Springer Proceedings in Mathematics & Statistics

Sofia Lambropoulou
Doros Theodorou
Petros Stefaneas
Louis H. Kauffman *Editors*

Algebraic Modeling of Topological and Computational Structures and Applications

THALES, Athens, Greece, July 1–3, 2015

EXTRAS ONLINE

Springer

Springer Proceedings in Mathematics & Statistics

Volume 219

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at http://www.springer.com/series/10533

Sofia Lambropoulou · Doros Theodorou
Petros Stefaneas · Louis H. Kauffman
Editors

Algebraic Modeling of Topological and Computational Structures and Applications

THALES, Athens, Greece, July 1–3, 2015

Springer

Editors
Sofia Lambropoulou
Department of Applied Mathematics
National Technical University of Athens
Athens
Greece

Petros Stefaneas
Department of Applied Mathematics
National Technical University of Athens
Athens
Greece

Doros Theodorou
School of Chemical Engineering
National Technical University of Athens
Athens
Greece

Louis H. Kauffman
Department of Mathematics
University of Illinois at Chicago
Chicago, IL
USA

Additional material to this book can be downloaded from http://extras.springer.com.

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-319-68102-3 ISBN 978-3-319-68103-0 (eBook)
https://doi.org/10.1007/978-3-319-68103-0

Library of Congress Control Number: 2017953823

Mathematics Subject Classification (2010): 00A, 00B, 03A, 03B, 03C, 05E, 11G, 12D, 13A, 14H, 16S, 20C, 20F, 37B, 42A, 55P, 55S, 57M, 57N, 57R, 68N, 68Q, 68R, 68T, 68U, 82-, 82D, 92B

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book comprises a collection of papers presented in the THALES workshop "Algebraic modeling of topological and computational structures and applications", that took place at the National Technical University of Athens (NTUA), 1–3 July 2015. This workshop disseminated the results of the research project THALES MIS 380154 with the same title, which was implemented from October 2011 until December 2015. The project has been co-financed by the European Union (European Social Fund—ESF) and Greek National Funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)—Research Funding Program: THALES: Reinforcement of the interdisciplinary and/or interinstitutional research and innovation.

The research project THALES was concerned with the study of topological and computational structures and applications, mainly with the use of algebra and especially with braid groups. Braids can be viewed as algebraic as well as topological objects and they play a crucial role in knot theory and in low-dimensional topology, in the study of homotopy groups, in reflection groups and C*-algebras, in statistical mechanics, in cryptography, and in Galois theory. The project consisted in three research programs (RP), corresponding to the three research groups involved:

RP1 "Algebraic modeling of topological structures"

RP2 "Algebraic modeling of applications"

RP3 "Algebraic modeling of computational structures"

RP1 aimed at the study of topological structures, such as knots, links, 3-manifolds and classical homotopy groups, using braid groups, classical and generalized, and their representations. Knot theory is the area of low-dimensional topology that deals with the problem of classification of embeddings of a circle or collections of circles into three-dimensional space. This problem is tackled by the construction of knot invariants. Within this project, new skein link invariants were extracted from the Yokonuma-Hecke algebras and other new (framization) knot algebras, via the celebrated method of V.F.R. Jones, which uses braid groups and the Alexander and Markov theorems. Another subproject was about the study of the mixed braid groups related to knots and links in specific 3-manifolds, and the construction of quotient

algebras of the mixed braid groups. Further, the HOMFLYPT skein module of the lens spaces was investigated via the Artin braid groups of B-type. The study of the knot theory of a 3-manifold renders information about the topological structure of the manifold, while the use of braids in the study provides more structure and more control on the moves for topological equivalence. Powerful algebraic and computational tools can then be employed. Our next focus was on special decomposable groups. In particular, the linearity of groups that fit into a short exact sequence with kernel a free group and cokernel a linear group was studied. Finally, the connections between modular invariant theory and certain unstable coalgebras over the mod p Steenrod algebra were investigated. This problem is related to the stable homotopy groups of spheres.

RP2 was concerned with the study of some novel application areas. One of them was the development of new measures of polymer entanglements. Polymer chains are long flexible molecules that impose topological constraints on each other, called entanglements, which affect the physical properties of the polymer. The Gauss linking number was extended to open polymeric chains in 1, 2 and 3 Periodic Boundary Conditions (PBC) and the simulation of polymeric chains through braid groups was proposed. Also, Fourier braids were introduced and PBC were represented by algebraic conditions. Furthermore, Turaev's knotoids were retaken, which are ideal for modeling, abstractly, polymers and biopolymers. The knotoids are a new chapter in classical knot theory that is worth studying for its own sake and that allows topological modeling of open-ended arcs in three-dimensional space (such as long chain molecules, proteins, DNA). Another application area was the molecular simulation of ionic liquids and their mixtures. Molecular simulation is a powerful tool for the study of physical systems, based on fundamental principles of statistical mechanics. This simulation manages to predict various properties of materials through the connection of their microscopic structure and their macroscopic properties. It can efficiently contribute to the production of new materials with desired properties. The modeling of many natural processes via topological surgery in 1, 2, and 3 dimensions was another application of the project. Topological surgery is a mathematical technique used for creating new manifolds out of known ones and, as we observed, it appears in nature in numerous, diverse processes of various scales as, for example, in the reconnection of cosmic magnetic lines, in DNA recombination, in the formation of tornadoes and of Falaco solitons, in drop coalescence, in cell mitosis, and in the formation of black holes. Inspired by such phenomena new theoretical concepts were introduced, which enhance topological surgery with the observed dynamics, and a connection with a 3D Lotka–Volterra dynamical system was also pinned down.

RP3 was concerned with the algebraic modeling of computational structures and applications. A subproject was about the unification of the well-known algebraic specification language CafeOBJ with the strong theorem prover Athena within a common interface. Other outputs included the development of novel techniques for system modeling and verification, as well as applications on the modeling of video and of musical structure. Algebraic modeling techniques were also applied in the geometry of curves over finite fields and applications to cryptography and coding

theory were investigated. The aim was the study of open problems referring to zeta functions of certain algebraic curves aiming to calculate the number of rational elements. Finally, another application was in the field of medical imaging, where a new medical PET (Positron Emission Tomography) data reconstruction algorithm was proposed for reconstructing sinograms to tomographic images. Also, various deformable methods for optic disk extraction in retinal images were studied and evaluated.

The research team of the project THALES comprised 56 researchers from universities and research institutions from all over the world. The Scientific Coordinator of the project and coordinator of the first RP was Professor Sofia Lambropoulou of the NTUA. Professor Doros Theodorou of the NTUA was the coordinator of the second RP and Assistant Professor Petros Stefaneas of the NTUA was the coordinator of the third RP. Finally, Professor Louis H. Kauffman of the University of Illinois at Chicago was the Invited Researcher of the project.

The THALES workshop was attended by more than 100 researchers. In total, 37 research talks and 17 posters were presented. The present book contains 23 chapters, arranged into three parts that correspond to the three RPs of the project.

Part I: **Algebraic Modeling of Topological Structures**

A knot algebra is an algebra obtained as a quotient of the group algebra of a braid group and endowed with a Markov trace; it can be thus used for the definition of knot invariants. Framization is a mechanism proposed recently by Juyumaya and Lambropoulou which consists of constructing a nontrivial extension of a knot algebra for the definition of framed knot invariants. The inspiring example of framization is the Yokonuma-Hecke algebra of type A, introduced by Yokonuma in the context of finite reductive groups as a generalization of the Iwahori–Hecke algebra of type A. The first five chapters of Part I are concerned with framizations of known knot algebras.

In Chap. 1 by Konstantinos Karvounis and Sofia Lambropoulou, the families of framed, classical, singular, and transverse link invariants defined via the Yokonuma-Hecke algebras of type A are presented. The Yokonuma-Hecke algebras comprise a family of algebras that generalize the classical Iwahori–Hecke algebra and that also support Markov traces that give rise to those families of link invariants. The family of classical link invariants is of special interest, since it contains the HOMFLYPT polynomial P and, moreover, it extends to a three-variable skein link invariant generalizing P and which is stronger than P.

Chapter 2 by Dimos Goundaroulis is a brief and comprehensive review of the construction of the framization of the Temperley–Lieb algebra of type A and its derived invariants for framed and classical links. Key elements of the representation theory of the involved quotient algebras are also included. The invariants for classical links are compared to the Jones polynomial and then they are generalized to a two-variable invariant that is stronger than the Jones polynomial.

In Chap. 3, Maria Chlouveraki studies the algebraic structure and the representation theory of the Yokonuma-Hecke algebra of type A, as well as of some

similar framizations of the affine Hecke algebra of type A, the Ariki-Koike algebra, and the Temperley–Lieb algebra.

Furthermore, in Chap. 4, Loic Poulain d'Andecy studies the affine Yokonuma-Hecke algebra, seeing it as a quotient of a certain braid group. A large family of Markov traces is constructed and, using well-known relations between braids and links, this way invariants for classical links and for links in the solid torus are produced. The study uses extensively a purely algebraic result, namely an isomorphism theorem relating affine Yokonuma-Hecke algebras with the usual affine Hecke algebras, which moreover allows one to deduce naturally some properties of the invariants.

In Chap. 5, Marcelo Flores gives a review about two framizations of the Hecke algebra of type B, and the results related to each of these. He begins by presenting the cyclotomic Yokonuma-Hecke algebra introduced by Chlouveraki and Poulain d'Andecy, which provides one of such framizations. Next, Flores focuses on a new framization defined recently by himself together with Juyumaya and Lambropoulou, and on the results obtained for this new algebra. The author concludes with a preliminary comparison between the isotopy invariants derived by both framizations.

In order to have a good understanding of knot theory in 3-manifolds, one should be able to visualize and understand link diagrams in these manifolds. In Chap. 6, Bostjan Gabrovsek and Maciej Mroczkowski present a survey of the so-called arrow diagrams, which are used for representing links in Seifert fibered spaces, a large class of 3-manifolds. Moreover, they show how to pass between some of the different types of diagrams found in the literature. Using arrow diagrams, the authors express the basis of the Kauffman bracket skein module and the HOMFLYPT skein module of some 3-manifolds and they present new bases for these skein modules for the solid torus and lens spaces.

In Chap. 7, Ioannis Diamantis and Sofia Lambropoulou present recent results toward the computation of the HOMFLYPT skein module of the lens spaces $L(p,1)$, $\mathcal{S}(L(p,1))$, using the braid approach. They describe first the HOMFLYPT skein module of the solid torus ST, $\mathcal{S}(ST)$, using the mixed braid group $B_{1,n}$, which is the Artin braid group of type B, and they present a new basis, Λ, for it, through which the braid band moves are naturally described. The authors then derive the relation between $\mathcal{S}(ST)$ and $\mathcal{S}(L(p,1))$ and show that in order to compute $\mathcal{S}(L(p,1))$ one needs to solve a controlled infinite system of equations obtained by performing all possible braid band moves on elements in the basis Λ.

In Chap. 8, Dimitrios Kodokostas and Sofia Lambropoulou define a tower of Hecke-type quotient algebras of the mixed braid group with two fixed strands, $B_{2,n}$. The groups $B_{2,n}$ are related to the knot theory of the handlebody of genus two, the complement of the 2-unlink in S^3 and the connected sums of two lens spaces. The authors focus on the algebras $H_{2,n}(q), n \in \mathbb{N}$ and review their work on extracting inductive spanning sets for these algebras, appropriate for constructing Markov traces. They also provide corresponding spanning sets for the other algebras defined in the paper and they conjecture all spanning sets to be linear bases for the algebras defined.

In Chap. 9, Valerij Bardakov proves that the kernel of the map from the braid group in the handlebody, $B_{g,n}$, to the classical braid group is a semi-direct product of free groups. Also, he introduces an analogue of the Hecke algebra for the braid group in the handlebody and formulates a conjecture on the basis of this algebra.

In Chap. 10, the research is twofold. First, Nondas Kechagias gives an invariant theoretic description for the $mod\,p$ cohomology of the stabilization functor on the basepoint of the zero sphere previously described by D. Quillen and M. Barratt—S. Priddy independently. Second, the author compares the $mod\,p$ Dyer-Lashof algebra with a co-free Steenrod coalgebra and he provides cogenerators along with corelations. His approach is connected with the Peterson conjecture relating the Dickson algebra with a quotient of a free unstable Steenrod algebra.

Part II: **Algebraic Modeling of Applications**

In Chap. 11, Eleni Panagiotou and Ken Millett extend the topological Gauss linking number to open chains in systems employing one-dimensional periodic boundary conditions to define periodic linking and periodic self-linking numbers. These give rise to a periodic linking matrix and associated eigenvalues that are applied to Olympic gels and tubular filamental structures.

The theory of knotoids, that was introduced in 2012 by Vladimir Turaev, was proposed as a new diagrammatic approach to classical knot theory and also as a generalization of classical knot theory. The notion of knotoids extends the notion of 1-1-tangles or long knots by having two distinct endpoints that may lie in any region of knotoid diagrams. This makes knotoids a natural domain for understanding physicality and topology of open-ended space curves and also for a transition and relationships with virtual knot theory. The height of a knotoid, that measures the distance between the endpoints of a knotoid, is an efficient tool for the classification of knotoids. In Chap. 12, Neslihan Gügümcü and Louis H. Kauffman review their results on the estimation of the height of a knotoid given by the affine index and the arrow polynomials. These polynomials, first defined in virtual knot theory, are here given definitions in the category of knotoids.

In Chap. 13, Stephan Klaus constructs braids by folding periodic complex valued functions. The method of finite Fourier approximation shows that braids are given by certain finite Laurent polynomials $g(z)$ such that an associated algebraic discriminant has no root on the unit circle. This condition is studied from the algebraic and topological viewpoint. Algebraic transformations of $g(z)$ correspond to geometric operations of the braid, for example, a Dehn twist. Moreover, this algebraic method allows one to construct braids in a thickened torus or in other spaces.

Chapter 14 is devoted to the computational study of ionic liquids at the molecular level. Ionic liquids are organic salts that are in the liquid state at room temperature and exhibit a fascinating combination of properties due to their dual ionic and organic nature that renders them ideal for use in a wide range of state-of-the-art applications. Niki Vergadou applies molecular simulation methods for the investigation of the microscopic phenomena that determine macroscopic properties of ionic liquids and for the study of their complex dynamics and structure.

In Chap. 15, Stathis Antoniou and Sofia Lambropoulou show that topological surgery happens in natural phenomena at various scales and analyze the common features of their underlying processes. These features are captured by schematic models and new definitions which provide the theoretical setting for describing the topological changes that occur. We believe that the understanding of their underpinning topology will lead to the better understanding of the phenomena themselves, as well as to new mathematical notions, which will in turn lead to new physical implications.

Part III: **Algebraic Modeling of Computational Structures**

Chapter 16 is a survey on the theory of curves with automorphisms. Several bounds on the order of the automorphism group curve are discussed, mostly related to the research of the authors, Jannis Antoniadis and Aristides Kontogeorgis. The authors' point of view is to relate the theory of (modular) representations of automorphism groups acting on natural objects related to the curve, such as (co)homology spaces and spaces of holomorphic (poly)differentials. The chapter concludes with applications to deformation theory of curves, liftings to characteristic zero and moduli problems.

Chapter 17 by Nicola Angius, Maria Dimarogkona, and Petros Stefaneas provides a first attempt at constructing semantic theories over institutions and examines the logical relations holding between different such theories. The results show that this approach can be very useful for theoretical computer science and may also contribute to the current philosophical debate regarding the semantic and the syntactic presentation of scientific theories.

Chapter 18 is a survey of results related to generic constructions and generic limits for semantic and syntactic cases. The authors Sergey Sudoplatof, Yiannis Kiouvrekis, and Petros Stefaneas consider both the pure model theory approach and the institutional approach.

Chapter 19 by Katerina Ksystra, Nikos Triantafyllou, and Petros Stefaneas contains some steps towards a verification framework based on the combination of the CafeOBJ algebraic specification language with the interactive theorem proving system Athena. The proposed framework combines two different specification and theorem proving systems, in order to facilitate the modeling and analysis of critical software systems.

Chapter 20 by Theodoros Mitsikas, Petros Stefaneas, and Iakovos Ouranos demonstrates the design of an innovative rule based approach for the Air Traffic Control regulations during the takeoff and landing phases, covering both current and future separation standards of ICAO and FAA. The rule base consists of the rules implementing the air traffic control regulations, and a database containing characteristics of airports and aircraft. The proposed rule base constitutes a flexible tool for the computation of the aircraft separation according to current and future regulations, useful in the fields of conflict detection, conflict avoidance, and scheduling aircraft landings. A further application will be for a decision support tool in real-time environments, guaranteeing the enforcement of all the separation standards.

In Chap. 21, Marianthi Bozapalidou demonstrates that every commutative affine musical contour actually simulates the classical one. Affine contours may be viewed as an abstraction of the notion of musical intervals and are closely related to sequential machines. Sequential machines are mathematical tools that can be depicted by finite directed graphs and are suitable to describe and represent various phenomena in music.

In Chap. 22, Antonios Kalambakas, Nikolaos Triantafyllou, Katerina Ksystra, and Petros Stefaneas indicate how hyperoperation can be utilized in order to formally represent and compare video content using algebraic semiotics. In this setup, pictures are defined as a specific type of rectangular graphs and the picture hyperoperation is given by virtue of the notion of the path inside such a picture.

Finally, in Chap. 23 Evi Karali presents and evaluates a new iterative algorithm for medical image reconstruction, under the name Image Space Weighted Least Squares (ISWLS). She then compares various snake or deformable methods, towards a modified self-affine mapping system technique, more suitable for weak edge detection. All methods were applied to glaucomatic retinal images with the purpose of segmenting the optical disk.

We thank all participants of the workshop and all contributors for their great efforts in making the workshop a success and this book possible.

Athens, Greece — Sofia Lambropoulou
Athens, Greece — Doros Theodorou
Athens, Greece — Petros Stefaneas
Chicago, USA — Louis H. Kauffman

Contents

Part I
Algebraic Modeling of Topological Structures

Chapter 1
Link Invariants from the Yokonuma–Hecke Algebras

Konstantinos Karvounis and Sofia Lambropoulou

Abstract The Yokonuma–Hecke algebras are naturally related to the framed braid group and they support a Markov trace. Consequently, invariants for various types of links (framed, classical, singular and transverse) are derived from these algebras. In this paper, we present results about these invariants and their properties. We focus, in particular, on the family of 2-variable classical link invariants that are not topologically equivalent to the HOMFLYPT polynomial and on the 3-variable classical link invariant that generalizes this family and the HOMFLYPT polynomial.

Keywords Classical braids · Framed braids · Yokonuma–Hecke algebras Markov trace · Framed knots and links · E–condition · Classical knots and links Transverse knots and links · Singular braid monoid · Singular knots and links HOMFLYPT polynomial

2010 Mathematics Subject Classification 57M27 · 57M25 · 20F36 · 20F38 · 20C08

Introduction

The first example of construction of link invariants via braid groups is the Jones polynomial [27]. It can be defined by means of a *knot algebra*, that is a triple (A, π, τ) where A is an algebra, π is a representation of the braid group in A and τ is a Markov trace on A. The Jones polynomial is obtained via the Jones' trace on the Temperley–Lieb algebra. This construction generalizes to the HOMFLYPT (or 2-variable Jones)

K. Karvounis
Institut für Mathematik, Universität Zürich,
Winterthurerstrasse 190, 8057 Zürich, Switzerland
e-mail: konstantinos.karvounis@math.uzh.ch

S. Lambropoulou (✉)
Department of Applied Mathematics, National Technical University of Athens,
Zografou Campus, 15780 Athens, Greece
e-mail: sofia@math.ntua.gr
URL: http://www.math.ntua.gr/~sofia

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_1

polynomial [19, 39, 43] using as a knot algebra the Iwahori–Hecke algebra of type A and the Ocneanu trace [27].

Some years ago, the Yokonuma–Hecke algebras of type A [47] received attention. These algebras have "framing" generators and they are naturally related to the framed braid group. We denote them as $Y_{d,n}(q)$, where n corresponds to the number of strands of the framed braid group $\mathscr{F}_n$, $d \in \mathbb{N}$ imposes a modular condition on the framing generators and q is a non-zero complex number. For $d = 1$ the algebra $Y_{1,n}$ coincides with the Iwahori–Hecke algebra. The representation theory of the Yokonuma–Hecke algebras has been extensively studied in [13, 45]. J. Juyumaya in [29] defined a unique Markov trace on the Yokonuma–Hecke algebras, denoted by tr_d, making $Y_{d,n}(q)$ into a knot algebra and a natural candidate for defining framed and classical link invariants. Surprisingly, the trace tr_d could not be directly re-scaled for the negative stabilization move of the framed braid equivalence. For this, a condition was needed to be imposed on the framing parameters of tr_d that, in turn, meant that they should satisfy a certain non-linear system of equations, called the E-*system*. As it was shown by P. Gérardin [31, Appendix], the solutions of the E-system are parametrized by the non-empty subsets D of $\mathbb{Z}/d\mathbb{Z}$.

Consequently, in [31] and in [32] an infinitum of framed and classical link invariants respectively were defined, parametrized by d and the subsets D. Both of these families of invariants contain the HOMFLYPT polynomial for $d = 1$. Furthermore, the Yokonuma–Hecke algebras were proved to be suitable for defining invariants for other classes of links, such as singular links [33] and transverse links [10].

With the classical link invariants in hand, the next natural question, which remained as a long-standing open problem, was whether these invariants are *topologically equivalent* to the HOMFLYPT polynomial P in the sense that they distinguish or not the same pairs of links. The first step was taken in [12], where it was proven that these invariants do not coincide with P except for the trivial cases when $|D| = 1$ and $q = \pm 1$. The Yokonuma–Hecke algebras have a quite complex quadratic relation for the braiding generators that involves some idempotent elements, denoted by e_i, that are sums of products of the framing generators. Computations were needed, and for this purpose a computer program was developed (see [35] and http://www.math.ntua.gr/~sofia/yokonuma). Note that the invariants have been originally defined using a different presentation of the Yokonuma–Hecke algebras, denoted by $Y_{d,n}(u)$, that used a more complicated quadratic relation than that of the presentation $Y_{d,n}(q)$. By comparing the classical invariants on various pairs of knots and links, a conjecture for the case of knots was formulated in [10] and later proved in [11]. In both papers the new presentation $Y_{d,n}(q)$ is used. More precisely, the classical link invariants from the Yokonuma–Hecke algebras coincide with P on *knots*. Note that, by this result it follows that these invariants are not topologically equivalent to the Kauffman polynomial.

In [12] the *specialized Juyumaya trace* $\mathrm{tr}_{d,D}$ that is the trace tr_d with the framing parameters specialized to a solution of the E-system was defined. Now, in [11] it has been proved that the trace $\mathrm{tr}_{d,D}$ can be computed for classical braids by five rules that involve the braiding generators of $Y_{d,n}$ and the idempotents e_i, instead of the framing generators. This result makes the calculations for the trace $\mathrm{tr}_{d,D}$ easier, since

the elements e_i can be considered as formal elements in the image of the classical braid group in $\mathrm{Y}_{d,n}(q)$. By the same result it also followed that the invariants $\Theta_{d,D}$ of classical links are actually parametrized by the natural numbers and can be simply denoted as Θ_d. In order to compare the invariants Θ_d to P on classical links, a new program was developed [35] that uses the five rules for the trace $\mathrm{tr}_{d,D}$ when applied to images of classical braids. This program facilitated the comparison of the invariants Θ_d and P on several P-equivalent pairs of links (that is pairs of links having the same HOMFLYPT polynomial) and, as it turned out, the invariants Θ_d are not topologically equivalent to P on *links* [11]. This fact was also proved theoretically in [11], since the invariants Θ_d satisfy the HOMFLYPT skein relation, but only on crossings involving different components, and this enabled a diagrammatic approach to the definition of the invariants Θ_d.

Remarkably, our examples of P-equivalent pairs of links distinguished by Θ_d were distinguished for all $d \geq 2$ [11]. Furthermore, in the five rules of $\mathrm{tr}_{d,D}$ when restricted to classical braids only the value $E_D := 1/|D|$ appears that depends only on the cardinality of the set D. This led to the hypothesis that the value E_D can be seen as a parameter, resulting in the construction of a new 3-variable classical skein link invariant $\Theta(q, \lambda, E)$ that is stronger than the HOMFLYPT polynomial [11]. Moreover, W.B.R. Lickorish provided in [11, Appendix B] a closed formula for the invariant Θ that shows that it is a complicated mixture of linking numbers and the values of P on sublinks of a given link, providing thus a topological interpretation for the invariant Θ (see also [36, 42]). Finally, the construction of the invariant Θ led to an analogous generalization of the Kauffman polynomial and to new state sum models, using the skein theoretical methods for Θ [36].

The interest in the Yokonuma–Hecke algebras led to the notion of framization of knot algebras [34], the Yokonuma–Hecke algebra being the basic example, as the framization of the classical Iwahori–Hecke algebra. Consequently, appropriate Temperley–Lieb-type quotients of the Yokonuma–Hecke algebras were constructed and studied in [22–25], see also [15, 16]. Furthermore, Yokonuma–Hecke algebras related to type B have been constructed [14, 18, 34], equipped with Markov traces, and related link invariants for the solid torus have been derived. Finally, a framization of the BMW algebra has also been defined and studied [4, 34].

In this paper we present results mainly from [10, 11] on the Yokonuma–Hecke algebras and link invariants derived from them. The paper is organized as follows. In Sect. 1.1 we define the Yokonuma–Hecke algebras and provide some facts about them. Then, in Sect. 1.2, we recall the definition of the Markov traces tr_d and $\mathrm{tr}_{d,D}$ on $\mathrm{Y}_{d,n}(q)$ and we discuss some properties that they satisfy. In Sect. 1.3 invariants for framed, classical, singular and transverse links are presented, while in Sect. 1.4 we study further the classical link invariants. In Sect. 1.5 the 3-variable classical link invariant Θ is presented and we discuss various ways to prove its well-definedness. Finally, in Sect. 1.6 we recall framed and classical link invariants derived from another presentation of the Yokonuma–Hecke algebras and we discuss their relation to the ones derived from the new presentation $\mathrm{Y}_{d,n}(q)$.

