk}], \\ \dfrac{\partial B_{ij}}{\partial t} = e^f \nabla_k (e^{-f} H^k{}_{ij}), \\ \dfrac{\partial A_i}{\partial t} = -e^f \nabla_k (e^{-f} F_i{}^k), \\ \dfrac{\partial f}{\partial t} = \chi - 2R - 3\Delta f + |\nabla f|^2 + \dfrac{1}{3} H^2 + \dfrac{3}{2} F^2. \end{cases}$$

The generalization to four-manifolds is probably more interesting. It may offer a systematic way to study four-manifolds.

The success of studying three-manifolds relies on a program proposed by Thurston, i.e., his geometrization conjecture. He conjectures and proves for several large classes of three-manifolds, that every three-manifold can be decomposed into pieces of three-manifolds of canonical metrics, i.e., those manifolds carrying one of the eight geometries of Thurston.

For four-dimension manifolds the critical points of S_1 might play a similar role as building blocks of smooth four-dimension manifolds. It would be interesting to study those critical points and to study what other four manifolds one can get by performing surgeries and gluing on those manifolds. We shall address this problem in the future.

As a first step, we shall show that the flow does exist. We shall also prove that the modified system of partial differential equations are strictly and uniformly parabolic.

The paper is organized as follows. Section 2 is devoted to the proof of local existences and uniqueness. In section 3 we study the monotonicity of S under the modified flow. In Section 4 we investigate the equations for the critical points of S and point out that fields F and H do not provide any help for the case of compact manifold but maybe play an important role for the noncompact case. In Section 5, we derive the evolution equations for the curvatures, which play an important role in our future study.

2. Local Existences and Uniqueness

In this section, we mainly establish the short-time existence and uniqueness result for the gradient flow (1), (2) and (3) on a compact 3-dimensional manifold M. It is known that the gradient flow (1), (2) and (3) is a system of second order nonlinear weakly parabolic partial differential equations.

By the proof of the local existence and uniqueness of the Ricci flow (for example see[3],[4]), we can obtain a modified evolution equations by the diffeomorphism φ of M, which is a strictly parabolic system. Then, by the standard theory of parabolic equations, the modified evolution equations has a uniqueness solution.

Let us choose a normal coordinate $\{x^i\}$ around a fixed point $x \in M$ such that $\dfrac{\partial g_{ij}}{\partial x^k} = 0$ and $g_{ij}(p) = \delta_{ij}$.

Theorem 2.1. *(Local existences and uniqueness) Let $(M, g_{ij}(x))$ be a three-dimensional compact Riemannian manifold. Then there exists a constant $T > 0$ such that the evolution equations*

$$\begin{cases} \dfrac{\partial g_{ij}}{\partial t} = -2[R_{ij} - \dfrac{1}{4} H_{ikl} H_j^{\ kl} - F_i^{\ k} F_{jk}], \\ \dfrac{\partial A_i}{\partial t} = -\nabla_k F_i^{\ k}, \\ \dfrac{\partial B_{ij}}{\partial t} = \nabla_k H^k_{\ ij}, \end{cases} \tag{1}$$

has a unique smooth solution on $M \times [0, T)$ for every initial fields.

Lemma 2.1. *For each gauge equivalent class of a gauge field A, there exists an A' such that $d(*A') = 0$.*

The lemma can be proved by the Hodge decomposition.

Proof. For each one-form A, by the Hodge decomposition, there exists an one-form A_0, a function α and a two-form β such that

$$A = A_0 + d\alpha + d^*\beta,$$

$$dA_0 = 0, \quad d^*A_0 = 0.$$

Let $A' = A - d\alpha$. A' is in the same gauge equivalent class of A. Since $d(*A_0) = 0$, $d(*d^*\beta) = 0$, then we have $d(*A') = 0$. □

Lemma 2.2. *The differential operator of the right hand of (2) with respect to the gauge equivalent class of a gauge field A is uniformly elliptic.*

Proof. Let $A = A_i dx^i$ be a gauge field. By Lemma 4.1 we can choose an A' in the gauge equivalent class of A such that $d(*A') = 0$. We still denote A' as A . Since $d(*A) = 0$, we have $dd^*A = 0$, then $\sum_{k=1}^{3} \dfrac{\partial^2 A_k}{\partial x^k \partial x^i} = 0, \forall\ i =$

$1, \cdots, 3$. Noting that $F = dA$ and $F_{ij} = \dfrac{\partial A_j}{\partial x^i} - \dfrac{\partial A_i}{\partial x^j}$, We have

$$\frac{\partial A_i}{\partial t} = -\nabla_k F_i^{\ k} = -\nabla_k (g^{kl} F_{il}) = -g^{kl}(\frac{\partial^2 A_l}{\partial x^k \partial x^i} - \frac{\partial^2 A_i}{\partial x^k \partial x^l}) = g^{kl} \frac{\partial^2 A_i}{\partial x^k \partial x^l}.$$

The right hand side of above equation is clearly elliptic at point x. If we apply a diffeomorphism to the metric it won't change the positivity property of the second order operator of the right hand side. □

Now let us consider the equation for B_{ij}.

Lemma 2.3. *For each gauge equivalent class of a B-field B, i.e., a two-form B on M, there exists a $B^{'}$ such that $d(*B^{'}) = 0$.*

Proof. Again we use the Hodge decomposition. For a two-form B, there exist a one-form α, a two-form B_0 and a three-form β such that

$$B = B_0 + d\alpha + d^* \beta,$$

$$dB_0 = 0, d^* B_0 = 0.$$

Let $B^{'} = B - d\alpha$. Since $B^{'}$ is in the same gauge equivalent class of B, we have $d(*B^{'}) = 0$. □

Lemma 2.4. *The differential operator of the right hand side of (3) with respect to the gauge equivalent class of a B-field B is uniformly elliptic.*

Proof. Let us consider the equation for B-field. Without loss of generality, we assume $d(*B) = 0$. Thus $dd^* B = 0$. Then $\sum_{k=1}^{3} (\dfrac{\partial^2 B_{ki}}{\partial x^k \partial x^j} + \dfrac{\partial^2 B_{jk}}{\partial x^k \partial x^i}) = 0, \forall\, i, j = 1, \cdots, 3$. We have

$$\frac{\partial B_{ij}}{\partial t} = \nabla_k H^k_{\ ij} = g^{kl}(\frac{\partial^2 B_{ij}}{\partial x^k \partial x^l} + \frac{\partial^2 B_{jl}}{\partial x^k \partial x^i} + \frac{\partial^2 B_{li}}{\partial x^k \partial x^j}) = g^{kl} \frac{\partial^2 B_{ij}}{\partial x^k \partial x^l}.$$

The right hand side is clearly elliptic at the point x. If we apply a diffeomorphism to the metric it does not change the positivity property of the second order operator of the right hand side. □

Suppose $\hat{g}_{ij}(x,t)$ is a solution of the equations (1), and $\varphi_t : M \to M$ is a family of diffeomorphisms of M. Let

$$g_{ij}(x,t) = \varphi_t^* \hat{g}_{ij}(x,t),$$

where φ_t^* is the pull-back operator of φ_t. We now want to find the evolution equations for the metric $g_{ij}(x,t)$.

Denote

$$y(x,t) = \varphi_t(x) = \{y^1(x,t), y^2(x,t), \cdots, y^n(x,t)\}$$

in local coordinates. Then

$$g_{ij}(x,t) = \frac{\partial y^\alpha}{\partial x^i}\frac{\partial y^\beta}{\partial x^j}\hat{g}_{\alpha\beta}(y,t)$$

and

$$\begin{aligned}\frac{\partial}{\partial t}g_{ij}(x,t) &= \frac{\partial}{\partial t}\left[\hat{g}_{\alpha\beta}(y,t)\cdot\frac{\partial y^\alpha}{\partial x^i}\cdot\frac{\partial y^\beta}{\partial x^j}\right] \\ &= \frac{\partial y^\alpha}{\partial x^i}\frac{\partial y^\beta}{\partial x^j}\frac{\partial}{\partial t}\hat{g}_{\alpha\beta}(y,t) + \frac{\partial y^\alpha}{\partial x^i}\frac{\partial y^\beta}{\partial x^j}\frac{\partial y^\gamma}{\partial t}\frac{\partial}{\partial y^\gamma}\hat{g}_{\alpha\beta}(y,t) \\ &\quad +\hat{g}_{\alpha\beta}(y,t)\frac{\partial}{\partial x^i}(\frac{\partial y^\alpha}{\partial t})\frac{\partial y^\beta}{\partial x^j} + \hat{g}_{\alpha\beta}(y,t)\frac{\partial y^\alpha}{\partial x^i}\frac{\partial}{\partial x^j}(\frac{\partial y^\beta}{\partial t}) \,.\end{aligned}$$

Since

$$\begin{aligned}\frac{\partial y^\alpha}{\partial x^i}\frac{\partial y^\beta}{\partial x^j}\frac{\partial y^\gamma}{\partial t}\frac{\partial}{\partial y^\gamma}\hat{g}_{\alpha\beta} &= \frac{\partial y^\alpha}{\partial x^i}\frac{\partial y^\beta}{\partial x^j}\frac{\partial y^\gamma}{\partial t}g_{kl}\frac{\partial}{\partial y^\gamma}(\frac{\partial x^k}{\partial y^\alpha}\frac{\partial x^l}{\partial y^\beta}) \\ &= \frac{\partial y^\beta}{\partial t}\frac{\partial^2 x^k}{\partial y^\alpha \partial y^\beta}\frac{\partial y^\alpha}{\partial x^i}g_{jk} + \frac{\partial y^\alpha}{\partial t}\frac{\partial^2 x^k}{\partial y^\alpha \partial y^\beta}\frac{\partial y^\beta}{\partial x^j}g_{ik} \,, \\ \Gamma^k_{jl} &= \frac{\partial y^\alpha}{\partial x^j}\frac{\partial y^\beta}{\partial x^l}\frac{\partial x^k}{\partial y^\gamma}\hat{\Gamma}^\gamma_{\alpha\beta} + \frac{\partial x^k}{\partial y^\alpha}\frac{\partial^2 y^\alpha}{\partial x^j \partial x^l} \,,\end{aligned}$$

then

$$\frac{\partial y^\alpha}{\partial x^i}\frac{\partial y^\beta}{\partial x^j}\hat{H}_{\alpha\rho\delta}\hat{H}_\beta{}^{\rho\delta} = H_{ikl}H_j{}^{kl}, \qquad \frac{\partial y^\alpha}{\partial x^i}\frac{\partial y^\beta}{\partial x^j}\hat{F}_\alpha{}^\rho\hat{F}_{\beta\rho} = F_i{}^kF_{jk} \,.$$

Therefore, in the normal coordinate, we have

$$\begin{aligned}\frac{\partial}{\partial t}g_{ij}(x,t) &= \frac{\partial}{\partial x^i}(\frac{\partial y^\alpha}{\partial t})\frac{\partial y^\beta}{\partial x^j}g_{kl}\frac{\partial x^k}{\partial y^\alpha}\frac{\partial x^l}{\partial y^\beta} + \frac{\partial y^\alpha}{\partial x^i}\frac{\partial}{\partial x^j}(\frac{\partial y^\beta}{\partial t})g_{kl}\frac{\partial x^k}{\partial y^\alpha}\frac{\partial x^l}{\partial y^\beta} \\ &\quad +\frac{\partial y^\alpha}{\partial t}\frac{\partial}{\partial x^i}(\frac{\partial x^k}{\partial y^\alpha})g_{jk} + \frac{\partial y^\beta}{\partial t}\frac{\partial}{\partial x^j}(\frac{\partial x^k}{\partial y^\beta})g_{ik} \\ &\quad +\frac{\partial y^\alpha}{\partial x^i}\frac{\partial y^\beta}{\partial x^j}\left[-2(\hat{R}_{\alpha\beta} - \frac{1}{4}\hat{H}_{\alpha\rho\delta}\hat{H}_\beta{}^{\rho\delta} - \hat{F}_\alpha{}^\rho\hat{F}_{\beta\rho})\right] \\ &= \frac{\partial}{\partial x^i}(\frac{\partial y^\alpha}{\partial t}\frac{\partial x^k}{\partial y^\alpha})g_{jk} + \frac{\partial}{\partial x^j}(\frac{\partial y^\beta}{\partial t}\frac{\partial x^k}{\partial y^\beta})g_{ik} - 2R_{ij} \\ &\quad +\frac{1}{2}H_{ikl}H_j{}^{kl} + 2F_i{}^kF_{jk} \\ &= -2R_{ij} + \nabla_i(\frac{\partial y^\alpha}{\partial t}\frac{\partial x^k}{\partial y^\alpha}g_{jk}) + \nabla_j(\frac{\partial y^\beta}{\partial t}\frac{\partial x^k}{\partial y^\beta}g_{ik}) \\ &\quad +\frac{1}{2}H_{ikl}H_j{}^{kl} + 2F_i{}^kF_{jk} \,.\end{aligned}$$

If we define $y(x,t) = \varphi_t(x)$ by the equations

$$\begin{cases} \dfrac{\partial y^\alpha}{\partial t} = \dfrac{\partial y^\alpha}{\partial x^k}(g^{jl}(\Gamma^k_{jl} - \tilde{\Gamma}^k_{jl})), \\ y^\alpha(x,0) = x^\alpha \end{cases} \tag{2}$$

and $V_i = g_{ik}g^{jl}(\Gamma^k_{jl} - \tilde{\Gamma}^k_{jl})$, we get the following evolution equations for the pull-back metric

$$\begin{cases} \dfrac{\partial}{\partial t} g_{ij}(x,t) = -2R_{ij} + \nabla_i V_j + \nabla_j V_i + \dfrac{1}{2} H_{ikl} H_j{}^{kl} + 2F_i{}^k F_{jk}, \\ g_{ij}(x,0) = \tilde{g}_{ij}(x), \end{cases} \tag{3}$$

where $\tilde{g}_{ij}(x)$ is the initial metric and $\tilde{\Gamma}^k_{jl}$ is the connection of the initial metric. The initial value problem (2) can be rewritten as

$$\begin{cases} \dfrac{\partial y^\alpha}{\partial t} = g^{jl}\left(\dfrac{\partial^2 y^\alpha}{\partial x^j \partial x^l} + \dfrac{\partial y^\beta}{\partial x^j}\dfrac{\partial y^\gamma}{\partial x^l}\tilde{\Gamma}^\alpha_{\beta\gamma} - \dfrac{\partial y^\alpha}{\partial x^k}\tilde{\Gamma}^k_{jl}\right), \\ y^\alpha(x,0) = x^\alpha. \end{cases} \tag{4}$$

Equation (4) is clearly a strictly parabolic system. Then, we have

$$\begin{aligned} \frac{\partial}{\partial t} g_{ij}(x,t) = {} & \frac{\partial}{\partial x^i}\left\{ g^{kl}\frac{\partial g_{kl}}{\partial x^j} \right\} - \frac{\partial}{\partial x^k}\left\{ g^{kl}\left(\frac{\partial g_{jl}}{\partial x^i} + \frac{\partial g_{il}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^l}\right)\right\} \\ & + \frac{\partial}{\partial x^i}\left\{ g_{jk} g^{pq} \tfrac{1}{2} g^{km}\left(\frac{\partial g_{mq}}{\partial x^p} + \frac{\partial g_{mp}}{\partial x^q} - \frac{\partial g_{pq}}{\partial x^m}\right)\right\} \\ & + \frac{\partial}{\partial x^j}\left\{ (g_{ik} g^{pq} \tfrac{1}{2} g^{km}\left(\frac{\partial g_{mq}}{\partial x^p} + \frac{\partial g_{mp}}{\partial x^q} - \frac{\partial g_{pq}}{\partial x^m}\right)\right\} \\ & + \frac{1}{2} H_{ikl} H_j{}^{kl} + 2F_i{}^k F_{jk} \\ = {} & g^{kl}\frac{\partial^2 g_{ij}}{\partial x^k \partial x^l} + \frac{1}{2} H_{ikl} H_j{}^{kl} + 2F_i{}^k F_{jk} . \end{aligned}$$

As a result, from the original equations, we can obtain

$$\begin{cases} \dfrac{\partial g_{ij}(x,t)}{\partial t} = g^{kl}\dfrac{\partial^2 g_{ij}}{\partial x^k \partial x^l} + \dfrac{1}{2} H_{ikl} H_j{}^{kl} + 2F_i{}^k F_{jk}, \\ \dfrac{\partial A_i}{\partial t} = g^{kl}\dfrac{\partial^2 A_i}{\partial x^k \partial x^l}, \\ \dfrac{\partial B_{ij}}{\partial t} = g^{kl}\dfrac{\partial^2 B_{ij}}{\partial x^k \partial x^l}. \end{cases} \tag{5}$$

Let

$$u_1 = g_{11}, u_2 = g_{12}, u_3 = g_{13}, u_4 = g_{22}, u_5 = g_{23}, u_6 = g_{33},$$
$$u_7 = A_1, u_8 = A_2, u_9 = A_3, u_{10} = B_{12}, u_{11} = B_{13}, u_{12} = B_{23} .$$

The above equations can be rewritten as the following form

$$\frac{\partial u_i}{\partial t} = \sum_{j\,k\,l} a_{ikjl} \frac{\partial^2 u_j}{\partial x^k \partial x^l} + (lower\ order\ terms)$$

$(k, l = 1, 2, 3;\ i, j = 1, 2, \cdots, 12)$, in which

$$a_{ikjl} = g^{kl}\ (j = i), \qquad a_{ikjl} = 0\ (j \neq i) \qquad (i = 1, \cdots, 12),$$

For arbitrary $\xi \in \mathbb{R}^{4\times 11} \setminus \{0\}$, we have

$$\sum_{ijkl} a_{ikjl} \xi^i_k \xi^j_l = \sum_{kl} \sum_i g^{kl} \xi^i_k \xi^i_l > 0.$$

Summarize the above discussions, we have the following lemma.

Lemma 2.5. *The differential operator of the right hand side of (5) with respect to the metric g is uniformly elliptic.*

Proof of Theorem 4.1. Noting Lemmas 4.2, 4.4, 4.5 and the compactness property of M, and using the standard theorem of partial differential equations (see,[1],[27]), we can immediately obtain the local existence of smooth solution of the modified system (5) with the initial value

$$g_{ij}(x, 0) = \tilde{g}_{ij}(x), \quad A_i(x, 0) = \tilde{A}_i(x), \quad B_{ij}(x, 0) = \tilde{B}_{ij}(x).$$

In turn the solution of the gradient flow (1) can be obtained from (4) (or (2)). The proof of the existence of smooth solution is completed.

Now we argue the uniqueness of the solution of the gradient flow (1).

By Lemmas 4.2, 4.4 and the standard theorem of partial differential equations, we can obtain the uniqueness of A and B. For any two solutions $\hat{g}^{(1)}_{ij}$ and $\hat{g}^{(2)}_{ij}$ of the gradient flow (1) with the same initial data, we can solve the initial value problem (4) (or (2)) to get two families $\varphi^{(1)}$ and $\varphi^{(2)}$ of diffeomorphisms of M. Thus we get two solutions

$$g^{(1)}_{ij}(\cdot, t) = (\varphi^{(1)}_t)^* \hat{g}^{(1)}_{ij}(\cdot, t), \qquad g^{(2)}_{ij}(\cdot, t) = (\varphi^{(2)}_t)^* \hat{g}^{(2)}_{ij}(\cdot, t),$$

to the modified evolution (5) equations with the same initial value $g_{ij}(x, 0) = \tilde{g}_{ij}(x)$. The uniqueness result for the strictly parabolic equation implies that $g^{(1)}_{ij} = g^{(2)}_{ij}$. Since the initial value problem (4) is clearly a strictly parabolic system, the corresponding solutions $\varphi^{(1)}$ and $\varphi^{(2)}$ of (4) must agree. Consequently, the metrics $\hat{g}^{(1)}_{ij}$ and $\hat{g}^{(2)}_{ij}$ must agree also. Thus, we have proved Theorem. $\square$

Remark 2.1. we are also able to consider a flow for a similar functional for a four-dimension manifold

$$S_1 = \int_M d^4x\sqrt{g}e^{-f}(\chi+R+4|\nabla\phi|^2)-\frac{\epsilon_H}{2}e^{-f}H\wedge *H-\epsilon_F e^{-f}F\wedge *F+\frac{e}{2}F\wedge F$$

where e is the Euler number $e(\eta)$ of the bundle η. The corresponding flow is given by

$$\begin{cases} \dfrac{\partial g_{ij}}{\partial t} = -2[R_{ij} + \nabla_i\nabla_j f - \dfrac{1}{4}H_{ikl}H_j{}^{kl} - F_i{}^k F_{jk}], \\ \dfrac{\partial B_{ij}}{\partial t} = e^f\nabla_k(e^{-f}H^k{}_{ij}), \\ \dfrac{\partial A_i}{\partial t} = -e^f\nabla_k(e^{-f}F_i{}^k), \\ \dfrac{\partial f}{\partial t} = \chi - 2R - 3\Delta f + |\nabla f|^2 + \dfrac{1}{3}H^2 + \dfrac{3}{2}F^2. \end{cases}$$

By the same argument, we can obtain the same results in sections 3-4.

3. The Monotonicity Formula

Let M be a n-dimensional compact Riemannian manifold with metric g_{ij} , the Levi-Civita connection is given by the Christoffel symbols

$$\Gamma^k_{ij} = \frac{1}{2}g^{kl}\left\{\frac{\partial g_{jl}}{\partial x^i} + \frac{\partial g_{il}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^l}\right\},$$

where (g^{ij}) is the inverse of (g_{ij}). The Riemannian curvature tensors read

$$R^k_{ijl} = \frac{\partial\Gamma^k_{jl}}{\partial x^i} - \frac{\partial\Gamma^k_{il}}{\partial x^j} + \Gamma^k_{ip}\Gamma^p_{jl} - \Gamma^k_{jp}\Gamma^p_{il}, \quad R_{ijkl} = g_{kp}R^p_{ijl}.$$

The Ricci tensor is the contraction

$$R_{ik} = g^{jl}R_{ijkl}$$

and the scalar curvature is

$$R = g^{ij}R_{ij}.$$

For each field we shall consider the gauge equivalent classes of fields. Two metrics g_1, g_2 are in the same equivalent class if and only if they are differ by a diffeomorphism, i.e., there exists a diffeomorphism $f : M \to M$ such that $g_2 = f^*g_1$. Two gauge fields A_1 and A_2 are equivalent if and only if there exists a function α on M such that $A_2 = A_1 + d\alpha$. Two B-fields B_1 and B_2 are equivalent if and only if there exists an one-form β on M such that $B_2 = B_1 + d\beta$.

From the first variation of S, we can obtain the flow equations

$$\begin{cases} \dfrac{\partial g_{ij}}{\partial t} = -2[R_{ij} + \nabla_i\nabla_j f - \dfrac{1}{4}H_{ikl}H_j{}^{kl} - F_i{}^k F_{jk}], \\ \dfrac{\partial B_{ij}}{\partial t} = e^f \nabla_k (e^{-f} H^k{}_{ij}), \\ \dfrac{\partial A_i}{\partial t} = -e^f \nabla_k (e^{-f} F_i{}^k), \\ \dfrac{\partial \phi}{\partial t} = \chi - 2R - 3\Delta f + |\nabla f|^2 + \dfrac{1}{3}H^2 + \dfrac{3}{2}F^2. \end{cases}$$

If φ_t is a one-parameter group of diffeomorphisms generated by a vector field ∇f, we have

$$\frac{\partial g_{ij}}{\partial t} = -2(R_{ij} - \frac{1}{4}H_{ikl}H_j{}^{kl} - F_{ik}F_j{}^k),$$

$$\frac{\partial A_i}{\partial t} = -\nabla_k F_i{}^k + \frac{\partial}{\partial x^i}(\nabla^k f A_k),$$

$$\frac{\partial B_{ij}}{\partial t} = \nabla_k H^k{}_{ij} + \frac{\partial}{\partial x^i}(\nabla^k f B_{kj}) + \frac{\partial}{\partial x^j}(\nabla^k f B_{ik}).$$

Let $\tilde{A} = A - d\beta$ where $\dfrac{\partial \beta}{\partial t} = \nabla^k f A_k$, then $\tilde{F} = F$ and

$$\frac{\partial \tilde{A}_i}{\partial t} = -\nabla_k \tilde{F}_i{}^k.$$

Similarly, let $\tilde{B} = B + d\omega$ where $\dfrac{\partial \omega_i}{\partial t} = \nabla^k f B_{ik}$, then

$$\frac{\partial \tilde{B}_{ij}}{\partial t} = \nabla_k (\tilde{H}^k{}_{ij}).$$

Because A and $\tilde{A}$ (B and $\tilde{B}$) are in the same gauge equivalent class, we still denote $\tilde{A}$ ($\tilde{B}$) as A (B). Now we consider the flow equation

$$\frac{\partial g_{ij}}{\partial t} = -2(R_{ij} - \frac{1}{4}H_{ikl}H_j{}^{kl} - F_{ik}F_j{}^k), \tag{1}$$

$$\frac{\partial A_i}{\partial t} = -\nabla_k F_i{}^k, \tag{2}$$

$$\frac{\partial B_{ij}}{\partial t} = \nabla_k (H^k{}_{ij}). \tag{3}$$

Theorem 3.1. *Let g_{ij}, A_i, B_{ij} and f evolve according to the coupled flow*

$$\begin{cases} \dfrac{\partial g_{ij}}{\partial t} = -2[R_{ij} - \dfrac{1}{4}H_{ikl}H_j{}^{kl} - F_i{}^k F_{jk}], \\ \dfrac{\partial B_{ij}}{\partial t} = \nabla_k H^k{}_{ij}, \\ \dfrac{\partial A_i}{\partial t} = -\nabla_k F_i{}^k, \\ \dfrac{\partial f}{\partial t} = \chi - 2R - 3\Delta f + 2|\nabla f|^2 + \dfrac{1}{3}H^2 + \dfrac{3}{2}F^2. \end{cases}$$

Then

$$\frac{dS}{dt} = \int \Big[(-\chi + R - |\nabla f|^2 + 2\Delta f - \frac{1}{12}H^2 - \frac{1}{2}F^2)^2 + 2(R_{ij} + \nabla_i\nabla_j f - \frac{1}{4}H_{ikl}H_j{}^{kl} - F_i{}^k F_{jk})^2 + 2(\nabla_k F_i{}^k - F_i{}^k \nabla_k f)^2 + \frac{1}{2}(\nabla_k H_k{}^{ij} - H_k{}^{ij}\nabla_k f)^2\Big] e^{-f} dV.$$

In particular S is nondecreasing in time and the monotonicity is strict unless we are on the critical points.

Proof.

$$\begin{aligned}
\frac{dS}{dt} &= \int d^3x\sqrt{g}e^{-f}(\frac{1}{2}g^{ij}\frac{\partial g_{ij}}{\partial t} - \frac{\partial f}{\partial t})(-\chi + R + 2\Delta f - |\nabla f|^2 - \frac{1}{12}H^2 - \frac{1}{2}F^2) \\
&\quad + \int d^3x\sqrt{g}e^{-f}\frac{\partial g_{ij}}{\partial t}(-R_{ij} - \nabla_i\nabla_j f + \frac{1}{4}H_{ikl}H_j{}^{kl} + F_i{}^k F_{jk}) \\
&\quad + \int d^3x\sqrt{g}e^{-f}\frac{\partial A_i}{\partial t}(-2\nabla_k(F_i{}^k e^{-f})e^f) + \frac{\partial B_{ij}}{\partial t}(\frac{1}{2}\nabla_k(H^k{}_{ij}e^{-f})e^f) \\
&= \int(\Delta f - |\nabla f|^2)(-\chi + R - |\nabla f|^2 + 2\Delta f - \frac{1}{12}H^2 - \frac{1}{2}F^2)e^{-f}dV \\
&\quad + \int[-\chi + R - |\nabla f|^2 + 2\Delta f - \frac{1}{12}H^2 - \frac{1}{2}F^2]^2 e^{-f}dV \\
&\quad + \int 2(R_{ij} + \nabla_i\nabla_j f - \frac{1}{4}H_{ikl}H_j{}^{kl} - F_{ik}F_j{}^k)^2 e^{-f}dV \\
&\quad + \int -2\nabla_i\nabla_j f(R_{ij} + \nabla_i\nabla_j f - \frac{1}{4}H_{ikl}H_j{}^{kl} - F_{ik}F_j{}^k)e^{-f}dV \\
&\quad + \int 2(\nabla_k F_i{}^k - F_i{}^k\nabla_k f)^2 e^{-f}dV + \int \frac{1}{2}(\nabla_k H^k{}_{ij} - H^k{}_{ij}\nabla_k f)^2 e^{-f}dV \\
&\quad + \int 2F_i{}^k\nabla_k f(\nabla_k F_i{}^k - F_i{}^k\nabla_k f)e^{-f}dV \\
&\quad + \int \frac{1}{2}H^k{}_{ij}\nabla_k f(\nabla_k H^k{}_{ij} - H^k{}_{ij}\nabla_k f)e^{-f}dV.
\end{aligned}$$

By the similar argument of Ricci flow, we have

$\int(\Delta f - |\nabla f|^2)(R - |\nabla f|^2 + 2\Delta f)e^{-f}dV = 2\int \nabla_i\nabla_j f(\nabla_i\nabla_j f + R_{ij})e^{-f}dV.$

And noting the following properties $\nabla_m F_{ij} + \nabla_j F_{mi} + \nabla_i F_{jm} = 0$, $\nabla_m H_{ijk} = \nabla_i H_{mjk} + \nabla_j H_{imk} + \nabla_k H_{ijm}$, we have

$$\begin{aligned}
&\int(\Delta f - |\nabla f|^2)(-\chi - \frac{1}{12}H^2 - \frac{1}{2}F^2)e^{-f}dV \\
=& \int g^{ij}(\nabla_i\nabla_j f - \nabla_i f\nabla_j f)(-\chi - \frac{1}{12}H^2 - \frac{1}{2}F^2)e^{-f}dV \\
=& \int g^{ij}\nabla_i f\nabla_j(\chi + \frac{1}{12}H^2 + \frac{1}{2}F^2)e^{-f}dV \\
=& \int g^{ij}\nabla_i f(\frac{1}{6}\nabla_j H_{pkl}H^{pkl} + \nabla_j F_{kl}F^{kl})e^{-f}dV \\
=& \int g^{ij}\nabla_i f(\frac{1}{6}(\nabla_p H_{jkl} + \nabla_k H_{pjl} + \nabla_l H_{pkj})H^{pkl} \\
&+ (-\nabla_k F_{lj} - \nabla_l F_{jk})F^{kl})e^{-f}dV \\
=& \int g^{ij}\nabla_i f(\frac{1}{2}\nabla_p H_{jkl}H^{pkl} + 2\nabla_k F_{jl}F^{kl})e^{-f}dV \\
=& \int(-\frac{1}{2}g^{ij}\nabla_p\nabla_i f H_{jkl}H^{pkl} - 2g^{ij}\nabla_k\nabla_i f F_{jl}F^{kl})e^{-f}dV \\
&+ \int \frac{1}{2}g^{ij}H_{jkl}\nabla_i f(-\nabla_p H^{pkl} + \nabla_p f H^{pkl})e^{-f}dV \\
&+ \int 2g^{ij}\nabla_i f F_{jl}(\nabla_k f F^{kl} - \nabla_k F^{kl})e^{-f}dV \\
=& \int 2\nabla_i\nabla_j f(-\frac{1}{4}H_{ikl}H_j{}^{kl} - F_{ik}F_j{}^k)e^{-f}dV \\
&+ \int \frac{1}{2}\nabla_k f H^k{}_{ij}(H^p{}_{ij}\nabla_p f - \nabla_p H^p{}_{ij})e^{-f}dV \\
&+ \int 2\nabla_k f F_i{}^k(F_i{}^k\nabla_k f - \nabla_k F_i{}^k)e^{-f}dV.
\end{aligned}$$

Combining with the above argument, we finish the proof.

Let $u = e^{-f}$ be the lowest eigenfunction of the Schrodinger operator, i.e.

$$(R - \frac{1}{12}H^2 - \frac{1}{2}F^2 - 4\Delta)u = \lambda u,$$

or,

$$R - \frac{1}{12}H^2 - \frac{1}{2}F^2 + 2\Delta f - |\nabla f|^2 = \lambda.$$

It minimizes the functional

$$S(g, A, B, f) = \int_M dV e^{-f/2}(R - \frac{1}{12}H^2 - \frac{1}{2}F^2 - 4\Delta)e^{-f/2} / \int_M e^{-f} dV.$$

We have a new functional

$$\lambda(g, A, B) = inf_{\{f | \int_M e^{-f} dV = 1\}} S(g, A, B, f).$$

Let $\lambda(t) = \lambda(g(t), A(t), B(t))$, we have

$$\frac{d\lambda}{dt} = \int_M (|R_{ij} + \nabla_i \nabla_j f - \frac{1}{4}H_{ikj}H_j^{kl} - F_{ik}F_j^k|^2 + \frac{1}{4}|\nabla^k H_{kij} - H_{kij}\nabla^k f|^2$$
$$+ |\nabla_k F_i^k - F_i^k \nabla_k f|^2)e^{-f} dV.$$

We have then (see also[9]):

1) $\lambda(t)$ is monotone, i.e. $\frac{d\lambda(t)}{dt} \geq 0$.

2) Critical points of (*) are the same as critical points of λ.

4. Critical points

Consider the functional

$$\begin{aligned} S &= \int_M d^3x\sqrt{g}e^{-f}(-\chi + R + |\nabla f|^2) - \frac{1}{2}e^{-f} H \wedge *H - e^{-f} F \wedge *F \\ &= \int d^3x\sqrt{g}e^{-f}(-\chi + R + |\nabla f|^2 - \frac{1}{12}H^2 - \frac{1}{2}F^2). \end{aligned} \tag{1}$$

Its first variation can be expressed as follows

$$\begin{aligned} \delta S &= \int d^3x\sqrt{g}e^{-f}(\frac{1}{2}g^{ij}\delta g_{ij} - \delta f)(-\chi + R + 2\Delta f - |\nabla f|^2 - \frac{1}{12}H^2 - \frac{1}{2}F^2) \\ &+ \int d^3x\sqrt{g}e^{-f}\delta g_{ij}(-R_{ij} - \nabla_i\nabla_j f + \frac{1}{4}H_{ikl}H_j{}^{kl} + F_i{}^k F_{jk}) \\ &+ \int d^3x\sqrt{g}e^{-f}\delta A_i(-2\nabla_k(F_i{}^k e^{-f})e^f) + \delta B_{ij}(\frac{1}{2}\nabla_k(H^k{}_{ij}e^{-f})e^f). \end{aligned} \tag{2}$$

The $U(1)$ gauge field A is a one-form potential whose field strength $F = dA$. The Wess-Zumino field B is a two-form potential whose field strength $H = dB, \eta$ is the volume form, f is a dilaton. And in 3-dimension manifold, the field strength is proportional to the Levi-Civita tensor $H_{\mu\nu\rho} = H(x)\eta_{\mu\nu\rho}$, where $H(x)$ is a scalar field and $\eta^{\mu\nu\rho} = \epsilon^{\mu\nu\rho}/\sqrt{g}$ is the completely skewsymmetric Levi-Civita tensor. Therefore, the critical points satisfy the following equations

$$R_{ij} + \nabla_i \nabla_j f - \frac{1}{4} H_{ikl} H_j{}^{kl} - F_i{}^k F_{jk} = 0, \tag{3}$$

$$\nabla_k (F_i{}^k e^{-f}) = 0, \tag{4}$$

$$\nabla_k (H^k{}_{ij} e^{-f}) = 0, \tag{5}$$

$$-\chi + R + 2\Delta f - |\nabla f|^2 - \frac{1}{12} H^2 - \frac{1}{2} F^2 = 0. \tag{6}$$

Suppose M is a compact Riemannian manifold. From (4) and (1.3), we can obtain $F = H = 0$ at the critical points of the general Ricci flow on M. In fact,

$$\begin{aligned}
\int_M F^2 e^{-f} dV &= \int_M F^{ij} F_{ij} e^{-f} dV = \int_M F^{ij} (\nabla_i A_j - \nabla_j A_i) e^{-f} dV \\
&= 2 \int_M F^{ij} \nabla_i A_j e^{-f} dV = -2 \int_M \nabla_i (F^{ij} e^{-f}) A_j dV = 0,
\end{aligned}$$

$$\begin{aligned}
\int_M H^2 e^{-f} dV &= \int_M H^{ijk} H_{ijk} e^{-f} dV \\
&= \int_M H^{ijk} (\nabla_k B_{ij} + \nabla_i B_{jk} + \nabla_j B_{ki}) e^{-f} dV \\
&= 3 \int_M H^{ijk} \nabla_i B_{jk} e^{-f} dV = -3 \int_M \nabla_i (H^{ijk} e^{-f}) B_{jk} dV = 0.
\end{aligned}$$

Remark: Although the fields F and H do not provide any help in the study of critical points of general Ricci flow for compact Riemannian manifold, they maybe play an important role for the noncompact case.

5. Evolution of Curvatures

By virtue of the curvature tensor evolution equations of the Ricci flow, we can obtain the curvature tensor evolution equations under the gradient flow (1). Let us choose a normal coordinate system $\{x^i\}$ around a fixed point $x \in M$ such that $\frac{\partial g_{ij}}{\partial x^k} = 0$ and $g_{ij}(p) = \delta_{ij}$.

Theorem 5.1. *Under the gradient flow (1), the curvature tensor satisfies the evolution equation*

$$\begin{aligned}\frac{\partial}{\partial t}R_{ijkl} &= \triangle R_{ijkl} + 2(B_{ijkl} - B_{ijlk} - B_{iljk} + B_{ikjl})\\
&- g^{pq}(R_{pjkl}R_{qi} + R_{ipkl}R_{qj} + R_{ijpl}R_{qk} + R_{ijkp}R_{ql})\\
&+ \frac{1}{4}[\nabla_i\nabla_l(H_{kpq}H_j{}^{pq}) - \nabla_i\nabla_k(H_{jpq}H_l{}^{pq})\\
&- \nabla_j\nabla_l(H_{kpq}H_i{}^{pq}) + \nabla_j\nabla_k(H_{ipq}H_l{}^{pq})]\\
&+ \frac{1}{4}g^{mn}(H_{kpq}H_m^{pq}R_{ijnl} + H_{mpq}H_l{}^{pq}R_{ijkn})\\
&+ \nabla_i\nabla_l(F_k{}^pF_{jp}) - \nabla_i\nabla_k(F_j{}^pF_{lp}) - \nabla_j\nabla_l(F_k{}^pF_{ip}) + \nabla_j\nabla_k(F_i{}^pF_{lp})\\
&+ g^{mn}(F_k{}^pF_{mp}R_{ijnl} + F_m^pF_{lp}R_{ijkn}),\end{aligned}$$

where $B_{ijkl} = g^{pr}g^{qs}R_{piqj}R_{rksl}$ *and* $\triangle$ *is the Laplacian with respect to the evolving metric.*

Proof. At the point $x \in M$, which we has chosen a normal coordinate system such that $\dfrac{\partial g_{ij}}{\partial x^k} = 0$, we compute

$$\begin{aligned}\frac{\partial}{\partial t}\Gamma^h_{jl} &= \frac{1}{2}\frac{\partial}{\partial t}g^{hm}\left(\frac{\partial g_{ml}}{\partial x^j} + \frac{\partial g_{mj}}{\partial x^l} - \frac{\partial g_{jl}}{\partial x^m}\right)\\
&+\frac{1}{2}g^{hm}\left[\frac{\partial}{\partial x^j}(\frac{\partial g_{ml}}{\partial t}) + \frac{\partial}{\partial x^l}(\frac{\partial g_{mj}}{\partial t}) - \frac{\partial}{\partial x^m}(\frac{\partial g_{jl}}{\partial t})\right],\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial t}R^h_{ijl} &= \frac{\partial}{\partial x^i}\left(\frac{\partial}{\partial t}\Gamma^h_{jl}\right) - \frac{\partial}{\partial x^j}\left(\frac{\partial}{\partial t}\Gamma^h_{il}\right)\\
&= -\frac{1}{2}g^{hp}g^{qm}\frac{\partial g_{pq}}{\partial t}\left(\frac{\partial^2 g_{ml}}{\partial x^i\partial x^j} + \frac{\partial^2 g_{mj}}{\partial x^i\partial x^l} - \frac{\partial^2 g_{jl}}{\partial x^i\partial x^m}\right)\\
&+ \frac{1}{2}g^{hp}g^{qm}\frac{\partial g_{pq}}{\partial t}\left(\frac{\partial^2 g_{ml}}{\partial x^j\partial x^i} + \frac{\partial^2 g_{mi}}{\partial x^j\partial x^l} - \frac{\partial^2 g_{il}}{\partial x^j\partial x^m}\right)\\
&+ \frac{1}{2}g^{hm}[\frac{\partial^2}{\partial x^i\partial x^l}\left(\frac{\partial g_{mj}}{\partial t}\right) - \frac{\partial^2}{\partial x^i\partial x^m}\left(\frac{\partial g_{jl}}{\partial t}\right)\\
&- \frac{\partial^2}{\partial x^j\partial x^l}\left(\frac{\partial g_{mi}}{\partial t}\right) + \frac{\partial^2}{\partial x^j\partial x^m}\left(\frac{\partial g_{il}}{\partial t}\right)]\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2} g^{hm} [\frac{\partial^2}{\partial x^i \partial x^l} \left(\frac{\partial g_{mj}}{\partial t} \right) - \frac{\partial^2}{\partial x^i \partial x^m} \left(\frac{\partial g_{jl}}{\partial t} \right) \\
&\quad - \frac{\partial^2}{\partial x^j \partial x^l} \left(\frac{\partial g_{mi}}{\partial t} \right) + \frac{\partial^2}{\partial x^j \partial x^m} \left(\frac{\partial g_{il}}{\partial t} \right)] - g^{hp} \frac{\partial g_{pq}}{\partial t} R^q_{ijl} \, , \\
\frac{\partial}{\partial t} R_{ijkl} &= \frac{\partial}{\partial t} R^h_{ijl} g_{kh} + R^h_{ijl} \frac{\partial}{\partial t} g_{kh} \\
&= \frac{1}{2} [\frac{\partial^2}{\partial x^i \partial x^l} \left(\frac{\partial g_{kj}}{\partial t} \right) \\
&\quad - \frac{\partial^2}{\partial x^i \partial x^k} \left(\frac{\partial g_{jl}}{\partial t} \right) - \frac{\partial^2}{\partial x^j \partial x^l} \left(\frac{\partial g_{ki}}{\partial t} \right) + \frac{\partial^2}{\partial x^j \partial x^k} \left(\frac{\partial g_{il}}{\partial t} \right)] \, ,
\end{aligned}
$$

then we have

$$
\begin{aligned}
\frac{\partial}{\partial t} R_{ijkl} &= \frac{\partial^2}{\partial x^i \partial x^k} R_{jl} - \frac{\partial^2}{\partial x^i \partial x^l} R_{kj} + \frac{\partial^2}{\partial x^j \partial x^l} R_{ki} - \frac{\partial^2}{\partial x^j \partial x^k} R_{il} \\
&\quad + \frac{1}{4} [\frac{\partial^2}{\partial x^i \partial x^l} (H_{kpq} H_j{}^{pq}) - \frac{\partial^2}{\partial x^i \partial x^k} (H_{jpq} H_l{}^{pq}) \\
&\quad - \frac{\partial^2}{\partial x^j \partial x^l} (H_{kpq} H_i{}^{pq}) + \frac{\partial^2}{\partial x^j \partial x^k} (H_{ipq} H_l{}^{pq})] \\
&\quad + \frac{\partial^2}{\partial x^i \partial x^l} (F_k{}^p F_{jp}) - \frac{\partial^2}{\partial x^i \partial x^k} (F_j{}^p F_{lp}) \\
&\quad - \frac{\partial^2}{\partial x^j \partial x^l} (F_k{}^p F_{ip}) + \frac{\partial^2}{\partial x^j \partial x^k} (F_i{}^p F_{lp}) \\
&\triangleq I_1 + \frac{1}{4} I_2 + I_3 .
\end{aligned}
$$

By the identity (see[3])

$$
\begin{aligned}
&\nabla_i \nabla_k R_{jl} - \nabla_i \nabla_l R_{jk} - \nabla_j \nabla_k R_{il} + \nabla_j \nabla_l R_{ik} \\
&= \Delta R_{ijkl} + 2(B_{ijkl} - B_{ijlk} - B_{iljk} + B_{ikjl}) - g^{pq}(R_{pjkl} R_{qi} + R_{ipkl} R_{qj})
\end{aligned}
$$

and

$$
\nabla_i \nabla_k R_{jl} = \frac{\partial^2 R_{jl}}{\partial x^i \partial x^k} - R_{ml} \frac{\partial}{\partial x^i} \Gamma^m_{kj} - R_{jm} \frac{\partial}{\partial x^i} \Gamma^m_{kl} \, ,
$$

we have

$$
\begin{aligned}
I_1 = {} & \nabla_i \nabla_k R_{jl} + R_{ml}\frac{\partial}{\partial x^i}\Gamma^m_{kj} + R_{jm}\frac{\partial}{\partial x^i}\Gamma^m_{kl} - \nabla_i \nabla_l R_{kj} \\
& - R_{km}\frac{\partial}{\partial x^i}\Gamma^m_{lj} - R_{mj}\frac{\partial}{\partial x^i}\Gamma^m_{lk} \\
& - \nabla_j \nabla_k R_{il} - R_{ml}\frac{\partial}{\partial x^j}\Gamma^m_{ki} - R_{im}\frac{\partial}{\partial x^j}\Gamma^m_{kl} + \nabla_j \nabla_l R_{ki} \\
& + R_{km}\frac{\partial}{\partial x^j}\Gamma^m_{li} + R_{mi}\frac{\partial}{\partial x^j}\Gamma^m_{lk} \\
= {} & \nabla_i \nabla_k R_{jl} - \nabla_i \nabla_l R_{jk} - \nabla_j \nabla_k R_{il} + \nabla_j \nabla_l R_{ik} - R_{km} R^m_{ijl} + R_{ml} R^m_{ijk} \\
= {} & \Delta R_{ijkl} + 2(B_{ijkl} - B_{ijlk} - B_{iljk} + B_{ikjl}) \\
& - g^{pq}(R_{pjkl}R_{qi} + R_{ipkl}R_{qj} + R_{ijpl}R_{qk} + R_{ijkp}R_{ql}),
\end{aligned}
$$

where $B_{ijkl} = g^{pr}g^{qs}R_{piqj}R_{rksl}$.

Now we compute I_2.

It is easily verified that

$$
\nabla_i \nabla_k (H_{jpq}H_l{}^{pq}) = \frac{\partial^2}{\partial x^i \partial x^k}(H_{jpq}H_l{}^{pq}) - H_{mpq}H_l{}^{pq}\frac{\partial}{\partial x^i}\Gamma^m_{kj} - H_{jpq}H_m^{pq}\frac{\partial}{\partial x^i}\Gamma^m_{kl} .
$$

As a result, we obtain

$$
\begin{aligned}
I_2 = {} & \nabla_i \nabla_l (H_{kpq}H_j{}^{pq}) + H_{mpq}H_j{}^{pq}\frac{\partial}{\partial x^i}\Gamma^m_{lk} \\
& + H_{kpq}H_m^{pq}\frac{\partial}{\partial x^i}\Gamma^m_{lj} - \nabla_i \nabla_k (H_{jpq}H_l{}^{pq}) \\
& - H_{mpq}H_l{}^{pq}\frac{\partial}{\partial x^i}\Gamma^m_{kj} - H_{jpq}H_m^{pq}\frac{\partial}{\partial x^i}\Gamma^m_{kl} \\
& - \nabla_j \nabla_l (H_{kpq}H_i{}^{pq}) - H_{mpq}H_i{}^{pq}\frac{\partial}{\partial x^j}\Gamma^m_{lk} \\
& - H_{kpq}H_m^{pq}\frac{\partial}{\partial x^j}\Gamma^m_{li} + \nabla_j \nabla_k (H_{ipq}H_l{}^{pq}) \\
& + H_{mpq}H_l{}^{pq}\frac{\partial}{\partial x^j}\Gamma^m_{ki} + H_{ipq}H_m^{pq}\frac{\partial}{\partial x^j}\Gamma^m_{kl} \\
= {} & \nabla_i \nabla_l (H_{kpq}H_j{}^{pq}) - \nabla_i \nabla_k (H_{jpq}H_l{}^{pq}) \\
& - \nabla_j \nabla_l (H_{kpq}H_i{}^{pq}) + \nabla_j \nabla_k (H_{ipq}H_l{}^{pq}) \\
& + H_{kpq}H_m^{pq}R^m_{ijl} + H_{mpq}H_l{}^{pq}R^m_{jik} \\
= {} & \nabla_i \nabla_l (H_{kpq}H_j{}^{pq}) - \nabla_i \nabla_k (H_{jpq}H_l{}^{pq}) \\
& - \nabla_j \nabla_l (H_{kpq}H_i{}^{pq}) + \nabla_j \nabla_k (H_{ipq}H_l{}^{pq}) \\
& + g^{mn}(H_{kpq}H_m^{pq}R_{ijnl} + H_{mpq}H_l{}^{pq}R_{ijkn}) .
\end{aligned}
$$

Now it remains to compute the last term. The following identity

$$\nabla_i\nabla_k(F_j{}^pF_{lp}) = \frac{\partial^2}{\partial x^i\partial x^k}(F_j{}^pF_{lp}) - F_m^pF_{lp}\frac{\partial}{\partial x^i}\Gamma_{kj}^m - F_j{}^pF_{mp}\frac{\partial}{\partial x^i}\Gamma_{kl}^m$$

yields

$$\begin{aligned}
I_3 = {} & \nabla_i\nabla_l(F_k{}^pF_{jp}) + F_m^pF_{jp}\frac{\partial}{\partial x^i}\Gamma_{lk}^m + F_k{}^pF_{mp}\frac{\partial}{\partial x^i}\Gamma_{lj}^m - \nabla_i\nabla_k(F_j{}^pF_{lp}) \\
& - F_m^pF_{lp}\frac{\partial}{\partial x^i}\Gamma_{kj}^m - F_j{}^pF_{mp}\frac{\partial}{\partial x^i}\Gamma_{kl}^m - \nabla_j\nabla_l(F_k{}^pF_{ip}) - F_m^pF_{ip}\frac{\partial}{\partial x^j}\Gamma_{lk}^m \\
& - F_k{}^pF_{mp}\frac{\partial}{\partial x^j}\Gamma_{li}^m + \nabla_j\nabla_k(F_i{}^pF_{lp}) + F_m^pF_{lp}\frac{\partial}{\partial x^j}\Gamma_{ki}^m + F_i{}^pF_{mp}\frac{\partial}{\partial x^j}\Gamma_{kl}^m \\
= {} & \nabla_i\nabla_l(F_k{}^pF_{jp}) - \nabla_i\nabla_k(F_j{}^pF_{lp}) - \nabla_j\nabla_l(F_k{}^pF_{ip}) + \nabla_j\nabla_k(F_i{}^pF_{lp}) \\
& + g^{mn}(F_k{}^pF_{mp}R_{ijnl} + F_m^pF_{lp}R_{ijkn}) .
\end{aligned}$$

Combining the above discussions, we complete the proof of the theorem. □

Theorem 5.2. *The Ricci curvature satisfies the following evolution equation*

$$\begin{aligned}
\frac{\partial}{\partial t}R_{ik} = {} & \triangle R_{ik} + 2g^{pr}g^{qs}R_{piqk}R_{rs} - 2g^{pq}R_{pi}R_{qk} \\
& + \frac{1}{4}g^{jl}[\nabla_i\nabla_l(H_{kpq}H_j{}^{pq}) \\
& - \nabla_i\nabla_k(H_{jpq}H_l{}^{pq}) - \nabla_j\nabla_l(H_{kpq}H_i{}^{pq}) + \nabla_j\nabla_k(H_{ipq}H_l{}^{pq})] \\
& + \frac{1}{4}g^{mn}(H_{kpq}H_m^{pq}R_{in} - g^{jl}H_{mpq}H_l{}^{pq}R_{ijkn}) \\
& + g^{jl}[\nabla_i\nabla_l(F_k{}^pF_{jp}) - \nabla_i\nabla_k(F_j{}^pF_{lp}) \\
& - \nabla_j\nabla_l(F_k{}^pF_{ip}) + \nabla_j\nabla_k(F_i{}^pF_{lp})] \\
& + g^{mn}(F_k{}^pF_{mp}R_{in} - g^{jl}F_m^pF_{lp}R_{ijkn}).
\end{aligned}$$

Proof. By Theorem 5.1, we can compute

$$
\begin{aligned}
\frac{\partial}{\partial t}R_{ik} &= \frac{\partial}{\partial t}R_{ijkl}g^{jl} + R_{ijkl}\frac{\partial}{\partial t}g^{jl} = \frac{\partial}{\partial t}R_{ijkl}g^{jl} - g^{jp}g^{lq}R_{ijkl}\frac{\partial}{\partial t}g_{pq} \\
&= g^{jl}[\triangle R_{ijkl} + 2(B_{ijkl} - B_{ijlk} - B_{iljk} + B_{ikjl}) \\
&\quad - g^{pq}(R_{pjkl}R_{qi} + R_{ipkl}R_{qj} + R_{ijpl}R_{qk} \\
&\quad + R_{ijkp}R_{ql})] + 2g^{jp}g^{lq}R_{ijkl}R_{pq} \\
&\quad + \frac{1}{4}g^{jl}[\nabla_i\nabla_l(H_{kpq}H_j{}^{pq}) - \nabla_i\nabla_k(H_{jpq}H_l{}^{pq}) \\
&\quad - \nabla_j\nabla_l(H_{kpq}H_i{}^{pq}) + \nabla_j\nabla_k(H_{ipq}H_l{}^{pq})] \\
&\quad + \frac{1}{4}g^{jl}g^{mn}(H_{kpq}H_m^{pq}R_{ijnl} + H_{mpq}H_l{}^{pq}R_{ijkn}) \\
&\quad - \frac{\epsilon_H}{2}R_{ijkl}g^{jp}g^{lq}H_{pmn}H_q{}^{mn} + g^{jl}[\nabla_i\nabla_l(F_k{}^pF_{jp}) \\
&\quad - \nabla_i\nabla_k(F_j{}^pF_{lp}) - \nabla_j\nabla_l(F_k{}^pF_{ip}) + \nabla_j\nabla_k(F_i{}^pF_{lp})] \\
&\quad + g^{mn}(F_k{}^pF_{mp}R_{in} + g^{jl}F_m^pF_{lp}R_{ijkn}) - 2\epsilon_F R_{ijkl}g^{jp}g^{lq}F_p{}^mF_{qm} \\
&= \triangle R_{ik} + 2g^{pr}g^{qs}R_{piqk}R_{rs} - 2g^{pq}R_{pi}R_{qk} \\
&\quad + \frac{1}{4}g^{jl}[\nabla_i\nabla_l(H_{kpq}H_j{}^{pq}) - \nabla_i\nabla_k(H_{jpq}H_l{}^{pq}) \\
&\quad - \nabla_j\nabla_l(H_{kpq}H_i{}^{pq}) + \nabla_j\nabla_k(H_{ipq}H_l{}^{pq})] \\
&\quad + \frac{1}{4}g^{mn}(H_{kpq}H_m^{pq}R_{in} - g^{jl}H_{mpq}H_l{}^{pq}R_{ijkn}) \\
&\quad + g^{jl}[\nabla_i\nabla_l(F_k{}^pF_{jp}) - \nabla_i\nabla_k(F_j{}^pF_{lp}) \\
&\quad - \nabla_j\nabla_l(F_k{}^pF_{ip}) + \nabla_j\nabla_k(F_i{}^pF_{lp})] \\
&\quad + g^{mn}(F_k{}^pF_{mp}R_{in} - g^{jl}F_m^pF_{lp}R_{ijkn}).
\end{aligned}
$$

□

Theorem 5.3. *The scalar curvature satisfies the following evolution equation*

$$
\begin{aligned}
\frac{\partial}{\partial t}R &= \triangle R + 2|Ric|^2 + \frac{1}{2}g^{jl}g^{ik}[\nabla_i\nabla_l(H_{kpq}H_j{}^{pq}) - \nabla_i\nabla_k(H_{jpq}H_l{}^{pq})] \\
&\quad + 2g^{jl}g^{ik}[\nabla_i\nabla_l(F_k{}^pF_{jp}) - \nabla_i\nabla_k(F_j{}^pF_{lp})] \\
&\quad - g^{ip}R_{ik}\left(\frac{1}{2}H_{pmn}H^{kmn} + 2F_{pm}F^{km}\right).
\end{aligned}
$$

Proof. By a direct calculation, we have

$$\begin{aligned}
\frac{\partial}{\partial t}R &= \frac{\partial}{\partial t}R_{ik}g^{ik} + R_{ik}\frac{\partial}{\partial t}g^{ik} = \frac{\partial}{\partial t}R_{ik}g^{ik} - R_{ik}g^{ip}g^{kq}\frac{\partial}{\partial t}g_{pq} \\
&= g^{ik}(\triangle R_{ik} + 2g^{pr}g^{qs}R_{piqk}R_{rs} - 2g^{pq}R_{pi}R_{qk}) + 2g^{ip}g^{kq}R_{ik}R_{pq} \\
&\quad + \frac{1}{4}g^{jl}g^{ik}[\nabla_i\nabla_l(H_{kpq}H_j{}^{pq}) - \nabla_i\nabla_k(H_{jpq}H_l{}^{pq}) \\
&\quad - \nabla_j\nabla_l(H_{kpq}H_i{}^{pq}) + \nabla_j\nabla_k(H_{ipq}H_l{}^{pq})] \\
&\quad + \frac{1}{4}g^{ik}g^{mn}(H_{kpq}H_m^{\ pq}R_{in} - g^{jl}H_{mpq}H_l{}^{pq}R_{ijkn}) \\
&\quad - \frac{1}{2}g^{ip}g^{kq}R_{ik}H_{pmn}H_q{}^{mn} \\
&\quad + g^{ik}g^{jl}[\nabla_i\nabla_l(F_k{}^pF_{jp}) - \nabla_i\nabla_k(F_j{}^pF_{lp}) \\
&\quad - \nabla_j\nabla_l(F_k{}^pF_{ip}) + \nabla_j\nabla_k(F_i{}^pF_{lp})] \\
&\quad + g^{ik}g^{mn}(F_k{}^pF_{mp}R_{in} - g^{jl}F_m^{\ p}F_{lp}R_{ijkn}) - 2g^{ip}g^{kq}R_{ik}F_p{}^mF_{qm} \\
&= \triangle R + 2|Ric|^2 + \frac{1}{2}g^{jl}g^{ik}[\nabla_i\nabla_l(H_{kpq}H_j{}^{pq}) - \nabla_i\nabla_k(H_{jpq}H_l{}^{pq})] \\
&\quad + 2g^{jl}g^{ik}[\nabla_i\nabla_l(F_k{}^pF_{jp}) - \nabla_i\nabla_k(F_j{}^pF_{lp})] \\
&\quad - g^{ip}R_{ik}\left(\frac{1}{2}H_{pmn}H^{kmn} + 2F_{pm}F^{km}\right). \qquad \square
\end{aligned}$$

Acknowledgements: The work of S. Hu was supported in part by the NNSF of China (Grant No. 10771203) and a renovation grant from the Chinese Academy of Sciences; the work of D. Kong was supported in part by the NNSF of China (Grant No. 10671124) and the NCET of China (Grant No. NCET-05-0390); the work of K. Liu was supported by the NSF and NSF of China.

References

1. H. Amann, *Quasilinear evolution equations and parabolic systems*, Tran. of AMS, **293** (1986), 191-227.
2. H. Amann, *Quasilinear parabolic systems under nonlinear boundary conditions*, Arch. Rat. Mech. Anal., Vol. **92** No. **2** (1986), 153-192.
3. H.-D. Cao and X.-P. Zhu, *A complete proof of the Poincaré and geometrization conjectures - application of the Hamilton-Perelman theory of the Ricci flow*, Asian J. Math., **10** (2006), 165-492.
4. D.DeTurck, *Deforming metrics in the direction of their Ricci tensors*, J. Differential Geom., 18 (1983), 157-162.
5. J. Gegenberg and G. Kunstatter, *Using 3D stringy gravity to understand the Thurston conjecture*, arxiv:hep-th/0306279, 2003.

6. J. Gegenberg, S. Vaidya and J. F. Vázquez-Poritz, *Thurston geometries from eleven dimensions*, arxiv:hep-th/0205276, 2002.
7. M. Giaquinta and G. Modica, *Local existence for quasilinear parabolic systems under nonlinear boundary conditions*, Annali di Matematica Pura ed Applicata, Vol. **149** Iss. **1**(1987), 41-59.
8. R. S. Hamilton, *Three manifolds with positive Ricci curvature*, J. Differential Geom., **17** (1982), 255-306.
9. T. Oliynyk, V. Suneeta, E. Woolgar, *A gradient flow for worldsheet nonlinear sigma models*, hep-th/0510239.
10. G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, math.DG/0211159.
11. W. P. Thurston, *Three-dimensional geometry and topology*, Vol.1, Edited by Silvio Levy, Princeton Mathematical Series, 35, Princeton University Press, Princeton, NJ, 1997.

Unitary Representations of the Artin Braid Groups and Quantum Algorithms for Colored Jones Polynomials and the Witten-Reshetikhin Invariant

Louis H. Kauffman and Samuel J. Lomonaco Jr.

Department of Mathematics, Statistics and Computer Science (m/c 249)
851 South Morgan Street, University of Illinois at Chicago
Chicago, Illinois 60607-7045, USA
** E-mail:kauffman@uic.edu*

Department of Computer Science and Electrical Engineering
University of Maryland Baltimore County
1000 Hilltop Circle, Baltimore, MD 21250, USA
** E-mail:lomonaco@comcast.net*

We review the q-deformed spin network approach to Topological Quantum Field Theory and apply these methods to produce unitary representations of the braid groups that are dense in the unitary groups. The simplest case of these models is the Fibonacci model, itself universal for quantum computation. We here formulate these braid group representations in a form suitable for computation and algebraic work. In particular, we give quantum algorithms for computing colored Jones polynomials and the Witten-Reshetikhin-Turaev invariant of three-manifolds.

Keywords: knots, links, braids, quantum computing, unitary transformation, spin networks, Jones polynomial, colored Jones polynonmials, Witten-Reshetikhin-Turaev invariant

1. INTRODUCTION

This paper describes the background for topological quantum computing in terms of Temperely – Lieb Recoupling Theory and gives an explicit description of the resulting unitary representations of the Artin braid group, including the Fibonacci model as the simplest case. This paper is a modified version of our papesr.[14,15] In particular, we give quantum algorithms for computing colored Jones polynomials and the Witten-Reshetikhin-Turaev invariant of three-manifolds.

We use a recoupling theory that generalizes standard angular momentum recoupling theory, generalizes the Penrose theory of spin networks and is inherently topological. Temperely – Lieb Recoupling Theory is based on the bracket polynomial model for the Jones polynomial. It is built in terms of diagrammatic combinatorial topology. The same structure can be explained in terms of the $SU(2)_q$ quantum group, and has relationships with functional integration and Witten's approach to topological quantum field theory. Nevertheless, the approach given here will be unrelentingly elementary. Elementary, does not necessarily mean simple. In this case an architecture is built from simple beginnings and this archictecture and its recoupling language can be applied to many things including: colored Jones polynomials, Witten–Reshetikhin–Turaev invariants of three manifolds, topological quantum field theory and quantum computing.

The contents of this paper are based upon the work in[13,15] and we shall refer to results from those papers.

In quantum computing, the application is most interesting because the recoupling theory yields representations of the Artin Braid group into unitary groups $U(n)$. These represententations are *dense* in the unitary group, and can be used to model quantum computation universally in terms of representations of the braid group. Hence the term: topological quantum computation.

In this paper, we outline the basics of the Temperely – Lieb Recoupling Theory, and show explicitly how unitary representations of the braid group arise from it. We will return to this subject in more detail in subsequent papers. In particular, we do not describe the context of anyonic models for quantum computation in this paper. Rather, we concentrate here on showing how naturally unitary representations of the braid group arise in the context of the Temperely – Lieb Theory. For the reader interested in the relevant background in anyonic topological quantum computing we recommend the following references {[2–6,16,17,19,20] }.

In the last section of this paper (Section 4) we show how these methods lead naturally to quantum algorithms for the computation of the colored Jones polynomials and the Witten-Reshetikhin-Turaev invariants of three-manifolds.

Here is a very condensed presentation of how unitary representations of the braid group are constructed via topological quantum field theoretic methods. For simplicity assmue that one has a single (mathematical) particle with label P that can interact with itself to produce either itself labeled P, or itself with the null label $*$. When $*$ interacts with P the result is always P. When $*$ interacts with $*$ the result is always $*$. One considers process spaces where a row of particles labeled P can successively interact, subject to the restriction that the end result is P. For example the space $V[(ab)c]$ denotes the space of interactions of three particles labeled P. The particles are placed in the positions a, b, c. Thus we begin with $(PP)P$. In a typical sequence of interactions, the first two P's interact to produce a $*$, and the $*$ interacts with P to produce P.

$$(PP)P \longrightarrow (*)P \longrightarrow P.$$

In another possibility, the first two P's interact to produce a P, and the P interacts with P to produce P.

$$(PP)P \longrightarrow (P)P \longrightarrow P.$$

It follows from this analysis that the space of linear combinations of processes $V[(ab)c]$ is two dimensional. The two processes we have just described can be taken to be the the qubit basis for this space. One obtains a representation of the three strand Artin braid group on $V[(ab)c]$ by assigning appropriate phase changes to each of the generating processes. One can think of these phases as corresponding to the interchange of the particles labeled a and b in the association $(ab)c$. The other operator for this representation corresponds to the interchange of b and c. This interchange is accomplished by a *unitary change of basis mapping*

$$F : V[(ab)c] \longrightarrow V[a(bc)].$$

If

$$A : V[(ab)c] \longrightarrow V[(ba)c : d]$$

is the first braiding operator (corresponding to an interchange of the first two particles in the association) then the second operator

$$B : V[(ab)c] \longrightarrow V[(ac)b]$$

is accomplished via the formula $B = F^{-1}AF$ where the A in this formula acts in the second vector space $V[a(bc)]$ to apply the phases for the interchange of b and c.

In this scheme, vector spaces corresponding to associated strings of particle interactions are interrelated by *recoupling transformations* that generalize the mapping F indicated above. A full representation of the Artin braid group on each space is defined in terms of the local intechange phase gates and the recoupling transfomations. These gates and transformations have to satisfy a number of identities in order to produce a well-defined representation of the braid group. These identities were discovered originally in relation to topological quantum field theory. In our approach the structure of phase gates and recoupling transformations arise naturally from the structure of the bracket model for the Jones polynomial.[8] Thus we obtain a knot-theoretic basis for topological quantum computing.

2. Spin Networks and Temperley – Lieb Recoupling Theory

In this section we discuss a combinatorial construction for spin networks that generalizes the original construction of Roger Penrose.[18] The result of this generalization is a structure that satisfies all the properties of a graphical $TQFT$ as described in our paper on braiding and universal quantum gates,[12] and specializes to classical angular momentum recoupling theory in the limit of its basic variable. The construction is based on the properties of the bracket polynomial.[9] A complete description of this theory can be found in the book "Temperley – Lieb Recoupling Theory and Invariants of Three-Manifolds" by Kauffman and Lins.[11]

The "q-deformed" spin networks that we construct here are based on the bracket polynomial relation. View 2.1 and 2.2.

In Figure 2.1 we indicate how the basic projector (symmetrizer, Jones-Wenzl projector) is constructed on the basis of the bracket polynomial expansion.[9] In this technology, a symmetrizer is a sum of tangles on n strands (for a chosen integer n). The tangles are made by summing over braid lifts of permutations in the symmetric group on n letters, as indicated in Figure 2.1. Each elementary braid is then expanded by the bracket polynomial relation, as indicated in Figure 2.1, so that the resulting sum consists of flat tangles without any crossings (these can be viewed as elements in the Temperley – Lieb algebra). The projectors have the property that the concatenation of a projector with itself is just that projector, and if you tie two

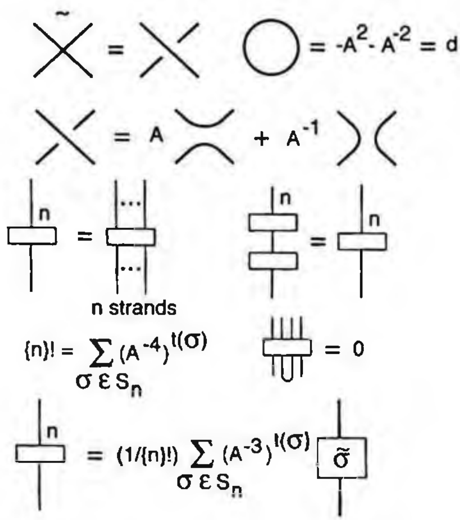


Fig. 2.1. **Basic Projectors**

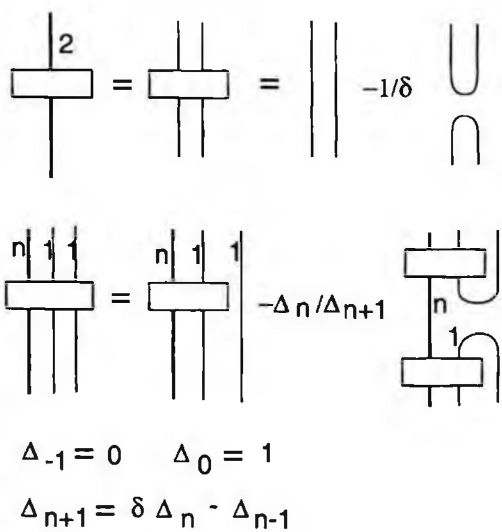


Fig. 2.2. **Two Strand Projector**

lines on the top or the bottom of a projector together, then the evaluation is zero. This general definition of projectors is very useful for this theory.

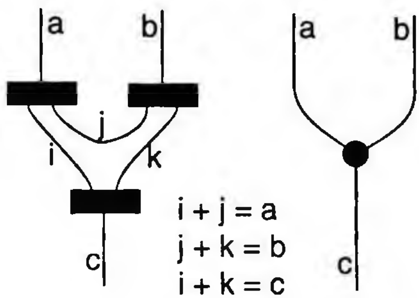


Fig. 2.3. **Trivalent Vertex**

The two-strand projector is shown in Figure 2.2. Here the formula for that projector is particularly simple. It is the sum of two parallel arcs and two turn-around arcs (with coefficient $-1/d$, with $d = -A^2 - A^{-2}$ is the loop value for the bracket polynomial. Figure 2.2 also shows the recursion formula for the general projector. This recursion formula is due to Jones and Wenzl and the projector in this form, developed as a sum in the Temperley - Lieb algebra (see Section 5 of this paper), is usually known as the *Jones–Wenzl projector.*

The projectors are combinatorial analogs of irreducible representations of a group (the original spin nets were based on $SU(2)$ and these deformed nets are based on the quantum group corresponding to SU(2)). As such the reader can think of them as "particles". The interactions of these particles are governed by how they can be tied together into three-vertices. See Figure 2.3. In Figure 2.3 we show how to tie three projectors, of a, b, c strands respectively, together to form a three-vertex. In order to accomplish this interaction, we must share lines between them as shown in that Figure so that there are non-negative integers i, j, k so that $a = i + j, b = j + k, c = i + k$. This is equivalent to the condition that $a + b + c$ is even and that the sum of any two of a, b, c is greater than or equal to the third. For example $a + b \geq c$. One can think of the vertex as a possible particle interaction where $[a]$ and $[b]$ interact to produce $[c]$. That is, any two of the legs of the vertex can be regarded as interacting to produce the third leg.

There is a basic orthogonality of three vertices as shown in Figure 2.4. Here if we tie two three-vertices together so that they form a "bubble" in the middle, then the resulting network with labels a and b on its free ends is a multiple of an a-line (meaning a line with an a-projector on it) or zero (if a is not equal to b). The multiple is compatible with the results of closing the diagram in the equation of Figure 2.4 so the the two free ends are identified with one another. On closure, as shown in the Figure, the left hand side of the equation becomes a Theta graph and the right hand side becomes a multiple of a "delta" where Δ_a denotes the bracket polynomial evaluation of the a-strand loop with a projector on it. The $\Theta(a, b, c)$ denotes the bracket evaluation of a theta graph made from three trivalent vertices and labeled with a, b, c on its edges.

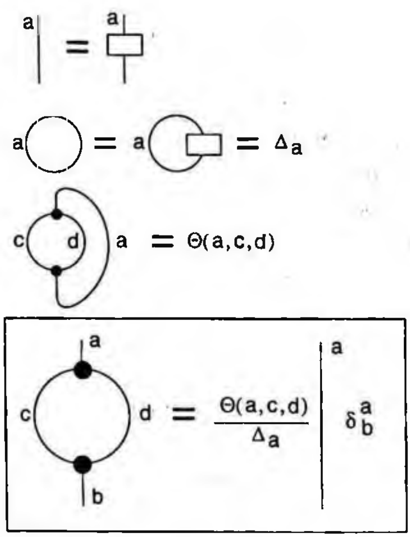


Fig. 2.4. Orthogonality of Trivalent Vertices

There is a recoupling formula in this theory in the form shown in Figure 2.5. Here there are "6-j symbols", recoupling coefficients that can be expressed, as shown in Figure 2.7, in terms of tetrahedral graph evaluations and theta graph evaluations. The tetrahedral graph is shown in Figure 2.6. One derives the formulas for these coefficients directly from the orthogonality relations for the trivalent vertices by closing the left hand side of the recoupling formula and using orthogonality to evaluate the right hand side. This is illustrated in Figure 2.7.

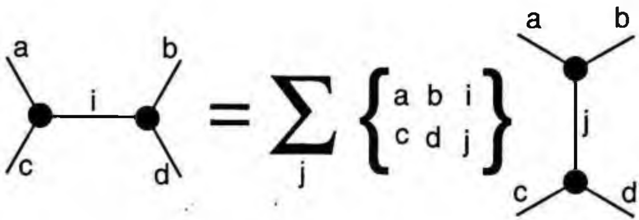


Fig. 2.5. **Recoupling Formula**

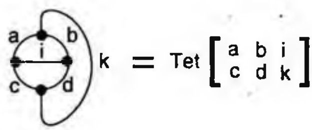
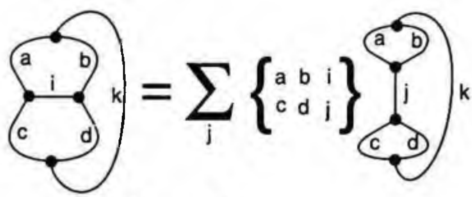


Fig. 2.6. **Tetrahedron Network**

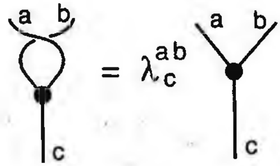


$$= \sum_j \left\{ \begin{matrix} a & b & i \\ c & d & j \end{matrix} \right\} \frac{\Theta(a,b,j)}{\Delta_j} \frac{\Theta(c,d,j)}{\Delta_j} \Delta_j \, \delta_j^k$$

$$= \left\{ \begin{matrix} a & b & i \\ c & d & k \end{matrix} \right\} \frac{\Theta(a,b,k)\,\Theta(c,d,k)}{\Delta_k}$$

$$\boxed{\left\{ \begin{matrix} a & b & i \\ c & d & k \end{matrix} \right\} = \frac{\mathrm{Tet}\left[\begin{matrix} a & b & i \\ c & d & k \end{matrix} \right] \Delta_k}{\Theta(a,b,k)\ \Theta(c,d,k)}}$$

Fig. 2.7. **Tetrahedron Formula for Recoupling Coefficients**



$$\lambda^{ab}_{c} = (-1)^{(a+b-c)/2} A^{(a'+b'-c')/2}$$

$$x' = x(x+2)$$

Fig. 2.8. **LocalBraidingFormula**

Finally, there is the braiding relation, as illustrated in Figure 2.8.

With the braiding relation in place, this q-deformed spin network theory satisfies the pentagon, hexagon and braiding naturality identities needed for a topological quantum field theory. All these identities follow naturally from the basic underlying topological construction of the bracket polynomial. One can apply the theory to many different situations.

2.1. *Evaluations*

In this section we discuss the structure of the evaluations for Δ_n and the theta and tetrahedral networks. We refer to[11] for the details behind these formulas. Recall that Δ_n is the bracket evaluation of the closure of the n-strand projector, as illustrated in Figure 2.4. For the bracket variable A, one finds that

$$\Delta_n = (-1)^n \frac{A^{2n+2} - A^{-2n-2}}{A^2 - A^{-2}}.$$

One sometimes writes the *quantum integer*

$$[n] = (-1)^{n-1}\Delta_{n-1} = \frac{A^{2n} - A^{-2n}}{A^2 - A^{-2}}.$$

If

$$A = e^{i\pi/2r}$$

where r is a positive integer, then

$$\Delta_n = (-1)^n \frac{sin((n+1)\pi/r)}{sin(\pi/r)}.$$

Here the corresponding quantum integer is

$$[n] = \frac{sin(n\pi/r)}{sin(\pi/r)}.$$

Note that $[n+1]$ is a positive real number for $n = 0, 1, 2, ...r-2$ and that $[r-1] = 0$.

The evaluation of the theta net is expressed in terms of quantum integers by the formula

$$\Theta(a,b,c) = (-1)^{m+n+p}\frac{[m+n+p+1]![n]![m]![p]!}{[m+n]![n+p]![p+m]!}$$

where

$$a = m+p, b = m+n, c = n+p.$$

Note that

$$(a+b+c)/2 = m+n+p.$$

When $A = e^{i\pi/2r}$, the recoupling theory becomes finite with the restriction that only three-vertices (labeled with a, b, c) are *admissible* when $a+b+c \leq 2r-4$. All the summations in the formulas for recoupling are restricted to admissible triples of this form.

2.2. *Symmetry and Unitarity*

The formula for the recoupling coefficients given in Figure 2.7 has less symmetry than is actually inherent in the structure of the situation. By multiplying all the vertices by an appropriate factor, we can reconfigure the formulas in this theory so that the revised recoupling transformation is orthogonal, in the sense that its transpose is equal to its inverse (compare with[7]). This is a very useful fact. It means that when the resulting matrices are real, then the recoupling transformations are unitary.

Figure 2.9 illustrates this modification of the three-vertex. Let $Vert[a,b,c]$ denote the original 3-vertex of the Temperley – Lieb recoupling theory. Let $ModVert[a,b,c]$ denote the modified vertex. Then we have the formula

$$ModVert[a,b,c] = \frac{\sqrt{\sqrt{\Delta_a\Delta_b\Delta_c}}}{\sqrt{\Theta(a,b,c)}}\, Vert[a,b,c].$$

Lemma. For the bracket evaluation at the root of unity $A = e^{i\pi/2r}$ the factor

$$f(a,b,c) = \frac{\sqrt{\sqrt{\Delta_a \Delta_b \Delta_c}}}{\sqrt{\Theta(a,b,c)}}$$

is real, and can be taken to be a positive real number for (a,b,c) admissible (i.e. with $a+b+c \leq 2r-4$).

Proof. See our basic reference.[13]

In[13] we show how this modification of the vertex affects the non-zero term of the orthogonality of trivalent vertices (compare with Figure 2.4). We refer to this as the "modified bubble identity." The coefficient in the modified bubble identity is

$$\sqrt{\frac{\Delta_b \Delta_c}{\Delta_a}} = (-1)^{(b+c-a)/2} \sqrt{\frac{[b+1][c+1]}{[a+1]}}$$

where (a,b,c) form an admissible triple. In particular $b+c-a$ is even and hence this factor can be taken to be positive real.

We rewrite the recoupling formula in this new basis and emphasize that the recoupling coefficients can be seen (for fixed external labels a,b,c,d) as a matrix transforming the horizontal "double-Y" basis to a vertically disposed double-Y basis. In Figure 2.10 and Figure 3.1 we have shown the form of this transformation,using the matrix notation

$$M[a,b,c,d]_{ij}$$

for the modified recoupling coefficients. In Figure 3.1 we show an explicit formula for these matrix elements. The proof of this formula follows directly from trivalent–vertex orthogonality (See Figure 2.4 and Figure 2.7.), and is given in.[13] The result shown in Figure 3.1 is the following formula for the recoupling matrix elements.

$$M[a,b,c,d]_{ij} = ModTet \begin{pmatrix} a & b & i \\ c & d & j \end{pmatrix} / \sqrt{\Delta_a \Delta_b \Delta_c \Delta_d}$$

where $\sqrt{\Delta_a \Delta_b \Delta_c \Delta_d}$ is short-hand for the product

$$\sqrt{\frac{\Delta_a \Delta_b}{\Delta_j}} \sqrt{\frac{\Delta_c \Delta_d}{\Delta_j}} \Delta_j$$

$$= (-1)^{(a+b-j)/2}(-1)^{(c+d-j)/2}(-1)^{j}\sqrt{\frac{[a+1][b+1]}{[j+1]}}\sqrt{\frac{[c+1][d+1]}{[j+1]}}[j+1]$$

$$= (-1)^{(a+b+c+d)/2}\sqrt{[a+1][b+1][c+1][d+1]}$$

In this form, since (a, b, j) and (c, d, j) are admissible triples, we see that this coefficient can be taken to be positive real, and its value is independent of the choice of i and j. The matrix $M[a, b, c, d]$ is real-valued.

It follows from Figure 2.10 (turn the diagrams by ninety degrees) that

$$M[a, b, c, d]^{-1} = M[b, d, a, c].$$

Figure 10 implies the formula

$$M[a, b, c, d]^{T} = M[b, d, a, c].$$

It follows from this formula that

$$M[a, b, c, d]^{T} = M[a, b, c, d]^{-1}.$$

Hence $M[a, b, c, d]$ is an orthogonal, real-valued matrix.

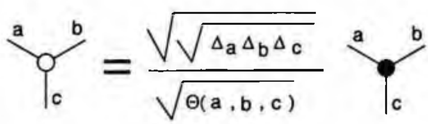


Fig. 2.9. **Modified Three Vertex**

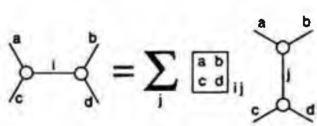


Fig. 2.10. **Modified Recoupling Formula**

Theorem. In the Temperley – Lieb theory we obtain unitary (in fact real orthogonal) recoupling transformations when the bracket variable A has the form $A = e^{i\pi/2r}$. Thus we obtain families of unitary representations of the Artin braid group from the recoupling theory at these roots of unity.

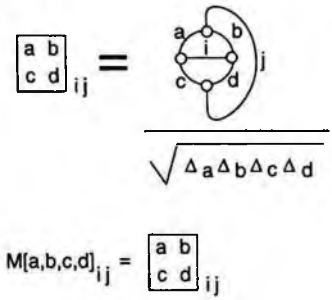


Fig. 2.11. **Modified Recoupling Matrix**

Proof. The proof is given by the discussion above and in.[13]

3. Explicit Form of the Braid Group Representations

In order to have an explicit form for the representations of the braid group that we have constructed we return to the description of the vector spaces in the introduction to this paper. Here we make this description of the vector spaces more precise as follows. We describe a vector space $V[(a_1a_2)a_3 : a_4]$ depending upon a choice of three input and output spins where (ab) denotes the possible outcome of two spin labels interacting at a trivalent vertex as in Figure 2.3. In that figure we see that (ab) can represent c (the remaining leg of the vertex) and that there is a range of values possible for c given by the constraints on i j and k as shown in that figure. Here we insist that the composite interaction $(a_1a_2)a_3$ shall equal a_4 so that the vector space $V[(a_1a_2)a_3 : a_4]$ corresponds to the left-hand tree shown in Figure 3.2. In that figure we indicate the recoupling mapping $F : V[(a_1a_2)a_3 : a_4] \longrightarrow V[a_1(a_2a_3) : a_4]$. The matrix form of F is composed from the recoupling matrix of Figure 3.1. In Figure 3.2 we have labeled $x = (a_1a_2)$ corresponding to one of the basis vectors in $V[(a_1a_2)a_3 : a_4]$. Similarly, we have $y = (a_2a_3)$ corresponding to one of the basis vectors in $V[(a_1a_2)a_3 : a_4]$. We let the corresponding vectors be denoted by $|x\rangle$ and $|y\rangle$ respectively. Then we can write

$$F|i\rangle = \Sigma_j F_{ji}|j\rangle$$

where j ranges over the admissible labels for the interaction of a_2 and a_3.

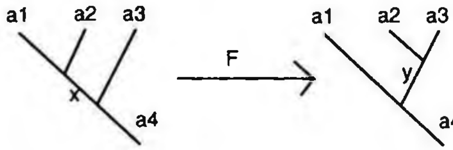


Fig. 3.1. **Recoupling Map** $F : V[(a_1a_2)a_3 : a_4] \longrightarrow V[a_1(a_2a_3) : a_4]$

To see how the three strand braid group acts on $V[(a_1a_2)a_3 : a_4]$, view Figure 13. If we let s_1 denote the generator of the three-stand braid group B_3 that twists the first two strands and s_2 denote the generator that twists the second two strands, then we see that s_1 acts directly at a trivalent vertex, giving the formula

$$s_1|x\rangle = \lambda(a_1, a_2, x)|x\rangle$$

where $\lambda(a_1, a_2, x) = \lambda_x^{a_1,a_2}$ is the braiding factor of Figure 2.8. On the other hand, we need to perform a recoupling in order to compute the action of s_2. As shown in Figure 3.3, we have

$$s_2|i\rangle = \Sigma_{kj} F_{kj}^{-1} \lambda(a_3, a_4, j) F_{ji}|k\rangle.$$

This gives a complete description of the representation of the three-strand braid group on the vector space $V[(a_1a_2)a_3 : a_4]$. Our next task is to generalize this to an abitrary "left-associated" tree.

We wish to consider larger left associated trees such as

$$V[((((a_1a_2)a_3)a_4)a_5) : a_6].$$

To this purpose it is useful to declare that a fully left-associated product may be written without parentheses. Thus we have

$$a_1a_2a_3a_4a_5 = ((((a_1a_2)a_3)a_4)a_5)$$

and

$$a_1(a_2a_3)a_4a_5 = (((a_1(a_2a_3))a_4)a_5).$$

Thus we have the recoupling transformation

$$F^2 : V[a_1a_2a_3a_4a_5 : a_6] \longrightarrow V[a_1(a_2a_3)a_4a_5 : a_6]$$

that will be used for the action of s_2 on the space $V[a_1a_2a_3a_4a_5 : a_6]$.