1.1 The Yokonuma–Hecke Algebra

In this section we recall the definition of the Yokonuma–Hecke algebra as a quotient of the framed braid group.

1.1.1 The Framed Braid Group and the Modular Framed Braid Group

The *framed braid group*, $\mathscr{F}_n \cong \mathbb{Z}^n \rtimes B_n$, is the group defined by the standard generators $\sigma_1, \ldots, \sigma_{n-1}$ of the classical braid group B_n together with the framing generators $t_1, \ldots, t_n$ (t_j indicates framing 1 on the j-th strand), subject to the relations:

$$\begin{array}{ll} (\mathrm{b}_1)\ \sigma_i\sigma_j\sigma_i = \sigma_j\sigma_i\sigma_j & \text{for } |i-j| = 1 \\ (\mathrm{b}_2)\quad \sigma_i\sigma_j = \sigma_j\sigma_i & \text{for } |i-j| > 1 \\ (\mathrm{f}_1)\quad t_i t_j = t_j t_i & \text{for all } i, j \\ (\mathrm{f}_2)\quad t_j\sigma_i = \sigma_i t_{s_i(j)} & \text{for all } i, j \end{array} \tag{1.1}$$

where $s_i(j)$ is the effect of the transposition $s_i := (i, i+1)$ on j. Relations (b_1) and (b_2) are the usual braid relations, while relations (f_1) and (f_2) involve the framing generators. Further, for a natural number d the *modular framed braid group*, denoted $\mathscr{F}_{d,n}$, can be defined as the group with the presentation of the framed braid group, but including also the modular relations:

$$(\mathrm{m})\quad t_j^d = 1 \quad \text{for all } j \tag{1.2}$$

Hence, $\mathscr{F}_{d,n} \cong (\mathbb{Z}/d\mathbb{Z})^n \rtimes B_n$. Geometrically, the elements of $\mathscr{F}_n$ (respectively $\mathscr{F}_{d,n}$) are classical braids on n strands with an integer (respectively an integer modulo d), the framing, attached to each strand. Further, due to relations (f_1) and (f_2), every framed braid α in $\mathscr{F}_{d,n}$ can be written in its *split form* as $\alpha = t_1^{k_1} \ldots t_n^{k_n}\sigma$, where $k_1, \ldots, k_{n-1} \in \mathbb{Z}$ and σ involves only the standard generators of B_n. The same holds also for the modular framed braid group.

For a fixed d we define the following elements e_i in the group algebra $\mathbb{C}\mathscr{F}_{d,n}$:

$$e_i := \frac{1}{d}\sum_{1\le s\le d} t_i^s t_{i+1}^{-s} \qquad (1 \le i \le n-1)$$

where $-s$ is considered modulo d. One can easily check that e_i is an idempotent: $e_i^2 = e_i$ and that $e_i\sigma_i = \sigma_i e_i$ for all i.

1.1.2 The Yokonuma–Hecke Algebras

Let $d \in \mathbb{N}$ and let $q \in \mathbb{C}\backslash\{0\}$ fixed. The *Yokonuma–Hecke algebra* q, denoted $\mathrm{Y}_{d,n}(q)$, is defined as the quotient of $\mathbb{C}\mathscr{F}_{d,n}$ by factoring through the ideal generated by the expressions: $\sigma_i^2 - 1 - (q - q^{-1})e_i\sigma_i$ for $1 \leq i \leq n-1$. We shall denote g_i the element in the algebra $\mathrm{Y}_{d,n}(q)$ corresponding to σ_i while we keep the same notation for t_j in the algebra $\mathrm{Y}_{d,n}(q)$. So, in $\mathrm{Y}_{d,n}(q)$ we have the following quadratic relations:

$$g_i^2 = 1 + (q - q^{-1})\, e_i\, g_i \qquad (1 \leq i \leq n-1). \tag{1.3}$$

The elements $g_i \in \mathrm{Y}_{d,n}(q)$ are invertible:

$$g_i^{-1} = g_i - (q - q^{-1})\, e_i \qquad (1 \leq i \leq n-1). \tag{1.4}$$

Further the elements $g_i \in \mathrm{Y}_{d,n}(q)$ satisfy the following relations:

Lemma 1.1 ([10, Lemma 1.1]) *Let $i \in \{1, \ldots, n-1\}$. Then:*

$$g_i^r = (1 - e_i)\, g_i + \left(\frac{q^r + q^{-r}}{q + q^{-1}}\right) e_i g_i + \left(\frac{q^{r-1} - q^{-r+1}}{q + q^{-1}}\right) e_i \qquad \textit{for } r \textit{ odd},$$

$$g_i^r = 1 - e_i + \left(\frac{q^r - q^{-r}}{q + q^{-1}}\right) e_i g_i + \left(\frac{q^{r-1} + q^{-r+1}}{q + q^{-1}}\right) e_i \qquad \textit{for } r \textit{ even}.$$

The Yokonuma–Hecke algebras were originally introduced by T. Yokonuma [47] in the representation theory of finite Chevalley groups and they are natural generalizations of the Iwahori–Hecke algebras $\mathrm{H}_n(q)$. Indeed, for $d = 1$ all framings are zero, so the corresponding elements of $\mathscr{F}_n$ are identified with elements in B_n; also we have $e_i = 1$, so the quadratic relation (1.3) becomes the well–known quadratic relation of the algebra $\mathrm{H}_n(q)$:

$$g_i^2 = 1 + (q - q^{-1})\, g_i \qquad (1 \leq i \leq n-1).$$

Thus, the algebra $\mathrm{Y}_{1,n}(q)$ coincides with the algebra $\mathrm{H}_n(q)$. The Yokonuma–Hecke algebras can be also regarded as unipotent algebras in the sense of [45]. The representation theory of these algebras has been studied in [13, 45]. In [13] a completely combinatorial approach is taken to the subject. Further, in [26] a decomposition of the Yokonuma–Hecke algebra is constructed, as a direct sum of matrix algebras with coefficients in tensor products of Iwahori–Hecke algebras of type A, which is a special case of a result of G. Lusztig [40].

Following [29, Sect. 3], the algebra $\mathrm{Y}_{d,n}(q)$ has linear dimension $d^n n!$ and the set

$$\mathscr{B}_n^{\mathrm{can}} = \left\{ t_1^{k_1} \ldots t_n^{k_n} (g_{i_1} \ldots g_{i_1 - r_1}) \cdots (g_{i_p} \ldots g_{i_p - r_p}) \;\middle|\; \begin{array}{l} k_1, \ldots, k_n \in \mathbb{Z}/d\mathbb{Z} \\ 1 \leq i_1 < \cdots < i_p \leq n-1 \end{array} \right\}$$

is a $\mathbb{C}$-linear basis for $\mathrm{Y}_{d,n}(q)$. This basis is called the *canonical* basis of $\mathrm{Y}_{d,n}(u)$. Note that, in each element of the standard basis, the highest index generator g_{n-1} appears at most once.

Now, the natural inclusions $\mathscr{F}_n \subset \mathscr{F}_{n+1}$ give rise to the algebra inclusions $\mathbb{C}\mathscr{F}_n \subset \mathbb{C}\mathscr{F}_{n+1}$, which in turn induce the algebra inclusions $\mathrm{Y}_{d,n}(q) \subset \mathrm{Y}_{d,n+1}(q)$ for $n \in \mathbb{N}$ (setting $\mathbb{C}\mathrm{Y}_{d,0}(q) := \mathbb{C}$). We can construct an inductive basis $\mathscr{B}_n^{\mathrm{ind}}$ for $\mathrm{Y}_{d,n}(q)$ in the following way: we set $B_0^{\mathrm{ind}} := \{1\}$ and

$$\mathscr{B}_{n+1}^{\mathrm{ind}} := \{w_n g_n g_{n-1} \dots g_i t_i^k,\ w_n t_{n+1}^k \mid 1 \leq i \leq n,\ k \in \mathbb{Z}/d\mathbb{Z},\ w_n \in \mathscr{B}_n^{\mathrm{ind}}\},$$

for all $n \in \mathbb{N}$.

Remark 1.1 In the papers [12, 29–34] another presentation was employed for the Yokonuma–Hecke algebra that was using a more complex quadratic relation giving rise to more computational difficulties. We refer to this extensively in Sect. 1.6.

Remark 1.2 By the fact that the classical braid group B_n embeds in $\mathscr{F}_{d,n}$ (and in $\mathscr{F}_n$) and by relations (1.1) (b_1, b_2), there is a natural homomorphism from B_n to $\mathrm{Y}_{d,n}$, treating the framing generators t_j's as formal elements in the algebra. So, the algebra $\mathrm{Y}_{d,n}$ can be also used in the study of classical knots and links.

Note 1.1 In this paper we will sometimes identify algebra monomials with their corresponding braid words.

1.2 Markov Traces on the Yokonuma–Hecke Algebras

In this section we recall the definition of a unique Markov trace defined on the algebras $\mathrm{Y}_{d,n}(q)$, as well as a necessary condition on the trace parameters, needed for obtaining framed link invariants.

1.2.1 The Juyumaya Trace

By the natural inclusions $\mathscr{F}_n \subset \mathscr{F}_{n+1}$, which induce the inclusions $\mathrm{Y}_{d,n}(q) \subset \mathrm{Y}_{d,n+1}(q)$, and using the inductive bases of the algebras $\mathrm{Y}_{d,n}(q)$ we have:

Theorem 1.1 ([29, Theorem 12]) *For $z, x_1, \dots, x_{d-1}$ indeterminates over $\mathbb{C}$ there exists a unique linear map*

$$\mathrm{tr}_d : \bigcup_{n \geq 0} \mathrm{Y}_{d,n}(q) \longrightarrow \mathbb{C}[z, x_1, \dots, x_{d-1}]$$

satisfying the rules:

$$\begin{array}{lllll}
(1) & \mathrm{tr}_d(\alpha\beta) & = \mathrm{tr}_d(\beta\alpha) & \alpha, \beta \in \mathrm{Y}_{d,n}(q) & \\
(2) & \mathrm{tr}_d(1) & = 1 & 1 \in \mathrm{Y}_{d,n}(q) & \\
(3) & \mathrm{tr}_d(\alpha g_n) & = z\,\mathrm{tr}_d(\alpha) & \alpha \in \mathrm{Y}_{d,n}(q) & \text{(Markov property)} \\
(4) & \mathrm{tr}_d(\alpha t_{n+1}^s) & = x_s\,\mathrm{tr}_d(\alpha) & \alpha \in \mathrm{Y}_{d,n}(q) & (1 \leq s \leq d-1).
\end{array}$$

Note that for $d = 1$ the trace restricts to the first three rules and it coincides with the Ocneanu trace τ on the Iwahori–Hecke algebras.

Moreover, the trace tr_d satisfies the following equality:

$$\mathrm{tr}_{d,D}(ae_ng_n) = z\,\mathrm{tr}_{d,D}(a) \qquad a \in \mathrm{Y}_{d,n}(q). \tag{1.5}$$

Note 1.2 In this paper we will sometimes write $\mathrm{tr}_d(\alpha)$ for a framed braid $\alpha \in \mathscr{F}_n$, by using the natural epimorphism of $\mathscr{F}_n$ onto $\mathrm{Y}_{d,n}$. Similarly, the same holds for a classical braid $\alpha \in B_n$ by using the natural homomorphism of B_n into $\mathrm{Y}_{d,n}$.

1.2.2 The E*–System*

Using the natural epimorphism of the framed braid group $\mathscr{F}_n$ onto $\mathrm{Y}_{d,n}(q)$, the trace tr_d and the Markov framed braid equivalence, comprising conjugation in the groups $\mathscr{F}_n$ and positive and negative stabilization and destabilization (see for example [37]), in [31] the authors tried to obtain a topological invariant for framed links after the method of V.F.R. Jones [27] (using for the algebra the presentation discussed in Sect. 1.6). This meant that tr_d would have to be normalized, so that the closed braids $\widehat{\alpha}$ and $\widehat{\alpha\sigma_n}$ ($\alpha \in \mathscr{F}_n$) be assigned the same value of the invariant, and re-scaled, so that the closed braids $\widehat{\alpha\sigma_n^{-1}}$ and $\widehat{\alpha\sigma_n}$ ($\alpha \in \mathscr{F}_n$) be also assigned the same value of the invariant. However, as it turned out, $\mathrm{tr}_d(\alpha g_n^{-1})$ does not factor through $\mathrm{tr}_d(\alpha)$, that is:

$$\mathrm{tr}_d(\alpha g_n^{-1}) \overset{(4)}{=} \mathrm{tr}_d(\alpha g_n) - (q - q^{-1})\,\mathrm{tr}_d(\alpha e_n) \neq \mathrm{tr}_d(g_n^{-1})\mathrm{tr}_d(\alpha). \tag{1.6}$$

The reason is that, although $\mathrm{tr}_d(\alpha g_n) = z\,\mathrm{tr}_d(\alpha)$, $\mathrm{tr}_d(\alpha e_n)$ does not factor through $\mathrm{tr}_d(\alpha)$, that is:

$$\mathrm{tr}_d(\alpha e_n) \neq \mathrm{tr}_d(e_n)\mathrm{tr}_d(\alpha). \tag{1.7}$$

This is due to the fact that:

$$\mathrm{tr}_d(\alpha t_n^k) \neq \mathrm{tr}_d(t_n^k)\mathrm{tr}_d(\alpha) \quad k = 1, \ldots, d-1. \tag{1.8}$$

Forcing

$$\mathrm{tr}_d(\alpha e_n) = \mathrm{tr}_d(e_n)\mathrm{tr}_d(\alpha) \tag{1.9}$$

yields that the trace parameters $x_1, \ldots, x_{d-1}$ have to satisfy the E–*system*, the non-linear system of equations in $\mathbb{C}$:

$$E^{(m)} = x_m E \qquad (1 \leq m \leq d-1)$$

where

$$E := E^{(0)} = \frac{1}{d}\sum_{s=0}^{d-1} x_s x_{d-s} = \mathrm{tr}_d(e_i) \qquad \text{and} \qquad E^{(m)} := \frac{1}{d}\sum_{s=0}^{d-1} x_{m+s} x_{d-s}\,,$$

where the sub-indices on the x_j's are regarded modulo d and $x_0 := 1$ (see [31]). As it was shown by P. Gérardin (in the Appendix of [31]), the solutions of the E–system are parametrized by the non-empty subsets of $\mathbb{Z}/d\mathbb{Z}$. For example, for every singleton subset $\{m\}$ of $\mathbb{Z}/d\mathbb{Z}$, we have a solution of the E–system given by:

$$\mathrm{x}_1 = \exp(2\pi m\sqrt{-1}/d) \qquad \text{and} \qquad \mathrm{x}_k = \mathrm{x}_1^k \quad \text{for } k = 2, \ldots, d-1. \tag{1.10}$$

1.2.3 The Specialized Trace

Let $X_D := (\mathrm{x}_1, \ldots, \mathrm{x}_{d-1})$ be a solution of the E–system parametrized by the non-empty subset D of $\mathbb{Z}/d\mathbb{Z}$. We shall call *specialized trace* the trace tr_d with the parameters $x_1, \ldots, x_{d-1}$ specialized to $\mathrm{x}_1, \ldots, \mathrm{x}_{d-1} \in \mathbb{C}$, and it shall be denoted $\mathrm{tr}_{d,D}$ (cf. [12]). More precisely,

$$\mathrm{tr}_{d,D} : \bigcup_n \mathrm{Y}_{d,n}(q) \longrightarrow \mathbb{C}[z]$$

is a Markov trace on the Yokonuma–Hecke algebra, satisfying the following rules:

$$\begin{array}{lrcll}
(1) & \mathrm{tr}_{d,D}(\alpha\beta) & = & \mathrm{tr}_{d,D}(\beta\alpha) & \alpha, \beta \in \mathrm{Y}_{d,n}(q) \\
(2) & \mathrm{tr}_{d,D}(1) & = & 1 & 1 \in \mathrm{Y}_{d,n}(q) \\
(3) & \mathrm{tr}_{d,D}(\alpha g_n) & = & z\,\mathrm{tr}_{d,D}(\alpha) & \alpha \in \mathrm{Y}_{d,n}(q) \quad \text{(Markov property)} \\
(4') & \mathrm{tr}_{d,D}(\alpha t_{n+1}^s) & = & \mathrm{x}_s\,\mathrm{tr}_{d,D}(\alpha) & \alpha \in \mathrm{Y}_{d,n}(q) \quad (1 \leq s \leq d-1).
\end{array}$$

The rules (1)–(3) are the same as in Theorem 1.1, while rule (4) is replaced by the rule (4′). As it turns out [32]:

$$E_D := \mathrm{tr}_{d,D}(e_i) = \frac{1}{|D|}, \tag{1.11}$$

where $|D|$ is the cardinality of the subset D. Note that $\mathrm{tr}_{1,\{0\}}$ coincides with tr_1 that in turn coincides with the Ocneanu trace τ.

1.2.4 Properties of the Markov Traces

We shall now give some properties of the traces tr_d and $\mathrm{tr}_{d,D}$ [10], analogous to known properties of the Ocneanu trace τ, by considering their behaviour under the operations below. Clearly, a property satisfied by tr_d is also satisfied by $\mathrm{tr}_{d,D}$ (and by τ), but the converse may not hold.

• *Inversion of braid words.* Inversion means that a braid word is written from right to left. For $\alpha = t_1^{k_1} \dots t_n^{k_n} \sigma_{i_1}^{l_1} \dots \sigma_{i_r}^{l_r} \in \mathscr{F}_n$, where $k_1, \dots, k_n, l_1, \dots, l_r \in \mathbb{Z}$, we denote by $\overleftarrow{\alpha}$ the inverted word, that is, $\overleftarrow{\alpha} = \sigma_{i_r}^{l_r} \dots \sigma_{i_1}^{l_1} t_n^{k_n} \dots t_1^{k_1}$. On the level of closed braids this operation corresponds to the change of orientation on all components of the resulting link. The operation can be extended linearly to elements of $\mathrm{Y}_{d,n}$. The trace tr_d (and consequently also the trace $\mathrm{tr}_{d,D}$) satisfies the following property:

$$\mathrm{tr}_d(\alpha) = \mathrm{tr}_d(\overleftarrow{\alpha}).$$

• *Split links.* Let $L = L_1 \sqcup \dots \sqcup L_m$ be a split framed link, where $L_1, \dots, L_m$ are framed links. Then there exists a braid word $\alpha = \alpha_1 \dots \alpha_m \in \mathscr{F}_n$, where $\alpha_i \in \mathscr{F}_{i_j} \setminus \mathscr{F}_{i_{j-1}+1}$ for some $1 \leq i_1 < \dots < i_k \leq n$ with $i_{j+1} - i_j > 1$ such that $\widehat{\alpha} = L$ and $\widehat{\alpha_i} = \sqcup_{k=1}^{i_{j-1}} U \sqcup L_i$ for each $i = 1, \dots, m$. The trace tr_d (and consequently also the trace $\mathrm{tr}_{d,D}$) satisfies the following property:

$$\mathrm{tr}_d(\alpha) = \mathrm{tr}_d(\alpha_1) \cdots \mathrm{tr}_d(\alpha_m) \tag{1.12}$$

• *Connected sums.* Let $\alpha \in \mathscr{F}_n$ and $\beta \in \mathscr{F}_m$ for some $n, m \in \mathbb{N}$. The *connected sum* of α and β is the word $\alpha \# \beta := \alpha^{[0]} \beta^{[n-1]}$ in the framed braid group $\mathscr{F}_{n+m-1}$, where $\alpha^{[0]}$ is the natural embedding of α in $\mathscr{F}_{n+m-1}$, while $\beta^{[n-1]}$ is the embedding of β in $\mathscr{F}_{n+m-1}$ induced by the following shifting of the indices: $\sigma_i \mapsto \sigma_{n+i-1}$ for $i \in \{1, \dots, m-1\}$ and $t_j \mapsto t_{n+j-1}$ for $j \in \{1, \dots, m\}$. Upon closing the braids, this operation corresponds to taking the connected sum of the resulting framed links. It is known that the Ocneanu trace is multiplicative under the connected sum operation, that is, $\tau(\alpha \# \beta) = \tau(\alpha)\, \tau(\beta)$ if $\alpha \in B_n$ and $\beta \in B_m$. On the other hand, the trace tr_d is *not multiplicative* under the connected sum operation, due to (1.8) and (1.7) (we have $\alpha \# t_1^k = \alpha t_n^k$ and, by linear extension, $\alpha \# e_1 = \alpha e_n$). Yet, the specialized trace $\mathrm{tr}_{d,D}$ is multiplicative on connected sums, due to the E–condition (1.9), but this is only true on the level of classical braids [10]. Namely:

$$\mathrm{tr}_{d,D}(\alpha \# \beta) = \mathrm{tr}_{d,D}(\alpha)\, \mathrm{tr}_{d,D}(\beta) \text{ for } \alpha \in B_n \text{ and } \beta \in B_m.$$

For framed braids this is true only when $E_D = 1$, that is, when the set D is singleton and hence the corresponding solution X_D of the E–system is described by (1.10). Namely, for $\alpha \in \mathscr{F}_n$ and $\beta \in \mathscr{F}_m$:

$$\mathrm{tr}_{d,D}(\alpha \# \beta) = \mathrm{tr}_{d,D}(\alpha) \mathrm{tr}_{d,D}(\beta) \Leftrightarrow \mathrm{x}_1^d = 1 \text{ and } \mathrm{x}_k = \mathrm{x}_1^k \text{ for } k = 1, \dots, d-1.$$

• *Mirror images.* Let us consider the group automorphism of B_n given by $\sigma_i \mapsto \sigma_i^{-1}$. For $\alpha \in B_n$, we denote by α^* the image of α via this automorphism. We call α^* the *mirror image* of α. On the level of closed braids this operation corresponds to switching all crossings. Note that the operation mirror image applies on classical braids or links. It is known that the Ocneanu trace satisfies a "mirroring property". Namely, $\tau(q,z)(\alpha^*) = \tau\left(q^{-1}, z-(q-q^{-1})\right)(\alpha)$. However, due to (1.6), the trace tr_d does not satisfy the mirroring property, but the specialized trace $\mathrm{tr}_{d,D}$ does. Namely, observe that $\mathrm{tr}_{d,D}(g_i^{-1}) = z-(q-q^{-1})E_D$ and set a new variable λ_D in place of z. By re-scaling σ_i to $\sqrt{\lambda_D}g_i$, so that $\mathrm{tr}_{d,D}(g_n^{-1}) = \lambda_D z$, we find

$$\lambda_D := \frac{z-(q-q^{-1})E_D}{z}. \tag{1.13}$$

If we solve (1.13) with respect to the variable z, we obtain

$$z = \frac{(q-q^{-1})\,E_D}{1-\lambda_D}.$$

Hence, the trace $\mathrm{tr}_{d,D}$ can be considered as a polynomial in the variables (q,z) or in the variables (q,λ_D) by the above change of variables. Using this notation, the trace $\mathrm{tr}_{d,D}$ satisfies the following property:

$$\mathrm{tr}_{d,D}(q,z)(\alpha^*) = \mathrm{tr}_{d,D}\left(q^{-1}, z-(q-q^{-1})E_D\right)(\alpha), \quad \text{for any } \alpha \in B_n,$$

or equivalently,

$$\mathrm{tr}_{d,D}(q,\lambda_D)(\alpha^*) = \mathrm{tr}_{d,D}\left(q^{-1}, \lambda_D^{-1}\right)(\alpha), \quad \text{for any } \alpha \in B_n.$$

1.2.5 *The Specialized Trace* $\mathrm{tr}_{d,D}$ *on Classical Braids*

Let $\alpha \in B_n$ be a classical braid. When calculating $\mathrm{tr}_{d,D}(\alpha)$, the framing generators t_j appear only in the form of the idempotents e_i due to the application of the quadratic relation (1.3). In this case, the fourth rule of the trace $\mathrm{tr}_{d,D}$ is not applied directly, but rather indirectly using the E–condition (1.9). It has been long conjectured by J. Juyumaya that the fourth rule of the trace $\mathrm{tr}_{d,D}$, when computed on classical braids, can be substituted by rules involving only the idempotents e_i (cf. [2, 28]). Indeed, this fact has been proved in [11].

Before we proceed to the statement (Theorem 1.2), we define the subalgebra $\mathrm{Y}_{d,n}(q)^{(\mathrm{br})}$ of $\mathrm{Y}_{d,n}(q)$ generated only by the braiding generators $g_1,\ldots,g_{n-1}$. The subalgebra $\mathrm{Y}_{d,n}(q)^{(\mathrm{br})}$ is also the image of the natural homomorphism $\delta : \mathbb{C}B_n \to \mathrm{Y}_{d,n}(q)$ defined by $\sigma_i \mapsto g_i$, since $g_i^{-1} \in \mathrm{Y}_{d,n}(q)^{(\mathrm{br})}$ for $i=1,\ldots,n-1$ [11]. Remarkably, for $q \neq 1$ it is also true that $e_i \in \mathrm{Y}_{d,n}(q)^{(\mathrm{br})}$ for $i=1,\ldots,n-1$ [11].

When computing the trace $\mathrm{tr}_{d,D}$ for classical braids, we restrict ourselves on the subalgebra $\mathrm{Y}_{d,n}(q)^{(\mathrm{br})}$. Then the fourth rule of the trace can be substituted by two new rules as follows:

Theorem 1.2 ([11, Theorem 4.3]) *The following rules are sufficient for computing the trace* $\mathrm{tr}_{d,D}$ *on* $\mathrm{Y}_{d,n}(q)^{(\mathrm{br})}$*:*

$$\begin{array}{lrcll}
\text{(i)} & \mathrm{tr}_{d,D}(ab) & = & \mathrm{tr}_{d,D}(ba) & a, b \in \mathrm{Y}_{d,n}(q)^{(\mathrm{br})} \\
\text{(ii)} & \mathrm{tr}_{d,D}(1) & = & 1 & 1 \in \mathrm{Y}_{d,n}(q)^{(\mathrm{br})} \\
\text{(iii)} & \mathrm{tr}_{d,D}(ag_n) & = & z\,\mathrm{tr}_{d,D}(a) & a \in \mathrm{Y}_{d,n}(q)^{(\mathrm{br})} \quad (\textit{Markov property}) \\
\text{(iv)} & \mathrm{tr}_{d,D}(ae_n) & = & E_D\,\mathrm{tr}_{d,D}(a) & a \in \mathrm{Y}_{d,n}(q)^{(\mathrm{br})} \\
\text{(v)} & \mathrm{tr}_{d,D}(ae_ng_n) & = & z\,\mathrm{tr}_{d,D}(a) & a \in \mathrm{Y}_{d,n}(q)^{(\mathrm{br})}.
\end{array}$$

As we have seen, rule (v) holds for the trace tr_d (1.5) and rule (iv) is the E–condition (1.9). There is no analogue of this theorem for the trace tr_d since it does not satisfy the E–condition (rule iv). Notice that the value of the trace $\mathrm{tr}_{d,D}$ does not depend on the specific solution of the E–system, but only on the cardinality of the subset $D \subseteq \mathbb{Z}/d\mathbb{Z}$ due to (1.11). This fact will have important consequences on the classical link invariants defined via the trace $\mathrm{tr}_{d,D}$.

1.3 Link Invariants from the Yokonuma–Hecke Algebras

Given a solution $X_D := (\mathrm{x}_1, \ldots, \mathrm{x}_{d-1})$ of the E–system invariants for various types of knots and links, such as framed, classical and singular, have been constructed from $\mathrm{tr}_{d,D}$ in [31–33]. The definitions of these invariants have been adapted in [10, 11] in view of the new presentation of $\mathrm{Y}_{d,n}(q)$. Moreover, we recall the construction of the transverse link invariants defined in [10].

1.3.1 Framed Links

Let $\mathscr{L}_f$ denote the set of oriented framed links. We set:

$$\Lambda_D := \frac{1}{z\sqrt{\lambda_D}}. \tag{1.14}$$

From the above and re-scaling σ_i to $\sqrt{\lambda_D}g_i$, so that $\mathrm{tr}_{d,D}(g_n^{-1}) = \lambda_D z$, we have the following Theorem, which is analogous to [31, Theorem 8]:

Theorem 1.3 ([10, Theorem 4.1]) *Given a solution* X_D *of the E–system, for any framed braid* $\alpha \in \mathscr{F}_n$ *we define for the framed link* $\widehat{\alpha} \in \mathscr{L}_f$*:*

$$\Phi_{d,D}(\widehat{\alpha}) = \Lambda_D^{n-1}(\sqrt{\lambda_D})^{\epsilon(\alpha)}\left(\mathrm{tr}_{d,D} \circ \gamma\right)(\alpha)$$

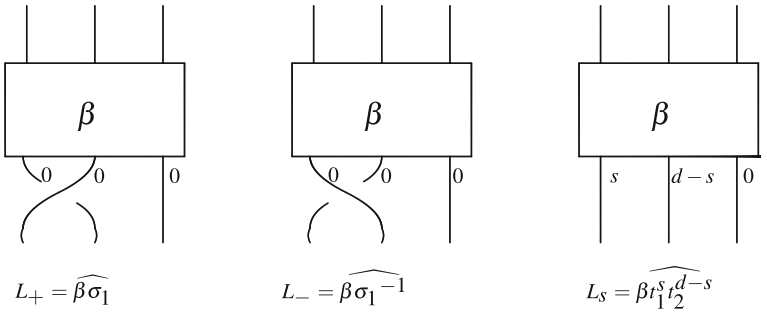


Fig. 1.1 The framed links of the skein relation in open braid form

where $\gamma : \mathbb{C}\mathscr{F}_n \longrightarrow \mathrm{Y}_{d,n}(q)$ *is the natural algebra homomorphism defined via:* $\sigma_i \mapsto g_i$ *and* $t_j^s \mapsto t_j^{s(\mathrm{mod}\, d)}$*, and* $\epsilon(\alpha)$ *is the algebraic sum of the exponents of the* σ_i*'s in* α*. Then the map* $\Phi_{d,D}(q, z)$ *is a 2-variable isotopy invariant of oriented framed links.*

Proposition 1.1 ([10, Proposition 4.2]) *The invariant* $\Phi_{d,D}$ *satisfies the following skein relation:*

$$\frac{1}{\sqrt{\lambda_D}}\Phi_{d,D}(L_+) - \sqrt{\lambda_D}\Phi_{d,D}(L_-) = \frac{q - q^{-1}}{d} \sum_{s=0}^{d-1} \Phi_{d,D}(L_s), \tag{1.15}$$

where the links L_+*,* L_- *and* L_s *are closures of the framed braids illustrated in Fig. 1.1.*

Remark 1.3 Note that, for every $d \in \mathbb{N}$, we have $2^d - 1$ distinct solutions of the E–system, so the above construction yields $2^d - 1$ isotopy invariants for framed links.

Remark 1.4 Due to the complicated computations for the trace $\mathrm{tr}_{d,D}$ (and hence for the invariants $\Phi_{d,D}$, two computer programs have been developed for this purpose. One has been developed by M. Chmutov in Maple [10] and the other by the first author in C#. Let $w \in \mathrm{Y}_{d,n}(q)$ be a word. Both of the programs apply iteratively the quadratic relation, breaking the word w into many simpler words. Then using the relations (b_1) and (f_2) of the braid group (which also hold in the algebra $\mathrm{Y}_{d,n}(q)$) the programs reduce w into words written in split form. Finally, the four rules of the trace $\mathrm{tr}_{d,D}$ are applied and the computation ends. Note that both programs have exponential complexity with respect to $r(w)$, where $r(w)$ is the number of indices of the braiding generators in w with powers different than 0 or 1.

1.3.2 Classical Links

Let $\mathscr{L}$ denote the set of oriented classical links. The classical braid group B_n injects into the framed braid group $\mathscr{F}_n$, whereby elements in B_n are viewed as framed braids

with all framings zero. So, by the classical Markov braid equivalence, comprising conjugation in the groups B_n and positive and negative stabilizations and destabilizations, and by the construction and notations above, we obtain isotopy invariants for oriented classical knots and links, where the t_j's are treated as formal generators. These invariants of classical links, which are analogous to those defined in [32] where the old presentation for the Yokonuma–Hecke algebra is used, are denoted as $\Theta_{d,D}$ and the restriction of $\gamma : \mathbb{C}\mathscr{F}_n \longrightarrow \mathrm{Y}_{d,n}(q)$ on $\mathbb{C}B_n$ is denoted as δ. Namely,

$$\Theta_{d,D}(\widehat{\alpha}) := \Lambda_D^{n-1}(\sqrt{\lambda_D})^{\epsilon(\alpha)}\,(\mathrm{tr}_{d,D} \circ \delta)(\alpha).$$

An important corollary of Theorem 1.2 is that the invariants $\Theta_{d,D}$ do not depend on d and D but only on the cardinality of the set D. Namely:

Proposition 1.2 ([11, Proposition 4.6]) *The values of the isotopy invariants $\Theta_{d,D}$ for classical links depend only on the cardinality $|D|$ of D. Hence, for a fixed d, we only obtain d invariants. Further, for d, d' positive integers with $d \leq d'$, we have $\Theta_{d,D} = \Theta_{d',D'}$ as long as $|D| = |D'|$. We deduce that, if $|D'| = d$, then $\Theta_{d',D'} = \Theta_{d,\mathbb{Z}/d\mathbb{Z}}$. Therefore, the invariants $\Theta_{d,D}$ can be parametrized by the natural numbers, setting $\Theta_d := \Theta_{d,\mathbb{Z}/d\mathbb{Z}}$ for all $d \in \mathbb{Z}_{>0}$.*

The invariants $\Theta_d(q, z)$ need to be compared with known invariants of classical links, especially with the HOMFLYPT polynomial. The HOMFLYPT polynomial $P(q, z)$ is a 2-variable isotopy invariant of oriented classical links that can be constructed from the Iwahori–Hecke algebras $\mathrm{H}_n(q)$ and the Ocneanu trace τ after re–scaling and normalizing τ [27]. In this paper we define P via the invariants Θ_d, since for $d = 1$ the algebras $\mathrm{H}_n(q)$ and $\mathrm{Y}_{1,n}(q)$ coincide and the traces τ, tr_1 and $\mathrm{tr}_{1,\{0\}}$ also coincide. Namely, we define:

$$P(\widehat{\alpha}) = \Theta_1(\widehat{\alpha}) = \left(\frac{1}{z\sqrt{\lambda_{\mathrm{H}}}}\right)^{n-1} (\sqrt{\lambda_{\mathrm{H}}})^{\epsilon(\alpha)} \left(\mathrm{tr}_{1,\{0\}} \circ \delta\right)(\alpha)$$

where $\lambda_{\mathrm{H}} := \frac{z-(q-q^{-1})}{z} = \lambda_{\{0\}}$. Further, recall that the HOMFLYPT polynomial satisfies the following skein relation [27]:

$$\frac{1}{\sqrt{\lambda_{\mathrm{H}}}}\, P(L_+) - \sqrt{\lambda_{\mathrm{H}}}\, P(L_-) = (q - q^{-1})\, P(L_0) \tag{1.16}$$

where L_+, L_-, L_0 is a Conway triple.

Contrary to the case of framed links, the skein relation of the invariants $\Phi_{d,D}(q, z)$ has no topological interpretation in the case of classical links since it introduces framings. This makes it very difficult to compare the invariants $\Theta_d(q, z)$ with the HOMFLYPT polynomial using diagrammatic methods. Further, on the algebraic level, there is no algebra homomorphism connecting the algebras and the traces [12]. Consequently, in [12] it is shown that for generic values of the parameters q, z the invariants $\Theta_d(q, z)$ do not coincide with the HOMFLYPT polynomial. In

fact, they only coincide in the trivial cases where $q = \pm 1$ or $\mathrm{tr}_d(e_i) = 1$. The last case implies that the solution of the E–system comprises the d-th roots of unity. Yet, the invariants $\Theta_d(q, z)$ could be topologically equivalent to the HOMFLYPT polynomial, in the sense that they distinguish or not distinguish the same pairs of knots and links. The topological comparison of Θ_d with P has been a long-standing open problem that was eventually answered in [11] and these results are presented in Sect. 1.4 below.

Remark 1.5 A computer program has been developed by the first author in Mathematica for computing the classical link invariants Θ_d [35]. This program uses Theorem 1.2 for computing the trace $\mathrm{tr}_{d,D}$ on classical braids, that is, the elements e_i are considered as formal elements instead of a sum of products of the framing generators. Further, it uses the new presentation $\mathrm{Y}_{d,n}(q)$ of the Yokonuma-Hecke algebra, whose quadratic relation is more economical for computations. Both facts result in lower computational complexity than the programs of Remark 1.4, however the computational complexity remains exponential (see [35] for more details).

1.3.3 Singular Links

Let $\mathscr{L}_{\mathscr{S}}$ denote the set of oriented singular links. Oriented singular links are represented by singular braids that form the singular braid monoids $\mathscr{S}B_n$ [5, 7, 44]. The singular braid monoid $\mathscr{S}B_n$ is generated by the classical braiding generators σ_i with their inverses, together with the elementary singular braids τ_i that are not invertible. In [33] a monoid homomorphism was constructed that we adapt here to the new presentation $\mathrm{Y}_{d,n}(q)$ of the Yokonuma–Hecke algebra, namely:

$$\begin{array}{rccc} \eta : & \mathscr{S}B_n & \longrightarrow & \mathrm{Y}_{d,n}(q) \\ & \sigma_i & \mapsto & g_i \\ & \tau_i & \mapsto & e_i \end{array} \tag{1.17}$$

Using the singular braid equivalence [21] (see also [38]), the map η and the specialized trace $\mathrm{tr}_{d,D}$ we obtain isotopy invariants for oriented singular links, analogous to the ones constructed in [33, Theorem 3.6], as follows:

Theorem 1.4 ([10, Theorem 4.8]) *For any singular braid $\alpha \in \mathscr{S}B_n$, we define*

$$\Psi_{d,D}(\widehat{\alpha}) := \Lambda_D^{n-1}(\sqrt{\lambda_D})^{\epsilon(\alpha)} \left(\mathrm{tr}_{d,D} \circ \eta\right)(\alpha) ,$$

where Λ_D, λ_D are as defined in (1.13) and (1.14), η is as defined in (1.17) and $\epsilon(\alpha)$ is the sum of the exponents of the generators σ_i and τ_i in the word α. Then the map $\Psi_{d,D}(q, z)$ is a 2-variable isotopy invariant of oriented singular links.

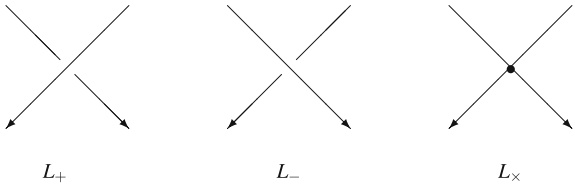


Fig. 1.2 The singular links L_+, L_- and $L_\times$

Moreover, in the image $\eta(\mathscr{S}B_n)$, we have

$$g_i - g_i^{-1} = (q - q^{-1})e_i \quad \text{for all } i = 1, \ldots, n-1,$$

which gives rise to the following skein relation (compare with [33]):

$$\frac{1}{\sqrt{\lambda_D}}\,\Psi_{d,D}(L_+) - \sqrt{\lambda_D}\,\Psi_{d,D}(L_-) = \frac{q - q^{-1}}{\sqrt{\lambda_D}}\,\Psi_{d,D}(L_\times)$$

where L_+, L_- and $L_\times$ are diagrams of three oriented singular links that are identical except for one crossing, where they are as in Fig. 1.2. Furthermore, the properties of the traces $\mathrm{tr}_{d,D}$ under inversion, split links, connected sums and mirror imaging carry through to the invariants $\Psi_{d,D}$.

1.3.4 Transverse Links

Another class of links that is naturally related to the Yokonuma–Hecke algebras is the class of *transverse links*, denoted by $\mathscr{L}_{\mathscr{T}}$. A transverse knot is represented by a smooth closed spacial curve that is nowhere tangent to planes of a special field of planes in $\mathbb{R}^3$ called *standard contact structure* (for the precise definition see for example [20]). Transverse links are naturally framed and oriented. Two links that are classically isotopic may be transversely non–equivalent. So, a topological type of framed links may consist of several different types in $\mathscr{L}_{\mathscr{T}}$. The problem is to find transverse invariants for such links.

In 1983, D. Bennequin [6] noted that the closed braid presentation of knots is convenient for describing transverse knots with the blackboard framing. For a knot K represented as a closed braid $\widehat{\alpha}$ with n strands, one can check that the self-linking number is equal to $sl(K) = \epsilon(\alpha) - n$, where $\epsilon(\alpha)$ is the sum of the exponents of the braiding generators σ_i in the word $\alpha \in B_n$ [6]. So, the transverse knot K defines naturally an element of the framed braid group $\alpha' := t_1^{sl(K)}\alpha \in \mathscr{F}_n$. This generalizes to transverse links in the obvious manner (using the self-linking of each component).

Further, S. Orevkov and V. Shevchishin [41] and independently N. Wrinkle [46] gave a transverse analogue of the Markov Theorem, comprising conjugation in the braid groups and *only positive* stabilizations and destabilizations: $\alpha \sim \alpha\,\sigma_n$, where $\alpha \in B_n$. Now, rule (3) of the definition of the trace tr_d (Theorem 1.1) tells us that tr_d

respects positive stabilizations. Moreover, the absence of the negative stabilization in the transverse braid equivalence resembles the problem of re-scaling of the trace tr_d with respect to the negative stabilization, recall (1.6), making the Yokonuma–Hecke algebras a natural algebraic object related to the class $\mathscr{L}_{\mathscr{T}}$ of transverse links. Let L be a transverse link represented by the closure $\widehat{\alpha}$ of a braid $\alpha \in B_n$, giving rise to the framed braid $\alpha' = t_1^{r_1} \dots t_n^{r_n} \alpha \in \mathscr{F}_n$, where $r_1 + \dots + r_n = sl(L)$. We define

$$M_d(\widehat{\alpha}) := \frac{1}{z^{n-1}} \mathrm{tr}_d(\alpha').$$

Theorem 1.5 ([10, Theorem 4.11]) *The map $M_d(q, z, x_1, \dots, x_{d-1})$ is a $(d+1)$-variable isotopy invariant of oriented transverse links.*

The properties of the trace tr_d under inversion and split links carry through to the invariants M_d.

Remark 1.6 Due to the transverse braid equivalence, M_d need not take the same value on the links $\widehat{\alpha\sigma_n}$ and $\widehat{\alpha\sigma_n^{-1}}$. Hence, the re-scaling map $\sigma_i \mapsto \sqrt{\lambda_D} g_i$ is not needed any more and by consequence a quantity analogous to λ_D is not introduced. However, if we make such a re-scaling and specialize $(x_1, \dots, x_{d-1})$ to the solution $X_D = (\mathrm{x}_1, \dots, \mathrm{x}_{d-1})$ of the E–system, then the corresponding invariant of transverse links would coincide with the invariants $\Phi_{d,D}(q, z)$ of oriented framed links from Theorem 1.3.

Our original hope was that the invariants M_d would distinguish the transverse knots of the same topological type and with the same Bennequin numbers [6]. However, this turned out not to be the case, due to the following reason: any quantum knot invariant can be expressed in terms of Vassiliev invariants in a standard way (see, for example, [7] or [8]); we show that the invariants M_d can be similarly expressed in terms of Vassiliev invariants.

Proposition 1.3 ([10, Proposition 6.1]) *Let us make a substitution $q = e^h$ into the transverse knot invariant $M_d(q, z, x_1, \dots, x_{d-1})$ and consider the Taylor expansion in the power series in h. For every $n \in \mathbb{N}$, the coefficient of h^n is a Vassiliev invariant of order $\leq n$.*

The above proposition implies that the invariant $M_d(e^h, z, x_1, \dots, x_{d-1})$ of transverse knots is covered by an (infinite) sequence of Vassiliev invariants. However, the Fuchs–Tabachnikov theorem [20, Theorem 5.6] claims that any transverse Vassiliev invariant turns out to be a topological Vassiliev invariant of framed knots. For further discussion, we refer the reader to [10].

1.4 The Classical Link Invariants

1.4.1 Behaviour of the Invariants Θ_d on Knots

Let L_1 and L_2 be two links. We will say that L_1 and L_2 are *Θ_d-equivalent* (respectively *P-equivalent*) if $\Theta_d(L_1) = \Theta_d(L_2)$ (respectively $P(L_1) = P(L_2)$). A long-standing question had been how the classical link invariants Θ_d compare to the HOMFLYPT polynomial and among themselves for various values of $d \geq 2$. The aim has been to find a pair of P-equivalent knots or links that are not Θ_d-equivalent for some values of $d \geq 2$.

The first computations (using the presentation $\mathrm{Y}_{d,n}(u)$) on several pairs of P-equivalent pairs of knots and links were disappointing. However, it became evident from the computations that the invariants Θ_d are related on *knots* to the HOMFLYPT polynomial via a change of variables. This conjecture (formulated in [10]) has been proved in [11] by comparing the traces $\mathrm{tr}_{d,D}$ and τ on braids whose closures are knots. In detail:

Proposition 1.4 ([11, Proposition 5.6]) *Let $\alpha \in B_n$ be a knot. Then*

$$\mathrm{tr}_{d,D}(q, z)(\alpha) = E_D^{n-1}\, \tau(q, z/E_D)(\alpha).$$

Now, using Proposition 1.4 the conjecture relating the invariants Θ_d and P could be proved:

Theorem 1.6 ([11, Theorem 5.8]) *Given a solution X_D of the E–system, for any braid $\alpha \in B_n$ such that $\widehat{\alpha}$ is a knot, we have:*

$$\Theta_d(q, z)(\widehat{\alpha}) = \Theta_1(q, z/E_D)(\widehat{\alpha}) = P(q, z/E_D)(\widehat{\alpha}),$$

or equivalently:

$$\Theta_d(q, \lambda_D)(\widehat{\alpha}) = \Theta_1(q, \lambda_D)(\widehat{\alpha}) = P(q, \lambda_D)(\widehat{\alpha}).$$

Note that the polynomials $\Theta_d(q, \lambda_D)$ and $P(q, \lambda_{\mathrm{H}})$ coincide on knots by considering substituting the variable λ_{H} with λ_D in P. Hence, the value E_D does not appear when computing the invariants Θ_d on knots.

1.4.2 Behaviour of the Invariants Θ_d on Split Links and Disjoint Union of Knots

Let L and L' be two links. The invariants Θ_d satisfy the following property for split links:

$$\Theta_d(L \sqcup L') = \frac{1-\lambda_D}{(q-q^{-1})\sqrt{\lambda_D}E_D}\,\Theta_d(L)\Theta_d(L').$$

This equality can easily be proven using the property of the trace on split links (1.12). Now, Theorem 1.6 can be generalized to disjoint unions of knots using the multiplicative property of the invariants Θ_d on split links and Proposition 1.4. In detail:

Theorem 1.7 ([11, Theorem 6.2]) *Given a solution X_D of the E–system, for any braid $\alpha \in B_n$ such that $\widehat{\alpha}$ is a disjoint union of k knots, we have*

$$\Theta_d(q,z)(\widehat{\alpha}) = E_D^{1-k}\Theta_1(q, z/E_D)(\widehat{\alpha}) = E_D^{1-k}P(q, z/E_D)(\widehat{\alpha}),$$

or equivalently:

$$\Theta_d(q,\lambda_D)(\widehat{\alpha}) = E_D^{1-k}\Theta_1(q,\lambda_D)(\widehat{\alpha}) = E_D^{1-k}P(q,\lambda_D)(\widehat{\alpha}).$$

Now the remaining question was the case of links that are not disjoint unions of knots. In this case there was not an apparent conjecture relating the invariants Θ_d and P or relating the traces $\mathrm{tr}_{d,D}$ and τ. On the contrary, computations of the traces $\mathrm{tr}_{d,D}$ and τ on simple examples of links indicated that the invariants Θ_d and P may not be related by a change of variables [11, Sect. 6.2]. Specifically, for a 2-component link, the invariants Θ_d seemed to depend not only on the value of P of the same link, but also on the value of P of the link with the two components unlinked [11, Sect. 6.2]. Also, the behaviour of the elements e_i when computing the trace $\mathrm{tr}_{d,D}$ on simple examples was complicating the comparison of Θ_d with P on links [11, Sect. 6.3].