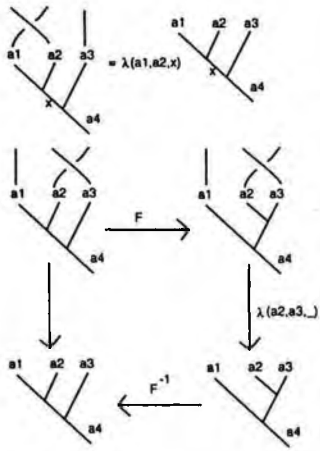


Fig. 3.2. **Action of the Braid Group**

In the general case we have the spaces $V[a_1a_2\cdots a_n : a_{n+1}]$ with basis elements $|x_2x_3\cdots x_{n-1}\rangle$ where (a_1a_2) has x_2 as an outcome, (x_2a_3) has x_3 as an outcome, and so on until $(x_{n-1}a_n)$ has a_{n+1} as an outcome. For articulating the braiding we need mappings

$$F^i : V[a_1a_2\cdots a_n : a_{n+1}] \longrightarrow V[a_1a_2\cdots a_{i-1}(a_ia_{i+1})a_{i+2}\cdots a_n : a_{n+1}].$$

The target space has the strands labeled i and $i+1$ combined at a vertex so that the braiding for s_i in the target space is local. We also need a basis for $V[a_1a_2\cdots a_{i-1}(a_ia_{i+1})a_{i+2}\cdots a_n : a_{n+1}]$. This is given by the kets $|y_2y_3\cdots y_{n-1}\rangle$ where

$$(a_1a_2) = y_2$$

$$\cdots$$

$$(y_{i-2}a_{i-1}) = y_{i+1}$$

$$(a_ia_{i+1}) = y_i$$

$$(y_{i+1}a_{i+2}) = y_{i+2}$$

$$\cdots$$

$$(y_{n-2}a_{n-1}) = y_{n-1}$$

We then have

$$s_i|x_2x_3\cdots x_{n-1}\rangle = (F^i)^{-1}\lambda(a_i,a_{i+1})F^i|x_2x_3\cdots x_{n-1}\rangle.$$

Here it is understood that

$$\lambda(a_i,a_{i+1})|y_2y_3\cdots y_{n-1}\rangle = \lambda(a_i,a_{i+1},y_i)|y_2y_3\cdots y_{n-1}\rangle,$$

where $\lambda(a,b,c)$ is defined as explained above. Finally, using the recoupling matrix formalism of Figure 2.10, we have

$$F^i|x_2x_3\cdots x_{n-1}\rangle = \Sigma_y M[a_i,a_{i+1},x_{i-1},x_{i+1}]_{yx_i}|x_2x_3\cdots x_{i-1}yx_{i+1}\cdots x_{n-1}\rangle.$$

This completes our description of the action of the braid group on these vector spaces.

3.1. *The Fibonacci Model*

In the Fibonacci model,[15] there is a single non-trivial recoupling matrix F.

$$F = \begin{pmatrix} 1/\Delta & 1/\sqrt{\Delta} \\ 1/\sqrt{\Delta} & -1/\Delta \end{pmatrix} = \begin{pmatrix} \tau & \sqrt{\tau} \\ \sqrt{\tau} & -\tau \end{pmatrix}$$

where $\Delta = \frac{1+\sqrt{5}}{2}$ is the golden ratio and $\tau = 1/\Delta$. The local braiding matrix is given by the formula below with $A = e^{3\pi i/5}$.

$$R = \begin{pmatrix} A^8 & 0 \\ 0 & -A^4 \end{pmatrix} = \begin{pmatrix} e^{4\pi i/5} & 0 \\ 0 & -e^{2\pi i/5} \end{pmatrix}.$$

The simplest example of a braid group representation arising from this theory is the representation of the three strand braid group generated by $s_1 = R$ and $s_2 = FRF$ (Remember that $F = F^T = F^{-1}$.). The matrices s_1 and s_2 are both unitary, and they generate a dense subset of $U(2)$, supplying the local unitary transformations needed for quantum computing.

In the Fibonacci model there are two labels, as we described in the introduction (see Figure 4.2): P and $*$. P can interact with itself to produce either P or $*$, while $*$ acts as an identity element. That is, $*$ interacts with P to produce only P, and $*$ interacts with $*$ to produce $*$. Let

$$V[n] = V[a_1a_2\cdots a_n : a_{n+1}] = V(PPP\cdots P : P).$$

The space $V[n]$ has basis vectors $|x_2x_3\cdots x_{n-1}\rangle$ where $\{x_2, x_3, \cdots x_{n-1}\}$ runs over all sequences of P's and $*$'s without consecutive $*$'s. The dimension of $V[n]$ is f_n, the n-th Fibonacci number: $f_1 = 1, f_2 = 1, f_3 = 2, f_4 = 3, f_5 = 5, f_6 = 8, \cdots$ and $f_{n+1} = f_n + f_{n-1}$.

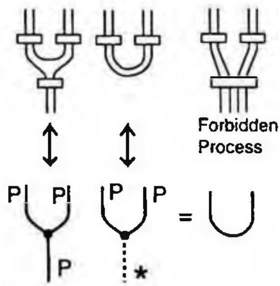


Fig. 3.3. **Fibonacci Vertices**

In terms of the matrix R, we have and $\lambda(*) = A^8$ and $\lambda(P) = -A^4$. The representation of the the braid group B_n on $V[n]$ is given by the formulas below (with $x_0 = x_n = P$ and $i = 1, 2, \cdots n-1$ and the matrix indices for F are $*$ and P corresponding to 0 and 1 respectively). We use the matrix $N = FRF$ below.

$$s_1|x_2x_3\cdots x_{n-1}\rangle = \lambda(x_2)|x_2x_3\cdots x_{n-1}\rangle,$$

and for $i \geq 2$:

$$s_i|x_2x_3\cdots x_{n-1}\rangle = \lambda(x_i)|x_2x_3\cdots x_{n-1}\rangle$$

if $x_{i-1} \neq P$ or $x_{i+1} \neq P$.

$$s_i|x_2x_3\cdots x_{n-1}\rangle = \Sigma_{\alpha=*,P}N_{\alpha,x_i}|x_2x_3\cdots x_{i-1}\,\alpha\,x_{i+1}\cdots x_{n-1}\rangle$$

if $x_{i-1} = x_{i+1} = P$.
These formulas make it possible to do full-scale computer experiments with the Fibonacci model and the generalizations of it that we have discussed. We will pursue this course in a subsequent paper. This model is universal for quantum computation.

4. Quantum Computation of Colored Jones Polynomials and the Witten-Reshetikhin-Turaev Invariant

In this section we make some brief comments on the quantum computation of colored Jones polynomials. This material will be expanded in a subsequent publication.

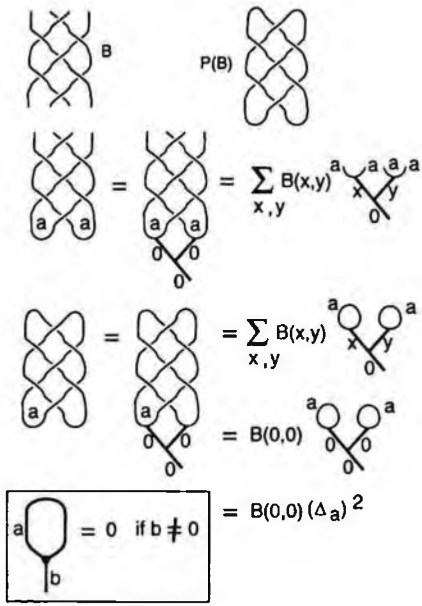


Fig. 4.1. **Evaluation of the Plat Closure of a Braid**

First, consider Figure 4.1. In that figure we illustrate the calculation of the evalutation of the *(a) - colored bracket polynomial* for the *plat closure* $P(B)$ of a braid B. The reader can infer the definition of the plat closure from Figure 15. One takes a braid on an even number of strands and closes the top strands with each other in a row of maxima. Similarly, the bottom strands are closed with a row of minima. It is not hard to see that any knot or link can be represented as the plat closure of some braid.

The (a) - colored bracket polynonmial of a link L, denoted $< L >_a$, is the evaluation of that link where each single strand has been replaced by a parallel strands and the insertion of Jones-Wenzl projector (as discussed in Section 2). We then see that we can use our discussion of the Temperley-Lieb recoupling theory to compute the value of the colored bracket polynomial for the plat closure PB. As shown in Figure 4.1, we regard the braid as acting on a process space $V_0^{a,a,\cdots,a}$ and take the case of the action on the vector v whose process space coordinates are all zero. Then the action of the braid takes the form

$$Bv(0,\cdots,0) = \Sigma_{x_1,\cdots,x_n} B(x_1,\cdots,x_n)v(x_1,\cdots,x_n)$$

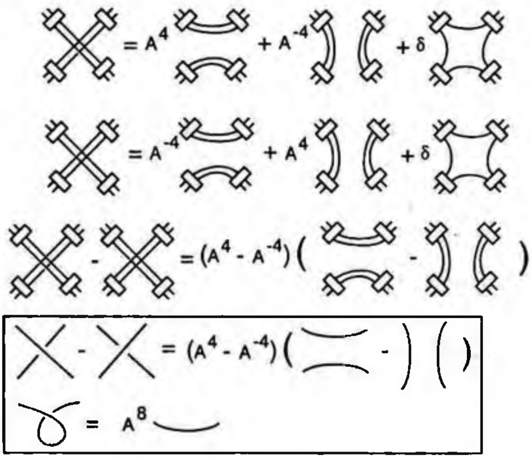


Fig. 4.2. Dubrovnik Polynomial Specialization at Two Strands

where $B(x_1, \cdots, x_n)$ denotes the matrix entries for this recoupling transformation and $v(x_1, \cdots, x_n)$ runs over a basis for the space $V_0^{a,a,\cdots,a}$. Here n is even and equal to the number of braid strands. In the figure we illustrate with $n = 4$. Then, as the figure shows, when we close the top of the braid action to form PB, we cut the sum down to the evaluation of just one term. In the general case we will get

$$< PB >_a = B(0, \cdots, 0)\Delta_a^{n/2}.$$

The calculation simplifies to this degree because of the vanishing of loops in the recoupling graphs. The vanishing result is stated in Figure 15.

The *colored Jones polynomials* are normalized versions of the colored bracket polymomials, differing just by a normalization factor.

In order to consider quantum computation of the colored bracket or colored Jones polynomials, we therefore can consider quantum computation of the matrix entries $B(0, \cdots, 0)$. These matrix entries in the case of the roots of unity $A = e^{i\pi/2r}$ and for the $a = 2$ Fibonacci model with $A = e^{3i\pi/5}$ are parts of the diagonal entries of the unitary transformation that represents the braid group on the process space $V_0^{a,a,\cdots,a}$. *We can obtain these matrix entries by using the Hadamard test as described in the subsection that concludes this section of the paper.* As a result we get relatively efficient quantum algoritms for the colored Jones polynomials at these roots of unity, in essentially the same framework as we described in section 3. We reserve discussion computational complexity of these algorithms, discussion of the factorization of the algorithm into elementary gates and comparison with the results of[1] to a subsequent publication. We point out here that in order to apply the algorithm for a colored Jones polynomial we only require the Hadamard test for a single entry of the unitary matrix that represents the braiding. This is a savings for the algorithm. The methods of Section 3 supply the necessary information for factoring the braiding representation into elementary gates.

These algorithms give not only quantum algorithms for computing the colored bracket and Jones polynomials, but also for computing the Witten-Reshetikhin-Turaev (WRT) invariants at the above roots of unity. The reason for this is that the WRT invariant, in unnormalized form is given as a finite sum of colored bracket polynomials:

$$WRT(L) = \Sigma_{a=0}^{r-2} \Delta_a < L >_a,$$

and so the same computation as shown in Figure 4.1 applies to the WRT. This means that we have, in principle, a quantum algorithm for the computation of the Witten functional integral[21] via this knot-theoretic combinatorial topology. It would be very interesting to understand a more direct approach to such a computation via quantum field theory and functional integration.

Finally, we note that in the case of the Fibonacci model, the (2)-colored bracket polynomial is a special case of the Dubrovnik version of the Kauffman polynomial.[10] See Figure 16 for diagammatics that resolve this fact. The skein relation for the Dubrovnik polynomial is boxed in this figure. Above the box, we show how the double strands with projectors reproduce this relation. This observation means that in the Fibonacci model, the natural underlying knot polynomial is a special evaluation of the Dubrovnik polynomial, and the Fibonacci model can be used to perform quantum computation for the values of this invariant.

4.1. *The Hadamard Test*

In order to (quantum) compute the trace of a unitary matrix U, one can use the *Hadamard test* to obtain the diagonal matrix elements $\langle\psi|U|\psi\rangle$ of U. The trace is then the sum of these matrix elements as $|\psi\rangle$ runs over an orthonormal basis for the vector space. We first obtain

$$\frac{1}{2} + \frac{1}{2} Re\langle\psi|U|\psi\rangle$$

as an expectation by applying the Hadamard gate H

$$H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

$$H|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

to the first qubit of

$$C_U \circ (H \otimes 1)|0\rangle|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes |\psi\rangle + |1\rangle \otimes U|\psi\rangle.$$

Here C_U denotes controlled U, acting as U when the control bit is $|1\rangle$ and the identity mapping when the control bit is $|0\rangle$. We measure the expectation for the first qubit $|0\rangle$ of the resulting state

$$\frac{1}{2}(H|0\rangle \otimes |\psi\rangle + H|1\rangle \otimes U|\psi\rangle) = \frac{1}{2}((|0\rangle + |1\rangle) \otimes |\psi\rangle + (|0\rangle - |1\rangle) \otimes U|\psi\rangle)$$

$$= \frac{1}{2}(|0\rangle \otimes (|\psi\rangle + U|\psi\rangle) + |1\rangle \otimes (|\psi\rangle - U|\psi\rangle)).$$

This expectation is

$$\frac{1}{2}((\langle\psi| + \langle\psi|U^\dagger)(|\psi\rangle + U|\psi\rangle) = \frac{1}{2} + \frac{1}{2} Re\langle\psi|U|\psi\rangle.$$

The imaginary part is obtained by applying the same procedure to

$$\frac{1}{\sqrt{2}}(|0\rangle \otimes |\psi\rangle - i|1\rangle \otimes U|\psi\rangle$$

Note that the Hadamard test enables this quantum computation to estimate the trace of any unitary matrix U by repeated trials that estimate individual matrix entries $\langle\psi|U|\psi\rangle$. We shall return to quantum algorithms for the Jones polynomial and other knot polynomials in a subsequent paper.

References

1. D. Aharonov, V. Jones, Z. Landau, A polynomial quantum algorithm for approximating the Jones polynomial, quant-ph/0511096.
2. M. Freedman, A magnetic model with a possible Chern-Simons phase, quant-ph/0110060v1 9 Oct 2001, (2001), preprint
3. M. Freedman, *Topological Views on Computational Complexity*, Documenta Mathematica - Extra Volume ICM, 1998, pp. 453–464.
4. M. Freedman, M. Larsen, and Z. Wang, A modular functor which is universal for quantum computation, quant-ph/0001108v2, 1 Feb 2000.
5. M. H. Freedman, A. Kitaev, Z. Wang, Simulation of topological field theories by quantum computers, *Commun. Math. Phys.*, **227**, 587-603 (2002), quant-ph/0001071.
6. M. Freedman, Quantum computation and the localization of modular functors, quant-ph/0003128.
7. C. Frohman and J. Barozynska, $SO(3)$-topological quantum field theory,*Comm. Anal. Geom.*, Vol. 4, (1996), No. 4, 589-679.
8. V.F.R. Jones, A polynomial invariant for links via von Neumann algebras, Bull. Amer. Math. Soc. **129** (1985), 103–112.
9. L.H. Kauffman, State models and the Jones polynomial, Topology **26** (1987), 395–407.
10. L. H. Kauffman, An invariant of regular isotopy, *Trans. Amer. Math. Soc.* **318** (1990), no. 2, 417–471.
11. L.H. Kauffman, *Temperley – Lieb Recoupling Theory and Invariants of Three-Manifolds*, Princeton University Press, Annals Studies **114** (1994).
12. L. H. Kauffman and S. J. Lomonaco Jr., Braiding Operators are Universal Quantum Gates, *New Journal of Physics* **6** (2004) 134, pp. 1-39.
13. L. H. Kauffman and S. J. Lomonaco Jr., Spin Networks and anyonic topological computing, In "Quantum Information and Quantum Computation IV", (Proceedings of Spie, April 17-19,2006) edited by E.J. Donkor, A.R. Pirich and H.E. Brandt, Volume 6244, Intl Soc. Opt. Eng., pp. 62440Y-1 to 62440Y-12.
14. L. H. Kauffman and S. J. Lomonaco Jr., Spin Networks and anyonic topological computing II, In "Quantum Information and Quantum Computation V", (Proceedings of Spie, April 10-12,2007) edited by E.J. Donkor, A.R. Pirich and H.E. Brandt, Volume 6573, Intl Soc. Opt. Eng., pp. 65730U-1 to 65730u-13.
15. L. H. Kauffman and S. J. Lomonaco Jr., q - Deformed Spin Networks, Knot Polynomials and Anyonic Topological Quantum Computation, *JKTR* Vol. 16, No. 3 (March 2007), pp. 267-332.
16. A. Kitaev, Anyons in an exactly solved model and beyond, *arXiv.cond-mat/0506438 v1 17 June 2005.*
17. A. Marzuoli and M. Rasetti, Spin network quantum simulator, *Physics Letters A* **306** (2002) 79–87.
18. R. Penrose, Angular momentum: An approach to Combinatorial Spacetime, In *Quantum Theory and Beyond*, edited by T. Bastin, Cambridge University Press (1969).
19. J. Preskill, Topological computing for beginners, (slide presentation), Lecture Notes for Chapter 9 - Physics 219 - Quantum Computation.

http://www.iqi.caltech.edu/ preskill/ph219
20. F. Wilczek, *Fractional Statistics and Anyon Superconductivity,* World Scientific Publishing Company (1990).
21. E. Witten, Quantum field Theory and the Jones Polynomial, *Commun. Math. Phys.*,vol. 121, 1989, pp. 351-399.

A New Approach to Deriving Recursion Relations for the Gromov-Witten Theory

Yon-Seo Kim

Department of Mathematics, University of Chicago, Chicago, Illinois 60637, USA
Email: yskim@math.uchicago.edu

Kefeng Liu

Center of Math Sciences, Zhejiang University, Hangzhou, Zhejiang 310027, China;
Department of Mathematics, UCLA, Los Angeles, CA 90095-1555, USA
Emails: liu@cms.zju.edu.cn, liu@math.ucla.edu

In this article, we obtain a universal method to compute the Gromov-Witten type invariants using the localization technique. This method can be applied to any natural cohomology class on the moduli space of curves $\overline{\mathcal{M}}_{g,n}$. As applications, we illustrate a new proof of the Witten's conjecture as well as the proof of Mariño-Vafa formula.

1. Introduction

The localization technique has been proved to be a useful tool in studying the Gromov-Witten invariants as was shown in the proof of the Mariño-Vafa formula and a new proof of the Witten conjecture/Kontsevich theorem. Especially when applied to the relative stable moduli, it allows us to express the Gromov-Witten invariants in terms of other invariants such as double Hurwitz numbers or another type of Gromov-Witten invariants. In this paper, we present a universal method to obtain recursion relations on the Gromov-Witten type invariants: let $\omega \in H^*(X)$ be any natural cohomology class on the target space X, we can determine the Gromov-Witten invariants

$$\int_{\overline{\mathcal{M}}_{g,n}} \psi_1^{k_1} \cdots \psi_n^{k_n} \omega$$

through the recursion relation obtained by localization technique. It is natural to expect that these invariants should depend only on the degeneration of the domain curves and how the cohomology class ω behaves under the degeneration. The method we present here strongly suggests that this is actually the case, as was suggested by the generalized Witten conjecture and the general Virasoro conjecture.

The rest of this paper is consisted as follows: In section 2, we briefly review the Gromov-Witten invariants, the moduli space of relative stable morphisms, and the virtual localization technique. In section 3, we present the universal approach to the Gromov-Witten invariants using localization technique. In sections 4 and 5, we present, as applications, a new proof of Witten conjecture and the proof of Mariño-Vafa formula. In section 6, we list several open problems to which we can apply this general approach.

2. Preliminaries

2.1. *The Gromov-Witten invariants*

Let X be a smooth projective variety and $\overline{\mathcal{M}}_{g,n}(X,\beta)$ be the moduli stack of n-pointed stable maps of genus g and degree β, i.e. it consists of maps

$$f : (C; x_1, \cdots, x_n) \longrightarrow X$$

such that

- C is a Riemann surface of arithmetic genus $g = h^1(C, \mathcal{O}_C)$ and n marked points $x_1, \cdots, x_n$ with only nodal singularities.
- An algebraic map $f : C \longrightarrow X$ such that $f_*[C] = \beta \in H_2(X, \mathbb{C})$.
- It admits no infinitesimal automorphisms fixing the marked points.

For each marked point x_i, consider the line bundle $\mathbb{L}_i$ over $\overline{\mathcal{M}}_{g,n}(X,\beta)$ whose fiber over $[C; x_1, \cdots, x_n] \in \overline{\mathcal{M}}_{g,n}(X,\beta)$ is the cotangent line $T^*_{x_i}C$ at the i-th marked point x_i. Then define *the ψ-class* as its first Chern-class, i.e. $\psi_i = c_1(\mathbb{L}_i)$. For each i, let $ev_i : \overline{\mathcal{M}}_{g,n}(X,\beta) \longrightarrow X$ be the evaluation map which sends x_i to its image $f(x_i) \in X$. The construction of *virtual fundamental class*, denoted by $[\overline{\mathcal{M}}_{g,n}(X,\beta)]^{vir}$, allows us to do the intersection theory on $\overline{\mathcal{M}}_{g,n}(X,\beta)$. The *Gromov-Witten invariants* are defined as the intersection numbers

$$\langle \tau_{k_1}(x_1) \cdots \tau_{k_n}(x_n) \rangle^X_{g,d} = \int_{[\overline{\mathcal{M}}_{g,n}(X,\beta)]^{vir}} \psi_1^{k_1} \cdots \psi_n^{k_n} \, \mathrm{ev}_1^*(Z_1) \cdots \mathrm{ev}_n^*(Z_n)$$

where $Z_1, \cdots, Z_n$ are cohomology classes of X.

2.2. Localization technique

In this section, we will briefly summarize various versions of localization formulas.[1,15,16] We start with the equivariant cohomology.

2.2.1. Equivariant Cohomology

Let G be a compact Lie group acting on M. The equivariant cohomology of M is defined as the ordinary cohomology of the space M_G obtained from a fixed universal G-bundle EG, by the mixing construction

$$M_G = EG \times_G M$$

Here, G acts on the right of EG and on the left of M, and the notation means that we identify $(pg, q) \sim (p, gq)$ for $p \in EG$, $q \in M$, $g \in G$. Hence M_G is the bundle with fibre M over the classifying space BG associated to the universal bundle $EG \longrightarrow BG$. We have natural projection map $\pi : M_G \longrightarrow BG$ and $\sigma : M_G \longrightarrow M/G$, which fits into the mixing diagram of Cartan and Borel:

$$\begin{tikzcd}
EG \arrow[d] & EG \times M \arrow[l] \arrow[r] \arrow[d] & M \arrow[d] \\
BG & E \times_G M \arrow[l, "\pi"'] \arrow[r, "\sigma"] & M/G
\end{tikzcd}$$

If G acts smoothly on M, then we have $M_G \cong M/G$. This is not true in general but it turns out that M_G is a better functorial construction and the proper homotopy theoretic quotient of M by G. In any case, the equivariant cohomology, denoted by $H^*_G(M)$, is defined by

$$H^*_G(M) = H^*(M_G)$$

and constitutes a contravariant functor from G-spaces to modules over the base ring $H^*_G := H^*_G(pt) = H^*(BG)$. The map σ defines a natural map $\sigma^* : H^*(M/G) \longrightarrow H^*_G(M)$ which is an isomorphism if G acts freely. The inclusion $i : M \longrightarrow M_G$ induces a natural map $i^* : H^*_G(M) \longrightarrow H^*(M)$.

2.2.2. Atiyah-Bott Localization Formula

Let $i : V \hookrightarrow M$ be a map of compact manifolds. The tubular neighborhood of V inside M can be identified with the normal bundle of V. On the total space of the normal bundle, there is the Thom form Φ_V which has compact support in the fibres and integrates to one in each fiber. Extending this form by zero gives a form in M, and multiplying by Φ_V provides a map

$H^*(V) \cong H^{*+k}(M, M\backslash V) \longrightarrow H^*(M)$. In particular, the cohomology class $1 \in H^0(V)$ is sent to the Thom class and this class restricts to the Euler class of the normal bundle $\mathcal{N}_{V/M}$ of V in M. Hence, we see that

$$i^* i_* 1 = e(\mathcal{N}_{V/M}).$$

This also holds in equivariant cohomology by the same argument applied to M_G. The theorem of Atiyah and Bott says that an inverse of the Euler class of the normal bundle always exists along the fixed locus of a group action. Precisely, $i^*/e(\mathcal{N}_{V/M})$ is the inverse of i_* in equivariant cohomology, i.e. for any equivariant class ϕ, we have a decomposition

$$\phi = \sum_F \frac{i_* i^* \phi}{e(\mathcal{N}_{F/M})}$$

where F runs over the fixed locus of G-action. In the integrated form, we have

$$\int_M \phi = \sum_F \int_F \frac{i^* \phi}{e(\mathcal{N}_{F/M})}$$

2.2.3. *Functorial Localization Formula*

Let X and Y be T-manifolds. Assume that $f : X \longrightarrow Y$ is a T-equivariant map, $j_E : E \hookrightarrow Y$ is a fixed component in Y, and $i_F : F \hookrightarrow f^{-1}(E)$ is a fixed component in X. For any equivariant class $\omega \in H_T^*(X)$, we have the commutative diagrams;

$$\begin{tikzcd} F \arrow[r, "i_F"] \arrow[d, "g=f|_F"'] & X \arrow[d, "f"] \\ E \arrow[r, "j_E"] & Y \end{tikzcd} \qquad \begin{tikzcd} \frac{i_F^*(\omega)}{e_T(F/X)} \arrow[d, "g_!"] & \omega \arrow[l, "i_F^*"'] \arrow[d, "f_!"] \\ g_!\left[\frac{i_F^*(\omega)}{e_T(F/X)}\right] & f_!(\omega) \arrow[l, "j_E^*"'] \end{tikzcd}$$

Applying the Atiyah-Bott localization formula with the naturality relation $f_!(\omega \cdot f^*\alpha) = f_!\omega \cdot \alpha$, we obtain *the functorial localization formula:*

$$g_!\left[\frac{i_F^*(\omega)}{e_T(F/X)}\right] = \frac{j_E^* f_!(\omega)}{e_T(E/Y)}$$

2.2.4. *Virtual Functorial Localization Formula*

The above functorial localization formula is also valid in the case where X and F are virtual fundamental classes. In this paper, we will use $[\overline{\mathcal{M}}_g(X \times$

$\mathbb{P}^1; X \times \{\infty\} \mid \beta; \mu)]^{vir}$ for X, and $[F_\Gamma]^{vir}$ for F. Hence for any equivariant class ω, we have:

$$\int_{[\overline{\mathcal{M}}_g(X \times \mathbb{P}^1; X \times \{\infty\} \mid \beta;\mu)]^{vir}} \omega = \sum_{F_\Gamma} \int_{[F_\Gamma]^{vir}} \frac{i_\Gamma^*(\omega)}{e_T(\mathcal{N}_{F_\Gamma})} \tag{1}$$

where $\mathcal{N}_{F_\Gamma}$ denotes the normal bundle of the fixed locus given by F_Γ in the relative stable moduli space $\overline{\mathcal{M}}_g(X \times \mathbb{P}^1; X \times \{\infty\} \mid \beta; \mu)$.

2.3. *The moduli space of relative stable morphisms*

In this section, we briefly summarize the definitions and results from[24–26,30] with minor modifications. Let X be a smooth projective variety and $D^1, \cdots, D^k$ be disjoint smooth divisors. For $\alpha = 1, \cdots, k$, define

$$\Delta(\mathcal{D}^\alpha)(m) = \Delta(\mathcal{D}^\alpha)_1 \cup \cdots \cup \Delta(\mathcal{D}^\alpha)_m$$

where $\Delta(\mathcal{D}^\alpha)_i \cong \mathbb{P}(\mathcal{O}_{\mathcal{D}^\alpha} \oplus \mathcal{N}_{\mathcal{D}^\alpha/X}) \to \mathcal{D}^\alpha$ for each i, α, and $\mathcal{N}_{\mathcal{D}^\alpha/X}$ denotes the normal sheaf of a subvariety $\mathcal{D}^\alpha$ in X. The projective line bundle $\Delta(\mathcal{D}^\alpha) \to \mathcal{D}^\alpha$ has two distinct sections

$$\mathcal{D}_0^\alpha = \mathbb{P}(\mathcal{O}_{\mathcal{D}^\alpha} \oplus 0), \qquad \mathcal{D}_\infty^\alpha = \mathbb{P}(0 \oplus \mathcal{N}_{\mathcal{D}^\alpha/X}).$$

We have $\mathcal{N}_{\mathcal{D}_0^\alpha/\Delta(\mathcal{D}^\alpha)} \cong \mathcal{N}_{\mathcal{D}^\alpha/X}^{-1}$ and $\mathcal{N}_{\mathcal{D}_\infty^\alpha/\Delta(\mathcal{D}^\alpha)} \cong \mathcal{N}_{\mathcal{D}^\alpha/X}$. Then $\Delta(\mathcal{D}^\alpha)(m)$ is constructed by gluing along the two distinct sections of $\Delta(\mathcal{D}^\alpha)_i$'s that correspond to two distinct sections $\mathcal{D}_0^\alpha$ and $\mathcal{D}_\infty^\alpha$. The $\mathbb{C}^*$-action on $\mathcal{O}_{\mathcal{D}^\alpha}$ induces a $\mathbb{C}^*$-action on $\Delta(\mathcal{D}^\alpha)$ such that $\Delta(\mathcal{D}^\alpha) \to \mathcal{D}^\alpha$ is $\mathbb{C}^*$-equivariant, where $\mathbb{C}^*$ acts on $\mathcal{D}^\alpha$ trivially. The two distinct sections $\mathcal{D}_0^\alpha$, $\mathcal{D}_\infty^\alpha$ are fixed under this $\mathbb{C}^*$-action. So there is a $(\mathbb{C}^*)^m$-action on $\Delta(\mathcal{D}^\alpha)(m)$ fixing $\mathcal{D}_0^\alpha, \cdots, \mathcal{D}_m^\alpha$, such that $\Delta(\mathcal{D}^\alpha)(m) \to \mathcal{D}^\alpha$ is $(\mathbb{C}^*)^m$-equivariant, where $(\mathbb{C}^*)^m$ acts on $\mathcal{D}^\alpha$ trivially. The variety

$$X[m^1, \cdots, m^k] = X \cup \bigcup_{\alpha=1}^{k} \Delta(\mathcal{D}^\alpha)(m^\alpha)$$

with normal crossing singularities is obtained by identifying $\mathcal{D}^\alpha \subset X$ with $\mathcal{D}_0^\alpha \subset \Delta(\mathcal{D}^\alpha)$ under the canonical isomorphism. There is a morphism

$$\pi[m^1, \cdots, m^k] : X[m^1, \cdots, m^k] \longrightarrow X$$

which contracts $\Delta(\mathcal{D}^\alpha)(m^\alpha)$ to $\mathcal{D}^\alpha$. The $(\mathbb{C}^*)^{\oplus m^\alpha}$-actions on $\Delta(\mathcal{D}^\alpha)(m^\alpha)$ give a $(\mathbb{C}^*)^{\oplus \sum m^\alpha}$ action on $X[m^1, \cdots, m^k]$ such that $\pi[m^1, \cdots, m^k]$ is $(\mathbb{C}^*)^{\oplus \sum m^\alpha}$-equivariant with respect to the trivial action on X. Let $\beta \in$

$H_2(X,\mathbb{Z})$ be a nonzero homology class and μ^α be a partition of d^α, for $\alpha = 1, \cdots, k$, where d^α's are defined to be

$$d^\alpha = \int_\beta c_1(\mathcal{O}(\mathcal{D}^\alpha)) \geq 0.$$

Define the *relative stable moduli*

$$\overline{\mathcal{M}}_g(X; D^1, \cdots, D^k \mid \beta; \mu^1, \cdots, \mu^k)$$

to be the moduli space of *relative stable morphisms*

$$f : (C; \{x_i^1\}_{i=1}^{l(\mu^1)}, \cdots, \{x_i^k\}_{i=1}^{l(\mu^k)}) \longrightarrow X[m^1, \cdots, m^k]$$

such that

(1) $(C; \{x_i^1\}_{i=1}^{l(\mu^1)}, \cdots, \{x_i^k\}_{i=1}^{l(\mu^k)})$ is a connected prestable curve of arithmetic genus g with $\sum_{\alpha=1}^k l(\mu^\alpha)$ marked points.
(2) $(\pi[m^1, \cdots, m^k] \circ f)_*[C] = \beta \in H_2(X, \mathbb{Z})$.
(3) As Cartier divisors, we have $f^{-1}(\mathcal{D}^\alpha_{(m^\alpha)}) = \sum_{i=1}^{l(\mu^\alpha)} \mu_i^\alpha x_i^\alpha$. In particular, if $d^\alpha = 0$, then $f^{-1}(\mathcal{D}^\alpha_{(m^\alpha)})$ is empty.
(4) The preimage of $\mathcal{D}_l^\alpha$ consists of nodes of C for $l = 0, \cdots, m^\alpha - 1$. If $f(y) \in \mathcal{D}_l^\alpha$ and C_1,C_2 are two irreducible components of C which intersect at y, then $f|_{C_1}$ and $f|_{C_2}$ have the same contact order to $D^\alpha_{(l)}$ at y.
(5) The automorphism group of f is finite.

The arguments in J.Li's papers[24-26] show that $\overline{\mathcal{M}}_g(X; D^\alpha \mid \beta; \mu^\alpha)$ is a separated, proper Deligne-Mumford stack which admits a perfect obstruction theory of virtual dimension

$$\int_\beta c_1(TX) + (1-g)(\dim X - 3) + \sum_{\alpha=1}^k (l(\mu^\alpha) - |\mu^\alpha|)$$

In order to perform the virtual localization computation on the relative stable moduli $\overline{\mathcal{M}}_g(X; D^1, \cdots, D^k \mid \beta; \mu^1, \cdots, \mu^k)$, we need to compute the Euler class of the tangent space T^1 and the obstruction space T^2 of $\overline{\mathcal{M}}_g(X; D^1, \cdots, D^k \mid \beta; \mu^1, \cdots, \mu^k)$ at the moduli point

$$f : (C; \{x_i^1\}_{i=1}^{l(\mu^1)}, \cdots, \{x_i^k\}_{i=1}^{l(\mu^k)}) \longrightarrow X[m^1, \cdots, m^k].$$

This can be done by the following two exact sequences:

$$\begin{aligned} 0 \longrightarrow & \mathrm{Ext}^0(\Omega_C(R), \mathcal{O}_C) \longrightarrow H^0(\mathbf{D}^\bullet) \longrightarrow T^1 \\ \longrightarrow & \mathrm{Ext}^1(\Omega_C(R), \mathcal{O}_C) \longrightarrow H^1(\mathbf{D}^\bullet) \longrightarrow T^2 \longrightarrow 0 \end{aligned}$$

$$
\begin{aligned}
0 &\to H^0\Big(C, f^*\big(\Omega_{X[m^1,\cdots,m^k]}(\sum_{\alpha=1}^{k}\log \mathcal{D}^{\alpha}_{m^\alpha})\big)^\vee\Big) \\
&\to H^0(\mathbf{D}^\bullet) \to \bigoplus_{\alpha=1}^{k}\bigoplus_{l=0}^{m^\alpha-1} H^0_{\mathrm{et}}(\mathbf{R}^{\alpha\bullet}_l) \\
&\to H^1\Big(C, f^*\big(\Omega_{X[m^1,\cdots,m^k]}(\sum_{\alpha=1}^{k}\log \mathcal{D}^{\alpha}_{m^\alpha})\big)^\vee\Big) \\
&\to H^1(\mathbf{D}^\bullet) \to \bigoplus_{\alpha=1}^{k}\bigoplus_{l=0}^{m^\alpha-1} H^1_{\mathrm{et}}(\mathbf{R}^{\alpha\bullet}_l) \to 0
\end{aligned}
$$

where

$$
R = \sum_{\alpha=1}^{k}\sum_{i=1}^{l(\mu^\alpha)} x_i^\alpha, \qquad H^1_{\mathrm{et}}(\mathbf{R}^{\alpha\bullet}_l) \cong H^0(\mathcal{D}^\alpha_l, L^\alpha_l)^{\oplus n^\alpha_l} / H^0(\mathcal{D}^\alpha_l, L^\alpha_l),
$$

$$
H^0_{\mathrm{et}}(\mathbf{R}^{\alpha\bullet}_l) \cong \bigoplus_{q\in f^{-1}(\mathcal{D}^\alpha_l)} T_q\big(f^{-1}(\Delta(\mathcal{D}^\alpha)_l)\big)\otimes T_q^*\big(f^{-1}(\Delta(\mathcal{D}^\alpha)_l)\big) \cong \mathbb{C}^{\oplus n^\alpha_l},
$$

and n^α_l is the number of nodes over $\mathcal{D}^\alpha_l$. Please refer to the paper by C.-C. Liu, K.Liu, and J.Zhou[30] for detailed notations.

3. A new approach to the Gromov-Witten theory

In this section, we will illustrate a new approach to derive a recursion relation for any natural cohomology class on $\overline{\mathcal{M}}_{g,n}(X,\beta)$. Precisely, let ω be a cohomology class on $\overline{\mathcal{M}}_{g,n}(X,\beta)$ that can be lifted to a equivariant class ω_T on $\overline{\mathcal{M}}_g(X\times\mathbb{P}^1; X\times\{\infty\}\mid\beta;\mu)$. We obtain a recursion relation for the Gromov-Witten invariants involving ω by the following two steps:

1) **Localization on the relative stable moduli:** We have a natural projection map

$$
p : \overline{\mathcal{M}}_g(X\times\mathbb{P}^1; X\times\{\infty\}\mid\beta;\mu) \longrightarrow \overline{\mathcal{M}}_g(\mathbb{P}^1;\{\infty\}\mid|\mu|;\mu)
$$

along with the branching morphism,[9]

$$
\mathrm{Br} : \overline{\mathcal{M}}_g(\mathbb{P}^1;\{\infty\}\mid|\mu|;\mu) \longrightarrow \mathrm{Sym}^r(\mathbb{P}^1) \cong \mathbb{P}^r
$$

Combining these morphisms, we obtain information for the Gromov-Witten invariants involving ω_T through:

$$F(u) = \int_{[\overline{\mathcal{M}}_g(X\times\mathbb{P}^1;X\times\{\infty\}\ \mid\ \beta;\mu)]^{vir}} \omega_T \left(\mathrm{Br}\circ p\right)^*\Big(\prod_{k\in B}(H-k)\Big)$$

Here, B is any subset of $\{0,1,2,\cdots,r\}$ and H is the hyperplane class of $H^*(\mathbb{P}^r)$ such that $H|_{p_k} = k$ for each fixed point p_k of $\mathbb{P}^r$ under the natural S^1-action. $F(u)$ is, in general, a polynomial of the equivariant parameter u. On the other hand, the virual functorial localization formula (1) applied on the relative stable moduli $\overline{\mathcal{M}}_g(X\times\mathbb{P}^1;X\times\{\infty\}\mid\beta;\mu)$ expresses $F(u)$ as the sum over all fixed locus of S^1-action. Precisely, we have the following expression:

$$F(u) = \sum_{l\in B^c}\Big(\prod_{k\in B}(l-k)\Big)\cdot\Gamma_l(u) \tag{1}$$

where $\Gamma_l(u)$ is the contribution from the fixed locus that are mapped to the fixed point p_l under $\mathrm{Br}\circ p$. These contributions have the following form:

$$\Gamma_l(u) = \sum\Big[\int_{\overline{\mathcal{M}}_{g,m}(X,\bar\beta)} i^*(\omega)\,\psi_1^{k_1}\cdots\psi_m^{k_m}\Big]\cdot\Big[\text{known data } C(\Gamma_l^k)\Big] \tag{2}$$

where $i^*(\omega)$ is the restriction of ω to the components of fixed locus. The coefficients $C(\Gamma_l^k)$ depends on the partition μ and the splitting-type of the fixed locus which is governed by the following Cut-and-Join operation:[44]

- **Cut-operation :** Geometrically this coreesponds to the pinching of the domain curve along a non-trivial cycle. In terms of localization computation, this corresponds to the Cut-operation on the partition μ:

$$\mu=(\cdots,\mu_i,\cdots)\longrightarrow\nu=(\cdots,p,q,\cdots)\quad ,p+q=\mu_i$$

- **Join-operation :** Geometrically this corresponds to the bubbling of the domain curve by pinching a cycle enscribing two marked points. In terms of localization computation, this corresponds to the Join-operation on the partition μ:

$$\mu=(\cdots,\mu_i,\mu_j,\cdots)\longrightarrow\eta=(\cdots,\mu_i+\mu_j,\cdots)$$

Moreover the fixed locus that are mapped to a fixed point p_l is precisely those curves that are obtained by performing the Cut-and-Join operation on the fixed locus that are mapped to the fixed point p_{l+1}. For example, there is a unique curve $\mathcal{C}_r$ of genus g and n-marked points that is mapped to p_r. And the fixed locus that are mapped to the branching point p_{r-1} are precisely

- † (Cut-of-type-I) A curve that is obtained by pinching a meridian of $\mathcal{C}_r$. This curve will have arithmetic genus $g-1$ and one more special point coming from the pinching.
- † (Cut-of-type-II) A curve that is obtained by pinching a longitude of $\mathcal{C}_r$. This curve will consist of two smooth components with genus g_1,g_2 such that $g_1 + g_2 = g$.
- † (Join) A curve that is obtained by pinching a cycle that enscribes two marked points. This curve will consist of two smooth components, one of which has genus g with one less marked points.

As the result, RHS of the relation (1) can be explicitly computed and consists of the Gromov-Witten invariants involving ω. This relation contains enough information to compute all Gromov-Witten invariants. However, we can extract more precise relations from (1) by using the asymptotic analysis as described below.

2) **Asymptotic Analysis:** The relation (1) holds for any given partition μ of any size. Hence it is natural to expect that, if we choose arbitraty μ, we should be able to extract relations on the Gromov-Witten invariants that are independent of the partition μ, i.e. relations between absolute Gromov-Witten invariants. This idea is realized by letting the size of μ to be arbitrarily large $|\mu| \to \infty$. Precisely we consider the following scaling limit of the partition μ:

$$\text{Write } \mu_i = N \cdot x_i \quad \text{where } N \in \mathbb{Z}, x_i \in \mathbb{Q} \quad \text{and let } N \to \infty$$

In the localization computation, we encounter the following type of combinatorial numbers that depend on the partition μ:

$$\prod_{i=1}^{l(\mu)} \frac{\mu_i^{\mu_i+k_i}}{\mu_i!}$$

Under the Cut-operation, this number will be replaced by the corresponding combinatorial number for ν. Especially the effect of the

Cut-operation is reflected through

$$\frac{\mu_i^{\mu_i+k_i}}{\mu_i!} \longrightarrow \sum_{p+q=\mu_i} \frac{p^{p+a}q^{q+b}}{p!q!} \tag{3}$$

where a and b depend on k_i and the splitting-type of fixed locus. The asymptotic behaviour of this combinatorial number is given by the following asymptotic formulas.[22]

• **Asymptotic Formula** : Let $a, b \in \mathbb{N}$ and $k \geq 0$. As $N \to \infty$, we have the following asymptotic behaviours

$$e^{-N} \sum_{p+q=N} \frac{p^{p+a}q^{q+b}}{p!q!} \longrightarrow \frac{1}{2}\left[\frac{(2a-1)!!(2b-1)!!}{2^{a+b}(a+b)!}\right]N^{a+b} + o(N^{a+b})$$

$$e^{-N} \sum_{p+q=N} \frac{p^{p+k+1}q^{q-1}}{p!q!} \longrightarrow \frac{n^{k+\frac{1}{2}}}{\sqrt{2\pi}} - \left[\frac{(2k+1)!!}{2^{k+1}k!}\right]N^{k} + o(N^{k}) \tag{4}$$

These asymptotic formulas are obtained through an application of the integration by parts and the Stirling's formula

$$N! \sim \sqrt{2\pi}e^{-N}N^{N+\frac{1}{2}}(1 + \frac{1}{12N} + \cdots).$$

This allows us to derive the limiting equation of the recursion relation under the scaling limit $N \longrightarrow \infty$. Moreover, the asymptotic behaviour does not depend on the specific partition-type of μ. Hence this allows us to extract relations between absolute Gromov-Witten invariants. Precisely, under the scaling limit $N \to \infty$, we obtain a stratification of the relation (1) with respect to the degree of N. This stratification gives us a system of recursion relations between absolute Gromov-Witten invariants.

4. Localization proof of the Witten conjecture

As an application of the new approach described in the previous section, we summarize the new proof of Witten's conjecture using localization method. Please refer to the paper by Y.-S. Kim and K. Liu[22] for details. The famous Witten conjecture[40] claims that stable intersection theory on moduli space is equivalent to the "hermitian matrix model" of two-dimensional gravity.

Precisely, E. Witten considered the generating function of the stable intersection theory on moduli space

$$F(t_0, t_1, \cdots) = \sum_{\{n_i\}} \prod_{i=0}^{\infty} \frac{t_i^{n_i}}{n_i!} \langle \tau_0^{n_0} \tau_1^{n_1} \tau_2^{n_2} \cdots \rangle.$$

and formulated the *Witten's conjecture* as follows: The generating function $F(t_0, t_1, \cdots)$ is determined by the following two constraints

(1) The object $U = \partial^2 F / \partial t_0^2$ obeys the KdV equations.

$$\frac{\partial U}{\partial t_n} = \frac{\partial}{\partial t_0} R_{n+1}(U, \dot{U}, \ddot{U}, \cdots),$$

where $\dot{U} = \partial U / \partial t_0$, $\ddot{U} = \partial^2 U / \partial t_0^2$, etc., are the derivatives of U with respect to t_0, and $R_{n+1}(U, \dot{U}, \ddot{U}, \cdots)$ are certain polynomials in U and its t_0 derivatives that are well-known in the theory of the KdV equations (and can be determined by recursion relations that are explicitly given).

(2) In addition, F obeys the "string equation,"

$$\frac{\partial F}{\partial t_0} = \frac{t_0^2}{2} + \sum_{i=0}^{\infty} t_{i+1} \frac{\partial F}{\partial t_i}.$$

Now there exist several different approaches to this conjecture:

1. M. Kontsevich[23] gave the first proof by constructing the main identity which relates the stable intersection theory on $\overline{\mathcal{M}}_{g,n}$ to its proper combinatorial model. The string partition function $\tau(t)$:

$$\tau(t) = \exp \sum_{g=0}^{\infty} \langle \exp \sum_{n} t_n \mathcal{O}_n \rangle_g$$

admits an integral representation which involves the following integral over $N \times N$ Hermitian matrix Y of the form[2]

$$\tau(Z) = \rho(Z)^{-1} \int dY \cdot \exp \mathrm{Tr}\Big[-\frac{1}{2} Z Y^2 + \frac{i}{6} Y^3 \Big]$$

where Z is a second $N \times N$ Hermitian matrix, and $\rho(Z)$ is the one-loop integral

$$\rho(Z) = \int dY \cdot \exp \Big[-\frac{1}{2} \mathrm{Tr}\, Z Y^2 \Big]$$

2. A. Okounkov-R. Pandharipande[35] gave another approach through the enumeration of branched covering of $\mathbb{P}^1$ using the ELSV-formula:[6]

$$\frac{H_{g,\mu} \cdot |\mathrm{Aut}\ \mu|}{(2g - 2 + |\mu| + l(\mu))!} = \prod_{i=1}^{l(\mu)} \frac{\mu_i^{\mu_i}}{\mu_i!} \int_{\overline{\mathcal{M}}_{g,l(\mu)}} \frac{\Lambda_g^{\vee}(1)}{\prod(1 - \mu_i \psi_i)}$$

3. M. Mirzhakani[34] derived the Virasoro constraints by connecting the stable intersection theory on $\overline{\mathcal{M}}_{g,n}$ to the Weil-Petersen volume and by using the McShane identity on the Weil-Petersen volume.
4. M. Kazarian-S. Lando[20] obtained an algebro-geometric proof by using the ELSV-formula and the PDEs which govern the generating series of Hurwitz numbers to derive the KdV-equation.

There are a couple of equivalent formulations for the Witten conjecture, namely the Virasoro constraints and the recursion relation for the correlation functions of topological gravity.

- **The Virasoro constraints:** The KdV-hierarchy can be expressed as linear homogeneous differential equations for the τ-function[2]

$$L_n \cdot \tau = 0, \qquad (n \geq -1)$$

where L_n denote the differential operators

$$L_{-1} = -\frac{1}{2}\frac{\partial}{\partial \tilde{t}_0} + \sum_{k=1}^{\infty}(k+\frac{1}{2})\tilde{t}_k \frac{\partial}{\partial \tilde{t}_{k-1}} + \frac{1}{4}\tilde{t}_0^2$$

$$L_0 = -\frac{1}{2}\frac{\partial}{\partial \tilde{t}_1} + \sum_{k=0}^{\infty}(k+\frac{1}{2})\tilde{t}_k \frac{\partial}{\partial \tilde{t}_k} + \frac{1}{16}$$

$$L_n = -\frac{1}{2}\frac{\partial}{\partial \tilde{t}_{n-1}} + \sum_{k=0}^{\infty}(k+\frac{1}{2})\tilde{t}_k \frac{\partial}{\partial \tilde{t}_{k+n}} + \frac{1}{4}\sum_{i=1}^{n}\frac{\partial^2}{\partial \tilde{t}_{i-1}\partial \tilde{t}_{n-i}}$$

- **The recursion relation for the correlation functions of topological gravity:** R. Dijkgraaf, E. Verlinde, and H. Verlinde derived,[2,3,39] through physical arguments, the following recursion relation for the correlation functions of topological gravity and showed that it is equivalent to the Virasoro constraints.

$$\begin{aligned}\langle \tilde{\sigma}_n \prod_{k\in S} \tilde{\sigma}_k \rangle_g &= \sum_{k\in S}(2k+1)\langle \tilde{\sigma}_{n+k-1} \prod_{l\neq k} \tilde{\sigma}_l \rangle_g + \frac{1}{2}\sum_{a+b=n-2} \langle \tilde{\sigma}_a \tilde{\sigma}_b \prod_{l\in S} \tilde{\sigma}_l \rangle_{g-1} \\ &+ \frac{1}{2}\sum_{S=X\cup Y, a+b=n-2, g_1+g_2=g} \langle \tilde{\sigma}_a \prod_{k\in X} \tilde{\sigma}_k \rangle_{g_1} \langle \tilde{\sigma}_b \prod_{l\in Y} \tilde{\sigma}_l \rangle_{g_2}\end{aligned} \tag{1}$$

where $\tilde{\sigma}_n = [(2n+1)!!]\sigma_n = [(2n+1)!!]\psi^n$ and

$$\langle \tilde{\sigma}_{k_1} \cdots \tilde{\sigma}_{k_l} \rangle_g = \left[\prod_{i=1}^{l}(2k_i+1)!!\right] \int_{\overline{\mathcal{M}}_{g,l}} \psi_1^{k_1} \cdots \psi_l^{k_l}$$

The above recursion relation (1) has the same degeneration type as that of the Cut-and-Join relation. It is proved[22] to be the limiting equation of the Cut-and-Join relation obtained by applying localization technique on the relative stable moduli $\overline{\mathcal{M}}_g(\mathbb{P}^1,\mu)$. We summarize the proof below:

Let ω be the trivial class and $B = \{0,1,2,\cdots,r-2\}$. Then the class $\omega\cdot \mathrm{Br}^*\prod_{k\in B}(H-k)$ has strictly less degree than the virtual dimension of the relative stable moduli $\overline{\mathcal{M}}_g(\mathbb{P}^1;\{\infty\}\mid|\mu|,\mu)$. Hence the relation (1) becomes

$$0 = r!\Gamma_r + (r-1)!\Gamma_{r-1}$$

As was explained in the previous section, the fixed curves that are mapped to p_{r-1} are precisely the curves obtained by performing the Cut-and-Join operation to the unique curve $\mathcal{C}_r$. This gives the following Cut-and-Join relation:[21,29]

$$r\Gamma_r = \sum_{i=1}^{n}\Big[\sum_{j\neq i}\frac{\mu_i+\mu_j}{1+\delta^{\mu_i}_{\mu_j}}\Gamma^{ij}_J + \sum_{p=1}^{\mu_i-1}\frac{p(\mu_i-p)}{1+\delta^{p}_{\mu_i-p}}\Big(\Gamma^{i,p}_{C1} + \sum_{g_1+g_2=g,\nu_1\cup\nu_2=\nu}\Gamma^{i,p}_{C2}\Big)\Big] \tag{2}$$

where $\Gamma_J, \Gamma_{C1}, \Gamma_{C2}$ denote the contributions from Join-curve, Cut-of-type-I, and Cut-of-type-II, respectively. Precisely they are defined as follows:

- The unique fixed curve that is mapped to the branching point p_r

$$\Gamma_r = \frac{1}{|\mathrm{Aut}\mu|}\prod_{i=1}^{n}\frac{\mu_i^{\mu_i}}{\mu_i!}\int_{\overline{\mathcal{M}}_{g,n}}\frac{\Lambda^\vee_g(1)}{\prod(1-\mu_i\psi_i)}$$

- Join curve that is obtained by joining i-th and j-th marked points:

$$\Gamma^{ij}_J = \frac{1}{|\mathrm{Aut}\eta|}\prod_{k=1}^{n-1}\frac{\eta_k^{\eta_k}}{\eta_k!}\int_{\overline{\mathcal{M}}_{g,n-1}}\frac{\Lambda^\vee_g(1)}{\prod(1-\eta_k\psi_k)}, \qquad \eta\in J_{ij}(\mu)$$

- Cut curve that is obtained by pinching around the i-th marked point:

$$\Gamma^{i}_{C1} = \frac{1}{|\mathrm{Aut}\nu|}\prod_{k=1}^{n+1}\frac{\nu_k^{\nu_k}}{\nu_k!}\int_{\overline{\mathcal{M}}_{g-1,n+1}}\frac{\Lambda^\vee_{g-1}(1)}{\prod(1-\nu_k\psi_k)}, \qquad \nu\in C_i(\mu)$$

- Cut curve that is obtained by splitting around the i-th marked point:

$$\Gamma^{i}_{C2} = \Big[\prod_{k=1}^{n+1}\frac{\nu_k^{\nu_k}}{\nu_k!}\Big]\prod_{s=1,2}\frac{1}{|\mathrm{Aut}\;\nu_s|}\int_{\overline{\mathcal{M}}_{g_s,n_s}}\frac{\Lambda^\vee_{g_s}(1)}{\prod(1-\nu_{s,k}\psi_k)}, \qquad \nu\in C_i(\mu)$$

where $\Lambda_g^\vee(u) = u^g - \lambda_1 u^{g-1} + \cdots + (-1)^g \lambda_g$ is the total Chern-class of the dual Hodge bundle. Applying the scaling limit $N \to \infty$ gives a stratification for $\Gamma_r, \Gamma_J, \Gamma_{C1}$, and Γ_{C2} with respect to N, i.e. we have expansions of the form

$$\Big[\prod_{i=1}^n \frac{\mu_i^{\mu_i}}{\mu_i!}\Big] \int_{\overline{\mathcal{M}}_{g,n}} \frac{\Lambda_g^\vee(1)}{\prod(1-\mu_i\psi_i)} = \sum_{(k_i)} \Big[\prod_{i=1}^n \frac{\mu_i^{\mu_i+k_i}}{\mu_i!}\Big] \int_{\overline{\mathcal{M}}_{g,n}} \prod \psi_i^{k_i} + o(N^d)$$

where d is the highest N-degree in the expression and $(k_i) = (k_1, \cdots, k_n)$ runs over the sequences of non-negative integers such that $\sum_{i=1}^n k_i = 3g - 3 + n = \dim_{\mathbb{C}} \overline{\mathcal{M}}_{g,n}$. Note that the top-degree terms consist of Hodge integrals of only ψ-classes since the total Chern-class of dual Hodge-bundle $\Lambda_g^\vee(1) = 1 - \lambda_1 + \cdots \pm \lambda_g$ do not involve the scaling parameter N. By applying the asymptotic formulas (4), we obtain a system of relations between linear Hodge integrals on $\overline{\mathcal{M}}_{g,n}$ from (2). The highest N-degree relation turns out to be a trivial one:

$$(\sum_{i=1}^n x_i) \prod \frac{x_i^{k_i - 1/2}}{\sqrt{2\pi}} \int_{\overline{\mathcal{M}}_{g,n}} \prod \psi_i^{k_i} - (\sum_{i=1}^n x_i) \prod \frac{x_i^{k_i - 1/2}}{\sqrt{2\pi}} \int_{\overline{\mathcal{M}}_{g,n}} \prod \psi_i^{k_i} = 0$$

The second-highest N-degree relation is the following:

$$\begin{aligned} 0 = \sum_{i=1}^n \Big[& \frac{(2k_i+1)!!}{2^{k_i+1}k_i!} x_i^{k_i} \prod_{j\neq i} \frac{x_j^{k_j-1/2}}{\sqrt{2\pi}} \int_{\overline{\mathcal{M}}_{g,n}} \prod \psi_j^{k_j} \quad - \\ & \sum_{k+l=k_i-2} \frac{(2k+1)!!(2l+1)!!}{2^{k_i+1}k_i!} x_i^{k_i} \prod_{j\neq i} \frac{x_j^{k_j-1/2}}{\sqrt{2\pi}} \Big(\int_{\overline{\mathcal{M}}_{g-1,n+1}} \psi_1^k \psi_2^l \prod \psi_j^{k_j} \\ & + \sum_{g_1+g_2=g, I\cup J=\{2,\cdots,n\}} \int_{\overline{\mathcal{M}}_{g_1,|I|+1}} \psi_1^k \prod_{j\in I} \psi_j^{k_j} \int_{\overline{\mathcal{M}}_{g_2,|J|+1}} \psi_1^l \prod_{j\in J} \psi_j^{k_j} \Big) \\ & - \sum_{j\neq i} \frac{(x_i+x_j)^{k_i+k_j-1/2}}{\sqrt{2\pi}} \prod_{l\neq i,j} \frac{x_l^{k_l-1/2}}{\sqrt{2\pi}} \int_{\overline{\mathcal{M}}_{g,n-1}} \psi^{k_i+k_j-1} \prod \psi_l^{k_l} \Big] \end{aligned} \tag{3}$$

This relation is identical to the recursion relation for the correlation functions of topological gravity (1) which can be seen as follows: Introduce formal variables $s_i \in \mathbb{R}_{>0}$ and recall the Laplace Transformation:

$$\int_0^\infty \frac{x^{k-1/2}}{\sqrt{2\pi}} e^{-x/2s} dx = (2k-1)!!\ s^{k+1/2}, \qquad \int_0^\infty x^k e^{-x/2s} dx = k!\ (2s)^{k+1}$$

After taking the Laplace transformation of (3), we recover the recursion relation for the correlation functions of topological gravity (1)

$$\begin{aligned}\langle\tilde{\sigma}_n\prod_{k\in S}\tilde{\sigma}_k\rangle_g &= \sum_{k\in S}(2k+1)\langle\tilde{\sigma}_{n+k-1}\prod_{l\neq k}\tilde{\sigma}_l\rangle_g+\frac{1}{2}\sum_{a+b=n-2}\langle\tilde{\sigma}_a\tilde{\sigma}_b\prod_{l\in S}\tilde{\sigma}_l\rangle_{g-1}\\ &+\frac{1}{2}\sum_{S=X\cup Y,a+b=n-2,g_1+g_2=g}\langle\tilde{\sigma}_a\prod_{k\in X}\tilde{\sigma}_k\rangle_{g_1}\langle\tilde{\sigma}_b\prod_{l\in Y}\tilde{\sigma}_l\rangle_{g_2}\end{aligned}$$

Since this recursion relation is equivalent to the Virasoro constraints and the Witten's conjecture, this finishes the proof of Witten's conjecture through localization technique. As a remark, the system of relations given by the stratification of (2) may give more identities between linear Hodge-integrals. For example, the third-highest N-degree relation verifies the following expression of λ_1-class in terms of κ_1-class and ψ-classes

$$12\,\lambda_1=\kappa_1+\delta-\sum\psi_i.$$

5. Proof of Mariño-Vafa formula

In this section, we summarize the survey note of the second author[31] about the recent proof of Mariño-Vafa formula.[29] Based on the string duality between open topological string theory on the deformed conifold T^*S^3 and the closed topological string theory on the resolved conifold, M. Mariño and C. Vafa[33] conjectured a closed formula about the generating series of the triple Hodge integrals for all genera and any number of marked points in terms of the Chern-Simons invariants, or equivalently in terms of the representations and combinatorics of symmetric groups. The precise statement is as follows:

The Mariño-Vafa conjecture is an identity between the geometry of the moduli spaces of stable curves and Chern-Simons knot invariants, or the combinatorics of the representation theory of symmetric groups. Let us first introduce the geometric side. For every partition $\mu=(\mu_1\geq\cdots\mu_{l(\mu)}\geq 1)$, we define the triple Hodge integral to be,

$$\mathcal{G}_{g,\mu}(\tau)=A(\tau)\cdot\int_{\overline{\mathcal{M}}_{g,l(\mu)}}\frac{\Lambda_g^\vee(1)\Lambda_g^\vee(-\tau-1)\Lambda_g^\vee(\tau)}{\prod_{i=1}^{l(\mu)}(1-\mu_i\psi_i)},$$

where the coefficient is given by

$$A(\tau) = -\frac{\sqrt{-1}^{|\mu|+l(\mu)}}{|Aut(\mu)|}[\tau(\tau+1)]^{l(\mu)-1}\prod_{i=1}^{l(\mu)}\frac{\prod_{a=1}^{\mu_i-1}(\mu_i\tau+a)}{(\mu_i-1)!}.$$

These expressions arise naturally from localization computations on the moduli spaces of relative stable maps into $\mathbf{P}^1$ with prescribed ramification type μ at ∞. We now introduce the generating series

$$\mathcal{G}_\mu(\lambda;\tau) = \sum_{g\geq 0}\lambda^{2g-2+l(\mu)}\mathcal{G}_{g,\mu}(\tau).$$

Introduce formal variables $p = (p_1, p_2, \ldots, p_n, \ldots)$, and define

$$p_\mu = p_{\mu_1}\cdots p_{\mu_{l(\mu)}}$$

for any partition μ. These p_{μ_j} correspond to $\operatorname{Tr} V^{\mu_j}$ in the notations of string theorists. The generating series for all genera and all possible marked points are defined to be

$$\mathcal{G}(\lambda;\tau;p)^\bullet = \exp\Big(\sum_{|\mu|\geq 1}\mathcal{G}_\mu(\lambda;\tau)p_\mu\Big),$$

which encode the complete information of the triple Hodge integrals.

Next we introduce the representation theoretical side. Let χ_μ denote the character of the irreducible representation of the symmetric group $S_{|\mu|}$, indexed by μ where $|\mu| = \sum_j \mu_j$. Let $C(\mu)$ denote the conjugacy class of $S_{|\mu|}$ indexed by μ. Introduce

$$\mathcal{W}_\mu(\lambda) = \prod_{1\leq a<b\leq l(\mu)}\frac{\sin[(\mu_a-\mu_b+b-a)\lambda/2]}{\sin[(b-a)\lambda/2]\prod_{i=1}^{l(\nu)}\prod_{v=1}^{\mu_i}2\sin[(v-i+l(\mu))\lambda/2]}.$$

This has an interpretation in terms of quantum dimension in Chern-Simons knot theory. We define the following generating series

$$\mathcal{R}(\lambda;\tau;p)^\bullet = \sum_{|\mu|\geq 0}\left(\sum_{|\nu|=|\mu|}\frac{\chi_\nu(C(\mu))}{z_\mu}e^{\sqrt{-1}(\tau+\frac{1}{2})\kappa_\nu\lambda/2}\mathcal{V}_\nu(\lambda)\right)p_\mu$$

where μ^i are sub-partitions of μ, $z_\mu = \prod_j \mu_j! j^{\mu_j}$, and

$$\kappa_\mu = |\mu| + \sum_i(\mu_i^2 - 2i\mu_i)$$

for a partition μ which is also standard for representation theory of symmetric groups. There is the relation $z_\mu = |Aut(\mu)|\mu_1 \cdots \mu_{l(\mu)}$. Finally we can give the precise statement of *the Mariño-Vafa conjecture:*

$$\textbf{Mariño-Vafa conjecture:} \qquad \mathcal{G}(\lambda;\tau;p)^\bullet = \mathcal{R}(\lambda;\tau;p)^\bullet$$

This conjecture was first proved by C.C. Liu, K. Liu, and J. Zhou[29] by showing that both sides have the same initial data, i.e.;

$$\mathcal{G}(\lambda, 0, p)^\bullet = \exp\Big(\sum_{d=1}^{\infty} \frac{p_d}{2d \sin(\frac{\lambda d}{2})}\Big) = \mathcal{R}(\lambda, 0, p)^\bullet$$

and satisfy the following Cut-and-Join relation: for $\Omega = \mathcal{G}^\bullet$, $\mathcal{R}^\bullet$, we have

$$\frac{\partial \Omega}{\partial \tau} = \frac{\sqrt{-1}\lambda}{2} \sum_{i,j\geq 1} \left(ijp_{i+j} \frac{\partial^2 \Omega}{\partial p_i \partial p_j} + (i+j)p_i p_j \frac{\partial \Omega}{\partial p_{i+j}} \right).$$

Since this Cut-and-Join relation completely determines Ω for any given initial condition, we conclude the identity of $\mathcal{G}^\bullet$ and $\mathcal{R}^\bullet$ which is the Mariño-Vafa conjecture.

Now let us explain how the new approach we illustrated in this paper applies to this case. Let $\pi : \mathcal{U}_{g,\mu} \to \overline{\mathcal{M}}_{g,0}(\mathbb{P}^1, \mu)$ and $P : \mathcal{T}_{g,\mu} \to \overline{\mathcal{M}}_{g,0}(\mathbb{P}^1, \mu)$ be the universal domain curve and the universal target, respectively. There is an evaluation map $F : \mathcal{U}_{g,\mu} \to \mathcal{T}_{g,\mu}$ and a contraction map $\bar{\pi} : \mathcal{T}_{g,\mu} \to \mathbb{P}^1$. Let $\mathcal{D}_{g,\mu} \subset \mathcal{U}_{g,\mu}$ be the divisor corresponding to the $l(\mu)$ marked points. Define

$$V_D = R^1\pi_*(\mathcal{O}_{\mathcal{U}_{g,\mu}}(-\mathcal{D}_{g,\mu})) \qquad \text{and} \qquad V_{D_d} = R^1\pi_*\tilde{F}^*\mathcal{O}_{\mathbb{P}^1}(-1),$$

where $\tilde{F} = \bar{\pi} \circ F : \mathcal{U}_{g,\mu} \to \mathbb{P}^1$. The fibers of V_D and V_{D_d} at

$$\big[f : (C, x_1, \ldots, x_{l(\mu)}) \to \mathbb{P}^1[m]\,\big] \in \overline{\mathcal{M}}_{g,0}(\mathbb{P}^1, \mu)$$

are $H^1(C, \mathcal{O}_C(-D))$ and $H^1(C, \tilde{f}^*\mathcal{O}_{\mathbb{P}^1}(-1))$, respectively, where $D = x_1 + \ldots + x_{l(\mu)}$, and $\tilde{f} = \pi[m] \circ f$. Note that $H^0(C, \mathcal{O}_C(-D)) = H^0(C, \tilde{f}^*\mathcal{O}_{\mathbb{P}^1}(-1)) = 0$, so V_D and V_{D_d} are vector bundles of rank $l(\mu)+g-1$ and $d+g-1$, respectively. The obstruction bundle

$$V = V_D \oplus V_{D_d}$$

is a vector bundle of rank $r = 2g - 2 + d + l(\mu) = \operatorname{vdim} \overline{\mathcal{M}}_g(\mathbb{P}^1, \mu)$. We integrate the equivariant Euler class of V over the relative stable moduli $\overline{\mathcal{M}}_g(\mathbb{P}^1, \mu)$ to obtain

$$K_\mu^\bullet(\lambda) = \int_{[\overline{\mathcal{M}}_g(\mathbb{P}^1,\mu)]^{vir}} e_T(V)$$

where $K^\bullet_\mu(\lambda)$ is of zero u-degree which depends on μ and λ. On the other hand, the localization computation gives a relation of the form (2). In this case, the 'known data' in (2) turns out to be the double Hurwitz numbers. Precisely, we reach the following convolution formula between triple Hodge integrals and double Hurwitz numbers

$$K^\bullet_\mu(\lambda) = \sum_{|\nu|=|\mu|} \mathcal{G}^\bullet_\nu(\lambda, \tau, p) z_\nu \Phi^\bullet_{\nu,\mu}(-i\tau\lambda)$$

where $\Phi^\bullet(\lambda)$ is a generating series of double Hurwitz numbers. This convolution formula can be inverted to give the convolution expression of Hodge integrals[28]

$$\mathcal{G}^\bullet(\lambda, \tau, p) = \sum_{|\mu|\geq 0} z_\mu K^\bullet_\mu(\lambda) \Phi^\bullet_\mu(i\tau\lambda, p, 1)$$

It is a direct consequence, from this expression, that $\mathcal{G}^\bullet$ satisfies the Cut-and-Join relation since the generating series of double Hurwitz numbers $\Phi^\bullet$ also satisfies it, hence finishing the proof of the Mariño-Vafa formula. Let us end this section with several consequences[44] of the Mariño-Vafa formula obtained by comparing the coefficients of τ in the Taylor expansions of the two expressions $\mathcal{G}^\bullet$ and $\mathcal{R}^\bullet$: It gives a simple proof of the λ_g-conjecture

$$\int_{\overline{\mathcal{M}}_{g,n}} \psi_1^{k_1} \cdots \psi_n^{k_n} \lambda_g = \binom{2g+n-3}{k_1, \cdots, k_n} \frac{2^{2g-1}-1}{2^{2g-1}} \cdot \frac{|B_{2g}|}{(2g)!},$$

and the following identities for Hodge integrals

$$\int_{\overline{\mathcal{M}}_g} \lambda_{g-1}^3 = \int_{\overline{\mathcal{M}}_g} \lambda_{g-2}\lambda_{g-1}\lambda_g = \frac{1}{2(2g-2)!} \frac{|B_{2g-2}|}{2g-2} \frac{|B_{2g}|}{2g},$$

$$\int_{\overline{\mathcal{M}}_{g,1}} \frac{\lambda_{g-1}}{1-\psi_1} = b_g \sum_{i=1}^{2g-1} \frac{1}{i} - \frac{1}{2} \sum_{\substack{g_1+g_2=g \\ g_1,g_2>0}} \frac{(2g_1-1)!(2g_2-1)!}{(2g-1)!} b_{g_1} b_{g_2},$$

where B_{2g} are Bernoulli numbers, $b_0 = 1$ and $b_g = \frac{2^{2g-1}-1}{2^{2g-1}} \frac{|B_{2g}|}{(2g)!}$ for $g > 0$.

6. Future Research problems

In this section, we list several open problems to which our new approach can be applied. Each of them has an equivalent formulation in the form of recursion relations which has the same structure as that of the Cut-and-Join relation.

6.1. *Generalized Witten conjecture*

Consider a series of integrable hierarchies KdV_r, where $r = 2, 3, \cdots$, called the generalized KdV, or Gelfand-Dickey hierarchies. E.Witten generalized his original conjecture,[42,43] suggesting that for each r there should exist moduli space and cohomology classes on them whose intersection numbers assemble into the formal τ-function of the KdV_r-hierarchy. Recently its first proof appeared in the paper by C.Faber, S.Shadrin, and D.Zvonkine[8] which relies on the equivalence of the formal and the geometric Gromov-Witten potentials under certain conditions. The corresponding moduli spaces of higher spin curves are constructed and the zero-genus case of the conjecture has been proved.[17,18] A brief idea of the construction is as follows:[38] Let $a_1, \cdots, a_n \in \{0, \cdots, r-1\}$ be integers assigned to the marked points $x_1, \cdots, x_n$ such that $2g - 2 - \sum a_i$ is divisible by r. On a smooth curve C, there are r^{2g} different line bundles $\mathcal{T}$ with an identification

$$\mathcal{T}^{\otimes r} \simeq K(-\sum a_i x_i).$$

The space of smooth curves endowed with such a line bundle $\mathcal{T}$ is denoted by $\mathcal{M}^{1/r}_{g;a_1,\cdots,a_n}$. The compactified space, denoted by $\overline{\mathcal{M}}^{1/r}_{g;a_1,\cdots,a_n}$, is constructed[17] and is called the *moduli space of stable r-spin curves*. The construction uses the *Jarvis-Vistoli twisted curves*, i.e. curves that are themselves endowed with an orbifold structure. It is a smooth stack with a finite projection mapping

$$p : \overline{\mathcal{M}}^{1/r}_{g;a_1,\cdots,a_n} \longrightarrow \overline{\mathcal{M}}_{g,n}.$$

Its analogue of the Gromov-Witten classes is also constructed[17] and is called a *virtual class* $c^{1/r}_{g,n}(\mathrm{a})$ in $H^\bullet(\overline{\mathcal{M}}^{1/r}_{g;a_1,\cdots,a_n})$. In the physics notation, we write

$$\langle \sigma_{m_1,a_1} \cdots \sigma_{m_n,a_n} \rangle_g = \int_{\overline{\mathcal{M}}^{1/r}_{g;a_1,\cdots,a_n}} c^{1/r}_{g,n}(\mathrm{a}) \psi_1^{m_1} \cdots \psi_n^{m_n}.$$

There is a conjectural recursion relation[27] for these intersection numbers:

$$\begin{aligned}\frac{h+1}{h} \langle \sigma_{m,1} \prod_{l=1}^{s} \sigma_{n_l,\alpha_l} \rangle_g = & \sum_l \left(n_l + \frac{\alpha_l}{h} \right) \langle \sigma_{m+n_l-1,\alpha_l} \prod_{j \neq l} \sigma_{n_j,\alpha_j} \rangle_g \\ & + \frac{1}{2} \sum_{n=2}^{m} \sum_{\alpha \in I} \Big[\langle \sigma_{n-2,\alpha} \sigma_{m-n,h-\alpha} \prod_{l=1}^{s} \sigma_{n_l,\alpha_l} \rangle_{g-1} \\ & + \sum_{\substack{S = X \cup Y \\ g = g_1 + g_2}} \langle \sigma_{n-2,\alpha} \prod_{l \in X} \sigma_{n_l,\alpha_l} \rangle_{g_1} \langle \sigma_{m-n,h-\alpha} \prod_{l \in Y} \sigma_{n_l,\alpha_l} \rangle_{g_2} \Big]\end{aligned}$$

where h is the dual Coxeter number such that $\alpha \in I$ if and only if $h-\alpha \in I$. The above recursion formula implies the generalized Witten conjecture, i.e. the generating series of these integrals with the Witten virtual class $c_{g,n}^{1/r}(\mathbf{a})$ satisfies the KdV$_r$-hierarchy where we denote $\mathbf{a} = (a_1, \cdots, a_n)$. Note that the above recursion formula has the same structure as that of the Dijkgraaf-Verlinde-Verlinde recursion relation which implies the original Witten conjecture. Our approach can be applied to this problem as follows: We first use the functorial localization formula on the moduli space of stable r-spin curves $\overline{\mathcal{M}}_{g,\mathbf{a}}^{1/r}(\mu, \mathbb{P}^1)$ into $\mathbb{P}^1$. Combining the natural projection to the moduli space of relative stable maps $\overline{\mathcal{M}}_g(\mu, \mathbb{P}^1)$ and the branch morphism, we obtain an equivariant morphism

$$\overline{\mathcal{M}}_{g,\mathbf{a}}^{1/r}(\mu, \mathbf{P}^1) \longrightarrow \overline{\mathcal{M}}_g(\mu, \mathbb{P}^1) \longrightarrow \mathbb{P}^r$$

The asymptotic analysis technique applied to the resulting cut-and-join equation will yield a recursion formula generalizing the cut-and-join equation for Hodge integrals involving the Witten virtual class $c_{g,n}^{1/r}(\mathrm{a})$ and the ψ-classes, which, in turn, should agree with the above conjectural recursion formula for higher spin intersection numbers.

6.2. *Faber's conjecture on Hodge integrals*

C. Faber[7] obtained a set of conjectures concerning the tautological Chow ring $R^\bullet(\mathcal{M}_g)$. The following identity on Hodge integrals is one of them:

$$\begin{aligned}\frac{(2g-3+n)!}{2^{2g-1}(2g-1)!} \cdot \frac{1}{\prod_{j=1}^k (2e_j-1)!!} &= \langle \tau_{e_1} \cdots \tau_{e_k} \tau_{2g} \rangle \\ &- \sum_{j=1}^k \langle \tau_{e_1} \cdots \tau_{e_{j-1}} \tau_{e_j+2g-1} \tau_{e_j+1} \cdots \tau_{e_k} \rangle \\ &+ \frac{1}{2} \sum_{j=0}^{2g-2} (-1)^j \langle \tau_{2g-2-j} \tau_j \tau_{e_1} \cdots \tau_{e_k} \rangle \\ &+ \frac{1}{2} \sum_{\underline{k}=I \amalg J} \sum_{j=0}^{2g-2} (-1)^j \langle \tau_j \prod_{i \in I} \tau_{e_i} \rangle \langle \tau_{2g-2-j} \prod_{i \in J} \tau_{e_i} \rangle\end{aligned}$$

As a remark, this conjectural identity implies,[11] through an application of the degree 0 Virasoro conjecture for $\mathbb{P}^2$, the following $\lambda_g \lambda_{g-1}$-conjecture

$$\left[\prod_{i=1}^n (2k_i-1)!!\right] \int_{\overline{\mathcal{M}}_{g,n}} \lambda_g \lambda_{g-1} \psi_1^{k_1} \cdots \psi_n^{k_n} = (2g-3+n)! \frac{|B_{2g}|}{2^{2g-1}(2g)!}$$

where B_{2g} denotes the Bernoulli number. A recent result of K.Liu and H.Xu[32] revealed that the constant term on the LHS is the third summation term on the RHS. This puts the above conjectural identity into a simpler equivalent recursion form

$$\langle \tau_{e_1} \cdots \tau_{e_k} \tau_{2g} \rangle = \sum_{j=1}^{k} \langle \tau_{e_1} \cdots \tau_{e_{j-1}} \tau_{e_j+2g-1} \tau_{e_j+1} \cdots \tau_{e_k} \rangle$$
$$- \frac{1}{2} \sum_{\underline{k}=I \amalg J} \sum_{j=0}^{2g-2} (-1)^j \langle \tau_j \prod_{i \in I} \tau_{e_i} \rangle \langle \tau_{2g-2-j} \prod_{i \in J} \tau_{e_i} \rangle$$

This recursion formula, in particular, is in the form of the Cut-and-Join equation where we perform the operation on the distinguished gravitational descendant term τ_{2g}. The fourth-highest N-degree relation given by the recursion relation obtained in the proof of Witten's conjecture strongly suggests this conjectural identity. The combinatorial techniques developed in a recent paper[14] may be used to simplify the technical difficulties arising from this approach. On the other hand, the two-partition Mariño-Vafa formula was proved through a cut-and-join equation for the involved two partitions, which has the same type of recursion formula. We can apply the asymptotic analysis to this two-partition equation to derive more generalized recursion formulas. Two-partition Mariño-Vafa formula was also proved by applying localization formula on moduli spaces of relative stable maps into a toric surface, which in principle indicates that the resulting formula should contain the same information as the Virasoro conjecture for surface. This is another approach to the above conjectural recursion formula.

6.3. *General Virasoro conjecture*

Let V be a non-singular projective variety. The *general Virasoro conjecture for V* asserts vanishing relations on the total Gromov-Witten potential

$$Z(V) = \exp\Big(\sum_{g \geq 0} \hbar^{g-1} \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k_1 \cdots k_n a_1 \cdots a_n} t_{k_n}^{a_n} \cdots t_{k_1}^{a_1} \langle \tau_{k_1}(\gamma_{a_1}) \cdots \tau_{k_n}(\gamma_{a_n}) \rangle_g^V \Big).$$

Precisely, there are differential operators L_k[4,5] which annihilates $Z(V)$, i.e.

$$L_k Z(V) = 0 \quad \text{for all } k \geq -1. \tag{1}$$

It is known that (1) is equivalent to the following recursion relation:[10]

$$
\begin{aligned}
0 = \sum_{i=0}^{k+1} \Big(\frac{1}{2} \sum_{m=i-k}^{-1} & (-1)^m \big[\mu_a + m + \frac{1}{2}\big]_i^k (R^i)^{ab} \big(\langle\langle \tau_{-m-1,a} \tau_{m+k-i,b} \rangle\rangle_{g-1}^V \\
& + \sum_{g_1+g_2=g} \langle\langle \tau_{-m-1,a} \rangle\rangle_{g_1}^V \langle\langle \tau_{m+k-i,b} \rangle\rangle_{g_2}^V \big) - \big[\frac{3-r}{2}\big]_i^k (R^i)_0^b \langle\langle \tau_{k-i+1,b} \rangle\rangle_g^V \\
& + \sum_{m=0}^{\infty} \big[\mu_a + m + \frac{1}{2}\big]_i^k (R^i)_a^b t_m^a \langle\langle \tau_{m+k-i,b} \rangle\rangle_g^V \Big) + \frac{\delta_{g,0}}{2} (R^{k+1})_{ab} t_0^a t_0^b \\
& + \delta_{k,0} \delta_{g,1} \rho(V)
\end{aligned}
$$

where R_a^b is the matrix associated to multiplication on the affine superspace $H(V)$ by the first Chern class $c_1(V)$ of V, defined by

$$R_a^b \gamma_b = c_1(V) \cup \gamma_a$$

and $\rho(V)$ is the characteristic number of V. This conjectural relation has the same structure as that of the Cut-and-Join relation except that the first Chern class of V is involved in it. The approach illustrated in this paper can be applied to this conjecture by considering a general relative stable moduli $\overline{\mathcal{M}}_g((d,\mu), V \times \mathbb{P}^1)$ relative to the divisor $V \times \{\infty\}$ and its natural projection map $\overline{\mathcal{M}}_g((d,\mu), V \times \mathbb{P}^1) \longrightarrow \overline{\mathcal{M}}_g(\mu, \mathbb{P}^1)$. Combining with the branch morphism, we get a equivariant morphism

$$\overline{\mathcal{M}}_g((d,\mu), V \times \mathbb{P}^1; V \times \infty) \longrightarrow \overline{\mathcal{M}}_g(\mu, \mathbb{P}^1) \longrightarrow \mathbb{P}^r$$

from which we get a general Cut-and-Join type formula involving the Chern classes of V. The asymptotic analysis method used in the proof of the Witten conjecture can be applied on this general cut-and-join formula. The resulting identity will be a quadratic recursion relation involving the first Chern class of V in agreement with the above conjectural identity. The main difficulty will be to understand the one-dimensional moduli space of relative stable maps to V with prescribed contact type at two divisors. Note that the special genus 0 case of the Virasoro conjecture has been previously proved. Givental and others[12] have announced the proof of the Virasoro conjecture for projective spaces and the Grassmannian manifolds. Our proposed method is quite different from the previous approaches, and it should prove the Virasoro conjecture for all genera and for general projective manifolds without any restriction.

References

1. M. Atiyah, R. Bott, *The moment map and equivariant cohomology*, Topology **23** (1984), no.1, 1-28.
2. R. Dijkgraaf, *Intersection Theory, Integrable Hierarchies and Topological Field Theory*, New symmetry principles in quantum field theory (Carge, 1991), 95–158, NATO Adv. Sci. Inst. Ser. B Phys., 295, Plenum, New York, 1992.
3. R. Dijkgraaf, E. Verlinde, H. Verlinde, *Notes on topological string theory and 2d quantum gravity*, in *String Theory and Quantum Gravity*, Proceedings of the Trieste Spring School 1990, M. Green *et al.*eds. (World-Scientific, 1991).
4. T. Eguchi, K. Hori, C.-S. Xiong, *Quantum cohomology and Virasoro algebra*, Phys. Lett. **B402** (1997), 71-80.
5. T. Eguchi, M. Jinzenji, C.-S. Xiong, *Quantum cohomology and free field representation*, Nucl. Phys. **B510** (1998), 608-622.
6. T. Ekedahl, S. Lando, M. Shapiro, A. Vainshtein, *Hurwitz numbers and intersections on moduli spaces of curves*, Invent. Math. **146** (2001), 297-327.
7. C. Faber, *A conjectural description of the tautological ring of the moduli space of curves*, preprint, math.AG/9711218
8. C. Faber, S. Shadrin, D. Zvonkine, *Tautological relations and the r-spin Witten conjecture*, preprint, math.AG/0612510
9. B. Fantecchi, R. Pandharipande, *Stable maps and branch divisors*, Compositio Math. **130** (2002), no.3, 345-364.
10. E. Getzler, *The Virasoro conjecture for Gromov-Witten invariants*, preprint, math.AG/9812026.
11. E. Getzler, R. Pandharipande, *Virasoro Constraints and the Chern classes of the Hodge bundle*, preprint, math.AG/9805114
12. A. Givental, *Gromov-Witten invariants and quantization of quadratic hamiltonians*, preprint, math.AG/0108100.
13. I.P. Goulden, D.M. Jackson, A. Vainshtein, *The number of ramified coverings of the sphere by the torus and surfaces of higher genera*, Ann. of Comb. 4 (2000), 27-46.
14. I.P. Goulden, D.M. Jackson, R. Vakil, *A short proof of λ_g conjecture without Gromov-Witten theory: Hurwitz theory and the moduli of curves*, preprint, math.AG/0604297
15. T. Graber, R. Pandharipande, *Localization of virtual classes*, Invent. Math. **135** (1999), no.2, 487-518
16. T. Graber, R. Vakil, *Relative virtual localization and vanishing of tautological classes on moduli spaces of curves*, preprint, math.AG/0309227.
17. T. Jarvis, T. Kimura, A. Vaintrob, *Moduli spaces of higher spin curves and Integrable hierarchies*, preprint, math.AG/9905034.
18. T. Jarvis, T. Kimura, A. Vaintrob, *Gravitational descendants and the moduli space of higher spin curves*, preprint, math.AG/0009066.
19. E. Katz, *Formalism for relative Gromov-Witten invariants*, preprint, math.AG/0507321.
20. M.E. Kazarian, S.K. Lando, *An algebro-geometric proof of Witten's conjecture*, MPIM-preprint, 2005-55.
21. Y.-S. Kim, *Computing Hodge integrals with one lambda-class*, Communica-

tions in Analysis and Geometry, Vol. 15, **no 3**.
22. Y.-S. Kim, K. Liu, *A simple proof of Witten conjecture through localization*, preprint, math.AG/0508384.
23. M. Kontsevich, *Intersection theory on the moduli space of curves and the matrix Airy function*, Comm. Math. Phys. **147** (1992), no. 1, 1–23.
24. J. Li, *Stable Morphisms to singular schemes and relative stable morphisms*, J. Diff. Geom. **57** (2001), 509-578.
25. J. Li, *Relative Gromov-Witten invariants and a degeneration formula of Gromov-Witten invariants*, J. Diff. Geom. **60** (2002), 199-293.
26. J. Li, *Lecture notes on GW-invariants*, preprint.
27. K. Li, *Recursion relations in topological gravity with minimal matter*, Nuclear Physics **B354** (1991), 725-739.
28. C.-C. Liu, *Formulae of One-Partition and Two-Partition Hodge Integrals*, preprint, math.AG/0502430.
29. C.-C. Liu, K. Liu, J. Zhou, *A proof of a conjecture of Mariño-Vafa on Hodge Integrals*, J. Differential Geom. **65** (2003), no. 2, 289–340.
30. C.-C. Liu, K. Liu, J. Zhou, *A formula of two-partition Hodge integrals*, preprint, math.AG/0310272.
31. K. Liu, *Localization and conjectures from string duality*, preprint. math.AG/0701057.
32. K. Liu, H. Xu, *Intersection numbers on $\overline{\mathcal{C}}_{g,n}$ and automorphisms of stable curves*, preprint, math.AG/0608209.
33. M. Mariño, C. Vafa, *Framed knots at large N*, Orbifolds in mathematics and physics (Madison, WI, 2001), 185–204, Contemp. Math., 310, Amer. Math. Soc., Providence, RI, 2002.
34. M. Mirzakhani, *Simple geodesics and Weil-Petersson volumes of moduli spaces of bordered Riemann surfaces*, preprint, 2003.
35. A. Okounkov, R. Pandharipande, *Gromov-Witten theory, Hurwitz numbers, and Matrix models, I*, preprint, math.AG/0101147.
36. A. Polishchuk, A. Vaintrob, *Algebraic construction of Witten's top Chern class*, Advances in algebraic geometry motivated by physics (Lowell, MA, 2000), pp. 229-249,
37. A. Polishchuk, *Witten's top Chern class on the moduli space of higher spin curves*, In "Frobenius manifolds", pp. 253-264, Aspects Math., vol. **E36**, Vieweg, Wiesbaden, 2004.
38. S. Shadrin, D. Zvonkine, *Intersection numbers with Witten's top Chern class*, preprint, math.AG/0601075.
39. E. Verlinde, H. Verlinde, *A solution of two-dimensional topological quantum gravity*, Nucl. Phys. **B348** (1990) 523.
40. E. Witten, *Two-dimensional gravity and intersection theory on moduli space*, Surveys in differential geometry (Cambridge, MA, 1990), 243–310, Lehigh Univ., Bethlehem, PA, 1991.
41. E. Witten, *On the Kontsevich model and other models of two-dimensional gravity*, Proceedings of the XXth International Conference on Differential Geometric Methods in Theoretical Physics, Vol. 1, 2 (New York, 1991), 176–216, World Sci. Publishing, River Edge, NJ, 1992. 32G15 (14H15 58F07 81T40).