1.4.3 Behaviour of the Invariants Θ_d on Links – A Special Skein Relation

In order to investigate the question of Θ_d equivalence on P-equivalent pairs of links, a diagrammatic computation was not possible, since the skein relation of the framed invariants $\Phi_{d,D}$ (1.15) involves framed links and hence, there is no topological interpretation when computing the invariants Θ_d. However, if the skein relation (1.15) is applied to a *crossing involving different components*, then all the framed links of the skein relation reduce to classical links. Also, the skein relation obtained is identical to the one of the HOMFLYPT polynomial. In detail:

Proposition 1.5 ([11, Proposition 6.8]) *Let $\beta \in \mathscr{F}_n$ and $i \in \{1,\ldots,n-1\}$. Let*

$$L_+ = \widehat{\beta\sigma_i}, \quad L_- = \widehat{\beta\sigma_i^{-1}} \quad \text{and} \quad L_0 = \widehat{\beta}.$$

Suppose we apply the skein relation (1.15) of $\Phi_{d,D}$ on L_+ on the crossing σ_i and that the i-th and $(i+1)$-st strands (at the region of the crossing) belong to different

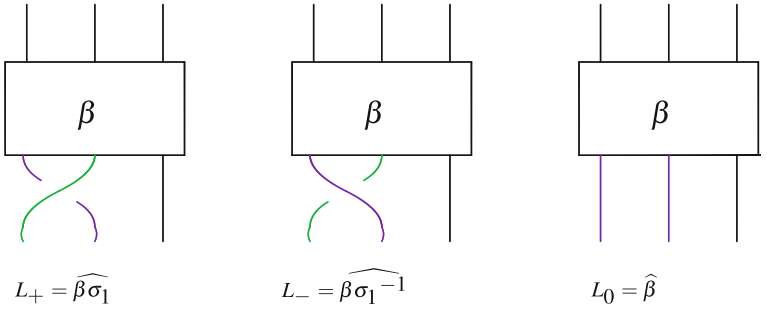


Fig. 1.3 The links in the special skein relation in open braid form

components. Then the skein relation reduces to the skein relation of the HOMFLYPT polynomial $P = P(q, \lambda_D)$ (1.16)

$$\frac{1}{\sqrt{\lambda_D}}\Phi_{d,D}(L_+) - \sqrt{\lambda_D}\Phi_{d,D}(L_-) = (q - q^{-1})\Phi_{d,D}(L_0), \tag{1.18}$$

see Fig. 1.3. Furthermore, if we take $\beta \in B_n$ *as a framed braid with all framings zero, then the above skein relation of* $\Phi_{d,D}$ *also holds for the invariants* Θ_d*, since it involves only classical links:*

$$\frac{1}{\sqrt{\lambda_D}}\Theta_d(L_+) - \sqrt{\lambda_D}\Theta_d(L_-) = (q - q^{-1})\Theta_d(L_0), \tag{1.19}$$

1.4.4 Behaviour of the Invariants Θ_d on Links – Θ_d as a Sum of HOMFLYPT Polynomials

This new special skein relation allows us to attack the problem diagrammatically. One can apply the special skein relation on mixed crossings resulting to a skein tree whose leaves consist of disjoint union of knots. Then using Theorem 1.7 one can compute the value of Θ_d on the initial link. Specifically, it has been proved inductively in [11], that this procedure can be applied:

Theorem 1.8 ([11, Theorem 6.16]) *For any* ℓ*-component link* L*, the value* $\Theta_d(L)$ *is a* $\mathbb{Q}[q^{\pm 1}, \sqrt{\lambda_D}^{\pm 1}]$*-linear combination of* $P(L)$ *and the values of* P *on disjoint unions of knots obtained by the skein relation:*

$$\Theta_d(L) = \sum_{k=1}^{\ell} E_D^{1-k} \sum_{\widehat{\alpha} \in \mathcal{N}(L)_k} c(\widehat{\alpha})\, P(\widehat{\alpha}) = P(L) + \sum_{k=2}^{\ell} (E_D^{1-k} - 1) \sum_{\widehat{\alpha} \in \mathcal{N}(L)_k} c(\widehat{\alpha})\, P(\widehat{\alpha}),$$

where $\mathscr{N}(L)_k$ denotes the set of all disjoint unions of k knots for $k = 1, \ldots, \ell$. Conversely, the value $P(L)$ is a $\mathbb{Q}[q^{\pm 1}, \sqrt{\lambda_D}^{\pm 1}]$-linear combination of $\Theta_d(L)$ and the values of Θ_d on disjoint unions of knots obtained by the skein relation:

$$P(L) = \sum_{k=2}^{\ell} E_D^{k-1} \sum_{\widehat{\alpha} \in \mathscr{N}(L)_k} c(\widehat{\alpha})\, \Theta_d(\widehat{\alpha}) = \Theta_d(L) + \sum_{k=2}^{\ell} (E_D^{k-1} - 1) \sum_{\widehat{\alpha} \in \mathscr{N}(L)_k} c(\widehat{\alpha})\, \Theta_d(\widehat{\alpha}).$$

Theorem 1.8 constitutes the first confirmation that Θ_d-equivalence does not necessarily imply P-equivalence and vice versa. Let now $d, d' \geq 2$ with $d \neq d'$. Using Theorem 1.8 one can write the invariants Θ_d and $\Theta_{d'}$ as a sum of HOMFLYPT polynomials and attempt to derive a relation connecting the two invariants. Indeed, this is possible and in fact is a generalization of Theorem 1.8:

Theorem 1.9 ([11, Theorem 6.18]) *Let $d, d' \in \mathbb{N}$. For any ℓ-component link L, the value $\Theta_{d'}(L)$ is an $\mathbb{Q}[q^{\pm 1}, \sqrt{\lambda_D}^{\pm 1}]$-linear combination of $\Theta_d(L)$ and the values of Θ_d on disjoint unions of knots obtained by the skein relation:*

$$\Theta_{d'}(L) = \Theta_d(L) + \sum_{k=2}^{\ell} \left(\left(\frac{E_D}{E_{D'}} \right)^{k-1} - 1 \right) \sum_{\widehat{\alpha} \in \mathscr{N}(L)_k} c(\widehat{\alpha})\, \Theta_d(\widehat{\alpha}).$$

It is immediate by Theorem 1.9 that Θ_d-equivalence does not necessarily imply $\Theta_{d'}$-equivalence for $d \neq d'$. However, for 2-component P-equivalent links the following result holds:

Theorem 1.10 ([11, Theorem 7.1]) *Let $d, d' \geq 2$ and let L_1 and L_2 be a pair of 2-component P-equivalent links. Then L_1 and L_2 are Θ_d-equivalent if and only if they are $\Theta_{d'}$-equivalent.*

1.4.5 Behaviour of the Invariants Θ_d on Links – A Skein-Theoretic Approach

Theorem 1.8 provides an algorithmic procedure to compute Θ_d diagrammatically as follows:

Step 1. Apply the skein relation of Proposition 1.5 on crossings linking different components until the link L is decomposed into disjoint unions of knots. An algorithmic process for achieving this is the following: we order the components of L and we select a starting point on each component. Starting from the chosen point of the first component and following its orientation we apply the skein relation on all mixed crossings we encounter, so that the arcs of this component are always overarcs. We proceed similarly with the second component changing all mixed crossing except for crossings involving the

first component, and so on. In the end we obtain the split version of the original link.

Step 2. Following Theorem 1.8 and its notation, we obtain

$$\Theta_d(L) = \sum_{k=1}^{\ell} E_D^{1-k} \sum_{\widehat{\alpha} \in \mathscr{N}(L)_k} c(\widehat{\alpha})\, P(\widehat{\alpha}).$$

Step 3. Apply the skein relation (1.16) of the HOMFLYPT polynomial to obtain the value of P on $\widehat{\alpha}$ at variables (q, λ_D), for all disjoint unions of knots $\widehat{\alpha} \in \mathscr{N}(L)_k$, $k = 1, \ldots, \ell$.

Since the invariants Θ_d are well-defined via braid methods, we obtain by Theorem 1.8 the following:

Theorem 1.11 ([11, Theorem 6.19]) *The invariants Θ_d can be completely defined via the HOMFLYPT skein relation.*

1.4.6 Behaviour of the Invariants Θ_d on Links – Comparison with the HOMFLYPT Polynomial on Links

Theorem 1.8 and the special skein relation (1.19) allow us to compare diagrammatically the invariants Θ_d to the HOMFLYPT polynomial on various links.

It is known that the HOMFLYPT polynomial, being a skein link invariant, does not distinguish *mutant* knots or links. The operation of mutation on a link diagram is defined by choosing a disk that intersects the diagram at exactly four points and then rotating 180° the 2-tangle encircled by the disk. Investigating with the use of the trace $\mathrm{tr}_{d,D}$ whether the invariants Θ_d distinguish mutant knots or links would be impossible. However, using the special skein relation, it is possible to prove the following result:

Proposition 1.6 ([11, Proposition 6.5]) *Let L and L' be two mutant links. Then $\Theta_d(L) = \Theta_d(L')$.*

In order to compare the invariants Θ_d to the HOMFLYPT polynomial on examples of P-equivalent pairs of links, not isotopic to each other as unoriented link, computations were needed in order to calculate the values of the invariants Θ_d on them. Using the data from [9], 89 pairs of such links up to 11 crossings were found and we computed, using the program of Remark 1.5 [35], the values of the invariants Θ_d on them using the program of Remark 1.5. Out of these 89 pairs, there are 6 pairs of P-equivalent links that are not Θ_d-equivalent for every $d \geq 2$ [11] (Table 1.1):

Table 1.1 Six P-equivalent pairs of 3-component links that are not Θ_d-equivalent

$L11n358\{0,1\}$	$L11n418\{0,0\}$
$L11a467\{0,1\}$	$L11a527\{0,0\}$
$L11n325\{1,1\}$	$L11n424\{0,0\}$
$L10n79\{1,1\}$	$L10n95\{1,0\}$
$L11a404\{1,1\}$	$L11a428\{0,1\}$
$L10n76\{1,1\}$	$L11n425\{1,0\}$

Specifically, for these pairs the differences of the polynomials have been computed [11]:

$$\Theta_d(L11n358\{0,1\}) - \Theta_d(L11n418\{0,0\}) = \frac{(E_D-1)(\lambda_D-1)(q-1)^2(q+1)^2\left(q^2-\lambda_D\right)\left(\lambda_D q^2-1\right)}{E_D\lambda_D^4 q^4},$$

$$\Theta_d(L11a467\{0,1\}) - \Theta_d(L11a527\{0,0\}) = \frac{(E_D-1)(\lambda_D-1)(q-1)^2(q+1)^2\left(q^2-\lambda_D\right)\left(\lambda_D q^2-1\right)}{E_D\lambda_D^4 q^4},$$

$$\Theta_d(L11n325\{1,1\}) - \Theta_d(L11n424\{0,0\}) = -\frac{(E_D-1)(\lambda_D-1)(q-1)^2(q+1)^2\left(q^2-\lambda_D\right)\left(\lambda_D q^2-1\right)}{E_D\lambda_D^3 q^4},$$

$$\Theta_d(L10n79\{1,1\}) - \Theta_d(L10n95\{1,0\}) = \frac{(E_D-1)(\lambda_D-1)(q-1)^2(q+1)^2\left(\lambda_D+\lambda_D q^4+\lambda_D q^2-q^2\right)}{E_D\lambda_D^4 q^4},$$

$$\Theta_d(L11a404\{1,1\}) - \Theta_d(L11a428\{0,1\}) = \frac{(E_D-1)(\lambda_D-1)(\lambda_D+1)(q-1)^2(q+1)^2\left(q^4-\lambda_D q^2+1\right)}{E_D q^4},$$

$$\Theta_d(L10n76\{1,1\}) - \Theta_d(L11n425\{1,0\}) = \frac{(E_D-1)(\lambda_D-1)(\lambda_D+1)(q-1)^2(q+1)^2}{E_D\lambda_D^3 q^2}.$$

Note that the factor $(E_D - 1)$ is common to all six pairs. This confirms that the pairs have the same HOMFLYPT polynomial, since for $E_D = 1$ the difference collapses to zero. Further, all the computations can be found on http://www.math.ntua.gr/~sofia/yokonuma. Except for the computational results, there is a diagrammatic proof for the pair of links $L11n358\{0,1\}$ and $L11n418\{0,0\}$ in [11]. Now, we can formulate the following immediate statement:

Theorem 1.12 ([11, Theorem 7.3]) *The invariants Θ_d are not topologically equivalent to the HOMFLYPT polynomial for any $d \geq 2$.*

The proof uses recursive applications of the special skein relation (1.19), in order to construct skein trees, where only disjoint union of knots appear as leaves, for both links. Then using the split link property of Θ_d (1.20), the value of Θ_d is written as a

sum of HOMFLYPT polynomials of knots. Finally, by recognizing the topological type of the knot diagrams and by finding the HOMFLYPT polynomial values for these knots, the calculation is completed. Note that the intrinsic difference in computing the invariants Θ_d and P is in the different values of these invariants on disjoint unions of knots. In particular, if K is a knot and U is the unknot, for the invariants Θ_d:

$$\Theta_d(K \sqcup U) = \frac{1-\lambda_D}{(q-q^{-1})\sqrt{\lambda_D}E_D}\Theta_d(K),$$

while for P we have that:

$$P(K \sqcup U) = \frac{1-\lambda_D}{(q-q^{-1})\sqrt{\lambda_D}}P(K).$$

Theorem 1.12 confirms that P-equivalence does not imply Θ_d-equivalence, but it cannot provide us with any indication as to whether or not the invariants Θ_d are strictly stronger than the HOMFLYPT polynomial.

1.5 A 3-Variable Generalization of the HOMFLYPT Polynomial

1.5.1 The Invariant $\Theta(q, \lambda, E)$

The program of Remark 1.5 considers the value E_D as a parameter. Moreover, the six pairs of links of Table 1.1 are distinguished by Θ_d for every $d \geq 2$, as seen of the difference of the values of Θ_d on each pair. The natural question arising is whether the value E_D can be considered as an indeterminate, allowing us to construct a link invariant generalizing both the HOMFLYPT polynomial and the invariants Θ_d. Indeed, in [11] such an invariant has been constructed:

Theorem 1.13 ([11, Theorem 8.1]) *Let q, λ, E be indeterminates. There exists a unique isotopy invariant of classical oriented links $\Theta : \mathscr{L} \to \mathbb{C}[q^{\pm 1}, \lambda^{\pm 1}, E^{\pm 1}]$ defined by the following rules:*

1. *For a disjoint union L of k knots, with $k \geq 1$, it holds that:*

$$\Theta(L) = E^{1-k}\, P(L).$$

2. *On crossings involving different components the following skein relation holds:*

$$\frac{1}{\sqrt{\lambda}}\,\Theta(L_+) - \sqrt{\lambda}\,\Theta(L_-) = (q-q^{-1})\,\Theta(L_0),$$

where L_+, L_-, L_0 is a Conway triple.

The invariant Θ distinguishes the 6 pairs of Table 1.1. Moreover, since Θ generalizes both the invariants Θ_d and the HOMFLYPT polynomial, we have the following:

Theorem 1.14 ([11, Theorem 8.2]) *The invariant $\Theta(q, \lambda, E)$ is stronger than the HOMFLYPT polynomial.*

Remark 1.7 By the above, the computer program of Remark 1.5 computes the invariant Θ.

As we shall see, the well-definedness of Θ can be proved using a variety of techniques, from diagrammatic ones to algebraic or combinatorial ones.

1.5.2 Properties of the Framed and Classical Invariants

The invariants $\Phi_{d,D}$ and Θ (and hence also the invariants Θ_d) satisfy properties analogous to the known ones of the HOMFLYPT polynomial, due to the behaviour of trace $\mathrm{tr}_{d,D}$ under inversion, split links, connected sums and mirror imaging. In detail:

- *Reversing orientation:*

$$\Phi_{d,D}(L) = \Phi_{d,D}(\overleftarrow{L}) \quad \text{and} \quad \Theta(L) = \Theta(\overleftarrow{L}),$$

  where $\overleftarrow{L}$ is the link L with reversed orientation on all components.
- *Split links:*

$$\begin{aligned} \Phi_{d,D}(L \sqcup L') &= \Lambda_D \, \Phi_{d,D}(L)\Phi_{d,D}(L') \\ \text{and} \quad \Theta(L \sqcup L') &= \frac{1-\lambda_D}{(q-q^{-1})\sqrt{\lambda_D}E}\,\Theta(L)\Theta(L'). \end{aligned} \tag{1.20}$$

- *Connected sums:*

$$\Phi_{d,D}(L\#L') = \Phi_{d,D}(L)\Phi_{d,D}(L') \quad \text{and} \quad \Theta(L\#L') = \Theta(L)\Theta(L'),$$

  where D is a subset of $\mathbb{Z}/d\mathbb{Z}$ such that $\mathrm{x}_1^d = 1$ and $\mathrm{x}_k = \mathrm{x}_1^k$ for all $k = 1, \ldots, d-1$.
- *Mirror images:*

$$\Phi_{d,D}(q, \lambda_D)(L^*) = \Phi_{d,D}(q^{-1}, \lambda_D^{-1})(L) \quad \text{and} \quad \Theta(q, \lambda_D)(L^*) = \Theta(q^{-1}, \lambda_D^{-1})(L),$$

  where L^* is the mirror image of L.

Notice that P satisfies the exact same properties as Θ except for split links, where the parameter E (or the value E_D) does not appear:

$$P(L \sqcup L') = \frac{1-\lambda_{\mathrm{H}}}{(q-q^{-1})\sqrt{\lambda_{\mathrm{H}}}} P(L)P(L').$$

1.5.3 A Closed Formula for Θ

Before we proceed to explain how the well-definedness of the invariant Θ can be proved, we provide a closed formula for Θ, proved by W.B.R. Lickorish. More precisely, the invariant Θ is a complicated mixture of linking numbers and HOMFLYPT polynomials of sublinks. In detail:

Theorem 1.15 ([11, Appendix B by W.B.R. Lickorish]) *Let L be an oriented link with n components. Then*

$$\Theta(L) = \sum_{k=1}^{n} \mu^{k-1} E_k \sum_{\pi} \lambda^{\nu(\pi)} P(\pi L) \tag{1.21}$$

where the second summation is over all partitions π of the components of L into k (unordered) subsets and $P(\pi L)$ denotes the product of the HOMFLYPT polynomials of the k sublinks of L defined by π. Furthermore, $\nu(\pi)$ is the sum of all linking numbers of pairs of components of L that are in distinct sets of π, $E_k = (E^{-1} - 1)(E^{-1} - 2)\cdots(E^{-1} - k + 1)$, with $E_1 = 1$, and $\mu = \frac{\lambda^{-1/2} - \lambda^{1/2}}{q - q^{-1}}$.

The above formula provides us with a topological interpretation of the invariant Θ. Specifically, the invariant Θ is *completely determined by the linking matrix of a link L and the values of P on each sublink of L*. For example, on the pair of links of Theorem 1.12 the invariant Θ detects a pair of 2-component sublinks that are not P-equivalent. Specifically, the link $L11n358\{0, 1\}$ contains a disjoint union of two unknots as a sublink, whereas $L11n418\{0, 0\}$ does not; hence, there is a pair of sublinks with different HOMFLYPT polynomials.

Theorem 1.15 is proved by W.B.R. Lickorish using the special skein relation and combinatorial tools. For more details, the reader can refer to [11, Appendix B]. The above result has also been proved in [42] using representation theory techniques.

Moreover, Theorem 1.15 allows us to investigate further the question of whether Θ_d-equivalence implies $\Theta_{d'}$-equivalence or vice versa. In detail:

Proposition 1.7 ([11, Proposition 8.9]) *Let L and L' be two n-component links that are not Θ-equivalent. Then they are not Θ_d-equivalent for $d \geq n$.*

The proof uses exclusively Theorem 1.15. In detail, for $d \geq n$ we do not lose any topological information, since all the coefficients E_k are not zero for all $k \in \{1, \ldots, n\}$. However, for $d < n$ the coefficients E_k for $k \geq d$ are zero and hence a pair of HOMFLYPT non-equivalent sublinks may not be detected. Hence, for a pair of

P-equivalent links L and L', either all the invariants Θ_d coincide and do not distinguish the links or there exists a $d \in \mathbb{N}$ with $2 \leq d < n$ such that $\Theta_d(L) \neq \Theta_d(L')$.

In [11] it is shown that Θ can indeed be defined by (1.21):

Theorem 1.16 ([11, Theorem 8.11]) *Suppose that the invariant Θ is defined by Eq. 1.21. Then, this definition is equivalent to the definition of Theorem 1.13.*

Note that the definition via (1.21) provides a quick way to compute the invariant Θ for a link L. One only needs to identify the sublinks of a link and compute Θ using the known values of P on the sublinks and the linking matrix of L.

1.5.4 The Well-Definedness of Θ

Chronologically, our first proof of the well-definedness of Θ was algebraic using the class of *tied links* [2]. However, there exist other methods for proving that Θ is well-defined. Summarizing, there are the following four equivalent methods:

1. *combinatorially*, via the closed formula of W.B.R. Lickorish (Theorem 1.15);
2. *skein-theoretically* based on Theorem 1.13 (a direct proof can be found in [36]);
3. *algebraically*, via the *algebra of braids and ties* $\mathscr{E}_n(q)$ [2] generated by $g_1, \ldots, g_{n-1}$ with the algebra of braids and ties.
4. *algebraically*, via the isomorphism of the subalgebra $\mathrm{Y}_{d,n}(q)^{(\mathrm{br})}$ of $\mathrm{Y}_{d,n}(q)$ generated by $g_1, \ldots, g_{n-1}$ with the algebra of braids and ties $\mathscr{E}_n(q)$ for $d \geq n$ [17].

All the above methods do not involve complicated constructions such as the E–system, even though Θ contains the invariants Θ_d where the E–system is needed. Moreover, the restriction $d \geq n$ of Theorem 1.17 does not obstruct the well-definedness of Θ: for a link L written as a braid in n strands one can always choose a suitable $d \geq n$.

Now we will provide some more insight on the last two algebraic methods for proving the well-definedness of Θ using *tied links*. Tied links were introduced and studied by F. Aicardi and J. Juyumaya in [2, 3]. A *tied link* is defined as a classical link L endowed with a set of ties, containing unordered pairs of points belonging to the components of L [3, Definition 1]. Diagrammatically, one can visualize a tie as a spring connecting two (not necessarily different) components of L. The endpoints of a tie are allowed to slide along the components that they are attached to. If two ties join the same two components, one of them can be removed, and any tie on a single component can be also removed. A tie that cannot be removed is called *essential*.

Tied link invariants can be constructed using either diagrammatic or algebraic methods. In [3] such an invariant is defined using both methods. Specifically, a tied link invariant is constructed with the use of a Markov trace on the *algebra of braids and ties* $\mathscr{E}_n(q)$. The algebra of braids and ties $\mathscr{E}_n(q)$ is defined as the algebra generated

by $g_1, \ldots, g_{n-1}, e_1, \ldots, e_{n-1}$ satisfying the following relations (cf. [2, Definition 1]):

$$\begin{aligned}
g_i g_j g_i &= g_j g_i g_j && \text{for } |i-j| = 1\\
g_i g_j &= g_j g_i && \text{for } |i-j| > 1\\
e_i e_j &= e_j e_i\\
e_i^2 &= e_i\\
e_i g_i &= g_i e_i\\
e_i g_j &= g_j e_i && \text{for } |i-j| > 1\\
e_i e_j g_i &= g_i e_i e_j && \text{for } |i-j| = 1\\
e_i g_j g_i &= g_j g_i e_j && \text{for } |i-j| = 1\\
g_i^2 &= 1 + (q - q^{-1})\, e_i g_i\,.
\end{aligned}$$

Diagrammatically, the generators g_i correspond to the classical braiding generators and the elements e_i correspond to ties connecting the i-th and the $(i+1)$-th strands. Note that the cancellation properties of ties mentioned above are reflected in the fact that the elements e_i are idempotents. In [2, 3], a different presentation $\mathscr{E}_n(u)$ is used, where the quadratic relation is changed.

The similarity between the algebra $\mathscr{E}_n(q)$ and $\mathrm{Y}_{d,n}(q)^{(\mathrm{br})}$ is obvious: both can be generated by the same generators and the generators of $\mathrm{Y}_{d,n}(q)^{(\mathrm{br})}$ satisfy the exact same relations. However, it is not evident that these relations are enough for the subalgebra $\mathrm{Y}_{d,n}(q)^{(\mathrm{br})}$ and whether the two algebras are isomorphic. In [17], the authors have provided a representation-theoretic proof for the isomorphism between the two algebras:

Theorem 1.17 ([17, Theorem 8]) *Suppose that $d \geq n$. Then the algebra $\mathscr{E}_n(q)$ is isomorphic to the subalgebra $\mathrm{Y}_{d,n}(q)^{(\mathrm{br})}$ of $\mathrm{Y}_{d,n}(q)$.*

Note also the similarity of the condition $d \geq n$ of Theorem 1.17 and Proposition 1.7.

Now, a Markov trace $\rho : \bigcup_{n\geq 0} \mathscr{E}_n(q) \to \mathbb{C}[q^{\pm 1}, z^{\pm 1}, E^{\pm 1}]$ can be defined satisfying the following rules (cf. [2, Theorem 3]):

$$\begin{array}{lrcll}
\text{(i)} & \rho(ab) & = & \rho(ba) & a, b \in \mathscr{E}_n(q)\\
\text{(ii)} & \rho(1) & = & 1 & 1 \in \mathscr{E}_n(q)\\
\text{(iii)} & \rho(ag_n) & = & z\,\rho(a) & a \in \mathscr{E}_n(q) \quad (\textit{Markov property})\\
\text{(iv)} & \rho(ae_n) & = & E\,\rho(a) & a \in \mathscr{E}_n(q)\\
\text{(v)} & \rho(ae_n g_n) & = & z\,\rho(a) & a \in \mathscr{E}_n(q)\,.
\end{array}$$

Notice the resemblance of the above rules with the five rules of $\mathrm{tr}_{d,D}$ in Theorem 1.2. Further, the trace ρ satisfies similar properties to those of $\mathrm{tr}_{d,D}$ [11]. In [3] the *tied braid monoid* TB_n is defined; it is generated by the braiding generators $\sigma_1, \ldots, \sigma_{n-1}$ and the generating ties $\eta_1, \ldots, \eta_{n-1}$, where η_i connects the i-th and the $i+1$-th strands of a tied braid. Denote by $\bar{\pi} : \mathbb{C}TB_n \to \mathscr{E}_n(q)$ the natural surjection defined by $\sigma_i \mapsto g_i$ and $\eta_i \mapsto e_i$. Then following the procedure of [3] an invariant $\overline{\Theta}$ can be defined as:

Theorem 1.18 ([11, Theorem 8.4]) *For any tied braid $\alpha \in TB_n$, we define*

$$\overline{\Theta}(\widehat{\alpha}) := \left(\frac{1}{z\sqrt{\lambda}}\right)^{n-1} \sqrt{\lambda}^{\epsilon(\alpha)} (\rho \circ \bar{\pi})(\alpha)\,,$$

where $\lambda = \frac{z-(q-q^{-1})E}{z}$ and $\epsilon(\alpha)$ is the sum of the exponents of the braiding generators σ_i in the word α. Then the map $\overline{\Theta}$ is a 3-variable isotopy invariant of oriented tied links.

Note that, for $E = 1$, $\overline{\Theta}$ specializes to the HOMFLYPT polynomial when restricted to classical links.

In [3], a similar invariant has been defined using the presentation $\mathscr{E}_n(q)$. This invariant is re-defined diagrammatically via a skein relation that applies to any crossing in the link diagram. It is proved to be well-defined via the standard Lickorish–Millett method [39]. The same can be done for $\overline{\Theta}$. Specifically, $\overline{\Theta}$ satisfies the following defining skein relation:

$$\frac{1}{\sqrt{\lambda}}\overline{\Theta}(L_+) - \sqrt{\lambda}\,\overline{\Theta}(L_-) = (q - q^{-1})\,\overline{\Theta}(L_{0,\sim}), \tag{1.22}$$

For the well-definedness of Θ one only needs to show that Θ coincides with $\overline{\Theta}$ on classical links and that $\overline{\Theta}$ satisfied the defining rules of Θ as in Theorem 1.13. Indeed, the skein relation (1.22) reduces to the special skein relation of the second rule of Theorem 1.13 and it also satisfies the property of Θ for disjoint unions of knots [11].

1.6 Other Invariants from the Yokonuma–Hecke Algebras

A well-known property of the HOMFLYPT polynomial P, as defined via the Iwahori–Hecke algebra, is that a transformation $g_i \mapsto cg_i$, where $c \in \mathbb{C}$, leads to a change of variables for P. In the case of the Yokonuma–Hecke algebras this is not always true, because we have more possibilities for applying a linear transformation on the braiding generators. Indeed, a transformation of the form $g_i \mapsto cg_i$ would work similarly for the framed link invariants $\Phi_{d,D}$ as for P. However, applying a transformation of the form $g_i \mapsto cg_i + c'e_i g_i$, with $c' \in \mathbb{C}$, does not lead to a change of variables for $\Phi_{d,D}$ but rather to new invariants, potentially topologically non-equivalent to the invariants $\Phi_{d,D}$. Indeed, as we shall see below, this transformation gives rise to an algebra isomorphic to $\mathrm{Y}_{d,n}(q)$ but with a different quadratic relation.

1.6.1 The Old Quadratic Relation

We begin by summarizing the construction of framed and classical links invariants using the other presentation of the Yokonuma–Hecke algebra that we shall denote

by $Y_{d,n}(u)$ and that was used in the papers [12, 29–34]. This old presentation was also used originally for defining the invariants for framed, classical, singular and transverse links described in [10, 11], recall previous sections, but then it was adapted to the presentation $Y_{d,n}(q)$ that has simpler quadratic relations.

The algebra $Y_{d,n}(u)$ is generated by the elements $\widetilde{g}_1, \ldots, \widetilde{g}_{n-1}$ and $t_1, \ldots, t_n$, satisfying relations (1.1) (with $\widetilde{g}_i$ corresponding to σ_i), (1.2) and the quadratic relations:

$$\widetilde{g}_i^{\,2} = 1 + (u-1)\, e_i + (u-1)\, e_i \widetilde{g}_i \qquad (1 \leq i \leq n-1). \tag{1.23}$$

The presentation $Y_{d,n}(q)$ used in this paper and in [10, 11, 35] was obtained in [13] by taking $u := q^2$ and $g_i := \widetilde{g}_i + (q^{-1} - 1)\, e_i \widetilde{g}_i$ (or, equivalently, $\widetilde{g}_i := g_i + (q-1)\, e_i g_i$).

In [26] a quadratic relation with two parameters is considered, which specializes to both the old and the new quadratic relation for the Yokonuma–Hecke algebra.

1.6.2 The Traces $\widetilde{\mathrm{tr}}_d$ and $\widetilde{\mathrm{tr}}_{d,D}$

Theorem 1.1 has been originally proved by J. Juyumaya using the presentation $Y_{d,n}(u)$ for the Yokonuma-Hecke algebra [29]. He proved that there exists a unique linear Markov trace $\widetilde{\mathrm{tr}}_d$ on $\bigcup_{n\geq 0} Y_{d,n}(u)$ defined inductively by the four rules of Theorem 1.1, where rule (3) is replaced by the rule:

$$(\widetilde{3})\ \widetilde{\mathrm{tr}}_d(a\widetilde{g}_n) = \widetilde{z}\,\widetilde{\mathrm{tr}}_d(a) \qquad a \in Y_{d,n}(u) \quad (Markov\ \ property)$$

for some indeterminate $\widetilde{z}$ over $\mathbb{C}$. Since his proof uses the inductive basis of $Y_{d,n}(u)$, it also works with the new quadratic relations (1.3), thus yielding Theorem 1.1. The E–condition and the E–system as presented in Sect. 1.2 were first defined and used in [31] in order to re-scale $\widetilde{\mathrm{tr}}_d$, and remain the same for $Y_{d,n}(q)$. In [12, Definition 3] the specialized trace $\widetilde{\mathrm{tr}}_{d,D}$ with parameter $\widetilde{z}$ is defined on $\bigcup_{n\geq 0} Y_{d,n}(u)$, satisfying the analogous rules: (1), (2), $(\widetilde{3})$ and (4′).

1.6.3 Related Invariants

Now, using the natural $\mathbb{C}$-algebra epimorphism from $\mathbb{C}\mathscr{F}_n$ onto $Y_{d,n}(u)$ given by $\sigma_i \mapsto \widetilde{g}_i$ and $t_j^k \mapsto t_j^{k(\mathrm{mod}\, d)}$ and abusing notation, one can define the trace $\widetilde{\mathrm{tr}}_d$ on the elements of $\mathbb{C}\mathscr{F}_n$, and thus, in particular, on the elements of $\mathscr{F}_n$. By normalizing and re-scaling the specialized trace $\widetilde{\mathrm{tr}}_{d,D}$, invariants $\Gamma_{d,D}(u, \widetilde{z})$ for oriented framed links are defined in [31, Theorem 8].

As it turned out [31, Proposition 7], the invariants $\Gamma_{d,D}$ satisfy the following skein relation, involving the braiding and the framing generators:

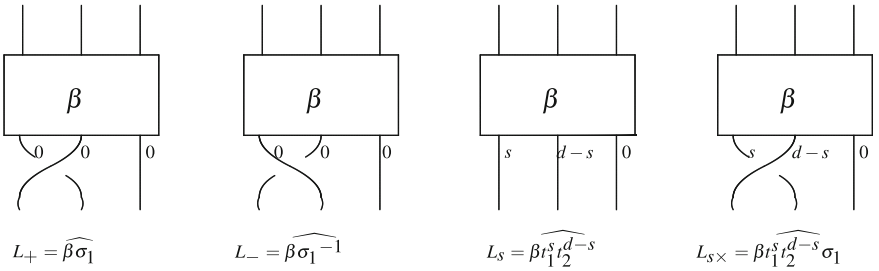


Fig. 1.4 The framed links of the skein relation in open braid form

$$\frac{1}{\sqrt{\widetilde{\lambda}_D}}\Gamma_{d,D}(L_+) - \sqrt{\widetilde{\lambda}_D}\Gamma_{d,D}(L_-) = \frac{1-u^{-1}}{d}\sum_{s=0}^{d-1}\Gamma_{d,D}(L_s) + \frac{1-u^{-1}}{d\sqrt{\widetilde{\lambda}_D}}\sum_{s=0}^{d-1}\Gamma_{d,D}(L_{s\times}) \tag{1.24}$$

where

$$\widetilde{\lambda}_D = \frac{\widetilde{z} - (u-1)E_D}{u\,\widetilde{z}},$$

and the links L_+, L_-, L_s and $L_{s\times}$ are illustrated in Fig. 1.4. Comparing this skein relation to the corresponding skein relation of $\Phi_{d,D}$ derived from $\mathrm{Y}_{d,n}(q)$ (1.15) we can see that terms involving $L_{s\times}$ are missing in (1.15).

Similarly to $\Phi_{d,D}(q,z)$, the invariants $\Gamma_{d,D}(u,\widetilde{z})$ become invariants of oriented classical links, denoted by $\Delta_{d,D}(u,\widetilde{z})$, when the traces $\widetilde{\mathrm{tr}}_d$, $\widetilde{\mathrm{tr}}_{d,D}$ are applied on the classical braid groups B_n and these are the ones studied in [12, 32]. Theorem 1.2 and all its consequences hold also for the specialized trace $\widetilde{\mathrm{tr}}_{d,D}$. In particular, the values of the classical link invariants $\Delta_{d,D}$ depend only on the cardinality $|D|$ of D, so they can be parametrized by the natural numbers, setting $\Delta_d := \Delta_{d,\mathbb{Z}/d\mathbb{Z}}$ for all $d \in \mathbb{Z}_{>0}$. For $d = 1$ we have that $\Delta_1 = \Delta_{1,\{0\}}(u,\widetilde{z}) = \Theta_{1,\{0\}}(q,z)$, the HOMFLYPT polynomial, for $u = q^2$ and $\widetilde{z} = qz$. This can be easily seen by comparing the skein relations for $\Gamma_{1,\{0\}}$ and $\Phi_{1,\{0\}}$. So, both families of invariants Δ_d and Θ_d include the HOMFLYPT polynomial as a special case.

However, the skein relation (1.24) of the invariants $\Gamma_{d,D}$ does not yield a special skein relation for the invariants Δ_d similar to (1.19) of Θ_d. Indeed, if the crossing of L_+ involves two different components, then so does the crossing of $L_{s\times}$ and so the framings in $L_{s\times}$ cannot be collected together. Consequently, Theorem 1.11 and all other results for Θ_d that depend on the special skein relation are not valid for the invariants Δ_d. Clearly, the diagrammatic analysis made for the invariants Θ_d on pairs of P-equivalent links cannot be implemented for the invariants Δ_d. Nevertheless, there are computational indications that the invariants Δ_d are not topologically equivalent to P. Concerning now the properties studied in Sect. 1.2.4, Δ_d has the same behaviour as Θ_d on links with reversed orientation, on split links, on connected sums and on mirror images. However, behaviour of Δ_d under mutation cannot be checked using the methods of Proposition 1.6. Furthermore, there is no reason that

the invariants $\Delta_d(u, \widetilde{z})$ and $\Theta_d(q, z)$ are topologically equivalent. In fact, there is computational evidence that they are not [1].

Moreover, using Theorem 1.17, E_D could be taken to be an indeterminate E, and $\widetilde{\mathrm{tr}}_d$ would be well-defined due to the isomorphism of $\mathrm{Y}_{d,n}(u)^{(\mathrm{br})}$ with the algebra of braids and ties $\mathscr{E}_n(u)$ for $d \geq n$; then $\widetilde{\mathrm{tr}}_d$ would coincide with the Markov trace on $\mathscr{E}_n(u)$ defined in [2]. More precisely, in [2, 3], F. Aicardi and J. Juyumaya worked on the algebra of braids and ties $\mathscr{E}_n(u)$, generated by elements $\widetilde{g}_1, \ldots, \widetilde{g}_{n-1}, e_1, \ldots, e_{n-1}$, with the braiding generators satisfying the old quadratic relations (1.23). Then they defined a Markov trace $\widetilde{\rho}$ on $\cup_{n\geq 0}\mathscr{E}_n(u)$ [2, Theorem 3] that gave rise to a 3-variable isotopy invariant of tied links, denoted by $\overline{\Delta}$. Our construction of $\overline{\Theta}$ (recall Sect. 1.5.4) is completely analogous to the construction of $\overline{\Delta}$. For $E = 1$, $\overline{\Delta}$ specializes to the HOMFLYPT polynomial when restricted to classical links. In [3], $\overline{\Delta}$ is re-defined diagrammatically via a skein relation that applies to any crossing in the link diagram. The invariant $\overline{\Delta}$ has not been identified topologically. One obstruction to this is the fact that the old quadratic relation is used for the algebra $\mathscr{E}_n(u)$. So, it was impossible to derive a special skein relation that only involves classical links (with no ties). Despite the fact that the algebras $\mathscr{E}_n(u)$ and $\mathscr{E}_n(q)$ are isomorphic, the invariants $\overline{\Delta}$ and $\overline{\Theta}$ are also not necessarily topologically equivalent [1] as we have already observed about the invariants Δ_d and Θ_d.

Restricting now $\overline{\Delta}$ to classical links, similarly to the proof of Theorem 1.13 [11], one can prove that the invariant $\overline{\Delta}$ satisfies the first rule of Theorem 1.13. Given also the isomorphism between the subalgebra $\mathrm{Y}_{d,n}^{(\mathrm{br})}(u)$ of $\mathrm{Y}_{d,n}(u)$ and the algebra of braids and ties $\mathscr{E}_n(u)$ for $d \geq n$, in this case the invariant $\overline{\Delta}$ contains the invariants Δ_d. Consequently, the invariants Δ_d are topologically equivalent to the HOMFLYPT polynomial on knots and on disjoint unions of knots.

Acknowledgements This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES: Reinforcement of the interdisciplinary and/or inter-institutional research and innovation.

References

1. Private communication with F. Aicardi
2. Aicardi, F., Juyumaya, J.: Markov trace on the algebra of braids and ties. Mosc. Math. J. **16**(3), 397–431 (2016)
3. Aicardi, F., Juyumaya, J.: Tied links. J. Knot Theory Ramif. **25**(9), 1641001 (2016)
4. Aicardi, F., Juyumaya, J.: Kauffman type invariants for tied links. arXiv:1607.04841 [math.GT]
5. Baez, J.C.: Link invariants of finite type and perturbation theory. Lett. Math. Phys. **26**(1), 43–51 (1992)
6. Bennequin, D.: Entrelacements et équations de Pfaffe. Asterisque **107–108**, 87–161 (1983)
7. Birman, J.S.: New points of view in knot theory. Bull. Am. Math. Soc. (N.S.) **28**(2), 253–287 (1993)
8. Chmutov, S., Duzhin, S., Mostovoy, Y.: Introduction to Vassiliev Knot Invariants. Cambridge University Press, Cambridge (2012)

9. Cha, J.C., Livingston, C.: LinkInfo: Table of Knot Invariants. http://www.indiana.edu/~linkinfo (2015). Accessed 16 April 2015
10. Chmutov, S., Jablan, S., Karvounis, K., Lambropoulou, S.: On the link invariants from the Yokonuma-Hecke algebras. J. Knot Theory Ramif. **25**(9), 1641004 (2016)
11. Chlouveraki, M., Juyumaya, J., Karvounis, K., Lambropoulou, S.: Identifying the invariants for classical knots and links from the Yokonuma–Hecke algebra, submitted for publication. arXiv: 1505.06666 [math.GT]
12. Chlouveraki, M., Lambropoulou, S.: The Yokonuma-Hecke algebras and the HOMFLYPT polynomial. J. Knot Theory Ramif. **22**(14), 1350080 (2013)
13. Chlouveraki, M., Poulain d'Andecy, L.: Representation theory of the Yokonuma-Hecke algebra. Adv. Math. **259**, 134–172 (2014)
14. Chlouveraki, M., Poulain d'Andecy, L.: Markov traces on affine and cyclotomic Yokonuma-Hecke algebras. Int. Math. Res. Not. **2016**(14), 4167–4228 (2016)
15. Chlouveraki, M., Pouchin, G.: Determination of the representations and a basis for the Yokonuma-Temperley-Lieb algebra. Algebras Represent. Theory **18**(2), 421–447 (2015)
16. Chlouveraki, M., Pouchin, G.: Representation theory and an isomorphism theorem for the Framisation of the Temperley–Lieb algebra. Math. Z., 11–24 (2016)
17. Espinoza, J., Ryom-Hansen, S.: Cell structures for the Yokonuma–Hecke algebra and the algebra of braids and ties, submitted for publication. arXiv:1506.00715
18. Flores, M., Juyumaya, J., Lambropoulou, S.: A Framization of the Hecke algebra of Type B. arXiv:1603.08487
19. Freyd, P., Hoste, J., Lickorish, W.B.R., Millett, K., Ocneanu, A., Yetter, D.: A new polynomial invariant of knots and links. Bull. Am. Math. Soc. (N.S.) **12**(2), 239–246 (1985)
20. Fuchs, D., Tabachnikov, S.: Invariants of Legendrian and transverse knots in the standard contact space. Topology **36**(5), 1025–1053 (1997)
21. Gemein, B.: Singular braids and Markov's theorem. J. Knot Theory Ramif. **6**(4), 441–454 (1997)
22. Goundaroulis, D., Juyumaya, J., Kontogeorgis, A., Lambropoulou, S.: The Yokonuma-Temperley-Lieb algebras. Banach Cent. Pub. **103**, 73–95 (2014)
23. Goundaroulis, D., Juyumaya, J., Kontogeorgis, A., Lambropoulou, S.: Framization of the Temperley–Lieb algebra, to appear in Mathematical Research Letters. arXiv:1304.7440
24. Goundaroulis, D., Lambropoulou, S.: Classical link invariants from the framization of the Iwahori–Hecke algebra and the Temperley–Lieb algebra of type A, to appear in J. Knot Theory Ramif. arXiv:1608.01812v1
25. Goundaroulis, D., Lambropoulou, S.: A new two-variable generalization of the Jones polynomial. arXiv:1608.01812v1
26. Jacon, N.: Poulain d'Andecy, L.: An isomorphism theorem for Yokonuma-Hecke algebras and applications to link invariants. Math. Z. **283**(1–2), 301–338 (2016)
27. Jones, V.F.R.: Hecke algebra representations of braid groups and link polynomials. Ann. Math. **126**(2), 335–388 (1987)
28. Juyumaya, J.: Another algebra from the Yokonuma–Hecke algebra. ICTP Preprint IC/1999/160
29. Juyumaya, J.: Markov trace on the Yokonuma-Hecke algebra. J. Knot Theory Ramif. **13**, 25–39 (2004)
30. Juyumaya, J., Lambropoulou, S.: p-adic framed braids. Topol. Appl. **154**, 149–191 (2013)
31. Juyumaya, J., Lambropoulou, S.: p-adic framed braids II. Adv. Math. **234**, 149–191 (2013)
32. Juyumaya, J., Lambropoulou, S.: An adelic extension of the Jones polynomial. In: Banagl, M., Vogel, D. (eds.) The Mathematics of Knots, Contributions in the Mathematical and Computational Sciences, Vol. 1, Springer, Berlin. arXiv:0909.2545v2 [math.GT]
33. Juyumaya, J., Lambropoulou, S.: An invariant for singular knots. J. Knot Theory Ramif. **18**(6), 825–840 (2009)
34. Juyumaya, J., Lambropoulou, S.: On the framization of knot algebras, in New Ideas in Low-Dimensional Topology. In: Kaufffman, L.H., Manturov, V. (eds.) Series Knots Everything. World Scientific Press, Singapore (2014)

35. Karvounis, K.: Enabling computations for link invariants coming from the Yokonuma-Hecke algebras. J. Knot Theory Ramif. **25**(9), 1641012 (2016)
36. Kauffman, L.H., Lambropoulou, S.: New invariants of links and their state sum models. arXiv:1703.03655 [math.GT]
37. Ko, K.H., Smolinsky, L.: The framed braid group and 3-manifolds. Proc. AMS **115**(2), 541–551 (1992)
38. Lambropoulou, S.: L-moves and Markov theorems. J. Knot Theory Ramif. **16**(10), 1–10 (2007)
39. Lickorish, W.B.R., Millett, K.: A polynomial invariant of oriented links. Topology **26**(1), 107–141 (1987)
40. Lusztig, G.: Character sheaves on disconnected groups VII. Represent. Theory **9**, 209266 (2005)
41. Orevkov, S.Y., Shevchishin, V.V.: Markov theorem for transversal links, J. Knot Theory Ramif. **12**(7) (2003) 905–913. Preprint arXiv:math/0112207v2 [math.GT]
42. Poulain d'Andecy, L., Wagner, E.: The HOMFLYPT polynomials of sublinks and the Yokonuma–Hecke algebras. Preprint arXiv:1606.00237v1 [math.GT]
43. Przytycki, J.H., Traczyk, P.: Invariants of links of Conway type. Kobe J. Math. **4**(2), 115–139 (1988)
44. Smolin, L.: Knot theory, loop space and the diffeomorphism group. New perspectives in canonical gravity, Monographs Textbooks Physics Science Lecture Notes, Vol. 5, pp. 245–266. Bibliopolis, Naples (1988)
45. Thiem, N.: Unipotent Hecke algebras of $\mathrm{GL}_n(\mathbb{F}_q)$. J. Algebra **284**, 559–577 (2005)
46. Wrinkle, N.: The Markov Theorem for Transverse Knots. Ph.D. thesis, Columbia University (2002). arXiv:math/0202055v1 [math.GT]
47. Yokonuma, T.: Sur la structure des anneaux de Hecke d'un groupe de Chevalley fini. C.R. Acad. Sci. Paris **264**, 344–347 (1967)

Chapter 2
A Survey on Temperley–Lieb-type Quotients from the Yokonuma–Hecke Algebras

Dimos Goundaroulis

Abstract In this survey we collect all results regarding the construction of the Framization of the Temperley–Lieb algebra of type A as a quotient algebra of the Yokonuma–Hecke algebra of type A. More precisely, we present all three possible quotient algebras the emerged during this construction and we discuss their dimension, linear bases, representation theory and the necessary and sufficient conditions for the unique Markov trace of the Yokonuma–Hecke algebra to factor through to each one of them. Further, we present the link invariants that are derived from each quotient algebra and we point out which quotient algebra provides the most natural definition for a framization of the Temperley–Lieb algebra. From the Framization of the Temperley–Lieb algebra we obtain new one-variable invariants for oriented classical links that, when compared to the Jones polynomial, they are not topologically equivalent since they distinguish more pairs of non isotopic oriented links. Finally, we discuss the generalization of the newly obtained invariants to a new two-variable invariant for oriented classical links that is stronger than the Jones polynomial.

2.1 Introduction

The Yokonuma–Hecke algebra was first introduced in the 60s by Yokonuma as a generalization of the Iwahori–Hecke algebra in the context of Chevalley groups [30]. In recent years, Juyumaya simplified the natural description by giving a presentation in terms of generators and relations [17–19]. A detailed overview of Juyumaya's approach can be found in [28, Preliminaries]. In this context, the Yokonuma–Hecke algebra of type A can be considered as a quotient of the framed braid group algebra over a two-sided ideal that is generated by a quadratic relation that involves certain weighted idempotent elements.

D. Goundaroulis (✉)
Center Intégratif de Génomique, Université de Lausanne,
Lausanne, Switzerland
e-mail: dimoklis.gkountaroulis@unil.ch

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_2

Throughout the past ten years the theory of Yokonuma–Hecke algebras received a significant amount of attention, mainly due to the concept of framization of knot algebras, a concept that was introduced by Juyumaya and Lambropoulou [24]. A knot algebra is an algebra that is involved in the construction of invariants of classical links via braid group representations [16]. To be more precise, a knot algebra A is a triplet (A, π, τ), where π is an appropriate representation of the braid group in A and τ is a Markov trace function defined on A. The Iwahori–Hecke algebra and the Temperley–Lieb algebra are the most known examples of knot algebras. On the other hand, the framization consists in an extension of a knot algebra via the addition of framing generators which gives rise to a new algebra that is related to framed braids and framed knots. The Yokonuma–Hecke algebra, $Y_{d,n}(u)$ is the basic example of this concept and it can be regarded as a framization of the Iwahori–Hecke algebra, $H_n(u)$ [20, 24]. With this in mind, Juyumaya and Lambropoulou proposed framizations of several knot algebras [23, 25] from which isotopy invariants for framed, classical and singular links were derived [20–22].

The breakthrough in this theory came while comparing the invariants for classical oriented knots and links from the Yokonuma–Hecke algebras to the HOMFLYPT polynomial. In [5], the use of a different presentation for $Y_{d,n}$ with parameter q instead of u and a different quadratic relation led to the proof that the derived two-variable invariants Θ_d are *not topologically equivalent to the HOMFLYPT polynomial on links* while they are topologically equivalent to the HOMFLYPT on knots. Furthermore, in the same work it was shown that the invariants Θ_d distinguish more pairs of non-isotopic oriented links than the HOMFLYPT polynomial. Moreover, it was shown that the invariants can be generalized to a 3-variable invariant Θ for oriented classical links that can be completely defined via the skein relation of the HOMFLYPT polynomial on crossings involving different components of the link and a set of initial conditions [5, 26]. The invariant Θ distinguishes the same pairs of HOMFLYPT-equivalent links as Θ_d, it is not topologically equivalent to the HOMFLYPT or the Kauffman polynomials and, thus, *it is stronger than the HOMFLYPT polynomial on links*.

One of the open problems in the concept of framization of knot algebras was the determination of a framization of the Temperley–Lieb algebra. If one considers the classical Temperley–Lieb algebra as it was introduced by Jones [16] that is, as a quotient of the Iwahori–Hecke algebra, it is immediately evident that desired framization will emerge as an appropriate quotient of the Yokonuma Hecke algebra. Contrary to the classical case such a candidate algebra is not unique. The study of these quotient algebras has been the topic of the author's Ph.D. thesis [11] which led to a series of results regarding their topological [12–15] as well as their algebraic properties [7, 8]. There are three potential candidates that can qualify as the framization of the Temperley–Lieb algebra: the *Yokonuma-Temperley–Lieb algebra* $YTL_{d,n}(u)$, the *Complex Reflection Temperley–Lieb algebra* $CTL_{d,n}(u)$ and the so-called *Framization of the Temperley–Lieb algebra* $FTL_{d,n}(u)$. The algebra $YTL_{d,n}(u)$ is too restricted and, as a consequence, the invariants for classical links from the algebra $YTL_{d,n}(u)$ just recover the Jones polynomial [12]. On the other hand, the algebra $CTL_{d,n}(u)$ is too large for our topological purposes and the derived link invariants coincide either with those from $Y_{d,n}(u)$ or with those from $FTL_{d,n}(u)$ [13]. Unfortunately, these two quotient algebras do not fit the topological purposes of deriving

new invariants for (framed) knots and links and, thus, they do not qualify as potential framizations of the Temperley–Lieb algebra. The third quotient algebra of $\mathrm{Y}_{d,n}(u)$, the Framization of the Temperley–Lieb algebra, $\mathrm{FTL}_{d,n}(u)$, lies between $\mathrm{YTL}_{d,n}(u)$ and $\mathrm{CTL}_{d,n}(u)$ and, as it will be made clear in Sect. 2.3, it turns out to be the right one [13]. The invariants θ_d for classical links from the algebras $\mathrm{FTL}_{d,n}$ adapted to a presentation with parameter q instead of u of $\mathrm{Y}_{d,n}(q)$, are proven to be *not topologically equivalent to the Jones polynomial on links* while they are topological equivalent to the Jones polynomial on knots [13]. Finally, in analogy to the invariants Θ_d, the invariants θ_d can be generalized to a new two-variable invariant of oriented classical links θ that is *stronger than the Jones polynomial* [15].