42. E. Witten, *Algebraic geometry associated with matrix models of two dimensional gravity*, Topological models in modern mathematics (Stony Brook, NY, 1991), Publish or Perish, Houston, TX (1993), 23569. 35.
43. E. Witten, *The N-matrix model and gauged WZW models*, Nucl. Phys. B371 (1992), no. 1, 19145.
44. J. Zhou, *Hodge integrals, Hurwitz numbers, and symmetric groups*, preprint, math.AG/0308024.

Link Groups and the A-B Slice Problem

Vyacheslav S. Krushkal

Department of Mathematics, University of Virginia,
Charlottesville, VA 22904-4137
Current address: Kavli Institute for Theoretical Physics
Santa Barbara, CA 93109
E-mail: krushkal@virginia.edu

Dedicated to the memory of Xiao-Song Lin

The $A-B$ slice problem is a reformulation of the topological 4−dimensional surgery conjecture in terms of decompositions of the 4−ball and link homotopy. We show that *link groups*, a recently developed invariant of 4−manifolds, provide an obstruction for the class of *model decompositions*, introduced by M. Freedman and X.-S. Lin. This unifies and extends the previously known partial obstructions in the $A-B$ slice program. As a consequence, link groups satisfy Alexander duality when restricted to the class of model decompositions, but not for general submanifolds of the 4−ball.

Keywords: 4−dimensional surgery, $A-B$ slice problem, Alexander duality, link homotopy, link groups.

1. Introduction

The surgery conjecture, a core ingredient in the geometric classification theory of topological 4−manifolds, remains an open problem for a large class of fundamental groups. The results to date in the subject: the disk embedding conjecture, and its corollaries – surgery and s-cobordism theorems for *good* groups[1,5,6,11] – show similarities of classification of topological 4−manifolds with the theory in higher dimensions. On the other hand, it has been conjectured[2] that surgery fails for (non-abelian) free fundamental groups.

The $A-B$ slice problem[3] is a reformulation of the surgery conjecture for free groups which seems most promising in terms of the search for an obstruction. In this approach one considers smooth codimension zero decompositions $D^4 = A_i \cup B_i$ of the 4−ball, extending the standard genus one Heegaard decomposition of the 3-sphere. (A precise definition is given in section 2, also see figure 2.1.) Then the problem is formulated in terms of the existence of disjoint embeddings of the submanifolds A_i, B_i in D^4 with a prescribed homotopically essential link in $S^3 = \partial D^4$ as the boundary condition. The central case corresponds to the link equal to the Borromean rings. The problem may be phrased in terms of the existence of a suitably formulated non-abelian Alexander duality in dimension 4. Recently this approach has been sharpened and now there is a precise, axiomatic description of what properties an obstruction, which in this context is an invariant of decompositions of D^4, should satisfy.

The $A - B$ slice formulation of surgery was introduced by Freedman[3] and further extensively studied by Freedman-Lin.[4] In particular, the latter paper introduced a family of *model decompositions* which appear to approximate, in a certain algebraic sense, an arbitrary decomposition $D^4 = A \cup B$. This family of decompositions is defined in section 4. In this paper we use *link groups* of 4−manifolds, recently introduced by the author,[8] to formulate an obstruction for the family of model decompositions:

Theorem 1.1. *Let L be the Borromean rings, or more generally any homotopically essential link in S^3. Then L is not $A - B$ slice where each decomposition $D^4 = A_i \cup B_i$ is a model decomposition.*

The invariant using link groups formulated in the proof unifies and generalizes the previously known partial obstructions[4,9] in the $A - B$ slice program. The definitions of link groups and the underlying geometric notion of *Bing cells* are given in section 3.

To place this result in the geometric context of link homotopy, it is convenient to introduce the notion of a robust 4−manifold. Recall that a link L in S^3 is *homotopically trivial*[12] if its components bound disjoint maps of disks in D^4. L is called homotopically essential otherwise. (The Borromean rings is a homotopically essential link with trivial linking numbers.) Let (M, γ) be a pair (4−manifold, embedded curve in ∂M). The pair (M, γ) is *robust* if whenever several copies (M_i, γ_i) are properly disjointly embedded in (D^4, S^3), the link formed by the curves $\{\gamma_i\}$ in S^3 is homotopically trivial. The following statement is a consequence of the proof of theorem 1.1:

Corollary 1.1. *Let $D^4 = A \cup B$ be a model decomposition. Then precisely one of the two parts A, B is robust.*

It is interesting to note that there exist decompositions where *neither* of the two sides is robust.[10] The following question relates this notion to the $A - B$ slice problem: *given a decomposition* $D^4 = A \cup B$, *is one of the given embeddings* $A \hookrightarrow D^4$, $B \hookrightarrow D^4$ *necessarily robust?* (The definition of a robust embedding $e\colon (M, \gamma) \hookrightarrow (D^4, S^3)$ is analogous to the definition of a robust pair above, with the additional requirement that each of the embeddings $(M_i, \gamma_i) \subset (D^4, S^3)$ is equivalent to e.)

In a certain sense, one is looking in the $A - B$ slice problem for an invariant of 4–manifolds which is more flexible than homotopy (so it satisfies a suitable version of Alexander duality), yet it should be more robust than homology – this is made precise using Milnor's theory of link homotopy. The subtlety of the problem is precisely in the interplay of these two requirements. Following this imprecise analogy, we show that link groups provide a step in construction of such a theory.

2. Surgery and the $A - B$ slice problem

The 4–dimensional topological surgery exact sequence (cf [FQ], Chapter 11), as well as the 5–dimensional topological s-cobordism theorem, are known to hold for a class of *good* fundamental groups. In the simply-connected case, this followed from Freedman's disk embedding theorem[1] allowing one to represent hyperbolic pairs in $\pi_2(M^4)$ by embedded spheres. Currently the class of good groups is known to include the groups of subexponential growth[6,11] and it is closed under extensions and direct limits. There is a specific conjecture for the failure of surgery for free groups:[2]

Conjecture 2.1. *There does not exist a topological* 4–*manifold* M, *homotopy equivalent to* $\vee^3 S^1$ *and with* ∂M *homeomorphic to* $\mathcal{S}^0(Wh(Bor))$, *the zero-framed surgery on the Whitehead double of the Borromean rings.*

In fact, this is one of a collection of canonical surgery problems with free fundamental groups, and solving them is equivalent to the unrestricted surgery theorem. The $A - B$ slice problem, introduced in ref. 3, is a reformulation of the surgery conjecture, and it may be roughly summarized as follows. Assuming on the contrary that the manifold M in the conjecture above exists, consider the compactification of the universal cover $\widetilde{M}$, which is homeomorphic to the 4–ball.[3] The group of covering transformations (the free group on three generators) acts on D^4 with a prescribed action on the boundary, and roughly speaking the $A - B$ slice problem is a program for finding an obstruction to the existence of such actions. Recall the definition of an $A - B$ slice link.[3,4]

Definition 2.1. A *decomposition* of D^4 is a pair of smooth compact codimension 0 submanifolds with boundary $A, B \subset D^4$, satisfying conditions (1) – (3) below. (Figure 2.1 gives a 2–dimensional example of a decomposition.) Denote

$$\partial^+ A = \partial A \cap \partial D^4, \quad \partial^+ B = \partial B \cap \partial D^4, \quad \partial A = \partial^+ A \cup \partial^- A, \quad \partial B = \partial^+ B \cup \partial^- B.$$

(1) $A \cup B = D^4$,
(2) $A \cap B = \partial^- A = \partial^- B$,
(3) $S^3 = \partial^+ A \cup \partial^+ B$ is the standard genus 1 Heegaard decomposition of S^3.

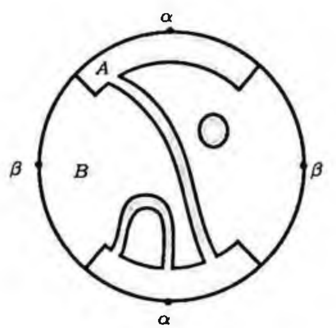


Fig. 2.1. A 2–dimensional analogue of a decomposition (A, α), (B, β): $D^2 = A \cup B$, A is shaded; (α, β) are linked 0–spheres in ∂D^2.

Definition 2.2. Given an n–component link $L = (l_1, \dots, l_n) \subset S^3$, let $D(L) = (l_1, l_1', \dots, l_n, l_n')$ denote the $2n$–component link obtained by adding an untwisted parallel copy L' to L. The link L is $A - B$ *slice* if there exist decompositions $(A_i, B_i), i = 1, \dots, n$ of D^4 and self-homeomorphisms α_i, β_i of D^4, $i = 1, \dots, n$ such that all sets in the collection $\alpha_1 A_1, \dots, \alpha_n A_n, \beta_1 B_1, \dots, \beta_n B_n$ are disjoint and satisfy the boundary data: $\alpha_i(\partial^+ A_i)$ is a tubular neighborhood of l_i and $\beta_i(\partial^+ B_i)$ is a tubular neighborhood of l_i', for each i.

The surgery conjecture holds for all groups if and only if the Borromean Rings (and the rest of the links in the canonical family of links) are $A - B$ slice.[3] Conjecture 2.1 above can therefore be reformulated as saying that the Borromean Rings are not $A - B$ slice.

As an elementary example, note that if a link L is $A-B$ slice where for each i the decomposition $D^4 = A_i \cup B_i$ consists of A_i = 2–handle $D^2 \times D^2$, and B_i = the collar on $\partial^+ B_i$, then L is actually *slice.*

Of course the Borromean Rings is not a slice (or homotopically trivial) link. However to show that a link is not $A-B$ slice, one needs to eliminate *all* choices for decompositions (A_i, B_i).

3. Link groups and Bing cells

In this section we recall the definition of Bing cells and link groups of 4–manifolds, denoted $\lambda(M^4)$, introduced in Ref. 8, in order to formulate the invariant I_λ used in the proof of theorem 1.1. The definition is inductive.

Definition 3.1. A *model Bing cell of height* 1 is a smooth 4-manifold C with boundary and with a specified attaching curve $\gamma \subset \partial C$, defined as follows. Consider a planar surface P with $k+1$ boundary components $\gamma, \alpha_1, \ldots, \alpha_k$ $(k \geq 0)$, and set $\overline{P} = P \times D^2$. Let $L_1, \ldots, L_k$ be a collection of links, $L_i \subset \alpha_i \times D^2$, $i = 1, \ldots, k$. Here for each i, L_i is the (possibly iterated) Bing double of the core α_i. Then C is obtained from $\overline{P}$ by attaching zero-framed 2-handles along the components of $L_1 \cup \ldots \cup L_k$.

The surface S (and its thickening $\overline{S}$) will be referred to at the *body* of C, and the 2-handles are the *handles* of C.

A *model Bing cell C of height h* is obtained from a model Bing cell of height $h-1$ by replacing its handles with Bing cells of height one. The *body* of C consists of all (thickenings of) its surface stages, except for the handles.

Figures 3.1, 3.2 give an example of a Bing cell of height 1: a schematic picture and a precise description in terms of a Kirby diagram. Here P is a pair of pants, and each link L_i is the Bing double of the core of the solid torus $\alpha_i \times D^2$, $i = 1, 2$.

Remark 3.1. To avoid a technical discussion, the definition presented here involves only the links L which are Bing doubles. To reflect this difference, we reserve for these objects the term *Bing cells* rather than the more general *flexible cells* discussed in Ref. 8. The definition in Ref. 8 involves more general homotopically essential links, however just the Bing doubles suffice for the applications in this paper.

Bing cells in a 4–manifold M are defined as maps of model Bing cells in M, subject to certain crucial disjointness requirements. (In particular, this

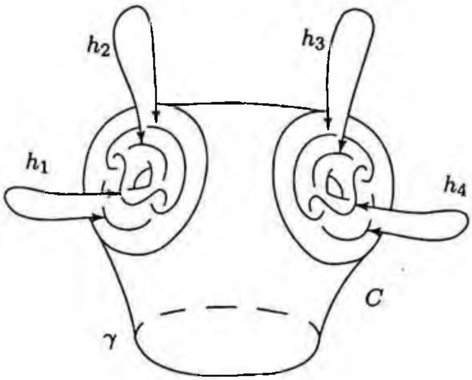


Fig. 3.1. Example of a model Bing cell of height 1: a schematic picture

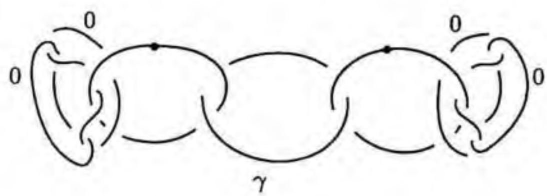


Fig. 3.2. A Kirby diagram of the model Bing cell in Figure 2.1

will be important for the discussion of model decompositions in section 4.) Roughly speaking, objects attached to different components of any given link L_i in the definition are required to be disjoint in M. To formulate this condition rigorously, recall the definition of the tree associated to a given Bing cell.

3.1. *The associated tree*

Given a Bing cell C, define the tree T_C inductively: suppose C has height 1. Then assign to the body surface P (say with $k+1$ boundary components) of C the cone T_P on $k+1$ points. Consider the vertex corresponding to the attaching circle γ of C as the root of T_P, and the other k vertices as the leaves of T_P. For each handle of C attach an edge to the corresponding leaf

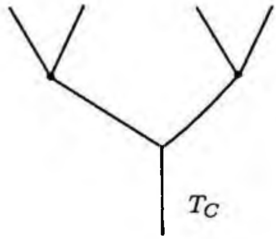


Fig. 3.3. The tree T_C associated to the Bing cell C in figures 3.1, 3.2.

of T_P. The leaves of the resulting tree T_C are in $1-1$ correspondence with the handles of C.

Suppose C has height $h > 1$, then it is obtained from a Bing cell C' of height $h-1$ by replacing the handles of C' with Bing cells $\{C_i\}$ of height 1. Assuming inductively that $T_{C'}$ is defined, one gets T_C by replacing the edges of $T_{C'}$ associated to the handles of C' with the trees corresponding to $\{C_i\}$. Figure 3.3 shows the tree associated to the Bing cell in figure 3.1.

Divide the vertices of T_C into two types: the vertices ("cone points") corresponding to body (planar) surfaces are *unmarked*; the rest of the vertices are *marked*. Therefore the valence of an unmarked vertex equals the number of boundary components of the corresponding planar surface. The marked vertices are in $1-1$ correspondence with the links L defining C, and the valence of a marked vertex is the number of components of L plus 1. It is convenient to consider the $1-$valent vertices of T_C: its root and leaves (corresponding to the handles of C) as unmarked. This terminology is useful in defining the maps of Bing cells below. The height of a Bing cell C may be read off from T_C as the maximal number of marked vertices along a geodesic joining a leaf of T_C to its root, where the maximum is taken over the leaves of T_C.

Definition 3.2. A *Bing cell* is a model Bing cell with a finite number of self-plumbings and plumbings among the handles and body surfaces of C, subject to the following disjointness requirement:

- Consider two surfaces A, B (they could be handles or body stages) of C. Let a, b be the corresponding vertices in T_C. (For body surfaces this is the corresponding unmarked cone point, for handles this is the associated leaf.) Consider the geodesic joining a, b in T_C, and look at its vertex c closest to the root of T_C – in other words, c is the first common ancestor of a, b. If c is a marked vertex then A, B are required to be disjoint.

In particular, self-plumbings of any handle and body surface are allowed. In the example shown in figures 3.1, 3.2 above, the handle h_1 is required to be disjoint from h_2, h_3 is disjoint from h_4; all other intersections are allowed.

A *Bing cell in a 4-manifold M* is an embedding of a Bing cell into M. We say that its image is a realization of C in M, and abusing the notation we denote its image in M also by C.

The main technical result of Ref. 8 shows how Bing cells fit in the context of Milnor's theory of link homotopy. This theorem is used in the analysis of the invariant I_λ below.

Theorem 3.1. *If the components of a link $L \subset S^3 = \partial D^4$ bound disjoint Bing cells in D^4 then L is homotopically trivial.*

Recall[12] that a link L in S^3 is *homotopically trivial* if L is homotopic to the unlink, so that different components stay disjoint during the homotopy. The theorem above builds on a classical result that if the components of L bound disjoint maps of *disks* in D^4 then L is homotopically trivial. The proof of theorem 3.1 is substantially more involved than the argument in the classical case. This is due to the topology of Bing cells which forces additional relations in the fundamental group of the complement. The main new technical ingredients in the proof are the *generalized Milnor group* and an obstruction which is well-defined in the presence of this additional indeterminacy.[8]

The *link groups* $\lambda_n(M)$ are defined as {based loops in a 4−manifold M} modulo loops bounding Bing cells of height n. These groups fit in a sequence of surjections

$$\pi_1(M) \longrightarrow \lambda_1(M) \longrightarrow \lambda_2(M) \longrightarrow \dots$$

The groups $\lambda_n(M)$ are topological but not in general homotopy invariants of M. In particular, they are not correlated with the first homology $H_1(M)$, or more generally with the quotients of $\pi_1(M)$ by the terms of its lower central or derived series. Define $\lambda(M)$ to be the direct limit of $\lambda_n(M)$. Given a pair (M, γ) where M is a 4−manifold and γ is a specified curve in ∂M, consider the invariant $I_\lambda(M, \gamma) \in \{0, 1\}$:

$$I_\lambda(M, \gamma) = 1 \;\text{ if }\; \gamma = 1 \in \lambda(M),$$

set $I_\lambda(M, \gamma) = 0$ otherwise. When the choice of the attaching circle γ of M is clear, we will abbreviate the notation to $I_\lambda(M)$.

Remark 3.2. For the interested reader we point out the "geometric duality" between Bing cells and gropes. Recall the definition:[5] A *grope* is a special pair (2-complex, circle). A grope has a *class* $k = 1, 2, \ldots, \infty$. For $k = 2$ a grope is a compact oriented surface Σ with a single boundary component. For $k > 2$ a *k-grope* is defined inductively as follow: Let $\{\alpha_i, \beta_i, i = 1, \ldots, \text{genus}\}$ be a standard symplectic basis of circles for Σ. For any positive integers p_i, q_i with $p_i + q_i \geq k$ and $p_{i_0} + q_{i_0} = k$ for at least one index i_0, a k-grope is formed by gluing p_i-gropes to each α_i and q_i-gropes to each β_i. A grope has a standard, "untwisted" 4−dimensional thickening, obtained by embedding it into $\mathbb{R}^3$, times I.

Consider a more general collection of 2-complexes, where at each stage one is allowed to attach several parallel copies of surfaces. Then one checks using Kirby calculus that model Bing cells are precisely complements in D^4 of standard embeddings of such generalized gropes. This observation is helpful in the analysis of the $A - B$ slice problem, where gropes play an important role, see section 4.

4. An obstruction for model decompositions.

In this section we show that the invariant I_λ defined above provides an obstruction for the family of model decompositions. We start the proof of theorem 1.1 by constructing the relevant decompositions of D^4. The simplest decomposition $D^4 = A \cup B$ where A is the 2−handle $D^2 \times D^2$ and B is just the collar on its attaching curve, was discussed in the introduction. Now consider the genus one surface S with a single boundary component α, and set $A_1 = S \times D^2$. Moreover, one has to specify its embedding into D^4 to determine the complementary side, B. Consider the standard embedding (take an embedding of the surface in S^3, push it into the 4−ball and take a regular neighborhood.) Note that given any decomposition, by Alexander duality the attaching curve of exactly one of the two sides vanishes in it homologically, at least rationally. Therefore the decomposition $D^4 = A_1 \cup B_1$ may be viewed as the first level of an "algebraic approximation" to an arbitrary decomposition. The general *model decomposition of height* 1 is analogous to the decomposition $D^4 = A_1 \cup B_1$, except that the surface S may have a higher genus.

Prop 4.1. *Let $A_1 = S \times D^2$, where S is the genus one surface with a single boundary component α. Consider the standard embedding $(A_1, \alpha \times \{0\}) \subset (D^4, S^3)$. Then the complement B_1 is obtained from the collar on its attaching curve, $S^1 \times D^2 \times I$, by attaching a pair of zero-framed 2−handles* *to the Bing double of the core of the solid torus $S^1 \times D^2 \times \{1\}$, figures 4.1, 4.2.*

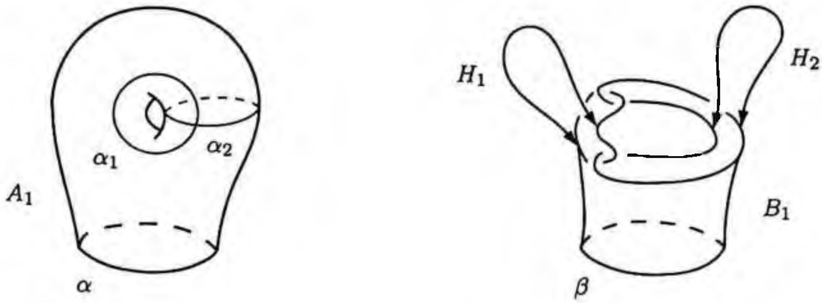


Fig. 4.1. A model decomposition $D^4 = A_1 \cup B_1$ of height 1: a schematic (spine) picture (figure 5) and a precise description in terms of Kirby diagrams, figure 4.6.

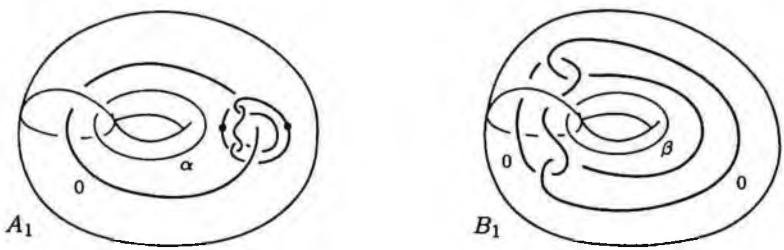


Fig. 4.2.

The proof is a standard exercise in Kirby calculus, see for example Ref. 4. A precise description of these 4–manifolds is given in terms of Kirby diagrams in figure 4.2. Rather than considering handle diagrams in the 3–sphere, it is convenient to draw them in the solid torus, so the 4–manifolds are obtained from $S^1 \times D^2 \times I$ by attaching the 1– and 2–handles as shown in the diagrams. To make sense of the "zero framing" of curves which are not null-homologous in the solid torus, recall that the solid torus is embedded into $S^3 = \partial D^4$ as the attaching region of a 4–manifold, and the 2–handle framings are defined using this embedding.

This example illustrates the general principle that (in all examples considered in this paper) the 1–handles of each side are in one-to-one correspondence with the 2–handles of the complement. This is true since the embeddings in D^4 considered here are all standard, and in particular each 2–handle is unknotted in D^4. The statement follows from the fact that 1–handles may be viewed as standard 2–handles removed from a collar, a standard technique in Kirby calculus (see Chapter 1 in Ref. 7.) Moreover, in each of our examples the attaching curve α on the A–side bounds a surface in A, so it has a zero framed 2–handle attached to the core of the solid torus. On the 3–manifold level, the zero surgery on this core transforms the solid torus corresponding to A into the solid torus corresponding to B. The Kirby diagram for B is obtained by taking the diagram for A, performing the surgery as above, and replacing all zeroes with dots, and conversely all dots with zeroes. (Note that the 2–handles in all our examples are zero-framed.)

Note that a distinguished pair of curves α_1, α_2, forming a symplectic basis in the surface S, is determined as the meridians (linking circles) to the cores of the 2–handles H_1, H_2 of B_1 in D^4. In other words, α_1, α_2 are fibers of the circle normal bundles over the cores of H_1, H_2 in D^4.

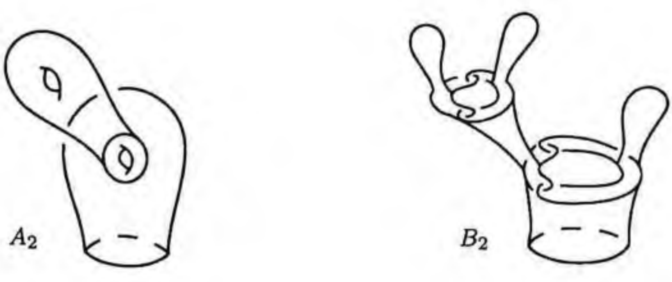


Fig. 4.3. A model decomposition $D^4 = A_2 \cup B_2$ of height 2.

An important observation[4] is that this construction may be iterated: consider the 2–handle H_1 in place of the original 4–ball. The pair of curves (α_1, the attaching circle β_1 of H_1) form the Hopf link in the boundary of H_1. As discussed in the beginning of this section, it is natural to consider two possibilities: either α_1 or β_1 bounds a surface in H_1. For simplicity of exposition, we again assume at this point that this is a surface of genus one. The first possibility (α_1 bounds) is shown in figure 4.3: note that in this decomposition one side, A_2, is a grope of height 3 (discussed in remark 3.2) and its complement B_2 is an example of a Bing cell.

Consider the second possibility: β_1 bounds a surface in H_1. As discussed above, its complement in H_1 is given by two zero-framed 2–handles attached to the Bing double of α_1. Assembling this data, consider the new decomposition $D^4 = A_2' \cup B_2'$, figures 4.4, 4.5. As above, the diagrams are drawn in solid tori (complements in S^3 of unknotted circles drawn dashed in the figures.) The decompositions $D^4 = A_2 \cup B_2$, $D^4 = A_2' \cup B_2'$ are examples of model decompositions of height 2. To get a general decomposition of this type, one also considers the alternative as above for the pair of curves α_2, β_2 in the 4−ball H_2. For simplicity of illustration, in the examples shown in figures 4.3 - 4.5 the curve β_2 bounds a surface of genus zero. One gets models of an arbitrary height by an iterated application of the construction above, and in general one considers (orientable) surfaces of an arbitrary genus at each stage. See figure 4.6 for examples of model decompositions of height 3.

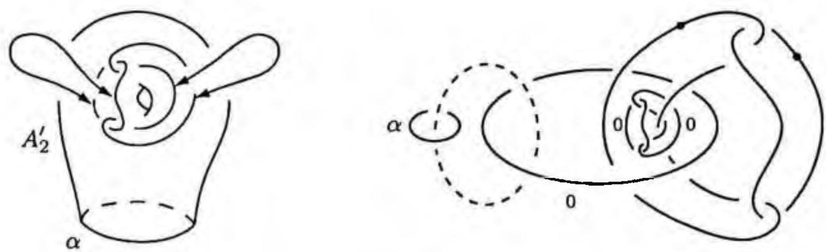


Fig. 4.4.

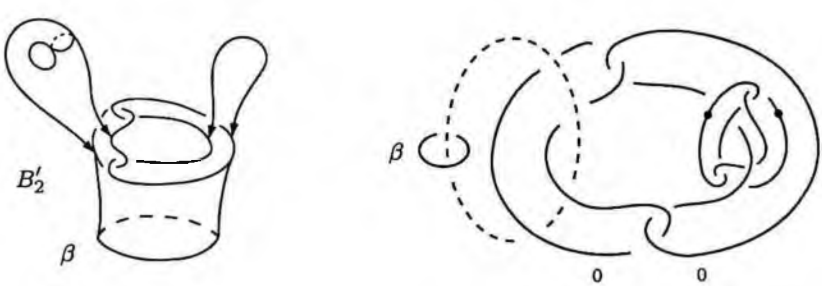


Fig. 4.5. Another example of a model decomposition $D^4 = A_2' \cup B_2'$ of height 2.

It follows from theorem 3.1 that the following lemma implies our main result, theorem 1.1:

Lemma 4.1. *Let $D^4 = A \cup B$ be a model decomposition. Then*

$$I_\lambda(A, \alpha) + I_\lambda(B, \beta) = 1.$$

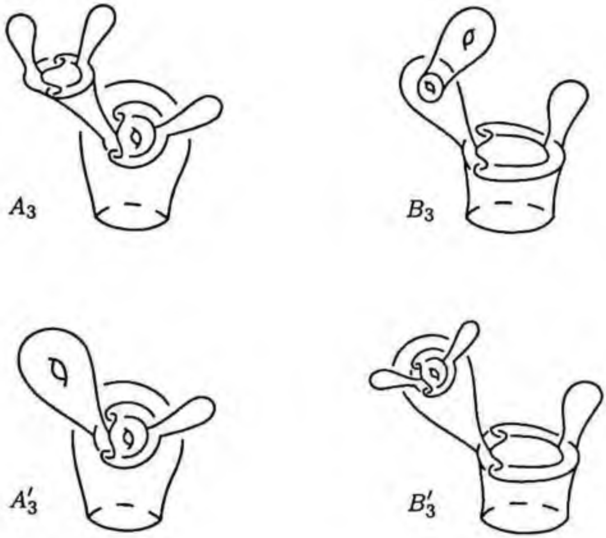


Fig. 4.6. Examples of model decompositions $D^4 = A_3 \cup B_3$, $D^4 = A'_3 \cup B'_3$ of height 3.

Indeed, suppose a link $L = (l_1, \dots, l_n)$ is $A - B$ slice where each decomposition $D^4 = A_i \cup B_i$, $i = 1, \dots, n$ is a model decomposition. According to lemma 4.1, the invariant I_λ of precisely one part of the decomposition equals 1. For each i, denote $C_i = A_i$ if $I_\lambda(A_i) = 1$ and $C_i = B_i$ otherwise. Let γ_i denote the attaching curve of C_i. It follows from the definition of I_λ that γ_i bounds a Bing cell in C_i. Since the collections $\{\alpha_i\}, \{\beta_i\}$ form the link L and its parallel copy, the collection of curves $(\gamma_1, \dots, \gamma_n)$ is isotopic to L. This contradicts theorem 3.1 since L is homotopically essential. This concludes the proof of theorem 1.1, assuming lemma 4.1.

Proof of lemma 4.1. It suffices to prove that given a model decomposition $D^4 = A \cup B$, either $\alpha = 1 \in \lambda(A)$ or $\beta = 1 \in \lambda(B)$. Then theorem 3.1 implies that precisely one of these two possibilities holds. The proof of the statement above is inductive. Given a model decomposition of height 1 (figure 4.1), observe that one of the two parts of the decomposition - the handlebody B_1 in the example in figure 4.1 - is a model Bing cell of height 1. (In this case the planar surface C in definition 3.1 is the annulus.) Therefore $\beta = 1 \in \lambda(B_1)$.

In the case that A_1 is a surface of genus $g > 1$, the handlebody description of B_1 consists of first taking g parallel copies of the core curve of the solid torus, Bing doubling them and then attaching zero-framed 2–handles to the resulting link. One observes that the attaching curve β still bounds a model Bing cell of height 1 in this handlebody, indeed there are g choices of Bing cells bounded by β.

Suppose lemma is proved for model decompositions of height $\leq n$, and let $D^4 = A \cup B$ be a model decomposition of height $n+1$. The attaching curve of either A or B is trivial in its first homology group. To be specific, assume $\alpha = 0 \in H_1(A;\mathbb{Z})$. First assume the surface Σ bounded by α has genus 1. Then A is obtained by attaching models A', A'' of height $\leq n$ to a symplectic basis of curves α_1, α_2 of Σ, figure 4.7. Similarly, using the notation of figure 4.1, B is obtained from the model B_1 of height 1 by replacing its 2–handles H_1, H_2 by two models B', B'' of height $\leq n$. Here $D^4 = A' \cup B', D^4 = A'' \cup B''$ are two decompositions for which lemma holds according to the inductive assumption. Therefore $I_\lambda(A') + I_\lambda(B') = I_\lambda(A'') + I_\lambda(B'') = 1$. Consider two cases:

Case 1: $I_\lambda(B') = I_\lambda(B'') = 1$

Case 2: At least one of $I_\lambda(A')$, $I_\lambda(A'')$ equals 1.

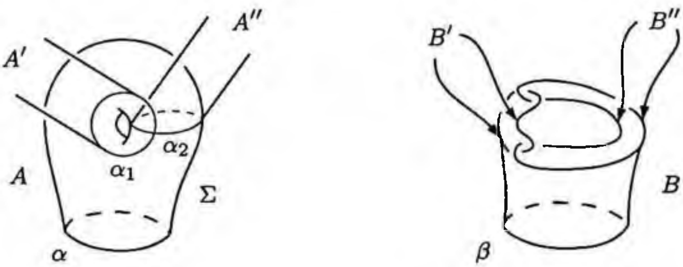


Fig. 4.7. Proof of lemma 4.1: the inductive step.

We claim that in the first case $I_\lambda(B) = 1$ and in the second case $I_\lambda(A) = 1$. Consider case 1. By assumption, the attaching curve β' of B' bounds a Bing cell C' in B', and similarly the attaching curve β'' bounds a Bing cell C'' in B''. Consider the handlebody C obtained from $S^1 \times D^2 \times I$ by attaching C', C'' to the Bing double of the core of the solid torus. The associated tree T_C is illustrated on the left in figure 4.8. (Note that the trees $T_{C'}$, $T_{C''}$ join in a *marked* vertex.) Since B' and B'' are disjoint, there are no $C' - C''$ intersections. (Note that such intersections are not allowed in the definition 3.2 of a Bing cell.) Therefore the attaching curve β bounds a Bing cell in B, and $I_\lambda(B,\beta) = 1$.

Fig. 4.8.

Consider the second case. Without loss of generality assume $I_\lambda(A') = 1$, so α_1 bounds a Bing cell C' in A'. Surger the first stage surface Σ along α_1, the result is a pair of pants whose boundary consists of α and two copies of α_1. Consider two copies of C' (denote them by C' and $\overline{C}'$) and perturb them so there are only finitely many intersections between surfaces in C' and surfaces in $\overline{C}'$. Consider the handlebody C assembled from the (pair of pants)$\times D^2$ with $C', \overline{C}'$ attached to it. The tree T_C associated to C is shown on the right in figure 4.8; observe that the trees $T_{C'}$, $T_{\overline{C}'}$ join in an *unmarked* vertex. Note that all $C' - \overline{C}'$ intersections are of the type allowed in definition 3.2, therefore α bounds a Bing cell in A, and $I_\lambda(A,\alpha) = 1$.

In the case when the surface Σ has genus $g > 1$ the proof is analogous to the genus one case discussed above. Specifically, A is obtained by attaching models A'_i, A''_i, $i = 1,\ldots,g$ to a symplectic basis of curves in Σ. The complements are denoted B'_i, B''_i. One observes that if there exists $1 \le i \le g$ such that $I_\lambda(B'_i) = I_\lambda(B''_i) = 1$, then $I_\lambda(B) = 1$. On the other hand, if for each i either $I_\lambda(A'_i)$ or $I_\lambda(A''_i)$ equals 1, then $I_\lambda(A) = 1$. This concludes the proof of lemma 4.1 and of theorem 1.1. □

Remark 4.1. In the example of the decomposition $D^4 = A'_2 \cup B'_2$ in figures 4.4, 4.5 the proof above shows that $I_\lambda(A'_2,\alpha) = 1$. One may find an explicit construction of a Bing cell bounded by α in A'_2 in the proof of [9, Lemma 7.3].

Acknowledgement

This research was supported in part by the NSF under grants DMS-0306934 and NSF PHY05-51164.

References

1. M. Freedman, *The topology of four-dimensional manifolds*, J. Differential Geom. 17(1982), 357-453.
2. M. Freedman, *The disk theorem for four-dimensional manifolds*, Proc. ICM Warsaw (1983), 647-663.
3. M. Freedman, *A geometric reformulation of four dimensional surgery*, Topology Appl., 24 (1986), 133-141.
4. M. Freedman and X.S. Lin, *On the* (A, B)*-slice problem*, Topology Vol. 28 (1989), 91-110.
5. M. Freedman and F. Quinn, *The topology of 4-manifolds*, Princeton Math. Series 39, Princeton, NJ, 1990.
6. M. Freedman and P. Teichner, 4-*Manifold Topology I*, Invent. Math. 122 (1995), 509-537.
7. R. Kirby, *Topology of* 4*−manifolds*, Lecture Notes in Mathematics 1374, Spriger-Verlag, 1989.
8. V. Krushkal, *Link groups of* 4*−manifolds*, arXiv:math.GT/0510507.
9. V. Krushkal, *On the relative slice problem and* 4*-dimensional topological surgery*, Math. Ann. 315 (1999), 363-396.
10. V. Krushkal, *A counterexample to the strong version of Freedman's conjecture*, Ann. Math., to appear; arXiv:math/0610865.
11. V. Krushkal and F. Quinn, *Subexponential groups in* 4*-manifold topology*, Geom. Topol. 4 (2000), 407-430.
12. J. Milnor, *Link Groups*, Ann. Math 59 (1954), 177-195.

Twisted L^2–Alexander–Conway Invariants for Knots

Weiping Li

Department of Mathematics
Oklahoma State University
Stillwater, OK 74078-1058, USA
E-mail: wli@math.okstate.edu

Weiping Zhang*

Chern Institute of Mathematics and LPMC
Nankai University
Tianjin 300071, China
E-mail: weiping@nankai.edu.cn

Dedicated to the memory of Professor Xiao-Song Lin

In this paper, we construct a knot invariant with $SL(2,\mathbb{C})$ representations by using the fundamental L^2–representation of the fundamental group of a knot complement, which may be thought of as an twisted L^2-Alexander(–Conway) invariant of the knot in S^3. Both L^2–Alexander and L^2-Alexander-Conway invariants define functions from the representation space of the knot group to nonnegative real numbers, and descend to functions from the $SL(2,\mathbb{C})$ character variety of the knot group. We also show that the twisted L^2-Alexander invariant is a L^2-Reidemeister torsion twisted by the $SL(2,\mathbb{C})$ representation. The L^2-Alexander (and L^2-Alexander-Conway) invariant twisted by $GL(n,\mathbb{C})$ representations of the knot group is given.

1. Introduction

This is a sequel that continues the study of the L^2-Alexander(-Conway) invariant defined by the authors in,[LZ,LZ2] where we gave a $\mathbb{C}^*$-parameterized L^2-invariant of knots from abelian representation tensoring with infinite dimensional representations of the knot group. As we mentioned earlier, such L^2-invariants with parameters and their twisted versions would play a role in the study of knots. It is natural to develop the L^2-Alexander-Conway invariant of knots with parameters through finite dimensional (non-abelian)

*Partially supported by MOEC and the 973 project.

representation tensoring with infinite dimensional representations of the knot group, especially through $SL(2,\mathbb{C})$-representations of hyperbolic knot groups.

In,[LZ,LZ2] we mainly concentrated on the most natural one which is associated with the universal covering of the knot complement and the $\mathbf{C}^*$ representations of the first homology group of the knot complement. It extends the one constructed by Lück[Lu2] who considers the case where the above $\mathbf{C}^*$ representation is the trivial representation of the first homology group of the knot complement. We call such an invariant the L^2-Alexander(-Conway) invariant.

Gukov[Gu] proposed a complex $SL(2,\mathbb{C})$ version of Witten's $SU(2)$ topological quantum field theory, and generalized the volume conjecture to $\mathbb{C}^*$-parameterized version with parameter lying on the zero locus of the A-polynomial from the $SL(2,\mathbb{C})$ character variety of the knot in S^3. In,[GuMu] Gukov and Murakami showed that the difference of their conjectures (in[Gu] and[Mu] respectively) comes from the different choices of polarization (different choices of the $SL(2,\mathbb{C})$ representations of the knot group). The work of the first author and Wang[LWa] shows that by focusing only on the regulator we can have a different generalized volume conjecture from that of Gukov ([Gu]) from the motivic point of view. In both cases, the character variety of the hyperbolic knot plays an essential role. In this paper, we construct the twisted L^2-invariant with parameter through a tensor of an infinite dimensional representation with a $SL(2,\mathbb{C})$ representations (a character). This extension fits naturally into the study of volume conjecture from the character variety point of view. In particular, we show that our twisted L^2-Alexander invariant is a twisted L^2-Reidemeister torsion of the knot complement. The essential ingredient for twisted $SL(2,\mathbb{C})$ L^2-invariants is the invertibility problem we solved in section 3. Hence it is natural to give the complete L^2-invariant twisted by $GL(n,\mathbb{C})$ representations.

This paper is organized as follows. In Section 2, recall the basic properties of the Fuglede-Kadison determinant for morphisms between free Hilbert modules over a group von Neumann algebra. In Section 3, for simplicity we construct the twisted L^2-Alexander(-Conway) invariant, with $SL(2,\mathbb{C})$-parameters, through the Wirtinger presentations of a knot. In Section 4, we interpret the L^2-Alexander(-Conway) invariant constructed in Section 3 through the L^2-Reidemeister torsion of the knot complement with a twisted $SL(2,\mathbb{C})$ flat bundle. In Section 5, we extend the L^2-Alexander invariant from the $SL(2,\mathbb{C})$ representation to its character. In Section 6, we show how to define the twisted L^2-Alexander(-Conway) invariant with any general $GL(n,\mathbb{C})$-representation parameters.

Acknowledgments Part of the work was done while the first author was visiting Chern Institute of Mathematics, Nankai Univeristy, he would like to thank Chern Institute for hospitality. The authors would like to thank Xiao-Song Lin for his interest in the earlier stage of this research. And his untimely death left us with a profound loss.

2. Group von Neumann algebra and Fuglede-Kadison determinant

In this section, we recall the definition and basic properties of the Fuglede-Kadison determinant, which will be used in the next section in our definition of the L^2-Alexander invariant. The basic reference is the comprehensive book of Lück.[Lu1]

Let Γ be a finitely generated discrete group. We assume that Γ is an infinite group.

Let $l^2(\Gamma)$ be the standard Hilbert space of squared summable formal sums over Γ with complex coefficients. Then every element in $l^2(\Gamma)$ can be written as

$$a=\sum_{\gamma\in\Gamma} a_\gamma\gamma,\quad a_\gamma\in\mathbf{C},\quad \text{with}\quad \sum_{\gamma\in\Gamma}|a_\gamma|^2<+\infty.$$

If $a=\sum_{\gamma\in\Gamma}a_\gamma\gamma$, $b=\sum_{\gamma\in\Gamma}b_\gamma\gamma$ are two elements in $l^2(\Gamma)$, their inner product is given by

$$\langle a,b\rangle=\sum_{\gamma\in\Gamma}a_\gamma\bar{b}_\gamma. \tag{1}$$

The left multiplication defines a natural unitary action of Γ on $l^2(\Gamma)$.

The *group von Neumann algebra* $\mathcal{N}(\Gamma)$ is the algebra of Γ-equivariant bounded linear operators from $l^2(\Gamma)$ to $l^2(\Gamma)$. The *von Neumann trace* on $\mathcal{N}(\Gamma)$ is defined by

$$\mathrm{Tr}_\tau:\mathcal{N}(\Gamma)\to\mathbf{C},\quad f\mapsto\langle f(e),e\rangle, \tag{2}$$

where $e\in\Gamma\subset l^2(\Gamma)$ is the unit element.

The right multiplication of Γ induces a natural isometric Γ–action of $l^2(\Gamma)$ on $l^2(\Gamma)$. Hence for any $\gamma\in\Gamma$, one can consider $\Gamma\subset\mathcal{N}(\Gamma)$. Moreover, for any $\gamma\in\Gamma\subset\mathcal{N}(\Gamma)$,

$$\mathrm{Tr}_\tau[\gamma]=1\ \text{ if }\ \gamma=e;\qquad \mathrm{Tr}_\tau[\gamma]=0\ \text{ if }\ \gamma\neq e. \tag{3}$$

For any positive integer n, set

$$l^2(\Gamma)^{[n]} = \underbrace{l^2(\Gamma) \oplus \cdots \oplus l^2(\Gamma)}_{n}.$$

We call it a free $\mathcal{N}(\Gamma)$-Hilbert module of rank n. The action of Γ on $l^2(\Gamma)$ (through left multiplications) induces a canonical action of Γ on $l^2(\Gamma)^{[n]}$. A morphism between two free $\mathcal{N}(\Gamma)$-Hilbert modules is a Γ-equivariant bounded linear map between them.

Let $f : l^2(\Gamma)^{[n]} \to l^2(\Gamma)^{[n]}$ be such a morphism. Let e_i, $i = 1, \cdots, n$, be the unit element in the i-th copy of $l^2(\Gamma)$ in $l^2(\Gamma)^{[n]}$. Then we can extend the von Neumann trace in (2) to define

$$\mathrm{Tr}_\tau[f] = \sum_{i=1}^{n} \langle f(e_i), e_i \rangle. \tag{4}$$

The Fuglede-Kadison determinant $\mathrm{Det}_\tau(f)$ of f can be defined as follows:

(i) If f is invertible and f^* is the adjoint of f, then define (cf. [Lu1, Lemma 3.15 (2)])

$$\mathrm{Det}_\tau(f) = \exp\left(\frac{1}{2}\mathrm{Tr}_\tau\left[\log\left(f^* f\right)\right]\right); \tag{5}$$

(ii) If f is injective, then define (cf. [Lu1, Lemma 3.15 (4), (5)])

$$\mathrm{Det}_\tau(f) = \lim_{\varepsilon \to 0+} \sqrt{\mathrm{Det}_\tau\left(f^* f + \varepsilon\right)} = \sqrt{\mathrm{Det}_\tau\left(f^* f\right)}. \tag{6}$$

If f is injective and $\mathrm{Det}_\tau(f) \neq 0$, then we say that f is of determinant class.

(iii) If both $f, g : l^2(\Gamma)^{[n]} \to l^2(\Gamma)^{[n]}$ are injective, then $g \circ f$ is also injective. Moreover (cf. [Lu1, Theorem 3.14 (1)]),[a]

$$\mathrm{Det}_\tau\left(g \circ f\right) = \mathrm{Det}_\tau(f) \cdot \mathrm{Det}_\tau(g). \tag{7}$$

(iv) If $f_1 : l^2(\Gamma)^{[n]} \to l^2(\Gamma)^{[n]}$, $f_2 : l^2(\Gamma)^{[m]} \to l^2(\Gamma)^{[m]}$ and $f_3 : l^2(\Gamma)^{[m]} \to l^2(\Gamma)^{[n]}$ be three morphisms such that f_1 and f_2 are injective. Then $\begin{pmatrix} f_1 & 0 \\ f_3 & f_2 \end{pmatrix} : l^2(\Gamma)^{[n+m]} \to l^2(\Gamma)^{[n+m]}$ is also injective. Moreover

[a]Indeed, in [Lu1, Theorem 3.14 (1)], it requires that f has dense image. Note that in finite von Neumann algebras the properties of an bounded operator to be injective and to have dense image are equivalent (see[DIX]).

(cf. [Lu1, Theorem 3.14 (2)]),[b]

$$\mathrm{Det}_\tau \begin{pmatrix} f_1 & 0 \\ f_3 & f_2 \end{pmatrix} = \mathrm{Det}_\tau(f_1) \cdot \mathrm{Det}_\tau(f_2). \tag{8}$$

(v) Let $f : l^2(\Gamma)^{[n]} \to l^2(\Gamma)^{[n]}$ be an invertible morphism. Then there exists a C^1 path f_u, $u \in [0,1]$, of invertible morphisms such that $f_0 = f$, $f_1 = \mathrm{Id}$. One then has (cf. [CFM, Theorem 1.10])

$$\log\left(\mathrm{Det}_\tau\left(f\right)\right) = -\mathrm{Re}\left(\int_0^1 \mathrm{Tr}_\tau\left[f_u^{-1}\frac{df_u}{du}\right] du\right). \tag{9}$$

3. Twisted $SL(2,\mathbb{C})$ L^2-Alexander invariant of a knot

Let $K \subset S^3$ be a knot. Let $\Gamma = \pi_1(S^3 \setminus K)$ denote the knot group. Let

$$P(\Gamma) = \langle x_1, \cdots, x_k | r_1, \cdots, r_{k-1}\rangle \tag{1}$$

be a Wirtinger presentation of Γ.

Let $F_k = \langle x_1, \cdots, x_k \rangle$ denote the free group of rank k.

Let $\phi : F_k \to \Gamma$ denote the canonical surjective homomorphism associated to (3.9). Then it induces a ring homomorphism

$$\widetilde{\phi} : \mathbf{Z}[F_k] \to \mathbf{Z}[\Gamma]. \tag{2}$$

Let β be a $SL(2,\mathbb{C})$ representation of the knot group Γ

$$\beta : \Gamma \to SL(2,\mathbb{C}). \tag{3}$$

Then $\beta(x_1), \beta(x_2), \cdots, \beta(x_k) \in SL(2,\mathbb{C})$ satisfy

$$\beta(r_1(x_1, \cdots, x_k)) = Id_{2\times 2}, \cdots, \beta(r_{k-1}(x_1, \cdots, x_k)) = Id_{2\times 2}. \tag{4}$$

Let $GL(l^2(\Gamma))$ denote the set of invertible elements in $\mathcal{N}(\Gamma)$. Let

$$\rho_\Gamma : \Gamma \to GL(l^2(\Gamma)) \tag{5}$$

denote the fundamental representation of Γ, which is given by the right multiplication of the elements in Γ. We denote the associated ring homomorphism of the integral ring $\mathbf{Z}(\Gamma)$ to $\mathcal{N}(\Gamma)$ by

$$\widetilde{\rho}_\Gamma : \mathbf{Z}[\Gamma] \to \mathcal{N}(\Gamma). \tag{6}$$

Let $\rho_\Gamma \otimes \beta$ be the tensor product representation of ρ_Γ and β. Hence for any $\gamma \in \Gamma$ we have $\rho_\Gamma(\gamma) \otimes \beta(\gamma) : l^2(\Gamma) \otimes \mathbb{C}^2 \to l^2(\Gamma) \otimes \mathbb{C}^2$. Let $\beta(\gamma) : \mathbb{C}^2 \to$

[b] We here replace the condition that f_1 has dense image to that f_1 is injective, which is possible as (7) now holds for f injective. Compare with the proof in [Lu1, page 135].

$\mathbb{C}^2$ be a $SL(2,\mathbb{C})$ matrix and $\mathbb{C}^2$ be the vector space with an ordered basis $\{e_1, e_2\}$. Hence $\beta(\gamma)(e_1) = a(\gamma)e_1 + b(\gamma)e_2, \beta(\gamma) = c(\gamma)e_1 + d(\gamma)e_2$ has its matrix form $\beta(\gamma) = \begin{pmatrix} a(\gamma) & c(\gamma) \\ b(\gamma) & d(\gamma) \end{pmatrix} \in SL(2,\mathbb{C})$ with $a(\gamma)d(\gamma) - b(\gamma)c(\gamma) = 1$. Let us identify

$$l^2(\Gamma) \otimes \mathbb{C}^2 \cong l^2(\Gamma)^{\otimes 2} \cong l^2(\Gamma) \otimes e_1 \oplus l^2(\Gamma) \otimes e_2. \tag{7}$$

Now the tensor product representation has the following

$$\rho_\Gamma \otimes \beta(\gamma)(a \otimes e_1) = \rho_\Gamma(\gamma)(a) \otimes \beta(\gamma)(e_1) = a(\gamma)\rho_\Gamma(\gamma)(a) \otimes e_1 + b(\gamma)\rho_\Gamma(\gamma)(a) \otimes e_2;$$

$$\rho_\Gamma \otimes \beta(\gamma)(a \otimes e_2) = \rho_\Gamma(\gamma)(a) \otimes \beta(\gamma)(e_2) = c(\gamma)\rho_\Gamma(\gamma)(a) \otimes e_1 + d(\gamma)\rho_\Gamma(\gamma)(a) \otimes e_2.$$

Hence the tensor product representation can be identified with the morphism from $l^2(\Gamma)^{\otimes 2} = l^2(\Gamma) \otimes e_1 \oplus l^2(\Gamma) \otimes e_2$ to $l^2(\Gamma) \otimes e_1 \oplus l^2(\Gamma) \otimes e_2$ as the following.

$$\begin{aligned} \rho_\Gamma \otimes \beta(\gamma) &= \begin{pmatrix} a(\gamma)\rho_\Gamma(\gamma) & c(\gamma)\rho_\Gamma(\gamma) \\ b(\gamma)\rho_\Gamma(\gamma) & d(\gamma)\rho_\Gamma(\gamma) \end{pmatrix} \\ &= \begin{pmatrix} \rho_\Gamma(\gamma) & 0 \\ 0 & \rho_\Gamma(\gamma) \end{pmatrix} \cdot \begin{pmatrix} a(\gamma) & c(\gamma) \\ b(\gamma) & d(\gamma) \end{pmatrix} \\ &= (\rho_\Gamma(\gamma) Id_{2\times 2}) \cdot \beta(\gamma). \end{aligned} \tag{8}$$

Let $\tilde{\beta} : \mathbf{Z}[\Gamma] \to M_{2\times 2}(\mathbb{C})$ be the induced ring homomorphism from the $SL(2,\mathbb{C})$ representation, where $M_{2\times 2}(\mathbb{C})$ is the 2×2 matrices. The induced ring homomorphism of the integral group rings for the tensor product representation is given by

$$\widetilde{\rho_\Gamma \otimes \beta} : \mathbf{Z}[\Gamma] \to \mathcal{N}(\Gamma) \otimes M_{2\times 2}(\mathbb{C}). \tag{9}$$

Let the composition of the ring homomorphism in (3.10) with the tensor product of ρ_Γ and the homomorphism in (9) be denoted by

$$\Psi = (\widetilde{\rho}_\Gamma \otimes \widetilde{\beta}) \circ \widetilde{\phi} : \mathbf{Z}[F_k] \to \mathcal{N}(\Gamma) \otimes M_{2\times 2}(\mathbb{C}). \tag{10}$$

Consider the morphism

$$A_{\rho_\Gamma \otimes \beta} : \underbrace{l^2(\Gamma)^{\otimes 2} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}}_{k-1} \to \underbrace{l^2(\Gamma)^{\otimes 2} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}}_{k} \tag{11}$$

which when written as a $(k-1) \times k$-matrix, the (i,j)-component is given by

$$A_{\rho_\Gamma \otimes \beta, (i,j)} = \Psi\left(\frac{\partial r_i}{\partial x_j}\right) \in \mathcal{N}(\Gamma) \otimes M_{2\times 2}(\mathbb{C}), \tag{12}$$

where $\frac{\partial r_i}{\partial x_j}$ is the standard Fox derivative and $l^2(\Gamma)^{\otimes 2}$ is identified in (7).

We call $A_{\rho_\Gamma\otimes\beta}$ the $SL(2,\mathbb{C})$ twisted L^2-Alexander matrix of the presentation $P(\Gamma)$ associated to the fundamental representation ρ_Γ and the $SL(2,\mathbb{C})$ representation β.

For any $1 \le j \le k$, let

$$A^j_{\rho_\Gamma\otimes\beta} : \underbrace{l^2(\Gamma)^{\otimes 2} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}}_{k-1} \to \underbrace{l^2(\Gamma)^{\otimes 2} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}}_{k-1} \tag{13}$$

denote the morphism obtained from $A_{\rho_\Gamma\otimes\beta}$ by removing the j-th column from its $(k-1) \times k$ matrix form.

The following result can be thought of as an L^2-analogue of [W, Lemma 2 and Lemma 3] or an extension of [LZ, Lemma 3.1].

Lemma 3.1. *(i) For any $1 \le j \le k$, $\Psi(x_j - 1) \in \mathcal{N}(\Gamma)$ is injective and has dense image. (ii) If one of the $A^j_{\rho_\Gamma\otimes\beta}$'s, $1 \le j \le k$, is injective, then every $A^j_{\rho_\Gamma\otimes\beta}$, $1 \le j \le k$, is injective. Moreover, in this case, for any $1 \le j < j' \le k$, one has*

$$\mathrm{Det}_\tau\left(A^j_{\rho_\Gamma\otimes\beta}\right)\mathrm{Det}_\tau\left(\Psi(x_{j'} - 1)\right) = \mathrm{Det}_\tau\left(A^{j'}_{\rho_\Gamma\otimes\beta}\right)\mathrm{Det}_\tau\left(\Psi(x_j - 1)\right). \tag{14}$$

Proof. (i) From (3.10), (4), (8) and (10), one sees that

$$\Psi(x_j - 1) = (\rho_\Gamma(\phi(x_j))Id_{2\times 2}) \cdot \beta(\phi(x_j)) - Id. \tag{15}$$

Clearly, $\gamma_j = \phi(x_j) \in \Gamma$ is of infinite order and $\beta(\gamma_j) \in SL(2,\mathbb{C})$.

Assume $a \in l^2(\Gamma)$ and $z_1, z_2 \in \mathbb{C}$ satisfies $((\rho_\Gamma(\phi(x_j))Id_{2\times 2}) \cdot \beta(\phi(x_j)) - Id)(a \otimes (z_1e_1 + z_2e_2)) = 0$. Then a direct verification shows that $(\rho_\Gamma(\gamma_j)Id_{2\times 2} - \beta(\gamma_j)^{-1})\beta(\gamma_j)(a \otimes (z_1e_1 + z_2e_2)) = 0$. Now we have identify $\beta(\gamma_j)$ as a morphism in $l^2(\Gamma)^{\otimes 2}$ with trivial action on $l^2(\Gamma)$ factor. For any $B \in SL(2,\mathbb{C})$ with $B(a \otimes (z_1e_1 + z_2e_2) = a \otimes (B(z_1e_1 + z_2e_2)) = 0$, we have $z_1 = z_2 = 0$, thus $a \otimes (z_1e_1 + z_2e_2) = 0$.

For $\beta(\gamma_j)$, let A be a $SL(2,\mathbb{C})$ matrix such that $A\beta(x_j)^{-1}A^{-1} = \begin{pmatrix} y_j & b_j \\ 0 & y_j^{-1} \end{pmatrix}$. Then we have $A(\rho_\Gamma(\gamma_j)Id_{2\times 2} - \beta(\gamma_j)^{-1})A^{-1} = (\rho_\Gamma(\gamma_j)Id_{2\times 2} - A\beta(\gamma_j)^{-1}A^{-1})$ and $\ker(\rho_\Gamma(\gamma_j)Id_{2\times 2} - \beta(\gamma_j)^{-1}) = \ker(\rho_\Gamma(\gamma_j)Id_{2\times 2} - A\beta(\gamma_j)^{-1}A^{-1})$ by the above. Now let $a \otimes (z_1e_1 + z_2e_2) \in \ker(\rho_\Gamma(\gamma_j)Id_{2\times 2} - A\beta(\gamma_j)^{-1}A^{-1})$. Thus we have

$$(\rho_\Gamma(\gamma_j)(a) - y_ja) \otimes z_1e_1 - b_ja \otimes z_2e_2 = 0$$
$$(\rho_\Gamma(\gamma_j)(a) - y_j^{-1}a) \otimes z_2e_2 = 0.$$

Hence we have either $z_2 = 0$ or $\rho_\Gamma(\gamma_j)(a) - y_j^{-1}(a) = 0$ from the last equality, and $z_1 = 0$ or $\rho_\Gamma(\gamma_j)(a) - y_ja = 0$. If $z_1 = z_2 = 0$, then we are done; if

$\rho_\Gamma(\gamma_j)(a) - y_j^{-1}(a) = 0$ and $y_j \in U(1)$, then $a = 0$ follows from Lemma 3.1 (i) of;[LZ] if $y_j \notin U(1)$,

$$a = y_j \rho_\Gamma(\gamma_j)(a) = y_j^2 \rho_\Gamma(\gamma_j^2)(a) = \cdots = y_j^n \rho_\Gamma(\gamma_j^n)(a) = \cdots$$

hence $a = 0$ for $\rho_\Gamma(\gamma)$ unitary representation; if $z_2 = 0$ and $\rho_\Gamma(\gamma_j)(a) - y_j a = 0$, we have $a = 0$ by the same method. Hence $\ker(\rho_\Gamma(\gamma_j) Id_{2\times 2} - A\beta(\gamma_j)^{-1}A^{-1}) = \{0\}$. Thus, we obtain

$$\ker((\rho_\Gamma(\phi(x_j)) Id_{2\times 2}) \cdot \beta(\phi(x_j)) - Id) = \{0\}. \tag{16}$$

Hence $\Psi(x_j - 1) \in \mathcal{N}(\Gamma)$ is injective and has dense image (see[DIX]). The dense image property follows from (16) as we have

$$\begin{aligned} &Im((\rho_\Gamma(\phi(x_j)) Id_{2\times 2}) \cdot \beta(\phi(x_j)) - Id)^\perp \\ &= \ker((\rho_\Gamma(\phi(x_j^{-1})) Id_{2\times 2}) \cdot \beta^*(\phi(x_j^{-1})) - Id) = \{0\}, \end{aligned}$$

where $\beta^*(\phi(x_j^{-1}))$ is the complex conjugate of $\beta(\phi(x_j^{-1}))$.

(ii) Without loss of generality, we may assume $j = 1$ and $j' = 2$. Since $r_i = 1$ in $\mathbf{Z}[F_k]$, it is easy to see that

$$\sum_{l=1}^{k} \frac{\partial r_i}{\partial x_l}(x_l - 1) = 0 \quad \text{in} \quad \mathbf{Z}[F_k], \tag{17}$$

from which we get

$$\sum_{l=1}^{k} \Psi\left(\frac{\partial r_i}{\partial x_l}\right) \Psi(x_l - 1) = 0. \tag{18}$$

Let $A_1,\ A_2 : \underbrace{l^2(\Gamma)^{\otimes 2} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}}_{k-1} \to \underbrace{l^2(\Gamma)^{\otimes 2} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}}_{k-1}$ be the endomorphisms such that

$$A_1 = \Psi(x_1 - 1)|_{l^2(\Gamma)^{\otimes 2}} \oplus \mathrm{Id}|_{l^2(\Gamma)^{\otimes 2}} \oplus \cdots \oplus \mathrm{Id}|_{l^2(\Gamma)^{\otimes 2}}, \tag{19}$$

while A_2, when represented in the $(k-1)\times(k-1)$ matrix form, is determined by

$$A_{2,(l,1)} = \Psi(x_{l+1} - 1), \quad A_{2,(i,j)} = \delta_{ij} \cdot (\mathrm{Id} : l^2(\Gamma)^{\otimes 2} \to l^2(\Gamma)^{\otimes 2}) \quad \text{for} \quad j \geq 2. \tag{20}$$

From (18)-(20), and from the definition of $A^j_{\rho_\Gamma \otimes \beta}$, one deduces that

$$A^1_{\rho_\Gamma \otimes \beta} \cdot A_2 = -A^2_{\rho_\Gamma \otimes \beta} \cdot A_1. \tag{21}$$

By (i), it is easy to see that both A_1 and A_2 are injective. Thus $A^1_{\rho_\Gamma\otimes\beta}$ is injective if and only if $A^2_{\rho_\Gamma\otimes\beta}$ is injective.

By (8), one finds for $j = 1,\ 2$,

$$\mathrm{Det}_\tau(A_j) = \mathrm{Det}_\tau\left(\Psi(x_j - 1)\right). \tag{22}$$

From (7), (21) and (22), one gets (7). Q.E.D.

Now recall that the spectral radius of a matrix A is defined to be the maximal of eigenvalues of the matrix A. For $\beta(x_j) \in SL(2,\mathbb{C})$, its spectral radius of $\beta(x_j)$ equals to $\max\{y_j, y_j^{-1}\}$ for its eigenvalues y_j, y_j^{-1}. The absolute spectral radius of the $\beta(x_j)$ is $\max\{|y_j|, |y_j^{-1}|\}$.

Prop 3.1. For any $1 \le j \le k$, the following identity holds,

$$\mathrm{Det}_\tau\left(\Psi(x_j - 1)\right) = \mathrm{Det}_\tau(\rho_\Gamma(\gamma_j) - y_j)\cdot \mathrm{Det}_\tau(\rho_\Gamma(\gamma_j) - y_j^{-1}) = \max\{|y_j|, |y_j^{-1}|\}, \tag{23}$$

where $\beta(\gamma_j) = A^{-1}\begin{pmatrix} y_j & b_j \\ 0 & y_j^{-1}\end{pmatrix}A$ for some $A \in SL(2,\mathbb{C})$. I.e., $\mathrm{Det}_\tau\left(\Psi(x_j - 1)\right)$ equals to the absolute spectral radius of $\beta(\phi(x_j)) = \beta(\gamma_j)$.

Proof. By (8) and the proof of Lemma 3.1 (i), we have

$$A(\Psi(x_j - 1)A^{-1} = \rho_\Gamma(\gamma_j)Id_{2\times 2} - A\beta(\gamma_j)A^{-1} = \begin{pmatrix} \rho_\Gamma(\gamma_j) - y_j & b_j \\ 0 & \rho_\Gamma(\gamma_j) - y_j^{-1}\end{pmatrix} \tag{24}$$

Hence $\mathrm{Det}_\tau\left(\Psi(x_j - 1)\right) = \mathrm{Det}_\tau(\rho_\Gamma(\gamma_j) - y_j)\cdot\mathrm{Det}_\tau(\rho_\Gamma(\gamma_j) - y_j^{-1})$ by the Fuglede-Kadison determinant properties (iii) and (iv) in section 2.

If $y_j \in U(1)$, we know $\mathrm{Det}_\tau(\rho_\Gamma(\gamma_j) - y_j) = \mathrm{Det}_\tau(\rho_\Gamma(\gamma_j) - y_j^{-1}) = 1$ by Proposition 3.2 of.[LZ] If $|y_j| > 1$, then we can compute

$$\begin{aligned}\mathrm{Det}_\tau(\rho_\Gamma(\gamma_j) - y_j) &= \mathrm{Det}_\tau(y_j(y_j^{-1})\rho_\Gamma(\gamma_j) - Id)\\ &= |y_j|\cdot \mathrm{Det}_\tau(y_j^{-1}\rho_\Gamma(\gamma_j) - Id) = |y_j|,\end{aligned}$$

where the last identity follows from the proof of Proposition 3.2 of.[LZ]

$$\begin{aligned}\mathrm{Det}_\tau\left(\Psi(x_j - 1)\right) &= \mathrm{Det}_\tau(\rho_\Gamma(\gamma_j) - y_j)\cdot \mathrm{Det}_\tau(\rho_\Gamma(\gamma_j) - y_j^{-1})\\ &= \max\{1, |y_j|\}\cdot\max\{1, |y_j^{-1}|\}\\ &= \max\{1, |y_j|, |y_j^{-1}\} = \max\{|y_j|, |y_j^{-1}|\}.\end{aligned}$$

Hence our result follows. Q.E.D.

Recall the following types of transformations for group presentations: (Ia) To replace one of the relators r_i by its inverse r_i^{-1}; (Ib) To replace one of the relators r_i by its conjugate wr_iw^{-1} ($w \in F_k$); (Ic) To replace one of the relators r_i by r_ir_j ($i \neq j$); (II) To add a new generator x and a new relator xw^{-1} for any word w in terms of previous generators. Two presentations are strongly Tietze equivalent if there is a finite sequence of operations of transformation types (Ia), (Ib), (Ic) and (II) and their inverse such that one presentation can be transformed to another one.

The following result may be thought of as an twisted L^2-analogue of [LZ, Proposition 3.4] and [W, Theorem 2].

Theorem 3.1. *(1) The quantity*

$$\Delta_K^{(2)}(\beta) = \frac{\mathrm{Det}_\tau\left(A^j_{\rho_\Gamma\otimes\beta}\right)}{\max\{|y_j|, |y_j^{-1}|\}} \tag{25}$$

does not depend on the choice of the Wirtinger presentation $P(\Gamma)$ in (3.9) and $j = 1, 2, \cdots, k$.

(2) The quantity

$$\Delta_K^{(2)}(\beta, \beta^{-1}) = \frac{\sqrt{\mathrm{Det}_\tau\left(A^j_{\rho_\Gamma\otimes\beta}\right)\cdot \mathrm{Det}_\tau\left(A^j_{\rho_\Gamma\otimes\beta^{-1}}\right)}}{\max\{|y_j|, |y_j^{-1}|\}} \tag{26}$$

does not depend on the choice of the Wirtinger presentation $P(\Gamma)$ in (3.9) and $j = 1, 2, \cdots, k$.

Proof. (1) By Lemma 3.1 and Proposition 3.1, we have $\frac{\mathrm{Det}_\tau\left(A^j_{\rho_\Gamma\otimes\beta}\right)}{\max\{|y_j|,|y_j^{-1}|\}}$ is independent of j for $1 \leq j \leq k$. Without loss of generality, we assume that $j = 1$. Since by [W, Lemma 6], all Wirtinger presentations of Γ are strongly Tietze equivalent in the sense of,[W] we need only to show that $\frac{\mathrm{Det}_\tau(A^1_{\rho_\Gamma\otimes\beta})}{\max\{|y_1|,|y_1^{-1}|\}}$ is invariant under the transformations defining the strong Tietze equivalence. This can be carried out in the same way as in [LZ, Proposition 3.4] and [W, Proof of Theorem 2]. We present the proof for completeness.

Type Ia). To replace one of the relators r_i by its inverse r_i^{-1}.

This amounts to change the homomorphism $A_{\rho_\Gamma\otimes\beta}$ to $-A_{\rho_\Gamma\otimes\beta}$, which clearly does not change $\mathrm{Det}_\tau(A^1_{\rho_\Gamma\otimes\beta})$.

Type Ib). Note that

$$\Psi\left(\frac{\partial}{\partial x_j}\left(wr_iw^{-1}\right)\right) = \Psi(w)\Psi\left(\frac{\partial r_i}{\partial x_j}\right), \quad 1 \leq j \leq k. \tag{27}$$

By proceeding as in the proof of Lemma 2.1(ii), one finds that $\mathrm{Det}_\tau(A^1_{\rho_\Gamma\otimes\beta})$ changes by a factor $\mathrm{Det}_\tau(\Psi(w))$. Now by (4) and (10) one sees that $\Psi(w) \in GL(l^2(\Gamma)^{\otimes 2})$ is a unitary operator $\rho_\Gamma(w)Id_{2\times 2}$ tensor with $\beta(w) \in SL(2,\mathbb{C})$, which implies

$$\begin{aligned}
\mathrm{Det}_\tau(\Psi(w)) &= \mathrm{Det}_\tau\begin{pmatrix} \rho_\Gamma(\phi(w))y(\phi(w)) & b(\phi(w)) \\ 0 & \rho_\Gamma(\phi(w))y(\phi(w))^{-1}\end{pmatrix} \\
&= \mathrm{Det}_\tau(\rho_\Gamma(\phi(w))y(\phi(w))) \cdot \mathrm{Det}_\tau(\rho_\Gamma(\phi(w))y(\phi(w))^{-1}) \\
&= |y(\phi(w)|\mathrm{Det}_\tau(\rho_\Gamma(\phi(w))) \cdot |y(\phi(w))^{-1}|\mathrm{Det}_\tau(\rho_\Gamma(\phi(w))) \\
&= 1,
\end{aligned}$$

where we use the diagonal form of $\beta(w)$ since the determinant is unchanged. Thus the Type Ib) transformation does not change $\mathrm{Det}_\tau(A^1_{\rho_\Gamma\otimes\beta})$.

Type Ic). For this transformation we have

$$\Psi\left(\frac{\partial}{\partial x_j}(r_i r_m)\right) = \Psi\left(\frac{\partial r_i}{\partial x_j}\right) + \Psi\left(\frac{\partial r_m}{\partial x_j}\right), \quad 1 \le j \le k. \tag{28}$$

Let $B : \underbrace{l^2(\Gamma)^{\otimes 2} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}}_{k-1} \to \underbrace{l^2(\Gamma)^{\otimes 2} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}}_{k-1}$ be the endomorphism which, when expressed through $(k-1)\times(k-1)$ matrix, takes the form $B_{i,m} = Id$, while otherwise $B_{s,t} = \delta_{st}\cdot(\mathrm{Id} : l^2(\Gamma)^{\otimes 2} \to l^2(\Gamma)^{\otimes 2})$. Then one finds that $A^1_{\rho_\Gamma\otimes\beta}$ changes to $BA^1_{\rho_\Gamma\otimes\beta}$. Note that $\mathrm{Det}_\tau(B) = 1$. Thus the type (Ic) transformation does not change $\mathrm{Det}_\tau(A^1_{\rho_\Gamma\otimes\beta})$.

Type II). In this case, the corresponding endomorphism

$$A'^1_{\rho_\Gamma\otimes\beta} : \underbrace{l^2(\Gamma)^{\otimes 2} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}}_{k-1} \oplus l^2(\Gamma)^{\otimes 2} \to \underbrace{l^2(\Gamma)^{\otimes 2} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}}_{k-1} \oplus l^2(\Gamma)^{\otimes 2} \tag{29}$$

can be written as $A^1_{\rho_\Gamma\otimes\beta} \oplus \mathrm{Id}_{l^2(\Gamma)^{\otimes 2}}$ plus a mapping from $\underbrace{l^2(\Gamma)^{\otimes 2} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}}_{k-1}$ to $l^2(\Gamma)^{\otimes 2}$. Thus, by (8),

$$\mathrm{Det}_\tau\left(A'^1_{\rho_\Gamma\otimes\beta}\right) = \mathrm{Det}_\tau\left(A^1_{\rho_\Gamma\otimes\beta}\right). \tag{30}$$

The type (II) transformation does not change

$$\mathrm{Det}_\tau(A^1_{\rho_\Gamma\otimes\beta})/\max\{|y_1|, |y_1^{-1}|\}.$$

Similarly, the inverse of transformation (Ia), (Ib), (Ic) or (II) does not change $\mathrm{Det}_\tau(A^1_{\rho_\Gamma\otimes\beta})/\max\{|y_1|, |y_1^{-1}|\}$.

Therefore $\mathrm{Det}_\tau(A^1_{\rho_\Gamma\otimes\beta})/\max\{|y_1|,|y_1^{-1}|\} = \Delta_K^{(2)}(\beta)$ is invariant under the strong Tietze equivalent transformations. By [W, Lemma 6], it does not depend on the Wirtinger presentations of Γ.

(2) From $\beta(\gamma_j) = A^{-1}\begin{pmatrix} y_j & b_j \\ 0 & y_j^{-1}\end{pmatrix} A$, we have

$$\beta(\gamma_j)^{-1} = A^{-1}\begin{pmatrix} y_j^{-1} & -b_j \\ 0 & y_j\end{pmatrix} A.$$

Hence $\mathrm{Det}_\tau(A^1_{\rho_\Gamma\otimes\beta^{-1}})/\max\{|y_1^{-1}|,|y_1|\} = \Delta_K^{(2)}(\beta^{-1})$ is invariant under the strong Tietze equivalent transformations. Therefore

$$\Delta_K^{(2)}(\beta,\beta^{-1}) = \sqrt{\Delta_K^{(2)}(\beta)\cdot\Delta_K^{(2)}(\beta^{-1})}$$

is independent of the choice of the Wirtinger presentation. Q.E.D.

By Lemma 3.1, Proposition 3.1 and Theorem 3.1, we see that $\Delta_K^{(2)}(\beta)$ depends only on the knot K and the representation β. Since its construction is closely related to the usual construction of the (twisted) Alexander polynomial of a knot, we make the following definition.

Definition 3.1. (i) We define $\Delta_K^{(2)}(\beta)$ to be the twisted L^2–Alexander invariant of the knot K in S^3 with $\beta:\Gamma\to SL(2,\mathbb{C})$, and $\Delta_K^{(2)}:\mathrm{Hom}(\Gamma,SL(2,\mathbb{C}))\to\mathbb{R}_+$.

(ii) We define $\Delta_K^{(2)}(\beta,\beta^{-1})$ to be the twisted L^2-Alexander-Conway invariant of the knot K in S^3 with $\beta,\beta^{-1}\in\mathrm{Hom}(\Gamma,SL(2,\mathbb{C}))$.

Remark 3.1. It is clear that $\Delta_K^{(2)}(\beta)$ and $\Delta_K^{(2)}(\beta,\beta^{-1})$ can be defined by using any presentation of Γ which is strongly Tietze equivalent to some Wirtinger presentation of Γ. We conjecture that any presentation of a knot group with deficiency one is strongly Tietze equivalent to a Wirtinger presentation of the knot group.

Remark 3.2. When $\beta = \mathrm{diag}(t,t^{-1})$ is a $U(1)$ representation of $SL(2,\mathbb{C})$, $\Delta_K^{(2)}(\beta) = \Delta_K^{(2)}(t)$ has been studied in our previous work [LZ, Definition 3.5], and that $\Delta_K^{(2)}(1)$ is equivalent to the L^2–Reidemeister torsion of $S^3\backslash K$. For $t\in\mathbb{C}^*$, $\Delta_K^{(2)}(\beta)$ is proportional to the definition [LZ, (7.2)] for the L^2-invariant associated to $\mathbb{C}^*$ representations, and $\Delta_K^{(2)}(\beta,\beta^{-1})$ is the L^2-Alexander-Conway invariant studied in [LZ2, Theorem 3.2].

Note that for any representation $\beta:\Gamma\to SL(2,\mathbb{C})$, there is a (pseudo) developing map for β which is a smooth equivariant map $\widetilde{f_\beta}:\widetilde{S^3\setminus K}\to\mathbf{H}^3$

which sends $\partial\widetilde{S^3 \setminus K}$ into S^2_∞ and $\mathrm{int}\widetilde{S^3 \setminus K}$ into $\mathbf{H}^3$, where the boundary has an equivariant cone structure on the neighborhood of $\partial\widetilde{S^3 \setminus K}$ in $\widetilde{S^3 \setminus K}$ and some extra conditions (see[CCGLS,Dun,FR] for more details). Let $\widetilde{f_\beta}(Vol_{\mathbf{H}^3})$ be the pullback of the volume form on $\mathbf{H}^3$. Hence the descending 3-form $\pi_*(\widetilde{f_\beta}(Vol_{\mathbf{H}^3}))$ is well-defined since the (pseudo) developing map is equivariant, where $\pi : \widetilde{S^3 \setminus K} \to S^3 \setminus K$. Then the volume of β is defined to be the integral of the descending 3-form over $S^3 \setminus K$. In [Dun, Lemma 2.5.2], *the volume of β is well-defined for the β whose character lies in an irreducible component of the character variety of $S^3 \setminus K$ which contains the character of a discrete faithful representation.* Our twisted L^2-Alexander(-Conway) invariant is an extension of volume by the work of.[LuS] It is natural to ask that if there is any link between these two functions $Vol : \mathrm{Hom}(\Gamma, SL(2,\mathbb{C})) \to \mathbb{R}_+$ (see[Dun,FR]) and $\Delta^{(2)}_K : \mathrm{Hom}(\Gamma, SL(2,\mathbb{C})) \to \mathbb{R}_+$.

Remark 3.3. In view of [W, Section 5], the above construction can also be applied to links.

4. Twisted L^2-Reidemeister torsion

We first recall the definition of the L^2-Reidemeister torsion ([CFM,Lu1]).

Let (C_*, ∂) be a finite length $\mathcal{N}(\Gamma)$-chain complex

$$(C_*, \partial) : 0 \to C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_1} C_0 \to 0, \tag{1}$$

where each C_i, $0 \le i \le n$, is a (finite rank) $\mathcal{N}(\Gamma)$ free Hilbert module. We make the assumption that (C_*, ∂) is weakly acyclic, i.e., for any $0 \le i \le n$, $\ker(\partial_i) = \overline{\mathrm{Im}(\partial_{i-1})}$ (usually one uses the terminology "L^2-acyclic").

For any $0 \le i \le n$, let $\partial_i^* : C_{i-1} \to C_i$ be the adjoint of $\partial_i : C_i \to C_{i-1}$. Then $\partial_i\partial_i^* : \overline{\mathrm{Im}(\partial_i)} \to \overline{\mathrm{Im}(\partial_i)}$ is injective.

We say that (C_*, ∂) is of determinant class if for any $0 \le i \le n$, $\partial_i\partial_i^* : \overline{\mathrm{Im}(\partial_i)} \to \overline{\mathrm{Im}(\partial_i)}$ is of determinant class. In this case, we define the L^2-Reidemeister torsion of (C_*, ∂) to be a real number $T^{(2)}(C_*, \partial)$ given by

$$\log T^{(2)}(C_*, \partial) = -\frac{1}{2}\sum_{i=0}^{n}(-1)^i \log \mathrm{Det}_\tau\left(\partial_i\partial_i^*|_{\overline{\mathrm{Im}(\partial_i)}}\right) \tag{2}$$

(cf. [Lu1, Definition 3.29]).

Let X be a finite cell complex and let $\rho : \pi_1(X) \to GL(H)$ be an $\mathcal{N}(\Gamma)$-linear representation of $\Gamma = \pi_1(X)$ on a (finite rank) free $\mathcal{N}(\Gamma)$ Hilbert module. Let $\widetilde{X}$ be the universal covering of X. Then the chain complex

$(C_*(\widetilde{X}) \otimes H, \widetilde{\partial})$ induces canonically a chain complex $(C_*(X, H_\rho), \partial_\rho)$ in the sense of (1), where $C_*(X, H_\rho) = (C_*(\widetilde{X}) \otimes_{\pi_1(X),\rho} H)$.

If $(C_*(X, H_\rho), \partial_\rho)$ is weakly acyclic and of determinant class, one can then define its L^2-Reidemeister torsion $T^{(2)}(C_*(X, H_\rho), \partial_\rho)$ as in (2).

Remark 4.1. If $\rho : \pi_1(X) \to GL(H)$ is unitary, then $T^{(2)}(C_*(X, H_\rho), \partial_\rho)$ is a well-defined piecewise linear invariant. See[Lu1,MA] for more information.

Let $P(\Gamma)$ be a Wirtinger representation of the knot group of a knot K as in (3.9), where r_i is the cross relation for each i.

Let W be a 2-dimensional cell complex constructed from one 0-cell p, k 1-cells $x_1, \cdots, x_k$ and $(k-1)$ 2-cells $D_1, \cdots, D_{k-1}$ with attaching maps given by $r_1, \cdots, r_{k-1}$. It is well-known that the knot complement $S^3 \setminus K$ collapses to the 2-dimensional complex W.

Let

$$\rho_\beta = \rho \otimes \beta : \Gamma \to GL\left(l^2(\Gamma)\right) \otimes SL(2, \mathbb{C}) = GL\left(l^2(\Gamma)^{\otimes 2}\right) \tag{3}$$

denote the representation of Γ obtained from the tensor product of the representations in (3.11), (4) and (5).

Then the $\mathcal{N}(\Gamma)$-chain complex $(C_*(S^3 \setminus K, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})$ is weakly acyclic if and only if $(C_*(W, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})$ is weakly acyclic. Moreover, by the simple homotopy invariance of the L^2-Reidemeister torsion (cf. [LuR, Corollary 3.12]), $(C_*(S^3 \setminus K, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})$ is of determinant class if and only if $(C_*(W, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})$ is of determinant class, and in this case,

$$T^{(2)}\left(C_*\left(S^3 \setminus K, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}\right), \partial_{\rho_\beta}\right) = T^{(2)}\left(C_*\left(W, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}\right), \partial_{\rho_\beta}\right). \tag{4}$$

Prop 4.1. The complex $(C_*(S^3 \setminus K, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})$ is weakly acyclic if and only if $A^1_{\rho_\Gamma \otimes \beta}$ defined in (6) is injective. Moreover, $A^1_{\rho_\Gamma \otimes \beta}$ is of determinant class if and only if $(C_*(S^3 \setminus K, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})$ is of determinant class and one has

$$T^{(2)}\left(C_*\left(S^3 \setminus K, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}\right), \partial_{\rho_\beta}\right) = \frac{1}{\Delta^{(2)}_K(\beta)} \tag{5}$$

Proof. By (4), we need only to compute the L^2-Reidemeister torsion of $(C_*(W, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})$. For any $m \in \mathbf{N}$, let $(l^2(\Gamma)^{\otimes 2})^{[m]}_{\rho_\beta}$ denote the $\mathcal{N}(\Gamma)$-Hilbert module of rank m,

$$(l^2(\Gamma)^{\otimes 2})^{[m]}_{\rho_\beta} = \underbrace{l^2(\Gamma)^{\otimes 2}_{\rho_\beta} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}_{\rho_\beta}}_{m}. \tag{6}$$

Then the chain complex $(C_*(W, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})$ is given as follows,

$$0 \longrightarrow (l^2(\Gamma)^{\otimes 2})^{[k-1]}_{\rho_\beta} \xrightarrow{\partial_2} (l^2(\Gamma)^{\otimes 2})^{[k]}_{\rho_\beta} \xrightarrow{\partial_1} (l^2(\Gamma)^{\otimes 2})_{\rho_\beta} \longrightarrow 0, \qquad (7)$$

where ∂_2, when expressed through a $(k-1) \times k$ matrix (with respect to the right matrix multiplications), is given by

$$\partial_2 = A_{\rho\otimes\beta} = \left(\Psi\left(\frac{\partial r_i}{\partial x_j}\right)\right)_{(k-1)\times k} \quad \text{with } 1 \le i \le k-1, \quad 1 \le j \le k, \quad (8)$$

while

$$\partial_1 = (\Psi(x_1 - 1), \cdots, \Psi(x_k - 1))^t. \qquad (9)$$

By Lemma 3.1, $\Psi(x_1 - 1)$ has dense image, by (9) one sees that the L^2-homology is given by

$$H_0^{(2)}(C_*(W, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta}) = l^2(\Gamma)^{\otimes 2}_{\rho_\beta} / \overline{\mathrm{Im}(\partial_1)} = 0.$$

On the other hand, it is clear that

$$\begin{aligned}\mathrm{rk}^{(2)}\left(H_0^{(2)}(C_*(W, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})\right) &- \mathrm{rk}^{(2)}\left(H_1^{(2)}(C_*(W, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})\right) \\ + \mathrm{rk}^{(2)}\left(H_2^{(2)}(C_*(W, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})\right) &= 2 - 2k + 2(k-1) = 0, \quad (10)\end{aligned}$$

where $\mathrm{rk}^{(2)}$ is the notation of von Neumann rank (dimension). By (9) and (10), one sees that $(C_*(W, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})$ is weakly acyclic if and only if ∂_2 is injective if and only if $\mathrm{rk}^{(2)}(H_2^{(2)}(C_*(W, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})) = 0$.