The outline of the paper is as follows: In Sect. 2.2 we introduce the necessary notations and we give a brief overview of all the required definitions and results such as: the Temperley–Lieb algebra, the Yokonuma–Hecke algebra, the E-system and the derived two-variable invariants for oriented framed and classical knots and links. In Sect. 2.3 we discuss three quotients of the Yokonuma–Hecke algebra as possible candidates for the framization of the Temperley–Lieb algebra. Moreover, we give all algebraic (linear basis, dimension, representation theory) as well as all topological (Markov trace, link invariants) results in the literature regarding each one of these quotient algebras. In Sect. 2.4 we describe how the invariants Θ_d and θ_d compare to the HOMFLYPT and Jones polynomials respectively. Finally, we discuss how the invariants Θ_d generalize to a new three-variable invariant for oriented classical links as well as the analogous generalization of the invariants θ_d to a new two-variable invariant for oriented classical links and we also describe closed combinatorial formulas for each one of the generalizations.

2.2 Preliminaries

In this section we will establish our notation and we will present the basic notions that will be used in the following sections.

2.2.1 Notations

We start by fixing two positive integers, d and n. Every algebra considered in this paper is an associative unital algebra over the field $\mathbb{C}(u)$, where u is an indeterminate. The *framed braid group* on n strands is defined as the semi-direct product of Artin's braid group B_n with n copies of $\mathbb{Z}$, namely: $\mathscr{F}_n = \mathbb{Z}^n \rtimes B_n$, where the action of the braid group B_n on $\mathbb{Z}^n$ is given by the permutation induced by a braid on the indices $\sigma_i t_j = t_{s_i(j)}\sigma_i$. By considering framings modulo d, the *modular framed braid group*, $\mathscr{F}_{d,n} = (\mathbb{Z}/d\mathbb{Z})^n \rtimes B_n$, is defined. Due to the above action a word w in $\mathscr{F}_n$ (resp. $\mathscr{F}_{d,n}$) has the *splitting property*, that is, it splits into the *framing* part and the *braiding* part $w = t_1^{a_1} \dots t_n^{a_n} \sigma$ where $\sigma \in B_n$ and $a_i \in \mathbb{Z}$ (resp. $\mathbb{Z}/d\mathbb{Z}$).

Finally, *a partition of* n, $\lambda = (\lambda_1, \ldots, \lambda_k)$, is a family of positive integers such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 1$ and $|\lambda| = \lambda_1 + \cdots + \lambda_k = n$. We identify every partition with its Young diagram, that is a left-justified array of k rows such that the jth row contains λ_j nodes for all $j = 1, \ldots, k$. A d-partition λ, or a Young d-diagram, of size n is a d-tuple of partitions such that the total number of nodes in the associated Young diagrams is equal to n. That is, we have $\lambda = (\lambda^{(1)}, \ldots, \lambda^{(d)})$ with $\lambda^{(1)}, \ldots, \lambda^{(d)}$, usual partitions such that $|\lambda^{(1)}| + \cdots + |\lambda^{(d)}| = n$.

2.2.2 The Temperley–Lieb algebra

For $n \geq 3$, the Temperley–Lieb algebra, $\mathrm{TL}_n(u)$, is the $\mathbb{C}(u)$-algebra that is generated by the elements $h_1, \ldots, h_{n-1}$ which are subject to the following relations:

$$\begin{aligned} h_i h_j &= h_j h_i \quad \text{for all} \quad |i-j| > 1 \\ h_i h_j h_i &= h_j h_i h_j \quad \text{for all} \quad |i-j| = 1 \\ h_i^2 &= u + (u-1)h_i \\ h_{i,i+1} &= 0, \end{aligned}$$

where $h_{i,j} := 1 + h_i + h_j + h_i h_j + h_j h_i + h_i h_j h_i$. Notice that the first three relations are the defining relations of the Iwahori–Hecke algebra, $\mathrm{H}_n(u)$, which is defined as the quotient of the algebra $\mathbb{C}(u)B_n$ over the two-sided ideal that is generated by the quadratic relations mentioned above. Thus, with this presentation, the algebra $\mathrm{TL}_n(u)$ can be considered as the quotient of $\mathrm{H}_n(u)$ over the two-sided ideal that is generated by the elements $h_{i,i+1} \in \mathrm{H}_n(u)$. It is not difficult to see that the defining ideal of $\mathrm{TL}_n(u)$ is principal and that is generated by the element $h_{1,2}$.

The algebra $\mathrm{H}_n(u)$ supports a unique Markov trace, the Ocneanu trace τ with parameter ζ [16, Theorem 5.1]. By normalizing and rescaling τ according to the braid equivalence, one obtains the *Homlypt polynomial* [16, Proposition 6.2], [10, 29]. Further, the trace τ factors through to the quotient algebra $\mathrm{TL}_n(u)$. The necessary and sufficient conditions for the factoring of τ provide a specialization for the trace parameter ζ which, in turn, gives rise to the *Jones polynomial* [16]:

$$V(u)(\widehat{\alpha}) = \left(-\frac{1+u}{\sqrt{u}}\right)^{n-1} \left(\sqrt{u}\right)^{\varepsilon(\alpha)} \tau(\pi(\alpha)),$$

where: $\alpha \in \cup_\infty B_n$, π is the natural epimorphism of $\mathbb{C}(u)B_n$ on $\mathrm{TL}_n(u)$ that sends the braid generator σ_i to h_i and $\varepsilon(\alpha)$ is the algebraic sum of the exponents of the σ_i's in α.

2.2.3 The Yokonuma–Hecke Algebra

The *Yokonuma–Hecke algebra* $\mathrm{Y}_{d,n}(u)$ [30] is defined as the quotient of the group algebra $\mathbb{C}(u)\mathscr{F}_{d,n}$ over the two-sided ideal generated by the elements:

$$\sigma_i^2 - 1 - (u-1)e_i - (u-1)e_i\sigma_i \quad \text{for all } i,$$

where $e_i := \frac{1}{d}\sum_{s=0}^{d-1} t_i^s t_{i+1}^{d-s}$, for $i = 1, \ldots, n-1$. The elements e_i in $\mathrm{Y}_{d,n}(q)$ are idempotents [18]. The generators of the ideal give rise to the following quadratic relations in $\mathrm{Y}_{d,n}(q)$:

$$g_i^2 = 1 + (u-1)e_i + (u-1)e_i g_i, \tag{2.1}$$

where g_i corresponds to σ_i. Moreover, (2.1) implies that the elements g_i are invertible with $g_i^{-1} = g_i - (u^{-1}-1)e_i + (u^{-1}-1)e_i g_i$, $1 \leq i \leq n-1$. The t_i's are called the *framing generators*, while the g_i's are called the *braiding generators* of $\mathrm{Y}_{d,n}(q)$. By its construction, the Yokonuma–Hecke algebra is considered as *the framization of the Iwahori–Hecke algebra*. Regarding its algebraic properties, the algebra $\mathrm{Y}_{d,n}(u)$ has the following standard linear basis [18]:

$$\{t_1^{a_1} \ldots t_n^{a_n} w \,|\, a_i \in \mathbb{Z}/d\mathbb{Z},\ w \in \mathscr{B}_{\mathrm{H}_n}\},$$

where $\mathscr{B}_{\mathrm{H}_n}$ is the standard basis of $\mathrm{H}_n(u)$. A simple counting argument implies that the dimension of the algebra $\mathrm{Y}_{d,n}(u)$ is equal to $n!\, d^n$. Further, the irreducible representations of $Y_{d,n}(u)$ over $\mathbb{C}(u)$, are parametrised by the d-partitions of n [4, Theorem 1].

One of the most important results regarding the Yokonuma–Hecke algebra lies in [18] where Juyumaya showed that $\mathrm{Y}_{d,n}(u)$ supports the following unique linear Markov trace function:

$$\mathrm{tr}_d : \cup_{n=1}^{\infty} \mathrm{Y}_{d,n}(u) \longrightarrow \mathbb{C}(u)[z, x_1, \ldots, x_{d-1}],$$

where $z, x_1, \ldots, x_{d-1}$ are indeterminates. The trace tr_d can be defined inductively on n by the following rules [18, Theorem 12]:

$$\begin{aligned} \mathrm{tr}_d(ab) &= \mathrm{tr}_d(ba) \\ \mathrm{tr}_d(1) &= 1 \\ \mathrm{tr}_d(ag_n) &= z\,\mathrm{tr}_d(a) \\ \mathrm{tr}_d(at_{n+1}^s) &= x_s \mathrm{tr}_d(a)\ (s = 1, \ldots, d-1), \end{aligned}$$

where $a, b \in \mathrm{Y}_{d,n}(u)$. Using the rules of tr_d and setting $x_0 := 1$, one deduces that $\mathrm{tr}_d(e_i)$ takes the same value for all i, indeed: $E := \mathrm{tr}_d(e_i) = \frac{1}{d}\sum_{s=0}^{d-1} x_s x_{d-s}$.

In order to define framed and classical link invariants via the trace tr_d, one should re-scale tr_d according to the framed braid equivalence [27]. Unfortunately, the trace tr_d is the only known trace that does not re-scale directly [24]. The E-*system* is the following system of non-linear equations

$$\sum_{s=0}^{d-1} x_{m+s}x_{d-s} = x_m \sum_{s=0}^{d-1} x_s x_{d-s} \qquad (1 \leq m \leq d-1),$$

that was introduced in order to find the necessary and sufficient conditions that needed to be applied on the parameters x_i of tr so that the re-scaling of tr_d would be possible [24]. We say that the $(d-1)$-tuple of complex numbers $(\mathrm{x}_1, \ldots, \mathrm{x}_{d-1})$ satisfies the E-*condition* if $\mathrm{x}_1, \ldots, \mathrm{x}_{d-1}$ are solutions of the E-system. The full set of solutions of the E-system is given by Paul Gérardin [24, Appendix] using tools of harmonic analysis on finite groups. More precisely, he interpreted the solution $(\mathrm{x}_1, \ldots, \mathrm{x}_d)$ of the E-system, as the complex function $x : \mathbb{Z}/d\mathbb{Z} \to \mathbb{C}$ that sends $k \mapsto \mathrm{x}_k$, $k \neq 0$ and $0 \mapsto 1$. Let now χ_m be the character of the group $\mathbb{Z}/d\mathbb{Z}$ and let $\mathbf{i}_m := \sum_{s=0}^{d-1} \chi_m(s)t^s$, for $m \in \mathbb{Z}/d\mathbb{Z} \in \mathbb{C}[\mathbb{Z}/d\mathbb{Z}]$. We then have that the solutions of the E-system are of the following form:

$$\mathrm{x}_s = \frac{1}{|D|} \sum_{m \in D} \mathbf{i}_m(s), \quad 1 \leq s \leq d-1,$$

where D is a non-empty subset of $\mathbb{Z}/d\mathbb{Z}$. Hence, the solutions of the E-system are parametrized by the non-empty subsets of $\mathbb{Z}/d\mathbb{Z}$. Two obvious solutions of the E-system are: the all-zero solution, that is $x_i = 0$, for all i, and when the x_i's are specialized to the dth roots of unity. For the rest of the paper we fix $X_D = \{\mathrm{x}_1, \ldots, \mathrm{x}_{d-1}\}$ to be a solution of the E-system parametrized by the non-empty subset D of $\mathbb{Z}/d\mathbb{Z}$. If we specialize the trace parameters x_i of tr_d to the values x_i we obtain the *specialized trace* $\mathrm{tr}_{d,D}$ with parameter z [5, 6].

By normalizing and re-scaling the specialized trace $\mathrm{tr}_{d,D}$, invariants for *framed links* are obtained [24]:

$$\Gamma_{d,D}(w,u)(\widehat{\alpha}) = \left(-\frac{(1-wu)|D|}{\sqrt{w}(1-u)}\right)^{n-1} \left(\sqrt{w}\right)^{\varepsilon(\alpha)} \mathrm{tr}_{d,D}(\gamma(\alpha)), \tag{2.2}$$

where: $w = \frac{z+(1-u)}{uz|D|}$ is the re-scaling factor, γ is the natural epimorphism of the framed braid group algebra $\mathbb{C}(u)\mathscr{F}_n$ on the algebra $\mathrm{Y}_{d,n}(u)$, and $\alpha \in \cup_\infty \mathscr{F}_n$. Further, by restricting the invariants $\Gamma_{d,D}(w,u)$ to *classical links*, seen as framed links with all framings zero, in [21] invariants of classical oriented links $\Delta_{d,D}(w,u)$ are obtained. In [6] it was proved that for generic values of the parameters u, z the invariants $\Delta_{d,D}(w,u)$ do not coincide with the HOMFLYPT polynomial except in the trivial cases $u = 1$ and $E_D = 1$.

2.3 The Three Possible Candidates

In this section we will present all results in the literature regarding the three possible quotient algebras that can be considered as candidates for the framization of the Temperley–Lieb algebra. In what follows, we will give the definitions and dimensions for each quotient algebra, describe their linear bases and representation theory and discuss the necessary and sufficient conditions so that the trace tr_d passes to each one of the quotient algebras. Finally, we will present the invariants for framed and classical links that are derived from each algebra.

2.3.1 Motivation Behind the Construction

Following the construction of the classical Temperley–Lieb algebra we would like to introduce an analogue of $\mathrm{TL}_n(u)$ in the context of framed knot algebras. Namely, to define a quotient of $\mathrm{Y}_{d,n}(u)$ over a two-sided ideal that is constructed from an appropriately chosen subgroup of the underlying group $C_{d,n} := (\mathbb{Z}/d\mathbb{Z})^n \rtimes S_n$ of $\mathrm{Y}_{d,n}(u)$. At this point two such subgroups emerge naturally. The first possibility is to consider the subgroups $\langle s_i, s_{i+1}\rangle$ of $C_{d,n}$ that are also related to the defining ideal of $\mathrm{TL}_n(u)$. The second possibility is to let the framing generators t_i be involved in the generating set of such a subgroup and consider the following subgroup of $C_{d,n}$:

$$C^i_{d,n} := \langle t_i, t_{i+1}, t_{i+2}\rangle \rtimes \langle s_i, s_{i+1}\rangle \quad \text{for all } i.$$

Therefore we can define at least two types of algebras which could be considered as analogues of the Temperley–Lieb algebras in the context of knot algebras with framing. The algebra that corresponds to the first possibility is the *Yokonuma-Temperley–Lieb algebra*, denoted by $\mathrm{YTL}_{d,n}(u)$, while the second is the *Complex Reflection Temperley–Lieb algebra*, $\mathrm{CTL}_{d,n}(u)$.

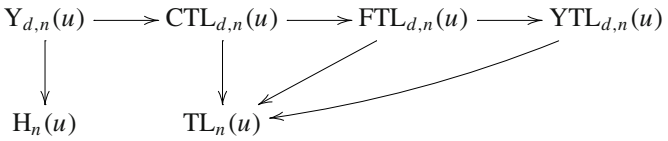
As mentioned in (2.2), new two-variable invariants for oriented framed knots and links are defined through the trace tr_d on the Yokonuma–Hecke algebra by imposing the E-system on the parameters $x_1, \ldots, x_{d-1}$ [24]. Hence, we expect that the framization of the Temperley–Lieb algebra will allow us to define one-variable specializations of the invariants derived from $\mathrm{Y}_{d,n}(u)$. Unfortunately, both quotients above are not satisfactory for this purpose. In the case of $\mathrm{YTL}_{d,n}(u)$, very strong conditions on the trace parameters must be applied in order for tr_d to pass through to the quotient algebra. Namely, the trace parameters x_i must be dth roots of unity, giving rise to obvious, special solutions of the E-system, which imply topologically loss of the framing information. However, the original Jones polynomial can be recovered from this quotient algebra. In the case of $\mathrm{CTL}_{d,n}(u)$, the quotient algebra is large enough so that the necessary and sufficient conditions such that tr_d passes to $\mathrm{CTL}_{d,n}(u)$ are, contrary to the case of $\mathrm{YTL}_{d,n}(u)$, too relaxed, especially on the trace

parameters x_i. So, in order to obtain knot invariants we would still need to impose the E-system on the trace parameters $x_1, \ldots, x_{d-1}$ as in the case of $\mathrm{Y}_{d,n}(u)$.

The discussion above indicated that the desired framization of the Temperley–Lieb algebra, for our topological purposes, could be an intermediate algebra between the quotient algebras $\mathrm{YTL}_{d,n}(u)$ and $\mathrm{CTL}_{d,n}(u)$. One may achieve this, by using for the defining ideal an intermediate subgroup that lies between $\langle s_i, s_{i+1}\rangle$ and $C^i_{d,n}$. More precisely, we define this framization as a quotient of the Yokonuma–Hecke algebra over an ideal that is constructed from the following subgroup of $C_{d,n}$:

$$H^i_{d,n} := \langle t_i t_{i+1}^{-1}, t_{i+1} t_{i+2}^{-1}\rangle \rtimes \langle s_i, s_{i+1}\rangle \quad \text{for all } i.$$

Thus, one obtains the so-called *Framization of the Temperley–Lieb algebra*, $\mathrm{FTL}_{d,n}(u)$. The relation between the three quotient algebras is given by the following commutative diagram of epimorphisms [13, Proposition 3]:



The Yokonuma-Temperley–Lieb algebra and its derived invariants were introduced and studied in [12], while its representation theory was studied in [7]. The algebras $\mathrm{FTL}_{d,n}(u)$, $\mathrm{CTL}_{d,n}(u)$ and their corresponding invariants were introduced in [13] and were further studied in [8, 14, 15].

2.3.2 The Yokonuma-Temperley–Lieb Algebra

For $n \geq 3$, the *Yokonuma-Temperley–Lieb algebra*, denoted by $\mathrm{YTL}_{d,n}(u)$, is defined as the quotient of $\mathrm{Y}_{d,n}(u)$ over the two-sided ideal that is generated by the elements:

$$g_{i,i+1} := 1 + g_i + g_{i+1} + g_i g_{i+1} + g_{i+1} g_i + g_i g_{i+1} g_i. \tag{2.3}$$

It is a straightforward computation to show that the defining ideal of $\mathrm{YTL}_{d,n}(u)$ is principal and is generated by the element $g_{1,2}$ [12, Lemma 4]. Thus, the algebra $\mathrm{YTL}_{d,n}(u)$ can be considered as the $\mathbb{C}(u)$-algebra that is generated by the elements $t_1, \ldots, t_n, g_1 \ldots, g_{n-1}$ that are subject to the defining relations of $\mathrm{Y}_{d,n}(u)$ and the relation $g_{1,2} = 0$ [12, Corollary 1]. Note also that for $d = 1$ the algebra $\mathrm{YTL}_{1,n}(u)$ coincides with $\mathrm{TL}_n(u)$.

Every word in the algebra $\mathrm{YTL}_{d,n}(u)$ inherits the splitting property from $\mathrm{Y}_{d,n}(u)$. For each fixed element in the braiding part, a set of linear dependency relations among the framing parts can be described which, in turn, lead to the extraction of a linear

basis for $\mathrm{YTL}_{d,n}(u)$ [7]. Using this technique, Chlouveraki and Pouchin proved in [7] that, for $n \geq 3$, the following set is a linear basis for $\mathrm{YTL}_{d,n}(u)$:

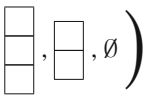
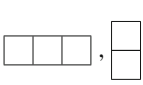
$$S_{d,n} = \left\{ t_1^{r_1} \dots t_n^{r_n} w \mid w \in \mathscr{B}_{\mathrm{TL}},\ (r_1, \dots, r_n) \in \mathscr{E}_{d,n}(w) \right\},$$

where $\mathscr{B}_{\mathrm{TL}}$ is the linear basis of the classical Temperley–Lieb algebra as computed by Jones in [16] and $\mathscr{E}_{d,n}(w)$ is a subset of $\{0, \dots, d-1\}^n$ that describes the exponents of the t_i's that correspond to the fixed braid word $w \in \mathrm{YTL}_{d,n}(u)$. For an explicit description of the set $\mathscr{E}_{d,n}(w)$, the reader is encouraged to consider [7, Propositions 9 and 11]. Subsequently, the dimension of the Yokonuma-Temperley–Lieb algebra can be computed, which is equal to:

$$\dim(\mathrm{YTL}_{d,n}(u)) = dc_n + \frac{d(d-1)}{2} \sum_{k=1}^{n-1} \binom{n}{k}^2,$$

where c_n is the nth Catalan number [7, Proposition 4].

By standard results in representation theory we have that the irreducible representations of $\mathrm{YTL}_{d,n}(u)$ are in bijection with those irreducible representations of $\mathrm{Y}_{d,n}(u)$ that respect the defining relation of $\mathrm{YTL}_{d,n}(u)$, which is $g_{1,2} = 0$. Specifically, the irreducible representations of $\mathrm{YTL}_{d,n}(u)$ are those representations of $\mathrm{Y}_{d,n}(u)$ who have at most two columns in total in the Young diagram of the parametrizing d-partition of n [7, Theorem 1]. In the following example, the first 3-partition of 5 parametrizes an irreducible representation of $\mathrm{YTL}_{3,5}(u)$ while the second one does not correspond to an irreducible representation of $\mathrm{YTL}_{3,5}(u)$:

$i.$  $ii.$ 

As mentioned in the introduction, the motivation behind the definition of a Temperley–Lieb type quotient from the Yokonuma–Hecke algebra was the construction of polynomial invariants for framed knots and links via the use of the trace tr_d of $\mathrm{Y}_{d,n}(u)$. Thus, one of the biggest challenges regarding the study of the algebra $\mathrm{YTL}_{d,n}(u)$ was the determination of the necessary and sufficient conditions for tr_d to factor through to the quotient algebra. By employing the methods that P. Gérardin used to describe the full set of solutions of the E-system [24, Appendix], the author together with Juyumaya, Kontogeorgis and Lambropoulou proved that following:

Theorem 2.1 ([12, Theorem 6]) *The trace* tr_d *passes to the quotient algebra* $\mathrm{YTL}_{d,n}(u)$ *if and only if the* x_i*'s are solutions of the* E*-system and one of the two cases holds:*

(i) the x_ℓ*'s are* $dt h$ *roots of unity and* $z = -\frac{1}{u+1}$ *or* $z = -1$,
(ii) the x_ℓ*'s are the solutions of the* E*-system that are parametrized by the set* $D = \{m_1, m_2 \mid 0 \leq m_1, m_2 \leq d-1 \text{ and } m_1 \neq m_2\}$ *and they are expressed as:*

$$x_\ell = \frac{1}{2}\left(\chi_{m_1}(t^\ell) + \chi_{m_2}(t^\ell)\right), \quad 0 \leq \ell \leq d-1,$$

where the χ_k's denote the characters of the group $\mathbb{Z}/d\mathbb{Z}$. In this case we have that $z = -\frac{1}{2}$.

Note that in both cases the x_i's are solutions of the E-system, as required by [24], in order to proceed with defining link invariants. We do not take into consideration case (i) for $z = -1$ and case (ii), where $z = -\frac{1}{2}$, since crucial braiding information is lost and therefore they are of no topological interest [12]. The only remaining case of interest is case (i) of Theorem 2.1, where the x_ℓ's are the dth roots of unity and $z = -\frac{1}{u+1}$. This implies that $E = 1$ and $w = u$ in (2.2). So, by [6, 16], the invariant $\Delta_{d,s}(u, u)$ coincides with the Jones polynomial. For this reason, the algebra $\mathrm{YTL}_{d,n}(u)$ is discarded as a potential framization of the Temperley–Lieb algebra.

2.3.3 The Complex Reflection Temperley–Lieb Algebra

We move on now with presenting the second natural definition of a potential framization of the Temperley–Lieb algebra. For $n \geq 3$, we define the *Complex Reflection Temperley–Lieb* algebra, denoted by $\mathrm{CTL}_{d,n}(u)$, as the quotient of the algebra $\mathrm{Y}_{d,n}(u)$ over the ideal that is generated by the elements

$$c_{i,i+1} := \sum_{\alpha,\beta,\gamma \in \mathbb{Z}/d\mathbb{Z}} t_i^\alpha t_{i+1}^\beta t_{i+2}^\gamma \, g_{i,i+1}. \tag{2.4}$$

In analogy to the algebra $\mathrm{YTL}_{d,n}(u)$, the defining ideal of $\mathrm{CTL}_{d,n}(u)$ can be shown to be principal and is generated by the single element $c_{1,2}$. Further, for $d = 1$, the algebra $\mathrm{CTL}_{1,n}(u)$ coincides with the algebra $\mathrm{TL}_n(u)$. The denomination Complex Reflection Temperley–Lieb algebra has to do with the fact that the underlying group of $\mathrm{CTL}_{d,n}(u)$ is isomorphic to the complex reflection group $G(d, 1, 3)$.

The Complex Reflection Temperley–Lieb algebra is isomorphic to a direct sum of matrix algebras over tensor products of Temperley–Lieb and Iwahori–Hecke algebras [8, Theorem 5.8]. This isomorphism, which we will denote by ϕ_n, will lead to the determination of a linear basis for $\mathrm{CTL}_{d,n}(u)$. More precisely, there exists an explicit isomorphism:

$$\phi_n \; : \bigoplus_{\mu \in \mathrm{Comp}_d(n)} \mathrm{Mat}_{m_\mu}\left(\mathrm{TL}_{\mu_1}(u) \otimes \mathrm{H}_{\mu_2}(u) \otimes \ldots \otimes \mathrm{H}_{\mu_d}(u)\right) \longrightarrow \mathrm{CTL}_{d,n}(u).$$

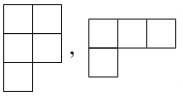
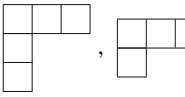
Then the following set is a linear basis for $\mathrm{CTL}_{d,n}(u)$ [8, Proposition 5.9]:

$$\left\{\phi_n\left(b_1 b_2 \ldots b_d \, \mathrm{M}_{k,l}\right) \mid b_1 \in \mathscr{B}_{\mathrm{TL}_{\mu_1}(u)}, b_i \in \mathscr{B}_{\mathrm{H}_{\mu_i}(u)} \text{ for all } i = 2, ..., d, \, 1 \leq k, l \leq m_\mu, \mu \in \mathrm{Comp}_d(n)\right\},$$

where $\mathscr{B}_{\mathrm{TL}_{\mu_1(u)}}$ is the linear basis of $\mathrm{TL}_{\mu_1(u)}$, $\mathscr{B}_{\mathrm{H}_{\mu_i}(u)}$ is the linear basis of H_{μ_i}, $\mathrm{M}_{k,l}$ is the elementary $m_\mu \times m_\mu$ matrix with 1 in position (k, l) and $\mu \in \mathrm{Comp}_d(n)$ is a d-composition of n, that is, $\mu = (\mu_1, \mu_2, \ldots, \mu_d) \in \mathbb{N}^d$ such that $\mu_1+\mu_2+\cdots+\mu_d = n$. Counting the elements of the above basis one can derive the dimension of the algebra $\mathrm{CTL}_{d,n}(u)$ [8, Theorem 5.5]. Indeed, if $c_k := \frac{1}{k+1}\binom{2k}{k}$ is the kth Catalan number, we have that:

$$\dim_{\mathbb{C}(u)}\mathrm{CTL}_{d,n}(u) = \sum_{\mu \in \mathrm{Comp}_d(n)} \left(\frac{n!}{\mu_1!\,\mu_2!\,\ldots\,\mu_d!}\right)^2 c_{\mu_1}\,\mu_2!\,\ldots\,\mu_d!$$

Let now $\lambda = (\lambda^{(1)}, \ldots, \lambda^{(d)})$ a d-partition of n. The irreducible representations of $\mathrm{CTL}_{d,n}(u)$ are those irreducible representations of $\mathrm{Y}_{d,n}(u)$ whose Young diagram of $\lambda^{(1)}$ has at most two columns [8, Theorem 5.3]. For instance, in the example given below, the first 2-Young diagram corresponds to an irreducible representation of $\mathrm{CTL}_{2,9}(u)$ while the second one does not:

$i.$  $ii.$  .

Next we present the necessary and sufficient conditions for the trace tr_d to factor through to the quotient algebra $\mathrm{CTL}_{d,n}(u)$. We have the following:

Theorem 2.2 ([13, Theorem 7]) *The trace* tr_d *passes to the quotient algebra* $\mathrm{CTL}_{d,n}(u)$ *if and only if the parameter* z *and the* x_i*'s are related through the equation:*

$$(u+1)z^2 \sum_{k\in\mathbb{Z}/d\mathbb{Z}} x_k + (u+2)z \sum_{k\in\mathbb{Z}/d\mathbb{Z}} E^{(k)} + \sum_{k\in\mathbb{Z}/d\mathbb{Z}} \mathrm{tr}(e_1^{(k)} e_2) = 0. \tag{2.5}$$

Notice now that the conditions of Theorem 2.2 do not include any solutions of the E-system. Thus, in order to obtain any well defined invariant from the algebras $\mathrm{CTL}_{d,n}(u)$ one has to impose the E-condition on the trace parameters x_i. Even by doing so, $\mathrm{CTL}_{d,n}(u)$ does not deliver any new invariants for framed or classical oriented knots and links. We have the following:

Proposition 2.1 ([13, Proposition 10]) *Let* X_D *be a solution of the* E*-system parametrized by the subset* D *of* $\mathbb{Z}/d\mathbb{Z}$*. The invariants derived from the algebra* $\mathrm{CTL}_{d,n}(u)$*:*

1. *if* $0 \in D$*, they coincide with the invariants derived from the algebra* $\mathrm{FTL}_{d,n}(u)$*,*
2. *if* $0 \notin D$*, they coincide with the invariants derived from the algebra* $\mathrm{Y}_{d,n}(u)$*.*

The above constitute the reasons for which the Complex Reflection Temperley–Lieb algebra is discarded as a potential candidate for the framization of the Temperley–Lieb algebra.

2.3.4 The Framization of the Temperley–Lieb algebra

For $n \geq 3$, the *Framization of the Temperley–Lieb algebra*, denoted by $\mathrm{FTL}_{d,n}(u)$, is defined as the quotient $\mathrm{Y}_{d,n}(u)$ over the two-sided ideal that is generated by the elements

$$r_{i,i+1} := \sum_{\alpha+\beta+\gamma=0} t_i^{\alpha} t_{i+1}^{\beta} t_{i+2}^{\gamma}\, g_{i,i+1}. \tag{2.6}$$

In analogy to the case of the other two quotient algebras, for $d = 1$ the algebra $\mathrm{FTL}_{1,n}(u)$ coincides with $\mathrm{TL}_n(u)$. Additionally, the defining ideal of $\mathrm{FTL}_{d,n}(u)$ is principal and is generated by the element $r_{1,2}$. Thus, in terms of generators and relations, $\mathrm{FTL}_{d,n}(u)$ is the $\mathbb{C}(u)$-algebra generated by the set $\{t_1, \ldots, t_n, g_1, \ldots, g_{n-1}\}$ whose elements are subject to the defining relations of $\mathrm{Y}_{d,n}(u)$ and the relation $r_{1,2} = 0$.

As in the case of $\mathrm{CTL}_{d,n}(u)$, the determination of a linear basis for the Framization of the Temperley–Lieb algebra will emerge from an isomorphism theorem for $\mathrm{FTL}_{d,n}(u)$. More precisely, the quotient algebra $\mathrm{FTL}_{d,n}(u)$ is isomorphic to a direct sum of matrix algebras over tensor products of Temperley–Lieb algebras [8, Theorem 4.3]. There exists an explicit isomorphism of $\mathbb{C}(u)$-algebras:

$$\widetilde{\phi}_n \;:\; \bigoplus_{\mu \in \mathrm{Comp}_d(n)} \mathrm{Mat}_{m_\mu}\left(\mathrm{TL}_{\mu_1}(u) \otimes \ldots \otimes \mathrm{TL}_{\mu_d}(u)\right) \longrightarrow \mathrm{FTL}_{d,n}(u),$$

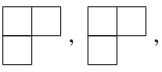
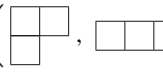
then the following set is a linear basis for the algebra $\mathrm{FTL}_{d,n}(u)$:

$$\left\{\widetilde{\phi}_n(b_1 \ldots b_d\, M_{k,l}) \mid b_i \in \mathscr{B}_{\mathrm{TL}_{\mu_i}(q)} \text{ for all } i = 1, \ldots d,\, 1 \leq k, l \leq m_\mu,\, \mu \in \mathrm{Comp}_d(n)\right\}.$$

By using a counting argument one can derive the dimension of the algebra $\mathrm{FTL}_{d,n}(u)$, which is equal to [8, Theorem 3.11]:

$$\dim_{\mathbb{C}(u)} \mathrm{FTL}_{d,n}(u) = \sum_{\mu \in \mathrm{Comp}_d(n)} \left(\frac{n!}{\mu_1!\, \mu_2!\, \ldots \mu_d!}\right)^2 c_{\mu_1}\, c_{\mu_2}\, \ldots\, c_{\mu_d}. \tag{2.7}$$

The irreducible representations of $\mathrm{FTL}_{d,n}(u)$ are those irreducible representations of $\mathrm{Y}_{d,n}(u)$ whose Young diagram of $\lambda^{(i)}$ has at most two columns, for $i = 1, 2, \ldots, d$. As in the previous examples, the first of the following 3-Young diagrams describes an irreducible representation of $\mathrm{FTL}_{3,7}(u)$ while the second does not:

$i.$  $\quad ii.$  .

We move on now to the necessary and sufficient conditions so that tr_d factors through to $\mathrm{FTL}_{d,n}(u)$.

Theorem 2.3 ([13, Theorem 6]) *The trace* tr *passes to* $\mathrm{FTL}_{d,n}(u)$ *if and only if the parameters of the trace* tr *satisfy:*

$$x_k = -z\left(\sum_{m\in \mathrm{Sup}_1}\chi_m(t^k) + (u+1)\sum_{m\in \mathrm{Sup}_2}\chi_m(t^k)\right) \quad \textit{and} \quad z = -\frac{1}{|\mathrm{Sup}_1| + (u+1)|\mathrm{Sup}_2|},$$

where χ_m *are the characters of the group* $\mathbb{Z}/d\mathbb{Z}$, $\mathrm{Sup}_1 \sqcup \mathrm{Sup}_2$ *(disjoint union) is the support of the Fourier transform of* x*, and* x *is the complex function on* $\mathbb{Z}/d\mathbb{Z}$*, that maps* 0 *to* 1 *and* k *to the trace parameter* x_k.

The intrinsic difference with the other two quotient algebras lies in the fact that the necessary and sufficient conditions of Theorem 2.3 include all solutions of the E-system. This observation is the main reason that led to the consideration of the quotient algebra $\mathrm{FTL}_{d,n}(u)$ as the most natural non-trivial analogue of the Temperley–Lieb algebra in the context of framization of knot algebras. If one lets either Sup_1 or Sup_2 to be the empty set, then the trace parameters x_k comprise a solution of the E-system. In this context, if Sup_1 is the empty set then $z = -\frac{1}{(u+1)|\mathrm{Sup}_2|}$ while if Sup_2 is the empty set then $z = -1/|\mathrm{Sup}_1|$ [13, Corollary 3]. Since for defining invariants for oriented (framed) knots and links only the cardinal $|D|$ of the parametrizing set D of a solution is needed, the solutions mentioned above cover all the possibilities. We do not take into consideration the case where $\mathrm{Sup}_2 = \emptyset$ and $z = -1/|\mathrm{Sup}_1|$ since important topological information is lost and thus basic pairs of knots are not distinguished [13, Remark 7]. For the remaining case, let X_D be a solution of the E-system, parametrized by the non-empty subset $D = \mathrm{Sup}_2$ of $\mathbb{Z}/d\mathbb{Z}$ and let $z = -\frac{1}{(u+1)|D|}$. We obtain from $\Gamma_{d,D}(w,u)$ the following new 1-variable framed link invariants:

$$\Gamma_{d,D}(u,u)(\widehat{\alpha}) := \left(-\tfrac{(1+u)|D|}{\sqrt{u}}\right)^{n-1}\left(\sqrt{u}\right)^{\varepsilon(\alpha)}\mathrm{tr}_{d,D}\left(\gamma(\alpha)\right), \tag{2.8}$$

for any $\alpha \in \cup_\infty \mathscr{F}_n$. Further, in analogy to the invariants of $\Gamma_{d,D}(w,u)$, if we restrict to framed links with all framings zero, we obtain from $\Gamma_{d,D}(u,u)$ new 1-variable invariants of classical links $\Delta_{d,D}(u,u)$. Additionally, for $d = 1$ the invariant $\Gamma_{d,D}(u,u)$ coincides with the Jones polynomial.

2.4 Comparisons and Generalizations

In this section we will present the comparisons of the invariants Θ_d and θ_d to the HOMFLYPT and the Jones polynomials respectively, and we will give generalizations for both of them.

2.4.1 The Invariants Θ_d and Their Generalization

In a recent development [5] it was proved that the classical link invariants derived from the Yokonuma–Hecke algebra are *not topologically equivalent to the HOMFLYPT polynomial on links* while they are topologically equivalent to the HOMFLYPT on knots. This was achieved by considering a different presentation for the algebra $\mathrm{Y}_{d,n}$ with parameter q instead of u and a different quadratic relation. More precisely, the algebra $\mathrm{Y}_{d,n}(q)$ is defined as the $\mathbb{C}(q)$-algebra that is generated by the elements $g'_1, \ldots, g'_{n-1}, t_1, \ldots, t_n$, which satisfy all relations of $\mathrm{Y}_{d,n}(u)$ except for the quadratic relation that is replaced with the following:

$$(g'_i)^2 = 1 + (q - q^{-1})e_i g'_i. \tag{2.9}$$

One can obtain this presentation from the one given in Sect. 2.2.3 by taking $u = q^2$ and

$$g_i = g'_i + (q-1)e_i g'_i \quad \text{(or, equivalently, } g'_i = g_i + (q^{-1} - 1)e_i g_i).$$

Thus, the following invariants of classical links were derived [5]:

$$\Theta_d(q, \lambda_d)(\widehat{\alpha}) = \left(\frac{1 - \lambda_d}{\sqrt{\lambda_d}(q - q^{-1})E_D}\right)^{n-1} \sqrt{\lambda_d}^{\varepsilon(\alpha)} \mathrm{tr}_{d,D}(\delta(\alpha)), \tag{2.10}$$

where $\alpha \in \cup_\infty B_n$, $E_D = 1/d$, $\varepsilon(a)$ is as in (2.2), δ is the natural homomorphism $\mathbb{C}(q)B_n \to \mathrm{Y}_{d,n}(q)$ and $\lambda_d = \frac{z' - (q - q^{-1})E_D}{z'}$ is the re-scaling factor for the trace $\mathrm{tr}_{d,D}$.

The invariants Θ_d depend only on $d \in \mathbb{N}$, that is, the cardinal of the subset D that parametrizes the solution of the E-system [5, Proposition 4.6]. Furthermore, the choice of the new presentation for $\mathrm{Y}_{d,n}$ revealed that the invariants Θ_d satisfy the HOMFLYPT skein relation on crossings between different components of a link L [5, Proposition 6.8]. Using this, one can prove that the invariants Θ_d distinguish more pairs of HOMFLYPT equivalent pair of non-isotopic oriented classical links [5, Sect. 7.2] and thus that Θ_d are *not topologically equivalent to the HOMFLYPT polynomial on links* [5, Theorem 7.3].

In [5] it has been shown skein-theoretically that the invariants for classical links Θ_d generalize to a new 3-variable invariant $\Theta(q, \lambda, E)$ for classical oriented links that can be defined uniquely by the following two rules:

1. On crossings between different components of an oriented classical link L the skein relation of the HOMFLYPT polynomial holds:

$$\frac{1}{\sqrt{\lambda_D}}\,\Theta(L_+) - \sqrt{\lambda_D}\,\Theta(L_-) = (q - q^{-1})\,\Theta(L_0),$$

   where L_+, L_- and L_0 is a Conway triple.
2. For a disjoint union of $\mathscr{K} = \sqcup_{i=1}^{r} K_i$ of r knots, with $r > 1$, it holds that:

$$\Theta(\mathscr{K}) = E^{1-r} \prod_{i=1}^{r} P(K_i),$$

where $P(K_i)$ is the value of the HOMFLYPT polynomial on K_i.

Algebraically, the well-definedness of the invariant Θ can be proved by using *the algebra of braids and ties*, $\mathscr{E}_n(q)$ [1]. The algebra $\mathscr{E}_n(q)$ supports a unique Markov trace ρ that gives rise to a 3-variable invariant for tied links $\overline{\Theta}(q, \lambda, E)$ which, in turn, restricts to an invariant of classical oriented links $\Theta(q, \lambda, E)$ [2, 3]. Alternatively, one can use the fact that, for $d \geq n$, $\mathscr{E}_n(q)$ is isomorphic to the subalgebra $\mathrm{Y}_{d,n}^{(\mathrm{br})}(q)$ of $\mathrm{Y}_{d,n}(q)$ that is generated only by the g_i's [9]. Note now that when computing the specialized trace $\mathrm{tr}_{d,D}$ of a braid word in B_n, the framing generators appear only when applying the quadratic or the inverse relation and only in the form of the idempotents e_i. In this case and by the E-condition, the last rule of the specialized trace: $\mathrm{tr}_{d,D}(at_{n+1}^s) = \mathrm{x}_s \mathrm{tr}_{d,D}(a)$, for $s = 1, \ldots, d-1$, can be substituted by the following two rules [5, Theorem 4.3]:

$$\mathrm{tr}_{d,D}(ae_n) = E_D \, \mathrm{tr}_{d,D}(a) \quad \text{and} \quad \mathrm{tr}_{d,D}(ae_n g_n) = z \, \mathrm{tr}_{d,D}(a),$$

where D is the non-empty subset of $\mathbb{Z}/d\mathbb{Z}$ that parametrizes a solution of the E-system. Consequently, if E_D is considered as an indeterminate, the specialized trace $\mathrm{tr}_{d,D}$ on $\mathrm{Y}_{d,n}^{(\mathrm{br})}(q)$ is well-defined since it coincides with the trace ρ on $\mathscr{E}_n(q)$ and, therefore, the invariant Θ can be constructed directly through $\mathrm{Y}_{d,n}^{(\mathrm{br})}(q)$ [5, Remark 4.18]. Conversely, one can recover the invariants Θ_d from Θ by specializing $E = 1/d$, $d \in \mathbb{N}$.

A self-contained diagrammatic proof for the well-definedness of the invariant Θ has been given in [26]. The invariant Θ distinguishes more pairs of non isotopic oriented links than the HOMFLYPT polynomial and thus it is stronger than the HOMFLYPT. We note also that, Θ is not topologically equivalent to the HOMFLYPT or the Kauffman polynomials.

Finally, it is worth noting that the invariant Θ can be described by the following closed combinatorial formula, namely:

Theorem 2.4 ([5, Appendix]) *Let L be an oriented link with n components, then:*

$$\Theta(q, \lambda, E)(L) = \sum_{k=1}^{m} \mu^{k-1} E_k \sum_{\pi} \lambda^{\nu(\pi)} P(\pi L), \tag{2.11}$$

where the second summation is over all partitions of π of the components of L into k (unordered) subsets and $P(\pi L)$ denotes the product of the HOMFLYPT polynomials of the k sublinks of L defined by π. Furthermore, $\nu(\pi)$ is the sum of all linking numbers of pairs of components of L that are distinct sets of π, $E_k = (E^{-1} - 1)(E^{-1} - 2) \ldots (E^{-1} - k + 1)$, with $E_1 = 1$ and $\mu = \frac{\lambda^{-1/2} - \lambda^{1/2}}{q - q^{-1}}$

2.4.2 The Invariants θ_d and Their Generalization

By adjusting the algebra $\mathrm{FTL}_{d,n}$ to the presentation that has parameter q and involves the quadratic relation (2.9), one can compare the derived invariants for classical oriented links to the Jones polynomial. In this context, the generator of the defining principal ideal of $\mathrm{FTL}_{d,n}$ is transformed to the following element of $\mathrm{Y}_{d,n}(q)$:

$$e_1e_2\Big(1+q(g_1'+g_2')+q^2(g_1'g_2'+g_2'g_1')+q^3g_1'g_2'g_1'\Big).$$

Note that the E-system and its solutions remain unaffected by this change of presentations. The values for the trace parameters z, however, are transformed to the following:

$$z'=-\frac{q^{-1}E_D}{q^2+1} \quad \text{or} \quad z'=-q^{-1}E_D.$$

The parameters z and z' are related through the equation: $z=qz'$. Again, the value $z'=-q^{-1}E_D$ is discarded. For the remaining values for z', we obtain from (2.10) the following 1-variable specialization of Θ_d:

$$\theta_d(q)(\widehat{\alpha}):=\left(-\frac{1+q^2}{qE_D}\right)^{n-1}q^{2\varepsilon(\alpha)}\mathrm{tr}_{d,D}(\delta(a))=\Theta_d(q,q^4)(\widehat{\alpha}),$$

where $\alpha\in\cup_\infty B_n$, d and E_D, $\varepsilon(a)$ and δ are as in (2.10). The invariants θ_d were proven to be topologically equivalent to the Jones polynomial on knots [13, Proposition 11], however, they are *topologically not equivalent to the Jones polynomial on links* [13, Theorem 9].

In [15] the author together with S. Lambropoulou has shown that the invariants θ_d generalize to a new 2-variable invariant θ for classical links. This generalization can be proved either algebraically or diagrammatically. Algebraically, this can be shown in two different ways. The first way is to consider the *partition Temperley–Lieb* algebra, $\mathrm{PTL}_n(q)$, which is a quotient of $\mathscr{E}_n(q)$ and determine the necessary and sufficient conditions such that the unique Markov trace ρ of $\mathscr{E}_n(q)$ factors through to $\mathrm{PTL}_n(q)$. These conditions give rise to a 2-variable invariant for classical links, $\theta(q,E)$ [15, Definition 1], that for $E=1/d$ coincides with θ_d. Alternatively, one can show that, for $d\geq n$, the subalgebra $\mathrm{FTL}_{d,n}^{(\mathrm{br})}(q)$ of $\mathrm{FTL}_{d,n}$ that is generated only by the braiding generators g_i is isomorphic to $\mathrm{PTL}_n(q)$ [14, Proposition 5]. Diagrammatically, one may consider the skein-theoretic definition of $\Theta(q,\lambda,E)$ and specialize $\lambda=q^4$. Thus, we obtain the following:

Theorem 2.5 ([15, Theorem 6]) *Let q, E be indeterminates. There exists a unique ambient isotopy invariant of classical oriented links*

$$\theta:\mathscr{L}\to\mathbb{C}[q^{\pm1},E^{\pm1}]$$

defined by the following rules:

1. *On crossings involving different components the following skein relation holds:*

$$q^{-2}\,\theta(L_+) - q^2\,\theta(L_-) = (q - q^{-1})\,\theta(L_0),$$

   where L_+, L_- and L_0 constitute a Conway triple.
2. *For a disjoint union $\mathscr{K} = \sqcup_{i=1}^{r} K_i$ of r knots, with $r > 1$, it holds that:*

$$\theta(\mathscr{K}) = E^{1-r} \prod_{i=1}^{r} V(K_i),$$

   where $V(K_i)$ is the value of the Jones polynomial on K_i.

All the properties of the invariant Θ carry through to θ [15] and so the invariant θ distinguishes the same pairs of non-isotopic oriented classical links as Θ. More precisely, in [5] six pairs of HOMFLYPT-equivalent non-isotopic oriented classical links were found to be distinguished by the invariants $\Theta(q, \lambda, E)$, which are all still distinguished by θ. Indeed we have that:

$$\theta(L11n358\{0,1\}) - \theta(L11n418\{0,0\}) = \frac{(1-E)(q-1)^5(q+1)^5(q^2+1)(q^2+q+1)(q^2-q+1)}{E\,q^{18}}$$

$$\theta(L11a467\{0,1\}) - \theta(L11a527\{0,0\}) = \frac{(1-E)(q-1)^5(q+1)^5(q^2+1)(q^2+q+1)(q^2-q+1)}{E\,q^{18}}$$

$$\theta(L11n325\{1,1\}) - \theta(L11n424\{0,0\}) = \frac{(E-1)(q-1)^5(q+1)^5(q^2+1)(q^2+q+1)(q^2-q+1)}{E\,q^{14}}$$

$$\theta(L10n79\{1,1\}) - \theta(L10n95\{1,0\}) = \frac{(E-1)(q^2-1)^3(q^8+2\,q^6+2\,q^4-1)}{E\,q^{18}}$$

$$\theta(L11a404\{1,1\}) - \theta(L11a428\{0,1\}) = \frac{(1-E)(q-1)^3(q+1)^3(q^2+1)(q^4+1)(q^6-q^4+1)}{E\,q^4}$$

$$\theta(L10n76\{1,1\}) - \theta(L11n425\{1,0\}) = \frac{(E-1)(q-1)^3(q+1)^3(q^2+1)(q^4+1)}{E\,q^{10}}.$$

The invariant $\theta(q, E)$ is not topologically equivalent to the HOMFLYPT or the Kauffman polynomials, it includes the family of invariants $\{\theta_d\}_{d\in\mathbb{N}}$ as well as the Jones polynomial and hence it is *stronger than the Jones polynomial* [15, Theorem 7].

Finally, the invariant θ can be described by a closed combinatorial formula, which is a corollary of Theorem 2.4. Indeed we have:

Corollary 2.1 *Let L be an oriented link with n components. Then:*

$$\theta(q, E)(L) = \sum_{k=1}^{m} (-1)^{k-1}(q + q^{-1})^{k-1} E_k \sum_{\pi} \lambda^{\nu(\pi)} V(\pi L),$$

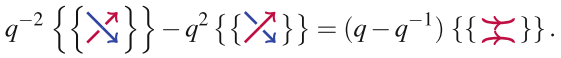
where π, $\nu(\pi)$, and E_k are as in Theorem 2.4, and $V(\pi L)$ denotes the product of the Jones polynomial of the k sublinks of L defined by π.

From Corollary 2.1 it is clear that the invariant θ depends on the orientations of the components of the link L, thus making it impossible to relate θ to the Kauffman bracket polynomial. However, as shown in [15, Theorem 7], θ can be expressed in terms of the oriented extension of the bracket polynomial. In particular, the author together with S. Lambropoulou defined in [15] the ambient isotopy link invariant $\{\{L\}\}$ of the link diagram L by the following two rules:

(1) For a disjoint union $\mathscr{K}^r := \sqcup_{i=1}^{r} K_i$, of r knots with $r \geq 1$, we have that:

$$\{\{\mathscr{K}^r\}\} := E^{1-r} \prod_{i=1}^{r} V(K_i), \tag{2.12}$$

(2) On crossings involving different components the skein relation of the Jones polynomial holds, namely:



(2.13)

Comparing (2.12) and (2.13) to Theorem 2.5, we deduce that $\{\{L\}\}$ coincides with the invariant $\theta(q, E)$.