Let $A' : \underbrace{l^2(\Gamma)^{\otimes 2}_{\rho_\beta} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}_{\rho_\beta}}_{k} \to \underbrace{l^2(\Gamma)^{\otimes 2}_{\rho_\beta} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}_{\rho_\beta}}_{k}$ be such that when expressed through the $k \times k$-matrix, one has

$$A'_{i,1} = \Psi(x_i - 1), \quad A'_{i,j} = \delta_{ij}\left\{\mathrm{Id} : l^2(\Gamma)^{\otimes 2}_{\rho_\beta} \to l^2(\Gamma)^{\otimes 2}_{\rho_\beta}\right\} \quad \text{for } 2 \le j \le k. \qquad (11)$$

From (18) and (11), one finds that the composition $A'\partial_2$, when expressed through the $(k-1) \times k$ matrix, takes the form

$$A'\partial_2 = \left(0, A^1_{\rho\Gamma\otimes\beta}\right) : \underbrace{l^2(\Gamma)^{\otimes 2}_{\rho_\beta} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}_{\rho_\beta}}_{k-1} \to \underbrace{l^2(\Gamma)^{\otimes 2}_{\rho_\beta} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}_{\rho_\beta}}_{k}. \qquad (12)$$

Note that $\Psi(x_1 - 1)$ is injective and has dense image. One sees easily that A' is also injective and has dense image by (11). By (12), we have ∂_2 is injective if and only if $A^1_{\rho\Gamma\otimes\beta}$ is injective. This proves that $(C_*(W, l^2(\Gamma)^{\otimes 2}_{\rho_\beta}), \partial_{\rho_\beta})$

(and thus $(C_*(S^3 \setminus K, l^2(\Gamma)^{\otimes 2}_{\rho\beta}), \partial_{\rho\beta})$) is weakly acyclic if and only if $A^1_{\rho\Gamma\otimes\beta}$ is injective.

In order to compute the L^2-Reidemeister torsion, one first observes that since here $\Psi(x_1 - 1)$ may not be invertible, the method in [Ki, Section 3] does not work directly. Here we follow what we did in [LZ, Proposition 5.1].

For any $\varepsilon > 0$, let $A'(\varepsilon) : \underbrace{l^2(\Gamma)^{\otimes 2}_{\rho\beta} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}_{\rho\beta}}_{k} \to \underbrace{l^2(\Gamma)^{\otimes 2}_{\rho\beta} \oplus \cdots \oplus l^2(\Gamma)^{\otimes 2}_{\rho\beta}}_{k}$ be such that when expressed through the $k \times k$-matrix, one has

$$A'_{1,1}(\varepsilon) = \Psi(x_1) - (1+\varepsilon)\left\{\mathrm{Id} : l^2(\Gamma_{\rho\beta})^{\otimes 2} \to l^2(\Gamma_{\rho\beta})^{\otimes 2}\right\}, \tag{13}$$

$$A'_{i,1}(\varepsilon) = \Psi(x_i) - \left\{\mathrm{Id} : l^2(\Gamma)^{\otimes 2}_{\rho\beta} \to l^2(\Gamma)^{\otimes 2}_{\rho\beta}\right\}, \quad 2 \leq i \leq k,$$

$$A'_{i,j}(\varepsilon) = \delta_{ij}\left\{\mathrm{Id} : l^2(\Gamma_{\rho\beta})^{\otimes 2} \to l^2(\Gamma_{\rho\beta})^{\otimes 2}\right\} \quad \text{for } 2 \leq j \leq k.$$

Clearly, for any $\varepsilon > 0$, $A'(\varepsilon)$ is invertible.

Let $(C_{*,\varepsilon}(W, l^2(\Gamma)^{\otimes 2}_{\rho\beta}), \partial_{\rho\beta})$ be the chain complex

$$0 \longrightarrow (l^2(\Gamma)^{\otimes 2})^{[k-1]}_{\rho\beta} \xrightarrow{\partial_2} (l^2(\Gamma)^{\otimes 2})^{[k]}_{\rho\beta,\varepsilon} \xrightarrow{\partial_1} l^2(\Gamma)^{\otimes 2}_{\rho\beta} \longrightarrow 0, \tag{14}$$

where $(l^2(\Gamma)^{\otimes 2})^{[k]}_{\rho\beta,\varepsilon}$ admits the new inner product $\langle\cdot,\cdot\rangle_\varepsilon$ in $l^2(\Gamma)^{\otimes 2}$ given by

$$\langle x, y\rangle_\varepsilon = \langle A'(\varepsilon)^* A'(\varepsilon) x, y\rangle. \tag{15}$$

By (13), we have, for any $\varepsilon > 0$,

$$\mathrm{Det}_\tau\left(A'(\varepsilon)\right) = \max\{1, (1+\varepsilon)|y_1|\} \cdot \max\{1, (1+\varepsilon)|y_1^{-1}|\}. \tag{16}$$

By (14)-(16) and [CFM, Proposition 3.11],

$$\max\{1,(1+\varepsilon)|y_1|\} \cdot \max\{1,(1+\varepsilon)|y_1^{-1}|\} T^{(2)}\left(C_{*,\varepsilon}(W, l^2(\Gamma)^{\otimes 2}_{\rho\beta}), \partial_{\rho\beta}\right) \tag{17}$$

$$= T^{(2)}\left(C_*(W, l^2(\Gamma)_{\rho\beta}), \partial_{\rho\beta}\right). \tag{18}$$

We now examine $T^{(2)}(C_{*,\varepsilon}(W, l^2(\Gamma)_{\rho_\alpha}), \partial_{\rho_\alpha})$ and its limit as $\varepsilon \to 0$. Let $\partial^*_{1,\varepsilon}$ denote the adjoint of ∂_1 with respect to the new inner product in $(C_{*,\varepsilon}(W, l^2(\Gamma)^{\otimes 2}_{\rho\beta}), \partial_{\rho\beta})$. Thus we get, by (15),

$$\langle \partial_1 x, y\rangle = \langle x, \partial_1^* y\rangle = \left\langle \left(A'(\varepsilon)^* A'(\varepsilon)\right)^{-1} x, \partial_1^* y\right\rangle_\varepsilon = \left\langle x, \left(A'(\varepsilon)^* A'(\varepsilon)\right)^{-1} \partial_1^* y\right\rangle_\varepsilon, \tag{19}$$

and $\partial_{1,\varepsilon}^* = (A'(\varepsilon)^* A'(\varepsilon))^{-1} \partial_1^* = A'(\varepsilon)^{-1} (A'(\varepsilon)^*)^{-1} \partial_1^*$. Therefore,

$$\partial_1 \partial_{1,\varepsilon}^* = \partial_1 A'(\varepsilon)^{-1} (A'(\varepsilon)^*)^{-1} \partial_1^*. \tag{20}$$

By (13), one deduces directly that $A'(\varepsilon)^{-1}$ can be written as

$$A'(\varepsilon)_{1,1}^{-1} = A'_{1,1}(\varepsilon)^{-1}, \tag{21}$$

$$A'(\varepsilon)_{i,1}^{-1} = -A'_{i,1}(\varepsilon) A'(\varepsilon)_{1,1}^{-1}, \quad 2 \leq i \leq k,$$

$$A'_{i,j}(\varepsilon)^{-1} = \delta_{ij} \left\{ \mathrm{Id} : l^2(\Gamma_{\rho_\beta})^{\otimes 2} \to l^2(\Gamma_{\rho_\beta})^{\otimes 2} \right\} \quad \text{for} \quad 2 \leq j \leq k.$$

By (9), (13) and (21), we have

$$\partial_1 A'(\varepsilon)^{-1} = \underbrace{(\mathrm{Id} + 2\varepsilon, 0 \cdots, 0)}_{k}{}^t - \varepsilon \left(A'(\varepsilon)_{1,1}^{-1}, \cdots, A'(\varepsilon)_{k,1}^{-1} \right)^t. \tag{22}$$

By (13), and (20)–(22), $\partial_1 \partial_{1,\varepsilon}^* =$

$$\left(\mathrm{Id} + \varepsilon A'(\varepsilon)_{1,1}^{-1}\right)^* \left(\mathrm{Id} + \varepsilon A'(\varepsilon)_{1,1}^{-1}\right) + \varepsilon^2 \sum_{j=2}^{k} \left(A'(\varepsilon)_{j,1}^{-1}\right)^* A'(\varepsilon)_{j,1}^{-1} =$$

$$\left(A'(\varepsilon)_{1,1}^{-1}\right)^* \left(\Psi(x_1 - 1)^* \Psi(x_1 - 1) + \varepsilon^2 \sum_{j=2}^{k} \Psi(x_j - 1)^* \Psi(x_j - 1) \right) \left(A'(\varepsilon)_{1,1}^{-1}\right). \tag{23}$$

By (13) and (21), one has that for any $\varepsilon > 0$,

$$\mathrm{Det}_\tau \left(A'(\varepsilon)_{1,1}^{-1}\right) = \frac{1}{(1+\varepsilon) \max\{|y_1|, |y_1^{-1}|\}}. \tag{24}$$

From (6), (7), Lemma 3.1, Proposition 3.1, (23) and (24), one finds

$$\lim_{\varepsilon \to 0^+} \mathrm{Det}_\tau \left(\partial_1 \partial_{1,\varepsilon}^*\right) = \lim_{\varepsilon \to 0^+} \frac{1}{(1+\varepsilon) \max\{|y_1|, |y_1^{-1}|\}^2} \cdot \max\{|y_1|, |y_1{}^{-1}|\}^2 = 1. \tag{25}$$

Similarly, by (15),

$$\langle \partial_2 x, y \rangle_\varepsilon = \langle A'(\varepsilon)^* A'(\varepsilon) \partial_2 x, y \rangle = \langle x, \partial_2^* A'(\varepsilon)^* A'(\varepsilon) y \rangle, \tag{26}$$

and the adjoint of ∂_2 with respect to the new inner product in $(C_{*,\varepsilon}(W, l^2(\Gamma)_{\rho_\beta}^{\otimes 2}), \partial_{\rho_\beta})$ is given by

$$\partial_{2,\varepsilon}^* = \partial_2^* A'(\varepsilon)^* A'(\varepsilon). \tag{27}$$

By (11), (12) and (27), we have

$$\partial_{2,\varepsilon}^*\partial_2 = \partial_2^* A'(\varepsilon)^* A'(\varepsilon)\partial_2 = A^1_{\rho r\otimes\beta}\left(A^1_{\rho r\otimes\beta}\right)^* \\ +\varepsilon^2\left(\Psi\left(\frac{\partial r_1}{\partial x_1}\right),\cdots,\Psi\left(\frac{\partial r_{k-1}}{\partial x_1}\right)\right)^t\left(\Psi\left(\frac{\partial r_1}{\partial x_1}\right)^*,\cdots,\Psi\left(\frac{\partial r_{k-1}}{\partial x_1}\right)^*\right). \tag{28}$$

By (6) and (28),

$$\lim_{\varepsilon\to 0+}\mathrm{Det}_\tau\left(\left(\partial_2\partial_{2,\varepsilon}^*\right)|_{\overline{\mathrm{Im}(\partial_2)}}\right) = \lim_{\varepsilon\to 0+}\mathrm{Det}_\tau\left(\partial_{2,\varepsilon}^*\partial_2\right) = \mathrm{Det}_\tau\left(A^1_{\rho r\otimes\beta}\right)^2. \tag{29}$$

By (2), (17), (25) and (29), one finds

$$T^{(2)}\left(C_*(W, l^2(\Gamma)^{\otimes 2}_{\rho\beta}), \partial_{\rho_\beta}\right) = \tag{30}$$

$$\lim_{\varepsilon\to 0+}\max\{1,(1+\varepsilon)|y_1|\}\cdot\max\{1,(1+\varepsilon)|y_1^{-1}|\}T^{(2)}\left(C_{*,\varepsilon}(W, l^2(\Gamma)^{\otimes 2}_{\rho\beta}), \partial_{\rho_\beta}\right)$$
$$= \lim_{\varepsilon\to 0+}\frac{\max\{1,(1+\varepsilon)|y_1|\}\cdot\max\{1,(1+\varepsilon)|y_1^{-1}|\}}{\mathrm{Det}_\tau\left(A^1_{\rho r\otimes\beta}\right)}$$
$$= \frac{\max\{|y_1|,|y_1^{-1}|\}}{\mathrm{Det}_\tau\left(A^1_{\rho r\otimes\beta}\right)}$$
$$= \frac{1}{\Delta_K^{(2)}(\beta)}.$$

and the result follows. Q.E.D.

Remark 4.2. We now have extended our previous result [LZ, Proposition 5.1] for the flat line bundle via $\alpha : \Gamma \to C^*$ to the twisted flat $SL(2,\mathbb{C})$ vector bundle via $\beta : \Gamma \to SL(2,\mathbb{C})$. Note that this flat $SL(2,\mathbb{C})$ bundle over $S^3 \setminus K$ plays an important role in understanding the volume conjecture beyond the leading order as pointed in.[GuMu] In,[GuMu] they claimed that the Ray-Singer torsion of the knot complement twisted by the flat connection associated to the representation $\beta : \Gamma \to SL(2,\mathbb{C})$ appears in the parameterized volume conjecture. Our L^2-Reidemeister torsion twisted by the representation $\beta : \Gamma \to SL(2,\mathbb{C})$ shows certainly close and possible role in the polarized volume conjecture.

5. Twisted L^2–Alexander(-Conway) invariant with parameters in character variety

We first recall the character variety and show that our twisted L^2–Alexander invariant is depending upon the character of the representation $\beta : \Gamma \to SL(2,\mathbb{C})$.

Let K be a knot in S^3 and M_K its complement. That is, $M_K = S^3 - N_K$ where N_K is the open tubular neighborhood of K in S^3. M_K is a compact 3-manifold with boundary $\partial M_K = T^2$ a torus. Denote by

$$R(M_K) = \mathrm{Hom}(\pi_1(M_K), SL_2(\mathbb{C})) = \mathrm{Hom}(\Gamma, SL(2,\mathbb{C})$$

$$R(\partial M_K) = \mathrm{Hom}(\pi_1(\partial M_K), SL_2(\mathbb{C})) = \mathrm{Hom}(\pi_1(T^2), SL(2,\mathbb{C})).$$

It is known that they are affine algebraic sets over the complex numbers $\mathbb{C}$ and so are the corresponding character varieties $X(M_K)$ and $X(\partial M_K)$ (See[CS]). We also have the canonical surjective morphisms $t : R(M_K) \longrightarrow X(M_K)$ and $t : R(\partial M_K) \longrightarrow X(\partial M_K)$ which map a representation to its character. The natural homomorphism $i : \pi_1(\partial M_K) \longrightarrow \pi_1(M_K)$ induces the restriction maps $r : X(M_K) \longrightarrow X(\partial M_K)$ and $r : R(M_K) \longrightarrow R(\partial M_K)$.

Note that $\pi_1(\partial M_K) = \mathbb{Z} \oplus \mathbb{Z}$ is generated by two classes, the meridian μ and the longitude λ as its generators. Let R_D be the subvariety of $R(\partial M_K)$ consisting of the diagonal representations. Then R_D is isomorphic to $\mathbb{C}^* \times \mathbb{C}^*$. Indeed, for $\rho \in R_D$, we obtain

$$\rho(\lambda) = \begin{bmatrix} l & 0 \\ 0 & l^{-1} \end{bmatrix} \text{ and } \rho(\mu) = \begin{bmatrix} m & 0 \\ 0 & m^{-1} \end{bmatrix},$$

then we assign the pair (l, m) to ρ. Clearly this is an isomorphism. We shall denote by t_D the restriction of the morphism $t : R(\partial M_K) \longrightarrow X(\partial M_K)$ on R_D.

Next we recall the definition of the A-polynomial of K which was introduced in.[CCGLS] Denote by $X'(M_K)$ the union of the irreducible components Y' of $X(M_K)$ such that the closure $\overline{r(Y')}$ in $X(\partial M_K)$ is 1-dimensional. For each component Z' of $X'(M_K)$, denote by Z the curve $t_D^{-1}(\overline{r(Y')}) \subset R_D$. We define D_K to be the union of the curves Z as Z' varies over all components of $X'(M_K)$. Via the above identification of R_D with $\mathbb{C}^* \times \mathbb{C}^*$, D_K is a curve in $\mathbb{C}^* \times \mathbb{C}^*$. Now by definition the *A-polynomial* $A(l, m)$ of K is the defining polynomial of the closure of D_K in $\mathbb{C} \times \mathbb{C}$.

From now on, we assume that K is a hyperbolic knot. Denote by $\beta_0 : \pi_1(M_K) \longrightarrow PSL_2(\mathbb{C})$ the discrete, faithful representation corresponding to the hyperbolic structure on M_K. Note that β_0 can be lifted to a $SL_2(\mathbb{C})$ representation. Moreover, there are exactly $|H^1(M_K;\mathbb{Z}_2)=\mathbb{Z}_2|=2$ such lifts.

Prop 5.1. If one of the $A^j_{\rho_\Gamma\otimes\beta}$ $(1\le j\le k)$ is injective and of determinant class for $\beta\in \mathrm{Hom}(\Gamma,SL(2,\mathbb{C}))$, then, for irreducible representations with $[\beta]=[\beta']\in X(M_K)$,

$$\Delta^{(2)}_K(\beta)=\Delta^{(2)}_K(\beta'),\qquad \Delta^{(2)}_K(\beta,\beta^{-1})=\Delta^{(2)}_K(\beta',(\beta')^{-1}).$$

Proof. Without loss of generality, we assume $j=1$ and simply compare $\mathrm{Det}_\tau(A^1_{\rho_\Gamma\otimes\beta})$ with $\mathrm{Det}_\tau(A^1_{\rho_\Gamma\otimes\beta'})$ since the absolute spectral radius is unchanged under the character map.

Let A_β be the morphism $A^1_{\rho_\Gamma\otimes\beta}$, and A^*_β be the adjoint operator of A_β. Note that two $SL(2,\mathbb{C})$ representations with the same character must be either equivalent or they are reducible. If β is irreducible and $[\beta]=[\beta']\in X(M_K)$, then we have β and β' are equivalent by [CS, Proposition 1.5.2]. Therefore

$$\mathrm{Det}_\tau(A^1_{\rho_\Gamma\otimes\beta})=\mathrm{Det}_\tau(A^1_{\rho_\Gamma\otimes\beta'}),$$

by the property of Fuglede-Kadison determinant in section 2. Hence the L^2-Alexander and L^2-Alexander-Conway invariants are unchanged under the character map. Q.E.D.

Remark 5.1. If β is reducible and $[\beta]=[\beta']$, then β' is also reducible by [CS, Corollary 1.2.2]. If both β and β' are non-abelian reducible, then $\mathrm{tr}\beta(c)=\mathrm{tr}\beta'(c)=2$ for each element of the commutator subgroup of Γ by [CS, Lemma 1.2.1]. Since $\beta[\Gamma,\Gamma]$ is normal in $\beta(\Gamma)$ and there is a unique 1-dimensional invariant subspace $L\subset\mathbb{C}^2$ of $\beta[\Gamma,\Gamma]$, we have the subspace L is fixed by $\beta(\Gamma)$. By [CS, Corollary 1.2.2], the same holds for β' since β' is reducible too. Then they both fix the same 1-dimensional subspace by changing the basis of $\mathbb{C}^2$. We have β and β' are equivalent under the conjugacy. If both β and β' are abelian reducible, then they are equivalent under the conjugacy. Hence the L^2-Alexander and L^2-Alexander-Conway invariants are unchanged under the character map.

Remark 5.2. Proposition 5.1 shows that the twisted L^2–Alexander invariant or the twisted L^2–Reidemeister torsion of the knot complement, can be reduced into the function of the character variety $X(M_K)$. Note that the volume function over the irreducible characters factor through a map to a zero-locus of A-polynomial by [Dun, Theorem 2.6]. Initially we are trying to push this further into the A-polynomial, but the difficulty relies on the understanding of the restriction map $r : X(M_K) \to X(\partial M_K)$ with the changing of Von Neuman group algebra and the Fuglede-Kadison determinant with respect to different von Neumann algebra.

Remark 5.3. If K is a hyperbolic knot, then the L^2-Reidemeister torsion of the knot complement determines the hyperbolic volume of the knot complement [LuS, Theorem 0.6] (cf. [Lu1, Theorem 4.3]). For our twisted L^2-Reidemeister torsion of the knot complement, is there any relation with the volume function parameterized by the zero locus of A-polynomial after identifying elements in the character variety ?

6. Twisted L^2-Alexander invariant with $GL(n,\mathbb{C})$ representations

In this section, we first replace the representation β considered in (3.11) and (4) by the representation

$$\beta' : \Gamma \to GL(n,\mathbb{C}) \tag{1}$$

with $\beta'(x_1), \cdots, \beta'(x_k) \in GL(n,\mathbb{C})$. Then by proceeding as in Section 1.6, we have identified

$$l^2(\Gamma) \otimes \mathbb{C}^n \cong l^2(\Gamma)^{\otimes n} \cong l^2(\Gamma) \otimes e_1 \oplus \cdots \oplus l^2(\Gamma) \otimes e_n. \tag{2}$$

By identifying the tensor representation of ρ_Γ and β', we have

$$\rho_\Gamma \otimes \beta'(\gamma) = (\rho_\Gamma(\gamma) Id_{n\times n}) \cdot \beta'(\gamma). \tag{3}$$

Consider the morphism

$$A_{\rho_\Gamma \otimes \beta'} : \underbrace{l^2(\Gamma)^{\otimes n} \oplus \cdots \oplus l^2(\Gamma)^{\otimes n}}_{k-1} \to \underbrace{l^2(\Gamma)^{\otimes n} \oplus \cdots \oplus l^2(\Gamma)^{\otimes n}}_{k} \tag{4}$$

which, when written as a $(k-1) \times k$-matrix, the (i,j)-component is given by

$$A_{\rho_\Gamma \otimes \beta', (i,j)} = \Psi\left(\frac{\partial r_i}{\partial x_j}\right) \in \mathcal{N}(\Gamma), \tag{5}$$

where $\frac{\partial r_i}{\partial x_j}$ is the standard Fox derivative and $l^2(\Gamma)^{\otimes n}$ is identified in (2).

We call $A_{\rho_\Gamma \otimes \beta'}$ the $GL(n,\mathbb{C})$ twisted L^2-Alexander matrix of the presentation $P(\Gamma)$ associated to the fundamental representation ρ_Γ and the $GL(n,\mathbb{C})$ representation β'.

For any $1 \leq j \leq k$, let

$$A^j_{\rho_\Gamma \otimes \beta'} : \underbrace{l^2(\Gamma)^{\otimes n} \oplus \cdots \oplus l^2(\Gamma)^{\otimes n}}_{k-1} \to \underbrace{l^2(\Gamma)^{\otimes n} \oplus \cdots \oplus l^2(\Gamma)^{\otimes n}}_{k-1} \tag{6}$$

denote the morphism obtained from $A_{\rho_\Gamma\otimes\beta'}$ by removing the j-th column from its matrix form.

Lemma 6.1. *(i) For any $1 \le j \le k$, $\Psi(x_j - 1) \in \mathcal{N}(\Gamma)$ is injective and has dense image. (ii) If one of the $A^j_{\rho_\Gamma\otimes\beta'}$'s, $1 \le j \le k$, is injective, then every $A^j_{\rho_\Gamma\otimes\beta'}$, $1 \le j \le k$, is injective. Moreover, in this case, for any $1 \le j < j' \le k$, one has*

$$\mathrm{Det}_\tau\left(A^j_{\rho_\Gamma\otimes\beta'}\right)\mathrm{Det}_\tau\left(\Psi(x_{j'} - 1)\right) = \mathrm{Det}_\tau\left(A^{j'}_{\rho_\Gamma\otimes\beta'}\right)\mathrm{Det}_\tau\left(\Psi(x_j - 1)\right). \quad (7)$$

Proof. Note that there exist an $A \in GL(n, \mathbb{C})$ such that $A\beta'(x_j)A^{-1} = \mathrm{diag}(y_j^1, y_j^2, \cdots, y_j^n)$. Then the rest follows from the same argument in the proof of Lemma 3.1. Q. E. D.

Theorem 6.1. *(i) For any $1 \le j \le k$, the following identity holds,*

$$\mathrm{Det}_\tau\left(\Psi(x_j - 1)\right) = \mathrm{Det}_\tau(\rho_\Gamma(\gamma_j) - y_j^1)\cdots\mathrm{Det}_\tau(\rho_\Gamma(\gamma_j) - y_j^n)$$

$$= \max\{|y_j^i| : 1 \le i \le n\},$$

where $A\beta'(x_j)A^{-1} = \mathrm{diag}(y_j^1, y_j^2, \cdots, y_j^n)$ for some $A \in GL(n, \mathbb{C})$.

(ii) The quantity

$$\Delta_K^{(2)}(\beta') = \frac{\mathrm{Det}_\tau\left(A^j_{\rho_\Gamma\otimes\beta'}\right)}{\max\{|y_j^i|, 1 \le i \le n\}} \quad (8)$$

does not depend on the choice of the Wirtinger presentation $P(\Gamma)$ in (3.9) and $j = 1, 2, \cdots, k$.

(iii) The quantity

$$\Delta_K^{(2)}(\beta', (\beta')^{-1}) = \sqrt{\frac{\mathrm{Det}_\tau\left(A^j_{\rho_\Gamma\otimes\beta'}\right)}{\max\{|y_j^i|, 1 \le i \le n\}} \cdot \frac{\mathrm{Det}_\tau\left(A^j_{\rho_\Gamma\otimes(\beta')^{-1}}\right)}{\max\{|y_j^i|^{-1}, 1 \le i \le n\}}}$$

does not depend on the choice of the Wirtinger presentation $P(\Gamma)$ in (3.9) and $j = 1, 2, \cdots, k$.

Proof. The proof is same as the proof of Proposition 3.1 and Theorem 3.1. Q.E.D.

Remark 6.1. If ρ_Γ is a trivial representation of Γ into $GL(l^2(\Gamma))$, then we have

$$\Delta_K^{(2)}(\beta') = \frac{|\mathrm{Det}\left(A^j_{Id\otimes\beta'}\right)|}{\max\{|y^i_j|, 1 \le i \le n\}},$$

where $\mathrm{Det}\left(A^j_{Id\otimes\beta'}\right)$ is the usual determinant and the absolute value is arised from the Fuglede-Kadison determinant. This relates our twisted L^2-Alexander invariant with the finite dimensional twisted Alexander invariant defined in.[L,W]

Remark 6.2. It would be interesting to know what our twisted L^2-Alexander(-Conway) invariant really measures when the the infinite dimensional representation ρ_Γ is not the fundamental representation of Γ.

References

BG. D. Bar-Natan and S. Garoufalidis, On the Melvin-Morton-Rozansky conjecture. *Invent. Math.* 1 (1996), 103-133.

CFM. A. Carey, M. Farber and V. Mathai, Determinant lines, von Neumann algebras and L^2 torsion. *J. Reine Angew. Math.* 484 (1997), 153-181.

CCGLS. Cooper, D., Culler, M., Gillet, H., Long, D.D., Shalen, P.B., *Plane curves associated to character varieties of 3-manifolds*, Invent. Math. 118(1994), 47-74.

CS. Culler, M., Shalen, P.B., *Varieties of group representations and splittings of 3-manifolds*, Ann. Math., (2) 117(1464), 109-146.

DIX. J. Dixmier, von Neumann Algebra, North-Holland Publishing Co., Amsterdam, 1981.

Dun. N. Dunfield, Cyclic surgery, degrees of maps of charcter curves, and volume rigidity for hyperbolic manifolds, Invent. Math. 136, 623-657 (1999).

FR. S. Francaviglia, Hyperbolic volume of representations of fundamental groups of cusped 3-manifolds, IMRN, No.9, 425-459 (2004).

FK. B. Fuglede and R. Kadison, Determinant theory in finite factors. *Ann. of Math.* (2), 55 (1952), 520-530.

GH. E. Ghate and E. Hironaka, The arithmetic and geometry of Salem numbers. *Bull. AMS.* 38 (2001), 293-314.

GKM. H. Goda, T. Kitano and T. Morifuji, Reidemeister torsion, twisted Alexander polynomial and fibered knots. *Preprint*, math.GT/0311155.

Gu. S. Gukov, Three-dimensional quantum gravity, Chern-Simons theory, and the A-polynomial. *Preprint*, hep-th/0306165.

GuMu. Gukov, S., Murakami, H., *SL(2,C) Chern-Simons theory and the asymptotic behavior of the colored Jones polynomial*, arXiv:math.GT/0608324.

JW. B. Jiang and S. Wang, Twisted topological invariants associated with representations. In *Topics in Knot Theory*, ed. M. E. Bozhüyük. Kluwer, Academic Publishers, Dordrecht, 1993, pp. 211-227.

K. R. M. Kashaev, The hyperbolic volume of knots from the quantum dilogarithm. *Lett. Math. Phys.* 39 (1997), 269-275.

KL. P. Kirk and C. Livingston, Twisted Alexander invariants, Reidemeister torsion, and Casson-Gordon invariants. *Topology* 38 (1999), 635-661.

Ki. T. Kitano, Twisted Alexander polynomial and Reidemeister torsion. *Pacific J. Math.* 174 (1996), 431-442.

L. X.-S. Lin, Representations of knot groups and twisted Alexander polynomials. *Acta Math. Sinica (English Series)* 17 (2001), 361-380.

LW. X.-S. Lin and Z. Wang, Random work on knot diagrams, colored Jones polynomial and Ihara-Selberg zeta function. *Knots, Braids, and Mapping Class Groups—Papers Dedicated to Joan S. Birman.* (New York, 1998), 107–121, AMS/IP Stud. Adv. Math., 24, Amer. Math. Soc., Providence, RI, 2001.

LZ. W. Li and W. Zhang, *An L^2-Alexander invariant for knots*, Commun. Contemp. Math., Vol 8 No 2(2006), 167-187.

LZ2. W. Li and W. Zhang, *An L^2-Alexander-Conway invariant for knots and the volume conjecture*, Differential Geometry and Physics, 303-312, Nankai Tracts Math., 10, Hackensack, NJ, 2006.

LWa. W. Li and Q. Wang, *On the generalized volume conjecture and regulator*, to appear in Commun. Contemp. Math..

Lu1. W. Lück, *L^2-Invariants: Theory and Applications to Geometry and K-Theory.* Springer-Verlag, 2002.

Lu2. W. Lück, L^2-torsion and 3-manifolds. *Low-Dimensional Topology.* Edited by K. Johannson, 72-107, International Press, 1994.

LuR. W. Lück and M. Rothenberg, Reidemeister torsion and the K-theory of von Neumann algebra. *K-Theory* 5 (1991), 213-264.

LuS. W. Lück and T. Schick, L^2-torsion of hyperbolic manifolds of finite volume. *Geom. Funct. Anal.* 9 (1999), 518-567.

MA. V. Mathai, L^2-analytic torsion. *J. Funct. Anal.*, 107(2) (1992), 369-386.

MM. P. Melvin and H. Morton, The colored Jones function. *Commun. Math. Phys.* 169 (1995), 501-520.

Mi. J. Milnor, A duality theorem for Reidemeister torsion. *Ann. of Math.* 76 (1962), 137-147.

Mu. H. Murakami, Mahler measure of the colored Jones polynomial and the volume conjecture. *Preprint*, math.GT/0206249.

MuM. H. Murakami and J. Murakami, The colored Jones polynomials and the simplicial volume of a knot. *Acta. Math.* 186 (2001), 85-104.

R. L. Rozansky, A contribution of the trivial connection to the Jones polynomial and Witten's invariant of 3-manifolds. *Commun. Math. Phys.* 175 (1996), 275-318.

RS. L. Rozansky and H. Saleur, Reidemeister torsion, the Alexander polynomial and $U(1,1)$ Chern-Simons theory. *J. Geom. Phys.* 13 (1994), 105-123.

W. M. Wada, Twisted Alexander polynomial for finitely presented groups. *Topology* 33 (1994), 241-256.

Wi. E. Witten, Quantum field theory and the Jones polynomial. *Commun. Math. Phys.* 121 (1989), 351-399.

Existence of Knots of Minimum Energy and Topological Growth Laws in the Faddeev Model

Fanghua Lin

Courant Institute of Mathematical Sciences
New York University
New York, New York 10012, USA

Yisong Yang

Department of Mathematics
Polytechnic University
Brooklyn, New York 11201, USA

Dedicated to the memory of Xiao-Song Lin

The Faddeev knots are the energy minimizers topologically stratified by the Hopf invariant in the Faddeev quantum field theory model governing the interaction of baryons and mesons. Recent progress made on the existence theory indicates that two growth laws expressed in terms of the Faddeev energy and the Hopf invariant are essential. The first one, called the Substantial Inequality, describes the energy splitting pattern in a minimization process and gives valuable information on compactness or convergence of a minimizing sequence. The second one, which will be shown to be universally valid for a broad range of energy functionals, ensures that knotted structures are preferred over multiple-soliton structures in high Hopf numbers.

1. Introduction

Prelude. During August 20 – 26, 2005, the 23rd International Conference on Differential Geometry Methods in Theoretical Physics was held at the Chern Institute of Mathematics. At the conference, we reported our work in a talk entitled "Faddeev Knots and Skyrme Solitons." Xiao-Song was in the audience and showed a lot of interest in our work. Since that time until May 2006, we had many conversations and email exchanges with Xiao-Song on this subject. Below, we first describe what were reported to Xiao-Song in 2005 and we then present some new development, which would please Xiao-Song if he were here with us today.

The concept of knots has important applications in science. In the past 100 years, mathematicians have made great progress in topological and combinatorial classifications of knots. In turn, the development of knot theory has also facilitated the advancement of mathematics in several of its frontiers, especially low-dimensional topology. In knot theory, an interesting problem concerns the existence of "ideal knots," which promises to provide a natural link between the geometric and topological contents of knotted structures. This problem has its origin in theoretical physics in which one wants to prove the existence and predict the properties of knots "based on a first principle approach".[59] In other words, one is interested in determining the detailed physical characteristics of a knot such as its energy (mass), geometric conformation, and topological identification, via conditions expressed in terms of temperature, viscosity, electromagnetic, nuclear, and possibly gravitational, interactions, which is also known as an Hamiltonian approach to knots as field-theoretical stable solitons. The Faddeev knots are such structures based on a first-principle approach and arise as knotted solitons in the Faddeev quantum field theory model.[9,10,27–30,59]

In normalized form, the action density of the Faddeev model over the standard $(3+1)$-dimensional Minkowski space of signature $(+---)$ reads

$$\mathcal{L} = \partial_\mu \mathbf{n} \cdot \partial^\mu \mathbf{n} - \frac{1}{2} F_{\mu\nu}(\mathbf{n}) F^{\mu\nu}(\mathbf{n}), \tag{1.1}$$

where the field $\mathbf{n} = (n_1, n_2, n_3)$ assumes its values in the unit 2-sphere in $\mathbb{R}^3$ and $F_{\mu\nu}(\mathbf{n}) = \mathbf{n} \cdot (\partial_\mu \mathbf{n} \wedge \partial_\nu \mathbf{n})$. Since $\mathbf{n}$ is parallel to $\partial_\mu \mathbf{n} \wedge \partial_\nu \mathbf{n}$, it is seen that $F_{\mu\nu}(\mathbf{n}) F^{\mu\nu}(\mathbf{n}) = (\partial_\mu \mathbf{n} \wedge \partial_\nu \mathbf{n}) \cdot (\partial^\mu \mathbf{n} \wedge \partial^\nu \mathbf{n})$,which may be identified with the well-known Skyrme term[35,55,68–71,83] when one embeds S^2 into $S^3 \approx SU(2)$. Hence, the Faddeev model may be viewed as a refined Skyrme model and the solution configurations of the former are the solution configurations of the latter with a restrained range.[23] In what follows, we shall only be interested in static fields which make the Faddeev energy

$$E(\mathbf{n}) = \int_{\mathbb{R}^3} \left\{ \sum_{j=1}^{3} |\partial_j \mathbf{n}|^2 + \frac{1}{2} \sum_{j,k=1}^{3} |F_{jk}(\mathbf{n})|^2 \right\} \mathrm{d}^3 x \tag{1.2}$$

finite. The finite-energy condition implies that $\mathbf{n}$ approaches a constant vector $\mathbf{n}_\infty$ at spatial infinity (of $\mathbb{R}^3$). Hence we may compactify $\mathbb{R}^3$ into S^3 and view the fields as maps from S^3 to S^2. As a consequence, we see that each finite-energy field configuration $\mathbf{n}$ is associated with an integer, $Q(\mathbf{n})$, in $\pi_3(S^2) = \mathbb{Z}$ (the set of all integers). In fact, such an integer $Q(\mathbf{n})$ is known as the Hopf invariant which has the following integral characterization: The differential form $F = F_{jk}(\mathbf{n}) \mathrm{d}x^j \wedge \mathrm{d}x^k$ $(j, k = 1, 2, 3)$ is closed in $\mathbb{R}^3$. Thus, there is a one form, $A = A_j dx^j$ so that $F = dA$. Then the Hopf charge $Q(\mathbf{n})$ of the map $\mathbf{n}$ may be evaluated by the integral

$$Q(\mathbf{n}) = \frac{1}{16\pi^2}\int_{\mathbb{R}^3} A \wedge F, \tag{1.3}$$

due to J. H. C. Whitehead,[81] which is a special form of the Chern–Simons invariant.[21,22]

The existence of the Faddeev knotted solitons are realized as the solutions to the problem

$$E_N \equiv \inf\{E(\mathbf{n}) \,|\, E(\mathbf{n}) < \infty,\ Q(\mathbf{n}) = N\}, \qquad N \in \mathbb{Z}, \tag{1.4}$$

referred to as the Faddeev Knot Problem.

Thus we encounter a direct minimization problem over the full space $\mathbb{R}^3$. In such a situation, a typical difficulty is that the minimizing sequence may fail to "concentrate" in a local region, which reminds us to look at what the concentration-compactness principle of P. L. Lions[53,54] can offer. A careful examination of the Faddeev Knots Problem indicates that we cannot make direct use of this method due to the lack of several key ingredients in the Faddeev energy (4) and in the Hopf–Whitehead topological integral (1.3).

A key tool we used was called later by us as "the Substantial Inequality" which may well be explained by what happens in a nuclear fission process: When a nucleus fissions, it splits into several smaller fragments. The sum of the masses of these fragments is less than the original mass. The "missing" mass has been converted into energy according to Einstein's equation.

On the other hand, in our general framework of minimization of a physical energy functional E subject to a topological constraint given by an integer invariant class $Q = N$, we may similarly expect an energy splitting of the configuration sequence into finitely many substantial constituents of topological charges $Q = N_s$ $(s = 1, 2, \cdots, k)$. We expect that the charge is conserved and the energy of the "particle" of charge N splits into the sum of energies E_{N_s} $(s = 1, 2, \cdots, k)$ of the "substantial particles" of respective charges N_s $(s = 1, 2, \cdots, k)$. Therefore, we expect to have

$$N = N_1 + N_2 + \cdots + N_k \quad \text{(charge conservation equality)}, \tag{1.5}$$

$$E_N \geq E_{N_1} + E_{N_2} + \cdots + E_{N_k} \quad \text{(energy conservation inequality)}. \tag{1.6}$$

Note that (1.6) is read as an energy conservation relation since possible extra energy may be needed for the substances or constituents of energies $E_{N_1}, E_{N_2}, \cdots, E_{N_k}$ to form a bound state or composite particle, of energy E_N, and, as a result, the composite particle may carry more energy than the sum of the energies of its substances or constituents. Hence we collectively call the above two relations "the Substantial Inequality" which spells out a first kind of topological growth law describing how energy and topology split in a general minimization process. The importance of this inequality is that it characterizes the situation when concentration occurs for a minimizing sequence. In other words, the charge-energy splitting above is nontrivial ($k \geq 2$) for a certain charge N if and only if concentration fails there.

To see how (1.5) and (1.6) can be used to quickly deduce an existence theorem for the Faddeev minimization problem (1.4) in 3 dimensions, we recall the topological lower bound

$$E(\mathbf{n}) \geq C|Q(\mathbf{n})|^{3/4} \tag{1.7}$$

established by Vakulenko and Kapitanski[77] where $C > 0$ is a universal constant. Hence $E_N > 0$ for any $N \neq 0$.

Define

$$\mathbb{S} = \{N \in \mathbb{Z} \setminus \{0\} \,|\, \text{the Faddeev Problem (1.4) has a solution at } N\}. \tag{1.8}$$

The Faddeev Knot Problem asks whether or not there holds $\mathbb{S} = \mathbb{Z}$. As a first step toward this question, we have

Theorem 1.1. *The set $\mathbb{S}$ is not empty.*

The proof[48] amounts to establishing the Substantial Inequality for the Faddeev energy (1.2) and noting that if $\mathbb{S}$ is empty, then the splitting expressed in (1.5) and (1.6) will continue forever, which contradicts the finiteness and positiveness of E_N for any N.

With (1.5), we can learn more about the soluble set $\mathbb{S}$. For example, choose $N_0 \in \mathbb{Z} \setminus \{0\}$ so that

$$E_{N_0} = \min\{E_N \,|\, N \in \mathbb{Z} \setminus \{0\}\}. \tag{1.9}$$

Then we must have $N_0 \in \mathbb{S}$ because a nontrivial splitting given in the Substantial Inequality will be impossible by the definition of N_0. Thus we can state[48]

Theorem 1.2. *The least energy point in the Faddeev energy spectrum $\{E_N \,|\, N \in \mathbb{Z} \setminus \{0\}\}$ is attainable, or $N_0 \in \mathbb{S}$.*

More knowledge about the set $\mathbb{S}$ can be deduced from the Substantial Inequality after we realize that the fractional exponent 3/4 in the lower bound (1.7) is in fact sharp by establishing[48] the sublinear energy upper bound

$$E_N \leq C_1|N|^{3/4}, \tag{1.10}$$

where C_1 is a universal positive constant (cf.[39] for some estimates for the value of C_1), which enables us[48] to obtain

Theorem 1.3. *The set $\mathbb{S}$ is an infinite subset of $\mathbb{Z}$.*

Here is a quick proof of the theorem. Otherwise assume that $\mathbb{S}$ is finite. Set $N^0 = \max\{N \in \mathbb{S}\}$ and let $N_0 \in \mathbb{S}$ be such that $E_{N_0} = \min\{E_N \mid N \in \mathbb{S}\}$, as defined earlier. Taking repeated decompositions if necessary, we may assume that all the integers $N_1, N_2, \cdots, N_k$ in (1.5) and (1.6) are in $\mathbb{S}$ already. Hence $|N_1|, |N_2|, \cdots, |N_k| \leq N^0$. Thus, in view of (1.5), we have $N \leq kN^0$; in view of (1.6), we have $E_N \geq kE_{N_0}$. Consequently, $E_N \geq (E_{N_0}/N^0)N$, which contradicts (1.10) when N is sufficiently large. Hence, the assumption that $\mathbb{S}$ is finite is false.

In the subsequent sections, we describe some new development on the existence of the Faddeev knots. In the next section, we show the existence of the Faddeev energy minimizers at the unit Hopf charge $Q = \pm 1$ and illustrate how to use the Substantial Inequality (1.5)–(1.6) and a suitable estimate on the upper bound on E_1 to arrive at a proof. In Section 3, we emphasize that relations given by (1.7) and (1.10) spell out a second kind of topological growth law which is seen to be fractionally-powered and universal in 3 dimensions and we discuss it in the context of other, well-known, topological and geometric growth laws in field theory. We also recall some recent studies on knot energies and knot invariants in knot theory community. In Section 4, we extend the Faddeev knot energy into general Hopf dimensions so that the configuration maps are from $\mathbb{R}^{4n-1}$ into S^{2n}. We are motivated from two considerations: First, in general dimensions, we will be able to achieve a deeper understanding on the fractional power in such topological growth law and single out its universal structures. Secondly, theoretical physics not only thrives in but also needs spaces of higher dimensions,[34,63,84] and a study of the knot energy of the Faddeev type in higher dimensions will be of interest. In Section 5, we present the fractionally-powered universal topological growth law in its most general form. In Section 6, we conclude by commenting on some future issues.

2. Faddeev Knots at Unit Charge

For the Faddeev Knot Problem, it has long been anticipated that the Faddeev energy (1.2) should attain its infimum at the unit charge $N = \pm 1$. Indeed, the numerical solutions of Battye and Sutcliffe[9,10] indicate that E_1 is the least positive energy point for the evaluated Faddeev energy for the

Hopf number from one to eight (more recent numerical work by Sutcliffe on the Faddeev knots of higher Hopf numbers is reported in[76]). Then, using the Substantial Inequality, we see that E_1 would be the least energy point among the entire Faddeev energy spectrum. Consequently, by Theorem 1.2, E_1 is attainable. However, a rigorous proof of this fact along such a line has been elusive because it is difficult to establish that E_1 is indeed the least positive energy point.

Using the classical Hopf map, Ward first estimated[78] that for the Faddeev energy (1.2) the energy E_1 has the upper estimate

$$E_1 \leq 32\sqrt{2}\pi^2. \tag{2.1}$$

For more details, see.[52]

Next, it can be shown[52] that the Vakulenko–Kapitanski lower bound (1.7) has the explicit form

$$E(\mathbf{n}) \geq 3^{3/8} 8\sqrt{2}\pi^2 |Q(\mathbf{n})|^{3/4}. \tag{2.2}$$

See also.[47,67] Combining (2.1), (2.2), and the Substantial Inequality (1.5)–(1.6), we have[52]

Theorem 2.1. *For the Faddeev energy (1.2), the energy $E_{\pm 1}$ is attainable.*

Here is a quick proof. Suppose that E_1 is not attainable. Then in the minimization process for E_1 concentration does not occur and there holds the nontrivial energy splitting in view of the substantial inequality: $E_1 \geq E_{N_1} + \cdots + E_{N_k}$, $1 = N_1 + \cdots + N_k, N_s \in \mathbb{Z} \setminus \{0\}, s = 1, \cdots, k$ with $k \geq 2$. Since each $E_{N_s} > 0$, we see from the fact $E_1 = E_{-1}$ that $N_s \neq \pm 1$ for $s = 1, \cdots, k$. Hence, one of the integers, $N_1, \cdots, N_k$, must be an odd number. Assume that N_1 is odd. Then $|N_1| \geq 3$. Of course, $|N_2| \geq 2$. Therefore we are led to

$$32\sqrt{2}\pi^2 \geq E_1 \geq E_{N_1} + E_{N_2} \geq 3^{3/8} 8\sqrt{2}\pi^2 (3^{3/4} + 2^{3/4}), \tag{2.3}$$

which is a contradiction and the proof of the theorem follows.

3. Growth Law Perspectives

We have seen that the fractionally-powered energy-topology growth law of the Faddeev model gives rise to a series of important consequences to the formation of knots and deserves refreshed close attention and study. In,[50] we showed that the growth law

$$E_N \sim |N|^{3/4} \tag{3.1}$$

for the Faddeev knot energy is universal in the sense that the topological growth factor $|N|^{3/4}$ may be proven to stay unaffected by the fine structure change of the energy. For example, when the L^2-gradient term in the Faddeev energy (1.2) is replaced by an L^p-gradient term so that the total energy is of the form

$$E(\mathbf{n}) = \int_{\mathbb{R}^3} \left\{ \sum_{j=1}^{3} |\partial_j \mathbf{n}|^p + \frac{1}{2} \sum_{j,k=1}^{3} |F_{jk}(\mathbf{n})|^2 \right\} \mathrm{d}x, \tag{3.2}$$

the asymptotic growth law (3.1) still holds provided that the power p satisfies $1 < p < 12/5$.

Note that the fractional-exponent topological growth law of the type (3.1) is uncommon in quantum field theory models. Indeed, most growth laws seen so far are linear, instead of being sublinear. As a comparison, it may be instructive to recall some well known problems.

Instantons. Consider an $SU(2)$-bundle over the standard 4-sphere (S^4, g) (g is the metric of S^4). The energy (action) functional governing an $su(2)$-valued gauge connection A is defined by

$$E(A) = -\int_{S^4} \mathrm{Tr}(F_A \wedge *F_A), \tag{3.3}$$

where $*$ is the Hodge dual induced from g and $F_A = \mathrm{d}A + A \wedge A$ is the curvature. One is interested in the global minimizers of (3.3) among the topological class that the associated second Chern or first Pontryagin invariant of the curvature is an integer,

$$c_2(F_A) = p_1(F_A) = -\frac{1}{8\pi^2} \int_{S^4} \mathrm{Tr}(F_A \wedge F_A) = N, \tag{3.4}$$

where $N \in \mathbb{Z}$. It is well known that for (3.3) subject to (3.4) there holds the following linear topological energy lower bound[1,8,65,82]

$$E(A) \geq 8\pi^2 |N|. \tag{3.5}$$

Monopoles. Use the above notation and consider an $SU(2)$-bundle over $\mathbb{R}^3$. Let ϕ be a scalar field which lies in the adjoint representation of $SU(2)$. The connection A induces the gauge-covariant derivative $D_A\phi = \mathrm{d}\phi + [A, \phi]$. The Yang–Mills–Higgs monopole energy may be written as

$$E(A,\phi) = \int_{\mathbb{R}^3} \left\{ -\frac{1}{2}\mathrm{Tr}(F_A \wedge *F_A) - \frac{1}{2}\mathrm{Tr}(D_A\phi \wedge *\overline{D_A\phi}) + *\frac{\lambda}{8}(|\phi|^2 - 1)^2 \right\} \tag{3.6}$$

where $\lambda \geq 0$ is a constant and ϕ obeys the boundary condition $|\phi(x)| \to 1$ as $|x| \to \infty$. Therefore, near infinity of $\mathbb{R}^3$, we may view ϕ as a map from S^2

into $SU(2)$ modulo $U(1)$. Since $SU(2)/U(1) \approx S^2$, ϕ may be represented by an element in the homotopy group $\pi_2(S^2) = \mathbb{Z}$, which is an integer. This integer, say N, is called the monopole number and can be represented by the integral

$$N = \frac{1}{4\pi} \int_{\mathbb{R}^3} \mathrm{Tr}(D_A\phi \wedge F_A). \tag{3.7}$$

When the monopole energy (3.6) is subject to the topological constraint (3.7), there holds the lower bound[43,65]

$$E(A, \phi) \geq 4\pi |N|, \tag{3.8}$$

which is again linear.

Vortices. The formulation is similar to that of monopoles. Let ξ be a complex line bundle over the Riemann surface S. Use u to denote a section $\xi \to S$. If A is a real-valued (Abelian) connection 1-form, then $D_A u = \mathrm{d}u - \mathrm{i}Au$ defines an induced connection and $F_A = \mathrm{d}A$ is the curvature 2-form. The Hamiltonian density $\mathcal{H}$ of the Abelian Higgs theory is written as

$$\mathcal{H}(u, A) = \frac{1}{2} * (F_A \wedge *F_A) + \frac{1}{2} * (D_A u \wedge *\overline{D_A u}) + \frac{\lambda}{8}(1 - |u|^2)^2, \tag{3.9}$$

where $\lambda > 0$. We are to find the global minimizers of the energy

$$E(u, A) = \int_S \mathcal{H}(u, A)\, \mathrm{d}V, \tag{3.10}$$

subject to the topological constraint

$$c_1(\xi) = \frac{1}{2\pi} \int_S F_A = N, \tag{3.11}$$

where $c_1(\xi)$ is the first Chern class of the line bundle and N is a given integer. In view of the procedure of Bogomol'nyi,[12] it can be shown that there holds the topological energy lower bound[43]

$$E(u, A) \geq \min\{1, \lambda\}\pi |N|. \tag{3.12}$$

Blackholes. Consider an isolated blackhole of mass $m_{\mathrm{BH}} > 0$ whose spacetime metric is known to be given by the Schwarzschild line element in terms of the spherical coordinates (θ, ϕ, ρ) as

$$\mathrm{d}s^2 = \left(1 - \frac{2Gm_{\mathrm{BH}}}{\rho}\right) \mathrm{d}t^2 - \left(1 - \frac{2Gm_{\mathrm{BH}}}{\rho}\right)^{-1} \mathrm{d}\rho^2 - \rho^2(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\phi^2), \tag{3.13}$$

where G is the Newton constant. It can be checked that the spatial slice at any fixed t has the property that its second fundamental form vanishes and that its ADM mass[7] is the same as the blackhole mass m_{BH}. In this case, the singular surface or the event horizon, Σ, of the blackhole is a sphere of radius $\rho_s = 2Gm_{\mathrm{BH}}$ whose surface area has the value

$$\mathrm{Area}(\Sigma) = 4\pi\rho_s^2 = 16\pi G^2 m_{\mathrm{BH}}^2. \tag{3.14}$$

The Penrose Conjecture[62] states that the total energy E of the spacetime, which is a more general concept than the ADM mass, is bounded from below by the total surface area of its apparent horizon Σ, which coincides with the event horizon in the case of a Schwarzschild blackhole, by

$$16\pi G^2 E^2 \geq \mathrm{Area}(\Sigma). \tag{3.15}$$

In the special case when the second fundamental form of the spatial slice vanishes, hence the total momentum is zero and the total gravitational energy reduces to the ADM mass, the relation (3.15) becomes

$$16\pi G^2 m_{\mathrm{ADM}}^2 \geq \mathrm{Area}(\Sigma), \tag{3.16}$$

which is referred to as the Riemannian Penrose Inequality, for which the lower bound may be saturated only in the Schwarzschild limit.[14–16,40,41]

These growth laws are geometrical rather than topological.

Ideal Knots. It has been an interesting question whether the energy infimum of a suitably defined knot energy evaluated over a given knot type may be used as a knot invariant. Indeed, Moffat[56] articulated to use the minimum knot energy as a new type of invariant for knots and links, further emphasized that any knot or link may be characterized by an "energy spectrum" – a set of positive real numbers determined solely by its topology, and proposed that the lowest energy provides a possible measure of knot or link complexity. Katritch *et al*[45] approached knot identification by considering the properties of specific geometric forms of knots which are defined as ideal so that for a knot with a given topology and assembled from a tube of uniform diameter, the ideal form is the geometrical configuration having the highest ratio of volume to surface area. Equivalently, this amounts to determining the shortest piece of tube that can be closed to form the knot. They reported their results of computer simulations showing a linear relationship between the length-to-diameter ratio, or the ropelength energy, and the (averaged) crossing number, of the knot and indicating the practicality of using ropelength energy to detect knot type. Buck[18] used the minimum ropelength energy of a knot to measure the complexity of the knot conformation and investigated the reported linear relationship between ropelength energy and the average crossing number of knots. He showed that a linear relationship cannot hold in general and the rope length required to tie an N-crossing knot or link varies at least between $N^{3/4}$ and N. Canterella *et al*[19] further showed that for any power $3/4 \leq p \leq 1$, there are infinite families of N-crossing knots and links which realize the minimum ropelength energy asymptotic relationship $E \sim N^p$ – that is, for each p, there are families of N-crossing knots and links whose minimum ropelength energy and p-powered crossing number ratio, E/N^p, remains bounded from below and above as $N \to \infty$. The common feature of these studies on the ideal or canonical conformations and complexity of knots and links is that they all originate from diagrammatic considerations of knotted space curves. Besides, it is often hard to obtain growth laws relating the knot energy and knot invariant in a sharp form. For example, for a link type with minimum crossing number N (topology) and minimum ropelength L_N (energy), the estimate

$$(4\pi N/11)^{3/4} \leq L_N \leq 24N^2 \tag{3.17}$$

was obtained in,[20] in which the left-hand side and right-hand side are quite far apart for large values of N. See[17,31,33,37,38,44,46,57,60,61,74] for other studies on various energetic and topological characteristics of diagrammatic knots.

4. Knot Energy in General Hopf Dimensions

Recall that the integral representation of the Hopf invariant by J. H. C. Whitehead[81] of the classical fibration $S^3 \to S^2$ can be extended to the general case of the fibration $S^{4n-1} \to S^{2n}$. More precisely, let $u : S^{4n-1} \to S^{2n}$ ($n \geq 1$) be a differentiable map. Then there is an integer representation of u in the homotopy group $\pi_{4n-1}(S^{2n})$, say $Q(u)$, called the generalized Hopf index of u, which has a similar integral representation as follows. Let Ω be a volume element of S^{2n} so that

$$|S^{2n}| \equiv \int_{S^{2n}} \Omega \tag{4.1}$$

is the total volume of S^{2n} and u^* the pullback map $\Lambda(S^{2n}) \to \Lambda(S^{4n-1})$ (a homorphism between the rings of differential forms). Since u^* commutes with d, we see that $du^*(\Omega) = 0$; since the de-Rham cohomology $H^{2n}(S^{4n-1}, \mathbb{R})$ is trivial, there is a $(2n-1)$-form v on S^{4n-1} so that $dv = u^*(\Omega)$. Of course, the normalized volume form $\tilde{\Omega} = |S^{2n}|^{-1}\Omega$ gives the unit volume and $\tilde{v} = |S^{2n}|^{-1}v$ satisfies $d\tilde{v} = u^*(\tilde{\Omega})$. Since $\tilde{\Omega}$ can be viewed also as an orientation class, $Q(u)$ may be represented as[36,42]

$$Q(u) = \int_{S^{4n-1}} \tilde{v} \wedge u^*(\tilde{\Omega}) = \frac{1}{|S^{2n}|^2} \int_{S^{4n-1}} v \wedge u^*(\Omega). \tag{4.2}$$

The conformal invariance of (4.2) enables us to come up with the Hopf invariant, $Q(u)$, for maps, u, from $\mathbb{R}^{4n-1}$ to S^{2n} which approach fixed directions at infinity, as

$$Q(u) = \frac{1}{|S^{2n}|^2} \int_{\mathbb{R}^{4n-1}} v \wedge u^*(\Omega), \quad \mathrm{d}v = u^*(\Omega). \tag{4.3}$$

With the above preparation, we introduce the generalized Faddeev knot energy over $\mathbb{R}^{4n-1}$ as

$$E(u) = \int_{\mathbb{R}^{4n-1}} \left\{ |\mathrm{d}u|^2 + \frac{1}{2}|u^*(\Omega)|^2 \right\} \mathrm{d}x, \tag{4.4}$$

and extend the Faddeev Knot Problem (1.4) into the form

$$E_N \equiv \inf\{E(u) \,|\, E(u) < \infty, Q(u) = N\}, \quad N \in \mathbb{Z}. \tag{4.5}$$

Theorem 4.1. *The generalized Faddeev energy (4.4) of a finite-energy map u and its Hopf invariant (4.3) are related by the inequality*

$$E(u) \geq C(n)|Q(u)|^{\frac{4n-1}{4n}}, \tag{4.6}$$

where the constant $C(n)$ in (4.6) has the explicit value

$$C(n) = 2(2n-1)(2[4n-3])^{-\frac{(4n-3)}{2(2n-1)}}(2n)^{\frac{1}{2(2n-1)}}(c_0|S^{2n}|^2)^{\frac{4n-1}{4n}}. \tag{4.7}$$

Here the constant c_0 in (4.7) is the best constant in the Sobolev inequality $c_0\|f\|_q \leq \|\nabla f\|_2$ over $\mathbb{R}^{4n-1}$ with q satisfying $1/q = 1/2 - 1/(4n-1) = (4n-3)/2(4n-1)$, given by the expression

$$c_0 = ([4n-1][4n-3])^{\frac{1}{2}} \left(\omega_{4n-1} \frac{\Gamma(2n-\frac{1}{2})\Gamma(2n+\frac{1}{2})}{\Gamma(4n-1)} \right)^{\frac{1}{(4n-1)}}, \tag{4.8}$$

with ω_m being the volume of the unit ball in $\mathbb{R}^m$.

Note that, when $n = 1$, we recover (2.2).

5. The Universal Growth Law

The importance of the growth laws relating knot energy and knot invariant prompts us to carry out a more thorough investigation. As a starting point, we consider the energy

$$E_p(u) = \int_{\mathbb{R}^{4n-1}} \left\{ |\mathrm{d}u|^p + \frac{1}{2}|u^*(\Omega)|^2 \right\} \mathrm{d}x, \tag{5.1}$$

which generalizes (4.4). Our first goal is to see how the exponent p affects the lower bound (4.6)–(4.7). We have

Theorem 5.1. *Suppose the exponent p in the knot energy (5.1) lies in the range*

$$1 < p < \frac{4n(4n-1)}{4n+1}. \tag{5.2}$$

Then there holds the universal fractionally-powered topological lower bound

$$E_p(u) \geq C(n,p)|Q(u)|^{\frac{4n-1}{4n}}, \tag{5.3}$$

where the positive constant $C(n,p)$ may be explicitly expressed as

$$(c_0|S^{2n}|^2)^{\frac{4n-1}{4n}}(2n)^{\frac{p}{2(4n-p)}}(4n-p)\cdot f(n,p), \tag{5.4}$$

where $f(n,p) = \left(\frac{4n}{(4n-1)(8n-p)-p(4n+1)}\right)^{\frac{(4n-1)(8n-p)-p(4n+1)}{8n(4n-p)}}$.

Note. (i) In the special case when $p = 2$, (5.4) reduces to (4.7), namely, $C(n,2) = C(n)$, as expected. (ii) The most restrictive range of p, as stated in (5.2), occurs at the bottom dimension $n = 1$. In this situation, we have $1 < p < 12/5$ as mentioned earlier. When n is larger, the range of p becomes bigger quickly. (iii) At the bottom dimension $n = 1$, an important choice for the L^p-gradient term is $p = 3$ which is based on a conformal invariance consideration. known to arise in the so-called Nicole type models in particle physics.[2,3,58,66,80] In,[6] the conformal invariance is designated directly on the Skyrme term. These cases are not covered in our range and deserve further investigation in 3 and higher dimensions.

See below.

Another, more surprising, property is that we can derive a topological upper bound of the form (1.10) in which the constant C_1 is independent of N but depends only on the details of the energy density when the knot energy is taken to be the most general form

$$E(u) = \int_{\mathbb{R}^{4n-1}} \mathcal{H}(\nabla u)\,\mathrm{d}x. \tag{5.5}$$

Here the energy density function $\mathcal{H}$ is only assumed to be continuous with respect to its arguments and satisfies the natural condition $\mathcal{H}(0) = 0$. We have

Theorem 5.2. *Let E be defined by (5.5). Then for any given integer N which may be realized as the value of the Hopf invariant, i.e., $Q(u) = N$ for some differentiable map $u : \mathbb{R}^{4n-1} \to S^{2n}$, and E_N defined as $E_N = \inf\{E(u)|E(u) < \infty, Q(u) = N\}$, we have the universal topological upper bound*

$$E_N \leq C|N|^{\frac{4n-1}{4n}}, \tag{5.6}$$

where $C > 0$ is a constant independent of N.

A proof of the above theorem can be found in[50] under the oversimplified assumption that the Hopf invariant may assume any integer values. In particular, our proof relied on using that the smallest positive Hopf number is 1. On the other hand, it is well known that the Hopf invariant[13,42] behaves rather differently in higher dimensions: (i) For $n = 1, 2, 4$, there are maps $S^{4n-1} \to S^{2n}$ of the Hopf invariant 1. In fact, there are maps with the Hopf invariant equal to any integer. (ii) Conversely, if there is a map $S^{4n-1} \to S^{2n}$ of the Hopf invariant 1, then $n = 1, 2, 4$. This statement is known as Theorem of Adams and Atiyah.[4,5,42] (iii) For any n, there is always a map $S^{4n-1} \to S^{2n}$ with the Hopf invariant equal to any even number. A modified complete proof of this theorem will appear elsewhere.

In summary, we have seen that for maps from $\mathbb{R}^{4n-1}$ into S^{2n} governed by the Faddeev type energy and stratified by the Hopf invariant assuming an integer value N, there holds the sharp universal growth law

$$E_N \sim N^{\frac{4n-1}{4n}}, \tag{5.7}$$

which has profound implication for knotted structures as energy minimizers to exist and is independent of the detailed properties of the energy functional.

6. Overlook

Of course, the ultimate goal of our study is to develop an existence theory for the Faddeev type knot problems.

Among these problems, an important and useful setting is when we consider the existence problem for maps from S^{4n-1} into S^{2n} (the compact-space setting). We can show that for any possible value N of the Hopf invariant Q there exists an energy minimizer among the constraint class $Q = N$.

The reason for the above rather strong statement to be true is quite simple: Since we are in a compact setting, the difficulty with failure of concentration disappears. The form of the integral representation of the Hopf invariant and the elliptic estimates enable us to show that the Hopf invariant is a preserved quantity in the limit of a minimizing sequence.

Although simple, this situation may be compared with the classical study on harmonic maps, especially the result known as the Theorem of Smith:[24,72,73] Every degree N class of the homotopy group $\pi_n(S^n) = \mathbb{Z}$ has a harmonic representative for $n \leq 7$. More precisely, for $n = 2$, the solutions are known explicitly and carry minimum Dirichlet (harmonic map) energy;[11] for $3 \leq n \leq 7$, the energy has infimum 0 which can easily be seen by a rescaling argument, and hence does not achieve its absolute minimum in any class of degree $N \neq 0$; for $n \geq 8$, the situation is not very well understood. See.[24] Generally speaking, minimization among a homotopy class is a difficult issue when there is a lack of a suitable integral representation for the class.

As described earlier, a crucial step in the proof that E_1 is attainable in the classical 3 dimensions amounts to show that the energy splitting does not occur at the unit charge $Q = 1$ by using the explicit energy lower bound obtained and a sufficiently good estimate for the energy E_1. Unlike in the Skyrme model case,[25,26,48,49] such an estimate requires some careful effort. In order to generalize this type of estimate to the higher dimensions over $\mathbb{R}^{4n-1}$ when $n = 2$ and $n = 4$ (say), recall that the classical Hopf map ($n = 1$) is defined via the fibration for which S^3 is viewed as lying in $\mathbb{C}^2$ and being "factored" out by unit complex multiplication to obtain $\mathbb{CP}^1 = S^2$, resulting in the fiber bundle $S^1 \to S^3 \to S^2$. For $n = 2$, we may view S^7 as lying in $\mathbb{H}^2$ (quaternionic 2-space) and factor it out by unit quaternion multiplication to get $\mathbb{HP}^1 = S^4$, resulting in the fiber bundle $S^3 \to S^7 \to S^4$. For $n = 4$, we may use octonions instead and obtain the fibration $S^7 \to S^{15} \to S^8$. These analogies indicate similar possible methods for estimating E_1 (when $n = 2$ and 4). Thus it may be hopeful to generalize Theorem 2.1 for dimension 3 to dimensions 7 and 15 and prove that the knot energy (4.4) for $n = 2$ and $n = 4$ attains its infimum at the unit Hopf charge $Q = \pm 1$. In the case when $n \neq 1, 2, 4$, the lowest positive Hopf charge is $Q = 2$ (the theorem of Adams and Atiyah[4,5]) and it is not known whether $E_{\pm 2}$ is attainable.

Theorem 5.2 suggests that the lower bound expressed in (5.3)–(5.4) may be valid for more general energy functionals than that given in (5.1). In particular, the exponent p may be allowed to assume other values than the confined range specified in (5.2). In this regard, an interesting model is the Nicole model already mentioned earlier governed by the conformally invariant energy

$$E_{\text{Nicole}}(u) = \int_{\mathbb{R}^{4n-1}} |\mathrm{d}u|^{4n-1}\,\mathrm{d}x, \quad u : \mathbb{R}^{4n-1} \to S^{2n}. \tag{6.1}$$

See[2,3,58] for some study when $n = 1$. It is clear that $p = 4n - 1$ does not satisfy (5.2) but it is not hard to establish the lower bound

$$E_{\text{Nicole}}(u) \geq C(n)|Q(u)|^{\frac{4n-1}{4n}} \tag{6.2}$$

for some constant $C(n) > 0$.

It will be interesting to know whether the energy (5.5) may be extended to contain a potential term which depends on the map u itself, rather than its derivatives, so that the model may be used for the situation with a broken vacuum symmetry. For this purpose, we consider the energy

$$E(u) = \int_{\mathbb{R}^{4n-1}} \{\mathcal{H}(\nabla u) + V(u)\}\, dx. \tag{6.3}$$

In applications, the potential energy density V scales differently from some terms in the kinetic energy density $\mathcal{H}$ in the integrand of (6.3) so that a critical point of (6.3) "partitions" the total energy in a form like

$$\begin{aligned} &c_1 \int_{\mathbb{R}^{4n-1}} V(u)\, dx + c_2 \int_{\mathbb{R}^{4n-1}} \{\text{some suitable terms in } \mathcal{H}(\nabla u)\}\, dx \\ &= c_3 \int_{\mathbb{R}^{4n-1}} \{\text{some other suitable terms in } \mathcal{H}(\nabla u)\}\, dx, \end{aligned} \tag{6.4}$$

where the nonnegative constants c_1, c_2, c_3 depends on n only so that $c_1, c_3 > 0$, the terms in $\mathcal{H}(\nabla u)$ on the left-hand side of (6.4) scale themselves as V, and the terms on the right-hand side of (6.4) scale themselves oppositely as V. In view of (6.4) and Theorem 5.2, we see that (5.6) is still valid for the more general energy (6.3).

It is seen that our progress made on the understanding of the topological growth laws described here has, to some extent, paved the road to the development of an existence theory for knotted structures governed by the Faddeev energy in general (Hopf) dimensions.

To end this talk, we would like to highlight and dramatize the universal relation (5.7) relating the knot energy and knot topology through the Faddeev quantum field theory as

$$\text{Physics} = \text{Topology},$$

which supplements the Einstein equation, Geometry = Physics, in General Relativity, and reiterates the theme of this conference.

Acknowledgments. Fanghua Lin was supported in part by NSF grant DMS–0201443. Yisong Yang was supported in part by NSF grant DMS–0406446.

References

1. A. Actor, Classical solutions of $SU(2)$ Yang–Mills theories, *Rev. Modern Phys.* **51** (1979) 461–525.

2. C Adam, J Sanchez-Guillen, Symmetries of generalized soliton models and submodels on target space S^2, *J. High Energy Phys.* 0501 (2005) 004.
3. C. Adam, J. Sanchez-Guillen, R.A. Vazquez, A. Wereszczynski, Investigation of the Nicole model, *J. Math. Phys.* **47** (2006) 052302.
4. J. F. Adams, On the non-existence of elements of Hopf invariant one, *Ann. Math.* **72** (1960) 20-104.
5. J. F. Adams and M. F. Atiyah, K-theory and the Hopf invariant, *Quarterly J. Math.* **17** (1966) 31-38.
6. H. Aratyn, L. A. Ferreira, and A. H. Zimerman, Exact static soliton solutions of (3+1)-dimensional integrable theory with nonzero Hopf numbers, *Phys. Rev. Lett.* **83** (1999) 1723–1726.
7. R. Arnowitt, S. Deser, and C. Misner, Coordinate invariance and energy expressions in general relativity, *Phys. Rev.* **122**, 997–1006 (1961).
8. M. F. Atiyah, V. G. Drinfeld, N. J. Hitchin, and Yu. I. Manin, Construction of instantons, *Phys. Lett.* A **65** (1978) 186–187.
9. R. A. Battye and P. M. Sutcliffe, Knots as stable solutions in a three-dimensional classical field theory, *Phys. Rev. Lett.* **81** (1998) 4798–4801.
10. R. A. Battye and P. M. Sutcliffe, Solitons, links and knots, *Proc. Roy. Soc.* A **455** (1999) 4305–4331.
11. A. A. Belavin and A. M. Polyakov, Metastable states of two-dimensional isotropic ferromagnets, *JETP Lett.* **22** (1975) 245–247.
12. E. B. Bogomol'nyi, The stability of classical solutions, *Sov. J. Nucl. Phys.* **24** (1976) 449–454.
13. R. Bott and L. W. Tu, *Differential Forms in Algebraic Topology*, Springer, New York, 1982.
14. H. L. Bray, Proof of the Riemannian Penrose inequality using the positive mass theorem, *J. Diff. Geom.* **59**, 177–267 (2001).
15. H. L. Bray, Black holes, geometric flows, and the Penrose inequality in general relativity, *Notices Amer. Math. Soc.* **49**, 1372–1381 (2002).
16. H. L. Bray and R. M. Schoen, Recent proofs of the Riemannian Penrose conjecture, *Current Developments in Mathematics*, Int. Press, Somerville, MA, 1999. pp. 1–36.
17. S. Bryson, M. H. Freedman, Z. X. He, and Z. H. Wang, Möbius invariance of knot energy, *Bull. Amer. Math. Soc.* (N.S.) **28** (1993) 99-103.
18. G. Buck, Four-thirds power law for knots and links, *Nature* **392** (1998) 238-239.
19. J. Cantarella, R. Kusner, and J. Sullivan, Tight knots deviate from linear relation, *Nature* **392** (1998) 237.
20. J. Cantarella, R. Kusner, and J. Sullivan, On the minimum ropelength of knots and links, *Invent. Math.* **150** (2002) 257–286.
21. S. S. Chern and J. Simons, Some cohomology classes in principal fiber bundles and their application to riemannian geometry, *Proc. Nat. Acad. Sci. U.S.A.* **68** (1971) 791–794.
22. S. S. Chern and J. Simons, Characteristic forms and geometric invariants, *Ann. of Math.* **99** (1974) 48–69.
23. Y. M. Cho, Monopoles and knots in Skyrme theory, *Phys. Rev. Lett.* **87**

(2001) 252001.
24. J. Eells and L. Lemaire, A report on harmonic maps, *Bull. Lond. Math. Soc.* **10** (1978) 1–68.
25. M. Esteban, A direct variational approach to Skyrme's model for meson fields, *Commun. Math. Phys.* **105** (1986) 571–591.
26. M. J. Esteban, A new setting for Skyrme's problem, *Variational Methods*, Birkhäuser, Boston, 1988. pp. 77–93.
27. L. Faddeev, Einstein and several contemporary tendencies in the theory of elementary particles, *Relativity, Quanta, and Cosmology*, vol. 1 (ed. M. Pantaleo and F. de Finis), 1979, pp. 247–266.
28. L. Faddeev, Knotted solitons, Proc. ICM2002, vol. 1, Beijing, August 2002, pp. 235–244.
29. L. Faddeev and A. J. Niemi, Stable knot-like structures in classical field theory, *Nature* **387** (1997) 58–61.
30. L. Faddeev and A. J. Niemi, Toroidal configurations as stable solitons, *Preprint*, hep-th/9705176.
31. M. H. Freedman, Z. X. He, and Z. H. Wang, Möbius energy of knots and unknots, *Ann. Math.* **139** (1994) 1–50.
32. T. Gisiger and M. B. Paranjape, Baby Skyrmion strings, *Phys. Lett.* B **384** (1996) 207–212.
33. O. Gonzalez and J. H. Maddocks, Global curvature, thickness, and the ideal shapes of knots, *Proc. Natl. Acad. Sci. USA* **96** (1999) 4769–4773.
34. M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory*, Volumes 1 and 2, Cambridge U. Press, 1987.
35. W. Greiner and J. A. Maruhn, *Nuclear Models*, Springer, Berlin, 1996.
36. W. Greub, S. Halperin, and R. Vanstone, *Connections, Curvature, and Cohomology*, Academic Press, New York and London, 1972.
37. M. Gromov, Homotopical effects of dilation, *J. Diff. Geom.* **13** (1978) 303–310.
38. M. Gromov, Filling Riemannian manifolds, *J. Diff. Geom.* **18** (1983) 1–147.
39. M. Hirayama, H. Yamakoshi, and Yamashita Estimation of the Lin-Yang bound of the least static energy of the Faddeev model, *Prog. Theor. Phys.* **116** (2006) 273–283.
40. G. Huisken and T. Ilmanen, The Riemannian Penrose inequality, *Int. Math. Res. Not.* **20**, 1045–1058 (1997).
41. G. Huisken and T. Ilmanen, The inverse mean curvature flow and the Riemannian Penrose inequality, *J. Diff. Geom.* **59**, 353–437 (2001).
42. D. Husemoller, *Fibre Bundles* (2nd ed.), Springer, New York, 1975.
43. A. Jaffe and C. H. Taubes, *Vortices and Monopoles* Birkhäuser, Boston, 1980.
44. E. J. Janse van Rensburg, A tutorial on knot energies, in *Physical and Numerical Models in Knot Theory*, pp. 19–44 (edited by J. A. Calvo *et al*), World Scientific, Singapore, 2005.
45. V. Katritch, J. Bednar, D. Michoud, R. G. Scharein, J. Dubochet, and A. Stasiak, Geometry and physics of knots, *Nature* **384** (1996) 142–145.
46. L. H. Kauffman, *Knots and Physics*, World Scientific, River Ridge, New Jersey, 2000.

47. A. Kundu and Yu P. Rybakov, Closed-vortex-type solitons with Hopf index, *J. Phys.* A Math. Gen. **15** (1982) 269–275.
48. F. H. Lin and Y. Yang, Existence of energy minimizers as stable knotted solitons in the Faddeev model, *Commun. Math. Phys.* **249** (2004) 273–303.
49. F. H. Lin and Y. Yang, Existence of 2D Skyrmions via concentration-compactness method, *Commun. Pure Appl. Math.* **57** (2004) 1332–1351.
50. F. H. Lin and Y. Yang, Static knot energy, Hopf charge, and universal growth law, *Nuclear Physics* B **747** (2006) 455–463.
51. F. H. Lin and Y. Yang, Faddeev knots, Skyrme solitons, and concentration-compactness, *Proceedings of the 23rd International Conference on Differential Geometry Methods in Theoretical Physics*, Nankai Tracts in Mathematics **10** (2006) pp. 313–322.
52. F. H. Lin and Y. Yang, Energy splitting, substantial inequality, and minimization for the Faddeev and Skyrme models, *Commun. Math. Phys.* **269** (2007) 137–152.
53. P. L. Lions, The concentration-compactness principle in the calculus of variations. Part I and Part II, *Ann. l'Insitut Henry Poincar'e - Analyse Non Lineaire* **1** (1984) 109–145; **1** (1984) 223–283.
54. P. L. Lions, The concentration-compactness principle in the calculus of variations. Part I and Part II, *Rev. Mat. Iber.* **1** (1985) 145–200; **2** (1985) 45–121.
55. V. G. Makhankov, Y. P. Rybakov, and V. I. Sanyuk, *The Skyrme Model*, Springer, Berlin and Heidelberg, 1993.
56. H. K. Moffat, The energy spectrum of knots and links, *Nature* **347** (1990) 367–369.
57. A. Nabutovsky, Non-recursive functions, knots "with thick ropes" and self-clenching "thick" hyperspheres, *Comm. Pure Appl. Math.* **48** (1995) 381–428.
58. D. A. Nicole, Solitons with non-vanishing Hopf index, *J. Phys. G: Nucl. Phys.* **4** (1978) 1363–1369.
59. A. J. Niemi, Hamiltonian approach to knotted solitons (a contributed chapter in[74]).
60. J. O'Hara, Energy of a knot, *Topology* **30** (1991) 241–247.
61. J. O'Hara, Family of energy functionals of knots, *Topology Appl.* **48** (1992) 147–161.
62. R. Penrose, Naked singularities, *Ann. New York Acad. Sci.* **224**, 125–134 (1973).
63. J. Polchinski, *String Theory*, Volumes 1 and 2, Cambridge U. Press, 1998.
64. M. K. Prasad and C. M. Sommerfield, Exact classical solutions for the 't Hooft monopole and the Julia–Zee dyon, *Phys. Rev. Lett.* **35** (1975) 760–762.
65. R. Rajaraman, *Solitons and Instantons*, North Holland, Amsterdam, 1982.
66. T. Riviere, Minimizing fibrations and p-harmonic maps in homotopy classes from S^3 to S^2, *Comm. Anal. Geom.* **6** (1998) 427–483.
67. S. V. Shabanov, On a low energy bound in a class of chiral field theories with solitons, *J. Math. Phys.* **43** (2002) 4127–4134.
68. T. H. R. Skyrme, A nonlinear field theory, *Proc. Roy. Soc.* A **260** (1961) 127–138.

69. T. H. R. Skyrme, Particle states of a quantized meson field, *Proc. Roy. Soc.* A **262** (1961) 237–245.
70. T. H. R. Skyrme, A unified field theory of mesons and baryons, *Nucl. Phys.* **31** (1962) 556–569.
71. T. H. R. Skyrme, The origins of Skyrmions, *Internat. J. Mod. Phys.* A **3** (1988) 2745–2751.
72. R. T. Smith, Harmonic mappings of spheres, *Bull. Amer. Math. Soc.* **78** (1972) 593–596.
73. R. T. Smith, Harmonic mappings of spheres, *Amer. J. Math.* **97** (1975) 364–385.
74. A. Stasiak, V. Katritch, and L. H. Kauffman (eds), *Ideal Knots*, World Scientific, Singapore, 1998.
75. D. W. Sumners, Lifting the curtain: using topology to probe the hidden action of enzymes, *Notices A. M. S.* **42** (1995) 528–537.
76. P. Sutcliffe, Knots in the Skyrme–Faddeev model, arXiv: 0705.1468.
77. A. F. Vakulenko and L. V. Kapitanski, Stability of solitons in S^2 nonlinear σ-model, *Sov. Phys. Dokl.* **24** (1979) 433–434.
78. R. S. Ward, Hopf solitons on S^3 and $\mathbb{R}^3$, *Nonlinearity* **12** (1999) 241–246.
79. R. S. Ward, private communication, 2005.
80. A. Wereszczynski, Toroidal solitons in Nicole-type models, *Eur. Phys. J.* C **41** (2005) 265-268
81. J. H. C. Whitehead, An expression of Hopf's invariant as an integral, *Proc. Nat. Acad. Sci. U. S. A.* **33** (1947) 117–123.
82. Y. Yang, *Solitons in Field Theory and Nonlinear Analysis*, Springer, New York, 2001.
83. I. Zahed and G. E. Brown, The Skyrme model, *Phys. Reports* **142** (1986) 1–102.
84. B. Zwiebach, *A First Course in String Theory*, Cambridge U. Press, 2004.

On Translation Invariant Symmetric Polynomials and Haldane's Conjecture

Jesse Liptrap

Department of Matheamtics
Unievrsity of California, Santa Barbara
Santa Barbara, CA 93106, USA
Email: jliptrap@math.ucsb.edu

We establish an isomorphism between the ring of translation invariant symmetric polynomials in n variables and the full polynomial ring in $n-1$ variables, over any field of characteristic 0. In addition, we give a counterexample to a conjecture of Haldane[3] regarding the structure of translation invariant symmetric polynomials. Our motivation is the fractional quantum Hall effect, where translation invariant (anti)symmetric complex polynomials in n variables characterize n-electron wavefunctions.

1. Introduction

A polynomial $p(z_1, \ldots, z_n)$ is *translation invariant* iff for all c we have

$$p(z_1 + c, \ldots, z_n + c) = p(z_1, \ldots, z_n)$$

Working over $\mathbb{C}$, such a polynomial might yield a quantum mechanical description of n particles in the plane. This is the nature of fractional quantum Hall states,[3] where in addition our polynomials must be symmetric or antisymmetric. Thus we are led to the study of translation invariant (anti)symmetric polynomials. Antisymmetric polynomials fortunately do not need special treatment, as they are merely symmetric polynomials multiplied by the Vandermonde determinant $\prod_{i<j}(z_i - z_j)$.

In this elementary note we present two results regarding the structure of translation invariant symmetric polynomials. Our first result is a simple description of the ring of all such polynomials. Our second result is a counterexample to Haldane's conjecture that every homogeneous translation invariant symmetric polynomial satisfies a certain physically convenient property. More precisely, to each symmetric polynomial p one associates a certain finite poset $B(p)$; Haldane conjectures that if p is homogeneous and translation invariant, then $B(p)$ has a maximum. We prove the conjecture for polynomials in fewer than four variables, indicate how to obtain counterexamples, and discuss whether a weakened version of the conjecture holds.

2. The ring of translation invariant symmetric polynomials

Let $R \subset F[z_1, \ldots, z_n]$ be the algebra of translation invariant symmetric polynomials over a field F of characteristic 0, and let $X^{S_n} \subset X = F[x_1, \ldots, x_n]$ be the algebra of symmetric polynomials in $x_1, \ldots, x_n$. Our theorem says that R written in center of mass coordinates becomes X^{S_n} modulo one degree of freedom.

Theorem 2.1. *The homomorphism $\sigma\colon X^{S_n} \to R$ given by $\sigma(x_i) = z_i - z_{avg}$ is a surjection with kernel $(x_1 + \cdots + x_n)$, where $z_{avg} = \frac{1}{n}\sum_{j=1}^{n} z_j$.*

An elementary proof of this theorem occupies section 4.

Since $\mathrm{char} F = 0$, any symmetric polynomial in X can be written uniquely as a polynomial in the power sum symmetric polynomials $x_1^k + \cdots + x_n^k$, where $1 \leq k \leq n$. In other words, the map

$$\theta\colon F[w_1, \ldots, w_n] \to X^{S_n}, \quad \theta(w_k) = x_1^k + \cdots + x_n^k$$

is an isomorphism of algebras. Note that we could define a different isomorphism θ using elementary symmetric polynomials or complete homogeneous symmetric polynomials. In any case, Theorem 2.1 implies that $\sigma\theta\colon F[w_1, \ldots, w_n] \to R$ is a surjection with kernel (w_1).

Corollary 2.1. *The map $\phi\colon F[w_2, \ldots, w_n] \to R$ given by*

$$\phi(w_k) = (z_1 - z_{avg})^k + \cdots + (z_n - z_{avg})^k$$

is an isomorphism of algebras.

Now let R^d be the vector space of all polynomials in R which are homogeneous of degree d. Since $\phi(w_k)$ is homogeneous of degree k, the isomorphism ϕ yields a basis for R^d, namely all

$$w_\lambda = \prod_{k=2}^{n} \phi(w_k)^{\lambda_k} \tag{1}$$

where λ is any partition of d in which the integer k appears λ_k times, with $\lambda_1 = 0$. Simon et al.[2] prove directly that the w_λ form a basis of R^d, whereas we have deduced this fact from the ring structure of R. Although[2] defines w_λ

using elementary symmetric polynomials rather than power sum symmetric polynomials, this difference is purely cosmetic. Since the dimension m_d of R^d equals the number of partitions of d into integers between 2 and n, it is easy to see that a generating function for m_d is given by

$$\sum_{d=0}^{\infty} m_d t^d = \prod_{s=2}^{n} \frac{1}{1-t^s}$$

where m_0 is defined to be 1.

Finally, we describe the vector space $A \subset Z$ of translation invariant antisymmetric polynomials. It is well known that any antisymmetric polynomial p can be written uniquely as Δq, where q is a symmetric polynomial and Δ is the Vandermonde determinant $\prod_{i<j}(z_i - z_j)$. Since Δ is translation invariant, we have $A = \Delta R$, which defines a vector space isomorphism from R to A. Note that Δ is homogeneous of degree $n(n-1)/2$.

3. Haldane's conjecture

Every symmetric polynomial is a unique linear combination of symmetrized monomials, which physicists like to call *boson occupation states*. We identify symmetrized monomials with multisets of natural numbers:

$$[l_1, \ldots, l_n] = \sum_{\sigma \in \mathrm{Sym}(n)} z_{\sigma(1)}^{l_1} \cdots z_{\sigma(n)}^{l_n}$$

For instance, the multiset $[5^1, 0^2] = [5, 0, 0]$ corresponds to the symmetrized monomial $2(z_1^5 + z_2^5 + z_3^5)$. *Squeezing* a symmetrized monomial $[l_1, \ldots, l_n]$ means decrementing l_i and incrementing l_j, for any pair of indices i, j such that $l_i > l_j + 1$. The set of symmetrized monomials becomes a poset under the *squeezing order*: for $a, b \in B(p)$, put $a \leq b$ iff a can be obtained from b by repeated squeezing. For a symmetric polynomial p, let $B(p)$ consist of every symmetrized monomial whose coefficient in p is nonzero. We view $B(p)$ as a poset under the squeezing order and refer to it as the *squeezing poset* of p.

Definition 3.1. A symmetric polynomial p is *Haldane* if $B(p)$ has a maximum.

Conjecture 3.1 (Haldane). *Every translation invariant homogeneous symmetric polynomial is Haldane.*

Remark 3.1. Since squeezing moves preserve homogeneous degree, every Haldane polynomial is homogeneous. Many homogeneous symmetric polynomials are not Haldane, such as $[3,3,0]+[4,1,1]$, but these might not be translation invariant.

Prop 3.1. Haldane's conjecture holds for polynomials of three or fewer variables.

Proof. Since R_1^d is empty for $d > 0$, Haldane's conjecture holds trivially for $n = 1$. For $n = 2$, every symmetrized monomial is of the form $[a, b]$ with $a + b = d$. Such symmetrized monomials are comparable with respect to the squeezing order, so Haldane's conjecture is automatic for $n = 2$.

The case $n = 3$ requires a bit more work. Define an algebra homomorphism $\tau : F[z_1, \ldots, z_n] \to F[z_1, \ldots, z_n, t]$ by

$$\tau(p)(z_1, \ldots, z_n, t) = p(z_1 + t, \ldots, z_n + t)$$

so that p is translation invariant iff $\tau(p) = p$. Define a family of linear maps $\tau_i : F[z_1, \ldots, z_n] \to F[z_1, \ldots, z_n]$ by

$$\tau(p) = \tau_0(p) + \tau_1(p)t + \ldots + \tau_d(p)t^d$$

so that p is translation invariant iff $\tau_i(p) = 0$ for all $i > 0$. It is easily checked that

$$\tau_1([a,b,c]) = a[a-1,b,c] + b[a,b-1,c] + c[a,b,c-1]$$

for all $a, b, c > 0$. Now suppose $[a, b, c]$ is a maximal element of the squeezing poset of some $p \in R_3^d$, with $a \geq b \geq c > 0$. Then $[a+1, b, c-1]$ and $[a, b+1, c-1]$ are not in $B(p)$. The above equation then implies that the coefficient of $[a, b, c]$ in p equals c times the coefficient of $[a, b, c-1]$ in $\tau_1(p)$. Thus the coefficient of $[a, b, c-1]$ in $\tau_1(p)$ is nonzero, contradicting our assumption that p is translation invariant. We conclude that every maximal element of $B(p)$ is of the form $[a, b, 0]$, where $a + b = d$. Since any two such elements are comparable, it follows that $B(p)$ has a maximum. Thus Haldane's conjecture holds in the case $n = 3$. $\square$

Any two symmetrized monomials written as weakly decreasing sequences of natural numbers can be compared lexicographically. The lexicographic order on symmetrized monomials linearizes the squeezing order. Let $R = R_n^d$ be the vector space of translation invariant symmetric homogeneous degree d polynomials in n variables, and let $L = L_n^d$ be the set of lexicographic maxima of squeezing posets of polynomials in R. Note that $|L_n^d| \leq \dim R_n^d$.

Lemma 3.1. *If L_n^d is not linearly ordered with respect to squeezing, then some polynomial in R_n^d is not Haldane.*

Proof. By the definition of L_n^d there exist $p_i \in R_n^d$ such that m_i is the lexicographic maximum of $B(p_i)$, where $i = 1, 2$. W.l.o.g. assume m_1 is lexicographically bigger than m_2, and let c_i be the coefficient of m_2 in p_i. Then for any scalar c, the polynomial $q = p_1 + cp_2 \in R_n^d$ has m_2 coefficient equal to $c_1 + cc_2$. Provided that $c \neq -c_1/c_2$, the squeezing poset of q contains both m_1 and m_2. Since m_1 is the lexicographic maximum of $B(q)$, and the lexicographic order refines the squeezing order, m_1 is a maximal element of $B(q)$ with respect to the squeezing order. Since m_1 and m_2 are incomparable, we conclude that $B(q)$ does not have a maximum with respect to the squeezing order, i.e. q is not Haldane. $\square$

It is a straightforward programming exercise to compute L_n^d using the basis for R_n^d given by formula 1. We find that L_4^d is linearly ordered with respect to squeezing for $d \leq 13$. Alas,

$$\begin{aligned} L_4^{14} = \{&[14,0,0,0],[12,2,0,0],[11,3,0,0],[10,4,0,0],\\ &[9,5,0,0],[8,6,0,0],[8,4,2,0],[7,7,0,0]\} \end{aligned}$$

is not linearly ordered because $[8,4,2,0]$ and $[7,7,0,0]$ are incomparable. Therefore by Lemma 3.1 some $q \in R_4^{14}$ is not Haldane. It is a straightforward programming exercise to construct such a q by following the proof of Lemma 3.1: one computes the basis of R_n^d given by equation 1 and solves two systems of linear equations. We get

$$\begin{aligned} q =& 3[8,4,2,0] - 3[8,4,1,1] - 3[8,3,3,0] + 6[8,3,2,1] - 3[8,2,2,2]\\ &+ 3[7,7,0,0] - 42[7,6,1,0] + 46[7,5,2,0] + 80[7,5,1,1] - 22[7,4,3,0]\\ &- 188[7,4,2,1] + 112[7,3,3,1] + 8[7,3,2,2] + 77[6,6,2,0] + 70[6,6,1,1]\\ &- 182[6,5,3,0] - 700[6,5,2,1] + 112[6,4,4,0] + 168[6,4,3,1] + 1078[6,4,2,2]\\ &- 728[6,3,3,2] + 5[5,5,4,0] + 1072[5,5,3,1] + 246[5,5,2,2] - 722[5,4,4,1]\\ &- 2976[5,4,3,2] + 1808[5,3,3,3] + 1805[4,4,4,2] - 1130[4,4,3,3] \end{aligned}$$

Prop 3.2. The polynomial q is a counterexample to Haldane's conjecture.

Proof. Observe that $[8,4,2,0]$ and $[7,7,0,0]$ are the maximal elements of $B(q)$, which is depicted in figure 3 (with arrows pointing from smaller to bigger elements). One checks by computer that q is translation invariant. Therefore, being symmetric and homogeneous, q is a counterexample to Haldane's conjecture. $\square$

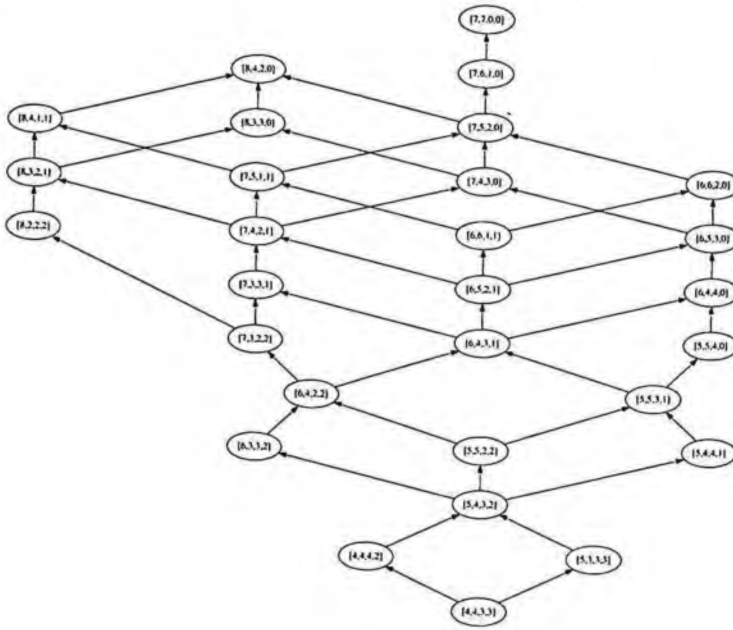


Fig. 3.1. Hasse diagram of $B(q)$

Remark 3.2. We hope that R_n^d nevertheless has a basis of Haldane polynomials. Computer evidence suggests that $|L_n^d| = \dim R_n^d$. Letting $L_n^d = \{l_1, \ldots, l_k\}$, we could then obtain a special basis $\{p_1, \ldots, p_k\}$ of R_n^d satisfying $B(p_i) \cap L_n^d = \{l_i\}$. Perhaps $\{p_1, \ldots, p_k\}$ would be a Haldane basis or could be used to construct one.

4. Proof of Theorem 2.1

Let $Z^{S_n} \subset Z = F[z_1, \ldots, z_n]$ be the algebra of symmetric polynomials, and let $Z^F \subset Z$ be the algebra of translation invariant polynomials, so that we are studying $R = Z^{S_n} \cap Z^F$. Let $X = F[x_1, \ldots, x_n]$, and let

$$\sigma \colon X \to Z^F, \quad \sigma(x_i) = z_i - z_{\text{avg}}$$

be the map from the main theorem. Our mission is to show that $\sigma(X^{S_n}) = R$ and $\ker \sigma = (x_{\text{avg}})$. Clearly σ is S_n-equivariant, implying $\sigma(X^{S_n}) \subseteq R$. For the reverse inclusion, given any $p(z_1, \ldots, z_n) \in R$, translation invariance yields

$$p(z_1, \ldots, z_n) = p(z_1 - z_{\text{avg}}, \ldots, z_n - z_{\text{avg}}) = \sigma(p(x_1, \ldots, x_n))$$

Thus $\sigma(X^{S_n}) = R$.

It remains only to show that $\ker\sigma = (x_{\text{avg}})$. We factor σ into two maps which are easier to study.

$$\begin{tikzcd} X \arrow[r, "\tau"'] \arrow[rr, bend left, "\sigma"] & Y \arrow[r, "\pi"'] & Z^F \end{tikzcd}$$

Let $Y = F[y_1, \dots, y_n]$, and define $\tau\colon X \to Y$, $\pi\colon Y \to Z^F$ by

$$\tau(x_i) = \frac{1}{n}\sum_{j=0}^{n-1}(n-1-j)y_{i+j}$$
$$\pi(y_i) = z_i - z_{i+1}$$

It is easy to check that the diagram commutes:

$$\begin{aligned} \pi\tau(x_i) &= \frac{1}{n}\sum_{j=0}^{n-1}(n-1-j)(z_{i+j} - z_{i+j+1}) \\ &= \frac{1}{n}\left((n-1)z_i - \sum_{j\neq i} z_j\right) \\ &= z_i - z_{\text{avg}} \end{aligned}$$

Thus $\sigma = \pi\tau$, so that we need only show $\tau^{-1}(\ker\pi) = (x_{\text{avg}})$.

First we check that τ is an isomorphism. Let $\hat{\tau} : Fx_1 + \cdots + Fx_n \to Fy_1 + \cdots + Fy_n$ be the linear map which extends to $\tau : X \to Y$, and let M be the matrix of $\hat{\tau}$ with respect to the evident bases. From the definition of τ we see that M is the $n \times n$ circulant matrix whose first column is given by the vector

$$v = \frac{1}{n}(n-1, n-2, \dots, 0)$$

Then $M^\top$ is the circulant matrix whose first row is v. Since $\text{char}F = 0$, the entries of v form a strictly decreasing sequence of nonnegative reals. Therefore by Theorem 3 of,[1] the matrix $M^\top$ is nonsingular. Hence M is nonsingular, showing that τ is an isomorphism via the following observation.

Observation 4.1. Suppose $f : F[a_1, \dots, a_n] \to F[b_1, \dots, b_n]$ is a homomorphism of polynomial rings which restricts to a linear map

$$\hat{f} : Fa_1 + \cdots + Fa_n \to Fb_1 + \cdots + Fb_n$$

Then f is an isomorphism of algebras iff $\hat{f}$ is an isomorphism of vector spaces.

Proof. Exercise in applying the universal property of polynomial rings. □

Now we complete the final step of the proof of Theorem 2.1, which is to show that $\tau^{-1}(\ker\pi) = (x_{\text{avg}})$. By the lemma below, $\ker\pi = (y_{\text{avg}})$. Since $\tau(x_{\text{avg}}) = \frac{(n-1)n}{2} y_{\text{avg}}$, and τ is an isomorphism, we have $\tau^{-1}((y_{\text{avg}})) = (x_{\text{avg}})$. Thus we are done, but for a final lemma.

Lemma 4.1. *$\ker\pi = (y_{avg})$, where $\pi : Y \to Z^F$ is defined by $\pi(y_i) = z_i - z_{i+1}$.*

Proof. Since $\pi(y_{\text{avg}}) = 0$, we have $(y_{\text{avg}}) \subseteq \ker\pi$. For the reverse inclusion, let $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ be the chain map depicted below

$$\begin{tikzcd}
0 \arrow[r] & (y_{\text{avg}}) \arrow[r] & Y \arrow[r, "\pi"] & Z^F \arrow[r] & 0 \\
0 \arrow[r] & (y_1) \arrow[r] \arrow[u, "\alpha_1"] & Y \arrow[r, "\pi'"'] \arrow[u, "\alpha_2"] & Y' \arrow[r] \arrow[u, "\alpha_3"] & 0
\end{tikzcd}$$

where $Y' = Y/(y_1) = F[y_2, \ldots, y_n]$, π' is the quotient projection, α_1 is the restriction of α_2 to (y_1), α_2 sends y_1 to y_{avg} while fixing the remaining variables, $\alpha_3(y_i) = z_i - z_{i+1}$, and the unlabelled nonzero maps are inclusions. We wish to show that the upper chain is exact. Since the lower chain is obviously exact, it suffices to check that α is a chain isomorphism.

Since the diagram commutes, it suffices to show that each component of α is an isomorphism. By Observation 4.1, α_2 is an isomorphism. Then so is α_1. For α_3, define $\beta : Y' \to Y'$ by $\beta(y_i) = y_i + \cdots + y_n$, $2 \leq i \leq n$. Again by Observation 4.1, β is an isomorphism. Thus it suffices to show that $\gamma = \alpha_3\beta$ is an isomorphism. Note that $\gamma : Y' \to Z^F$ sends y_i to $z_i - z_1$.

For any $p(z_1, \ldots, z_n) \in Z^F$, translation invariance yields

$$\begin{aligned}
p(z_1, \ldots, z_n) &= p(0, z_2 - z_1, \ldots, z_n - z_1) \\
&= p(0, \gamma(y_2), \ldots, \gamma(y_n)) \\
&= \gamma(p(0, y_2, \ldots, y_n))
\end{aligned}$$

Thus γ is surjective. Now suppose that

$$0 = \gamma(q(y_2, \ldots, y_n)) = q(z_2 - z_1, \ldots, z_n - z_1)$$

Then in particular $q(z_2 - z_1, \ldots, z_n - z_1)$ is zero modulo (z_1), implying that $q(z_2, \ldots, z_n) = 0$, showing that γ is injective. Thus γ is an isomorphism, proving the lemma. □

Acknowledgements

The author would like to thank Tobias Hagge, Thomas Howard, Joules Nahas, and Zhenghan Wang for useful discussions.

References

1. D. Geller, I. Kra, S. Popescu, and S. Simanca, On circulant matrices. www.math.sunysb.edu/sorin/eprints/circulant.pdf.
2. S. H. Simon, E. H. Rezayi, and N. R. Cooper, *Pseudopotentials for multi-particle interactions in the quantum Hall regime, Phys. Rev. B*, 75 (195306), 2007, cond-mat/0701260.
3. X.-G. Wen, and Z. Wang, A classification of symmetric polynomials of infinite variables – a construction of Abelian and non-Abelian quantum Hall states, arXiv:0801.3291.

Additional Gradings in Khovanov Homology

Vassily Olegovich Manturov

Moscow State Regional University, Russia
Email:vomanturov@yahoo.es

Dedicated to the memory of Xiao-Song Lin

The main goal of the present paper is to construct new invariants of knots with additional structure by adding new gradings to the Khovanov complex. The ideas given below work in the case of virtual knots, closed braids and some other cases of knots with additional structure. The source of our additional grading may be topological or combinatorial; it is axiomatised for many partial cases. As a byproduct, this leads to a complex which in some cases coincides (up to grading renormalisation) with the usual Khovanov complex and in some other cases with the Lee-Rasmussen complex.

The grading we are going to construct behaves well with respect to some generalisations of the Khovanov homology, e.g., Frobenius extensions. These new homology theories give sharper estimates for some knot characteristics, such as minimal crossing number, atom genus, slice genus, etc.

Our gradings generate a natural filtration on the usual Khovanov complex. There exists a spectral sequence starting with our homology and converging to the (graded version associated with) usual Khovanov homology.

1. Introduction

In the last few years, the invention of link homology (Khovanov homology, Ozsváth-Szabó invariants, and also papers by Rasmussen, Khovanov-Rozansky, Manolescu-Ozsváth-Sarkar-Thurston and others) brought many constructions from algebraic topology to knot theory and low-dimensional topology.

Such theories take a representative of a low-dimensional diagram (say, knot diagram or Heegaard diagram of a 3-manifold) and associate a certain complex with this. The homology of this complex is independent of the choice of representative, thus the homology defines an invariant of knot (resp., 3-manifold, knot in a manifold). Such algebraic complexes have different gradings, and this allows one to construct filtrations and spectral sequences. The behaviour of such spectral sequences is often closely connected to some topological property of knots/3-manifolds. A nice example is the work of Rasmussen[Ras] estimating the Seifert genus from Khovanov homology and giving a simple proof of Milnor's conjecture. Another example is the work by K.Kawamura,[Kaw] who sharpened the Morton-Franks-Williams estimate for the braiding index.

There is also an approach to estimate the minimal crossing number, see.[Ma8]

We shall mainly concentrate on Khovanov homology. In a sequence of recent papers, the author generalized Khovanov's theory from knots in $\mathbf{R}^3$ to knots in arbitrary thickened 2-surfaces (up to stabilisation, giving virtual knots (by Kauffman,[KaV]) or twisted knots (by Bourgoin,[Bou])).

Virtual knots, besides their "knottedness" also carry some information about the topology of the underlying surface.

Thus, it would be quite natural to take into account some topological data to introduce into Khovanov homology to make the latter stronger. This idea was also used in the paper by Asaeda, Przytycki, Sikora.[APS] We shall discuss the interaction between the present work and the work[APS] later.

The main idea goes as follows: Assume we have a well-defined complex made out of some knot diagram. Consider the chain spaces $\mathcal{C}(K)$ and the differential ∂. It turns out that in some cases it is possible to introduce a new grading gr that splits the differential ∂ into two parts $\partial = \partial' + \partial''$ in such a way that:

(1) ∂' preserves the new grading, whence ∂'' increases the new grading;
(2) (C, ∂') is a well-defined complex;
(3) the homology of (C, ∂') is invariant (under Reidemeister moves);
(4) there is a spectral sequence with $E^1 = H(C, \partial)$ converging to the (graded group associated with) usual Khovanov homology (the latter differential is taken with respect to ∂).

The new gradings have a topological nature: they correspond to cohomology classes.

This will guarantee that the complex is well defined. However, the gradings may be of any other (say, combinatorial) nature; the only thing we need is that for the Kauffman bracket states, there are two sorts of circles which behave nicely with respect to the Reidemeister moves.

The latter condition guarantees not only that the complex is well defined (that is, ∂' is indeed a differential) but also the invariance under

Reidemeister moves.

Varying this construction, one can construct further complexes with differentials of type $\partial' + \lambda\partial''$, where λ can be a coefficient or some operator.

The outline of the present paper is the following. In the next section, we define the Kauffman bracket, virtual knots, and Khovanov homology (with arbitrary coefficients) for virtual links and classical links (which actually constitute a proper part of virtual links).

Section 3 will be devoted to our main example: categorifying the Bourgoin invariant with the only one new grading corresponding to the first Stiefel-Whitney class for oriented thickenings of non-orientable surfaces.

The proof of the invariance theorem is given in section 4; it indeed contains all ingredients for the proof of the main theorem to follow in section 5, where we have multiple gradings of various types and present more examples.

Section 5 also contains the axiomatics for these new gradings and examples what they can be applied to: braids, cables, tangles, long knots etc.

Section 6 devoted is to a generalisation of Khovanov's Frobenius structure. From this point of view, one can think of Lee's homology as a partial case of Khovanov's Frobenius theory as well as our new theory. As a byproduct, we present yet another definition of the Khovanov theory where the usual gradings are treated from our "dotted grading viewpoint".

In section 7, we focus on gradings and filtrations. We discuss the Frobenius construction due to Khovanov, which is then followed by spectral sequences, and Lee-Rasmussen invariants.

Section 8 is devoted to applications of the theory constructed and generalisations of some classical constructions in this context

Section 9 is devoted to the discussion and open questions.

1.1. *Acknowledgments*

I am very grateful to O.Ya.Viro for various stimulating discussions. During the writing process of that paper, I also had very useful discussions with L.H.Kauffman, V.A.Vassiliev, M.Khovanov, J.H.Przytycki, A.T.Fomenko, V.G.Turaev.

A large part of this work was written during my stay at the Chern Institute of Mathematics, Nankai University, People's Republic of China. I express my gratitude to the Institute especially to Prof. Mo-Lin Ge for the nice creative atmosphere.

2. Preliminaries: Virtual knots, Kauffman bracket, atoms, and Khovanov homology

We think of knot diagrams[a] as a collection of (classical) crossings on the plane somehow connected by arcs, see Fig. 2.1.

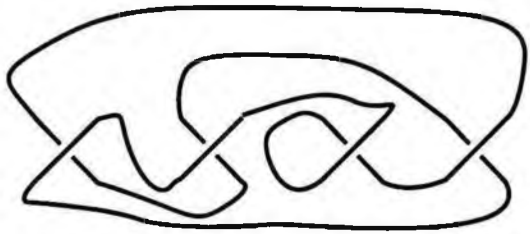


Fig. 2.1. A knot diagram

Knots are such diagrams modulo Reidemeister moves, but sometimes it happens that for a given setup of crossings we are unable to connect them by arcs in an appropriate way; this will lead to a diagram called *virtual* with "additional crossings" encircled, see Fig. 2.4.

2.1. *Atoms and Knots*

A four-valent planar graph Γ generates a natural checkerboard colouring of the plane by two colours (adjacent components of the complement $\mathbf{R}^2\backslash\Gamma$ have different colours).

This construction perfectly describes the role played by *alternating diagrams* of classical knots. Recall that a link diagram is *alternating* if while walking along any component we alternate over= and underpasses. Another definition of an alternating link diagram sounds as follows: fix a checkerboard colouring of the plane (one of the two possible colourings). Then, for every vertex the colour of the region corresponding to the angle swept by going from the overpass to the underpass in the counterclockwise direction is the same.

Thus, planar graphs with natural colourings somehow correspond to alternating diagrams of knots and links on the plane: starting with a graph and a colouring, we may fix the rule for making crossings: if two edges share a black angle, then the we decree the left one (with respect to the clockwise

[a]We refer both to knots and links by using a generic term "knot".

direction) to form an overcrossing, and the right one to be an undercrossing, see Fig. 2.2. Thus, colouring a couple of two opposite angles corresponds to a choice of a pair of opposite edges to form an overcrossing and vice versa.

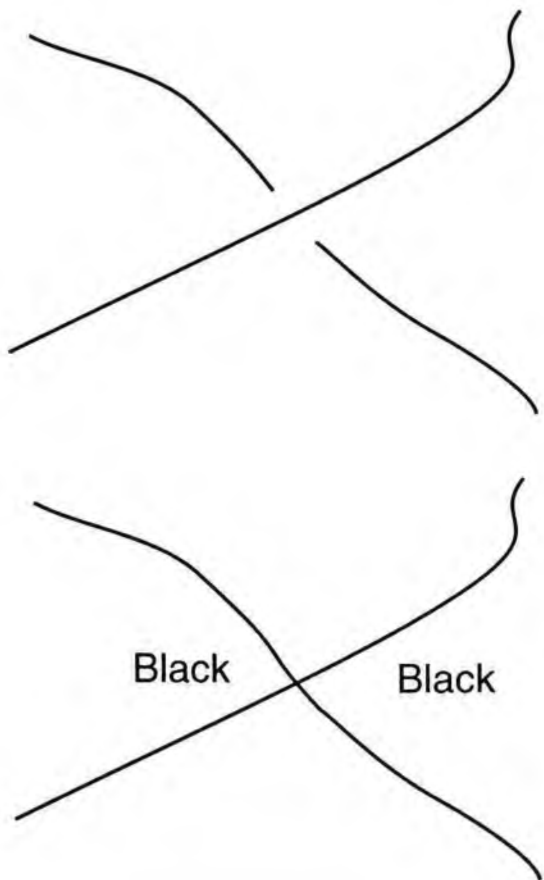


Fig. 2.2. A crossing corresponding to a vertex of an atom

Now, if we take an arbitrary link diagram and try to establish the colouring of angles according to the rule described above, we see that generally it is impossible unless the initial diagram is alternating: we can just get a region on the plane where colourings at two adjacent angles disagree. So, alternating diagrams perfectly match colourings of the 2-sphere (think of S^2 as a one-point compactification of $\mathbf{R}^2$). For an arbitrary link, we may try to take colours and attach cells to them in a way that the colours would agree, namely, the circuits for attaching two-cells are chosen to be those rotating circuits, where we always turn inside the angle of one colour.

This leads to the notion of *atom*. An *atom* is a pair (M, Γ) of a 2-manifold M and a graph Γ embedded M together with a colouring of $M \backslash \Gamma$ in a checkerboard manner. Here Γ is called the *frame* of the atom, whence by *genus* (resp., *Euler characteristic*) of the atom we mean that of M.

Note that the atom genus is also called the *Turaev genus*,[Tu]

Certainly, such a colouring exists if and only if Γ represents the trivial $\mathbf{Z}_2$ homology class in M.

Thus, gluing cells to some turning circuits on the diagram, we get an atom, where the shadow of the knot plays the role of the frame. Note that the structure of opposite half-edges on the plane coincides with that on the surface of the atom.

Now, we see that atoms on the sphere are precisely those corresponding to alternating link diagrams, whence non-alternating link diagrams lead to atoms on surfaces of a higher genus.

In some sense, the genus of the atom is a measure of how far a link diagram is from an alternating one, which leads to generalisations of the celebrated Kauffman-Murasugi theorem, see[Ia] and to some estimates concerning the Khovanov homology.[Ma8]

Having an atom, we may try to embed its frame in $\mathbf{R}^2$ in such a way that the structure of opposite half-edges at vertices is preserved. Then we can take the "black angle" structure of the atom to restore the crossings on the plane.

In[Ma0] it is proved that the link isotopy type does not depend on the particular choice of embedding of the frame into $\mathbf{R}^2$ with the structure of opposite edges preserved. The reason is that such embeddings are quite rigid.

The atoms whose frame is embeddable in the plane with opposite half-edge structure preserved are called *height* or vertical.

However, not all atoms can be obtained from some classical knots. Some abstract atoms may be quite complicated for its frame to be embeddable into $\mathbf{R}^2$ with the opposite half-edges structure preserved. However, if it is impossible to *immerse* a graph in $\mathbf{R}^2$, we may embed it by marking artifacts of the embedding (we assume the embedding to be generic) by small circles.

A *virtual diagram* is a four-valent graph on the plane with two types of crossings: classical or (for which we mark which pair of opposite edges form an overpass) and virtual (which are just marked by a circled crossing).

A *virtual link* is an equivalence class of virtual diagrams modulo generalised Reidemeister moves. The later consist of usual Reidemeister moves and the detour move. The detour move removes an arc virtually connecting some points A and B (that is, having no classical crossings inside) restores another connection between A and B with several virtual intersections and self-intersections, see Fig. 2.3.

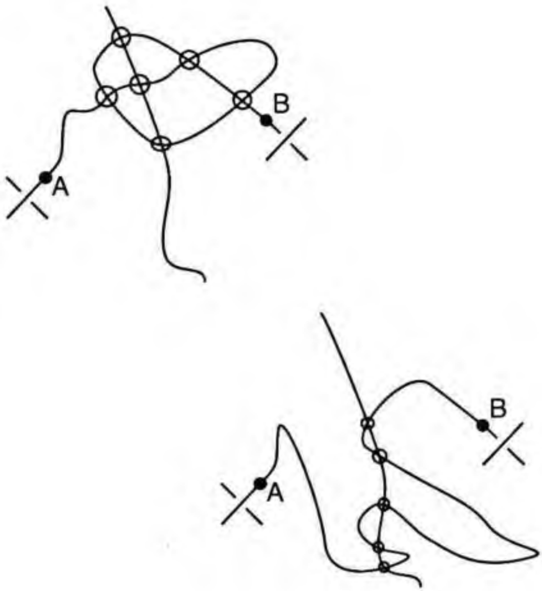


Fig. 2.3. The detour move

This move just means that it is inessential to indicate *which curves connect classical crossings*, it is important only to know how these crossings are paired.

Considering these diagrams modulo usual Reidemeister moves and the detour moves (see ahead), we get what are called *virtual knots*. The detour move is the move removing an arc (possibly, with self-intersections) containing only virtual crossing, and adding another arc connecting the same points elsewhere.

Virtual knots, being defined diagrammatically, have a topological interpretation. They correspond to knots in thickened surfaces $S_g \times I$ with fixed I-bundle structure (later we will also talk about oriented thickenings of non-orientable surfaces) up to stabilisations/destabilisations. Projecting S_g

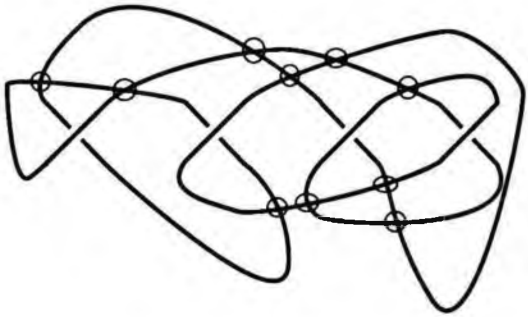
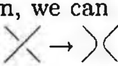
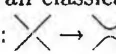


Fig. 2.4. A virtual diagram

to $\mathbf{R}^2$ (with the condition, however, that all neighbourhoods of crossings are projected with respect to the orientation, we get from a generic diagram on S_g a diagram on $\mathbf{R}^2$: besides the usual crossings arising naturally as projections of classical crossings, we get virtual crossings, which arise as artefacts of the projection: two strands lie in different places on S_g but they intersect on the plane because they are forced to do so.

Having a (virtual) knot diagram, we can smooth all classical crossings of it in the following two ways: A :  and B : .

Thus, for a diagram L with n classical crossings we have 2^n *states*. Every state is a way of smoothing all (classical) crossings. Enumerate all classical crossings by $1, \ldots, n$. Then the states can be regarded as vertices of the discrete cube $\{0, 1\}^n$, where 0 and 1 correspond to the A-smoothing and the B-smoothing, respectively. In each state we have a collection of circles representing an unlink. We call this cube *the state cube* of the diagram L.

Then any for any state s we have its height $\beta(s)$ being the number of crossings smoothed negatively, $\alpha(s) = n - \beta(s)$ being the number of crossings smoothed positively, and the number γ of closed circles.

Then the Kauffman bracket is defined as

$$\sum_s a^{\alpha(s)-\beta(s)}(-a^2 - a^{-2})^{\gamma(s)-1} \tag{2.1}$$

This bracket is invariant under all Reidemeister moves except for the first one.

The normalisation $X(K) = (-a)^{-3w(K)}\langle K \rangle$, where w is the writhe number, leads to the definition of the Jones polynomial.

The Kauffman bracket satisfies the usual relation

$$\langle \times \rangle = a\langle \asymp \rangle + a^{-1}\langle)(\rangle \tag{2.2}$$

After a little variable change and renormalisation, the Kauffman bracket can be rewritten in the following form:

$$\langle \times \rangle = \langle \asymp \rangle - q\langle)(\rangle \tag{2.3}$$

with $\langle \bigcirc \rangle = (q + q^{-1})$.

Here we consider bigraded complexes $\mathcal{C}^{ij}$ with *height* (homological grading) i and *quantum grading* j; the differential preserves the quantum grading and increases the height by 1. The height and grading shift operations are defined as $(\mathcal{C}[k]\{l\})^{ij} = \mathcal{C}[i-k]\{j-l\}$.

This form is used as the starting point for the Khovanov homology. Namely, we regard the factors $(q+q^{-1})$ as graded dimensions of the module $V = \{1, X\}, \deg 1 = 1, \deg X = -1$ over some ring R, and the height $\beta(s)$ plays the role of homological dimension. Then, if we define the chain space $[[K]]_k$ of homological dimension k to be the direct sum over all vertices of $\beta = k$ of $V^{\gamma(s)}\{k\}$ (here $\{\cdot\}$ is the quantum grading shift), then the alternating sum of graded dimensions of $[[K]]_k$, is precisely equal to the (modified) Kauffman bracket.

Thus, if we define a differential on $[[K]]$ preserving the grading and increasing the homological dimension by 1, the Euler characteristic of that complex would be precisely the Kauffman bracket.

Remark 2.1. Later on, we shall not care about the normalisation of the complexes by degree and height shifts to make their homology invariant under the Reidemeister moves. It is done exactly as in.[Kh]

We have defined the *state cube* consisting of circles and carrying no information how these circles interact. Turning to Khovanov homology, we shall deal with the same cube remembering the information about the circle bifurcation. Later on, we refer to it as a *bifurcation cube.*

The chain spaces of the complex are well defined. However, the problem of finding a differential ∂ in the general case of virtual knots, is not very easy. To define the differential, we have to pay attention to different isomorphism classes of the chain space identified by using some local bases.

The differential acts on the chain space as follows: it takes a chain corresponding to a certain vertex of the bifurcation cube to some chains corresponding to all adjacent vertices with greater homological degree. That is, the differential is a sum of *partial differentials*, each partial differential acts along an edge of the cube. Every partial differential corresponds to some direction and is associated with some classical crossing of the diagram.

With each circle of each state, we associate the tensor power of the space V of graded dimension $q + q^{-1}$, however, with no prefixed basis. With a collection of circles, we shall associate the *exterior power* of this space, as follows. With each state s of height b, we associate a basis consisting of $2^{\gamma(s)}$ chains. Now, we order the circles in the state s arbitrarily, fix an arbitrary orientation on them and associate with each such circle either 1 or X. With any such choice, consisting of a state, an ordering of oriented circles and a set of elements 1 and X, we associate a chain of the complex. We can also associate elements ± 1 or $\pm X$ with any circle, which also defines a chain of our complex; this chain differs from the corresponding chain with 1 and X by a corresponding sign. Furthermore, we identify the chains according to the following rule: the orientation change for one circle leads to a sign change of a chain if this circle is marked by $\pm X$ and does not change sign if the circle is marked by ± 1; the permutation of circles multiplies the chain by the sign of corresponding permutation. This would correspond to taking exterior product of vector spaces (graded modules) instead of their symmetric product.

Then for a state with l circles, we get a vector space (module) of dimension 2^l. All these chains have homological dimension b. We set the grading of these chains equals b plus the number of circles marked by ± 1 minus the number of circles marked by $\pm X$.

Let us now define the partial differentials of our complex. First, we think of each classical crossing so that its edges are oriented upwards, as in Fig. 2.5, upper right picture.

Choose a certain state of a virtual link diagram $L \subset \mathcal{M}$. Choose a classical crossing U of L. We say that in a state s a state circle γ is incident to a classical crossing X if at least one of the two local parts of smoothed crossing X belongs to γ. Consider all circles γ incident to U. Fix some orientation of these circles according to the orientation of the edge emanating in the upward-right direction and opposite to the orientation of the edge coming from the bottom left, see Fig. 2.5. Such an orientation is well defined except for the case when one edge corresponding to a vertex of the cube, takes one circle to one circle. In such situation, we shall not define the local basis $\{1, X\}$; we set the partial differential corresponding to the edge, to be zero.

In the other situations, the edge of the cube corresponding to the partial differential either increases or decreases the number of circles. This means

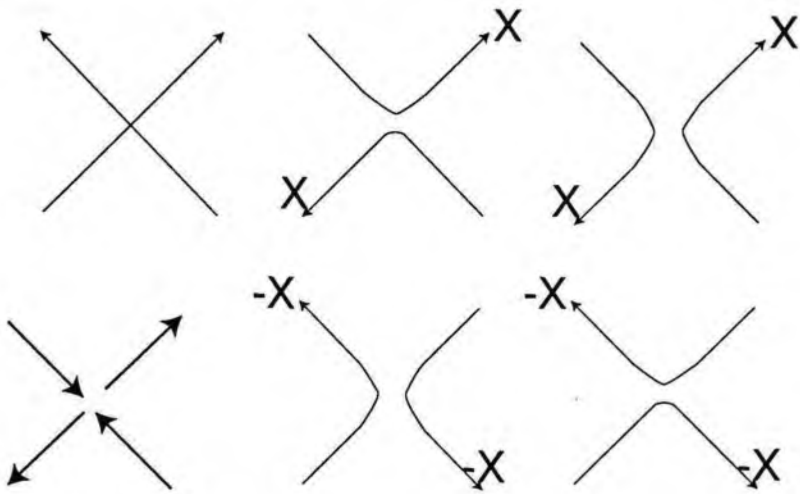


Fig. 2.5. Setting the local basis for a crossing

that at the corresponding crossing the local bifurcation either takes two circles into one or takes one circle into two. If we deal with two circles incident to a crossing from opposite signs, we order them in such a way that the upper (resp., left) one is the first one; the lower (resp., right) one is the second; here the notions "left, right, upper, lower" are chosen according to the rule for identifying the crossing neighbourhood with Fig. 2.5. Furthermore, for defining the partial differentials of types m and Δ (which correspond to decreasing/increasing the number of circles by one) we assume that the circles we deal with are in the very initial poisitions in our ordered tensor product; this can always be achieved by a preliminary permutation, which, possibly leads to a sign change. Now, let us define the partial differential locally according to the prescribed choice of generators at crossings and the prescribed ordering.

Now, we describe the partial differentials ∂' from[Ma6] without new gradings. If we set $\Delta(1) = 1_1 \wedge X_2 + X_1 \wedge 1_2; \Delta(X) = X_1 \wedge X_2$ and $m(1_1 \wedge 1_2) = 1; m(X_1 \wedge 1_2) = m(1_1 \wedge X_2) = X; m(X_1 \wedge X_2) = 0$, define the partial differential ∂' according to the rule $\partial'(\alpha \wedge \beta) = m(\alpha) \wedge \beta$ (in the case we deal with a $2 \to 1$-buifurcation, where α denotes the first two circles α) or $\partial'(\alpha \wedge \beta) = \Delta(\alpha) \wedge \beta$ (when one circle marked by α bifurcates to two ones); here by β we mean an ordered set of oriented circles, not incident to the given crossings; the marks on these circles ± 1 and $\pm X$ are given.

Theorem 2.1.[Ma6] *$[[K]]$ is a well-defined complex with respect to ∂; after a small grading shift and a height shift, the homology is invariant under generalised Reidemeister moves.*

Later, when we have new gradings, the differential will be defined just by projecting this differential to the grading-preserving subspace, namely, $\tilde{\partial}'\alpha = \mathrm{pr}_{\deg=\deg\alpha}\partial'\alpha$, where $\mathrm{pr}_{\deg=\deg\alpha}$ is the projection to the subspace having all additional gradings the same as α. After all, we shall define ∂ as the sum of partial differentials $\tilde{\partial}'$. We will get a set of graded groups Kh'_H with differential ∂. This differential increases the height (homological grading), preserves the grading, and does not change the additional gradings.

Remark 2.2. The homology theory described above is initially constructed out of planar diagrams; thus, it represents a homology theory for links in thickened surfaces *modulo stabilisation*; that is, this homology theory "does not feel" removable handles. However, when we impose new gradings, we will have to fix the thickened surface, since we will deal with its homology groups. The homology of the new complex to be constructed for such thickened surfaces, frankly speaking, would not be a virtual link invariant. It would rather be an obstruction for links in thickened surfaces to decrease the underlying genus of the corresponding surface.

2.2. *Usual Khovanov homology*

For the case of classical knot theory (and also some parts of virtual knot theory) the above setup is actually not needed for constructing Khovanov homology. One can get the chain spaces generated by tensor powers of V with appropriate grading and degree shifts, as it was done in the original Khovanov paper.[Kh] Namely, one takes just the symmetric tensor power $V^{\otimes k}$ for a vertex of a cube with k circles in the corresponding state. One also need not care about signs: the type-X generators are chosen once forever. Then it allows to construct partial differentials just by using some concrete formulae for Δ and m. The main difficulty we had to overcome was the case of $1 \to 1$-type partial differentials. If no such $1 \to 1$-bifurcations occur then the original construction works straightforwardly. Namely, after splicing some minus signs, these formulae lead to a well defined complex whose homology is the usual Khovanov homology.

3. Bourgoin's twisted knots. Additional gradings

Assume for some category (knots, virtual knots, braids, tangles) we have a well-defined Kauffman bracket. That is, we have a set of (classical) crossings, which can be smoothed so that the formula (2.3) can be applied.

Consider the following generalisation of virtual knots (proposed by Mario Bourgoin, see[Bou]).

We consider knots in oriented thickenings of 2-surfaces, the latter not necessarily orientable. Namely, we take a 2-surface M and fix the I-bundle $\mathcal{M}$ over M which is oriented as a total fibration space, and keep both the orientation and the I-bundle structure fixed.

We consider knots and links in such surfaces up to isotopy and stabilisation/destabilisation and refer to them as *twisted links*. Virtual links constitute a proper part of twisted links.[Bou]

Note that this theory encloses as a partial case the theory of knots in $\mathbb{R}P^3$, since $\mathbb{R}P^3\backslash\{*\}$ is nothing but the oriented thickening of $\mathbb{R}P^2$.

Any link in $\mathcal{M}$ has a projection to the base space, the latter being a four-valent graph.

Since the space $\mathcal{M}$ is orientable (and even oriented), there is a canonical way for defining the A-smoothing and the B-smoothing with respect to the orientation. Thus, the formula (2.3) gives a well-defined Kauffman bracket for such objects, which turns out to be invariant; the proof is standard, see, e.g.[Ma3]

Moreover, **the approach described in the previous section gives a well-defined Khovanov homology theory**. To this end, we have to establish the chain space and the differentials.

Fix a cell decomposition of M with exactly one 2-cell C and choose a canonical "upward" direction for C. Then we can treat every crossing as a classical one, that is, identify its neighbourhood with the local picture shown in Fig. 2.5.

This allows to define $[[K]]$ literally as above, and we get the following

Theorem 3.1. *For twisted knots the complex $[[K]]$ is a well-defined complex with respect to ∂'; after a small grading shift and a height shift, the homology is invariant under isotopy (the orientation of the ambient space remains fixed together with the I-bundle structure); the differential ∂' increases the homological grading by* 1 *and preserves the quantum grading.*

As shown in,[Ma6] the homology of this complex does not depend on the choice of C and the upward orientation.

We should mention, that there have been a lot of generalisations of the

Kauffman bracket, see e.g. Kauffman-Dye,[DK2] Manturov,[Ma9] Miyazawa.[Miy]

Each of these generalisations introduces something new to the formula for the Kauffman bracket of either topological or combinatorial nature.

Bourgoin proposed the following generalization of the Kauffman bracket for such surfaces.

$$\sum_s a^{\alpha(s)-\beta(s)} M^{\gamma''(s)}(-a^2-a^{-2})^{\gamma'(s)} \tag{3.1}$$

where γ' and γ'' correspond to the number of orienting/non-orienting circles in the state s, respectively.

The goal of the present section is to describe how to categorify this invariant and then see *which further examples will fit into the construction.*

In the Khovanov setup, we had $(q+q^{-1})$ instead of $(-a^2-a^{-2})$. What should we have instead of M?

What should be the vector space categorifying this variable. As can be seen from Khovanov's algebraic reasonings, see,[Kh2] the space corresponding to one circle should be two-dimensional.

To preserve the similarity with the initial picture, it is convenient to make one generator (1) of this space having quantum grading +1 and the other one (which might be X or $-X$) having quantum grading to -1.

This is the point where *new* gradings come into play: with every non-orienting circle in a Kauffman state, we associate the space of graded dimension $qg^{-1}+q^{-1}g$, where g corresponds to the new grading. At the uncategorified level, this just means $M = qg^{-1}+q^{-1}g$, and thus we lose no information.

At the categorified level, this means that we introduce a new grading for the chain spaces: for every non-orienting loop we associate a Z-grading equal to 1 if this loop is marked by X and -1 if this loop is marked by 1. For orienting loops, we have no new gradings.

Now, let us define the new grading (g-grading) for the complex $[[K]]$ as the sum of all new gradings over all non-orienting circles.

Denote the obtained chain space by $[[K]]_g$; this is actually nothing but $[[K]]$ with new grading imposed.

Notation. Further on, we shall mark all labels belonging to non-orienting circles by a point, that is, we write $\dot{1}$ and $\dot{X}$ for labels 1 and X on non-orienting circles.

Here we give an example how one smoothing with dots gets reconstructed into another smoothing; we put dots over some circles which correspond to "non-orienting" curves, see Fig. 3.1.

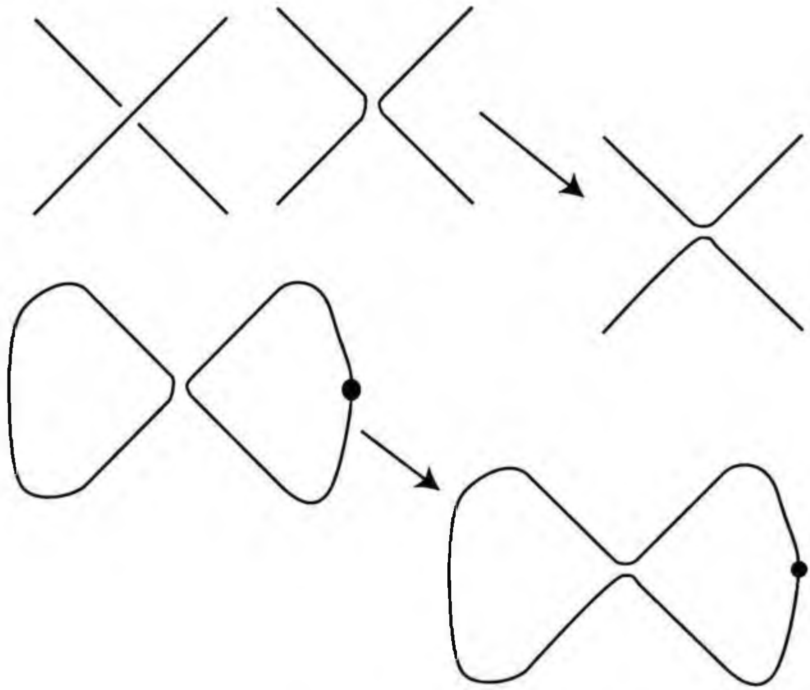


Fig. 3.1. Dotted circles

Let us look how the differentials in $[[K]]$ behave with respect to the new grading g. It is easy to see that

Lemma 3.1. *The differential ∂ can be uniquely represented as $\partial' + \partial''$, where ∂' preserves the new grading, and ∂'' increases the new grading by 2.*

Indeed, one can check all m-type and Δ-type partial differentials, and see that $\dot{1} \wedge \dot{1} \to 1$, $\dot{1} \wedge X \to \dot{X}$, $X \to \dot{X} \wedge \dot{X}$ are all increasing the grading by 2, whence the partial differential $\dot{1} \to X \wedge \dot{1} + 1 \wedge \dot{X}$ splits into two parts, where the first one preserves the new grading, and the second one increases that by 2.

From Lemma 3.1 we get

Lemma 3.2. $[[K]]_g$ *is a well defined triply graded complex with respect to the differential ∂'.*

Proof. Indeed, (∂'^2) is just the projection of $\partial^2 = 0$ to the grading-preserving subspace. □

Luckily, it turns out that the homology of $[[K]]_g$ is invariant after the same grading and degree shift (for old gradings) as in the usual case of classical knots ([Kh]) or virtual knots with oriented atoms.[Ma3] We shall show this more generally in the next section.

4. Additional grading: the general case

The goal of our section is the following. Assume we have a space of knots (braids, tangles, etc.) with a well-defined Kauffman bracket and Khovanov homology. We wish to mark some circles in Kauffman's states by dots (analogously to non-orienting cirlces in Bourgoin's case) thus defining the new "dotted gradings": the dotted grading for a chain in a given state is defined as the number of all $\dot{X}$ minus the number of all $\dot{1}$. Then we split the usual Khovanov differential ∂ into two parts: the one ∂' preserving the dotted grading and the one ∂'' changing the dotted grading.

What are the properties this dotting should satisfy if we want the grading to satisfy the following:

(1) The complex $[[K]]_g$ is well defined;
(2) Its homology (after some height and degree shift) is invariant under isotopy (combinatorial equivalence, Reidemeister moves).

The answer to the first question is easy: we just need that ∂'' either always increase the new grading or always decrease the new grading. Then it will guarantee $\partial'^2 = 0$.

But if we want the dots on circles to behave just as in the case of Bourgoin so that the rules for multiplication and comultiplication (with respect to the new grading) are:

$$m(1 \wedge 1) = 1; m(1 \wedge X) = X; m(X \wedge 1) = X; m(X \wedge X) = 0$$

$$m(\dot{1} \wedge 1) = \dot{1}; m(\dot{1} \wedge X) = 0; m(\dot{X} \wedge 1) = \dot{X}; m(\dot{X} \wedge X) = 0$$

$$m(1 \wedge \dot{1}) = \dot{1}; m(1 \wedge \dot{X}) = \dot{X}; m(X \wedge \dot{1}) = 0; m(X \wedge \dot{X}) = 0$$

$$m(\dot{1} \wedge \dot{1}) = 0; m(\dot{1} \wedge \dot{X}) = X; m(\dot{X} \wedge \dot{1}) = X; m(\dot{X} \wedge \dot{X}) = 0$$

and

$$\Delta(1) = 1 \wedge X + X \wedge 1$$

or

$$\Delta(1) = \dot{1} \wedge \dot{X} + dX \wedge 1$$

(depending on whether the output circles are dotted)

$$\Delta(X) = X \wedge X$$

or (when both output circles are dotted)

$$\Delta(X) = 0.$$

$$\Delta(\dot{1}) = \dot{1} \wedge X$$

or

$$\Delta(\dot{1}) = \dot{X} \wedge \dot{1}$$

(depending on which of the two output circles is dotted)

$$\Delta(\dot{X}) = X \wedge \dot{X}$$

or

$$\Delta(X) = \dot{X} \wedge X$$

(depending on which of the two output circles is dotted).

The operators m and Δ above are just as before (in the categorification of Bourgoin's invariant), however, with the reasons for putting dots completely forgotten.

Nevertheless, to have precisely this dotting, we need that *the dotting of circles is additive modulo* $\mathbf{Z}_2$, that is, if we have a $2 \to 1$ bifurcation, then the number of dots for the two circles is congruent modulo 2 to the number of dots for the one circle (analogously for $1 \to 2$-bifurcations). We also require that this dotting is preserved under $1 \to 1$-bifurcations, that is analogous to the fact that if a surgery transforms one circle to one circle then this circle should necessarily be unorienting both before and after the surgery.

The conditions above is enough for the complex $[[K]]_g$ to be well defined.

Now, in order to have the invariance under the Reidemeister moves, we have to restore the proof picture of Khovanov (or of[Ma6]).

The invariance under the first Reidemeister move is based on the following two which should held when adding a small curl:

(1) the mapping Δ is injective
(2) the mapping m is surjective.

In fact, the last two conditions hold when the small circle is not dotted. Indeed, consider the complex

$$[[\times]] = \left([[\times]] \xrightarrow{\Delta} [[\times]]\{1\}\right). \tag{4.1}$$

The usual argument goes as follows: the complex in the right hand side contains an m-type partial differential, which is surjective. Thus, the complex $[[\times]]$ is killed, and what remains from $[[\times]]$ is precisely (after a suitable normalisation) the homology of $[[\times]]$.

But Δ is injective because for any $l \in 1, X$ we have $\Delta(l) = l \wedge X + \langle somemess\rangle$, where the second term X in $l \wedge X$ corresponds to the small circle.

But in our situation with dotted circles, this happens only if *the small circle is not dotted.* But if the small circle is dotted, it would lead, say, to $\Delta : X \to 0$, because $\dot{X} \wedge \dot{X}$ has another dotted grading (greater by 2 than the grading of X).

An analogous situation happens with

$$[[\times]] = \left([[\times]] \xrightarrow{m} [[\times]]\{1\}\right). \tag{4.2}$$

Here we need that the mapping m be surjective; actually, it would suffice that the multiplication by 1 on the small circle is the identity. But this happens if and only if the small circle is not dotted, that is, we have 1, not $\dot{1}$.

Quite similar things happen for the second and for the third Reidemeister moves. The necessary conditions can be summarised as follows:

The small circles which appear for the second and the third Reidemeister move should not be dotted.

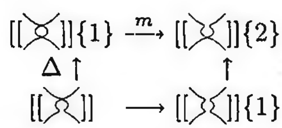
The explanation comes a bit later. Now, we see that this condition is obviously satisfied when the dotting comes from a cohomology class, and not necessarily the Stiefel-Whitney cohomology class for non-orientable surface. Any homology class should do.

Thus (modulo some explanations given below) we have proved the following

Theorem 4.1. *Let $\mathcal{M} \to M$ be a fibration with I-fibre so that $\mathcal{M}$ is orientable and M is a 2-surface. Let h be a Z_2-cohomology class and let g be the corresponding dotting. Consider the corresponding grading on $[[K]]$. Then for a link $K \subset \mathcal{M}$ the homology of $[[K]]_g$ is invariant under isotopy of K in M (with both the orientation of M and the I-bundle structure fixed) up to some shifts of the usual (quantum) grading and height (homological grading).*

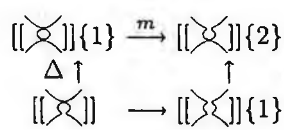
4.1. *Explanation for the second and the third moves*

We have the following picture for the Reidemeister move for :



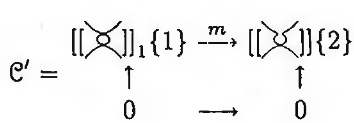
(4.3)

Here we use the notation $\{\cdot\}$ for the degree shifts, see page 296.



(4.4)

This complex contains the subcomlex $\mathcal{C}'$:



(4.5)

if *the small circle is not dotted.*

Here and further 1 denotes the mark on the small circle.

Then the acyclicity of $\mathcal{C}'$ is evident.

Factoring $\mathcal{C}$ by $\mathcal{C}'$, we get:

$$\begin{tikzcd} {[[\;]]\{1\}/_{1=0}} \arrow[r] & 0 \\ {[[\;]]} \arrow[u, "\Delta"] \arrow[r] & {[[\,)(\,]]\{1\}} \arrow[u] \end{tikzcd} \quad . \tag{4.6}$$

In the last complex, the mapping Δ directed upwards, is an isomorphism (when our small circle is not dotted). Thus the initial complex has the same homology group as $[[\,)(\,]]$. This proves the invariance under Ω_2.

The argument for Ω_3 is standard as well; it relies on the invariance under Ω_2 and thus we should also require that the small circle is not dotted.

5. More gradings; more examples

We have listed the necessary conditions for the dotting to give such a grading that the homology of $[[K]]_g$ is invariant (up to some shifts); the conditions are quite natural: additivity of dots modulo $\mathbf{Z}_2$ and triviality of small circles for all types of Reidemeister moves. We have actually missed one condition we assumed without saying. Namely, *in the pictures corresponding to the Reidemeister moves, the similar arcs are dotted similarly.*

This means, for example, that for the second Reidemeister move the smoothing)(gives two branches which should have the same dotting as the two branches of)(. The same follows for all the three moves.

Thus, we introduce the *dotting axiomatics.* Namely, assume we have some class of objects with Reidemeister moves, Kauffman bracket and the Khovanov homology (in the usual setup or in the setup of[Ma6]). Assume its circles can be dotted in such a way that the following conditions hold:

(1) The dotting of circles is additive with respect to $2 \to 1$ and $1 \to 2$-bifurcations, and it is preserved under $1 \to 1$-bifurcations.
(2) Similar curves for similar smoothings of the RHS and the LHS of any Reidemeister move have the same dotting
and
(3) Small circles appearing for the first, the second, and the third Reidemeister moves are not dotted.

Let us call the conditions above *the dotting conditions.*

Theorem 5.1. *Assume there is a theory with Khovanov complex* $([[K]], \partial)$ *such that the Kauffman states can be dotted so that the dotting conditions hold. Define* $[[K]]_g$ *as before (see page 301).*

1) Then the homology of $[[K]]_g$ is invariant (up to a degree shift and a height shift).

2) For any operator λ on the ground ring, the complex $[[K]]_g$ is well defined with respect to the differential $\partial' + \lambda\partial''$, and the corresponding homology is invariant (up to well-known shifts).

3) Moreover, if we have several dottings $g_1, g_2, \dots, g_k$ so that for each of them the dotting condition holds, then the homology of the complex $K_{g_1,\dots,g_k}$ with differential $\partial_{g_1,\dots,g_k}$ defined to be the projection of ∂ to the subspace preserving all the gradings, is invariant.

Proof. The first part of the theorem follows from the reasonings above.

Now, for the differential $\tilde{\partial} = \partial' + \lambda\partial''$ we have $(\tilde{\partial})^2 = \partial'^2 + \lambda(\partial'\partial'' + \partial''\partial') + \lambda^2\partial''^2$; the expression in the right hand side gives the projections of $(\partial)^2 = (\partial' + \partial'')^2$ to three subspaces of corresponding gradings taken with some coefficients (here $1, \lambda, \lambda^2$). Since $(\partial)^2 = 0$, all projections are zeroes. The invariance of the homology is proved as above. The main thing is that the mapping m is surjective and Δ is injective.

The proof of the last statement is analogous to the proof with only one grading. Again, it is enough to mention that m remains surjective and Δ remains injective. $\square$

5.1. *Examples*

One example (already published in the note[Ma7]) deals with the following situation. Consider a fixed oriented thickened surface $\mathcal{M}$ which is the total space of an I-fibre bundle over some 2-manifold M, not necessarily orientable. We assume the orientation of $\mathcal{M}$ and the I-bundle structure fixed.

Consider all $\mathbf{Z}_2$-cohomology classes $H^1(\mathcal{M})$ (there are finitely many of them). For knots in $\mathcal{M}$, each of these classes generates a dotting for circles (see page 301) in the Kauffman states, thus, it defines gradings for $[[K]]$. Call these gradings *additional* (with respect to the two usual Khovanov gradings). Denote the obtained complex by $[[K]]_{gg}$ and the projection of the differential ∂ by ∂_{gg}.

Theorem 5.2. *The homology of $[[K]]_{gg}$ with respect to ∂_{gg} is an invariant of K.*

Consider the category T of (classical or virtual) tangles with $2k$ open ends. Then the construction above allows to make the following dotting on the states of the Kauffman bracket.

Fix some number l and mark some of the tangle ends by some of l colours $1, 2 \ldots, l$.

Couple the endpoints of the tangle in an arbitrary way (so that any tangle closes into a classical or virtual knot).

Having done this, for any tangle $t \in T$, we can consider its closure $Cl(t)$. It acquires a dotting from l colours, thus we get l additional gradings for the Khovanov complex; denote the obtained complex by $[[Cl(t)]]_{dd}$, and denote the corresponding differential by ∂_{dd}.

From the above, we get the following

Theorem 5.3. *For any fixed endpoint coupling, the homology of $[[Cl(t)]]_{dd}$ is an invariant of t.*

A particular case of this refers to long classical (and virtual) knots.

Namely, if we deal with *long* virtual knots, this grading will lead to a new invariants. Note that *long* virtual knots do not coincide with *compact* virtual knots, see e.g.,[Ia4] There are non-trivial long virtual knots (and tangles) having only trivial classical closures. Say, it is easy to construct two classical $2-2$-tangle with the same classical closures and different virtual closures.

As for classical knots, thinking of them from the "long" point of view seems to be very prospective. In our case, if we take long classical knots and put one dot on one end, thus defining a new grading. This will split the usual Khovanov differential ∂ into $\partial' + \partial''$. The only circle which can support the new grading is the one obtained by closing the only long arc. It exists in every state, and it can be marked either by $\dot{X}$ or by $\dot{1}$. If we just take ∂', then it would split the initial Khovanov complex into two parts: the one with $\dot{X}$ and the one with $\dot{1}$ with no differential acting from one part to another.

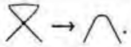
This is nothing but the usual *reduced Khovanov homology*.

However, if we take not just ∂', but $\partial' + \lambda\partial''$ for some ring R where λ is a zero divisor (say, 2 in the ring $\mathbf{Z}_4$).

This defines new invariants of ordinary knots (or links with one marked component).

However, it seems to be much more interesting when we pass from usual long knots to cables. Namely, having a long classical knot (assume it to be *framed*), we can take its n-cabling. Then for any dotting and for any closure the new homology groups will be invariants of the initial (long) classical knot.

One more example refers to *rigid virtual knots*. We consider virtual knot diagrams up to all Reidemeister moves and all detours preserving the

Whitney index of the curve. Namely, we prohibit the following *first virtual Reidemeister moves*: . Rigid virtual knots are of interest because all quantum invariants of classical knots (which can not be generalised for generic virtual knots) can be generalised in full totality for rigid virtual knots.

For such knots, since the first virtual Reidemeister move is forbidden, in any Kauffman state for any circle the number of self-intersections modulo 2 for such circles is invariant. It defines well a dotting, thus giving one new grading for rigid virtual knots.

5.2. *Braids*

It is a very intriguing question to get new gradings for classical knots (without going to long knots).

We are not going to consider braids just as a partial case of tangles and put various dots on the ends of the braid. We think of a braid as a source of constructing knot invariants via Markov moves.

Thus, a *closed braid* can be viewed of as a special kind of link in a thickened annulus $S^1 \times I \times I$. This annulus has non-trivial cohomology group $H^1(S^1 \times I \times I, \mathbf{Z}_2) = \mathbf{Z}_2$. From this we get an additional grading, thus having a complex $[[Cl(B)]]_g$ with differential ∂'; here $Cl(B)$ is the closure of a braid B. It is obvious that the homology of this complex is well defined not only under braid isotopies, but also under braid conjugations, since they preserve the closure.

Thus, in order to get a knot invariant, we have to overcome the second Markov move (adding a new loop). Unfortunately, if K' is obtained from K by a second Markov move then the homology of $Cl(K')$ should not coincide with the homology of $Cl(K)$. The reason is that the move we perform is the first Reidemeister move, and the small circle that appears is dotted.

However, this allows to extract the difficulty for proving the invariance of the the new dotted (grading) homology for knots in its *pure form*: the only obstacle we get is the first Reidemeister move.

Hopefully, the homology of this space with extra gradings behaves in a predictable manner under the Markov move, maybe, after some stabilisations.

We shall return to this question while speaking about filtrations and spectral sequences.

5.3. *Further gradings*

The construction above takes into account only $\mathbf{Z}_2$-homology classes (unlinke the construction of[APS]) where the *homotopy* information of Kauffman state circles was taken into account to construct a grading.

More homology information can be taken into account in the following manner.

Assume we have only one non-trivial cohomology class (say, we live on the thickened annulus or deal with long knots with one dot on one end).

Then such an object has $H^1 = \mathbf{Z}$. Previously, we were using only the $\mathbf{Z}_2$ information for constructing our differentials.

We shall now use the $\mathbf{Z}$-cohomology information to introduce the *secondary gradings* as follows.

If the usual grading coming from the $\mathbf{Z}_2$-cohomology class is non-trivial, then we decree the secondary grading to be zero. If the first grading is trivial, then we look at the value of the cohomology class not over $\mathbf{Z}_2$, but over $\mathbf{Z}_4$ and then we set the secondary grading to be 0 if the cohomology class is trivial modulo $\mathbf{Z}_4$ and 1 if it is equal to 2 modulo $\mathbf{Z}_4$. Analogously, in the case when the primary and the secondary gradings are both zero, we define the ternary grading to be 1 or 0 depending on the value of the $\mathbf{Z}_8$-cohomology (of course, if one of them was not zero, we set all further gradings to be zero).

This defines a family of further gradings on circles which answers the question what is the maximal power of 2, the corresponding value of the cohomology is equal to. For instance, such gradings can be all zeroes (say, if the circle is trivial) or $(1,0,0,\dots)$ or $(0,1,0,0,\dots)$ or $(0,0,1,0,0,\dots)$, etc.

These gradings define corresponding dottings and gradings for all elements 1 and X (as before, we count the gradings for X with plus, and the gradings for 1 with minus).

This defines a multigrading on the complex (chain set) $[[K]]$. Denote the obtained chain set by $[[K]]_{mg}$. The usual differential ∂ for $[[K]]$ splits into two parts: the one ∂' preserving the new multigrading and the one ∂'' not preserving the grading.

Lemma 5.1. *For any of the new gradings, the differential ∂'' either preserves it or increases it.*

Proof. Indeed, assume we have a bifurcation $2 \to 1$ or $1 \to 2$. Such a bifurcation may behave in two ways with respect to the new gradings on circles: either it preserves the total set (sum) of gradings (each considered modulo $\mathbf{Z}_2$) as in the case $(1,0,\dots)\wedge(1,0,\dots,) \to (0,0\dots)$, or it changes it, as in the case $(1,0,\dots,)\wedge(1,0,\dots) \to (0,1,\dots)$. In the second case the parity in one grading (in our case, the second) is violated, thus, ∂' equals zero.

In the first case we may think that our differential behaves in the same way with *all* the gradings separately, which returns us to the case of different gradings coming from different cohomology classes. □

The above reasonings lead us to the following

Theorem 5.4. *The homology of* $[[K]]_{mg}$ *with respect to* ∂' *is an invariant in the corresponding category.*

Analogously, one may consider the case when we have H^1 of rank greater than one.

6. Khovanov's Frobenius theory

The Khovanov theory for classical knots has some natural generalisations, some of them were first discovered by Khovanov. Here we briefly discuss the generalisation of them for the case of *knots in thickened surfaces and additional gradings.* The corresponding results without additional gradings were published in.[Ma3,Ma6]

Let $\mathcal{R}, \mathcal{A}$ be commutative rings, and let $\iota : \mathcal{R} \to \mathcal{A}$ be an embedding, such that $\iota(1) = 1$. The restriction functor mapping $\mathcal{A}$-modules to $\mathcal{R}$-modules has a right conjugate and a left conjugate: the induction functor $Ind(M) = \mathcal{A} \otimes_{\mathcal{R}} \mathcal{M}$ and the coinduction functor $CoInd(M) = Hom_{\mathcal{R}}(\mathcal{A}, \mathcal{M})$. One says that ι is a Frobenius embedding if these two functors are isomorphic. Equivalently: the embedding ι is Frobenius, if the restriction function has a two-sided dual functor. In this case one says also that the ring $\mathcal{A}$ is a *Frobenius extension* of $\mathcal{R}$ by means of ι.

In,[Kh2] Khovanov asked the question: to find a couple of linear spaces $(\mathcal{A}, \mathcal{R})$ such that, taking $\mathcal{R}$ as the basic coefficient ring and a Frobenius extension $\mathcal{A}$ over $\mathcal{R}$ as the homology ring of the unknot, we would be able to construct a link homology theory "in the same way" as the usual homology theory.

Here "in the same manner" means that we consider the state cube, where at each vertex we put a tensor power of $\mathcal{A}$ (over $\mathcal{R}$), corresponding to the number of circles in the given state, and define the partial differentials by means of m and Δ (multiplication and comultiplication), and then put signs on the edges of the cube and normalise the whole construction by height and grading shifts (he did not use wedge product or involution in the Frobenius algebra).

Khovanov showed that the invariance under the first Reidemeister move requires that $\mathcal{A}$ is a two-dimensional module over $\mathcal{R}$ and gave necessary an sufficeint conditions for the existence of such an invariant link homology theory.

Note that in the present section we shall mainly work with the classical setup and notation of Khovanov, that is, we use symmetric tensor powers and then add minus signs to the cube, thus restricting ourselves for the case when no $1 \to 1$-bifurcations in the state cube occur. We have partially generalised Khovanov Frobenius theory for the case of arbitrary virtual knots, and we shall return to that case in the end of the present section.

In,[Kh2] it is also shown that any link homology theory of such sort can be obtained by means of some operations (basis change, twisting and duality) from the following solution called *universal*:

(1) $\mathcal{R} = \mathbf{Z}[h, t]$.
(2) $\mathcal{A} = \mathcal{R}[X]/(X^2 - hX - t)$,
(3) $\deg X = 2, \deg h = 2, \deg t = 4$;
(4) $\Delta(1) = 1 \otimes X + X \otimes 1 - h1 \otimes 1$
(5) $\Delta(X) = X \otimes X + t1 \otimes 1$.

As we see, the multiplication in the algebra $\mathcal{A}$ preserves the grading, and the comultiplication increases this by 2.

We omit the normalisation regulating the corresponding gradings.

First note that this Frobenius theory contains (as an important partial case) the Lee-Rasmussen theory, see,[Lee,Ras] when we specify $t = h = 1$. The Lee-Rasmussen theory, has one grading less: indeed, the differentials here do not respect the quantum grading.

We call the theory constucted above the *universal* $(\mathcal{R}, \mathcal{A})$*-construction.* The corresponding homology of a (classical) link L is be denoted by $Kh_U(L)$.

The main question we address in the following section is: *how to split the differentials above into ∂' and ∂''?*

Note that if we introduce the new grading just by dotting and then counting the number of $\dot{X}$ minus the number of $\dot{1}$, the partial differential corresponding to ∂ [we call it ∂ as well; abusing notation] which is some tensor product (or wedge product) of one Δ or one μ with the identity operator, would not behave so nicely with respect to the new grading. Namely, the mapping Δ may take X to the sum $\dot{X} \wedge \dot{X} + \dot{1} \wedge \dot{1}$, see Fig. 6.1.

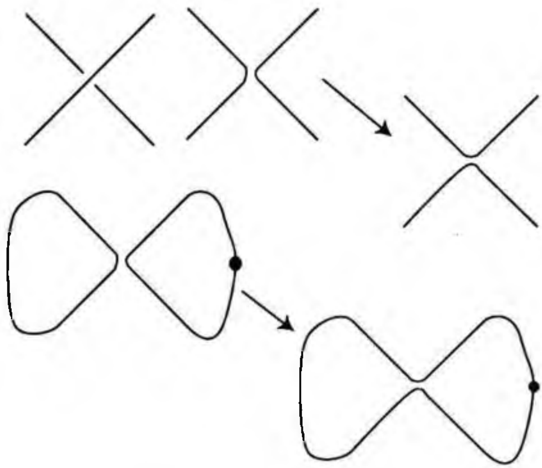


Fig. 6.1. The mapping Δ

The mapping to the first term *increases* the grading whence the mapping to the second term *decreases* it.

Thus, we have to repair the dotted grading. The correct answer is: define the dotted grading gr as the difference between $\#\dot{X} - \#\dot{1}$ plus half the total degree of monomials in t and h.

There is a trick with λ, which goes as follows. Denote the usual Khovanov differential by ∂, and denote the "Frobenius addition" containing h and t by ∂_F so that we totally have $\partial + \partial_F$. According to our rules, if some circles are dotted, and the Khovanov (Frobenius) theory is well established then we can introduce the new "dotted grading" gr as before, which splits the differential into two parts $\partial = \partial' + \partial''$.

Theorem 6.1. *Consider the basic ring* $\mathbf{Z}[h, t, \lambda | \lambda h = \lambda t = 0]$. *Then the homology of the Khovanov Frobenius complex with respect to the differential* $\partial_F + \lambda\partial''$ *is invariant.*

The proof goes as follows. We only need to mention that is that the square of this differential equals zero, because in the expression $(\partial_F + \lambda\partial'')^2$ the interaction between the "Frobenius part" of ∂_F and $\lambda\partial''$ gets cancelled. This proves that the complex is well defined with respect to the differential ∂''. However, one of our goals is to approach the Lee-Rasmussen theory, which is defined over $\mathbf{Q}$ with $t = 1$, $h = 0$. For these purposes, the approach above is not satisfactory.

Then, the terms in the differential corresponding to the "usual" multiplication and comultiplication (without new t and h) behave as before. Also, we know the behaviour of the grading when we have no dotted circles; it is correlated by degrees of h and t.

Consider the remaining cases.

$m : \dot{X} \otimes \dot{X} \to t \cdot 1, m : \dot{X} \otimes X \to t \cdot \dot{1}\Delta : \dot{X} \to t \cdot 1 \otimes 1, \Delta : \dot{X} \to \dot{1} \wedge \dot{1}.$

Let us look at our dotted grading more carefully. Denote the former dotted grading by gr', and let us construct the true dotted grading gr by varying gr'.

We count the usual quantum grading. It is equal to $tot(1) - tot(X) + h$, where $tot(1)$ is the total number of circles marked by 1 or by $\dot{1}$, $tot(X)$ is the total number of circles marked by X or by $\dot{X}$, and h is the height. Then we set

$$gr = gr' + \frac{tot(1) - tot(X) + h}{2} = \frac{\#\dot{X} + \#1 - \#\dot{1} - \#X + h}{2}.$$

Lemma 6.1. *The differential ∂ defined above can be split into summands each of which either preserves gr or increases it by 2.*

The proof follows from a direct calculation.

Then it is possible to split ∂ into ∂' (preserving the grading) and ∂'' increasing that by 2, and consider the dotted homology of $[[K]]_g$ with respect to ∂'. This homology will be invariant.

If we look at the differential ∂' more carefully, we will see that the new "Frobenius" mappings vanish when they are applied to sets of usual (not dotted) circles.

Namely, for the mapping $X \otimes X \to t \cdot 1$ we have: gr' does not change, whence the usual grading [coming from counting $tot(1) - tot(X) + h$] increases.

This means, that if we have no dots at all, the differential ∂' coincides with the usual Khovanov differential (without h and t).

Considering the Lee-Rasmussen theory for $t = 1$, $h = 0$, we get a complex $[[K]]_{LR}$ with a differential ∂_{LR} which coincides with the usual Khovanov differential in the case of classical knots. Note that the complex $[[K]]_{LR}$ has two gradings: the height and the grading gr (the quantum grading was lost).

However, in the dotted picture, this differential has some other interesting terms, like $\dot{X} \otimes \dot{X} \to 1$.

6.1. *Yet another definition of the Khovanov homology*

If we look at the complex constructed above from in the case we have no additional (dotted) gradings at all, we see that the new grading prohibits exactly those parts of the differential ∂_Φ which deal with t: e.g., $X\times X\to t\cdot 1$ does not change the dotted grading, but it does change the usual quantum grading if we forget about t.

Thus, *the definition above with $t=1$ leads to the usual Khovanov homology if no circle is dotted.*

On the other hand, if many circles are dotted, this is a sort of Lee-Rasmussen homology theory.

It is interesting that we can use a mixture to get another definition of the Khovanov homology theory. Namely, take a knot diagram K and put dots on circles in an arbitrary way. Then for every dotted circle change the notation: replace $\dot{1}$ by $\dot{X}$ and vice versa. The resulting complex would be precisely the Khovanov complex up to some renormalisation in the new grading which becomes coincident with the usual quantum grading.

This effect is interesting because it allows one to handle the situation with braids: whenever we perform the second Markov move, we replace $\dot{1}$ by $\dot{X}$, which leads to the injectivity of Δ and surjectivity of m. Unfortunately, this gives us no new homology theory, but it allows one to look at the usual Khovanov homology from another point of view.

6.2. *Khovanov Frobenius theory modulo $\mathbf{Z}_2$ in the general case*

The aim of the present section is to define the differential ∂_F generalizing the theory described above for the case of arbitrary virtual knots in the $\mathbf{Z}_2$ case. We shall describe the difficulties that occur in the general case of arbitrary virtual knots.

The main difficulty here is to define the differential corresponding to the $1\to 1$-bifurcation.

We start up with the chain structure of the complex. First, we assume for simplicity $h=0$, the case of generic h will be considered afterwards.

We deal with the ring $R=\mathbf{Z}[t]$, where t has grading 4.

With every circle in every Kauffman state we associate the graded module V over R freely generated by 1 of grading 0 and X of grading 2 (t has grading 4, as above). The generator 1 is assumed to be fixed for any circle; the generator X depends on the orientation of the circle as before.

With each Kauffman state with n corresponding circles, we associate the n-th exterior power of V, and we define the following operations "muliplication and comultiplication" just as before, however, corrected by terms containing h:

$m(1_1 \wedge 1_2) = 1, m(X_1 \wedge 1_2) = m(1_1 \wedge X_2) = X,$
$m(X_1 \wedge X_2) = 0$
$\Delta(1) = 1_1 \wedge X_2 + X_1 \wedge 1_2$
$\Delta(X) = X_1 \otimes X_2 + t 1_1 \otimes 1_2$, where in the definition of partial differentials it is assumed (as before) that we deal with the first two circles in the tensor product, and the first one is left (resp., upper), whence the second one is left (resp., lower).

For all $1 \to 1$-bifurcations, we set the partial differential to be equal to zero.

For all other bifurcations ($2 \to 1$ or $1 \to 2$), we define the partial differential ∂ just as in section 3.

Denote the resulting set of chain spaces for a given virtual knot diagram K by $[[K]]_t$.

Theorem 6.2. *The differential ∂ defines a complex on $[[K]]_t$, so that the homology of $[[K]]_t$ with respect to ∂ is an invariant of the link K.*

The well-definiteness proof actually repeats the main points of[Kh2] together with those in:[Ma6] one should consider all 2-faces of the corresponding cube and prove that they anticommute. The proof of the invariance under Reidemeister moves follows from the surjectivity of m and injectivity of Δ.

However, here we do not touch on the variable h. The reason why the construction proposed in[Ma6] behaves nicely when we add the variable t is the following: both in the usual Khovanov homology theory and in the Frobenius theory with some t and $h = 0$, the involution on the space $V = \{1, X\}$ defined by $1 \mapsto 1, X \mapsto -X$ behaves well with respect to the operations Δ and m: it changes signs of Δ and preserves the sign of m.

However, when we add a new variable h, we will not see this effect any more: the mapping Δ takes $1 \wedge 1 \to 1 \wedge X + X \wedge 1 - h \cdot 1 \wedge 1$. Here the involution $X \to -X$ changes the sign of one part $(1 \wedge X + X \wedge 1)$ and preserves the other part $(h \cdot 1 \wedge 1)$.

Also, the routine check of the well-definiteness (as in[Ma6]) of the complex, that is, anti-commutativity of the 2-faces of the cube, leads to an example shown below (we are citing,[Ma6] see Fig. 6.2) for the case $t = 0$.

First, consider the case $t = 0$. For the lower composition, we have the identical zero map by definition. Substituting X into the upper composition, we get $\pm X \wedge X$ at the first step and 0 at the second step. Substituting 1,

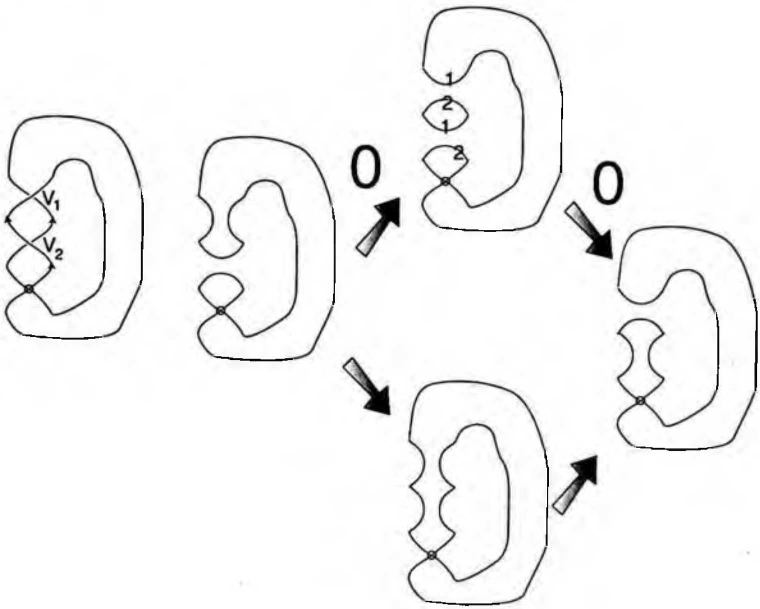


Fig. 6.2. A face of the cube

we first get $1_1 \wedge X_2 + X_1 \wedge 1_2$ here the index refers to the number of circle (the first circle is the big one), and the second index refers to the crossing number. While passing to the second crossing V_2 the circles change their roles: the first circle becomes the lower one, and the second circle becomes the upper one. Moreover, for the first circle we get a basis change: X maps to $-X$. Thus we get $-X\wedge 1+1\wedge X$, which is taken to zero by the multiplication m. Now, we have to check what happens for general t. Subsituting 1 to the upper composition, we will get no terms with t at all. Substituting X, we shall first get (besides $X_1 \wedge X_2$) also $t \cdot 1_1 \wedge 1_1$. Passing to the second crossing and multiplying, these terms will give $t \cdot 1 \wedge 1$ and $-t \cdot 1 \wedge 1$, which cancel each other.

The example above is in fact the key example of;[Ma6] it works without any changes when $h = 0$ (because t does not appear in the comultiplication of 1 or in the multiplication of $1 \wedge X$).

But in the case $h \neq 0$ it does appear, and this would lead to the fact that the $1 \to 1$-bifurcation should not be zero any more. We will in fact need to introduce a new variable being the square root of h.

On the other hand, h itself should be treated in a special way so that the multiplication m and comultiplication Δ behave nicely with respect to $1 \mapsto 1, X \mapsto -X$.

We shall consider this problem in a separate publication.

6.2.1. *The $\mathbf{Z}_2$-case*

We first consider the $\mathbf{Z}_2$-case solution given in.[Ma3] First note that there is no difference between $\wedge$ and $\otimes$, and we shall use the notation $\otimes$.

This time we do not set the $1 \to 1$-type partial differentials to be zero; we define this partial differential by some mapping $I : V \to V$, the matrix I will be defined later.

Here will show how the square root of h appears. Of course, in this case we shall not need exterior products and control the signs. Consider the basic ring of coefficients $\mathbf{Z}_2[t, c]$ with $\deg t = 4, \deg c = 1$ (we assume $c^2 = h$). Now, consider Fig. 6.2. We have the following situation: in the lower composition we have two maps corresponding to $1 \to 1$ bifurcations, thus the corresponding matrix should look like $I \cdot I$. Return to Fig. 6.2 in the upper part we have the composition of two mappings Δ and then m. Starting with 1, we get $\Delta(1) = 1 \otimes X + X \otimes 1 + h1 \times 1$. Multiplying, we see that $X \otimes 1$ and $1 \otimes X$ cancel each other, and the only remaining term is $h \cdot 1$. Now, if we start with X, we get $X \to X \otimes X + t \cdot 1 \otimes 1$. After the multiplication, we get $hX + t + t = h \cdot X$ (we are dealing with the $\mathbf{Z}_2$ case). Now we see that the corresponding transformation matrix looks like

$$\begin{pmatrix} 1 \\ X \end{pmatrix} \mapsto \begin{pmatrix} h & 0 \\ 0 & h \end{pmatrix} \cdot \begin{pmatrix} 1 \\ X \end{pmatrix} \tag{6.1}$$

For this scalar matrix $h \cdot Id$ we set the matrix corresponding to the $1 \to 1$-mapping to be $c \cdot Id$, and then any face of the bifurcation cube corresponding to Fig. 6.2 will (anti)commute. Then it is not difficult to see (see[Ma3]) that with this *scalar* $1 \to 1$-bifurcation matrix, all other faces (anti)commute as well.

Now, the dotted gradings gr appear straigthforwardly by counting *monomials in t and c* and correcting gr' by using this monomials. Denote the obtained homology by $Kh(K)_{tc}$.

Note that the degree of c is 1, so we will have *half-integer gradings*. This immeadiately leads to the following

Theorem 6.3. *If $Kh(K)_{tc}$ has a non trivial homology of half-integer additional grading then K has no diagram with orientable corresponding atom. In particular, the knot K is not classical.*

7. Gradings or filtrations? The spectral sequence

Since the works of Lee[Lee] and Rasmussen,[Ras] spectral sequences play a significant role in knot homology. Sometimes it turns out that studying convergence of a spectral sequence leads to some interesting and deep invariants such as Rasmussen's invariant, which is applicable to estimating the Seifert genus and the 4-ball genus of classical links.

The Lee-Rasmussen spectral sequence starts with the Khovanov homology and ends up with some two-term homology which carries a nice piece of information.

Recently (see[BN3]), it was discovered that the spectral sequence of Lee-Rasmussen does not converge after E_3-term, and that there are some nice torsions in Khovanov homology which survive after the E_3-term of the spectral sequence.

Our goal here is to construct a spectral sequence from the "complicated" theory with new dotted gradings to the "simple" (Khovanov) theory. Thus, in some sense our spectral sequence will behave with respect to the usual Khovanov homology as Khovanov homology itself behaves with respect to the Rasmussen homology.

It would also be very interesting to inspect two spectral sequences converging from the "complicated" theory to the Rasmussen theory.

The argument of the present section is standard. In all cases described above when we deal with one new (dotted) grading, the *old* differential $\partial = \partial' + \partial''$ in the complex $[[K]]_g$ does not decrease the new grading.

Thus, let us introduce the (dotted) *filtration* on the chain spaces as follows: we set $[[K]]_g^n = \{c \in [[K]]_g | gr(c) \geq n\}$. Then we have $[[K]]_g^{\infty} \subset \dots [[K]]_g^2 \subset [[K]]_g^1 \subset [[K]]_g^0 \subset [[K]]_g^{-1} \subset \dots \subset [[K]]_g^{-\infty}$.

The usual differential ∂ respects this filtration. This leads to the following

Theorem 7.1. *For any field of coefficients, there is a spectral sequence whose E_1-term is isomorphic to $[[K]]$ with the first differential ∂', the E_2-term isomorphic to the homology of $[[K]]_g$, so that this spectral sequence converges to the usual Khovanov homology (with respect to ∂).*

The argument proving this theorem is standard. We also conjecture that all terms of this spectral sequence are invariants (of knots, braids, tangles) in the corresponding category.

It would be very interesting to know whether some terms of the spectral sequence described above survive after the braid stabilsations. In this case we would be able to hope to construct gradings for usual knots without going into the long category.

Returning to the Lee-Rasmussen theory, we see that in the dotted case, we have two complexes: the usual Khovanov complex and the complex $([[K]]_{LR}, \partial_{LR})$ with homology $H(K)_{LR}$. They coincide in the case when we have no dotting, but they differ in the case when we have dotting.

Quite in the usual manner one proves

Theorem 7.2. *For the field* $\mathbf{Q}$, *there is a spectral sequence whose* E_1*-term is isomorphic to* $[[K]]_{LR}$ *with the first differential* ∂_{LR}, *the* E_2*-term isomorphic to the homology* $H(K)_{LR}$, *so that this spectral sequence converges to the Lee-Rasmussen homology.*

Thus, two bigraded homology theories (the usual Khovanov theory with height and quantum grading) and the one described above (with height and dotted grading) both converge to the Lee-Rasmussen theory.

It is known that the Lee-Rasmussen theory give nice invariants (quantum gradings of the two surviving elements). It would be interesting to compare the convergence of the spectral sequence describing above: what is the meaning of the dotted grading of surviving elements?

8. Applications

The theory above has some obvious applications coming from the definitions. Thus, if we work for knots in thickened surfaces, there is a natural question whether such a knot can be destablised, i.e., some handles of the surface are nugatory, or, in other words, the representative of the knot given by this surface is minimal. The surface M has $\mathbf{Z}_2$-homology group of rank k, and if they are all used as gradings of some homology groups of a knot in $M\tilde{\times}I$, then the knot can not be destabilised.

Corollary 8.1. *If a set of* additional *gradings of non-trivial groups of* $Kh_{gg}(K)$ *forms a subset in* $\mathbf{R}^k$ *not belonging to any hyperplane passing through zero, then the link* K *does not admit destabilisation, i.e., there is no surface* M' *of smaller genus obtained from* M *by a destabilisation so that the link* K *lies in the natural fibration over* M' *generated by* $\mathcal{M} \to M$.

Analogously, the dotted grading can be used for estimating the number of virtual crossings of a rigid virtual knot diagram.

Also, we mention (without any details, however) the facts which generalise straightforwardly for the case of new gradings:

(1) The homological length of the complex does not exceed the number of classical crossings.
(2) The spanning tree of Wehrli[Weh] and Champanerkar-Kofman[ChK] saying that the Khovanov homology can be obtained from a complex with a smaller chain group. This leads to the estimation for the thickness: $Th(Kh(K)) \leq 2+g$, where g is the genus of the atom corresponding to the diagram K.
Here the thickness estimates the number of diagonals with slope 2 on the plane with height and quantum gradings serving as coordinates.
The same estimates can be obtained for our complex with new gradings when looking at the *diagonals with respect to the former gradings.* This leads to

Theorem 8.1. *For any knot K, the thickness of the dotted Khovanov homology $Th(Kh_g(K)) \leq 2+g$, where g is the genus of the atom corresponding to any diagram of K.*

Together with the lemma saying that $span\langle K\rangle \leq 4n$, where n is the number of classical crossings, we get sharper estimates for the number of crossings.
(3) The Bar-Natan topological picture[BN2] for tangles and cobordisms, see also.[TuTu] We need to generalize Bar-Natan's topological category and construct a functor from it to our category. We shall discuss this in a separate publication.
(4) Rasmussen's estimates for the genus of a spanning surface; here we must, indicate the category of cobordisms, say, for knots in $M \times I$ we should consider spanning surfaces in $M \times I \times I$.

9. The relation to other papers

This paper generalises many constructions. First of all, we would like to mention the work,[APS] the work[Kh2] and the work.[Ma6]

In fact, the idea of taking new gradings counting X and 1 on non-trivial circles with opposite sides was originally used in.[APS] However, we used this approach for a more general situation. For instance, the grading there was *necessary to construct the Khovanov homology itself*; without it, the Khovanov theory for knots in thickened surfaces does not exist; even with it, it does not exist for knots in thickened $\mathbf{R}P^2$. We have taken the approach from[Ma6] with twisted coefficient as the basement for our homology theory (that allows us to give a fair generalisation of Khovanov's theory for virtual and twisted knots without any new gradings), and then introduced new gradings similar to those ones by M.Asaeda, J.Przytycki and A.Sikora.

They used integral homology or even homotopy classes to define the gradings. This was quite difficult for making it more algebraic.

We have axiomatized this approach taking the $\mathbf{Z}_2$-cohomology (or just dotting) making it applicable to many other situations.

On the other hand, classical knot diagrams considered up to braid-like moves also admit "dotting". This leads to a class of "knot-like" objects where not all equivalences are allowed; they were studied in,[AuFied] and their generalisation of the Khovanov homology turned out to be a partial case of ours.

Finally, we would like to mention a very recent paper by Ozsvath, Rasmussen, and Szabo[ORS] where "odd Khovanov homology" was introduced.

Like us they also used exterior vector product instead of symmetric products (the idea first appeared in[Ma6]) but with a different goal: they constructed another (odd) Khovanov homology with the same chain space, whence we rearranged the usual Khovanov homology making it working for virtual knots. Certainly, their construction enjoys many properties of the usual Khovanov complex (like thickness estimate in terms of atoms). We shall discuss the "odd Khovanov homology for virtual knots" and additional gradings for odd Khovanov homology in separate papers.

References

APS. Asaeda, M., Przytycki, J., Sikora, A. (2004), Categorification of the Kauffman bracket skein module of I-bundles over surfaces, *Algebraic and Geometric Topology*, **4**, No. 52, pp. 1177-1210.

AuFied. Audoux, B., Fiedler, Th. (2005), A Jones polynomial for braid-like isotopies of oriented links and its categorification, *Algebraic and Geometric Topology Volume* 5, pp. 15351553.

BN. Bar–Natan, D. (2002), On Khovanov's categorification of the Jones polynomial, *Algebraic and Geometric Topology*, **2**(16), pp. 337–370.

BN2. Bar–Natan, D. (2004), Khovanov's homology for tangles and cobordisms, arXiv:mat.GT/0410495.

BN3. Bar-Natan, D. (2007), Fast Khovanov homology computations, *Journal of Knot Theory and Its Ramifications*, **16** (3), pp. 243-256.

Bou. Bourgoin, M. O., Twisted Link Theory, arxiv: math. GT/0608233

ChK. Champanerkar, A., Kofman, I., Spanning trees and Khovanov homology, arxiv: math. GT/0607510

Dro. Drobotukhina Yu.V. (1991), An Analogue of the Jones-Kauffman poynomial for links in $\mathbf{R}^3$ and a generalisation of the Kauffman-Murasugi Theorem, *Algebra and Analysis*, **2**(3), pp. 613-630.

DK2. Dye, H.A., Kauffman, L.H. (2004), Minimal Surface Representation of Virtual Knots and Links, arXiv:math. GT/0401035 v1.

F. Fomenko A. T. (1991), The theory of multidimensional integrable hamiltonian systems (with arbitrary many degrees of freedom). Molecular table of all integrable systems with two degrees of freedom, *Adv. Sov. Math*, *6*, pp. 1-35.

FKM. Fenn, R.A, Kauffman, L.H, and Manturov, V.O. (2005), Virtual Knots: Unsolved Problems, *Fundamenta Mathematicae*, Proceedings of the Conference "Knots in Poland-2003", **188**.

GPV. Goussarov M., Polyak M., and Viro O., Finite type invariants of classical and virtual links// Topology. 2000. V. 39. P. 1045–1068.

Jac. Jacobsson, M. (2002), An invariant of link cobordisms from Khovanov's homology theory, arXiv:mat.GT/0206303 v1.

JKS. Jaeger, F., Kauffman, L.H., and H. Saleur (1994), The Conway Polynomial in S^3 and Thickened Surfaces: A new Determinant Formulation, *J. Combin. Theory. Ser. B.*, **61**, pp. 237-259.

KaV. Kauffman L.H., Virtual knot theory, Eur. J. Combinatorics. 1999. V. 20, N. 7. P. 662-690.

Kaw. Kawamuro, K.// Khovanov-Rozansky homology and the braid index of a knot,arXiv.Math:GT/0707.1130v1

Kh. Khovanov, M. (1997), A categorification of the Jones polynomial, *Duke Math. J*,**101** (3), pp.359-426.

Kh2. Khovanov, M. (2004), Link homology and Frobenius extensions, Arxiv.Math:GT/0411447

KhR1. Khovanov, M., Rozansky, L., Matrix Factorizations and Link Homology, Arxiv.Math:GT/0401268

KhR2. Khovanov, M., Rozansky, L.,Matrix Factorizations and Link Homology II, Arxiv.Math:GT/0505056

KK. Kamada, N. and Kamada, S. (2000), Abstract link diagrams and virtual knots, *Journal of Knot Theory and Its Ramifications*, **9** (1), pp. 93–109.

Kup. Kuperberg, G. (2002), What is a Virtual Link?, www.arXiv.org, math-GT/ 0208039, *Algebraic and Geometric Topology*, 2003, **3**, 587-591.

Lee. Lee, E.S. (2003) On Khovanov invariant for alternating links, arXiv: math.GT/0210213.

Ma0. Manturov, V.O. (2000), Bifurcations, Atoms, and Knots, *Moscow Univ. Math. Bull.*, **1**, pp. 3-8.

Ia. Manturov, V.O., Teoriya Uzlov (Knot Theory, In Russian), *RCD*, M.-Izhevsk, 2005.

Ma1. Manturov, V.O. (2004), The Khovanov polynomial for Virtual Knots, *Russ. Acad. Sci. Doklady*, **398**, N. 1., pp. 11-15.

Ma2. Manturov V.O. (2006), The Khovanov Complex and Minimal Knot diagrams, *Russ. Acad. Sci. Doklady*, **406**, (3), pp. 308-311.

Ma3. Manturov V.O. (2005), The Khovanov complex for virtual knots, *Fundamental and applied mathematics*, **11**, N. 4., pp. 127-152 (in Russian).

Ia4. Manturov V.O. (2005), On Long Virtual Knots, *Russ. Acad. Sci. Doklady*, **401** (5), pp. 595-598.

Ma6. Manturov V.O. (2007), Khovanov Homology for Virtual Knots with Arbitrary Coefficients, *Russ. Acad. Sci. Izvestiya*, **71**, N. 5, pp. 111-148.

Ma7. Manturov V.O. (2007) Additional gradings in the Khovanov Complex for Thickened Surfaces, *Russ. Acad. Sci. Doklady*, to appear.

Ma8. Manturov, V.O., Minimal diagrams of classical knots, ArXiv:GT/ 0501510.

Ma9. Manturov, V.O. (2003), Kauffman–like polynomial and curves in 2–surfaces, *J. Knot Theory Ramifications*, **12**, (8), pp.1145-1153.

Miy. Miyazawa, Y. (2006), Magnetic Graphs and an Invariant for Virtual Links, *J. Knot Theory & Ramifications*, **15** (10), pp. 1319-1334.

Oht. Ohtsuki, T. (2002), Quantum Invariants, *World Scientific*, Singapore.

ORS. Ozsváth, P., Rasmussen, J., Szabó, Z.(2007), Odd Khovanov homology,// www.arxiv.org/math-qa/0710.4300

Ras. Rasmussen, J. A. (2004), Khovanov Homology and the slice genus,ArXivMath:/GT. O402131.

Ras2. Rasmussen, J., Some Differentials on Khovanov-Rozansky Homology (2006), arXiv: math. GT/0607544

Shu. Shumakovitch, A.//www.arxiv.org/math-gt/ 0405474.

Tu. Turaev, V.G. (1987), A simple proof of the Murasugi and Kauffman theorems on alternating links. *Enseign. Math.* (2), 33(3-4):203-225.

TuTu. Turaev, V.G., Turner, P.(2005), Link homology and unoriented topological quantum field theory, //www.arxiv.org/math-gt/0506229 v1.

Viro. Viro, O., Virtual links and orientations of chord diagrams, *Proceedings of the Gökova Conference-2005*, International Press, pp. 187-212.

Weh. Wehrli, S.,A spanning tree model for Khovanov homology, arxiv: math. GT/0409328

Towards the Large N Duality between the Chern-Simons Gauge Theory and the Topological String Theory

Pan Peng

Department of Mathematics
Harvard University
One Oxford Street
Cambridge, MA 02138
Email: ppeng@math.harvard.edu

We survey some recent developments in the large N duality between the Chern-Simons gauge theory and the topological string theory. Some related problems and its applications to different areas of mathematics are also discussed. We dedicate this paper to the memory of Professor X.-S. Lin.

1. Introduction

The Gromov-Witten theory has been extensively studied in the past two decades. Computing Gromov-Witten invariants, especially at high genera, is always a challenging problem in the Gromov-Witten theory. In the case of lower genera, localization method or mirror symmetry might be applied to the computation. However, if the genus goes higher and higher, these two methods will involve more and more terms in the computation and thus become essentially impractical.

A complete answer can be given by the large N duality between the Chern-Simons gauge theory and the open Gromov-Witten theory. For example, in the case of toric Calabi-Yau threefolds, the topological vertex theory[1,17] expresses Gromov-Witten invariants as the combinatorial data of Chern-Simons invariants of certain torus link.

The first important bridge connecting gauge theory and string theory went back to t' Hooft's work in 1974, where he proposed that gauge theory can be identified as an $1/N$ expansion in string theory. It was in 1990, Witten made a very important step relating the Chern-Simons gauge theory on a three dimensional manifold as a topological string theory on its cotangent bundle. The final picture was merged by R. Gopakumar and C.

Vafa in 1998, in which they conjectured that at large N, topological string A-model of T^*S^3 with N D-branes is equivalent to the topological string theory on the resolved conifold $\mathcal{O}(-1) \oplus \mathcal{O}(-1) \to \mathbb{P}^1$. In 2000, H. Ooguri and C. Vafa described the topological string theory on the resolved conifold in terms of Chern-Simons invariants of knots.

To study this duality conjecture, an important observation is that, in both the Chern-Simons gauge theory and the topological string theory, a system of ODEs called cut-and-join equation is satisfied. In the Chern-Simons gauge theory, the cut-and-join equation reflects the deformation of the choice of framing, while, in the topological string theory, it naturally corresponds to the deformation of holomorphic curves mapped into the target space. Therefore, the duality can be approached by proving the uniqueness of the solution of the cut-and-join equation under certain condition.

Based on the large N Chern-Simons/topologial string (CS/TS) duality, in a series of papers, J.M.F. Labastida, M. Marino, H. Ooguri and C. Vafa conjectured the existence of a series of certain integer invariants which reveals the deep structure of quantum group invariants of links and integrality structure of the topological string theory. The highly nontrivial check of this integrality phenomenon gives a strong evidence of the large N Chern-Simons/topological string duality.

Motivated by the large N Chern-Simons/topological string duality, there are a lot of related problems should have their mathematical consequences and the corresponding mathematical foundation should be built. All these relations stimulate the development of geometry and topology in a very profound way. It's not surprising to see more connections between quite different areas of mathematics can be motivated from this duality picture.

This paper is organized as follows. In section 2, we define quantum group invariants of links. The large N Chern-Simons/topological string conjecture is stated in section 3. In section 4, some related results have been discussed. In the last section, we discuss some related problems and applications.

2. Quantum Group Invariants

2.1. *Partition and symmetric function*

A *partition* λ is a finite sequence of positive integers $(\lambda_1, \lambda_2, \cdots)$ such that

$$\lambda_1 \geq \lambda_2 \geq \cdots .$$

The total number of parts in λ is called the length of λ and denoted by $\ell(\lambda)$. We use $m_i(\lambda)$ to denote the number of times that i occurs in λ. The degree of λ is defined to be

$$|\lambda| = \sum_i \lambda_i \,.$$

If $|\lambda| = d$, we say λ is a partition of d. We also use notation $\lambda \vdash d$. The automorphism group of λ, $Aut\lambda$, contains all the permutations that permute parts of λ while still keeping it as a partition. Obviously, the order of $Aut\lambda$ is given by

$$|Aut\lambda| = \prod_i m_i(\lambda)! \,.$$

There is another way to rewrite a partition λ in the following format

$$(1^{m_1(\lambda)} 2^{m_2(\lambda)} \cdots)\,.$$

A traditional way to visualize a partition is to identify a partition as a *Young diagram*. The Young diagram of λ is a 2-dimensional graph with λ_j boxes on the j-th row, $j = 1, 2, ..., \ell(\lambda)$. All the boxes are put to fit the left-top corner of a rectangle. For example

$$(5, 4, 2, 2, 1) = (12^2 45) = 54221.$$

For a given partition λ, denote by λ^t the conjugate partition of λ. The Young diagram of λ^t is transpose to the Young diagram of λ: the number of boxes on j-th column of λ^t equals to the number of boxes on j-th row of λ, where $1 \le j \le \ell(\lambda)$.

By convention, we regard a Young diagram with no box as the partition of 0 and use notation (0). Denote by $\mathcal{P}$ the set of all partitions. We can define an operation " $\cup$ " on $\mathcal{P}$. Given two partitions λ and μ, $\lambda \cup \mu$ is the partition by putting all the parts of λ and μ together to form a new partition. For example

$$(12^2 3) \cup (15) = (1^2 2^2 35).$$

Using Young diagram, it looks like

$$3221 \quad \cup \quad 51 \quad = \quad 532211\,.$$

The following number associated with a partition λ is used throughout this paper,

$$\mathfrak{z}_\lambda = \prod_j j^{m_j(\lambda)} m_j(\lambda)!\,, \qquad\qquad \kappa_\lambda = \sum_j \lambda_j(\lambda_j - 2j + 1)\,.$$

It's easy to see that

$$\kappa_\lambda = -\kappa_{\lambda^t}. \tag{2.1}$$

A power symmetric function of a sequence of variables $x = (x_1, x_2, ...)$ is defined as follows

$$p_n(x) = \sum_i x_i^n.$$

For a partition λ,

$$p_\lambda(x) = \prod_{j=1}^{\ell(\lambda)} p_{\lambda_j}(x).$$

It is well-known that every irreducible representation of symmetric group can be labeled by a partition. Let χ_λ be the character of the irreducible representation corresponding to λ. Each conjugate class of symmetric group can also be represented by a partition μ such that the permutation in the conjugate class has cycles of length $\mu_1, \ldots, \mu_{\ell(\mu)}$. Schur function s_λ is determined by

$$s_\lambda(x) = \sum_{|\mu|=|\lambda|} \frac{\chi_\lambda(C_\mu)}{\mathfrak{z}_\mu} p_\mu(x) \tag{2.2}$$

where C_μ is the conjugate class of symmetric group corresponding to partition μ.

2.2. *Quantum group invariants of links*

Let $\mathcal{L}$ be a link with L components $\mathcal{K}_\alpha$, $\alpha = 1, \ldots, L$, represented by the closure of an element of braid group $\mathcal{B}_m$. We associate to each $\mathcal{K}_\alpha$ an irreducible representation R_α of quantized universal enveloping algebra $U_q(\mathfrak{sl}(N, \mathbb{C}))$, labeled by its highest weight Λ_α. Denote the corresponding module by V_{Λ_α}. The j-th strand in the braid will be associated with the irreducible module $V_j = V_{\Lambda_\alpha}$, if this strand belongs to the component $\mathcal{K}_\alpha$. The braiding is defined through the following *universal R-matrix* of $U_q(\mathfrak{sl}_N)$

$$\mathcal{R} = q^{\frac{1}{2}\sum_{i,j} C_{ij}^{-1} H_i \otimes H_j} \prod_{\text{positive root } \beta} \exp_q[(1 - q^{-1}) E_\beta \otimes F_\beta].$$

Here $\{H_i, E_i, F_i\}$ are the generators of $U_q(\mathfrak{sl}_N)$, (C_{ij}) is the Cartan matrix and

$$\exp_q(x) = \sum_{k=0}^{\infty} q^{\frac{1}{4}k(k+1)} \frac{x^k}{\{k\}_q!},$$

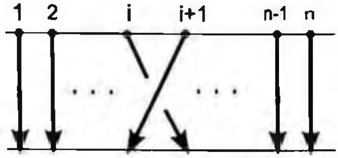
where

$$\{k\}_q = \frac{q^{-k/2} - q^{k/2}}{q^{-1/2} - q^{1/2}}, \qquad \{k\}_q! = \prod_{j=1}^{k} \{j\}_q .$$

Define *braiding* by $\check{\mathcal{R}} = P_{12}\mathcal{R}$, where $P_{12}(v \otimes w) = w \otimes v$.

Now for a given link $\mathcal{L}$ of L components, one chooses a closed braid representative in braid group $\mathcal{B}_m$ whose closure is $\mathcal{L}$. In the case of no confusion, we also use $\mathcal{L}$ to refer its braid representative in $\mathcal{B}_m$. We will assign each crossing by the braiding as follows. Let U, V be two $U_q(\mathfrak{sl}_N)$-modules labeling two outgoing strands of the crossing, the braiding $\check{R}_{U,V}$ (resp. $\check{R}^{-1}_{V,U}$) is assigned as

The above assignment will give a representation of $\mathcal{B}_m$ on $U_q(\mathfrak{g})$-module $V_1 \otimes \cdots \otimes V_m$. Namely, for any generator, $\sigma_i \in \mathcal{B}_m$,



define*

$$h(\sigma_i) = \underset{V_1}{\mathrm{id}} \otimes \cdots \otimes \check{\mathcal{R}}_{V_i, V_{i+1}} \otimes \cdots \otimes id_{V_N} .$$

Therefore, any link $\mathcal{L}$ will provide an isomorphism

$$h(\mathcal{L}) \in \mathrm{End}_{U_q(\mathfrak{sl}_N)}(V_1 \otimes \cdots \otimes V_m) .$$

Let $K_{2\rho}$ be the enhancement of $\check{\mathcal{R}}$ in the sense of,[31] where ρ is the half-sum of all positive roots of $\mathfrak{sl}_N$. The irreducible representation R_α is labeled by the corresponding partition A^α.

*In the case of σ_i^{-1}, use $\check{\mathcal{R}}^{-1}_{V_{i+1}, V_i}$ instead.

Definition 2.1. Given L labeling partitions $A^1, \dots, A^L$, the quantum group invariant of $\mathcal{L}$ is defined as follows:

$$W_{(A^1,\dots,A^L)}(\mathcal{L}) = q^{d(\mathcal{L})}\mathrm{Tr}_{V_1\otimes\cdots\otimes V_m}(K_{2\rho}\circ h(\mathcal{L})),$$

where

$$d(\mathcal{L}) = -\frac{1}{2}\sum_{\alpha=1}^{L}\omega(\mathcal{K}_\alpha)(\Lambda_\alpha, \Lambda_\alpha + 2\rho) + \frac{1}{N}\sum_{\alpha<\beta}\mathrm{lk}(\mathcal{K}_\alpha, \mathcal{K}_\beta)|A^\alpha|\cdot|A^\beta|,$$

and $\mathrm{lk}(\mathcal{K}_\alpha, \mathcal{K}_\beta)$ is the linking number of components $\mathcal{K}_\alpha$ and $\mathcal{K}_\beta$. A substitution of $t = q^N$ is used to give a two-variable framing independent link invariant.

3. Large N CS/TS duality

The Chern-Simons partition function associated to a link $\mathcal{L}$ is defined to be the following generating function:

$$Z_{CS}(\mathcal{L}; q, t; x) = 1 + \sum_{A^1,\cdots,A^L} W_{(A^1,\cdots,A^L)}(q,t)\prod_{\alpha=1}^{L} s_{A^\alpha}(x^\alpha)$$

where s_A is the Schur function. Free energy is defined to be

$$F = \log Z_{CS}.$$

Quantum group invariants of links can be expressed as vacuum expectation value of Wilson loops which admit a large N expansion in physics. It can also be interpreted as a string theory expansion (also see[12] for more details). The geometric picture of $f_{(A^1,\dots,A^L)}$ is proposed in.[14] One can rewrite the free energy as

$$F = \sum_{\mu^1,\dots,\mu^L} \sqrt{-1}^{\sum_\alpha \ell(\mu^\alpha)} \sum_{g=0}^{\infty} \lambda^{2g-2+\sum_\alpha \ell(\mu^\alpha)} F_{g,(\mu^1,\dots,\mu^L)}(t) \prod_\alpha p_{\mu^\alpha}.$$

The quantities $F_{g,(\mu^1,\dots,\mu^L)}(t)$ can be interpreted in terms of the Gromov-Witten invariants of Riemann surface with boundaries. It was conjectured in[27] that for every link $\mathcal{L}$ in S^3, one can canonically associate a lagrangian submanifold $\mathcal{C}_\mathcal{L}$ in the *resolved conifold* X

$$\mathcal{O}(-1)\oplus\mathcal{O}(-1)\to\mathbb{P}^1,$$

with $b_1(\mathcal{C}_\mathcal{L}) = L$, the number of components of $\mathcal{L}$. The construction of such lagrangian submanifolds by C. H. Taubes[2] can be served as a candidate.

Let γ_α, $\alpha = 1, \dots, L$, be one-cycles representing a basis for $H_1(\mathcal{C}_\mathcal{L}, \mathbb{Z})$ and $\beta \in H_2(X, \mathcal{C}_\mathcal{L}; \mathbb{Z})$. Consider the holomorphic map $f : \Sigma_{g,h} \to X$ from genus g Riemann surface with $h = \sum_\alpha \ell(\mu^\alpha)$ holes to the resolved conifold X such that $f_*[\Sigma_{g,h}] = \beta$ and the boundary mapped to the cycle γ_α with the prescribed winding number. The "number" of such maps will be denoted by $K^\beta_{g,(\mu^1,\dots,\mu^L)}$. These numbers are the open-string analogue of Gromov-Witten invariants†. $F_{g,(\mu^1,\dots,\mu^L)}(t,\tau)$ can thus be expressed as the following generating function of open Gromov-Witten invariants:

$$F_{g,(\mu^1,\dots,\mu^L)}(t,\tau) = \sum_\beta K^\beta_{g,(\mu^1,\dots,\mu^L)} e^{\int_\beta \omega}$$

where ω is the Kähler class of the Calabi-Yau threefold X, and

$$t = e^{\int_{\mathbb{P}^1} \omega}.$$

One can write $t^Q = \int_\beta \omega$, where Q is in general a half-integer.

4. Known results

4.1. *Mariño-Vafa formula*

The first important example is of course the case of the unknot. In,[24] based on the large N Chern-Simons/topological string duality, M. Mariño and C. Vafa proposed a formula relating the Chern-Simons invariants of the unknot (in this case, it is the quantum dimension), to certain generating series of Hodge integrals. For a mathematical proof, please refer to.[20]

Let $\overline{M}_{g,n}$ denote the Deligne-Mumford moduli stack of stable curves of genus g with n marked points. Let $\pi : \overline{M}_{g,n+1} \to \overline{M}_{g,n}$ be the universal curve, and let ω_π be the relative dualizing sheaf. The Hodge bundle $\mathbb{E} = \pi_*\omega_\pi$ is a rank g vector bundle over $\overline{M}_{g,n}$. Let $s_i : \overline{M}_{g,n} \to \overline{M}_{g,n+1}$ denote the section of π which corresponds to the i-th marked point, and let $\mathbb{L}_i = s_i^*\omega_\pi$. A Hodge integral is an integral of the form

$$\int_{\overline{M}_{g,n}} \psi_1^{j_1} \cdots \psi_n^{j_n} \lambda_1^{k_1} \cdots \lambda_g^{k_g}$$

where $\psi_i = c_1(\mathbb{L}_i)$ is the first Chern class of $\mathbb{L}_i$, and $\lambda_j = c_j(\mathbb{E})$ is the j-th Chern class of the Hodge bundle. Let

$$\Lambda_g^\vee(u) = u^g - \lambda_1 u + \cdots + (-1)^g \lambda_g$$

be the Chern polynomial of $\mathbb{E}^\vee$, the dual of the Hodge bundle.

†The definition of open Gromov-Witten invariants is still at large in general.

Define

$$\mathcal{C}_{g,\mu}(\tau) = -\frac{\sqrt{-1}^{\ell(\mu)}}{|Aut(\mu)|}[\tau(\tau+1)]^{\ell(\mu)-1}\prod_{i=1}^{\ell(\mu)}\frac{\prod_{a=1}^{\mu_i-1}(\mu_i\tau+a)}{(\mu_i-1)!} \cdot \int_{\overline{M}_{g,\ell(\mu)}}\frac{\Lambda_g^\vee(1)\Lambda_g^\vee(-\tau-1)\Lambda_g^\vee(\tau)}{\prod_{i=1}^{\ell(\mu)}(1-\mu_i\psi_i)}.$$

The Mariño-Vafa formula gives the following identity:

Theorem 4.1 ([20]).

$$\sum_\mu p_\mu(x)\sum_{g\geq 0} u^{2g-2+\ell(\mu)}\mathcal{C}_{g,\mu}(\tau) = \log\Big(1+\sum_A s_A(q^\rho)s_A(x)\Big) \tag{4.1}$$

where $q^\rho = (q^{-1/2}, q^{-3/2}, \cdots, q^{-n+1/2}, \cdots)$.

As a remark to the Mariño-Vafa formula, one may see that it provides a very strong tool to study the intersection theory of the moduli space of curves.

4.2. *Topological vertex theory*

Topological vertex theory was proposed in[1] to give a complete solution to the topological string theory on toric Calabi-Yau threefolds. A mathematical version of the topological vertex theory is given in,[17] which is approached by the relative Gromov-Witten theory.[15,16] We will adapt this version to give a rough picture of topological vertex theory (for more details, please refer to[17]).

A Calabi-Yau threefold X is *toric* if it contains an algebraic torus $(\mathbb{C}^*)^3$ as an open dense subset and the $(\mathbb{C}^*)^3$ action can be extended to X. Let X^1 be the union of all one-dimensional $(\mathbb{C}^*)^3$orbit closures in X, X^0 the union of $(\mathbb{C}^*)^3$ fixed points. Naturally assume that X^1 is connected and X^0 is not empty. Given $p \in X^0$, $(\mathbb{C}^*)^3$ acts on T_pX and $\wedge^3 T_pX$, where T_pX is the tangent space of X at p. The action of $(\mathbb{C}^*)^3$ on $\wedge^3 T_pX$ gives an irreducible character $\alpha : (\mathbb{C}^*)^3 \to \mathbb{C}^*$. α is independent of choice of p due to Calabi-Yau condition and connectedness of X^1. Define $T = \mathrm{Ker}\alpha \cong (\mathbb{C}^*)^2$. Let $T_\mathbb{R} \cong U(1)^2$ be the maximal compact subgroup of T and $\mu : X \longrightarrow \mathfrak{t}_\mathbb{R}^\vee$ be the moment map of the $T_\mathbb{R}$-action on X, where $\mathfrak{t}_\mathbb{R}^\vee$ is the dual of the Lie algebra of $T_\mathbb{R}$.

The image of X^1 gives a planar trivalent graph Γ. Each vertex of Γ corresponds to a fix points of T, and each edge of Γ corresponds to an irreducible component C^e of X^1.

Let $\mathcal{M}^{\bullet}_{\chi}(X, \beta)$ be the moduli space of stable maps from possibly disconnected domains to X within the class $\beta \in H_2(X, \mathbb{Z})$ satisfying the constraint $2\chi(\mathcal{O}_X) = \chi$. Define the generating function of disconnected Gromov-Witten invariants of degree β summing for all genera:

$$Z_\beta(u) = \sum_{\chi} u^{\chi} \int_{[\mathcal{M}^{\bullet}_{\chi}(X,\beta)]^{vir}} 1 . \tag{4.2}$$

Topological partition function of X is defined as follows:

$$Z_{top\, str} = 1 + \sum_{\beta \neq 0} Z_\beta(u) Q^\beta .$$

To compute Gromov-Witten invariants of toric Calabi-Yau threefold X, one can degenerate X into two relative Calabi-Yau threefold (Y_1, D) and (Y_2, D) with normal crossing singularity along divisor D. The Gromov-Witten invariants of X can be obtained as a combination of the relative Gromov-Witten invariants of (Y_1, D) and (Y_2, D).[15,16] This procedure can keep going on until all the pieces are indecomposable ones. All these indecomposable pieces are precisely the topological vertex as proposed in physics which amplitudes are certain weighted Chern-Simons invariants. This gives the following formula:

$$Z_\beta(q) = \sum \prod_e w_e \prod_v w_v . \tag{4.3}$$

Here w_e is some combinatorial formula associated with the inner edge e of the toric diagram Γ, and w_v is the three-partition Hodge integral associated with the vertex v except for a possible negative sign determined by its profile. Three-partition Hodge integral is related to the topological vertex amplitudes by a convolution formula proved in.[17] Given three partitions A, B, C, the topological vertex amplitude is give by

$$\begin{aligned} \mathcal{W}_{A,B,C}(q) = q^{-(\kappa_A - 2\kappa_B - \frac{1}{2}\kappa_C)/2} \sum c^{\nu^+}_{(\nu^1)^t B} c^{A}_{(\eta)^t \nu^1} c^{C}_{\eta^3 (\nu^3)^t} \\ \times q^{(-2\kappa_{\nu^+} - \frac{\kappa_{\nu^3}}{2})/2} \mathcal{W}_{\nu^+, C}(q) \frac{1}{\mathfrak{z}_\mu} \chi_{\eta^1}(\mu) \chi_{\eta^3}(2\mu) . \end{aligned} \tag{4.4}$$

Therefore, this closed form of topological string partition gives a complete solution to the topological string theory on any given toric Calabi-Yau threefolds.

4.3. *Labastida-Mariño-Ooguri-Vafa Conjecture*

The Chern-Simons partition function of $\mathcal{L}$ is a generating function of quantum group invariants of links given by

$$Z_{\mathrm{CS}}(\mathcal{L}; q, t) = \sum_{A^1,\ldots,A^L} W_{(A^1,\ldots,A^L)}(\mathcal{L}; q, t) \prod_{\alpha=1}^{L} s_{A^\alpha}(x^\alpha) \tag{4.5}$$

for any arbitrarily chosen sequence of variables

$$x^\alpha = (x_1^\alpha, x_2^\alpha, \ldots,),$$

where $s_{A^\alpha}(x^\alpha)$ is the Schur function.

Free energy is defined to be:

$$F = \log Z_{\mathrm{CS}}.$$

Use plethystic exponential, one can obtain‡

$$F = \sum_{d=1}^{\infty} \sum_{A^1,\ldots,A^L} \frac{1}{d} f_{(A^1,\ldots,A^L)}(q^d, t^d) \prod_{\alpha=1}^{L} s_{A^\alpha}((x^\alpha)^d), \tag{4.6}$$

where

$$(x^\alpha)^d = ((x_1^\alpha)^d, (x_2^\alpha)^d, \ldots).$$

Based on the duality between Chern-Simons gauge theory and topological string theory, J.M.F. Labastida, M. Mariño, H. Ooguri, C. Vafa conjectured that $f_{\vec{A}}$ have the following highly nontrivial structures.

For any A, $B \in \mathcal{P}$, define the following function

$$M_{AB}(q) = \sum_{\mu} \frac{\chi_A(C_\mu)\chi_B(C_\mu)}{\mathfrak{z}_\mu} \prod_{j=1}^{\ell(\mu)} (q^{-\mu_j/2} - q^{\mu_j/2}). \tag{4.7}$$

Conjecture 4.1 (LMOV). *For any* $(A^1, \ldots, A^L) \in \mathcal{P}^L$,

(i). *there exist* $P_{(B^1,\ldots,B^L)}(q,t)$ *for* $(B^1, \ldots, B^L) \in \mathcal{P}^L$, *such that*

$$f_{(A^1,\ldots,A^L)}(q,t) = \sum_{|B^\alpha|=|A^\alpha|} P_{(B^1,\ldots,B^L)}(q,t) \prod_{\alpha=1}^{L} M_{A^\alpha B^\alpha}(q). \tag{4.8}$$

Furthermore, $P_{(B^1,\ldots,B^L)}(q,t)$ *has the following expansion:*

$$P_{(B^1,\ldots,B^L)}(q,t) = \sum_{g=0}^{\infty} \sum_{Q \in \mathbb{Z}/2} N_{(B^1,\ldots,B^L);\,g,Q} (q^{-1/2} - q^{1/2})^{2g-2} t^Q. \tag{4.9}$$

‡It also gives a definition of $f_{(A^1,\ldots,A^L)}$.

(ii). $N_{(B^1,\ldots,B^L);g,Q}$ *are integers.*

In a joint work with K. Liu, we proved the following theorem:

Theorem 4.2 ([22]). *Notations as above, we have:*

$$(q^{\frac{1}{2}} - q^{-\frac{1}{2}})^2 P_{(B^1,\ldots,B^L)}(q,t) \in \mathbb{Z}[(q^{1/2} - q^{-1/2})^2, t^{\pm 1/2}].$$

The above theorem not only implies the integrality in the LMOV conjecture, but also shows these integer invariants vanish at large g and Q. It also gives a very strong evidence of the duality between Chern-Simons theory and topological string theory. The existence of algebraic structure of $P_{(B^1,\ldots,B^L)}(q,t)$ implies a deep structure of quantum group invariants of links. The pole order of $f_{(A^1,\ldots,A^L)}(q,t)$ at $q = 1$ a priori tends to go to infinity as the degree of the labeling irreducible representation goes higher and higher. However, the LMOV conjecture claims that the pole at $q = 1$ is at most of order 2, which implies many miracle cancelations happened.

As a direct corollary, we consider the case that all the labeling irreducible representations are fundamental ones. The quantum group invariant reduces to HOMFLY polynomial. Simply apply the cancelation at the lowest order, we obtain the following theorem by Lickorish and Millett which was originally proved through rather complicated Skein analysis:

Theorem 4.3 ([18]). *Let $\mathcal{L}$ be a link with L components. Its HOMFLY polynomial,*

$$P_{\mathcal{L}}(q,t) = \sum_{g\geq 0} p^{\mathcal{L}}_{2g+1-L}(t)\left(q^{-\frac{1}{2}} - q^{\frac{1}{2}}\right)^{2g+1-L},$$

satisfies:

$$p^{\mathcal{L}}_{1-L}(t) = t^{-\mathrm{lk}}\left(t^{-\frac{1}{2}} - t^{\frac{1}{2}}\right)^{L-1} \prod_{\alpha=1}^{L} p_0^{\mathcal{K}_\alpha}(t).$$

Here $p_0^{\mathcal{K}_\alpha}(t)$ is HOMFLY polynomial of the α-th component of the link $\mathcal{L}$ with $q = 1$.

4.4. $U(N)$ *Chern-Simons gauge theory*

It is in fact more nature to consider $U(N)$ Chern-Simons gauge theory. Chern-Simons vevs change under a change of framing, which will give a change on the invariants similarly defined in the LMOV conjecture related to the BPS degeneracies of domain walls in a geometry with different IR behavior. In the LMOV conjecture, the integrality of $N_{(B^1,\dots,B^L);g,Q}$ is already highly nontrivial, and it provides one of the major evidences for the duality proposed in.[4] The integrality predictions for any framing are even more surprising.[24] In a joint work with K. Liu, we prove the integrality holds for any framing change[23] in the setting of $U(N)$ Chern-Simons gauge theory.

4.5. *Uniqueness of the cut-and-join system*

A very important feature in both Chern-Simons gauge theory and topological string theory is that they both satisfy a system of non-linear ODEs, which is called cut-and-join equations. In the Chern-Simons gauge theory, the cut-and-join equation corresponds to the deformation of the framing, while in the topological string theory, it reflects deformations of Riemann surface mapped into the given Calabi-Yau threefold. Therefore, the large N Chern-Simons/topological string duality should be reduced to the uniqueness of this cut-and-join system in the end of the day.

In,[30] we study this system and pride an interesting condition to the uniqueness of the cut-and-join system.

Theorem 4.4 ([30]). *Associated to any link $\mathcal{L}$ of L components, $Z_1(\mathcal{L})$ and $Z_2(\mathcal{L})$ can be expressed as the following:*

$$Z_1(\mathcal{L}) = 1 + \sum_{\mu^1,\dots,\mu^L} R_{(\mu^1,\dots,\mu^L)}(q,t;\tau_1,\dots,\tau_L)\prod_{\alpha=1}^{L} p_{\mu^\alpha}(x_\alpha),$$

$$Z_2(\mathcal{L}) = 1 + \sum_{\mu^1,\dots,\mu^L} G_{(\mu^1,\dots,\mu^L)}(q,t;\tau_1,\dots,\tau_L)\prod_{\alpha=1}^{L} p_{\mu^\alpha}(x_\alpha)$$

satisfy the following cut-and-join equation:

$$\frac{\partial Z_k}{\partial \tau_\alpha} = \sum_{i,j\geq 1}\left((i+j)p_i^\alpha p_j^\alpha \frac{\partial}{\partial p_{i+j}^\alpha} + ijp_{i+j}^\alpha \frac{\partial^2}{\partial p_i^\alpha \partial p_j^\alpha}\right) Z_k$$

where $k = 1, 2$, $1 \leq \alpha \leq L$, and $p_i^\alpha = p_i(x^\alpha)$. If $R_{(1,\dots,1)} = G_{(1,\dots,1)}$ for any link $\mathcal{L}$, then $Z_1 = Z_2$ for any link $\mathcal{L}$.

If we specify all the labeling irreducible representation of a given link $\mathcal{L}$ to be the fundamental representation, the quantum group invariants of links reduces to the famous HOMFLY polynomial of $\mathcal{L}$. However, we known that HOMFLY polynomial can be recursively determined by the skein relation. Therefore, the above theorem will imply that if one could prove that in the topological string theory, such skein relation is satisfied by the generating function of the open Gromov-Witten invariants, one could conclude the duality between Chern-Simons gauge theory and topological string theory at large N.

5. Concluding remarks

It is of course very important to construct open Gromov-Witten theory. D. Joyce's work[10] will put this foundation a solid ground. We expect a general method to prove that the partition function satisfies the cut-and-join equation on the geometry side.

There are also many interesting applications to knot theory from the large N Chern-Simons/topological string duality conjecture. As one may find the geometry of the moduli space of stable maps from Riemann surfaces to Calabi-Yau threefolds reveals further structure of three-dimensional topology. For example, volume conjecture was proposed by Kashaev in[11] and reformulated by.[25] It relates the volume of hyperbolic 3-manifolds to the limits of quantum invariants. This conjecture was later generalized to complex case[26] and to incomplete hyperbolic structures.[8] The study of the volume conjecture is still staying at a rather primitive stage. We expect that certain vanishing phenomenon in the open Gromov-Witten theory will give a deep characterization of these limits of quantum invariants.

As mentioned above, quantum group invariants satisfy skein relation which must have some implications on topological string side as mentioned in.[12] One could also rephrase a lot of unanswered questions in the knot theory in terms of the open Gromov-Witten theory and vice versa. We hope that the relation between knot theory and open Gromov-Witten theory will be explored much more in detail in the future.

References

1. M. Aganagic, A. Klemm, M. Mariño and C. Vafa, *The topological vertex*, arxiv.org: hep-th/0305132.
2. C.H. Taubes, *Lagrangians for the Gopakumar-Vafa conjecture*, Adv. Theor. Math. Phys. 5 (2001) 139-163.
3. P. Freyd, D. Yetter, J. Hoste, W.B.R. Lickorish, K. Millet and A. Ocneanu, *A new polynomial invariant of knots and links*, Bull. Amer. Math. Soc. 12 (1985) 239.
4. R. Gopakumar and C. Vafa, *On the gauge theory/geometry correspondence*, Adv. Theor. Math. Phys. 3 (1999) 1415.
5. R. Gopakumar and C. Vafa. *M-theory and topological strings, I*, arxiv: 9809187.

6. R. Gopakumar and C. Vafa, *M-theory and topological strings-II*, arxiv.org: hep-th/9812127.
7. I.P. Goulden, *A differential operator for symmetric functions and the combinatorics of multiplying transpositions*, Trans. Amer. math. Soc. 344 (1994), 421.
8. S. Gukov, *Three-dimensional quantm gravity, Chern-Simons theory, and the A-polynomial*, Comm. Math. Phys. 255 (2005), no. 3, 577.
9. V.F.R Jones, *Hecke algebras representations of braid groups and link polynomials*, Ann. of Math. 126 (1987) 335.
10. D. Joyce, *Kuranishi bordism and Kuranishi homology*, arxiv.org: math.SG/0707.3572.
11. R.M. Kashaev, *"The hyperbolic volume of knots from the quantum dilogarithm"*, Lett. Math. Phys. 39 (1997), no. 3, 269.
12. J.M.F. Labastida and M. Mariño, *A new point of view in the ehory of knot and link invariants*, J. Knot Theory Ramif. 11 (2002), 173.
13. J.M.F. Labastida and M. Mariño, *Polynomial invariants for torus knots and topological strings*, Comm. Math. Phys. 217 (2001), no. 2, 423.
14. J.M.F. Labastida, M. Marino and C. Vafa, *Knots, links and branes at large N*, J. High Energy Phys. 2000, no. 11, Paper 7.
15. J. Li, *Stable morphisms to singular schemes and relative stable morphisms*, J. Differential Geom. 57 (2001), no. 3, 509–578.
16. J. Li, *A Degeneration formula of GW-invariants*, J. Differential Geom. 60 (2002), no. 2, 199.
17. J. Li, C.-C. Liu, K. Liu and J. Zhou, *A mathematical theory of the Topological Vertex*, math.AG/0408426.
18. W.B.R Lickorish and K.C. Millett, *A polynomial invariant of oriented links*, Topology 26 (1987) 107.
19. X.-S. Lin and H. Zheng, *On the Hecke algebra and the colored HOMFLY polynomial*, math.QA/0601267.
20. C.-C. Liu, K. Liu and J. Zhou, *A proof of a conjecture of Mariño-Vafa on Hodge integrals*, J. Differential Geom. 65 (2003), no. 2, 289.
21. C.-C. Liu, K. Liu and J. Zhou, *A formula of two-partition Hodge integrals*, J. Amer. Math. Soc. 20 (2007), no. 1, 149.
22. K. Liu and P. Peng, *Proof of the Labastida-Marino-Ooguri-Vafa Conjecture*, arXiv: 0704.1526.
23. Kefeng Liu and Pan Peng, *Framed knot and integrality structure in the $U(N)$ Chern-Simons gauge theory*, preprint.
24. M. Mariño and C. Vafa, *Framed knots at large N*, *Orbifolds in mathematics and physics* (Madison, WI, 2001), 185–204, Contemp. Math., 310, Amer. Math. Soc., Providence, RI, 2002.
25. H. Murakami and J. Murakami, *"The colored Jones polynomials and the simplicial volume of a knot"* Acta Math. 186 (2001), no. 1, 85.
26. H. Murakami, J. Murakami, M. Okamoto, T. Takata and Y. Yokota, *"Kashaev's conjecture and the Chern-Simons invariants of knots and links"*, Experiment. Math. 11 (2202), no. 3, 427.
27. H. Ooguri and C. Vafa, *Knot invariants and topological strings*, Nucl. Phys.

B 577 (2000), 419.

28. P. Peng, *A simple proof of Gopakumar-Vafa conjecture for local toric Calabi-Yau manifolds*, to appear, Commun. Math. Phys.
29. P. Peng, *Integrality structure in the Gromov-Witten theory*, Preprint.
30. P. Peng, *Uniqueness of the cut-and-join system*, preprint.
31. M. Rosso and V. Jones, *On the invariants of torus knots derived from quantum groups*, J. Knot Theory Ramifications 2 (1993), 97.
32. V.G. Turaev, *The Yang-Baxter equation and invariants of links*, Invent. Math. 92 (1988), 527.
33. E. Witten, *Quantum field theory and the Jones polynomial*, Commun. Math. Phys. 121 (1989) 351.
34. E. Witten, *Chern-Simons gauge theory as a string theory*, arxiv.org: hel-th/9207094, in *The Floer memorial volume*, The Floer memorial volume, 637, Progr. Math., 133, Birkhäuser, Basel, 1995.

On the Heegaard Genera of 3-manifolds Containing Non-separating Surfaces

Ruifeng Qiu*

Fengchun Lei†

Department of Mathematics
Dalian University of Technology
Dalian 116024, China
Emails: qiurf@dlut.edu.cn, ffcclei@yahoo.com.cn

Dedicated to the memory of Professor Xiao-Song Lin

Abstract. Let M be a compact orientable 3-manifold which contains a non-separating closed incompressible surface F. Let $M^{'} = M - \eta(F)$ where $\eta(F)$ is an open regular neighborhood of F in M. In the paper we show that if $M^{'}$ has a Heegaard splitting $V^{'} \cup_{S^{'}} W^{'}$ with $d(S^{'}) > 2g(M^{'})$, then $g(M) \geq g(M^{'}) - g(F)$. Furthermore, if F is a torus, then $g(M) \geq g(M^{'}) + 1$.

1. Introducion

Let M be a compact orientable 3-manifold. If there is a closed surface S which cuts M into two compression bodies V and W with $S = \partial_+ W = \partial_+ V$, then we say M has a Heegaard splitting, denoted by $M = V \cup_S W$; and S is called a Heegaard surface of M. Moreover, if the genus $g(S)$ of S is minimal among all Heegaard splittings of M, then $g(S)$ is called the genus of M, denoted by $g(M)$. If there are essential disks $B \subset V$ and $D \subset W$ such that $\partial B \cap \partial D = \emptyset$, then $V \cup_S W$ is said to be weakly reducible. Otherwise, it is said to be strongly irreducible.

Let $M = V \cup_S W$ be a Heegaard splitting. The distance between two essential simple closed curves α and β on S, denoted by $d(\alpha, \beta)$, is the smallest integer $n \geq 0$ so that there is a sequence of essential simple closed curves $\alpha_0 = \alpha, \ldots, \alpha_n = \beta$ on S such that α_{i-1} is disjoint from α_i for

*Supported by a grant of NSFC (No. 10625102)
†Supported by a grant of NSFC (No. 10571034)

$1 \leq i \leq n$. The distance of the Heegaard splitting $W \cup_S V$ is $d(S) = Min\{d(\alpha, \beta)\}$, where α bounds a disk in V and β bounds a disk in W. This was first introduced by Hempel. See [4]. It is clear that $V \cup_S W$ is reducible if and only if $d(S) = 0$, and $V \cup_S W$ is weakly reducible if and only if $d(S) \leq 1$.

Let F be either a properly embedded surface in a 3-manifold M or a sub-surface of ∂M. If there is an essential curve on F which bounds a disk in M with its interior disjoint from F or F is a 2-sphere which bounds a 3-ball in M, then we say F is compressible; otherwise, F is said to be incompressible. If F is an incompressible surface not parallel to ∂M, then F is said to be essential. A 3-manifold is said to be reducible if it contains an incompressible 2-sphere; otherwise, it is said to be irreducible.

Let F be an essential closed surface in a 3-manifold M. Assume that F is separating in M. Then F cuts M into two manifolds M_1 and M_2. Now if $M_i = V_i \cup_{S_i} W_i$ is a Heegaard splitting, $i = 1, 2$, then M has a natural Heegaard splitting $V \cup_S W$ called the amalgamation of $V_1 \cup_{S_1} W_1$ and $V_2 \cup_{S_2} W_2$ with $g(S) \leq g(S_1) + g(S_2) - g(F)$. From this point of view, $g(M) \leq g(M_1) + g(M_2) - g(F)$. There are some examples to show that it is possible that $g(M) = g(M_1) + g(M_2) - g(F) - n$ for any integer n. See [7] and.[16] J. Johnson [5] proved that $g(M) \geq 1/5(g(M_1) + g(M_2) - g(F))$ when M_1 and M_2 are anannular. In general, J. Schultens[15] proved that $g(M) \geq 1/5(g(M_1) + g(M_2) - g(F) - N(M_1) - N(M_2))$ where $N(M_i)$ is the number of pairwise disjoint non-parallel essential annuli in M_i. Furthermore, some sufficient conditions for $g(M) = g(M_1) + g(M_2) - g(F)$ have been given. See [1], [6] and [9].

Assume now that F is non-separating in M. Let $\eta(F)$ ($N(F)$) be an open (closed) regular neighborhood of F in M. We denote by F_1 and F_2 the two boundary components of $N(F)$. Let $M' = M - \eta(F)$ and $M' = V' \cup_{S'} W'$ be a Heegaard splitting such that $F_1, F_2 \subset \partial_- V'$. Then M has a natural Heegaard splitting $V \cup_S W$ called the self-amalgamation of $V' \cup_{S'} W'$ with $g(S) = g(S') + 1$. In this case, $g(M) \leq g(M', F_1 \cup F_2) + 1$. Furthermore, if M is homeomorphic to $F \times S^1$, then $g(M) = g(M', F_1 \cup F_2) + 1$. See [13]. In this paper, we shall give a lower bound of $g(M)$ when M' has a high distance Heegaard splitting. The main result is the following:

Theorem 1. Let M be a compact orientable 3-manifold, and F a non-separating incompressible closed surface in M. Let $M' = M - \eta(F)$. If M' has a Heegaard splitting $V' \cup_{S'} W'$ with $d(S') \geq 2g(M')$, then $g(M) \geq g(M') - g(F)$.

Corollary 2. Under the assumptions of Theorem 1, if F is a torus, then $g(M) \geq g(M^{'}) + 1$.

2. An extension of Schultens's Lemma

Let $M = V \cup_S W$ be a strongly irreducible Heegaard splitting, and F be a collection of essential surfaces in M. F is called a good separating system if $M - F$ contains two components M_1 and M_2; furthermore, for any subset $F^{'}$, $M - F^{'}$ contains only one component. By Schultens's lemma, S can be isotoped to intersect F in essential simple closed curves on both F and S. In this section, we will give an extension of Schultens's lemma. The main argument is due to [6].

Lemma 2.1. Let $M = V \cup_S W$ be a strongly irreducible Heegaard splitting, and F a good separating system in M which cuts M into two manifolds M_1 and M_2. Then S can be isotoped so that

(1) each of $S \cap M_1$ and $S \cap M_2$ is incompressible or

(2) one of $S \cap M_1$ and $S \cap M_2$, say $S \cap M_1$, is incompressible while all components of $S^{'} \cap M_2$ are incompressible except one bicompressible component.

(3) one of $S \cap M_1$ and $S \cap M_2$, say $S \cap M_1$, is incompressible while $S \cap M_2$ is compressible. Furthermore, there is a Heegaard surface $S^{'}$ isotopic to S such that

(i) $S^{'} \cap M_1$ is compressible while $S^{'} \cap M_2$ is incompressible, and

(ii) $S^{'}$ is obtained by ∂-compressing S in M_2 only one time.

Remark on Lemma 2.1. Bachman, Schleimer and Sedgwick [1] gave an extension of Schultens's lemma similar to Lemma 2.1 when F is connected and closed.

Proof. Let $\{H_1, H_2\} = \{W, V\}$. If each of $S \cap M_1$ and $S \cap M_2$ is incompressible, then Lemma 2.1(1) holds. If one of $S \cap M_1$ and $S \cap M_2$ is bicompressible, then, since $V \cup_S W$ is strongly irreducible, Lemma 2.1(2) holds. We may assume that

(1) one of $S \cap M_1$ and $S \cap M_2$ is compressible in $M_1 \cap H_1$ or $M_2 \cap H_1$.

(2) $S \cap M_i$ is incompressible in $M_i \cap H_2$ for $i = 1, 2$.

Since F is a collection of essential surfaces in M, H_1 and H_2 are nontrivial compression bodies. Let D is an essential disk of H_2 such that $|D \cap F|$ is minimal among all essential disks in H_2. By Assumption (2), $|D \cap F| > 0$. Furthermore, we may assume that

(3) S is a strongly irreducible Heegaard surface such that $|D \cap F|$ is minimal among all Heegaard surfaces isotopic to S and satisfying Assumptions (1) and (2).

Let a be an outermost component of $D \cap F$ on D. This means that a, together with an arc b on $\partial D(\subset S)$, bounds a disk B in D which lies in either $M_1 \cap H_2$ or $M_2 \cap H_2$ such that $B \cap F = a$. By the minimality of $|D \cap F|$, B is a ∂-compressing disk of $S \cap M_1$ or $S \cap M_2$.

Now there are two cases:

Case 1. $D \subset M_2 \cap H_2$ and $S \cap M_1$ is compressible in $M_1 \cap H_1$.

By Assumption (2), $S \cap M_2$ is incompressible in $M_2 \cap H_2$.

Now let S' be the Heegaard surface of M obtained by ∂-compressing S along D. We denote by H_1' and H_2' the two components of $M - S'$. We may assume that $M_1 \cap H_1 \subset M_1 \cap H_1'$. Since $S \cap M_1$ is compressible in $M_1 \cap H_1$, $S' \cap M_1$ is compressible in $M_1 \cap H_1'$, and $S' \cap M_2$ is incompressible $M_2 \cap H_2'$. Now if $S' \cap M_1$ is compressible in $M_1 \cap H_2'$, then Lemma 2.1(2) holds.

Suppose that $S' \cap M_1$ is incompressible in $M_1 \cap H_2'$. Then $S' \cap M_i$ is either incompressible or compressible in $M_i \cap H_1'$ but not bicompressible. Now $D \cap H_2'$ is an essential disk in H_2'. But $|D \cap H_2' \cap F| = |D \cap F| - 1$. This contradicts Assumptions (1), (2) and (3).

Case 2. $D \subset M_2 \cap H_2$, $S \cap M_2$ is compressible in $M_2 \cap H_1$, and $S \cap M_1$ is incompressible in $M_1 \cap H_1$.

By Assumption (2), $S \cap M_1$ is incompressible in $M_1 \cap H_2$. Hence $S \cap M_1$ is incompressible in M_1. Similarly, let S' be the Heegaard surface of M obtained by ∂-compressing S along D. We denote by H_1' and H_2' the two components of $M - S'$. We may assume that $M_1 \cap H_1 \subset M_1 \cap H_1'$. By Assumption (2), $S \cap M_2$ is incompressible in $M_2 \cap H_2$. Hence $S' \cap M_2$ is incompressible in $M_2 \cap H_2'$. If $S' \cap M_2$ is incompressible in $M_2 \cap H_1'$, then Lemma 2.1(3) holds.

Suppose that $S' \cap M_2$ is compressible in $M_2 \cap H_1'$. Since S' is also a strongly irreducible Heegaard surface, $S' \cap M_1$ is incompressible in $M_1 \cap H_2'$. But $|D \cap H_2' \cap F| = |D \cap F| - 1$. This contradicts Assumptions (1), (2) and (3). Q.E.D.

3. The proof of Theorem 1

Lemma 3.1. (1).[3] Let $M = V \cup_S W$ be a Heegaard splitting, and F be an incompressible surface in M. Then either F can be isotoped to be disjoint from S or $d(S) \leq 2 - \chi(F)$.
(2).[12] Let $V \cup_S W$ and $V^* \cup_{S^*} W^*$ be two Heegaard splittings for M. Then either $d(S) \leq 2g(S^*)$ or $V^* \cup_{S^*} W^*$ is a stabilization or ∂-stabilization of $V \cup_S W$.

The proofs of Theorem 1 and Corollary 2. Assume that F is a non-separating incompressible closed surface of genus at least one in M. We denote by $\eta(F)$ and $N(F)$ the open and closed regular neighborhoods of F in M, F_1 and F_2 the two boundary components of $N(F)$. Let $M' = M - \eta(F)$. We assume that M' has a Heegaard splitting $V' \cup_{S'} W'$ with $d(S') > 2g(M')$. By Lemma 3.1, $g(M') = g(S')$. By Haken's lemma, M' and M are irreducible.

Now let $M = V \cup_S W$ be a minimal Heegaard splitting of M. Then $V \cup_S W$ is irreducible.

Claim 1. If $V \cup_S W$ is weakly reducible, then $g(S) \geq g(M') + 1$.

Proof. Now $V \cup_S W$ has a thin position as

$$V \cup_S W = (V_1' \cup_{S_1'} W_1') \cup_{H_1} \ldots \cup_{H_{n-1}} (V_n' \cup_{S_n'} W_n') \qquad (*)$$

where $n \geq 2$, each component of $H_1, \ldots, H_{n-1}$ is an incompressible closed surface in M, and $V_i \cup_{S_i} W_i$ is a strongly irreducible non-trivial Heegaard splitting for $1 \leq i \leq n$. See [11].

Suppose that $g(S) < g(M') + 1$. By Lemma 3.1(2), $g(M', F_1 \cup F_2) \leq g(M') + g(F)$. Note that $g(H_i) \leq g(S)$. By Lemma 3.1(1), H_i is disjoint joint from M'. Hence each component of $H_1, \ldots, H_{n-1}$ is parallel to F. Hence one of the manifolds $M_1, M_2, \ldots, M_n$ is homeomorphic to M', say M_1, and each of $M_2, \ldots, M_n$ is homeomorphic to $F \times I$. This means that $g(S) = g(M', F_1 \cup F_2) + 1$. See [8]. Q.E.D. (Claim 1)

Now suppose that $V \cup_S W$ is strongly irreducible, and $g(S) < g(M') - g(F)$.

Claim 2. S can be isotoped so that $S \cap M'$ is bicompressible while $S \cap N(F)$ is incompressible.

Proof. Let $\mathfrak{F} = F_1 \cup F_2$. Then $\mathfrak{F}$ is a good separating system of M. By Lemma 2.1, there are three cases:

Case 1. $S \cap M'$ and $S \cap N(F)$ are incompressible.

Since $g(S) < g(M') - g(F)$ and $d(S') > 2g(M')$, by Lemma 3.1(1), S can be isotoped to be disjoint from M'. This means that a compression body contains an essential closed surface, a contradiction.

Case 2. one of $S \cap M'$ and $S \cap N(F)$ is bicompressible while the other is incompressible.

By the argument in Case 1, $S \cap M'$ in bicompressible and $S \cap N(F)$ is incompressible.

Case 3. $S \cap M'$ is compressible while $S \cap N(F)$ is incompressible. Furthermore, there is a Heegaard surface S^* isotopic to S such that $S^* \cap M'$ is incompressible while $S^* \cap N(F)$ is compressible.

Again by the argument in Case 1, this is impossible. Hence Claim 2 holds. Q.E.D. (Claim 2)

By Claim 2, we may assume that $S \cap M'$ is bicompressible while $S \cap N(F)$

is incompressible. Furthermore, we assume that $|S \cap \mathfrak{F}|$ is minimal among all Heegaard surfaces isotopic to S and satisfying the above condition.

Claim 3. $S \cap M'$ contains only one component.

Proof. Since $V \cup_S W$ is strongly irreducible, there is one component of $S \cap M'$, say C, which is bicompressible, and $S \cap M' - C$ is incompressible. Suppose that $S \cap M' - X$ contains at least one component C'. Then, by Lemma 3.1(1), C' is parallel to a surface in $\mathfrak{F}$. Hence C' can be disjoint from M'. This contradicts the assumption on $|S \cap \mathfrak{F}|$. Q.E.D. (Claim 3)

The following argument is essentially due to [10].

By Claim 3, $S \cap M'$ is connected and bicompressible. Let S_V be the surface obtained by maximally compressing $S \cap M'$ into $M' \cap V$, and S_W be the surface obtained by maximally compressing $S \cap M'$ into $M' \cap W$. By the nested lemma, S_V and S_W are incompressible in M'. This means that there are n essential disks $D_1, \ldots, D_n$ in V such that $S_V = (S \cap M' - \cup_{i=1}^n D_i \times [0,1]) \cup_{i=1}^n D_i \times \{0,1\}$. By Lemma 3.1(1), each component of S_V and S_W, say C, is parallel to one component of $\mathfrak{F} - \partial C$, say C^*, in M'. We denote by W_C the handlebody bounded by C and C^*.

Claim 4. If C_1 and C_2 are two components of S_V (S_W), then $W_{C_1} \cap W_{C_2} = \emptyset$.

Proof. Suppose that there are two components of S_V, say C_1 and C_2, such that $W_{C_1} \cap W_{C_2} \neq \emptyset$. Then $C_2 \subset W_{C_1}$. Since C_2 is incompressible in W_{C_1}, C_2 is parallel to a surface $C' \subset C_1^*$ in W_{C_1}. Let $C_1 \times I$ be a regular neighborhood of C_1 in W_{C_1} such that $C_1 \times \{0\} = C_1$. Now there are two cases:

Case 1. $C_1 \times (0,1] \subset intV$.

Now $D_i \times [0,1]$ is disjoint from $C_1 \times \{1\}$ for each $1 \leq i \leq n$. Since $C_2 \subset W_{C_1}$, $S \cap M'$ is not connected, contradicting to Claim 3.

Case 2. $C_1 \times I \subset W \cup_{i=1}^n D_i \times [0,1]$.

Now let $C_1 \times [-1,0]$ be a regular neighborhood of C_1 in V such that $C_1 \times [-1,0) \subset intV$. Then $D_i \times [0,1]$ is disjoint from $C_1 \times \{-1\}$. Hence either all the components of S_V lie in W_{C_1} or $S \cap M'$ is not connected. Suppose now that all the components of S_V lie in W_{C_1}. Then S can be isotoped to be disjoint from $\mathfrak{F}$, a contradiction. Q.E.D. (Claim 4)

By Claim 4, each component of S_V is parallel to a component of $\mathfrak{F} - S \cap \mathfrak{F}$, say C_*, which lies in V. Similarly, each component of S_W is parallel to a component of $\mathfrak{F} - S \cap \mathfrak{F}$ which lies in W.

Let $S_1' = (S \cap M') \cup (\mathfrak{F} \cap V)$, and $S_2' = (S \cap M') \cup (\mathfrak{F} \cap W)$. Then S_1' and S_2' are two Heegaard surfaces of M'. By Lemma 3.1(2), $g(S_1'), g(S_2') \geq g(S')$. Note that $\chi(\mathfrak{F}) = \chi(F_1) + \chi(F_2) = 2\chi(F)$, $\chi(\mathfrak{F}) = \chi(\mathfrak{F} \cap V) + \chi(\mathfrak{F} \cap W)$. Now we may assume that $\chi(\mathfrak{F}\cap V) \geq 1/2\chi(\mathfrak{F}) = \chi(F)$. Hence $\chi(S) = \chi(S\cap M')+ \chi(S\cap N(F)) \leq \chi(S_2') - \chi(F) + \chi(S\cap N(F))$. Hence $g(S) \geq g(M') - g(F)$. Thus Theorem 1 holds.

Now suppose that F is a torus. Then each component of $\mathfrak{F}\cap V$ and $\mathfrak{F}\cap W$ is an annulus. In this case, $d(S_1'), d(S_2') \leq 2$. For details, see [10]. Hence $g(S_1'), g(S_2') \geq g(S') + 1$. Thus Corollary 2 holds. Q.E.D.

References

1. D. Bachman, S. Schleimer and E. Sedgwick, Sweepouts of amalgamated 3-manifolds, Algebr. Geom. Topol. 6(2006) 171-194.
2. A. Casson and C. McA Gordon, Reducing Heegaard splittings, Topology Appl. 27(1987) 275-283.
3. K. Hartshorn, Heegaard splittings of Haken manifolds have bounded distance, Pacific J. Math. 204(2002) 61-75.
4. J. Hempel, 3-manifolds as viewed from the curve complex, Topology 40(2001) 631-657.
5. K. Johannson, Topology and combinatorics of 3-manifolds, LMN 1599, ISBN 3-540-59063-3, Springer-Verlag, Berlin Heideberg (1995).
6. T. Kobayashi and R. Qiu, The amalgamation of high distance Heegaard splittings is unstabilized, Preprint.
7. Kobayashi, R. Qiu, Y. Rieck and S. Wang, Separating incompressible surfaces and stabilizations of Heegaard splittings, Math. Proc. Cambridge Philos. Soc. 137(2004) 633-643.
8. T. Kobayashi and Y. Rieck, Heegaard genus of the connected sum of m-small knots, Commu. Anal. Geom. 14(2006), 1037-1077.
9. M. Lackenby, The Heegaard genus of amalgamated 3-manifolds, Geom. Dedicata 109(2004) 139-145.
10. R. Qiu, K. Du, J. Ma and X. M. Zhang, Distance and the Heegaard genera of annular 3-manifolds, Preprint.
11. M. Scharlemann and A. Thompson, Thin position for 3-manifolds. Geometric Topology (Haifa, 1992) 231-238, Contemp. Math., 164, Amer. Math. Soc., Providence, RI, 1994.
12. M. Scharlemann and M. Tomova, Alternate Heegaard genus bounds distance, Geom. Topol. 10(2006) 593-617.
13. J. Schultens, The classification of Heegaard splittings for (compact orientable surfaces)$\times S^1$, Proc. London Math. Soc. 67 (1993), 425-448
14. J. Schultens, Additivity of Tunnel number for small knots, Comment. Math. Helv. 75(2000) 353-367.
15. J. Schultens, Heegaard genus formula for Haken 3-manifolds, Geom. Delicata 119(2006), 49-68.
16. J. Schultens and R. Weidman, Destabilizing amalgamated Heegaard splittings, preprint, ArXiv: math.GT/0510386.

A Geometric Categorification of Representations of $U_q(sl_2)$

H. Zheng
Department of Mathematics, Zhongshan University,
Guangzhou, Guangdong 510275, China
E-mail: zhenghao@mail.sysu.edu.cn

The purpose of this short note it to give an elementary description of a recent work[18] of the author on the geometric categorification of tensor products of $U_q(sl_2)$-modules. That is, we show the main ideas without going into specific techniques.

Keywords: Quantum sl_2; Geometric categorification; Intersection complex.

1. Introduction

The term categorification was invented by Crane.[5,6] It refers to the process of finding category-theoretic analogues of set-theoretic concepts, so that one recovers the set-theoretic structures from the Grothendieck group of the categories involved.

Listed below are some typical examples of abelian categorification. See Ref. 15 for a recent review.

set	category
$\mathbb{Z}$	vector spaces
$\mathbb{Z}[t, t^{-1}]$	graded vector spaces
abelian group	abelian category
module	abelian category & exact endofunctors

More precisely, categorification of a module M over an algebra A means lifting the module M to an additive or abelian category $\mathcal{C}$ and, accordingly, lifting the algebra A to a collection of endofunctors of $\mathcal{C}$ as well as functor isomorphisms among them, in such a way that the Grothendieck group of $\mathcal{C}$ recovers the module M and the endofunctors and the isomorphisms among them recover the module structure of M.

The possibility of categorifying representations of quantum enveloping algebras was first observed from their canonical bases introduced by Lusztig[16] and Kashiwara:[12] canonical bases have many pretty nice properties necessary for categorification, such as integrality and positivity.

However, the categorification task turned out to be a rather difficult one. For treatments of tensor products of the fundamental $U_q(sl_2)$-module, see Bernstein-Frenkel-Khovanov[2] and Chuang-Rouquier.[4]

Very recently, Frenkel-Khovanov-Stroppel[8] succeeded in categorifying tensor products of general finite-dimensional $U_q(sl_2)$-modules, by using at full length many deep results in representation theory of Lie algebras.

Later, in Ref. 18 the author fulfilled two tasks in a purely geometric way. One is the same as Frenkel-Khovanov-Stroppel's work; the other is the categorification of R-matrices among the tensor products of $U_q(sl_2)$-modules. In this note we will give an elementary description of the first part of this work.

2. Quantum sl_2

The quantum enveloping algebra[13] $U_q(sl_2)$ is the $\mathbb{Q}(q)$-algebra defined by the generators

$$K, K^{-1}, E, F$$

and the relations

$$\begin{aligned} &KK^{-1} = K^{-1}K = 1, \\ &KE = q^2 EK, \\ &KF = q^{-2} FK, \\ &EF - FE = \frac{K - K^{-1}}{q - q^{-1}}. \end{aligned}$$

We can rewrite the last relation in an equivalent form

$$qEF + q^{-1}FE + K^{-1} = q^{-1}EF + qFE + K.$$

The simple $U_q(sl_2)$-modules are parameterized by their dimensions. More precisely, for a nonnegative integer d, let Λ_d be the $\mathbb{Q}(q)$-linear space spanned by

$$\{v_0, v_1, \ldots, v_d\}.$$

Then the followings endow Λ_d with a $U_q(sl_2)$-module structure (we define

$v_{-1} = v_{d+1} = 0$).

$$\begin{aligned} Kv_r &= q^{d-2r} v_r, \\ Ev_r &= [d-r+1]_q v_{r-1}, \\ Fv_r &= [r+1]_q v_{r+1}, \end{aligned}$$

where

$$[n]_q = \frac{q^n - q^{-n}}{q - q^{-1}}.$$

For a composition $\mathbf{d} = (d_1, d_2, \ldots, d_l)$ of d (i.e. a sequence of nonnegative integers summing up to d), we have a tensor product of $U_{\mathcal{A}}$-modules

$$\Lambda_{\mathbf{d}} = \Lambda_{d_1} \otimes \Lambda_{d_1} \otimes \cdots \Lambda_{d_l}.$$

It has a standard basis

$$\{v_{r_1} \otimes v_{r_2} \otimes \cdots \otimes v_{r_l} \mid 0 \le r_i \le d_i\}.$$

3. A toy model: simple $U_q(sl_2)$-modules

As a warm-up, we illustrate by a simple example how to categorify $U_q(sl_2)$-modules through geometric approaches.

The geometric settings is as follows. Recall that Λ_d is the $\mathbb{Q}(q)$-linear space spaned by

$$\{v_0, v_1, \ldots, v_d\}.$$

Correspondingly, we have Grassmannian varieties,

$$X_d^r = \{V \subset \mathbb{C}^d \mid \dim V = r\}, \qquad 0 \le r \le d.$$

Consider the partial flag varieties

$$X_d^{r,r+1} = \{V_1 \subset V_2 \subset \mathbb{C}^d \mid \dim V = r,\ \dim V_2 = r+1\}.$$

The obvious projections

$$X_d^r \xleftarrow{\quad p \quad} X_d^{r,r+1} \xrightarrow{\quad p' \quad} X_d^{r+1}$$

endow $X_d^{r,r+1}$ with a $\mathbb{P}^{d-r-1}$-bundle structure over X_d^r and a $\mathbb{P}^r$-bundle structure over X_d^{r+1}. In this way, both cohomology rings $H^\bullet(X_d^r, \mathbb{C})$ and $H^\bullet(X_d^{r+1}, \mathbb{C})$ act on the cohomology groups $H^\bullet(X_d^{r,r+1}, \mathbb{C})$.

To categorify the simple $U_q(sl_2)$-module Λ_d, we need to construct an abelian category as well as a collection of exact endofunctors.

First, we define a finite-dimensional graded $\mathbb{C}$-algebra

$$A^\bullet = \oplus_r H^\bullet(X_d^r, \mathbb{C}).$$

The abelian category involved is the category $A^\bullet$-mof of finite-dimensional graded $A^\bullet$-modules. By the Grothendieck group of $A^\bullet$-mof, we mean the $\mathbb{Q}(q)$-linear space $\mathrm{G}(A^\bullet\text{-mof})$ which is defined by the generators each for an isomorphism class of graded $A^\bullet$-modules and the relations

- $[M^\bullet] = [M'^\bullet] + [M''^\bullet]$, for a short exact sequence $M'^\bullet \hookrightarrow M^\bullet \twoheadrightarrow M''^\bullet$;
- $[M^{\bullet+1}] = q^{-1}[M^\bullet]$, for $M^\bullet \in A^\bullet$-mof.

Notice the isomorphism

$$\begin{array}{ccc} \Lambda_d & \xrightarrow{\sim} & \mathrm{G}(A^\bullet\text{-mof}) \\ v_r & \mapsto & [H^\bullet(X_d^r, \mathbb{C})]. \end{array}$$

Next, we describe the exact endofunctors. Recall that each flat $A^\bullet$-bimodule defines an exact functor of $A^\bullet$-mof by tensoring on the left. We have following flat (indeed projective) $A^\bullet$-bimodules

$$\begin{aligned} \mathcal{K}^\bullet &= \oplus_r H^{\bullet-d+2r}(X_d^r, \mathbb{C}), \\ \mathcal{K}^{-1\bullet} &= \oplus_r H^{\bullet+d-2r}(X_d^r, \mathbb{C}), \\ \mathcal{E}^\bullet &= \oplus_r H^{\bullet+d-r-1}(X_d^{r,r+1}, \mathbb{C}), \\ \mathcal{F}^\bullet &= \oplus_r H^{\bullet+r}(X_d^{r,r+1}, \mathbb{C}), \end{aligned}$$

in which $H^\bullet(X_d^{r,r+1}, \mathbb{C})$ is regraded as an $H^\bullet(X_d^r, \mathbb{C})$-$H^\bullet(X_d^{r+1}, \mathbb{C})$-bimodule for $\mathcal{E}^\bullet$ and as an $H^\bullet(X_d^{r+1}, \mathbb{C})$-$H^\bullet(X_d^r, \mathbb{C})$-bimodule for $\mathcal{F}^\bullet$. These are the exact endofunctors we need.

A straightforward computation shows the following isomorphisms of $A^\bullet$-bimodules (hence isomorphisms of endofunctors of $A^\bullet$-mof)

$$\begin{aligned} &\mathcal{K}^\bullet \otimes \mathcal{K}^{-1\bullet} \cong \mathcal{K}^{-1\bullet} \otimes \mathcal{K}^\bullet \cong A^\bullet, \\ &\mathcal{K}^\bullet \otimes \mathcal{E}^\bullet \cong A^{\bullet-2} \otimes \mathcal{E}^\bullet \otimes \mathcal{K}^\bullet, \\ &\mathcal{K}^\bullet \otimes \mathcal{F}^\bullet \cong A^{\bullet+2} \otimes \mathcal{F}^\bullet \otimes \mathcal{K}^\bullet, \end{aligned}$$

and

$$\begin{aligned} &A^{\bullet-1} \otimes \mathcal{E}^\bullet \otimes \mathcal{F}^\bullet \oplus A^{\bullet+1} \otimes \mathcal{F}^\bullet \otimes \mathcal{E}^\bullet \oplus \mathcal{K}^\bullet \\ \cong\ &A^{\bullet+1} \otimes \mathcal{E}^\bullet \otimes \mathcal{F}^\bullet \oplus A^{\bullet-1} \otimes \mathcal{F}^\bullet \otimes \mathcal{E}^\bullet \oplus \mathcal{K}^{-1\bullet} \end{aligned}$$

Comparing them with the defining relations of $U_q(sl_2)$

$$\begin{aligned}
&KK^{-1} = K^{-1}K = 1,\\
&KE = q^2EK,\\
&KF = q^{-2}FK,\\
&qEF + q^{-1}FE + K^{-1} = q^{-1}EF + qFE + K,
\end{aligned}$$

we conclude that the abelian category $A^\bullet$-mof and the exact functors

$$\mathcal{K}^\bullet \otimes -, \quad \mathcal{K}^{-1\bullet} \otimes -, \quad \mathcal{E}^\bullet \otimes -, \quad \mathcal{F}^\bullet \otimes -$$

categorify the $U_q(sl_2)$-module Λ_d.

Before proceeding to categorify tensor products of $U_q(sl_2)$-modules, we give a sheaf-theoretic interpretation of the above constructions.

Notice the algebra isomorphism

$$H^\bullet(X_d^r, \mathbb{C}) \quad = \quad \mathrm{Ext}^\bullet_{Sh(X_d^r)}(\mathbb{C}_{X_d^r}, \mathbb{C}_{X_d^r}).$$

where $Sh(\cdot)$ denotes the category of $\mathbb{C}$-sheaves, $\mathbb{C}_{X_d^r}$ denotes the constant sheaf on X_d^r. The multiplication in the right hand side is given by the Yoneda product.

Besides, we have a linear isomorphism

$$\begin{aligned}
H^\bullet(X_d^{r,r+1}, \mathbb{C}) &= \mathrm{Ext}^\bullet_{Sh(X_d^{r,r+1})}(\mathbb{C}_{X_d^{r,r+1}}, \mathbb{C}_{X_d^{r,r+1}})\\
&= \mathrm{Ext}^\bullet_{Sh(X_d^{r,r+1})}(p^*\mathbb{C}_{X_d^r}, p'^*\mathbb{C}_{X_d^{r+1}})
\end{aligned}$$

such that the canonical actions of the algebras $H^\bullet(X_d^r, \mathbb{C})$, $H^\bullet(X_d^{r+1}, \mathbb{C})$ on the left hand side and the canonical actions of the algebras $\mathrm{Ext}^\bullet_{Sh(X_d^r)}(\mathbb{C}_{X_d^r}, \mathbb{C}_{X_d^r})$, $\mathrm{Ext}^\bullet_{Sh(X_d^{r+1})}(\mathbb{C}_{X_d^{r+1}}, \mathbb{C}_{X_d^{r+1}})$ on the right hand side are compatible with the above algebra isomorphisms.

To summarize, we can rewrite as follows.

$$\begin{aligned}
A^\bullet &= \oplus_r \mathrm{Ext}^\bullet_{Sh(X_d^r)}(\mathbb{C}_{X_d^r}, \mathbb{C}_{X_d^r}),\\
\mathcal{K}^\bullet &= \oplus_r \mathrm{Ext}^{\bullet-d+2r}_{Sh(X_d^r)}(\mathbb{C}_{X_d^r}, \mathbb{C}_{X_d^r}),\\
\mathcal{E}^\bullet &= \oplus_r \mathrm{Ext}^{\bullet+d-r-1}_{Sh(X_d^{r,r+1})}(p^*\mathbb{C}_{X_d^r}, p'^*\mathbb{C}_{X_d^{r+1}}),\\
\mathcal{F}^\bullet &= \oplus_r \mathrm{Ext}^{\bullet+r}_{Sh(X_d^{r,r+1})}(p'^*\mathbb{C}_{X_d^{r+1}}, p^*\mathbb{C}_{X_d^r}),
\end{aligned}$$

With slight modifications to the expressions above, we will use them in Section 5 to categorify tensor products of $U_q(sl_2)$-modules.

4. Intersection complex

The notion of intersection complex backdated to Goresky-MacPherson[10,11] trying to seek a new homology theory for singular spaces such that important properties of usual homology theory for smooth manifolds, such as the Poincaré duality, still hold.

Intersection complexes of certain singular varieties (Shurbert varieties) will provide us key ingredients for our geometric categorification.

For simplicity, we assume X is a projective complex variety of pure dimension n. Choose a stratification

$$X = S_0 \sqcup S_1 \sqcup \cdots \sqcup S_n$$

in which each stratum S_i is a smooth subvariety of dimension $n - i$. One obvious choice is that S_0 is the regular part of X and, inductively, S_i is the regular part of $X \setminus S_{i-1}$.

Let $C^\bullet(X)$ be the chain complex for computing the usual cohomology $H^\bullet(X, \mathbb{C})$ (either the simplicial chain complex or the singular chain complex). We say a cochain $\xi \in C^i(X)$ is *allowable* if the following *transversality condition* is satisfied.

$$\dim_{\mathbb{R}}(|\xi| \cap S_k) \leq i - k - 1, \qquad \text{for } k = 1, 2, \ldots, n.$$

The *intersection chain complex* is the subcomplex of $C^\bullet(X)$ formed by

$$IC^i(X) = \{\xi \in C^i(X) \mid \text{both } \xi,\ \partial\xi \text{ are allowable}\}.$$

Its homology defines the *intersection cohomology* $IH^\bullet(X)$ of X.

For example, the usual cohomology groups and the intersection cohomology groups of $\mathbb{P}^1 \vee \mathbb{P}^1$ are computed as follows.

i	$H^i(\mathbb{P}^1 \vee \mathbb{P}^1, \mathbb{C})$	$IH^i(\mathbb{P}^1 \vee \mathbb{P}^1)$
0	$\mathbb{C}$	$\mathbb{C} \oplus \mathbb{C}$
1	0	0
2	$\mathbb{C} \oplus \mathbb{C}$	$\mathbb{C} \oplus \mathbb{C}$

Note that the Poincaré duality is violated in the usual cohomology but persists in the intersetion cohomology.

This phenomenon is indeed a general fact.

Theorem 4.1 (Goresky-MacPherson[10,11]). *(1) $IH^i(X)$ is independent of the stratification $\{S_i\}_{0 \leq i \leq n}$, hence is a topological invariant of X.*

(2) Poincaré duality holds

$$IH^i(X) \cong IH^{2n-i}(X).$$

Corollary 4.1. *$IH^i(X) = H^i(X, \mathbb{C})$ if X is smooth.*

The intersection cohomology can be better understood and treated in the sheaf-theoretic setting. Recall that the usual cohomology $H^\bullet(X, \mathbb{C})$ is the hyper-cohomology of the constant sheaf $\mathbb{C}_X \in Sh(X)$. It turns out that the intersection cohomology $IH^\bullet(X)$ is the hyper-cohomology of a perverse sheaf $\mathcal{IC}(X) \in Perv(X)$, the *intersection complex* of X. More generally, for every closed subvariety S of X, there is a perverse sheaf $\mathcal{IC}(S) \in Perv(X)$ (supported on S), whose hyper-cohomology computes $IH^\bullet(S)$.

The subject of perverse sheaf has been too far from the scope of this note. We refer the readers to Refs. 7,9 for accessible introductions and to Refs. 1,3 for further details. We only mention here that the category of perverse sheaves $Perv(X)$ has many resemblance with $Sh(X)$, so it is helpful to keep in mind that $Perv(X), \mathcal{IC}(X)$ are analogues of $Sh(X), \mathbb{C}_X$ just as $IH^\bullet(X)$ is an analogue of $H^\bullet(X, \mathbb{C})$.

5. Tensor products of $U_q(sl_2)$-modules

Let the varieties X_d^r, $X_d^{r,r+1}$ and the morphisms p, p' be the same as Section 3. We fix a composition $\mathbf{d} = (d_1, d_2, \ldots, d_l)$ of $d = \sum d_i$. Recall that the tensor product

$$\Lambda_{\mathbf{d}} = \Lambda_{d_1} \otimes \Lambda_{d_2} \otimes \cdots \otimes \Lambda_{d_l}$$

has a standard basis

$$\{v_{r_1} \otimes v_{r_2} \otimes \cdots \otimes v_{r_l} \mid 0 \leq r_i \leq d_i\}.$$

The general linear group $GL(\mathbb{C}^d)$ acts on the Grassmannians X_d^r, $0 \leq r \leq d$ and so does its parabolic subgroup

$$P_{\mathbf{d}} = \left\{ \begin{pmatrix} P_1 & & & * \\ & P_2 & & \\ & & \ddots & \\ 0 & & & P_l \end{pmatrix} \mid P_i \in GL(\mathbb{C}^{d_i}) \right\}.$$

A key observation is the obvious one-one correspondence

$$\{P_{\mathbf{d}}\text{-orbits of } \sqcup_r X_d^r\} \leftrightarrow \text{the standard basis of } \Lambda_{\mathbf{d}}.$$

The closures of the $P_{\mathbf{d}}$-orbits are specific examples of Schubert varieties; they are singular varieties, in general.

Let L^r denote the direct sum of the intersection complexes of the $P_{\mathbf{d}}$-orbits of X_d^r

$$L^r = \oplus_S \mathcal{IC}(\overline{S}) \in Perv(X_d^r).$$

Define a finite-dimensional graded (non-commutative in general!) $\mathbb{C}$-algebra

$$A^\bullet = \oplus_r \mathrm{Ext}^\bullet_{Perv(X_d^r)}(L^r, L^r),$$

as well as $A^\bullet$-bimodules

$$\begin{aligned} \mathcal{K}^\bullet &= \oplus_r \mathrm{Ext}^{\bullet-d+2r}_{Perv(X_d^r)}(L^r, L^r), \\ \mathcal{E}^\bullet &= \oplus_r \mathrm{Ext}^{\bullet+d-r-1}_{Perv(X_d^{r,r+1})}(p^* L^r, p'^* L^{r+1}), \\ \mathcal{F}^\bullet &= \oplus_r \mathrm{Ext}^{\bullet+r}_{Perv(X_d^{r,r+1})}(p'^* L^{r+1}, p^* L^r). \end{aligned}$$

The main result of [18, Section 3] can be stated as follows.

Theorem 5.1. *There is a canonical linear isomorphism*

$$\Lambda_{\mathbf{d}} \cong \mathrm{G}(A^\bullet\text{-}mof) \tag{$*$}$$

and isomorphisms of projective $A^\bullet$-bimodules

$$\begin{aligned} \mathcal{K}^\bullet \otimes \mathcal{K}^{-1\bullet} &\cong \mathcal{K}^{-1\bullet} \otimes \mathcal{K}^\bullet \cong A^\bullet, \\ \mathcal{K}^\bullet \otimes \mathcal{E}^\bullet &\cong A^{\bullet-2} \otimes \mathcal{E}^\bullet \otimes \mathcal{K}^\bullet, \\ \mathcal{K}^\bullet \otimes \mathcal{F}^\bullet &\cong A^{\bullet+2} \otimes \mathcal{F}^\bullet \otimes \mathcal{K}^\bullet, \end{aligned}$$

and

$$\begin{aligned} & A^{\bullet-1} \otimes \mathcal{E}^\bullet \otimes \mathcal{F}^\bullet \ \oplus \ A^{\bullet+1} \otimes \mathcal{F}^\bullet \otimes \mathcal{E}^\bullet \ \oplus \ \mathcal{K}^\bullet \\ \cong \ & A^{\bullet+1} \otimes \mathcal{E}^\bullet \otimes \mathcal{F}^\bullet \ \oplus \ A^{\bullet-1} \otimes \mathcal{F}^\bullet \otimes \mathcal{E}^\bullet \ \oplus \ \mathcal{K}^{-1\bullet}. \end{aligned}$$

Therefore, the abelian category $A^\bullet$-mof and the exact functors

$$\mathcal{K}^\bullet \otimes -, \quad \mathcal{K}^{-1\bullet} \otimes -, \quad \mathcal{E}^\bullet \otimes -, \quad \mathcal{F}^\bullet \otimes -$$

categorify the $U_q(sl_2)$-module $\Lambda_{\mathbf{d}}$.

We conclude this note by several remarks. The linear isomorphism $(*)$ from the above theorem is non-trivial. It can be expressed in terms of parabolic Kazhdan-Lusztig polynomials. The elements

$$[\mathrm{Ext}^\bullet_{Perv(X_d^r)}(L^r, \mathcal{IC}(\overline{S}))], \quad 0 \le r \le d, \ S \in \{P_{\mathbf{d}}\text{-orbits of } X_d^r\}$$

form a basis of $\Lambda_{\mathbf{d}}$, which coincides with the canonical basis introduced by Lusztig.[17]

The bimodule isomorphisms from the above theorem were proved by reduction to sheaf isomorphisms. The main technique involved is the usage of the Decomposition Theorem of Beilinson-Bernstein-Deligne-Gabber.[1]

The work Ref. 18 was partly motivated by a desire to understand Khovanov homology[14] of knots and links. The treatment was inspired by Lusztig's geometric construction of canonical basis of quantum enveloping algebras.[16]

Acknowledgement

I am grateful to the organizers of the conference in memory of Lin Xiao-Song held at Nankai University, both for their invitation and the opportunity to make a contribution to this proceeding. It was Lin Xiao-Song that first introduced to me Khovanov homology of knots and links, which eventually led me to the present work on categorification.

References

1. A. Beilinson, J. Bernstein and P. Deligne, Faisceaux pervers, *Astérisque* **100**, (1982).
2. J. Bernstein, I. Frenkel and M. Khovanov, A categorification of the Temperley-Lieb algebra and Schur quotients of $U(sl_2)$ via projective and Zuckerman functors, *Selecta Math. (N.S.)* **5**, 199–241 (1999).
3. A. Borel, *Intersection Cohomology* (Birkhäuser, Boston, 1984).
4. J. Chuang and R. Rouquier, Derived equivalences for symmetric groups and sl_2-categorification, arXiv:math/0407205.
5. L. Crane, Clock and category: is quantum gravity algebraic?, *J. Math. Phys.* **36**, 6180–6193 (1995).
6. L. Crane and I. Frenkel, Four dimensional topological quantum field theory, Hopf categories, and the canonical bases, *J. Math. Phys.* **35**, 5136–5154 (1994).
7. A. Dimca, *Sheaves in Topology* (Springer, Berlin 2004).
8. I. Frenkel, M. Khovanov and C. Stroppel, A categorification of finite-dimensional irreducible representations of quantum $sl(2)$ and their tensor products, arXiv:math.QA/0511467.
9. S. I. Gelfand and Yu. I. Manin, *Homological algebra* (Springer, Berlin, New York, 1999).
10. M. Goresky and R. MacPherson, Intersection homology theory, *Topology* **19**, 135–162 (1980).
11. M. Goresky and R. MacPherson, Intersection homology II, *Invent. Math.* **72**, 77–129 (1983).
12. M. Kashiwara, On crystal bases of the q-analogue of universal enveloping algebras, *Duke Math. J.* **63**, 465–516 (1991).
13. C. Kassel, *Quantum Groups* (Springer-Verlag, New York, 1995).
14. M. Khovanov, A categorification of the Jones polynomial, *Duke Math. J.* **101**, 359–426 (2000).
15. M. Khovanov, V. Mazorchuk and C. Stroppel, A brief review of abelian categorifications, arXiv:math/0702746.
16. G. Lusztig, Quivers, perverse sheaves, and quantized enveloping algebras, *J. Amer. Math. Soc.* **4**, 365–421 (1991).
17. G. Lusztig, *Introduction to Quantum Groups* (Birkhäuser, Boston, 1993).
18. H. Zheng, A geometric categorification of tensor products of $U_q(sl_2)$-modules, arXiv:math/0705.2630.

PART B

Xiao-Song Lin's Unpublished Papers

Editors' Notes

In this part, we included 5 unpublished papers of Xiao-Song Lin. The recent paper on *the motion group of the unlink and its representations* starts an interesting generalization of the braid groups to higher dimensions. It has been used by other mathematicians to study statistics of extended objects in dimension 4. Some papers dated back to 1991. We did not correct mistakes or update the papers. Readers may continue to find ideas in these papers interesting and discover new theorems following his ideas, which will be the best way to remember Xiao-Song.

Markov Theorems for Links in 3-Manifolds

We describe here a Markov theorem for links in $S^2 \times S^1$ using the braid groups corresponding to the Coxeter groups of type B_ℓ. See the theorem in section 3. We will work on a more general setting. The only reason here to single out $S^2 \times S^1$ is its simplicity. It is possible that all the other Artin groups will turn out to be useful in the study of links in 3-manifolds.

1. Coxeter groups and their corresponding braid groups

Let W be a Coxeter group. It has presentation

$$W = \langle r \in R; r^2 =, (rs)^{m_{rs}} = 1 \rangle$$

where $m_{rs} = m_{sr} \geq 3$ for $r, s \in R$ and $r \neq s$. Then, B_W, the braid group corresponding to W, is given by

$$B_W = \langle \sigma_r; r \in R, \sigma_r \sigma_s \cdots = \sigma_s \sigma r \cdots \rangle.$$

These groups have been studied extensively (see,[BrD]). They are called Artin groups by some authors.

The classical braid group with n strands, $\mathfrak{B}_n$, is the braid group corresponding to the Coxeter group whose Dynkin diagram is of the type A_{n-1}. We will study the braid group corresponding to the Coxeter group with the type B_n Dynkin diagram (or the braid group of type B_n for simplicity).

Denote the generators of the braid group of type B_n by $\sigma_0, \sigma_1, \ldots, \sigma_{n-1}$. Then the generating relations are

(1) $\sigma_i \sigma_j = \sigma_j \sigma_i$ if $|i - j| \geq 2$;
(2) $\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}$ for $i = 1, \ldots, n-2$; and
(3) $\sigma_0 \sigma_1 \sigma_0 \sigma_1 = \sigma_1 \sigma_0 \sigma_1 \sigma_0$.

We will denote by $\mathfrak{C}_n$ the braid group of type B_n. Then it is easy to see that we have natural inclusions

$$\mathfrak{C}_2 \subset \mathfrak{C}_3 \subset \cdots \subset \mathfrak{C}_n \subset \mathfrak{C}_{n+1} \subset \cdots.$$

We define a homomorphism $\phi : \mathfrak{C}_n \to C_{n+1}$ which will be useful later. We define

$$\begin{aligned}\phi(\sigma_0) &= \sigma_1\sigma_0\sigma_1,\\ \phi(\sigma_1) &= \sigma_2\\ &\vdots\\ \phi(\sigma_{n-1}) &= \sigma_n.\end{aligned}$$

Lemma 1.1. *ϕ extends to a homomorphism $\mathfrak{C}_n \to \mathfrak{C}_{n+1}$.*

Proof. It is easy to check that the relations (1) and (2) are preserved under the map ϕ. Let us check the relation (3). We have

$$\begin{aligned}\phi(\sigma_0\sigma_1\sigma_0\sigma_1) &= \sigma_1\sigma_0\sigma_1\sigma_2\sigma_1\sigma_0\sigma_1\sigma_2\\ &= \sigma_1\sigma_0\sigma_2\sigma_1\sigma_2\sigma_0\sigma_1\sigma_2\\ &= \sigma_1\sigma_2\sigma_0\sigma_1\sigma_0\sigma_2\sigma_1\sigma_2\\ &= \sigma_1\sigma_2\sigma_0\sigma_1\sigma_0\sigma_1\sigma_2\sigma_1,\end{aligned}$$

and

$$\begin{aligned}\phi(\sigma_1\sigma_0\sigma_1\sigma_0) &= \sigma_2\sigma_1\sigma_0\sigma_1\sigma_2\sigma_1\sigma_0\sigma_1\\ &= \sigma_2\sigma_1\sigma_0\sigma_2\sigma_1\sigma_2\sigma_0\sigma_1\\ &= \sigma_2\sigma_1\sigma_2\sigma_0\sigma_1\sigma_0\sigma_2\sigma_1\\ &= \sigma_1\sigma_2\sigma_1\sigma_0\sigma_1\sigma_0\sigma_2\sigma_1\\ &= \sigma_1\sigma_2\sigma_0\sigma_1\sigma_0\sigma_1\sigma_2\sigma_1.\end{aligned}$$

So ϕ extends to a homomorphism $\mathfrak{C}_n \to \mathfrak{C}_{n+1}$ □

We will see that this homomorphism $\phi : \mathfrak{C}_n \to \mathfrak{C}_{n+1}$ is quite natural when the following geometrical interpretation of the group $\mathfrak{C}_n$ is established. And it turns out that ϕ is a homomorphism.

Let $\mathfrak{B}_{n+1}$ be the classical braid generated by $\alpha_0, \alpha_1, \ldots, \alpha_{n-1}$. The generator α_i is represented by a braid.

It is easy to see that the two braids are isotopic where the left braid represents $\alpha_0^2\alpha_1\alpha_0^2\alpha_1$ and the right braid represents $\alpha_1\alpha_0^2\alpha_1\alpha_0^2$. Thus, we can define a homomorphism $\mathfrak{C}_n \to \mathfrak{B}_{n+1}$ by sending σ_0 to α_0^2 and sending σ_i to α_i for $i = 1, 2, \ldots, n-1$.

Proposition 1.1. *The homomorphism $\mathfrak{C}_n \to \mathfrak{B}_{n+1}$ defined above is a monomorphism, i.e. the subgroup of $\mathfrak{B}_{n+1}$ generated by $\alpha_0^2, \alpha_1, \ldots, \alpha_{n-1}$ is the braid group of type B_n*

Proof. For any finite Coxeter group W of rank ℓ, let $\mathcal{A}$ be the set of complexified reflecting hyperplanes of W of $\mathbb{C}^\ell$. Write

$$M_W = \mathbb{C}^\ell \setminus \bigcup_{H \in \mathcal{A}} H$$

for the corresponding hyperplane complement on which W acts freely. Then, we have $B_W = \pi_1(M_W/W)$. See[Br] or.[D]

Let W be the Coxeter group of type B_n generated by $r_0, r_1, \ldots, r_{n-1}$. The action of W of $\mathbb{C}^n$ is generated by

$$r_0 : (z_1, z_2, \ldots, z_n) \mapsto (-z_1, z_2, \ldots, z_n)$$

and permutations of the coordinates given by

$$r_i : (\ldots, z_i, z_{i+1}, \ldots) \mapsto (\ldots, z_{i+1}, z_i, \ldots)$$

for $i = 1, \ldots, n-1$. So

$$M_W = \{(z_1, \ldots, z_n) \in (\mathbb{C}^*)^n ; z_i \neq \pm z_j \text{ if } i \neq j\}$$

where $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$.

The normal subgroup N of W generated by $(z_1, z_2, \ldots, z_n) \mapsto (-z_1, z_2, \ldots, z_n)$ is abelian and W/N is isomorphic to the symmetric group S_n. Thus, S_n acts freely on M_W/N. Let $\mathbb{C}^*/\sim$ be the quotient of $\mathbb{C}^*$ by the reflection $z \mapsto -z$. Then the space M_W/N is the product of n copies of $\mathbb{C}^*/\sim$ and S_n acts on it by permuting the factors.

Let

$$X = \{(z_1, \ldots, z_n) \in (\mathbb{C}^*)^n ; z_i \neq z_j \text{ if } i \neq j\}.$$

Then the symmetric group S_n also acts freely on X. The diffeomorphism $\mathbb{C}^*/\sim \to \mathbb{C}^*$ extends to an S_n-equivalent diffeomorphism $M_W/N \to X$. Thus, we get an isomorphism

$$\pi_1(M_W/W) \to \pi_1(X/S_n).$$

If we interpret a classical braid in $\mathfrak{B}_{n+1}$ as an ambient isotopy of $n+1$ distinct points in the complex plain $\mathbb{C}$, and element in $\pi_1(X/S_n)$ can be thought of as a braid in $\mathfrak{B}_{n+1}$ with the point $0 \in \mathbb{C}$ keeping fixed. Thus $\pi_1(X/S_n)$ is the subgroup of $\mathfrak{B}_{n+1}$ generated by $\alpha_0^2, \alpha_1, \ldots, \alpha_{n-1}$. It is also quite clear that $\alpha_0^2, \alpha_1, \ldots, \alpha_{n-1}$ are the standard generators of the braid group of type B_n. This proves our theorem $\square$

It seems that X/S_n is a more natural configuration space for $\mathfrak{C}_n$. And it is not hard to write down directly a presentation of $\pi_1(X/S_n)$. To illustrate this, let us work out the case $n = 2$ explicitly.

In the case $n = 2$, X is the complement of the complex curves $z_1 = 0$, $z_2 = 0$ and $z_1 = z_2$ in $\mathbb{C}^2 = \{(z_1, z_2)\}$. The symmetric group S_2 is generated by the reflection $\tau : (z_1, z_2) \mapsto (z_2, z_1)$.

Let $B^4 = \{(z_1, z_2); |z_1|^2 + |z_2|^2 \leq 1\}$ be the unit ball and $\partial B^4 = S^3 = \{(z_1, z_2); |z_1|^2 + |z_2|^2 = 1\}$ be the unit sphere. Let

$$Y = \Sigma^3 \setminus (S^3 \cap (\{z_1 = 0\} \cup \{z_2 = 0\} \cup \{z_1 = z_2\})).$$

Then X and Y are equivariantly homotopy equivalent. So we have

$$\pi_1(X/\tau) \cong \pi_1(Y/\tau).$$

Notice that Y is the complement of the link in the left part of Figure 1.3 in S^3. Moreover, τ restricted on S^3 is the standard rotation of 180 degrees with $S^3 \cap \{z_1 = z_2\}$ as the axis. Thus Y/τ is the complement of the link in the right part of Figure 1.3 in S^3.

Let us write down the Wirtinger presentation for $\pi_1(Y/\tau)$. In this presentation, we have generators a, b, c, d, e and relations

(1) $bca^{-1}c^{-1} = 1$,
(2) $eb^{-1}e^{-1}a = 1$,
(3) $c^{-1}a^{-1}da = 1$,
(4) $cd^{-1}c^{-1}e = 1$,
(5) $ae^{-1}a^{-1}c = 1$.

From the relations (1), (3) and (4), we get

$$\begin{aligned} b &= cac^{-1}, \\ d &= aca^{-1}, \\ e &= cdc^{-1} = caca^{-1}c^{-1}. \end{aligned}$$

So (2) becomes

$$\begin{aligned} 1 &= eb^{-1}e^{-1}a \\ &= caca^{-1}c^{-1}ca^{-1}c^{-1}cac^{-1}a^{-1}c^{-1}a \\ &= caca^{-1}c^{-1}a^{-1}c^{-1}a, \end{aligned}$$

or

$$caca = acac.$$

Similarly, (5) is equivalent to

$$caca = acac.$$

Back to $\pi_1(X/\tau)$, we have $\alpha_0^2 = c$ and $\alpha_1 = a$. So $\pi_1(X/\tau)$ has a presentation

$$\langle \alpha_0^2, \alpha_1 : \alpha_0^2\alpha_1\alpha_0^2\alpha 1 = \alpha_1\alpha_0^2\alpha_1\alpha_0^2 \rangle.$$

Now the homomorphism $\phi : \mathfrak{C}_n \to \mathfrak{C}_{n+1}$ can be obtained from the following geometrical operation: First split the 0-th strand into two strands without twisting, and name the resulting two endpoints by 0 and 1. Then rename the original i-th endpoint by $i+1$ for $i = 1, 2, \ldots, n-1$. This operation changes an n-braid into an $(n+1)$-braid and gives rise to the homomorphism $\phi : \mathfrak{C}_n \to \mathfrak{C}_{n+1}$. It is easy to see now that ϕ is a monomorphism.

2. Study links via open book decompositions

Let M be a closed, connected 3-manifold. An open book decomposition of M consists of a collection A of disjoint circles, called the *binding*, and a fibration $p : M \setminus A \to S^1$. The fibres are called the *pages*. We also assume that the fibration is well-behaved near A, i.e. that A has a tubular neighborhood $A \times D^2$ so that p restricted to $A \times (D^2 \setminus \{0\})$ is the map $(x, y) \to y/|y|$. Thus, a fibred link in M is a special case of an open book decomposition. The difference here is that for an open book decomposition, more than one component of the boundary of a page (fibre) may meet a single component of the binding (link).

On S^3, there is a standard open book decomposition with disk pages. The binding of this open book decomposition is an unknot in S^3.

In,[S] R. Skora developed a theory of braids in an arbitrary 3-manifold with respect to a fibred link in that 3-manifold. There is no essential difficulty in adapting Skora's theory for the general case.

Let M be a closed, connected 3-manifold with an open book decomposition. Denote by $A_1, \ldots, A_k$ the circles in the binding of this open book decomposition.

Definition 2.1. A piecewise transverse link in M is an oriented link $L \subset M \setminus \bigcup_i A_i$ together with a decomposition of L into closed segments $s_1, \ldots, s_n$ such that

(1) $L = \bigcup_i s_i$;
(2) $\mathrm{Int}(s_i) \cap \mathrm{Int}(s_j) = \emptyset$ for $i \neq j$; and
(3) each s_i is transverse to the pages.

Each segment s_i inherits an orientation from L. A segment is called *increasing* if its orientation is consistent with that of S^1 under the projection p, and is called *decreasing* otherwise.

The *height* of a piecewise transverse link is the number of decreasing segments.

Let us reconstruct M as follows: There is a compact surface F and an automorphism $\psi : F \to F$ such that

$$\overline{M \setminus \bigcup_i (A_i \times D^2)}$$

is the mapping torus of F, i.e.

$$\overline{M \setminus \bigcup_i (A_i \times D^2)} = \frac{F \times [0,1]}{\sim}$$

where $(x,0) \sim (\psi(x),1)$ for all $x \in F$.

Let $b_1, \ldots, b_\ell$ be the boundary components of F. Let D_i be a meridian disk of $A_i \times D^2$. Then it is clear that ∂D_i intersects some $b_j \times 0$ at least once. From this observation, the following lemma is obvious.

Lemma 2.1. *With an appropriate orientation, we can decompose ∂D_i to get a piecewise transverse link with zero height.*

We have some special isotopies between two piecewise transverse links.

Definition 2.2. Let L, L' be two piecewise transverse links. Let s be an increasing segment of L and s'_1, s'_2 increasing segments of L', such that $L' = (L \setminus s) \cup (s'_1 \cup s'_2)$. Assume further that there is a disk D in $M \setminus \bigcup_i A_i$ such that $D \cap L = s$ and $D \cap L' = s'_1 \cup s'_2$. Then we say that L' results from L by an isotopy of type $\mathcal{H}$

Definition 2.3. Let L, L' be two piecewise transverse links. Let s be an increasing segment of L and s' an increasing segment of L', such that $L' = (L \setminus s) \cup s'$. Assume further that there is a disk D as shown in Figure 2.2 which intersects the binding transversally at a single point on A_i. Then we say that L' results from L by an isotopy of type $\mathcal{W}_i$.

With these definitions, we have the following propositions.

Proposition 2.1. *Let L be an oriented link in M, then L is isotopic to a link which can be decomposed into a piecewise transverse link of zero height.*

Proposition 2.2. *Let L, L' be two piecewise transverse links of zero height. Suppose they are isotopic in M. Then there is a sequence of piecewise transverse links*

$$L = L_1, \ldots, L_j, \ldots, L_m = L'$$

such that for each $j = 1, \ldots, m-1$, L_{j+1} results from L_j by an isotopy of type $\mathcal{H}^{\pm}$ or $\mathcal{W}_i^{\pm}$, $i = 1, \ldots, n$.

The proof of these two propositions is basically the same as in.[S] So we will omit it and only point out the following key observation.

Let L and L' be two piecewise transverse links. Suppose there are increasing segments s and s' of L and L' separately such that $L \setminus s = L' \setminus s'$ and there is a disk D with $D \cap L = s$ and $D \cap L' = s'$. Moreover, D intersects the binding transversally at two points, one is on A_i and the other on A_j. Then L' results from L by two isotopies, one is of type $\mathcal{W}_i^{-1}$ and the other of type $\mathcal{W}_j$. Actually, D can be decomposed into two disks D_1 and D_2. The disk D_1 gives us an isotopy of type $\mathcal{W}_i$ whereas the disk D_2 gives us an isotopy of type $\mathcal{W}_j$.

The advantage here of using open book decompositions rather than fibred knots is that for any closed, orientable 3-manifold, there is an open book decomposition whose pages are planar surfaces. This makes it quite easy to formulate Markov theorems for links in such manifolds using various braid groups, at least in some special cases such as $S^2 \times S^1$. Let us discuss the general case first.

Let M be a closed, connected and orientable 3-manifold. Then, M can be obtained by performing Dehn surgery on a link K in S^3. We can assume that the surgery coefficients are all integers. Denote by K^* the link in M dual to K.

Consider the standard open book decomposition of S^3 wth an unknot J as the binding. We can assume K s a closed braid with respect to the braid axis J. Then

$$\overline{S^3 \setminus ((K \times D^2) \cup (J \times D^2))}$$

is a fibration on S^1 whose fibre is a compact, connected planar surface. The boundary of a fibre consists of meridians of $K \times D^2$ and $J \times \{z\}$ with $z \in \partial D^2$. Notice that the surgery on K with the resulting manifold M is done by gluing a disk to each component of $\partial(K \times D^2$ along longitudes of $K \times D^2$ (i.e. a simple closed curve on a component of $\partial(K \times D^2)$ which intersects a corresponding meridian transversally at a single point) and then fill up the sphere boundary components with balls. Thus, the fibration on $\overline{S^3 \setminus ((K \times D^2) \cup (J \times D^2))}$ can be extended to

$$\overline{M \setminus (K^* \cup J)}$$

which satisfies the requirement for an open book decomposition near $K^* \cup J$. So we get an open book decomposition of M with the binding $K^* \cup J$. The pages are connected planar surfaces. We now have the following theorem.

Theorem 2.1. *Each link in M is isotopic to a transverse link in $\overline{S^3 \setminus (K \cup J)}$, i.e. it intersects each fibre transversally. Two transverse links are isotopic in M if and only if one results from the other by a sequence of the following two operations or their inverses:*

(1) An isotopy in $\overline{S^3 \setminus (K \cup J)}$ preserving the fibration;
(2) An operation on transverse links described in Figure 2.4.

The operation on transverse links described in Figure 2.4 is analogous to the classical Markov move in the way that the usual braid axis is replaced by any circle in the binding $K \cup J$ of the open book decomposition. Notice that when K is empty, $M = S^3$. In this case, our theorem is the classical Markov theorem (see[B]).

3. The case of $S^2 \times S^1$

$S^2 \times S^1$ can be obtained by performing a 0-framing surgery on an unknot in S^3. Call this unknot K. Let J be another unknot in S^3 such that $K \cup J$ is the Hopf link. Then $\overline{S^3 \setminus (K \cup J)}$ is the product of an open annulus with S^1. Let L be a transverse link in $\overline{S^3 \setminus (K \cup J)}$. Let n be the number of intersection points of L with a fibre. Then we can think of L as representing an element in $\mathfrak{C}_n$ where K corresponding to the 0-th strand. The theorem in section 2 leads to the following definition.

Definition 3.1. A Markov move on

$$\bigcup_{n=2}^{\infty} \mathfrak{C}_n$$

is one of the following operations or its inverse:

(1) change $\beta \in \mathfrak{C}_n$ to one of its conjugates in $\mathfrak{C}_n$;
(2) change $\beta \in \mathfrak{C}_n$ to $\beta\sigma_n^{\pm 1} \in \mathfrak{C}_{n+1}$; and
(3) change $\beta \in \mathfrak{C}_n$ to $\sigma_1^{\pm 1}\phi(\beta) \in \mathfrak{C}_{n+1}$.

Definition 3.2. Let

$$\beta, \beta' \in \bigcup_{n=2}^{\infty} \mathfrak{C}_n,$$

then β and β' are Markov equivalent if there is a sequence of elements in $\bigcup \mathfrak{C}_n$:

$$\beta = \beta_1, \ldots, \beta_j, \ldots, \beta_m = \beta'$$

such that β_{j+1} results from β_j by a Markov move for each $j = 1, \ldots, m-1$.

Now we have

Theorem 3.1. *Isotopy classes of oriented links in $S^2 \times S^1$ are in one-to-one correspondence with Markov equivalent classes of $\bigcup \mathfrak{C}_n$.*

We end up with some remarks.

For links in the lens space $L(p,1)$, we can also get a Markov theorem using $\bigcup \mathfrak{C}_n$ just by replacing the third type Markov move with the following operation:

$(3)_p$ change $\beta \in \mathfrak{C}_n$ to $\sigma_0^p \sigma_1^{\pm 1} \phi(\beta) \in \mathfrak{C}_{n+1}$.

In particular, $L(1, \pm 1) = S^3$. So we can get a Markov theorem for links in S^3 using $\bigcup \mathfrak{C}_n$ instead of $\bigcup \mathfrak{B}_n$.

Let us call the equivalence relation on $\bigcup \mathfrak{C}_n$ generated by (1), (2) and $(3)_p$ the *p-Markov equivalence relation.* Using link polynomials for links in S^3, we can get a class function for (± 1)-Markov classes of $\bigcup \mathfrak{C}_n$. With some modifications, it is possible that this class function will give rise to a class function for p-Markov classes of $\bigcup \mathfrak{C}_n$. This wil give us polynomials invariants for links in $L(p,1)$.

On the other hand, C. Squier[Sq] has generalized the classical Burau representations fo the braid groups $\mathfrak{B}_n$ to all other Artin groups. In particular, we have concrete matrix representations for the groups $\mathfrak{C}_n$. With some modifications, it is also possible that the characteristic polynomials of these matrix representations will give rise to polynomials invariants for links in $L(p,1)$.

One more posisble further approach is to construct "Ocneanu trace" for the other Hecke algebras, in particular, for the Hecke algebras of type B_ℓ. Also, a detailed study of representations f the braid groups of type B_ℓ which arise from the Hecke algebra of type B_ℓ is desirable. See.[J]

We understand that E. Witten has already constructed polynomial invariants for links in 3-manifolds using the machinery of quantum field theory.[W] Our purpose here is to point out that it is possible to find some more accessible ways of constructing such invariants, at least for some simple 3-manifolds.

References

B. J. Birman, *Braids, links and mapping class groups*, Ann. of Math. Studies, vol. 82, Princeton Univ. Press, Princeton, New Jersey, 1974.

Br. E. Brieskon, *Sur les groupes de tresses (d'aprés V. I. Arnold)*, Sém. Bourbaki, 401(1971), Lecture Notes in Math., vol. 317, 1973.

D. P. Deligne, *Les immeubles des groupes de tresses généralisés*, Invent. Math. **17** (1972), 273–302.

J. V. F. R. Jones, *Hecke algebra representations of braid groups and link polynomials*, Ann. of Math. **126** (1987), 335–388.

S. R. Skora, *Braids in 3-manifolds*, preprint.

Sq. C. Squier, *Matrix representations of Artin groups*, preprint.

W. E. Witten, *Quantum field theory and the Jones polynomial*, preprint.

Vertex Models, Quantum Groups, and Vassiliev's Knot Invariants

1. Introduction

In,[2] a deep connection between the seminal work of V. F. R. Jones[4] and that of V. A. Vassiliev[12] has been established. We will try to explore the full extent of this connection in the present paper.

It was soon realized after Jones' revolutionary discovery of his famous knot polynomial in 1985 that behind it there stands a fast developing algebraic theory, namely, the theory of quantum groups. Via the machinery of quantum groups, not only can the original Jones, HOMFLY, and Kauffman polynomials be derived, there actually is a way of systematically producing families of knot polynomials (see, for example,,[11],[6] and[7]). It is unfortunate, though, that in this approach, knots, rather than being viewed as imbedded circles in the 3-space in the most natural perspective, are taken to be equivalence classes of certain combinatorial objects, namely knot diagrams. Vassiliev in[12] introduced a new way to look at knots as topological objects which led to a scheme of producing (finite, in a certain sense) numerical knot invariants relying on firm topological foundations. In,[2] it was shown that the HOMFLY polynomial and Kauffman polynomial (thought of as families of knot polynomials, with the original Jones polynomial as a member of the HOMFLY family) can be put into Vassiliev's picture in a very natural way. However, the proof there relies heavily on the recursive formulae of the HOMFLY and Kauffman polynomials. It reveals no connection between the algebraic structure behind the HOMFLY and Kauffman polynomials and Vassiliev's theory. Such a connection is what we want to study in this paper.

Recall the notion of (circular) $[i]$-configurations in.[2] They are patterns of pairing $2i$ points on an oriented circle. Let $\mathcal{V}_i$ be the vector space (over $\mathbb{C}$, say) spanned by all $[i]$-configurations and $\mathcal{V} = \bigoplus \mathcal{V}_i$. A *VBL-funtional* is a linear functional on $\mathcal{V}$ satisfying two conditions. The first condition says that if an $[i]$-configuration contains a pair consisting of two adjacent points on the circle, then the value of a VBL-functional on this $[i]$-configuration should be zero. The second condition is more substantial. It plays a role just like the one played by the Yang-Baxter equation in Jones' theory. See Definition 3.1 for details. In Vassiliev's theory, if f is a VBL-functional, then $f|\mathcal{V}_i$ determines an element of the term $E_1^{i,i}$ in a spectral sequence constructed for the cohomology of the so-called knot space. We have the following theorem.

Theorem 1.1. *There is a VBL-functional associated to every irreducible representation of a simple Lie algebra.*

One approach to this theorem already appeared in the thesis of Dror Bar-Natan,[1] which studied the perturbative theory of Witten's Chern-Simons path integrals.[13] This seems to tell the other aspect of a whole story. Notice that in Witten's theory of Chern-Simons path integrals, knots are also treated as topological (geometric) objects.

Our approach to this theorem is based on the representation theory of quantum universal enveloping algebras of simple Lie algebras. The similarity between Bar-Natan's approach and ours will be transparent. Whereas Bar-Natan deals with the first-order approximations of Chern-Simons path integrals, we will deal with the first-order approximations of quantum universal enveloping algebra of simple Lie algebras. On the other hand, our approach seems to be more direct and simpler. It will also suggest a states model for VBL-functionals.

The states model for VBL-functionals seems to provide a nice way of understanding the underlying algebraic structure of Vassiliev's theory. With this states model, we can produce VBL-functionals directly from finite dimensional irreducible representations of quantum universal enveloping algebras of simple Lie algebras without going through the formalism of constructing the generalized Jones polynomials. In spite of that, the proof of Theorem 0.1 is best understood in terms of the relation between generalized Jones polynomials and VBL-functionals established in this paper.

There are two minor problems which will not be discussed in detail in this paper. In the terminology of,[2] if f is a VBL-functional, then $f|\mathcal{V}_i$ corresponds to the top row of an actuality table of order i. An actuality table of order i determines a Vassiliev knot invariant of order i. Our first problem is for a VBL-functional f given by Theorem 0.1, whether $f|\mathcal{V}_i$ can always be extended to a Vassiliev knot invariant of order i, or whether we can complete an actuality table if its top row is given as $f|\mathcal{V}_i$. In Bar-Natan's approach, the answer to this problem is yes in a lot of concrete cases. In general, an affirmative answer requires the related R-matrix to satisfy an additional equation. See the discussion after Theorem 2.11. Another minor problem is whether the VBL-functionals provided by Bar-Natan's approach and our approach are essentially the same. Direct calculation is possible but tedious. Yet it seems to be hard to believe that these two approaches would not produce essentially the same VBL-functionals.

The major problem here is whether the set of all VBL-functionals is spanned by those produced via the theory of quantum groups or the theory of Chern-Simons path integrals. This problem seems to be very difficult and a new insight is certainly needed if one wants to make a breakthrough. One anticipation is that our states model for VBL-functionals would probably play an important role in the solution of this problem. After all, we tend to believe that Vasiliev's theory is the topological counterpart of the quantum group formalism (algebraic) and Witten's formalism (geometric and analytic) of constructing knot invariants.

This paper is organized as follows. In section 1, we will review some basic facts about quantum groups and their finite dimensional irreducible representations. We will make an observation (Lemma 1.3) about the universal R-matrix which turns out to be quite important in the following discussion. In section 2, we will discuss Jones' vertex models and establish some facts about vertex models derived from irreducible representations of simple Lie algebras. These facts will be useful in section 3. Finally, in section 3, a states model for VBL-functionals will be introduced and we will show how to get VBL-functionals from the knot invariants derived via the machinery of quantum groups.

2. The quantum group $U_h\mathfrak{g}$ and its finite dimensional representations

For every simply complex Lie algebra $\mathfrak{g}$, there is a natural deformation of its universal enveloping algebra $U\mathfrak{g}$ as a Hopf algebra over the formal power series over $\mathbb{C}$. We denote this deformation by $U_h\mathfrak{g}$, which is called a *quantum universal enveloping algebra* or *quantum group*. Following,[3] we define $U_h\mathfrak{g}$ in terms of generators and relations.

Let $\mathfrak{g}$ be a simple Lie algebra of rank l, (a_{ij}) its Cartan matrix, and d_i the length of the ith root.

As a $\mathbb{C}[[h]]$-algebra, $U_h\mathfrak{g}$ is generated by dl elements $H_i, X_i^{\pm}, 1 \le i \le l$

subject to the following relations:

$$[H_i, H_j] = 0,$$
$$[H_i, X_j^{\pm}] = \pm a_{ij} X_j^{\pm},$$
$$[X_i^+, X_j^-] = \delta_{ij} \frac{\sinh((h/2)H_i)}{\sinh(h/2)},$$

and, for $i \neq j$, let $q_i = \exp(hd_i/2)$:

$$\sum_{k=0}^{1-a_{ij}} (-1)^k \binom{1-a_{ij}}{k}_{q_i} q_i^{-k(1-a_{ij}-k)/2} (X_i^{\pm})^k X_j^{\pm} (X_i^{\pm})^{1-a_{ij}-k} = 0$$

where

$$\binom{n}{k}_q = \frac{(q^n - 1)\cdots(q^{n-k+1}-1)}{(q^k-1)\cdots(q-1)}.$$

The comultiplication Δ is defined by

$$\Delta(H_i) = H_i \otimes 1 + 1 \otimes H_i,$$
$$\Delta(X_i^{\pm}) = X_i^{\pm} \otimes \exp((h/4)H_i) + \exp((h/4)H_i) \otimes X_i^{\pm}.$$

This comultiplication gives rise to a Hopf algebra structure on $U_h\mathfrak{g}$. This Hopf algebra is certainly non-commutative and non-cocommutative. Notice that $U_h\mathfrak{g} \equiv U\mathfrak{g} \bmod h$ (by which we mean $U_h\mathfrak{g}/hU_h\mathfrak{g} \cong U\mathfrak{g}$) whereas the reduced Hopf algebra structure on $U\mathfrak{g}$ is still non-commutative but cocommutatative. Nevertheless, $U_h\mathfrak{g}$ is *quasitriangle* (quasitriangular?) in the sense that there is an invertible element $R \in U_h\mathfrak{g} \otimes U_h\mathfrak{g}$ such that

$$R\Delta(g)R^{-1} = P\Delta(g), \forall g \in U_h\mathfrak{g}, \tag{1}$$

where $P \in \mathrm{End}(U_h\mathfrak{g} \otimes U_h\mathfrak{g})$ is the permutation; and

$$(\Delta \otimes \mathrm{id})(R) = R_{13}R_{23}, (\mathrm{id} \otimes \Delta)(R) = R_{13}R_{12}, \tag{2}$$

where $R_{12} = \sum g_1 \otimes g_2 \otimes 1 \in (U_h\mathfrak{g})^{\otimes 3}$ if $R = \sum g_1 \otimes g_2$ and so forth.

Here the property (1.2) implies the Yang-Baxter equation

$$R_{12}R_{13}R_{23} = R_{23}R_{13}R_{12}.$$

Lemma 2.1. *$R \equiv 1 \otimes 1$ mod h.*

Proof. We have an explicit formula for $R \in U_h\mathfrak{g} \otimes U_h\mathfrak{g}$ given by Kirillov-Reshetikhin and Levendorskii-Soibelman (see[10]). From that formula, Lemma 1.3 is quite easy to see. □

Notice that Lemma 1.3 is consistent with the cocommutativity of the Hopf algebra $U\mathfrak{g}$.

Finite dimensional representations of $U_h\mathfrak{g}$ have been studied by several authors. See the references in.[10] Here we think of h as a generic complex variable so that $U_h\mathfrak{g}$ is an algebra over $\mathbb{C}$. Rosso[8] showed that every finite dimensional representation of $U_h\mathfrak{g}$ is completely reducible. Moreover, we can deform every finite dimensional irreducible representation of $U\mathfrak{g}$ to a finite dimensional irreducible representation of $U_h\mathfrak{g}$ and all finite dimensional irreducible representations of $U_h\mathfrak{g}$ are essentially obtained by deforming finite dimensional irreducible representations of $U\mathfrak{g}$ (after possibly tensoring by a 1-dimensional representation).

Let (ρ, V) be a finite dimensional irreducible representation of $U_h\mathfrak{g}$. As indicated by Rosso,[9] there is a basis $\mathbf{e}$ of V such that

$$\mathrm{Mat}(\rho(Tg), \mathrm{e}) = {}^{t}\mathrm{Mat}(\rho(g), \mathrm{e}), \forall g \in U_h\mathfrak{g},$$

where $T : U_h\mathfrak{g} \to U_h\mathfrak{g}$ is the antiautomorphism defined by

$$\begin{aligned} TH_i &= H_i, \\ TX_i^+ &= \frac{\sinh(h/2)}{\sinh(hd_i/2)} X_i^- . \\ TX_i^- &= \frac{\sinh(hd_i/2)}{\sinh(h/2)} X_i^+ . \end{aligned}$$

Such a basis is unique up to a scalar. We call it a *privileged basis* with respect to ρ.

3. Vertex models and quantum groups

In,[5] Jones introduced the notion of vertex models and indicated how to derive regular isotopy invariants of knots from vertex models. See also.[11]

A *vertex model* consists of a vector space V together with a privileged basis e, two families of linear operations $R_\pm(\lambda) = R_\pm(\lambda, h) \in \mathrm{End}(V \otimes V)$ where $\lambda \in (0, \pi)$ is called the *spectral parameter* and $L(\lambda) \in \mathrm{End}(V)$ which is diagonal over the basis $\mathbf{e}$. The triple $\nu = \{\mathrm{e}, R_\pm(\lambda), L(\lambda)\}$ has the following properties:

(3.1) $R_\pm(\lambda + \delta) = (L(-\delta) \otimes \mathrm{id}) R_\pm(\lambda)(L(-\delta) \otimes \mathrm{id})^{-1} = (\mathrm{id} \otimes L(\delta)) R_\pm(\lambda)(\mathrm{id} \otimes L(\delta))^{-1}$.

(3.2) $\check{R}_-(0)\check{R}_+(0) = \mathrm{id}$, where $\check{R}_\pm(\lambda) = PR_\pm(\lambda)$ and $P \in \mathrm{End}(V \otimes V)$ is the permutation.

(3.3) Let $\mathbf{e} = \{\dots, a, b, \dots, x, y, \dots\}$. We write $A \in \mathrm{End}(V \otimes V)$ as

$$A(a \otimes x) = \sum_{b,y} A(a, b|x, y) b \otimes y.$$

The linear operators A^{t_1} and A^{t_2} are then defined to be

$$A^{t_1}(a \otimes x) = \sum_{b,y} A(b, a|x, y) b \otimes y,$$

$$A^{t_2}(a \otimes x) = \sum_{b,y} A(a, b)|y, x) b \otimes y.$$

We should have

$$PR_+^{t_1}(\pi) PR_-^{t_2}(\pi) = \mathrm{id}\,.$$

(3.4) For $A \in \mathrm{End}(V \otimes V)$ as in (2.3), we write $A_{12}, A_{13}, A_{23} \in \mathrm{End}(V \otimes V \otimes V)$ as

$$A_{12}(a \otimes x \otimes c) = \sum_{b,y} A(a, b|x, y) b \otimes y \otimes c$$

and so forth. Then we should have

$$R_{12}(\lambda) R_{13}(\lambda + \delta) R_{23}(\delta) = R_{23}(\delta) R_{13}(\lambda + \delta) R_{12}(\lambda).$$

From (3.1) and (3.2), we see that $R_\pm(\lambda)$ is determined by $R_\pm = R_\pm(0)$ and R_- is determined by R_+.

Proposition 3.1 (Rosso). *Let (ρ, V) be a finite dimensional irreducible representation of $U_h\mathfrak{g}$. Let* $\mathbf{e}$ *be the privileged basis of V with respect to ρ. Let*

(1) $R_+ = \rho \otimes \rho(R)$; and
(2) $L(\lambda) = \rho(\exp(-(h/2\pi)\lambda H_\alpha))$, where α is a half of the sum of all positive roots.

Then $\nu = \{\mathbf{e}, R_\pm, L(\lambda)\}$ is a vertex model.

See.[9]
We write

$$R_\pm(a \otimes x) = \sum_{b,y} R_\pm(a, b|x, y)(\lambda) b \otimes y,$$

$$L(\lambda)(a) = \exp(-(h/2\pi)\lambda\xi_a)a.$$

Let

$$R_\pm(a, b|x, y)(\lambda, h) = R_\pm^0(a, b|x, y)(\lambda) + R_\pm^1(a, b|x, y)(\lambda)(h) + O(h^2).$$

Lemma 3.1. *We have*

$$R^0_\pm(a,b|x,y)(\lambda) = \delta(a,b|x,y) = \begin{cases} 1 & \textit{if } a=b \textit{ and } x=y, \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. This is just a corollary of Lemma 1.1 that $R \equiv 1 \otimes 1 \bmod h$. $\square$

Let

$$\Gamma(a,b|x,y)(\lambda) = \lim_{h\to 0} \frac{R_+(a,b|x,y)(\lambda) - R_-(a,b|x,y)(\lambda)}{h}.$$

Lemma 3.2. $\Gamma(a,b|x,y)(\lambda) = \Gamma(a,b|x,y)(0)$

Proof. By (2.1) we have

$$R_\pm(a,b|x,y)(\lambda) = R_\pm(a,b|x,y)(0)\exp(-(h/2\pi)(\xi_a - \xi_b)\lambda).$$

So

$$\begin{aligned} &\lim_{h\to 0} \frac{R_+(a,b|x,y)(\lambda) - R_-(a,b|x,y)(\lambda)}{h} \\ &\quad = \lim_{h\to 0} \frac{R_+(a,b|x,y)(0) - R_-(a,b|x,y)(0)}{h} \exp(-(h/2\pi)(\xi_a - \xi_b)\lambda) \\ &\quad = \lim_{h\to 0} \frac{R_+(a,b|x,y)(0) - R_-(a,b|x,y)(0)}{h}. \end{aligned}$$ $\square$

We will denote $R_\pm(a,b|x,y) = R_\pm(a,b|x,y)(0)$ and $R^1_\pm(a,b|x,y) = R^1_\pm(a,b|x,y)(0)$. We also denote $\Gamma(a,b|x,y) = \Gamma(a,b|x,y)(0)$. Then

$$\Gamma(a,b|x,y)(0) = R^1_+(a,b|x,y) - R^1_-(a,b|x,y).$$

Lemma 3.3. $R^1_+(a,b|x,y) - R^1_-(a,b|x,y) = 0.$

Proof. From (2.2), we have

$$\sum_{b,y} R_+(a,b|x,y)R_-(y,z|b,c) = \delta(a,c|x,z).$$

Since

$$R_\pm(a,b|x,y) = \delta(a,b|x,y) + R^1_\pm(a,b|x,y)h + O(h),$$

we get

$$\sum_{b,y} \left[R^1_+(a,b|x,y)\delta(y,z|b,c) + \delta(a,b|x,y)R^1_-(y,z|b,c)\right] = 0$$

or

$$R^1_+(a,b|x,y) + R^1_-(x,y|a,b) = 0. \qquad \square$$

Corollary 3.1. *We have*

$$\Gamma(a,b|x,y) = R^1_+(a,b|x,y) + R^1_+(x,y|a,b).$$

In particular,

$$\Gamma(a,b|x,y) = \Gamma(x,y|a,b).$$

Proposition 3.2. *If the vertex model is derived from an irreducible representation of $U_h\mathfrak{g}$ which is the deformation of an irreducible representation of $U\mathfrak{g}$, then the contraction of $\Gamma(a,b|x,y)$ is a scalar matrix, i.e.*

$$\sum_b \Gamma(a,b|b,y) = \mu\delta(a,y)$$

for a certain constant μ.

Proof. Let ρ be an irreducible representation of $U_h\mathfrak{g}$ which is the deformation of an irreducible representation ρ' of $U\mathfrak{g}$. If we write

$$\rho\otimes\rho(R)(a\otimes x) = \sum_{b,y} R_+(a,b|x,y)b\otimes,$$

then

$$\rho\otimes\rho(P(R))(a\otimes x) = \sum_{b,y} R_+(x,y|a,b)b\otimes y.$$

Thus,

$$\begin{aligned} Q(a\otimes x) &= \rho\otimes\rho(P(R)R)(a\otimes x) \\ &= \sum_{b,c,y,z} R_+(y,z|b,c)R_+(a,b|x,y)c\otimes z \end{aligned}$$

and

$$\begin{aligned} Q(a,c|x,z) &= \sum_{b,y} R_+(y,z|b,c)R_+(a,b|x,y) \\ &= \sum_{b,y} \delta(y,z|b,c)\delta(a,b|x,y) \\ &\quad + h\sum_{b,y}\left[\delta(y,z|b,c)R^1_+(a,b|x,y) + R^1_+(y,z|b,c)\delta(a,b|x,y)\right] + O(h^2) \\ &= \delta(a,c|x,z) + h\left[R^1_+(a,c|x,z) + R^1_+(x,z|a,c)\right] + O(h^2). \end{aligned}$$

Therefore,

$$Q = \mathrm{id} + \Gamma h + O(h^2).$$

The condition

$$R\Delta(g)R^{-1} = P\Delta(g), \forall g \in U_h\mathfrak{g}$$

implies

$$P(R)R\Delta(g) = \Delta(g)P(R)R, \forall g \in U_h\mathfrak{g}.$$

We have

$$\rho \otimes \rho(\Delta(g)) = G \otimes \mathrm{id} + \mathrm{id} \otimes G + O(h)$$

where $G = \rho'(g)$ and $g \in U\mathfrak{g}$. Thus

$$\Gamma(G \otimes \mathrm{id} + \mathrm{id} \otimes G) = (G \otimes \mathrm{id} + \mathrm{id} \otimes G)\Gamma.$$

Let the matrix of G be $G(a, x)$. Then the identity

$$\sum_{b,x} G(a,b)\Gamma(b,x|x,z) = \sum_{b,x} \Gamma(a,b|b,x)G(x,z)$$

can be easily verified using the previous identity. Since ρ' is irreducible, we see that $\sum_b \Gamma(a,b|b,x)$ is a scalar matrix by Schur's lemma. □

Given a vertex model $\nu = \{\mathbf{e}, R_\pm, L(\lambda)\}$, Jones defined in[5] a partition function associated with a knot diagram $\mathbf{K}$ which turns out to be a regular isotopy invariant of $\mathbf{K}$.

A *state* of the knot diagram $\mathbf{K}$ is an assignment of an element of $\mathbf{e}$ to each edge of $\mathbf{K}$. At each crossing of $\mathbf{K}$, denote the "ingoing" angle measured in radians by $\lambda \in (0, \pi)$. We then can define the partition function

$$Z_{\mathbf{K}}^{\nu} = \sum_{\text{states}} \left(\prod_{\text{crossings}} R_\pm(a,b|x,y)(\lambda) \right) \exp\left(\frac{h}{2\pi} \int \xi_a d\theta \right). \tag{1}$$

Here ξ_a defines a locally constant function on $\mathbf{K}$ for each state, and $d\theta$ is the pull-back to $\mathbf{K}$ of the angle form on S^1 via the Gauss mapping $\mathbf{K} \to S^1$.

Theorem 3.1 (Jones). *Z_K^ν is independent of λ and defines a regular isotopy invariant of K.*

As pointed out by Jones, if the vertex model satisfies the following additional condition that

$$\sum_a R_+(a,b|x,a)e^{h\xi_a} = \sum_a R_+(x,a|a,b)e^{-h\xi_a} = \delta(b,x),$$

then $Z^{\nu}_{\mathbf{K}}$ is an isotopy invariant of $\mathbf{K}$. It is reasonable to conjecture that if the vertex model is derived from an irreducible representation of $U\mathfrak{g}$, then the above equation should always be true provided that $\delta(b,x)$ is replaced by a constant multiple of it. This occurs quite often, and Proposition 2.10 provides evidence for this conjecture. If this conjecture is true, we would always be able to change the vertex model multiplying $R_+(a,b|x,y)$ by a constant factor so that the resulting partition function is an isotopy invariant.

4. VBL-functionals and their states models

Let us first recall the definition of (circular) $[i]$-configurations and $\langle i\rangle$-configurations in.[2] A $[i]$-configuration is a pairing of $2i$ points on the oriented circle, and a $\langle i\rangle$-configuration consists of a $[i-2]$-configuration and a triple of points on the circle distinct from the underlying $2i-4$ points of that $[i-2]$-configuration. Two $[i]$- or $\langle i\rangle$-configurations are the same if they match up to an orientation-preserving homeomorphism of the circle.

The diagram of a $[i]$-configuration α consists of an oriented circle with $2i$ points on it representing the underlying point set of α and, for each pair of α, a line segment connecting the two points in that pair. The diagram of a $\langle i\rangle$-configuration β is similar with the exception that the points in the triple of β have nothing to be attached to.

Denote by $\mathcal{V}_i$ the vector space (over $\mathbb{C}$) spanned by all $[i]$-configurations, and let $\mathcal{V} = \bigoplus \mathcal{V}_i$. We will consider linear functionals on $\mathcal{V}$.

Suppose β is a $\langle i\rangle$-configuration. One can obtain six $[i]$-configurations from β by first splitting a point in the triple of β into two adjacent points and then pairing the resulting points with the remaining two points in that triple, where we use β_{rs} to denote the resulting $[i]$-configuration. It is understood that there is a common underlying $[i-2]$-configuration in each picture and all the circles are oriented counterclockwise.

Definition 4.1. A VBL-functional f is a linear functional on $\mathcal{V}$ with the following properties:

(4.1) For each $[i]$-configuration α with a pair consisting of two adjacent points on the circle, $f(\alpha)=0$; and

(4.2) For each $\langle i\rangle$-configuration β, we have

$$f(\beta_{10}) - f(\beta_{11}) = f(\beta_{20}) - f(\beta_{21}) = f(\beta_{30}) - f(\beta_{31}).$$

The condition (4.2) was derived by Birman-Lin[2] from the work of Vassiliev[12] on the topology of the discriminant of the space of maps $S^1 \to \mathbb{R}^3$. If

f is a VBL-functional, then $f|_{\mathcal{V}_i}$ determines a stabilized relative homology class in th discriminant. See[2] for details.

Let $\mathbf{e} = \{\ldots, a, b, \ldots, x, y, \ldots\}$ be a finite set, and α an $[i]$-configuration. We will call the pairs in α *vertices* and the arcs between points in the underlying point set on the circle *edges*. A *state* of α is an assignment of elements in $\mathbf{e}$ to edges of α.

Definition 4.2. The pair $\nu = \{\mathbf{e}, \Omega\}$ is called a vertex model for configurations, or simply vertex model, where Ω is a matrix with entries $\Omega(a, b|x, y)$, if it satisfies the following conditions:

(4.4) $\Omega(a, b|x, y) = \Omega(x, y|a, b)$;

(4.5) $\sum_b \Omega(a, b|b, x) = 0$;

(4.6) $\sum_b [\Omega(a, b|x, z)\Omega(b, c|r, t) - \Omega(a, b|r, t)\Omega(b, c|x, z)] = \sum_b [\Omega(a, c|r, b)\Omega(b, t|x, z) - \Omega(a, c|b, t)\Omega(r, b|x, z)]$.

Remark 4.1. Let V be the vector space spanned by $\mathbf{e}$ and $\Omega \in \mathrm{End}(V \otimes V)$. Then we can write the conditions (4.4), (4.5), and (4.6) in an invariant form. In particular, (4.6) can be written as

$$\Omega_{12}\Omega 13 - \Omega_{13}\Omega_{12} = \Omega_{13}\Omega_{23} - \Omega_{23}\Omega_{13}.$$

Suppose we have a matrix Γ with entries $\Gamma(a, b|x, y)$ satisfying (4.4) and (4.6). Instead of (4.5), it satisfies the following condition:

(4.5') $\sum_b \Gamma(a, b|b, x) = \mu\delta(a, x)$

with μ a certain constant. Then we can let

$$\Omega(a, b|c, d) = \Gamma(a, b|x, y) - \mu\delta(a, b|x, y),$$

and it should be easy to verify that $\Omega(a, b|x, y)$ satisfy (4.4), (4.5) and (4.6).

Let $\nu = \{\mathbf{e}, \Omega\}$ be a vertex model. Let α be an $[i]$-configuration. We may define a partition function on α in the following way. For a state of α, the weight of a vertex shown in Figure 3.3 is $\Omega(a, b|x, y)$. Then define

$$Z^{\nu}_{\alpha} = \sum_{\text{states}} \prod_{\text{vertices}} \Omega(a, b|x, y).$$

Notice that this partition function is well-defined because of (4.4).

Remark 4.2. If a vertex model for configurations is given in the invariant form $\nu = \{V, \Omega\}$, it should be easy to see that Z^{ν}_{α} is independent of the choice of basis of V.

Proposition 4.1. *The linear functional f on $\mathcal{V}$ defined by*

$$f(\alpha) = Z^{\nu}_{\alpha}$$

is a VBL-functional.

Proof. The equation (4.5) is used for verifying the first condition in the definition of VBL-functionals and (4.6) for the second. □

Now suppose $\nu = \{\mathbf{e}, R_\pm, L(\lambda)\}$ is a vertex model for knot diagrams derived from an irreducible representation of $U\mathfrak{g}$. Let

$$\Gamma(a,b|x,y) = R_+(a,b|x,y) - R_-(a,b|x,y).$$

Then $\Gamma(a,b|x,y)$ satisfy (4.4), (4.5') and (4.6). The first two equations follow from Corollary 2.9 and Proposition 2.10. The last equation, as it can be verified directly, is just a consequence of the Yang-Baxter equation. This shows that we can get a VBL-functional from each irreducible representation of $U\mathfrak{g}$ as we stated in Theorem 0.1.

We will not present the direct verification of (4.6) for $\Gamma(a,b|x,y)$ since it is quite tedious. The proof of Theorem 0.1 seems to be best understood in terms of the relation, which we are going to establish, between the VBL-functional and the knot invariant both obtained from the vertex $\nu = \{\mathrm{e}, R_\pm, L(\lambda)\}$.

Let $\mathbf{K}$ be a knot diagram. We can collapse i crossings of $\mathbf{K}$ into double points to get a diagram of an immersed S^1. Denote the resulting diagram by $\overline{K}$. This diagram of immersed S^1 determines an $[i]$-configuration as follows. Consider the preimage of the double points of that immersion. It consists of $2i$ points on S^1. There is a natural pairing among these $2i$ points: Two points are paired if they are mapped to a common double point. This gives rise to an $[i]$-configuration α. We say that $\overline{K}$ *respects* α. For any $[i]$-configuration α, there is a diagram of immersed S^1 respecting α. For details, see.[2]

For each double point of $\overline{K}$, there are two ways to resolve it to crossings. Depending on whether the resulting crossing is positive or negative, we will call such a resolution positive or negative, respectively. Thus, we can resolve $\overline{K}$ into 2^i knot diagrams. Denote them by $\overline{K}_p$, $p = 1, \ldots, 2^i$. Let

$$\epsilon_p = (-1)^{\text{number of negative resolutions in } \overline{K}_p}$$

and we will call $\epsilon_p \overline{K}_p$ a *signed resolution* of $\overline{K}$.

Proposition 4.2. *Let $\bar{f}(\alpha) = \sum_{states} \prod_{vertices} \Gamma(a,b|x,y)$, where the sum is over all states of α. Then, we have*

$$\sum_{p=1}^{2^i} \epsilon_p Z^{\nu}_{\overline{K}_p} = \bar{f}(\alpha) h^i + O(h^{i+1}).$$

As a consequence, the linear functional $\tilde{f}$ on $\mathcal{V}$ satisfies (4.2).

Proof. We can express the left hand side of (3.11) in another form:

$$\sum_p \epsilon_p Z^{\nu}_{\overline{K}_p} = \sum_{\text{states}} \left(\prod_{\text{double points}} (R_+(a,b|x,y)(\lambda) - R_-(a,b|x,y)(\lambda)) \right.$$

$$\left. \cdot \prod_{\text{crossings}} R_{\pm}(a,b|x,y)(\lambda) \right) \exp((h/2\pi) \int_{\overline{K}} \xi_a d\theta).$$

It is understood that the sum is over all states of $\overline{K}$ and a state of $\overline{K}$ is as usual an assignment of elements of $\mathbf{e}$ to edges of $\overline{K}$. We say that a state of $\overline{K}$ induces a state of the $[i]$-configuration α if it respects it at every crossing of $\overline{K}$, the two edges on the over-crossing strand have the same assignment and so do the edges on the under-crossing strand. Since

$$R_+(a,b|x,y)(\lambda) - R_-(a,b|x,y)(\lambda) = h\Gamma(a,b|x,y) + O(h^2)$$

and

$$R_{\pm}(a,b|x,y)(\lambda) = \delta(a,b|x,y) + O(h),$$

we see that a state of $\overline{K}$ may contribute to the right hand side of (3.12) mod h^{i+1} only when it induces a state of α. Moreover, the contribution of such a state made to the coefficient of h^i in the right hand side of (3.12) is the same as the contribution the induced state of α made to $\tilde{f}(\alpha)$. Also, the set of all states of α is the same as the set of these induced by states of $\overline{K}$. Thus, (3.11) is true.

Now it becomes quite easy to verify (4.2) for $\tilde{f}$. We only need to compare the eight signed resolutions of the diagrams in Figure 3.4(a) and (b) respectively. These two sets of signed resolutions are the same module Reidemeister moves of type III. This verifies (3.3) for $\tilde{f}$. □

We notice that (3.11) was derived in[2] for the special cases of the HOMFLY and Kauffman polynomials using recursive formulae.

Now since $\Gamma(a,b|x,y)$ satisfy (4.5'), we may change $R_{\pm}$ to

$$R'_{\pm} = e^{\mp \mu h/2} R_{\pm}.$$

Then $\nu' = \{\mathbf{e}, R'_{\pm}, L(\lambda)\}$ is still a vertex model with

$$\Gamma'(a,b|x,y) = \Gamma(a,b|x,y) - \mu\delta(a,b|x,y).$$

Thus,

$$f(\alpha) = \sum_{\text{states}} \prod_{\text{vertices}} \Gamma'(a, b|x, y)$$

is the desired VBL functional.

References

1. D. Bar-Natan, *Perturbative aspects of the Chern-Simons topological quantum field theory*, Ph.D. thesis, Princeton University, 1991.
2. J. S. Birman and X.-S. Lin, *Knot polynomials and Vassiliev's invariants*, preprint, Columbia University, 1991.
3. V. G. Drinfeld, *Quantum groups* Proceedings of I. C. M., Berkeley, Vol. 1, 798–820, 1986.
4. V. F. R. Jones, *Hecke algebra representations of braid groups and link polynomials*, Ann. of Math. **126** (1987), 335–388.
5. V. F. R. Jones, *On knot invariants related to some statistical mechanical models*, Pacific Jour. of Math. **137** (1989), 311–334.
6. N. Yu. Reshetikhin, *Quantum universal enveloping algebras, the Yang-Baxter equation and invariants of links*, I and II, LOMI preprints, Leningrad, 1988.
7. B. Yu. Reshetikhin and V. G. Turaev, *Ribbon graphs and their invariants derived from quantum groups*, Commun. Math. Phys. **127** (1990), 1–26.
8. M. Rosso, *Finite dimensional representations of the quantum analog of the universal enveloping algebra of a complex simple Lie algebra*, Commun. Math. Phys. **117** (1988), 581–593.
9. M. Rosso, *Groupes quantiques et modèles à vertex de V. Jones en théorie des nœudes*
10. M. Rosso, *Représentations des groups quantiques*, Séminaire Bourbaki, 43ème année, 1990-91, no. 744.
11. V. G. Turaev, *The Yang-Baxter equation and invariants of links.* Invent. Math. **92** (1988), 527–553.
12. V. A. Vassiliev, *Cohomology of knot spaces*, Theory of Singularities and Its Applications (ed. V. I. Arnold), Advances in Soviet Math., Vol. 1, AMS, 1990.
13. E. Witten, *Quantum field theory and Jones polynomial*, Commun. Math. Phys. **121** (1989), 351–399.

Knot Invariants and Iterated Integrals

We give precise formulae for the coefficients of Drinfeld's KZ associator in terms of iterated integrals over the unit interval. These formulae are used to calculate Kontsevich's universal knot invariant for $(2,p)$-torus knots up to the 4th order.

There are already several combinatorial descriptions of Kontsevich's universal knot invariant[Ko] in terms of Drinfeld's work on quasi-triangular quasi-Hopf algebras. See,[B2],[C],[LM3] and.[P] Drinfeld's work was presented in,[D1 D2] We only mention here that the category of representations of a quasi-triangular quasi-Hopf algebra is a tensorial category, from which one can construct framed link invariants (see e.g.,[AC] and[RT]). The essential structure of a quasi-triangular quasi-Hopf algebra $\mathcal{A}$ is determined by two objects. One is an element $R \in \mathcal{A} \otimes \mathcal{A}$, called the *R-matrix*, which measures the non-commutativity of $\mathcal{A}$ and the other is an element $\Phi \in \mathcal{A} \otimes \mathcal{A} \otimes \mathcal{A}$, called the *associator*, which measures the non-associativity of $\mathcal{A}$. For the purpose of constructing link invariants, one can always choose R to be very simple; all of the difficulties lie in constructing Φ. In,[D2] Drinfeld constructed an associator Φ_{KZ} using the monodromy of the formal Knizhnik-Zamolochikov connection. He also suggested a combinatorial construction which would yield an associator with rational coefficients. A detailed discussion of a combinatorial construction of such a pair (R, Φ) appeared in.[B2] Also, it was proved in[LM2] that the coefficients of Φ_{KZ} are determined by multiple ζ-numbers.

In Section 1 of this note, we give precise formulae expressing the coefficients of Φ_{KZ} as iterated integrals on the unit interval. Our calculation of the first few coefficients of Φ_{KZ} using these formulae suggests that $\log \Phi_{\mathrm{KZ}}$ might admit a very beautiful expression. In Section 2, we review briefly the combinatorial formalism of Kontsevich's universal knot invariant, essentially following the approach via non-associative tangles (see[B2] and[LM3]). Finally, in Section 3, we exhibit some calulations in simple cases. Hopefully, these calculations will stimulate further interest in this subject.

We would like to thank Dror Bar-Natan for his comments.

1. The associator Φ_{KZ}

In,[D1],[D2] Drinfeld considered the following differential equation

$$G'(t) = \left(\frac{A}{t} + \frac{B}{t-1}\right) G(t) \tag{1}$$

where A, B are commuting symbols, $G(t)$ is a formal power series in A, B with coefficients that are analytic functions of $t, 0 < t < 1$. Geometrically, a solution to (1) is the monodromy of a flat formal connection (the so-called Knizhnik-Zamolochikov connection) on the configuration space of 3 distinct points in $\mathbb{C}$ along the path $(0, t, 1), 0 < t < 1$. Or it is the monodromy of the connection

$$A\frac{dz}{z} + B\frac{dz}{z-1}$$

on $\mathbb{C} \setminus \{0, 1\}$ with values in $\mathbb{C}[[A, B]]$ along the path $z = t, 0 < t < 1$.

Let $G_1(t)$ and $G_2(t)$ be solutions to (1) with the following fixed asymptotic behaviors when $t \to 0$ and $t \to 1$ respectively:

$$G_1(t) \sim t^A = e^{A \log t} \text{ as } t \to 0,$$
$$G_2(t) \sim (1-t)^B = e^{B \log(1-t)} \text{ as } t \to 1.$$

Then Drinfeld's KZ associator is defined to be

$$\Phi_{KZ} = \Phi_{KZ}(A, B) = G_2^{-1}(t) G_1(t)$$

as formal power series in A, B independent of t.

We use another expression for Φ_{KZ}, which appeared in.[LM2] For $a \in (0, 1)$, there is a unique solution to (1) with $G(a) = 1$. We denote the value of this solution at $b \in (0, 1)$ by $\mathbb{Z}_a^b$. Then

$$\Phi_{KZ} = \lim_{\varepsilon \to 0} e^{-B \log \varepsilon} \mathbb{Z}_\varepsilon^{1-\varepsilon} e^{A \log \varepsilon}. \tag{2}$$

Suppose $\xi_1, \xi_2, \ldots, \xi_k$ are 1-forms on $[a, b]$ with values in $\mathbb{C}[[A, B]]$. As the usual iterated integrals (see[Ch]), we denote

$$\int_a^b \xi_k \cdots \xi_2 \xi_1 = \int_{a \le t_1 \le t_2 \le \cdots \le t_k \le b} \xi_k(t_k) \wedge \cdots \wedge \xi_2(t_2) \wedge \xi_1(t_1)$$
$$= \int_a^b \xi_k(t_k) \int_a^{t_{k-1}} \xi_{k-1}(t_{k-1}) \cdots \int_a^{t_3} \xi_2(t_2) \int_a^{t_2} \xi_1(t_1) \in \mathbb{C}[[A, B]].$$

Notice that our order of integrands is reversed compared with that in.[Ch]

Let

$$\omega_0 = \frac{A\,dt}{t},$$
$$\overline{\omega}_0 = \frac{e^{-B\log(1-t)}A\,dt}{t},$$
$$\omega_1 = \frac{B\,dt}{t-1},$$
$$\overline{\omega}_1 = \frac{Be^{A\log t}\,dt}{t-1}.$$

For $W = A^{p_n}B^{q_n}\cdots A^{p_1}B^{q_1}$, we denote

$$I_W(a,b) = \int_a^b \omega_0^{p_n}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1}.$$

Then

$$\mathbb{Z}_a^b = 1 + \sum_W I_W(a,b),$$

where W runs over all monomials in A, B.

Lemma 1.1. *We have*

$$\mathbb{Z}_a^b = e^{A\log b}e^{-A\log a} + \sum_{q_1\neq 0}\left(\int_a^b \omega_0^{p_n}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1-1}\overline{\omega}_1\right)e^{-A\log a}$$
$$= e^{B\log(1-b)}e^{B\log(1-a)} + \sum_{p_n\neq 0} e^{B\log(1-b)}\left(\int_a^b \overline{\omega}_0\omega_0^{p_n-1}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1}\right).$$

Proof. These expressions for $\mathbb{Z}_a^b$ are derived from the formulae

$$\int_a^b \omega_0^k = \frac{1}{k!}(A\log b - A\log a)^k$$
$$\int_a^b \omega_1^k = \frac{1}{k!}(B\log(1-b) - B\log(1-a))^k$$

and integration by parts. □

Lemma 1.2. *We have* $\mathbb{Z}_a^b = \mathbb{Z}_c^b\mathbb{Z}_a^c$.

Proof. This is the usual property of iterated integrals. □

We now use these two lemmas to calculate (2).

First, we choose an arbitrary $w \in (0,1)$. Then,

$$
\begin{aligned}
\mathbb{Z}_{\varepsilon}^{1-\varepsilon} &= \mathbb{Z}_{w}^{1-\varepsilon}\mathbb{Z}_{\varepsilon}^{w} \\
&= \left(e^{B\log\varepsilon}e^{-B\log(1-w)} + \sum_{p_n\neq 0} e^{B\log\varepsilon}\left(\int_w^{1-\varepsilon} \overline{\omega}_0\omega_0^{p_n-1}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1}\right)\right) \\
&\quad\cdot\left(e^{A\log w}e^{-A\log\varepsilon} + \sum_{q_1\neq 0}\left(\int_\varepsilon^{w} \omega_0^{p_n}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1-1}\overline{\omega}_1\right)e^{-A\log\varepsilon}\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\Phi_{\mathrm{KZ}} &= \lim_{\varepsilon\to 0} e^{B\log\varepsilon}\mathbb{Z}_{\varepsilon}^{1-\varepsilon}e^{A\log\varepsilon} \\
&= \left(e^{B\log(1-w)} + \sum_{p_n\neq 0}\left(\int_w^{1} \overline{\omega}_0\omega_0^{p_n-1}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1}\right)\right) \\
&\quad\cdot\left(e^{A\log w} + \sum_{q_1\neq 0}\left(\int_0^{w} \omega_0^{p_n}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1-1}\overline{\omega}_1\right)\right) \\
&= [e^{-B\log(1-w)} + \sum_{p_n\neq 0, q_1\neq 0}\left(\int_w^{1} \overline{\omega}_0\omega_0^{p_n-1}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1-1}\overline{\omega}_1\right)e^{-A\log w} \\
&\quad+\left(\int_w^1 \overline{\omega}_0 e^{A\log t}\right)e^{-A\log w}]\cdot\left(e^{A\log w} + \sum_{q_1\neq 0}\int_0^{w} \omega_0^{p_n}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1-1}\overline{\omega}_1\right) \\
&= [e^{-B\log(1-w)}e^{A\log w} + \sum_{p_n\neq 0, q_1\neq 0}\left(\int_w^{1} \overline{\omega}_0\omega_0^{p_n-1}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1-1}\overline{\omega}_1\right) \\
&\quad+\int_w^1 \overline{\omega}_0 e^{A\log t}]\cdot\left(1 + e^{-A\log w}\sum_{q_1\neq 0}\int_0^{w} \omega_0^{p_n}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1-1}\overline{\omega}_1\right).
\end{aligned}
$$

Notice that

$$
\begin{aligned}
&e^{-B\log(1-w)}e^{A\log w} + \int_w^1 \overline{\omega}_0 e^{A\log t} \\
&\quad= (e^{-B\log(1-w)}-1)(e^{A\log w}-1) + e^{-B\log(1-w)} + e^{A\log w} - 1 \\
&\qquad+\int_w^1 \frac{e^{-B\log(1-t)}-1}{t}Ae^{A\log t}\,dt + \int_w^1 \frac{A}{t}e^{A\log t}\,dt \\
&\quad= (e^{-B\log(1-w)}-1)(e^{A\log w}-1) + e^{-B\log(1-w)} \\
&\qquad+\int_w^1 \frac{e^{-B\log(1-t)}-1}{t}Ae^{A\log t}\,dt
\end{aligned}
$$

and that the limit of the last expression above when $w \to 0$ is

$$1 + \int_0^1 \frac{e^{-B\log(1-t)} - 1}{t} A e^{A\log t}\, dt.$$

Moreover,

$$\lim_{w\to 0} e^{-A\log w} \int_0^w \omega_0^{p_n}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1-1}\overline{\omega}_0 = 0.$$

Combining all of these calculations, we obtain the following theorem.

Theorem 1.1. *We have*

$$\Phi_{KZ} = 1 + \int_0^1 \frac{e^{-B\log(1-t)} - 1}{t} A e^{A\log t}\, dt \qquad (3)$$
$$+ \sum_{p_n\neq 0, q_1\neq 0} \int_0^1 \overline{\omega}_0\omega_0^{p_n-1}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1-1}\overline{\omega}_1.$$

Before we present our calculation of the first few coefficients of Φ_{KZ} using (3), we recall the so-called multiple ζ-numbers. They are defined as

$$\zeta(i_1,\ldots,i_k) = \sum_{0<m_1<\cdots<m_k} \frac{1}{m_1^{i_1}\cdots m_k^{i_k}}.$$

If we write

$$\int_0^1 \omega_0^{p_n}\omega_1^{q_n}\cdots\omega_0^{p_1}\omega_1^{q_1} = (-1)^{q_1+\cdots+q_n}\tau(p_n,q_n,\ldots,p_1,q_1)A^{p_n}B^{q_n}\cdots A^{p_1}B^{q_1},$$

then

$$\tau(p_n,q_n,\ldots,p_1,q_1) = \zeta(\underbrace{1,\ldots,1}_{p_n-1},q_n+1,\underbrace{1,\ldots,1}_{p_{n-1}-1},q_{n-1}+1,\ldots,\underbrace{1,\ldots,1}_{p_1-1},q_1+1).$$

See.[LM1] The following result is obtained by a straightforward calculation using (3).

Theorem 1.2. *We have*

$$\begin{aligned}\Phi_{KZ} = 1 &- \zeta(2)[A,B] - \zeta(3)[A,[A,B]] - \zeta(3)[B,[A,B]] \\ &- \zeta(4)[A,[A,[A,B]]] - \zeta(4)[B,[B,[A,B]]] \\ &- \zeta(1,3)[A,[B,[A,B]]] + \frac{1}{2}\zeta(2)^2[A,B]^2 \\ &+ \textit{higher order terms}.\end{aligned} \qquad (4)$$

It is known that $\log \Phi_{\mathrm{KZ}}$ is a Lie element in A, B. Theorem 1.4 implies that

$$\begin{aligned}\log \Phi_{\mathrm{KZ}} = & -\zeta(2)[A,B] - \zeta(3)[A,[A,B]] - \zeta(3)[B,[A,B]] \\ & -\zeta(4)[A,[A,[A,B]]] - \zeta(4)[B,[B,[A,B]]] \\ & -\zeta(1,3)[A,[B,[A,B]]] + \text{higher order Lie elements.}\end{aligned} \tag{5}$$

It will be interesting to see how far this pattern can be sustained. In fact, it is proved in[LM2] that all coefficients of $\log \Phi_{\mathrm{KZ}}$ are $\mathbb{Q}$-linear combinations of multiple ζ-numbers. R. Hain also has a direct proof of this fact.[H]

2. Non-associative tangles

We set up the combinatorial formalism for calculating Kontsevich's universal knot invariant in this section. Although we essentially follow the approach using non-associative tangles,[B2],[LM3] we would like to emphasize here the distinction between concordance and cobordism. A tangle is a concordance between two 0-dimensional compact submanifolds of $\mathbb{R}^2$. The underlying 1-dimensional compact manifold of this concordance is a cobordism between those two 0-dimensional manifolds. Certainly, a cobordism can be realized by many different concordances. Nevertheless, we may think of the underlying cobordism as the "0th order approximation" of the concordance in question. We may also have "higher order approximations" using the functor between the category of tangles and the category of chord diagrams originated in Kontsevich's construction of his universal knot invariant. Our account here will be very brief and the reader is referred to[B2] and[LM3] for more details.

The objects in the category of non-associative tangles **NAT** are finite ordered sets of oriented points $v_1, \ldots, v_n$ in the plane together with a *parenthesization* on the word $w = v_1 \ldots v_n$. There is only one parenthesization on the empty word or words of length 1. Inductively, if w_1 and w_2 are parenthesized words of length k and l respectively, then $(w_1 w_2)$ is a parenthesization of the word $w_1 w_2$ (forgetting the parenthesizations) of length $k+l$. For example, there are exactly two different parenthesizations on the word $v_1 v_2 v_3$, namely $((v_1 v_2) v_3)$ and $(v_1 (v_2 v_3))$.

The morphisms in **NAT** are generated by *elementary tangles* $\cap$, $\cup$, $\overline{N}$, N, X^+, and X^-. Their domains and targets are specified below:

$\cap$: domain $w = (\cdots (v_i v_{i+1}) \cdots)$, target w' is obtained from w by deleting $(v_i v_{i+1})$ with the inherited parenthesization;

$\cup$: the reverse of $\cap$;

$\overline{N}$: domain $w = (\cdots((w_1w_2)w_3)\cdots)$ where w_1, w_2, and w_3 are parenthesized words, target $w' = (\cdots(w_1(w_2w_3))\cdots)$;

N: the reverse of $\overline{N}$;

X^+: domain $w = (\cdots(v_iv_{i+1})\cdots)$, target $w' = w$ with the orientations of v_i and v_{i+1} switched;

X^-: the reverse of X^+.

Moreover, the strands in an elementary tangle are oriented consistently with the orientations of points in its domain and target.

In general, morphisms of **NAT** are products of elementary tangles modulo isotopies of tangles. Forgetting the parenthesizations, morphisms in **NAT** are simply *concordances* of compact, oriented 0-dimensional submanifolds of the plane with ordered points. It is easy to see that **NAT** is equivalent to the usual category of tangles.

We have another category, **CD**, the *category of chord diagrams*. The objects in **CD** are compact, oriented 0-dimensional manifolds with ordered points. A chord diagram is a *cobordism* by a compact, oriented 1-manifold between two objects of **CD** together with an ambient isotopy class of finitely many pairs of distinct points in the interior of this cobordism. A pair of points in a chord diagram is indicated by a dashed chord connecting these points. Morphisms of **CD** will then be elements of the completed, graded (with grading coming from the number of chords) $\mathbb{C}$-vector spaces generated by chord diagrams subject to the so-called 4-term relations and framing-independence relations (see[B1] and[BL]).

Let $\mathcal{D}_1 \subset \mathrm{Mor}(\emptyset, \emptyset)$ be the completed subspace generated by chord diagrams on a single oriented circle. Suppose δ is a chord diagram with a marked component of the underlying cobordism. Then, for every $\phi \in \mathcal{D}_1$, we may join ϕ to the marked component of δ to get a well-defined connected sum $\delta\#\phi$ thought of as a morphism of **CD**. In particular, $\mathcal{D}_1$ itself is a completed, graded $\mathbb{C}$-algebra.

There is a functor from **NAT** to **CD**. It sends objects in **NAT** to objects in **CD** by forgetting the parenthesization. For morphisms, we only need to define a functor

$$\widehat{Z} : \mathrm{Mor}(\mathbf{NAT}) \to \mathrm{Mor}(\mathbf{CD})$$

for elementary tangles. Some more notation is needed in order to describe the images of elementary tangles under the functor $\widehat{Z}$.

For a tangle T, we denote by $|T|$ the underlying cobordism of T. We have $|X^+| = |X^-|$ and $|N| = |\overline{N}|$. A *trivial cobordism* is simply a product cobordism. We will denote by β^{ij} the *signed* chord diagram on a trivial cobordism β with exactly one chord connecting its ith and jth components, $i \neq j$. To determine the sign, think of β^{ij} as a morphism from $\{v_1, \dots, v_n\}$ to itself. If the orientations of v_i and v_j are the same, the sign of β^{ij} will be $+$. Otherwise, it is $-$.

We now define:

(1) $Z(\cap) = |\cap|$, $Z(\cup) = |\cup|$;

(2) let $\beta = |\overline{N}| = |N|$, then

$$Z(\overline{N}) = \Phi\left(\frac{1}{2\pi\sqrt{-1}}\sum_{i\in I, j\in J}\beta^{ij}, \frac{1}{2\pi\sqrt{-1}}\sum_{j\in J, k\in K}\beta^{jk}\right),$$

$$Z(N) = \Phi\left(\frac{1}{2\pi\sqrt{-1}}\sum_{j\in J, k\in K}\beta^{jk}, \frac{1}{2\pi\sqrt{-1}}\sum_{i\in I, j\in J}\beta^{ij}\right),$$

where $\Phi = \Phi_{\mathrm{KZ}}$ and I, J, K are sets of indices of w_1, w_2, w_3 respectively.

(3) let $\sigma = |X^+| = |X^-| \in \mathrm{Mor}(\{v_1, \dots, v_n\}_{\mathrm{dom}}, \{v_1, \dots, v_n\}_{\mathrm{tar}})$, where the target is obtained from the domain by switching the orientations of v_i and v_{i+1}, and let τ be the trivial cobordism of $\{v_1, \dots, v_n\}_{\mathrm{dom}}$, then

$$Z(X^{\pm}) = \sigma e^{\pm\frac{1}{2}\tau^{i(i+1)}}.$$

For a tangle $T = E_1 \cdots E_k$ represented as a product of elementary tangles $E_1, \dots, E_k$, we define

$$Z(T) = Z(E_1) \cdots Z(E_k).$$

It turns out that Z is invariant under isotopy of tangles preserving the number of maximal points. To get invariants under isotopy, we have to normalize Z in the following way.

Let ∞ be the tangle and

$$\phi_0 = Z(\infty) \in \mathfrak{D}_1.$$

Now for an arbitrary tangle T, we first mark its components by $1, 2, \dots, r$. Let s_i be the number of maximal points on the ith component. Then

$$\widehat{Z}(T) = Z(T)\#\phi_0^{-s_1}\#\cdots\#\phi_0^{-s_r}$$

where $\#\phi_0^{-s_i}$ is the connected sum onto the ith component of T.

Theorem 2.1. $\widehat{Z} : \mathrm{Mor}(\boldsymbol{NAT}) \to \mathrm{Mor}(\boldsymbol{CD})$ *is well-defined.*

A fundamental problem in this field is the question of whether the functor $\widehat{Z}$ is faithful.

Notice that the functor $\widehat{Z}$ will send the morphism in $\mathrm{Mor}(\emptyset, \emptyset)$ represented by a round circle to $\phi_0^{-1} \in \mathcal{D}_1$. This does not agree with the usual normalization. If one is only interested in knot invariants, it is better to use the following normalization:

$$\widetilde{H}(K) = Z(K)\#\phi_0^{-s+1}$$

where s is the number of maximal points on the knot K thought of as a morphism in $\mathrm{Mor}(\emptyset, \emptyset)$. This is Kontsevich's universal knot invariant (see[Ko] and[B1]).

3. Some calculations

Let $\mathcal{D}_1^{(n)}$ be the subspace of $\mathcal{D}_1$ spanned by chord diagrams on an oriented circle with n chords. We list bases for $\mathcal{D}_1^{(n)}, n \leq 4$. We will denote by $\mathcal{D}^{(i,j)}$ the chord diagram encoded by (i, j). See,[BL] Figures 9, 10, 11 for how to reduce every chord diagram with ≤ 4 chords to a linear combination of these base chord diagrams. For example,

$$\mathcal{D}^{(2,1)}\#\mathcal{D}^{(2,1)} = \mathcal{D}^{(4,1)} + 2\mathcal{D}^{(4,2)} - 3\mathcal{D}^{(4,3)}.$$

Consider the knot $K_{2,p}$, where p is an odd integer. We calculate $Z(K_{2,p})$ using (4):

$$\begin{aligned} Z(K_{2,p}) = {} & |\cap^{14}| \cdot |\cap^{23}| \cdot \Phi\left(\frac{1}{2\pi\sqrt{-1}}\beta^{12}, \frac{1}{2\pi\sqrt{-1}}\beta^{23}\right) \cdot \sigma \cdot e^{\frac{p}{2}\tau^{12}} \\ & \cdot \Phi\left(\frac{1}{2\pi\sqrt{-1}}\gamma^{23}, \frac{1}{2\pi\sqrt{-1}}\gamma^{12}\right) \cdot |\cup^{23}| \cdot |\cup^{14}| \\ = {} & \mathcal{D}^{(0,1)} + \left(\frac{p^2}{8} - \frac{\zeta(2)}{2\pi^2}\mathcal{D}^{(2,1)}\right) + \left(\frac{p^3}{24} - \frac{p\zeta(2)}{4\pi^2}\right)\mathcal{D}^{(3,1)} + \cdots . \end{aligned}$$

Notice that $K_{2,1}$ and $K_{2,-1}$ are isotopic via a level-preserving isotopy. Thus, the coefficient of $\mathcal{D}^{(3,1)}$ in $Z(K_{2,p})$ must be zero when $p = \pm 1$. As a consequence, we get the value of $\zeta(2)$:

$$\zeta(2) = \frac{\pi^2}{6}.$$

See,[LM1LM2] for more about the relationship between values of multiple ζ-numbers and the HOMFLY and Kauffman polynomials.

Since the tangle ∞ is isotopic to $K_{2,1}$ via a level-preserving isotopy, we get

$$\phi_0 = Z(\infty) \equiv \mathcal{D}^{(0,1)} + \frac{1}{24}\mathcal{D}^{(2,1)} \bmod \bigoplus_{n\geq 4} \mathcal{D}_1^{(n)}.$$

Thus we have

$$\widehat{Z}(K_{2,p}) = \mathcal{D}^{(0,1)} + \frac{p^2-1}{8}\mathcal{D}^{(2,1)} + \frac{p(p^2-1)}{24}\mathcal{D}^{(3,1)} + \cdots .$$

To calculate the order 4 term in $\widehat{Z}(K_{2,p})$, let us make the following observation. If

$$\phi_0 = \mathcal{D}^{(0,1)} + \frac{1}{24}\mathcal{D}^{(2,1)} + \alpha + \cdots ,$$

where $\alpha \in \mathcal{D}_1^{(4)}$, then

$$\begin{aligned}
\phi_0^{-1} &= \mathcal{D}^{(0,1)} - \frac{1}{24}\mathcal{D}^{(2,1)} - \alpha + \left(\frac{1}{24}\right)^2 \mathcal{D}^{(2,1)}\#\mathcal{D}^{(2,1)} + \cdots \\
&= \mathcal{D}^{(0,1)} - \frac{1}{24}\mathcal{D}^{(2,1)} - \alpha + \left(\frac{1}{24}\right)^2 (\mathcal{D}^{(4,1)} + 2\mathcal{D}^{(4,2)} - 3\mathcal{D}^{(4,3)}) + \cdots .
\end{aligned}$$

Thus, when calculating the order 4 term in $\widehat{Z}(K_{2,p})$, we don't need to take care of terms having nothing to do with τ^{12} since they appear in both $Z(K_{2,p})$ and ϕ_0^{-1} with opposite signs. Also, the "imaginary" part of the order 4 term, i.e. the part consisting of those chord diagrams whose coefficients are imaginary numbers, is zero. So, the order 4 term in $\widehat{Z}(K_{2,p})$ is

$$\begin{aligned}
&| \cap^{14} || \cap^{23} || [\frac{-\zeta(2)}{24}\frac{p^2-1}{2!2^2}([\beta^{12},\beta^{23}]\sigma(\tau^{12})^2 + \sigma(\tau^{12})^2[\gamma^{23},\gamma^{12}]) \\
&\quad + \frac{p^4-1}{4!2^4}\sigma(\tau^{12})^4] | \cup^{23} || \cup^{14} | \\
&\quad + \left(\left(\frac{1}{24}\right)^2 - \left(\frac{1}{24}\right)\left(\frac{p^2}{8} - \frac{1}{12}\right)\right)\left(\mathcal{D}^{(4,1)} + 2\mathcal{D}^{(4,2)} - 3\mathcal{D}^{(4,3)}\right) \\
&= \frac{1}{24}\frac{p^2-1}{2!2^2}(2\mathcal{D}^{(4,1)} - 2\mathcal{D}^{(4,3)}) + \frac{p^4-1}{4!2^4}\mathcal{D}^{(4,1)} \\
&\quad + \frac{1-p^2}{24\cdot 8}(\mathcal{D}^{(4,1)} + 2\mathcal{D}^{(4,2)} - 3\mathcal{D}^{(4,3)}) \\
&= \frac{p^4-1+2(p^2-1)}{384}\mathcal{D}^{(4,1)} - \frac{p^2-1}{96}\mathcal{D}^{(4,2)} + \frac{p^2-1}{192}\mathcal{D}^{(4,3)}.
\end{aligned}$$

Therefore we have

$$\begin{aligned}\widehat{Z}(K_{2,p}) &= \mathcal{D}^{(0,1)} + \frac{p^2-1}{8}\mathcal{D}^{(2,1)} + \frac{p(p^2-1)}{24}\mathcal{D}^{(3,1)} \\ &\quad + \frac{p^4+2p^2-3}{384}\mathcal{D}^{(4,1)} - \frac{p^2-1}{96}\mathcal{D}^{(4,2)} + \frac{p^2-1}{192}\mathcal{D}^{(4,3)} + \cdots .\end{aligned}$$

When $p = 3$, $K_{2,3}$ is the right trefoil knot. When $p = -3$, $K_{2,p}$ is the left trefoil knot. We have

$$\widehat{Z}(K_{2,3}) = \mathcal{D}^{(0,1)} + \mathcal{D}^{(2,1)} + \mathcal{D}^{(3,1)} + \frac{1}{4}\mathcal{D}^{(4,1)} - \frac{1}{12}\mathcal{D}^{(4,2)} + \frac{1}{24}\mathcal{D}^{(4,3)} + \cdots$$

and

$$\widehat{Z}(K_{2,-3}) = \mathcal{D}^{(0,1)} + \mathcal{D}^{(2,1)} - \mathcal{D}^{(3,1)} + \frac{1}{4}\mathcal{D}^{(4,1)} - \frac{1}{12}\mathcal{D}^{(4,2)} + \frac{1}{24}\mathcal{D}^{(4,3)} + \cdots .$$

References

AC. D. Altschuler and A. Coste, *Quasi-quantum groups, knots, three-manifolds and topological field theory*, Commun. Math. Phys. **150**(1992), 83–107.

B1. D. Bar-Natan, *On the Vassiliev knot invariants*, Topology (to appear).

B2. D. Bar-Natan, *Non-associative tangles*, Proceedings of the 1993 George International Topology Conference (to appear).

BL. J. Birman and X.-S. Lin, *Knot polynomials and Vassiliev's invariants*, Invent. Math. **111**(1993), 225–270.

C. P. Cartier, *Construction combinatorire des invariants de Vassiliev-Kontsevich des nœuds*, C. R. Acad. Sci. Paris **316**(1993), 1205–1210.

Ch. K.-T. Chen, *Iterated path integrals*, Bull. of Amer. Math. Soc. **83**(1977), 831–879.

D1. V. G. Drinfeld, *On quasi-Hopf algebras*, Leningrad Math. J. **1**(1990), 1419–1457.

D2. V. G. Drinfeld, *On quasi-triangular quasi-Hopf algebras and a group closely related with* Gal($\overline{\mathbb{Q}}/\mathbb{Q}$), Leningrad Math. J. **2**(1990), 829–860.

H. R. Hain, private communication, March, 1994.

Ko. M. Kontsevich, *Vassiliev's knot invariants*, Adv. Sov. Math. **16**(1993), 137–150.

LM1. T. Q. T. Le and J. Murakami, *Kontsevich's integral for HOMFLY polynomial and relation between values of multiple ζ-functions*, preprint, 1993.

LM2. T. Q. T. Le and J. Murakami, *Kontsevich's integral for Kauffman polynomial*, preprint, 1993.

LM3. T. Q. T. Le and J. Murakami, *The universal Vassiliev-Kontsevich link invariant*, preprint, 1993.

P. S. Piunikhin, *Combinatorial expression for universal Vassiliev link invariant*, preprint, 1993.

RT. N. Yu. Reshetikhin and V. G. Turaev, *Ribbon graphs and their invariants derived from quantum groups*, Commun. Math. Phys. **127**(1990), 262–288.

Invariants of Legendrian Knots

1. Introduction

Legendrian knots are imbeddings of a circle in a contact 3-manifold which is tangent to the contact plane at every point. For certain standard compact 3-manifolds, invariants of Legendrian knots other than the ordinary knot type are available. Usually, one my have "index-type" invariants which characterize the homotopy types of Legendrian curves. One may also have a "self-linking number" coming from the fact that a Legendrian knot admits a natural framing. So the key question in the study of Legendrian knots seems to be whether two Legendrian knots with the same ordinary framed knot type and the same "indices" are necessarily Legendrian isotopic. See.[E] This question is still open and the purpose of this article is to offer a new approach to this question. This new approach leads to a combinatorial question (Question 7.1) of the same natural as the geometric question about Legendrian knots. It is very likely that these two questions, the geometric one and the combinatorial one, are in fact equivalent. We provide also some evidence supporting a negative answer to the combinatorial question.

We will only deal with Legendrian knots in the space of cooriented contact elements in the plane here. In this case, as index-type invariants, we have the winding number of the normal angel and the so-called Maslov index. The self-linking number was defined only recently by Arnold, who denoted it by J^+. See.[A] Our approach is based on the construction of a Legendrian isotopy invariant generalizing the Kontsevich integral.[Kon] The invariant takes values in the completion of a graded vector space spanned by "dotted chord diagram". One may recover the Maslov index and Arnold's J^+ from our invariant. Although we don't have a Vassiliev-type theory for Legendrian knots yet, the relations among dotted chord diagrams may be visualized through some local moves on fronts (projections) of Legendrian knots. Moreover, these relations imply magically the flatness of a formal connection generalizing the formal Knizhnik-Zamolodchikov connection. We therefore believe that these relations must have some deep topological origin. (Our relations include the so-called 4-term relations in the theory of Vassiliev knot invariants whose topological meaning is quite clear by now.)

The generalization of the formal Knizhnik-Zamolodchikov connetion takes into account cusp points on a general front (projection) of a Legendrian knot by using the delta function. A discussion about the so-called Mathai-Quillen formalism with S. Wu was of great help for us to perceive such a generalization of the formal Knizhnik-Zamolodchikov connection. We are also grateful to O. Viro for some helpful discussions in the early stage of this work and to J. Birman and Y. Eliashberg for their interest in this work.

This article presents only a sketch of our construction. A detailed exposition will appear elsewhere.

2. Legendrian knots and Legendrian braids

We start with some definitions. According to Arnold,[A] a *contact element* on the plane is a line in a tangent plane. The *coorientation* of a contact element is the choice of one of two half-planes into which it divides the tangent plane. Thus, the manifold M of all cooriented contact elements of the plane can be identified with $\mathbb{R}^2 \times S^1$, where $(x, y, \phi \pmod{2\pi}) \in \mathbb{R}^2 \times S^1$ is identified with the tangent line of $\mathbb{R}^2$ at $(x, y) \in \mathbb{R}^2$ perpendicular to its *normal vector* $\mathbf{n} = (\cos\phi, \sin\phi)$.

The manifold M of all coordinated contact elements of the plane is naturally a contact manifold. Under the identification $M = \mathbb{R}^2 \times S^1$, the contact form on M can be written as

$$\omega = (\cos\phi)dx + (\sin\phi)dy.$$

A *Legendrian curve* in M is an immersion $l : S^1 \to M$ such that the tangent vector of l is annihilated by ω everywhere. We will also sometimes call a segment of such an immersion a Legendrian curve. Two Legendrian curves are *Legendrian homotopic* if they can be connected by a (smooth) path of Legendrian curves.

If a Legendrian curve is an imbedding, it is called a *Legendrian knot.* Two Legendrian knots are *Legendrian homotopic* if they can be connected by a (smooth) path of Legendrian knots. A Legendrian isotopic class of a Legendtrian knot is called a *Legendrian knot type.*

The notion of Legendrian curves and Legendrian knots can certainly be applied to any contact 3-manifold.

There are two basic integer invariants of Legendrian curves in the manifold M of all cooriented contact elements of the plane. One is called the *index* and the other the *Maslov index*. They are defined as follows.

Let $l : S^1 \to M = \mathbb{R}^2 \times S^1$ be an immersion. Composing the map l with the projection of $\mathbb{R}^2 \times S^1$ onto the S^1 factor, we get a map $S^1 \to S^1$. Then the index of l, $\mathrm{ind}(l)$, is simply the degree of this map $S^1 \to S^1$.

To define the Maslov index $\mu(l)$ of a Legendrian curve l, we consider the completely non-integrable plane field $\ker(\omega)$ on M. It is oriented by the orientation of M and coorientations of contact elements (their normal vectors). The union of projective spaces of each plane in $\ker(\omega)$ is a (trivial) S^1-bundle over M. The tangent field along the Legendrian curve l determines a section of this S^1-bundle over l and $\mu(l)$ is the Euler number of this section.

Theorem 2.1 (Gromov). *Two Legendrian curves are Legendrian homotopic if and only if they have the same index and Maslov index.*

Let $l : S^1 \to M = \mathbb{R}^2 \times S^1$ be a Legendrian knot. The composition of a map l and the projection of $\mathbb{R}^2 \times S^1$ onto the $\mathbb{R}^2$ factor is called the *front*, $f = f(l)$, of l. The front f of a Legendrian knot is *generic* if

(1) the only singular points (where f fails to be an immersion) are of cusp form, i.e. $f(t) = (t^2, t^3)$ near $t = 0$ up to local diffeomorphism;

(We will then call $f(0)$ a *cusp point* and f will have only finitely many cusp points. An open curve between neighboring cusp points is called a *branch*.)

(2) cusp points are all different;
(3) branches are transverse to each other;
(4) no triple intersections among branches;
(5) no cusp points lie on branches.

In the case of the standard contact structure on $\mathbb{R}^3$, the following basic theorems were formulated by Eliashberg and their proofs were given by Swiatkowski.[S] Here, we use their parallel versions in the case of the space of all cooriented contact elements in the plane. We refer the reader to[S] for more detailed statements of these theorems.

Theorem 2.2. *The space of Legendrian knots with generic fronts is open and dense in the space of all Legendrian knots with topology induced from* $C^\infty(S^1, M)$.

Theorem 2.3. *Two generic fronts represent the same Legendrian knot type iff we can pass from one to the other by a finite sequence of moves of the following four types:*

(0) composition with an orientation preserving diffeomorphism of the plane or reparametrization;
(1) creation or elimination of a "swallow tail";
(2) passage of a cusp point through a branch;
(3) passage of double points through a branch.

Along a generic front $f = f(l)$, we have a smooth normal field given by $\mathbf{n}(t) = (\cos\phi(t), \sin\phi(t))$ if $l(t) = (x(t), y(t), \phi(t))$. This is called a *coorientation* of the front. It can be lifted to a normal field along the Legendrian knot l. Thus a Legendrian knot type determines a ordinary framed knot type. The following seems to be the major question in the study of Legendrian knots.

Is it true that two Legendrian knots are Legendrian isotopic if and only if they are isotopic as ordinary framed knots and have the same index and Maslov index?

It is known that Question 2.1 has a positive answer when the ordinary knot type is trivial. See[E] and.[F]

The winding number of the normal field $\mathbf{n}(t)$ of a cooriented front $f = f(l)$ is the index of the Legendrian knot l. Certainly, a front f also has an orientation coming form the orientation of its domain S^1. A branch of a front is *positive* (or *negative*) if

$$-x' \sin\phi + y' \cos\phi > 0 \text{ (or } < 0)$$

on that branch.

Walking along a front in the direction of its orientation, the sign of branches will change when we pass through a cusp point. Also, the normal angle $\phi(t)$ may increase or decrease when we pass through a cusp point. The sign of a cusp point is determined by the following rules:

sign change of branches	normal angle	sign of cusp point
$- \to +$	increase	$+$
$- \to +$	decrease	$-$
$+ \to -$	increase	$-$
$+ \to -$	decrease	$+$

Let μ_+ (or μ_-) be the number of positive (or negative) cusp points on a front $f = f(l)$. Then

$$\mu(l) = \mu_+ - \mu_-.$$

Recently, Arnold defined a "self-linking number" J^+ for Legendrian knots in the contact manifold of all cooriented contact elements in the plane. See.[A] Up to a certain normalization, the invariant J^+ is characterized by the local property that J^+ changes by a constant ± 2 only under a "dangerous self-tangency perestrokia". Here a self-tangency is a point where two branches of a front are tangent to each other with distinct curvatures. It is dangerous if the coorientations of these two branches agree at that point. And a "perestrokia" here means the process of changing one generic cooriented front to another by passing through the hyperplane of a certain kind of non-generic cooriented front once transversely. A combinatorial description of the Legendrian knot invariant J^+ has been given by Polyak.[Po]

We now identify $\mathbb{R}^2$ with the complex plane $\mathbb{C}$. Suppose $z = z(t) \in \mathbb{C}$ is a smooth curve which is an immersion everywhere except at finitely many places where it has cusp points. We assume further that z'' is never zero in the interior of the curve, i.e. there is no inflection point in the interior of this curve. Then we can reparametrize this curve by its normal angle and we will have

$$u(\phi) = -ie^{i\phi}\frac{dz}{d\phi} \in \mathbb{R} \tag{1}$$

where $i = \sqrt{-1}$. Write $z(\phi) = x(\phi) + iy(\phi)$. Then

$$\frac{dx}{d\phi} = -u(\phi)\sin\phi, \frac{dy}{d\phi} = u(\phi)\cos\phi. \tag{2}$$

So the curve $z = z(\phi)$ is determined by the real valued function $u(\phi)$ up to the initial position. Moreover, the condition that $z' = 0$ only at finitely many places and z'' is never zero is equivalent to the condition that $u = 0$ only at finitely many places and $u' \neq 0$ when $u = 0$. We will call such a curve *regular* if in addition $z' = 0$ only in the interior of the curve.

Let $\phi_0, \phi_1 \in S^1$. Denote by $S^1_{[\phi_0,\phi_1]}$ the arc on S^1 corresponding to angles between ϕ_0 and ϕ_1 in the direction of S^1. We denote

$$M_{[\phi_0,\phi_1]} = \mathbb{C} \times S^1_{[\phi_0,\phi_1]}$$

under the identification $M = \mathbb{C} \times S^1$.

Definition 2.1. A Legendrian braid in $M_{[\phi_0,\phi_1]}$ is a collection $b = \{b_1, \ldots, b_n\}$ of n strands

$$b_\alpha = \{(z_\alpha(\phi), \phi) : \phi_0 \leq \phi \leq \phi_1\} \subset M_{[\phi_0,\phi_1]},$$

$\alpha = 1, \ldots, n$, such that

(1) $z_\alpha(phi) \neq z_\beta(\phi)$ for all $\phi \in S^1_{[\phi_0,\phi_1]}$ if $\alpha \neq \beta$;
(2) $z_\alpha(\phi) \in \mathbb{C}$ is a regular curve for each α.

Notice that each strand in a Legendrian braid is a Legendrian curve.

In the following definition, z^{rev} refers to the curve obtained form going backward along the curve z.

Definition 2.2. A *generalized Legendrian braid* in $M_{[\phi_0,\phi_1]}$ is a collection $b = \{b_1, \ldots, b_n\}$ of n strands

$$b_\alpha = \{(z_\alpha(\phi), \phi) : \phi_0 \leq \phi \leq \phi_1\} \subset M_{[\phi_0,\phi_1]},$$

$\alpha = 1, \ldots, n$, such that

(1) these strands are all disjoint except for some distinct pairs b_α and b_β, we may have $z_\alpha(\phi_0) = z_\beta(\phi_0)$ or $z_\alpha(\phi_1) = z_\beta(\phi_1)$;
(2) if $z_\alpha(\phi_1) = z_\beta(\phi_1)$ (or $z_\alpha(\phi_0) = z_\beta(\phi_0)$), then the curve $z_\alpha * z_\beta^{\text{rev}}$ (or $z_\alpha^{\text{rev}} * z_\beta$) is a smooth curve with an inflection point $z_\alpha(\phi_1)$ (or $z_\beta(\phi_0)$);
(3) $z_\alpha(\phi) \in \mathbb{C}$ is a regular curve for each i.

Two (generalized) Legendrian braids are *Legendrian isotopic* if they can be connected by a path of (generalized) Legendrian braids, such that throughout the isotopy, the $z_\alpha(\phi_{0,1})$'s remain fixed.

We will call a real valued function $u(\phi)$ defined on $[\phi_0, \phi_1]$ *regular* if it passes $u = 0$ transversely in (ϕ_0, ϕ_1). A deformation of a regular function $u(\phi, t)$ is *admissible* if $u(\phi, t)$ is regular for each fixed t and $u(\phi_{0,1}, t) = u(\phi_{0,1}, 0)$ for all t.

A regular front $z(\phi)$ determines a regular function $u(\phi)$. Moreover, a Legendrian isotopy $z(\phi, t)$ of the Legendrian braid $(z(\phi), \phi)$ except a creation or elimination of a swallow tail determines an admissible deformation $u(\phi, t)$.

Let l be a Legendrian knot with a generic front f. We may perturb f a little so that it remains generic and $f'' = 0$ (or $\phi' = 0$) only at finitely many places. There are inflection points on f and they must lie in branches. Then we have

Lemma 2.1. *We may decompose M into*

$$M = M_{[\phi_0,\phi_1]} \cup M_{[\phi_1,\phi_2]} \cup \cdots \cup M_{[\phi_k,\phi_0]}$$

such that $l \cap M_{[\phi_i,\phi_{i+1}]}$ is a (generalized) Legendrian braid for each i mod k.

We will call the decomposition of a Legendrian knot described in the above lemma a *braid decomposition* of l. Such a braid decomposition of a generic front is basically determined by the position of inflection points on the front.

Let f be a generic front with finitely many inflection points. An inflection point on f is called a *local maximum point* if $\phi'' < 0$ at this point. Otherwise, it is a *local minimum point.*

Theorem 2.4. *We may change one braid decomposition of a Legendrian knot type to another by a finite sequence of moves of the following types:*

(1) Legendrian isotopy of a (generalized) Legendrian braid;
(2) combine two (generalized) Legendrian braids into one or vice versa;
(3) change the normal angle or the position of an inflection point on a branch slightly;
(4) creation or elimination of a pair of inflection points on the same branch which are close enough to each other.

3. Extended configuration spaces and a flat formal connection

Let $u \in \mathbb{R}$ be a real variable. Let

$$H(u) = \begin{cases} \frac{1}{2} & \text{if } u \geq 0 \\ -\frac{1}{2} & \text{if } u < 0 \end{cases}$$

be (a shifted version of) the unit step function. Let $\delta(u)$ be the delta function. Then $H'(u) = \delta(u)$ in the sense of distributions.

Consider the distribution-valued 1-form $\delta(u)du$ on $\mathbb{R}$. It can be thought of as a representative of the Thom class of $(\mathbb{R}, \mathbb{R} \setminus 0)$.

Let $u = u(\phi)$ be a regular function on $[\phi_0, \phi_1]$. Let

$$\epsilon_\phi = \begin{cases} 1 & \text{if } u(\phi) = 0 \text{ and } u'(\phi) > 0 \\ -1 & \text{if } u(\phi) = 0 \text{ and } u'(\phi) < 0. \end{cases}$$

Then we have

$$\int_{\phi_0}^{\phi_1} \delta(u)du = \sum_{u(\phi)=0} \epsilon_\phi$$

where the left side is a Lebesgue integral.

Let

$$C_n = \{(z_1, \ldots, z_\alpha, \ldots, z_n) \in \mathbb{C}^n : z_\alpha \neq z_\beta, 1 \leq \alpha < \beta \leq n\}$$

be the configuration space of n distinct ordered points in the complex plane $\mathbb{C}$. The *extended configuration space* is simply

$$C_n \times \mathbb{R}^n.$$

Let $X_{\alpha\beta} = X_{\beta\alpha}$ and Y_α be formal non-commutative variables and consider the following formal 1-form on $C_n \times \mathbb{R}^n$ with values in the algebra $\mathbb{C}[[X_{\alpha\beta}, Y_\alpha]]$ of formal power series:

$$\begin{aligned}\Omega = &\sum_{1\le\alpha<\beta\le n} \frac{1}{2\pi i} X_{\alpha\beta} \frac{dz_\alpha - dz_\beta}{z_\alpha - z_\beta} + \sum_{1\le\alpha\le n} Y_\alpha \delta(u_\alpha) du_\alpha \\ &+ \sum_{1\le\alpha<\beta\le n} \frac{1}{2\pi i}([Y_\alpha, X_{\alpha\beta}]H(u_\alpha)\frac{dz_\alpha - dz_\beta}{z_\alpha - z_\beta} \\ &+ [Y_\beta, X_{\alpha\beta}]H(u_\beta)\frac{dz_\alpha - dz_\beta}{z_\alpha - z_\beta}).\end{aligned}$$

We may calculate the holonomy of Ω on paths in $C_n \times \mathbb{R}^n$ whose $\mathbb{R}^n$ components are regular functions using the Lebesgue integral. The following theorem shows that when the holonomy takes values in an appropriate quotient algebra of $\mathbb{C}[[X_{\alpha\beta}, Y_\alpha]]$, it is invariant under a deformation which induces an admissible deformation of the $\mathbb{R}^n$ components.

Theorem 3.1. *Assume that the formal non-commutative variables $X_{\alpha\beta}$ and Y_α satisfy the following relations:*

$$\begin{cases} [X_{\alpha\beta}, X_{\rho\tau}] = 0] & \text{if } \alpha,\beta,\rho,\tau \text{ are distinct} \\ [X_{\alpha\beta}, X_{\alpha\rho} + X_{\rho\beta}] = 0 & \text{if } \alpha,\beta,\rho \text{ are distinct} \end{cases} \tag{1}$$

$$\begin{cases} [Y_\alpha, Y_\beta] = 0 & \text{if } \alpha,\beta \text{ are distinct} \\ [X_{\alpha\beta}, Y_\rho] = 0 & \text{if } \alpha,\beta,\rho \text{ are distinct} \\ [[Y_\rho, X_{\alpha\beta}], Y_\tau] = 0 & \text{if } \{\rho,\tau\} = \{\alpha,\beta\} \end{cases} \tag{2}$$

Then Ω is flat.

The relation (3.1) is Kohno's infinitesimal pure braid relation.[Koh] It is closely related with the 4-term relation in the theory of Vassiliev invariants[B-L, V]. Both the infinitesimal pure braid relation and the 4-term relation reflect local structures of discriminants of some function spaces. We don't know the topological meaning of the relation (3.2).

Let b be a Legendrian braid. It determines a path in C_n:

$$b = \{(z_1(\phi), \ldots, z_n(\phi)) : \phi_0 \le \phi \le \phi_1\}.$$

Let

$$u_\alpha(\phi) = ie^{i\phi}\frac{dz_\alpha}{d\phi},$$

then

$$B = \{(z_1(\phi), \ldots, z_n(\phi), u_1(\phi), \ldots, u_n(\phi)) : \phi_0 \le \phi \le \phi_1\}$$

is a path in $C_n \times \mathbb{R}^n$ whose $\mathbb{R}^n$ components are regular functions. Let

$$Z(b) = 1 + \sum_{m=1}^{\infty} \int_{\phi_0 \le \varphi_1 \le \cdots \le \varphi_m \le \phi_1} (B^*\Omega)(\varphi_m) \wedge \cdots \wedge (B^*\Omega)(\varphi_1) \in \mathbb{C}[[X_{\alpha\beta}, Y_\alpha]] \tag{3}$$

be the holonomy of Ω along B. Since Ω is flat, $Z(b)$ is invariant under an admissible deformation of the u_α's. One may also show that $Z(b)$ is invariant under a creation or elimination of a swallow tail on b. So $Z(b)$ is invariant under Legendrian isotopy of Legendrian braids.

We may use "dotted chord diagrams" to depict monomials in $X_{\alpha\beta}$ and Y_α. Take a collection L of n ordered line segments, say, of the same length 1. We may think of them as a collection of vertical line segments from height ϕ_0 to height ϕ_1. A *dotted chord diagram* on L is a decoration on L by finitely many horizontal chords running from one line segment in L to another and finitely many dots on line segments in L. Assume, for the moment, that different objects in this decoration have different heights. We will denote by a dot on the αth segment by Y_α and a chord from the αth segment to the βth one by $X_{\alpha\beta}$, for $\alpha \neq \beta$. Then, we may record such a decoration by a monomial in $X_{\alpha\beta}$ and Y_α. To have a 1-1 correspondence between decorations and monomials, we should be allowed to shift dots and chords up and down as long as there is no such time when two objets have the same height. Furthermore, for being able to view a dotted chord diagram as an element in the algebra generated by $X_{\alpha\beta}$ and Y_α subject to the relations (3.1) and (3.2), we should be able to

(1) shift dots and chords up and down as long as dots and end points of chords do not touch each other (corresponding to the first equation in (3.1) and the first and second equations of (3.2)); and
(2) have linear relations in the vector space spanned by decorated diagrams (corresponding to the second equation in (3.1) and the third equation in (3.2)).

Notice that if b is a Legendrian braid in a braid decomposition of a Legendrian knot, each strand of b has an orientation coming from the orientation of the Legendrian knot. On the other hand, each strand has a natural orientation along which ϕ is increasing. Let sign(α) be 1 or -1 according to whether these two orientations on the αth strand agree or not.

Taking the matter of different orientations on strands of b into consideration, we may modify Ω by a change of the formal variables:

$$X_{\alpha\beta} \to \operatorname{sign}(\alpha)\operatorname{sign}(\beta)X_{\alpha\beta}$$

We will always use this modified Ω from now on, except in Section 6 where we will sketch a proof of Theorem 3.1.

When b is a generalized Legendrian braid, some integrals in (3.3) will diverge. One way to deal with those divergent terms is to identify them first and then simply drop them off.

Let us look at the degree m term in (3.3):

$$\int_{phi_0 \le \varphi_1 \le \cdots \le \varphi_m \le \phi_1} (B^*\Omega)(\varphi_m) \wedge \cdots \wedge (B^*\Omega)(\varphi_1)$$

This quantity is a homogeneous polynomial in $X_{\alpha\beta}$, Y_α, $[Y_\alpha, X_{\alpha\beta}]$ and $[Y_\beta, X_{\alpha\beta}]$ of degree m with iterated integrals as coefficient. Let b be a generalized Legendrian braid. If

$$z_\alpha(\phi_0) = z_\beta(\phi_0) \text{ (or } z_\alpha(\phi_1) = z_\beta(\phi_1)),$$

then the iterated integral corresponding to a monomial is divergent if and only if the right-most (or left-most) variable in this monomial is either $X_{\alpha\beta}$ or $[Y_\alpha, X_{\alpha\beta}]$ or $[Y_\beta, X_{\alpha\beta}]$. So we will drop off these terms and still denote the resulting formal power series in $\mathbb{C}[[X_{\alpha\beta}, Y_\alpha]]$ by $Z(b)$ for a generalized Legendrian braid b. We summarize the discussion here into the following theorem.

Theorem 3.2. *$Z(b) \in \mathbb{C}[[X_{\alpha\beta}, Y_\alpha]]$ is a Legendrian isotopy invariant of (generalized) Legendrian braids.*

4. An invariant of Legendrian knots

Assume now that l is a Legendrian knot with a braid decomposition. We apply the invariant Z to each (generalized) Legendrian knot in this decomposition. In a way similar to the construction of Kontsevich's integral, we may form the "cyclic product" of these $Z(b)$'s, keeping the cyclic order of (generalized) Legendrian braids b's in the braid decomposition of l. The result of this cyclic product, denoted by $Z(l)$, is a formal series in dotted chord diagrams on an oriented circle.

By a *dotted chord diagram* on an oriented circle, we mean a decoration of the circle by finitely many dots and chords running from one point on the circle to the other. The dots and end points of chords are all distinct. We are allowed to move these dots and the end points of chords on the circle as long as they remain distinct. We define the *degree* of a dotted chord diagram to be the sum of the number of dots and the number of chords on the circle. Let $\mathcal{D}$ be the completion of the graded vector space spanned by these decorated circles subject to some relations, where we should think of these line segments now as disjoint arcs on the circle. Then, if l is a Legendrian knot with a braid decomposition, $Z(l) \in \mathcal{D}$.

The last thing we need to do in constructing an invariant of Legendrian knots is the stabilization corresponding to the last move in Theorem 2.4: creation or elimination of a pair of inflection points on the same branch which are close enough to each other. We have to distinguish the cases on positive and negative branches.

Let l be a Legendrian knot with a braid decomposition. We may assume that different cusp points lie in different Legendrian braids in the braid decomposition. Moreover, we may assume that if a Legendrian braid b contains a cusp point, then

$$Z(b) = e^{\pm \bullet_a}. \tag{1}$$

The notation $\bullet_a$ stands for a doted chord diagram on the collection of line segments in $Z(b)$ where the only decoratino is a dot on the line segment corresponding to the stand b_α having a cusp point on it. The sign is the sign of that cusp point.

Assume that $Z(l)$ is in such kind of a particular form where cusp points are concentrated as in (4.1). Let $\infty_\pm$ be generalized Legendrian braids where the sign is the sign of the front (there is no cusp point on it). We may assume that $\infty_\pm$ is very thin such that $Z(\infty_\pm)$ can be thought of as concentrated at a point. Consider a branch of l and let s be the number of local maximum points on that branch. Then in the stabilization of $Z(l)$, we will stick $Z(\infty_\pm)^{-s}$ to that branch for every dotted chord diagram in $Z(l)$, where the sign in $\infty_\pm$ should agree with the sign of the branch it is stuck to. More precisely, for every dotted chord diagram in $Z(l)$ where cusp points are concentrated as in (4.1), we cut a branch open at a point away form dots and end points of chords on that branch and glue in $Z(\infty_\pm)^{-s}$ to get the circle back. It turns out that the stabilization is independent of where we stick $Z(\infty_\pm)^{-s}$ to on a branch. Stabilize all other branches in the same way and we eventually get an element

$$\widehat{Z}(l) \in \mathcal{D}.$$

Theorem 4.1. *$\widehat{Z}(l) \in \mathcal{D}$ is a Legendrian isotopy invariant of Legendrian knots.*

It is easy to see that the Maslov index of a Legendrian knot l, $\mu(l)$, is the coefficient of the dotted chord diagram on a circle with a single dot. Moreover, the coefficient in $\widehat{Z}(l)$ of the dotted chord diagram changes by a constant under a "dangerous self-tangency perestrokia", it is proportional to J^+ up to a constant summand. We expect that the non-commutativity of $X_{\alpha\beta}$ and Y_α (see relation (3.2)) indicates that $\widehat{Z}(l)$ may contain more information than the Maslov index and the ordinary framed knot type of l. See the discussion in the last section.

5. Relation with Kontsevich integral

It is clear that if a (generalized) Legendrian braid b has no cusp point on it, $Z(b)$ is just the usual Kontsevich integral with each chord $X_{\alpha\beta}$ replaced by

$$X_{\alpha\beta} + \operatorname{sign}(b_\alpha)\frac{1}{2}[Y_\alpha, X_{\alpha\beta}] + \operatorname{sign}(b_\beta)\frac{1}{2}[Y_\beta, X_{\alpha\beta}]$$

where $\operatorname{sign}(b_\alpha)$ is the sign of b_α as a branch. In particular, $Z(\infty_\pm)$ is completely determined by Drinfeld's associator.

Compare with Kontsevich integral, e.g., see,[BN1, BN2, C, P, L-M1, L-M2, Kas] it will be desirable to have answers to the following questions.

Is there a combinatorial construction of the Legendrian knot invariant $\widehat{Z}$?

Is there a theory of weight systems on dotted chord diagrams?

We hope that positive answers to these two questions will make the computation of our invariant $Z(l)$ easy and eventually lead to an (presumably negative) answer to Question 2.1.

To complete the whole picture, we hope to have a positive answer to the following question (compare with[VB-L]):

Is there a theory of finite type invariants for Legendrian knots?

6. A sketch of the proof of Theorem 3.1

We will sketch a proof of the flatness of the formal connection Ω here, assuming the relations (3.1) and (3.2).

To get $d\Omega - \Omega\wedge\Omega = 0$, it suffices to check that each part of $d\Omega - \Omega\wedge\Omega$ in the following three cases vanishes. For simplicity, we will use $1, 2, 3$ instead of general $\alpha, \beta\rho$ for indices. Also we will denote

$$\omega_{12} = \frac{1}{2\pi i}\frac{dz_1 - dz_2}{z_1 - z_2}, \text{ etc.}$$

in the following calculation.

Case I.

$$\begin{aligned} &[[Y_1, X_{12}], X_{23}]H(u_1)\omega_{12}\wedge\omega_{23}+ \\ &[X_{23}, [Y_1, X_{13}]]H(u_1)\omega_{23}\wedge\omega_{13}+ \\ &[[Y_1, X_{13}], X_{12}]H(u_1)\omega_{13}\wedge\omega_{12}+ \\ &[X_{13}, [Y_1, X_{12}]]H(u_1)\omega_{13}\wedge\omega_{12} = 0. \end{aligned} \tag{1}$$

First we have

$$\begin{aligned} [[Y_1, X_{12}], X_{23}] &= Y_1X_{12}X_{23} - X_{12}Y_1X_{23} - X_{23}Y_1X_{12} + X_{23}X_{12}Y_1 \\ &= Y_1X_{12}X_{23} - X_{12}X_{23}Y_1 - Y_1X_{23}X_{12} - X_{23}X_{12}Y_1 \\ &= Y_1[X_{12}, X_{23}] - [X_{12}, X_{23}]Y_1 \\ &= [Y_1, [X_{12}, X_{23}]] \end{aligned}$$

Similarly we have

$$[X_{23}, [Y_1, X_{13}]] = [Y_1, [X_{23}, X_{13}]].$$

Thus,

$$[X_{23}, [Y_1, X_{13}]] = [[Y_1, X_{12}], X_{23}]. \tag{2}$$

Next, we have

$$\begin{aligned} &[[Y_1, X_{13}], X_{12}] + [X_{13}, [Y_1, X_{12}]] \\ &\quad = Y_1X_{13}X_{12} - X_{13}Y_1X_{12} - X_{12}Y_1X_{13} + X_{12}X_{13}Y_1 \\ &\qquad\quad + X_{13}Y_1X_{12} - X_{13}X_{12}Y_1 - Y_1X_{12}X_{13} + X_{12}Y_1X_{13} \\ &\quad = Y_1X_{13}X_{12} - X_{13}X_{12}Y_1 - Y_1X_{12}X_{13} + X_{12}X_{13}Y_1 \\ &\quad = [Y_1, [X_{13}, X_{12}]] \end{aligned}$$

Thus

$$[[Y_1, X_{13}], X_{12}] + [X_{13}, [Y_1, X_{12}]] = [[Y_1, X_{12}], X_{23}]. \tag{3}$$

Combining (6.2), (6.3) and Arnold's identity

$$\omega_{12}\wedge\omega_{23} + \omega_{23}\wedge\omega_{13} + \omega_{13}\wedge\omega_{12} = 0,$$

we get (6.1).

Case II.

$$\begin{aligned}&[[Y_1, X_{12}], [Y_2, X_{23}]]H(u_1)H(u_2)\omega_{12}\wedge\omega_{23}+\\&[[Y_2, X_{23}], [Y_1, X_{13}]]H(u_1)H(u_2)\omega_{23}\wedge\omega_{13}+\\&[[Y_1, X_{13}], [Y_1, X_{12}]]H(u_1)H(u_2)\omega_{13}\wedge\omega_{12}=0.\end{aligned}\tag{4}$$

We have

$$\begin{aligned}&[[Y_1, X_{12}], [Y_2, X_{23}]]\\&\quad= [Y_1, X_{12}]Y_2X_{23} - [Y_1, X_{12}]X_{23}Y_2 - Y_2X_{23}[Y_1, X_{12}] + X_{23}Y_2[Y_1, X_{12}]\\&\quad= Y_2[Y_1, X_{12}]X_{23} - [Y_1, X_{12}]X_{23}Y_2 - Y_2X_{23}[Y_1, X_{12}] + X_{23}[Y_1, X_{23}]Y_2\\&\quad= Y_2[[Y_1, X_{12}], X_{23}] - [[Y_1, X_{12}], X_{23}]Y_2\end{aligned}$$

and

$$\begin{aligned}&[[Y_2, X_{23}], [Y_1, X_{13}]]\\&\quad= Y_2X_{23}[Y_1, X_{13}] - X_{23}Y_2[Y_1, X_{13}] - [Y_1, X_{13}]Y_2X_{23} + [Y_1, X_{13}]X_{23}Y_2\\&\quad= Y_2X_{23}[Y_1, X_{13}] - X_{23}[Y_1, X_{13}]Y_2 - Y_2[Y_1, X_{13}]X_{23} + [Y_1, X_{13}]X_{23}Y_2\\&\quad= Y_2[X_{23}, [Y_1, X_{13}]] - [X_{23}, [Y_1, X_{13}]]Y_2.\end{aligned}$$

Therefore

$$[[Y_1, X_{12}], [Y_2, X_{23}]] = [[Y_2, X_{23}], [Y_1, X_{13}]].\tag{5}$$

Similarly

$$[[Y_1, X_{12}], [Y_2, X_{23}]] = [[Y_1, X_{13}], [Y_2, X_{12}]].\tag{6}$$

We see that (6.5) and (6.6) together with Arnold's identity imply (6.4).

Case III.

$$[[Y_2, X_{12}], [Y_2, X_{23}]]H(u_2)H(u_2)\omega_{12}\wedge\omega_{23}=0.\tag{7}$$

We have

$$\begin{aligned}&[[Y_2, X_{12}], [Y_2, X_{23}]] + [[Y_2, X_{12}], [Y_2, X_{23}]]\\&\qquad= [Y_2, X_{12}]Y_2X_{23} - [Y_2, X_{12}]X_{23}Y_2 - Y_2X_{23}[Y_2, X_{12}] + X_{23}Y_2[Y_2, X_{23}]\\&\qquad\qquad + Y_2X_{12}[Y_2, X_{23}] - X_{12}Y_2[Y_2, X_{23}] - [Y_2, X_{23}]Y_2X_{12}\\&\qquad\qquad + [Y_2, X_{23}]X_{23}Y_2\\&\quad= Y_2[[Y_2, X_{12}], X_{23}] - [[Y_2, X_{12}], X_{23}]Y_2\\&\qquad + Y_2[X_{12}, [Y_2, X_{23}]] - [X_{12}, [Y_2, X_{23}]]Y_2\\&\quad= Y_2[X_{13}, [Y_2, X_{12}]] - [X_{13}, [Y_2, X_{12}]]Y_2 = 0.\end{aligned}$$

So (6.7) holds.

7. An irreducible weight system on dotted chord diagrams

By a *weight system*, we mean a linear functional on (the completion of) the graded vector space $\mathcal{D}$ of dotted chord diagrams on the circle. In the following definition, if C is a dotted chord diagram, then $\bar{C}$ will be the chord diagram obtained from C by deleting all dots. We will call $\bar{C}$ the *underlying chord diagram* of C.

Definition 7.1. A weight system w is called *reducible* if for any two dotted chord diagrams C_1 and C_2 with the same number of dots, $w(\bar{C}_1) = w(\bar{C}_2)$ implies $w(C_1) = w(C_2)$. Otherwise, w is called irreducible.

Thus, a reducible weight system is completely determined by its restriction on chord diagrams without dots and the number of dots. The existence of irreducible weight systems is related with the question of whether the graded vector space $\mathcal{D}$ is "decomposable" in the following sense.

There are two graded subspaces in $\mathcal{D}$: one is spanned by chord diagrams without dots and the other is spanned by dotted chord diagrams without chords. We denote the former by $\mathcal{D}^{\mathrm{c}}$ and the latter by $\mathcal{D}^{\mathrm{d}}$.

Is it true that $\mathcal{D} \cong \mathcal{D}^{\mathrm{c}} \otimes \mathcal{D}^{\mathrm{d}}$ as graded vector spaces?

We are unable to answer this question. On the other hand, it is quite easy to come up with a $\mathbb{Z}_2$-valued irreducible weight system on the graded abelian group $\mathcal{D}_{\mathbb{Z}}$, where $\mathcal{D}_{\mathbb{Z}}$ is the abelian group generated by dotted chord diagrams on the circle subject to some relations. This weight system is analogous to the weight system of the Alexander polynomial for ordinary knots.

Let C be a dotted chord diagram on the circle. Think of the circle as one of the boundary components of an oriented annulus with consistent orientation. Then replace each chord by a very thin riboon such that

(1) there is no dot in the region where these ribbons are stuck to the circle;
(2) these ribbons are all disjoint; and
(3) the resulting compact surface is orientable.

The circle now splits into a collection $s(C)$ of circles decorated with dots. Let A be an abelian group. Pick an element $d_p \in A$ for each $p = 0, 1, 2, \ldots$. We define

$$w(C) = \sum d_p \in A,$$

where in the summation there is a d_p for each circle decorated with p dots in $s(C)$.

Theorem 7.1. *If $d_{p+2} - 2d_{p+1} + d_p \in A$ is a 2-torsion independent of p, then the linear extension $w : \mathcal{D}_{\mathbb{Z}} \to A$ is well-defined.*

For example, we may take $A = \mathbb{Z}_2$ and $d_{p+2} + d_p = 1 \in \mathbb{Z}_2$. It is quite easy to find two dotted chord diagrams C_1 and C_2 with the same underlying chord diagram and $w(C_1) = d_2 + d_0 = 1$ but $w(C_2) = d_1 + d_1 = 0$. In other words, the weight system w is irreducible.

References

A. V. Arnold, Topological Invariants of Plane Curves and Caustics, Univ. Lec. Series, vol. 5, Amer. Math. Soc., 1995.

BN1. D. Bar-Natan, *On the Vassiliev knot invariants*, Topology, **34**(1995), pp. 423–471.

BN2. D. Bar-Natan, *Non-associative tangles*, to appear.

B-L. J. Birman and X.-S. Lin, *Knot polynomials and Vassiliev's invariants*. Invent. Math., **111**(1993), pp. 225–270.

C. P. Cartier, *Construction combinatoire des invariants de Vassiliev-Kontsevich des nœuds.*, C. R. Acad. Sci. Paris, **319**(1993), series 1, pp. 1205–1210 F.

E. Y. Eliashberg, *Legendrian and transversal knots in tight contact 3-manifolds*, Topological Methods in Modern Mathematics (Stony Book, NY, 1991), Publish or Perish, 1993.

F. M. Fraser, Thesis, Stanford University, 1994.

Koh. T. Kohno, *Monodromy representations of braid groups and Yang-Baxter equations*, Ann. Inst. Fourier, **37**(1987), pp. 139–160.

Kon. M. Kontsevich, *Vassiliev's knot invariant*, Adv. Sov. Math., **16**(1993), part 2, pp. 137–150.

Kas. C. Kassel, Quantum Groups, Graduate Texts in Mathematics, Springer-Verlag, 1994.

L-M1. T. Q. T. Le and J. Murakami, *Representations of the category of tangles by Kontsevich's iterated integral*, Max-Planck Institute preprint, 1993.

L-M2. T. Q. T. Le and J. Murakami, *The universal Vassliev-Kontsevich invariant for framed oriented links*, Max-Planck Institute für Mathematik, preprint, 1993.

P. S. Piunikhin, *Combinatorial expression for universal Vassiliev link invariant*, Harvard University, preprint, 1993.

Po. M. Polyak, *Invariants of plane curves and fronts via Gauss diagrams*, Max-Planck Institute für Mathematik, preprint, 1995.

S. J. Swiatkowski, *On the isotopy of Legendrian knots*, Ann. Clo. Ana. Geo., **10**(1992), pp. 195–207.

V. V. Vassiliev, *Cohomology of knot spaces*, Theory of Singularities and Its Application (ed. V. Arnold), Amer. Math. Soc., 1990.

The Motion Group of the Unlink and its Representations

A loop braid is formed by an isotopy of a finite collection of disjoint "small loops" in the 3-space. We show that loop braids form a finitely presented group, called the loop braid group. This extends the phenomenon of braiding from 2+1 dimension to 3 + 1 dimension.

1. Introduction

The braid group was introduced by E. Artin in the 1926.[1] In the last 30 or more years, the notion of braiding has become indispensable for many fields of mathematics and mathematical physics including number theory, representation theory, algebraic geometry, and conformal and quantum field theory. The classical book[2] presents a thorough treatment of the braid group and its applications in knot theory. See[3] for a survey of the current state of the art in the study of the braid group and its role in knot theory, as well as an extensive list of references to the literature. In recent years, the braid group has also played a prominent role in the study of quantum Hall effect and quantum computing.

Recall that the braid group describes the topology of the motion of distinct points in the 2-dimensional plane. A basic fact is that there are infinitely many ways, non-homotopic to each other, to exchange the positions of two particles in the plane. In the presence of n particles in the plane, the totality of topologically different ways to exchange the position of these particles is given by the braid group B_n. The braid group B_n is an infinite extension of the symmetric group S_n. When we have at least three particles in the plane, some different combinations of various ways to exchange the position of particles turn out to be topologically equivalent, which gives rise to braiding relations. Figure 1 describes the most important basic braiding relation. In the picture, the vertical direction is the direction of the time of the motion. So what we see are world lines in 2+1 dimensional space-time traveled by particles as they move in the plane to exchange their positions.

Contrary to the plane, the topology of the motion of distinct points in the 3-space is trivial, in the sense that up to homotopy, there is only one way to exchange the positions of two particles in the 3-space. The basic and simple idea of this paper is that in dimension 3, one should replace distinct points by disjoint small loops in order to have non-trivial topology. To be more specific, we consider a collection of disjoint loops in the 3-space such that they bound disjoint disks. These disks will allow us to shrink the loops to be arbitrarily small without touching each other. It is in this sense that we call them small loops. In the motion of such a collection of disjoint small loops, we allow a small loop to pass through the interior of the disks bounded by other small loops. We introduce the *loop braid group* to describe precisely topologically different ways to exchange the positions of these small loop under admissible motion. The loop braid group LB_n for n disjoint small loops is determined by a finite set of elementary loop braids as generators and a finite set of relations among these elementary loop braids. In order words, the loop braid group LB_n has a finite presentation. This is the main result of this paper. Figure 2 illustrates the most basic loop braid relation in the loop braid group. See Theorem 3.1 and Remark 3.1. Once again, the vertical direction is the direction of the time of the motion, and what we see are world lines in 3+1 dimensional space-time traveled by small loops in the 3-space.

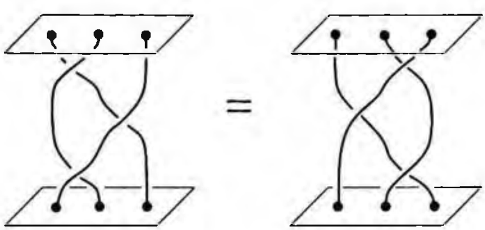


Fig. 1.1. Braid relation.

2. Configuration space of disjoint small loops

We denote by S^1 and D^2 the unit circle and unit disk, respectively, such that $\partial D^2 = S^1$.

Let $\delta : D^2 \longrightarrow \mathbb{R}^3$ be a smooth embedding. We call $l = \delta|S^1 : S^1 \longrightarrow \mathbb{R}^3$ a *small loop* in $\mathbb{R}^3$. Note that by definition, a small loop $l : S^1 \longrightarrow \mathbb{R}^3$ has an embedding $\delta : D^2 \longrightarrow \mathbb{R}^3$ as its extension. But such an extension is far from unique. Let $(l_1, l_2, \ldots, l_n)$ be a collection of small loops in $\mathbb{R}^3$. We say

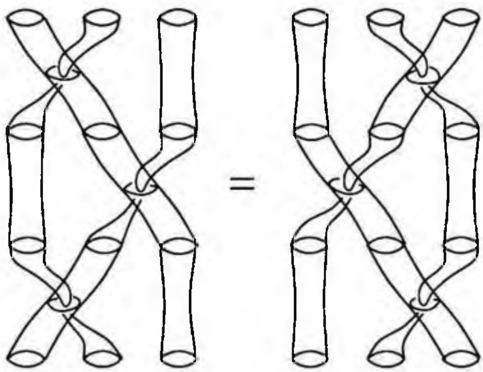


Fig. 1.2. Loop braid relation.

that this is a collection of *disjoint small loops* if each $l_i : S^1 \longrightarrow \mathbb{R}^3$ has an extension $\delta_i : D^2 \longrightarrow \mathbb{R}^3$ and $\delta_i(D^2) \cap \delta_j(D^2) = \emptyset$ for $i \neq j$.

Denote by $\mathcal{C}_n$ the space of all collections of disjoint small loops in $\mathbb{R}^3$ with n components equipped with the usual compact-open topology. A continuous path in this space $\mathcal{C}_n$ corresponds to an isotopy of a collection of disjoint small loops $(l_1, l_2, \ldots, l_n)$ in $\mathbb{R}^3$. Let δ_i be the defining extension of l_i. Then this isotopy can be thought of as to move the disks $\delta_i(D^2)$ in $\mathbb{R}^3$ with the condition $\delta_i(\partial D^2) \cap \delta_j(\partial D^2) = \emptyset$ for $i \neq j$ kept preserved all the time. A path in $\mathcal{C}_n$ or an isotopy of $(l_1, l_2 \ldots, l_n)$ will be denoted by $(l_1, l_2, \ldots, l_n)_t$.

Let (x, y, z) be a Cartesian coordinate system of $\mathbb{R}^3$. A collection of disjoint small loops $(l_1, l_2, \ldots, l_n)$ is called *horizontal* if each $\delta_i(D^2)$ lies in a horizontal plane $z =$ constant. Furthermore, we require that the positive normal direction of $\delta_i(D^2)$ agrees with the positive z-direction form every i. A path $(l_1, l_2, \ldots, l_n)_t$ is horizontal if $(l_1, l_2, \ldots, l_n)_t$ is horizontal for every t.

Lemma 2.1. *Suppose that $(l_1, l_2, \ldots, l_n)_t$, $t \in [0, 1]$, is a path in $\mathcal{C}_n$ such that both $(l_1, l_2, \ldots, l_n)_0$ and $(l_1, l_2, \ldots, l_n)_1$ are horizontal. Then the path is path-homotopic to an horizontal path in $\mathcal{C}_n$.*

Let $(l_1, l_2, \ldots, l_n)_t$ be a path in $\mathcal{C}_n$ corresponding to an isotopy of $(l_1, l_2, \ldots, l_n) = (l_1, l_2, \ldots, l_n)_0$. Let δ_i be the defining extension of l_i. We assume that during the entire isotopy, we have $\delta_i(D^2) \cap \delta_j(D^2) = \emptyset$ for $i \neq j$. If at the end of the isotopy, l_i is moved to $l_{\tau(i)}$, $i = 1, 2, \ldots, n$, for some $\tau \in S_n$, then we call this path $(l_1, l_2 \ldots, l_n)_t$ a *permutation path* associated with τ.

Lemma 2.2. *Two permutation paths in $\mathcal{C}_n$ associated with the same permutation τ are path-homotopic.*

Lemma 2.3. *Consider an isotopy that moves $(l_1, l_2, \ldots, l_n)$ to $(l_{\tau(1)}, l_{\tau(2)}, \ldots, l_{\tau(n)})$, $\tau \in S_n$, as a path in $\mathcal{C}_n$. Then, this path is path-homotopic to the joint of a closed path and a permutation path, and such a decomposition is unique up to path-homotopy.*

Lemma 2.4. *Up to path-homotopy in $\mathcal{C}_n$, the joint of a closed path from $(l_1, l_2, \ldots, l_n)$ to itself and a permutation path associated with $\tau \in S_n$ can change order after renaming that closed path as a closed path from $(l_{\tau(1)}, l_{\tau(2)}, \ldots, l_{\tau(n)})$ to itself.*

A *loop braid* is an isotopy from $(l_1, l_2, \ldots, l_n)$ to $(l_{\tau(1)}, l_{\tau(2)}, \ldots, l_{\tau(n)})$, for $\tau \in S_n$, of collections of disjoint small loops. Loop braids are in one-one correspondence with path in $\mathcal{C}_n$ from $(l_1, l_2, \ldots, l_n)$ to $(l_{\tau(1)}, l_{\tau(2)}, \ldots, l_{\tau(n)})$. Two loop braids from $(l_1, l_2, \ldots, l_n)$ to $(l_{\tau(1)}, l_{\tau(2)}, \ldots, l_{\tau(n)})$ are isotopic if the corresponding paths in $\mathcal{C}_n$ are path-homotopic. By the previous lemmas, it suffices to consider only loop braids from $(l_1, l_2, \ldots, l_n)$ to itself, and we may further assume that $(l_1, l_2, \ldots, l_n)_t$ is horizontal for all t. The set of isotopy classes of such loop braids is the same as the fundamental group of $\mathcal{C}_n$ with the base point $(l_1, l_2, \ldots, l_n)$. We will call this group the *loop braid group* and denote it by LB_n.

3. Elementary loop braids and relations among them

We will also use $(l_1, l_2, \ldots, l_n)$ to denote the subset $l_1(S^1) \cup l_2(S^1) \cup \cdots \cup l_n(S^1)$ of $\mathbb{R}^3$, and $(l_1\, l_2, \ldots, \hat{l_i}, \ldots, l_n)$ means to drop the i-th component of this collection of disjoint small loops. Note that $\pi_1(\mathbb{R}^3 \backslash (l_1\, l_2, \ldots, \hat{l_i}, \ldots, l_n))$ is a free group of rank $n-1$. Denote by x_{ij}, $j \in \{1, 2 \ldots, \hat{i}, \ldots, n\}$, the set of standard generators of this free group. An *elementary loop braid* σ_{ij} is an isotopy of $(l_1, l_2, \ldots, l_n)$ to itself specified by (1) it moves l_i to itself such that the trajectory of $\delta_i(0)$ represents x_{ij}, and (2) it does not move all other components of $(l_1, l_2, \ldots, l_n)$.

Theorem 3.1. *Elementary loop braids σ_{ij}, $i \neq j$, $i, j \in \{1, 2, \ldots, n\}$, satisfy the following relations:*

(1) $\sigma_{ij}\sigma_{km} = \sigma_{km}\sigma_{ij}$, if i, j, k, m are all distinct;

(2) $\sigma_{ik}\sigma_{jk} = \sigma_{jk}\sigma_{ik}$, if i, j, k are all distinct;

and

(3) $\sigma_{ij}\sigma_{kj}\sigma_{ik} = \sigma_{ik}\sigma_{kj}\sigma_{ij}$, if i, j, k are all distinct.

Proof. The first two sets of relations (1) and (2) are easy to see. We check the third set of relations. Without loss of generality, we consider only (l_1, l_2, l_3).

Denote by x, y the generators of $\pi_1(\mathbb{R}^3 \setminus (l_2, l_3))$, where x is represented by a path going through l_2 and y a path going through l_3. The elementary loop braid σ_{32} is the isotopy from (l_2, l_3) to itself where l_3 goes through l_2. This isotopy induces an automorphism ϕ of $\pi_1(\mathbb{R}^3 \setminus (l_2, l_3))$. We have $\phi(x) = x$ and $\phi(y) = xyx^{-1}$.

The joint of elementary loop braids $\sigma_{12}\sigma_{32}\sigma_{13}$ can be written intuitively as $x * \phi * y$. We calculate as follows:

$$x * \phi * y = x * \phi * y * \phi^{-1} * \phi = x * (x^{-1}yx) * \phi = y * x * \phi = y * \phi * x.$$

In the above line of calculation, t he first equality is to insert a trivial loop braid $\phi^{-1} * \phi$, the second equality is to comb y through ϕ on its left side and delete $\phi * \phi^{-1}$, the third equality is to delete xx^{-1}, and the last equality comes from the relation (2): $\sigma_{12}\sigma_{32} = \sigma_{32}\sigma_{12}$. Thus we have

$$\sigma_{12}\sigma_{32}\sigma_{13} = \sigma_{13}\sigma_{32}\sigma_{12}. \qquad \square$$

Remark 3.1. Since closed paths and permutation paths in $\mathcal{C}_n$ commute with each other up to path-homotopy and renaming of closed paths by permutation paths involved, it is easy to see that relation (3) in Theorem 3.1 depicted in Figure 3 and the loop braid relation depicted in Figure 2 are equivalent.

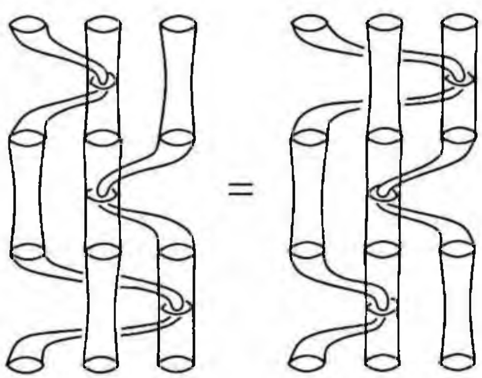


Fig. 3.1. Relation (3) in Theorem 3.1.

Remark 3.2. Let F_n be the free group generated by $x_1, x_2, \ldots, x_n$. Each elementary loop braid σ_{ij} induces an automorphism ϕ_{ij} of F_n given by

$$\begin{aligned} \phi_{ij}(x_k) &= x_k, \quad k \neq i, j \\ \phi_{ij}(x_i) &= x_j x_i x_j^{-1}, \\ \phi_{ij}(x_j) &= x_j. \end{aligned}$$

One can check directly that these automorphisms of F_n satisfy the relations (1), (2), and (3) in Theorem 3.1.

4. Presentation of loop braid group

We may follow the classical approach to show that the loop braid group LB_n admits a finite presentation where the generators are elementary loop braids σ_{ij}, $i \neq j$, and relations are these given in Theorem 3.1.

If we drop the first small loop l_1 in $(l_1, l_2, \ldots, l_n)$, we get a epimorphism

$$LB_n \longrightarrow LB_{n-1}.$$

Let G be the kernel of this epimorphism. Note first that σ_{1i}, $i = 2, \ldots, n$, generate a free group G' of rank $n - 1$. Second, σ_{i1}, $i = 2, \ldots, n$ generate a free abelian group G'' of rank $n - 1$.

Lemma 4.1. *G is isomorphic to the free product of G' and G''.*

We illustrate the situation by consider the loop braids of (l_2, l_3) used in the proof of Theorem 3.1. The loop braid σ_{32} induces an automorphism ϕ of the free group generated by x, y:

$$\phi(x) = x \quad \text{and} \quad \phi(y) = xyx^{-1}.$$

Similarly, the loop braid σ_{23} induces an automorphism ψ:

$$\psi(x) = yxy^{-1} \quad \text{and} \quad \psi(y) = y.$$

Let $\phi^{n_1}\psi^{m_1}\cdots\phi^{n_r}\psi^{m_r}$ be a reduced word in ϕ and ψ. Then

$$\begin{aligned} &\phi^{n_1}\psi^{m_1}\cdots\phi^{n_r}\psi^{m_r}(x) \\ &\quad = x^{n_1}y^{m_1}\cdots x^{n_r}y^{m_r}xy^{-m_r}x^{-n_r}\cdots y^{-m_1}x^{-n_1}. \end{aligned}$$

Thus, if $m_r \neq 0$, then $\phi^{n_1}\psi^{m_1}\cdots\phi^{n_r}\psi^{m_r} \neq 1$. Similarly, let $\psi^{m_1}\phi^{n_1}\cdots\psi^{m_r}\phi^{n_r}$ be a reduced word in ϕ and ψ such that $n_r \neq 0$, then $\psi^{m_1}\phi^{n_1}\cdots\psi^{m_r}\phi^{n_r} \neq 1$. This means that ϕ and ψ generate a free group. This proves the case of Lemma 4.1 when $n = 2$.

Theorem 4.1. *The loop braid group LB_n admits a presentation with generators σ_{ij}, $i \neq j$, $i, j = 1, 2 \ldots, n$ and relations (1), (2), and (3) in Theorem 3.1.*

Proof. Consider the exact sequence

$$1 \longrightarrow G \longrightarrow LB_n \longrightarrow LB_{n-1} \longrightarrow 1.$$

We can think of LB_{n-1} naturally as a subgroup of LB_n. The conjugation action of LB_{n-1} on G is by combing. Since a presentation of G is known by Lemma 4.1, and inductively, LB_{n-1} admits the presentation given by this theorem, the loop group LB_n is generated by σ_{ij}, $i \neq j$ and $i, j = 1, 2 \ldots, n$, and subject to relations in G, relations in LB_{n-1}, and the relations describing the conjugation action of LB_{n-1} on G. By the calculation in the proof of Theorem 3.1, the relations describing the conjugation action of LB_{n-1} on σ_{1i} and σ_{i1} are equivalent to the relations (1), (2), and (3) that involve σ_{1i} and σ_{i1}, respectively. Thus, LB_n has the given presentation. □

References

1. E. Artin, *Theory of braids.* Ann. Math. **48**(1947), 101–126.
2. J. Birman, Braids, Links, and Mapping Class Groups. Ann. Math. Studies **82**, Princeton University Press, 1976.
3. J. Birman and T. Brendle, *Braids: a survey.* to appear in the Handbook of Knot Theory, edited by W. Menasco and M. Thistlethwaite.
4. A. Brownstein and R. Lee, *Cohomology of the group of motions of n strings in 3-space.* Mapping class groups and moduli spaces of Riemann surfaces (Gttingen, 1991/Seattle, WA, 1991), 51–61, Contemp. Math., 150.
5. D.M. Dahm, *A generalisation of braid theory.* Princeton Thesis, 1962.
6. D.L. Goldsmith, *The theory of motion groups.* Michigan Math. J. **28**(1981), no. 1, 3–17.
7. D.L. Goldsmith, *Motion of links in the 3-sphere.* Math. Scand. **50** (1982), no. 2, 167–205.
8. L. McCool, *On basis-conjugating automorphisms of free groups.* Canad. J. Math. **38** (1986), no. 6, 1525–1529.
9. T. Fiedler, *Isotopy invariants for smooth tori in 4-manifolds.* Topology **40** (2001), no. 6, 1415–1435.
10. R. L. Rubinsztein, *On the group of motions of oriented, unlinked and unknotted circles in* $\mathbb{R}^3$, *I.* Preprint, Uppsala University, 2002.
11. S. Surya, *Cyclic statistics in three dimensions.* J. Math. Phys. **45** (2004), no. 6, 2515–2525.
12. R.J. Szabo, *Topological field theory and quantum holonomy representations of motion groups.* Annals Phys. **280**(2000), 163–208
13. F. Wattenberg, *Differentiable motions of unknotted, unlinked circles in 3-space.* Math. Scand. **30**(1972), 107–135.

PART C

Lin Award, Speeches and Writings

Lin Award at Beijing University

The Xiao-Song Lin Award at Beijing University was established by the family of Xiao-Song Lin in 2007. A cash prize is awarded each year to a graduating undergraduate student at Beijing University who has demonstrated exceptional scholarship in mathematics. The first recipient of the Xiao-Song Lin Award was Hongbin Sun.

1. Goal

Xiao-Song Lin was an alumnus of Beijing University. For many years, he played an active role in the growth and development of the mathematics community at Beijing University and in China. During the 1990s, he returned in China to deliver lectures every year. In remembrance of Xiao-Song Lin, his accomplishments, and his contributions to mathematics, especially in low-dimensional topology and knot theory, his family established the Xiao-Song Lin award in the College of Mathematics, Beijing University. The goal of the Xiao-Song Lin Award is to encourage undergraduate students to work hard, to pursue mathematics passionately, and to make their own contributions to the development of mathematics in China, like Xiao-Song Lin did.

2. Eligibility

Every year, two or three graduates will be recommended by the professors of the College of Mathematics at Beijing University. The recipient of the award will be chosen by the selection committee. Upon the approval of the Xiao-Song Lin fund members, a certificate and financial award from the Xiao-Song Lin fund will be presented by the provost of the College of Mathematics at Beijing University.

3. Xiao-Song Lin fund members

Jian-Pin He	— Wife of Xiao-Song Lin
Kevin Lin	— Son of Xiao-Song Lin
Zhiwen Li	— Former Ph.D student of Xiao-Song Lin

4. Timeline and cash prize

The award will be presented for 51 years, from 2007 to 2057. The year 2057 will be the 100th anniversary of Xiao-Song Lin's birth. The financial award for the nth year will be $1000 + 50(n - 1)$ US dollars.

Selected Speeches at the Funeral

A farewell/funeral service held by Xiao-Song's family, friends, and colleagues took place on Friday, Jan. 19 from 2:00pm to 3:30pm at Bobbitt Memorial Chapel, 1299 E. Highland Ave. Michael Freedman, Gang Tian, and Zhenghan Wang spoke at the funeral, and their speeches were included below.

Michael Freedman's Speech:

Thank you...thank you for letting me be here. We knew Lin as a man of really wonderful and great courage, a very gentle man, firm on principle, but extraordinarily kind and careful to detail. Very easy going. But when things mattered, he was firm.

I apologize for my speaking voice it's hard to follow "Amazing Grace" (a song).

You know, I'll call Xiao-Song, Lin, and I'll tell you why at the end, but Lin would tell me about his past in China when he was a graduate student. And for me it was incredible, I couldn't visualize this, his teen years. He was apparently, in the Cultural Revolution, he was working in a steel factory, and he would manipulate these huge ladles of molten metal above his head. He would have metal bars sliding along a track and poured them, you know . . . it sounded like Dickens through the 19th century. And I sometimes thought that some of the steel got into him and became the strength of him.

I felt with Lin, he was one of my first PhD students, but I felt toward him more like my mountaineering colleagues, because we kind of ignored his thesis and we went off on an expedition; an intellectual one. We were trying to do something that in the end we couldn't do; which is very risky business for young mathematicians. If you try as hard as you can to do something, this was called the A-B slice problem. If you try very hard and you fail, there was a risk. I mean, you know just as in mountaineering, you can fall off a cliff. In mathematics you can as good as fall off a cliff. If you try something very difficult and you fail, that can be the end. So we started to crawl down from this mountain alive. We started going through the snowstorms. But it was an adventure. It's still an open problem for students in the audience who maybe want to think about. It was a real bonding experience with Lin.

And I used to call him up, sometimes late at night, because I had an idea that I wanted to bounce off him. He was very good at certain calculations I didn't really know how to do and he could know something and I would feel like it and I could write them up.

Now I will tell you why I call him Lin: because if I called, and somebody answered his house and I said to them "can I speak to Xiao-Song?" They wouldn't know who I was talking about because my intonation wouldn't be correct. Even Lin wouldn't work too well. I tell you, "Can I speak to Lin?"; "Who?" "Lin," (in different pronunciation), "who?" "Lin, Lin, Lin, Lin, Lin." Eventually, I get it. And they would get it. That's probably Tian. was it? (Someone in the audience said something).

So I have to apologize for this actually, when I was a small child I had a high fever, a pneumatic fever. Well the fever would improve my ability to do math and damage my hearing a little bit. Not very much, but enough that the subtleties. So you know I was never able to penetrate Chinese as slight as the first two names. Lin was as close as I could come. Lin very graciously allowed me to use his surname and I think he thought we were always close for it.

I just wanted to say a couple more words about those early days. So in the math department in UCSD where we were; of course Yau was there, Professor Yau. And you know I was an observing young man; I looked around and I realized that Yau was onto something. That if you had these incredibly smart students, well that was good. It leveraged what you could do quite a bit. Heck you didn't need to think quite so much. So Lin was sort of my Rick Schoen, my Tian. You know I could turn over the worse of it to him and expect by morning that something would have happened. There would be real progress as well. I don't want to give the impression that's all the work that we did, but what sticks in my mind is that expedition that we completed.

So I just want to say that I love Lin, and thank you for letting me be here.

Gang Tian's Speech:

Like all of you, I feel very sad to lose a special friend: Xiao-Song Lin.

I met Xiao-Song in early 1982 when we both went to Beijing University for our graduate studies. I was lucky to have shared a dormitory room with Xiao-Song in Building 29. It was a small room, so there were only three of us in the room, rather than the usual four. In fact, Xiao-Song and I shared a bunk bed. To my surprise, we all came from Nanjing, although we have never previously met. Xiao-Song and Ying-Qing Wu (the other person in the room) studied topology. I was doing calculus of variations. As we all know, Xiao-Song is a very pleasant person to be with. At that time, we did not know much and were all eager to gain new mathematical knowledge. We talked a lot. Of course, Xiao-Song became fond of knots. I remember that he often lay in bed and thought about mathematics. I guess that he was trying to visualize various knots.

In the Fall of 1983, the mathematics department of Beijing University recommended four people to study for their PhDs abroad. Xiao-Song and I were selected. I chose to study geometry with Prof. Yau, while Xiao-Song chose to study topology with Prof. Freedman. They were both then in UC San Diego. On September 10, 1984, we took the same flight to San Diego. I had the luck to go studying with Xiao-Song in San Diego for three more years. There were so many memorable moments and events.

When we just arrived in San Diego, the university arranged for us to stay in a host family house for one week. The host was a very kind, old lady. Her house sat on the top of a hill close to downtown La Jolla. It had an incredible view of Pacific Ocean. We often stood in the yard and watched the beautiful sunset. We talked. It was a completely new continent. There were many unknowns ahead of us. We missed our families. China was not fully open at that time. We did not know when we could go back to our home country and see our families again. A few months later, Xiao-Song told me that his wife, Jian-Pin, had been pregnant before he came to the States. I thought that Xiao-Song must have also wondered when he would see his son, Hai-Jian. But we both agreed that the future would be bright and felt hopeful.

During the day, we took the bus to school. One day after dinner at school, we decided to walk back to the host house. Quickly, we discovered that it was not so trivial. It was not because of the distance. Since we did not have a map and were in a totally new place, we found that many paths were topologically equivalent. Of course, we did not have any cell phones then. Even though we tried to use our memory to find a geodesic back home, it was only theoretically easy and very difficult in practice. Many small roads we tried ended up in someone else's home or back to where we started. After many failures, we eventually got home. It was already 2AM. We felt so ashamed to wake up the lady to let us in.

During the first year in San Diego, we shared an apartment in UCSD's student housing. Our lives then were simple, but pleasant and full of wonderful things. We made many friends. We learned a lot of mathematics. We adapted to a new life in America. Xiao-Song often told me his new findings in topology, knots, the A-B slice problem, and so on. I tried to tell him about some geometric analysis. Apparently, I did not do a good job because Xiao-Song never really got involved in doing much analysis in his research.

Each weekend we spent half a day to write a long letter to our wives. Of course, I never knew what Xiao-Song wrote to his wife, but we did often share with each other the good news we received from our families back at home. Xiao-Song had his first son in that year. I remember that he was so excited and offered to cook a few dishes for dinner. In fact, for two and half a years as graduate students at Beijing University, even though we shared a dormitory room, we did not talk much else other than mathematics. In the August of 1984, Xiao-Song invited me to visit his family in Suzhou. There I met his wife Jian-Pin. One month later, when Jian-Pin came to see Xiao-Song off at the airport, our families met. Since then, we all became friends. After one year, China opened its door fully and my wife was able to join me in San Diego. So I had to move to a bigger apartment. It was the end of our primitive social lives. My classmate Fangyang Zheng joked that the disappearance of the primitive societies was due to the appearance of families.

Xiao-Song always did things in his own way. It took him five tries to pass his driving license's road test. In fact, his driving skill was very good. Maybe he was thinking about knots during the tests. He bought a Fort Pinto as his first car. It was an interesting car with a funny shape. It was very unpopular among Chinese students, but Xiao-Song didn't seem to mind. In fact, I did not remember anyone else who bought the same car among graduate students at UCSD.

Xiao-Song was a very responsible person and a trustable friend. Last June, I forwarded to him a paper submitted to the journal Communication in Contemporary Mathematics. Shortly after, I heard that he was diagnosed with cancer. I wondered what I should do with the paper. To my surprise, in a few days, he processed the paper. I was moved. He did it after taking so many exhausting medical tests, not to mention the stress of learning that he had a life-threatening disease. Later, when he realized that he would no longer have the energy to continue as a chief editor, he thought carefully who could replace him and wrote us a long email about it. He made an excellent arrangement for continuing the journal which he had founded.

We all know that Xiao-Song was a very pleasant friend. Every time we met, after a brief standard conversation, he always said "I found something interesting", usually in mathematics. For many years, I learned many new things from him this way. Last November, when I came to see him, lying in the bed and suffering great pain caused by the spreading of cancer, he again told me that he found something interesting in mathematics. He had a few math books next to him. He talked about quantum computing. He suggested a few names to give lecture series. I admired his incredible courage and love for mathematics. I thought that Xiao-Song would win in his fighting against the terrible disease. He is a hero in my mind.

Xiao-Song did outstanding research in topology. He also did an excellent job in spreading mathematical knowledge. For many years, he went back to China to give lecture series. He was involved in organizing many summer schools in mathematics. He was also involved in organizing summer camps for talented high school students in China and taught classes. Many young people there benefited from his efforts. He made very significant contributions in the development of Chinese mathematics. His contributions will be remembered.

Xiao-Song, we miss you, your smile, your friendship and the interesting mathematics that you shared with us. Your mathematics and achievements will be with us forever.

May Xiao-Song rest in peace.

Zhenghan Wang's Speech:

We are here to celebrate Xiao-Song's life. He brings us together today, just as he brought so many people together during his life. I speak for Feng Luo, Zhengxu He, Wenxiang Wang, and Yiping Wang, who are very close friends of Xiao-Song.

Xiao-Song was much more than a friend, he was like a big brother to us. When we came to America decades ago, we left behind our parents and friends. In Xiao-Song's home—a home of mathematics and love, we found a new home with a big brother to lean on. Xiao-Song is a man of few words with a famous smile, a smile that warmed your heart instantly and melted away your worries. His few words could brighten an entire room.

His passion for mathematics did not diminish even during the most difficult time of his life. When I came back from China last summer, I went to visit him in September. At that time, he already had trouble walking. It took him 30 minutes to walk from the bedside to the bathroom. While Jian-Pin was preparing lunch, we were talking about mathematics. He showed me a letter from an amateur mathematician which explained a wood knot puzzle. Then he gave me a wood knot puzzle to solve that he bought during a trip. I failed the challenge. He took the puzzle, and quickly unknotted the arcs. His signature hand movements painted the world's most beautiful knots.

Xiao-Song dedicated his life to the discovery of new mathematics. In the dictionary of Xiao-Song Lin, life is the same as mathematics. During the last week of his life, he only woke up intermittently. During a moment of consciousness, he mumbled, "I had solved the problem".

Words are not enough to describe Xiao-Song's humbleness and kindness. In November Xiao-Song was bed-bound at this critical juncture of his life, and I went to visit him again. I would drive back from Riverside to Santa Barbara during the night. About 9:00pm, he reminded me to leave as he worried that it might be too late. As I left, and before I closed the hallway door, I turned around. There again, he was looking at me smiling. But his right hand was clutching hard onto the bed sidebar, and his face was red. Both were indications of acute pain. He told Jian-Pin not to keep me long. This was my last sight of Xiao-Song.

Xiao-Song drew people to him like a magnet. In December we solicited writings from friends to encourage him. We were privileged to read so many letters written to him. The number of lives that have been touched by Xiao-Song is just astounding. The book, when finished, will show you the love that all kinds of friends poured out for him.

In September Xiao-Song wrote a letter to a friend.

"*When I first came to America from a closed society, I experienced many culture shocks. But it was the spirit of freedom that shocked me the most. Over the years, I became convinced that in order to pursue something of eternal value, you have to free yourself from other irrelevant thoughts. Only then can you be creative and original. I think it is reasonable to call such a spirit the American spirit, because it could be seen in all aspects of life throughout American history.*"

Xiao-Song was the finest combination of this American spirit and the ancient Chinese culture. He came from a paradise on earth-Suzhou, and he would go to heaven for a rest.

Nothing said more about the bravest woman I know, a proud mathematician's wife, Jian-Pin, than Xiao-Song's own words: I have no regrets in life.

Xiao-Song, you live in our hearts forever!

Thank you all.

Freedman's Writings on Lin

Xiao-Song studied under Michael Freedman from 1984-1988 at UCSD. In September 2006, Freedman wrote Xiao-Song on the inside cover of the Soviet book, `How the Steel was Tempered`*, to encourage him. Then in December 2006, he wrote Xiao-Song again during the critical juncture of Xiao-Song's life. Below are the letters.*

1. On the inside cover of: How the Steel was Tempered

Dear Lin,

This seems to be one of the many great stories of struggle and passion. I've read dozens in the mountaineering genera—the pacifist's field of battle. But I think all this Human struggle—scurrying to and fro—on what at the time seem critical missions, is cut from a single cloth. Very little of what is won survives, but maybe the stone of our world is eventually polished down a fraction by Millennia of our discordant efforts. You and I can take pleasure in having carved into a more durable medium—mathematics. There will ever after be string links and finite type invariants. As I learn physics I become increasingly religious in the sense of feeling connected to all other things. It feels as if we are ripples on a lake rather than separate entities. Even time-as Einstein wrote-is "only a stubbornly persistent illusion."

Best Wishes,

Mike

2. For the LinBook

Dear Lin,

Permit me to write you, your friends, and Family all at the same time a kind of "open letter". If I count correctly you were the second student that I took on for a Ph.D. (Fred Hickling being my first), but I recall that you and he may have graduated in the same year so you are also my "first student". Well, I probably gave you a lot of useless advice being inexperienced. This was a problem I had my whole life: since I didn't have a normal education, I never quite felt comfortable dispensing education - I had no idea, particularly in the beginning, what was normal, expected or proper. So we just winged it, you and I, and it did turn out fine. You wrote a great thesis, invented string links, collaborated with Nathan Habbeger, and did as much as anyone else has in the last 25 years to unravel the A-B slice problem. Sometimes when I think of our work on that problem I feel like an old time mountaineer stormed off a high peak just short of the summit. Intense efforts have a cost, I don't know how the experience affected you, but the effort, like a marathon run too hard, left me unable to concentrate fully on topology. I felt beaten. Happily we both recovered and found good things to think about. I'm glad we both ended up in the quantum world. I appreciated your coming up to Redmond - was that 1998? - and reversing the roles. Then you were my teacher as I was trying to catch up in quantum topology in order to think about computing. Some people generate a mythology. You have some unintended talent in this area. I think my stories about you have some foundation in truth but perhaps I have exaggerated the imagery, or perhaps not. Often in explaining to students that they should not be discouraged, how that can catch and exceed their child prodigy peers (if the mood blows hard into their sails) I like to mention Kevin Walker growing up in South Carolina where, in his assessment, "no one knew calculus within a fifty mile radius of my high school." But my favorite image in young Lin weathering the cultural revolution: As my story goes you are this tiny human figure in a chthonian steel foundry manipulating with tongs and pikes giant buckets of molten metal gimbaled precariously over your head. I think this is true.

When I think of what you have accomplished I am awed: getting to America, adapting to America, raising your beautiful family, instilling inquiry and insight into your sons, developing your university and your journal, your beautiful work in mathematics, and all the while staying close and available to your many friends. All these things reflect your courage and imagination. These are the values in which I continue to trust. With all my respect and love and the best wishes of my family,

Mike

Jian-Pin He's Speech on July 27, 2007

We are here today to commemorate Xiao-Song's 50th birthday. I feel very honored that all of you are here to remember Xiao-Song and to celebrate his birthday with me and my family. Since 1993, Xiao-Song spent almost every summer in China, and so I was unable to be with him on his birthday for many years.

I still have difficulty believing that Xiao-Song has actually left us. Sometimes I feel like it is just a dream that he is gone. I feel like Xiao-Song is just on a trip, giving a lecture at a far away university, and that he will be home soon.

I vividly remember the day that Xiao-Song first received his diagnosis. We were shocked, but at the same time, he was so brave and calm. In the months that followed, he showed us an amazing amount of strength and courage. He was able to enjoy life and to maintain a positive outlook, even as he accepted that the end of his life was near. I admired Xiao-Song so much! Everyone agreed that Xiao-Song deserved a miracle and deserved to live. All of your support helped to keep Xiao-Song going during the final period of his life, and has helped me to get through this most difficult time of my life.

Xiao-Song was a wonderful husband, the best husband I could have possibly asked for. While we were dating, he told me that his dream was to become a successful mathematician, and that he believed that we were the perfect match and that our marriage would be one that everyone would look up to.

Xiao-Song and I both grew up in Suzhou, China. Due to the Cultural Revolution, we did not get the chance to enter college after graduating from high school. Instead, we were both assigned to work at the same factory on the same day. That was the day we first met. We worked in that same factory for three years. After the Cultural Revolution ended, and after taking many difficult college entrance exams, we were assigned by the government to the same college. The funny thing about this was that neither of us even applied to that college. But it was a fortunate coincidence.

We were married on January 18, 1984. During our 23 years of marriage, we loved, respected, understood, supported, and relied on each other unconditionally. We talked about everything, and the level of trust between us was incredible. We were meant for each other; we were truly soul mates.

Xiao-Song was a dedicated mathematician. He told me that when he was writing papers and doing research, his goal was to make his work accessible and to help others enjoy the beauty of math. He also told me that the greatest reward for him was that his work could create new research paths for students and other mathematicians. Xiao-Song also loved teaching. He was a patient teacher who always made sure to spend extra time explaining subtle and difficult concepts. He was known by his students to draw beautiful geometric pictures and diagrams during his lectures. I believe that anyone who attended his lectures could see his passion for teaching.

During the last period of his life, he was still thinking about math. While he was unconscious, his hands would sometimes move about in the air, as if he was writing on a chalkboard. On January 1, 2007, he awoke for a few seconds and murmured, "I solved that problem." We can only now wonder what problem he managed to solve.

I would like to thank the Chern Institute of Mathematics for organizing and hosting this conference in memory of my beloved husband. Special thanks go to Professor Mo-Lin Ge and Professor Weiping Zhang for their efforts and their hospitality. I would also like to thank Professor Zhenghan Wang and the organizing committee for making this meeting so memorable. This meeting is truly a wonderful tribute to Xiao-Song and I appreciate it very much.

Xiao-Song left us prematurely, and there are no words that can express my deep sadness and sorrow from this tragic loss. All good memories of Xiao-Song will carry me and keep me going. I am so proud of Xiao-Song's accomplishments and proud of being his wife, a mathematician's wife.

Xiao-Song will live in our hearts forever!

Jian-Pin He

TOPOLOGY AND PHYSICS

Proceedings of the Nankai International Conference in Memory of Xiao-Song Lin

This unique volume, resulting from a conference at the Chern Institute of Mathematics dedicated to the memory of Xiao-Song Lin, presents a broad connection between topology and physics as exemplified by the relationship between low-dimensional topology and quantum field theory.

The volume includes works on picture (2+1)-TQFTs and their applications to quantum computing, Berry phase and Yang–Baxterization of the braid relation, finite type invariant of knots, categorification and Khovanov homology, Gromov–Witten type invariants, twisted Alexander polynomials, Faddeev knots, generalized Ricci flow, Calabi–Yau problems for CR manifolds, Milnor's conjecture on volume of simplexes, Heegaard genera of 3-manifolds, and the (A,B)-slice problem. It also includes five unpublished papers of Xiao-Song Lin and various speeches related to the memorial conference.

World Scientific
www.worldscientific.com
6907 hc
ISSN 1793-1118
ISBN-13 978-981-281-910-9
ISBN-10 981-281-910-X
9 789812 819109