References

1. Aicardi, F., Juyumaya, J.: An algebra involving braids and ties. Preprint ICTP IC/2000/179, Trieste (2000)
2. Aicardi, F., Juyumaya, J.: Markov trace on the algebra of braids and ties. Moscow Math. J. **16**, 397–431 (2016)
3. Aicardi, F., Juyumaya, J.: Tied links. J. Knot Theory Ramific. **25**(9), 1641001 (2016)
4. Chlouveraki, M., Poulain d'Andecy, L.: Representation theory of the Yokonuma–Hecke. Adv. Math. **259**, 134–172 (2014)
5. Chlouveraki, M., Juyumaya, J., Karvounis, K., Lambropoulou, S.: Identifying the invariants for classical knots and links from the Yokonuma–Hecke algebras, submitted for publication (2015). arXiv:1505.06666
6. Chlouveraki, M., Lambropoulou, S.: The Yokonuma–Hecke algebras and the Homflypt polynomial. J. Knot Theory and Its Ramif. **22** (2013)
7. Chlouveraki, M., Pouchin, G.: Determination of the representations and a basis for the Yokonuma-Temperley-Lieb algebra, Algebras Represent. Theory **18** (2015)
8. Chlouveraki, M., Pouchin, G.: Representation theory and an isomorphism theorem for the Framisation of the Temperley-Lieb algebra. Math. Z. **285**(3), 1357–1380 (2017)
9. Espinoza, J., Ryom-Hansen, S.: Cell structures for the Yokonuma–Hecke algebra and the algebra of braids and ties, submitted for publication (2016). arXiv:1506.00715
10. Freyd, P., Yetter, D., Hoste, J., Lickorish, W., Millett, K., Ocneanu, A.: A new polynomial invariant of knots and links. Bull. AMS **12**, 239–246 (1985)

11. Goundaroulis, D.: Framization of the Temperley-Lieb algebra and related link invariants, Ph.D. thesis, Department of Mathematics, National Technical University of Athens, 1 (2014)
12. Goundaroulis, D., Juyumaya, J., Kontogeorgis, A., Lambropoulou, S.: The Yokonuma-Temperley-Lieb algebra. Banach Center Pub. **103**, 73–95 (2014)
13. Goundaroulis, D., Juyumaya, J., Kontogeorgis, A., Lambropoulou, S.: Framization of the Temperley-Lieb algebra. Math. Res. Letters **24**(7), 299–345 (2017)
14. Goundaroulis, D., Lambropoulou, S.: Classical link invariants from the framizations of the Iwahori-Hecke algebra and the Temperley-Lieb algebra of type A. J. Knot Theory Ramific. **26**(9), 1743005 (2017)
15. Goundaroulis, D., Lambropoulou, S.: A new two-variable generalization of the Jones polynomial. Submitted for publication (2016). arXiv:1608.01812 [math.GT]
16. Jones, V.: Hecke algebra representations of braid groups and link polynomials. Ann. Math. **126**, 335–388 (1987)
17. Juyumaya, J.: Sur les nouveaux générateurs de l'algèbre de Hecke $(g, u, 1)$. J. Algebra **204**, 40–68 (1998)
18. Juyumaya, J.: Markov trace on the Yokonuma–Hecke algebra. J. Knot Theory Ramif. **13**, 25–39 (2004)
19. Juyumaya, J., Kannan, S.: Braid relations in the Yokonuma–Hecke algebra. J. Algebra **239**, 272–297 (2001)
20. Juyumaya, J., Lambropoulou, S.: p-adic framed braids. Topol. Appl. **154**, 1804–1826 (2007)
21. Juyumaya, J., Lambropoulou, S.: An adelic extension of the Jones polynomial. In: Banagl, M., Vogel, D. (eds.) The Mathematics of Knots. Contributions in the Mathematical and Computational Sciences, vol. 1, pp. 825–840. Springer (2009)
22. Juyumaya, J., Lambropoulou, S.: An invariant for singular knots. J. Knot Theory Ramif. **18**, 825–840 (2009)
23. Juyumaya, J., Lambropoulou, S.: Modular framization of the BMW algebra (2013). arXiv:1007.0092v1 [math.GT]
24. Juyumaya, J., Lambropoulou, S.: p-adic framed braids II. Adv. Math. **234**, 149–191 (2013)
25. Juyumaya, J., Lambropoulou, S.: On the framization of knot algebras. In: Kauffman, L., Manturov, V. (eds.) New Ideas in Low-dimensional Topology. Series on Knots and everything. World Scientific (2014)
26. Kauffman, L.H., Lambropoulou, S.: New invariants of links and their state sum models (submitted for publication) (2016). arXiv: 1703.03655
27. Ko, K., Smolinsky, L.: The framed braid group and 3-manifolds. Proc. AMS **115**, 541–551 (1992)
28. Marin, I.: Artin groups and the Yokonuma–Hecke algebra. Int. Math. Res. Notices, rnx007 (2017). https://doi.org/10.1093/imrn/rnx007
29. Przytycki, J.H., Traczyk, P.: Invariants of links of Conway type. Kobe J. Math. **4**, 115–139 (1987)
30. Yokonuma, T.: Sur la structure des anneux de Hecke d'un group de Chevalley fin. C.R. Acad. Sc. Paris **264**, 344–347 (1967)

Chapter 3
Representation Theory of Framisations of Knot Algebras

Maria Chlouveraki

Abstract We study the algebraic structure and the representation theory of the Yokonuma–Hecke algebra of type A, its generalisations, the affine and cyclotomic Yokonuma–Hecke algebras, and its Temperley–Lieb type quotients, the Yokonuma–Temperley–Lieb algebra, the Framisation of the Temperley–Lieb algebra and the Complex Reflection Temperley–Lieb algebra.

3.1 Introduction

A knot algebra is an algebra obtained as a quotient of the group algebra of a braid group and endowed with a Markov trace. Using Jones's method [2, 3] of normalising and re-scaling the trace according to the braid equivalence, these algebras can be used for the definition of knot invariants. Some known examples are the Temperley–Lieb algebra (Jones polynomial), the Iwahori–Hecke algebra of type A (Homflypt polynomial), the Iwahori–Hecke algebra of type B (Geck–Lambropoulou invariants), the Ariki–Koike algebras and the affine Hecke algebra of type A (Lambropoulou invariants), the singular Hecke algebra (Kauffman–Vogel and Paris–Rabenda invariants), the BMW algebra (Kauffman polynomial) and the Rook algebra (Alexander polynomial).

The author would like to thank the organisers of the conference "Algebraic modeling of topological and computational structures and applications", and in particular Professor Sofia Lambropoulou, who has inspired this research. A grateful acknowledgement to the referee whose suggestions led us to improve this paper, as well as [1]. This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALIS.

M. Chlouveraki (✉)
Laboratoire de Mathématiques UVSQ, Université de Versailles,
Bâtiment Fermat, 45 Avenue des États-Unis, 78035 Versailles Cedex, France
e-mail: maria.chlouveraki@uvsq.fr

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_3

Modular framisation (or simply framisation) is a mechanism proposed recently by Juyumaya and Lambropoulou [4] which consists of constructing a non-trivial extension of a knot algebra via the addition of the so-called "framing" generators, each of which is a generator of a cyclic group. In this way we obtain a new algebra which is related to framed braids and framed knots.

The inspiring example of framisation is the Yokonuma–Hecke algebra of type A. Yokonuma–Hecke algebras were introduced by Yokonuma [5] in the context of finite reductive groups as generalisations of Iwahori–Hecke algebras. Given a finite reductive group G, the Iwahori–Hecke algebra is the endomorphism ring of the permutation representation of G with respect to a Borel subgroup, while the Yokonuma–Hecke algebra is the endomorphism ring of the permutation representation of G with respect to a maximal unipotent subgroup.

In recent years, the presentation of the Yokonuma–Hecke algebra of type A has been transformed in [1, 6–9] to the one that we will use here. This new presentation is given by generators and relations, depending on two positive integers, d and n, and a parameter q. For $q = p^m$ and $d = p^m - 1$, where p is a prime number and m is a positive integer, the Yokonuma–Hecke algebra of type A, denoted by $\mathrm{Y}_{d,n}(q)$, is the endomorphism ring of the permutation representation of $\mathrm{GL}_n(\mathbb{F}_q)$ with respect to a maximal unipotent subgroup. The algebra $\mathrm{Y}_{d,n}(q)$ can be viewed as a framisation of the Iwahori–Hecke algebra $\mathcal{H}_n(q) \cong \mathrm{Y}_{1,n}(q)$, whose presentation we deform by adding the framing generators $t_1, \ldots, t_n$, which generate the finite abelian group $(\mathbb{Z}/d\mathbb{Z})^n$. Thus, $\mathrm{Y}_{d,n}(q)$ can be obtained as a quotient of the group algebra of both the framed braid group $\mathbb{Z}^n \rtimes B_n$ and the modular framed braid group $(\mathbb{Z}/d\mathbb{Z})^n \rtimes B_n$, where B_n denotes the classical braid group on n strands. Finally, $\mathrm{Y}_{d,n}(q)$ can be also obtained as a deformation of the group algebra of the complex reflection group $G(d, 1, n) \cong (\mathbb{Z}/d\mathbb{Z})^n \rtimes \mathfrak{S}_n$, different from the Ariki–Koike algebra [10].

Juyumaya [8] has defined a Markov trace on $\mathrm{Y}_{d,n}(q)$, which has been subsequently used by himself and Lambropoulou for the definition of 2-variable isotopy invariants for framed [11, 12], classical [13] and singular [14] knots and links, after Jones's method. The invariants for classical links are simply the invariants for framed links restricted to links with all framings equal to 0 (that is, the links obtained as closures of elements of B_n). Using the new presentation for $\mathrm{Y}_{d,n}(q)$ established in [6], we have recently proved that the classical link invariants obtained from the Yokonuma–Hecke algebra are not topologically equivalent to the Homflypt polynomial [15]. This implies that framisations of knot algebras are very useful to topologists not only for the construction of framed link invariants, but also for the construction of new classical link invariants.

We have now introduced and studied many interesting new algebras, which are obtained as framisations of other important knot algebras. First of all, we have the affine and cyclotomic Yokonuma–Hecke algebras [16], which generalise respectively the affine Hecke algebras of type A and the Ariki–Koike algebras. In fact, we have shown in [17] that affine Yokonuma–Hecke algebras appear also naturally in the study of p-adic reductive groups, arising from a construction analogous to the one used by Yokonuma, while cyclotomic Yokonuma–Hecke algebras give rise to both Ariki–Koike algebras and classical Yokonuma–Hecke algebras of type A, both as

quotients and as particular cases. Further, we have three possible framisations of the Temperley–Lieb algebra, all obtained as quotients of $Y_{d,n}(q)$ by a suitable two-sided ideal: the Yokonuma–Temperley–Lieb algebra [18], the Framisation of the Temperley–Lieb algebra [19] and the Complex Reflection Temperley–Lieb algebra [19]. Our study of the structure and the representation theory of these three algebras in [1, 20] indicates that the Framisation of the Temperley–Lieb algebra is the most natural analogue of the Temperley–Lieb algebra in this setting.

Now, all algebras mentioned above are endowed with Markov traces (see [16] for the affine and cyclotomic Yokonuma–Hecke algebras, and [18, 19] for the Temperley–Lieb quotients of the Yokonuma–Hecke algebra), which can be used for the definition of knot invariants after Jones's method. In view of the results of [15], we have concluded that the invariants for links in the solid torus obtained from the affine and cyclotomic Yokonuma–Hecke algebras in [16] are not topologically equivalent to the invariants obtained from the affine and cyclotomic Hecke algebras in [21–23], whereas the 1-variable invariants obtained from the Framisation of the Temperley–Lieb algebra in [19] are not topologically equivalent to the Jones polynomial.

In view of the importance of the algebras mentioned above to both algebraists and topologists, in this paper we will study their algebra structure and representation theory. We will provide explicit combinatorial formulas for their irreducible representations, compute their dimensions, construct bases for them and give semisimplicity criteria. We will also discuss their symmetric algebra structure.

3.2 Symmetric Algebras

Let R be a commutative integral domain and let A be an R-algebra, free and finitely generated as an R-module. If R' is a commutative integral domain containing R, we will write $R'A$ for $R' \otimes_R A$ and we will denote by $\mathrm{Irr}(R'A)$ the set of irreducible representations of $R'A$.

A *symmetrising form* on the algebra A is a linear map $\tau : A \to R$ such that

(a) $\tau(ab) = \tau(ba)$ for all $a, b \in A$, that is, τ is a trace function, and
(b) the map $\widehat{\tau} : A \to \mathrm{Hom}_R(A, R)$, $a \mapsto (x \mapsto \tau(ax))$ is an isomorphism of A-bimodules.

If there exists a symmetrising form on A, we say that A is a *symmetric* algebra.

Example 3.1 Let G be a finite group. The linear map $\tau : \mathbb{Z}[G] \to \mathbb{Z}$ defined by $\tau(1) = 1$ and $\tau(g) = 0$ for all $g \in G \setminus \{1\}$ is a symmetrising form on $\mathbb{Z}[G]$; it is called the *canonical symmetrising form* on $\mathbb{Z}[G]$.

Suppose that there exists a symmetrising form τ on A. Let K be a field containing R such that the algebra KA is split. The map τ can be extended to KA by extension of scalars. If $V \in \mathrm{Irr}(KA)$ and χ_V denotes the corresponding irreducible character,

then $\widehat{\tau}^{-1}(\chi_V)$ belongs to the centre of KA [24, Lemma 7.1.7]. Schur's lemma implies that $\widehat{\tau}^{-1}(\chi_V)$ acts as a scalar on V; we define this scalar to be the *Schur element* associated with V and denote it by s_V. We have $s_V \in R_K$, where R_K denotes the integral closure of R in K [24, Proposition 7.3.9].

Example 3.2 Let G be a finite group and let τ be the canonical symmetrising form on $A := \mathbb{Z}[G]$. If K is an algebraically closed field of characteristic 0, then KA is a split semisimple algebra and $s_V = |G|/\chi_V(1)$ for all $V \in \mathrm{Irr}(KA)$.

Following [24, Theorem 7.2.6], we have that the algebra KA is semisimple if and only if $s_V \neq 0$ for all $V \in \mathrm{Irr}(KA)$. If this is the case,

$$\tau = \sum_{V \in \mathrm{Irr}(KA)} \frac{1}{s_V} \chi_V.$$

From now on, we assume that R is integrally closed in K. Let $\theta : R \to L$ be a ring homomorphism into a field L such that L is the field of fractions of $\theta(R)$. We call such a ring homomorphism a *specialisation* of R. Schur elements can be then used to determine whether the algebra LA is semisimple as follows [24, Theorem 7.4.7]:

Theorem 3.1 *Assume that KA and LA are split and that A is symmetric with symmetrising form τ. For any simple KA-module V, let $s_V \in R$ be the Schur element with respect to τ. Then LA is semisimple if and only if $\theta(s_V) \neq 0$ for all $V \in \mathrm{Irr}(KA)$.*

Finally, if LA is semisimple, we have the following famous result known as "Tits's deformation theorem". For its proof, the reader may refer, for example, to [24, Theorem 7.4.6].

Theorem 3.2 *Assume that KA and LA are split. If LA is semisimple, then KA is also semisimple and we have a bijection* $\mathrm{Irr(KA)} \leftrightarrow \mathrm{Irr(LA)}$.

3.3 Yokonuma–Hecke Algebras

Let $n \in \mathbb{N}$, $d \in \mathbb{N}^*$. Let q be an indeterminate. The *Yokonuma–Hecke algebra (of type A)*, denoted by $\mathrm{Y}_{d,n}(q)$, is an associative $\mathbb{C}[q, q^{-1}]$-algebra generated by the elements

$$g_1, \ldots, g_{n-1}, t_1, \ldots, t_n$$

subject to the following relations:

$$\begin{array}{lrll}
(\mathrm{b}_1) & g_i g_j &= g_j g_i & \text{for all } i, j = 1, \ldots, n-1 \text{ such that } |i-j| > 1, \\
(\mathrm{b}_2) & g_i g_{i+1} g_i &= g_{i+1} g_i g_{i+1} & \text{for all } i = 1, \ldots, n-2, \\
(\mathrm{f}_1) & t_i t_j &= t_j t_i & \text{for all } i, j = 1, \ldots, n, \\
(\mathrm{f}_2) & t_j g_i &= g_i t_{s_i(j)} & \text{for all } i = 1, \ldots, n-1 \text{ and } j = 1, \ldots, n, \\
(\mathrm{f}_3) & t_j^d &= 1 & \text{for all } j = 1, \ldots, n,
\end{array} \tag{3.3.1}$$

where s_i is the transposition $(i, i+1)$, together with the quadratic relations:

$$g_i^2 = q + (q-1)\, e_i\, g_i \qquad \text{for all } i = 1, \ldots, n-1, \tag{3.3.2}$$

where

$$e_i := \frac{1}{d} \sum_{s=0}^{d-1} t_i^s t_{i+1}^{-s}. \tag{3.3.3}$$

It is easily verified that the elements e_i are idempotents in $\mathrm{Y}_{d,n}(q)$. Also, that the elements g_i are invertible, with

$$g_i^{-1} = q^{-1} g_i + (q^{-1} - 1)\, e_i \qquad \text{for all } i = 1, \ldots, n-1. \tag{3.3.4}$$

If we specialise q to 1, the defining relations (3.3.1)–(3.3.2) become the defining relations for the complex reflection group $G(d, 1, n) \cong (\mathbb{Z}/d\mathbb{Z}) \wr \mathfrak{S}_n \cong (\mathbb{Z}/d\mathbb{Z})^n \rtimes \mathfrak{S}_n$. Thus, the algebra $\mathrm{Y}_{d,n}(q)$ is a deformation of the group algebra over $\mathbb{C}$ of $G(d, 1, n)$. Moreover, for $d = 1$, the Yokonuma–Hecke algebra $\mathrm{Y}_{1,n}(q)$ coincides with the Iwahori–Hecke algebra $\mathcal{H}_n(q)$ of type A, and thus, for $d = 1$ and q specialised to 1, we obtain the group algebra over $\mathbb{C}$ of the symmetric group $\mathfrak{S}_n$.

Remark 3.1 Note that in all the papers prior to [6], the algebra $\mathrm{Y}_{d,n}(q)$ is generated by elements $\overline{g}_1, \ldots, \overline{g}_{n-1}, t_1, \ldots, t_n$ satisfying relations (3.3.1) and the quadratic relations:

$$\overline{g}_i^2 = 1 + (q-1)\, e_i + (q-1)\, e_i\, \overline{g}_i \qquad \text{for all } i = 1, \ldots, n-1. \tag{3.3.5}$$

This presentation changed in [6], where we considered $\mathrm{Y}_{d,n}(q)$ defined over $\mathbb{C}[q^{1/2}, q^{-1/2}]$ and generated by elements $\widetilde{g}_1, \ldots, \widetilde{g}_{n-1}, t_1, \ldots, t_n$ satisfying relations (3.3.1) and the quadratic relations:

$$\widetilde{g}_i^2 = 1 + (q^{1/2} - q^{-1/2})\, e_i\, \widetilde{g}_i \qquad \text{for all } i = 1, \ldots, n-1. \tag{3.3.6}$$

By taking $\overline{g}_i := \widetilde{g}_i + (q^{1/2} - 1)\, e_i \widetilde{g}_i$ (and thus, $\widetilde{g}_i = \overline{g}_i + (q^{-1/2} - 1)\, e_i\, \overline{g}_i$), we obtain the old presentation of the Yokonuma–Hecke algebra. By taking $g_i := q^{1/2} \widetilde{g}_i$ we obtain our presentation of the Yokonuma–Hecke algebra.

Now let $w \in \mathfrak{S}_n$, and let $w = s_{i_1} s_{i_2} \ldots s_{i_r}$ be a reduced expression for w. Since the generators g_i of the Yokonuma–Hecke algebra satisfy the same braid relations as the generators of $\mathfrak{S}_n$, Matsumoto's theorem (see, for example, [24, Theorem 1.2.2])

implies that the element $g_w := g_{i_1} g_{i_2} \dots g_{i_r}$ is well defined, that is, it does not depend on the choice of the reduced expression of w.

Juyumaya [8] has proved that the following set is a $\mathbb{C}[q, q^{-1}]$-basis of $\mathrm{Y}_{d,n}(q)$:

$$\mathcal{B}^{\mathrm{H}}_{d,n} := \left\{ t_1^{r_1} \dots t_n^{r_n} g_w \;\middle|\; w \in \mathfrak{S}_n,\ 0 \leqslant r_j \leqslant d-1 \text{ for all } j = 1, 2, \dots, n \right\}. \tag{3.3.7}$$

In particular, $\mathrm{Y}_{d,n}(q)$ is a free $\mathbb{C}[q, q^{-1}]$-module of rank $d^n n!$.

The representation theory of Yokonuma–Hecke algebras has been first studied by Thiem [25–27] in the general context of unipotent Hecke algebras. The generality of his results and the new presentation for $\mathrm{Y}_{d,n}(q)$ has led us to develop in [6] a combinatorial approach to the representation theory of the Yokonuma–Hecke algebra of type A, in terms of d-partitions and standard d-tableaux.

3.3.1 Combinatorics of d-Partitions and Standard d-tableaux

A *partition* $\lambda = (\lambda_1, \dots, \lambda_h)$ is a family of positive integers such that $\lambda_1 \geqslant \lambda_2 \geqslant \dots \geqslant \lambda_h \geqslant 1$. We write $|\lambda| := \sum_{i=1}^{h} \lambda_i$ and we say that λ is a partition of n if $n = |\lambda|$. We denote by $\mathcal{P}(n)$ the set of partitions of n. We define the set of nodes $[\lambda]$ of λ to be the set

$$[\lambda] := \{(x, y) \mid 1 \leqslant x \leqslant h,\ 1 \leqslant y \leqslant \lambda_x\}.$$

We identify partitions with their Young diagrams: the Young diagram of λ is a left-justified array of h rows such that the i-th row contains λ_i boxes (nodes) for all $i = 1, \dots, h$.

A *d-partition* of n is an ordered d-tuple $\boldsymbol{\lambda} = (\lambda^{(0)}, \lambda^{(1)}, \dots, \lambda^{(d-1)})$ of partitions such that $|\boldsymbol{\lambda}| := \sum_{i=0}^{d-1} |\lambda^{(i)}| = n$. We denote by $\mathcal{P}(d, n)$ the set of d-partitions of n. The empty multipartition, denoted by $\boldsymbol{\emptyset}$, is a d-tuple of empty partitions. A node $\boldsymbol{\theta}$ of $\boldsymbol{\lambda}$ is a triple (x, y, i), where $0 \leqslant i \leqslant d-1$ and (x, y) is a node of the partition $\lambda^{(i)}$. We define $\mathrm{p}(\boldsymbol{\theta}) := i$ to be the *position* of $\boldsymbol{\theta}$ and $\mathrm{c}(\boldsymbol{\theta}) := q^{y-x}$ to be the *quantum content* of $\boldsymbol{\theta}$.

A *d-tableau of shape* $\boldsymbol{\lambda}$ is a bijection between the set $\{1, \dots, n\}$ and the set of nodes of $\boldsymbol{\lambda}$. In other words, a d-tableau of shape $\boldsymbol{\lambda}$ is obtained by placing the numbers $1, \dots, n$ in the nodes of $\boldsymbol{\lambda}$. The *size* of a d-tableau of shape $\boldsymbol{\lambda}$ is n, that is, the size of $\boldsymbol{\lambda}$. A d-tableau is *standard* if its entries increase along any row and down any column of every diagram in $\boldsymbol{\lambda}$. For $d = 1$, a standard 1-tableau is a usual standard tableau.

For a d-tableau $\mathcal{T}$, we denote respectively by $\mathrm{p}(\mathcal{T}|i)$ and $\mathrm{c}(\mathcal{T}|i)$ the position and the quantum content of the node with the number i in it. For example, for the standard 3-tableau $\mathcal{T} = \left(\begin{array}{|c|c|}\hline 1 & 3 \\ \hline \end{array},\ \varnothing,\ \begin{array}{|c|}\hline 2 \\ \hline \end{array} \right)$ of size 3, we have

$$\mathrm{p}(\mathcal{T}|1) = 0\,,\ \ \mathrm{p}(\mathcal{T}|2) = 2\,,\ \ \mathrm{p}(\mathcal{T}|3) = 0 \quad \text{and} \quad \mathrm{c}(\mathcal{T}|1) = 1\,,\ \ \mathrm{c}(\mathcal{T}|2) = 1\,,\ \ \mathrm{c}(\mathcal{T}|3) = q\,.$$

For any d-tableau $\mathcal{T}$ of size n and any permutation $\sigma \in \mathfrak{S}_n$, we denote by $\mathcal{T}^\sigma$ the d-tableau obtained from $\mathcal{T}$ by applying the permutation σ on the numbers contained in the nodes of $\mathcal{T}$. We have

$$\mathrm{p}(\mathcal{T}^\sigma|i) = \mathrm{p}\big(\mathcal{T}|\sigma^{-1}(i)\big) \quad \text{and} \quad \mathrm{c}(\mathcal{T}^\sigma|i) = \mathrm{c}\big(\mathcal{T}|\sigma^{-1}(i)\big) \qquad \text{for all } i = 1, \dots, n.$$

Note that if the d-tableau $\mathcal{T}$ is standard, the d-tableau $\mathcal{T}^\sigma$ is not necessarily standard.

3.3.2 Representation Theory of Yokonuma–Hecke Algebras

Let $\boldsymbol{\lambda} \in \mathcal{P}(d, n)$, and let $V_{\boldsymbol{\lambda}}$ be a $\mathbb{C}(q)$-vector space with a basis $\{\mathbf{v}_{\mathcal{T}}\}$ indexed by the standard d-tableaux of shape $\boldsymbol{\lambda}$. We set $\mathbf{v}_{\mathcal{T}} := 0$ for any non-standard d-tableau $\mathcal{T}$ of shape $\boldsymbol{\lambda}$. By [6, Proposition 5 and Theorem 1] and [1, Theorem 3.7], we have the following description of the irreducible representations of $\mathbb{C}(q)\mathrm{Y}_{d,n}(q)$:

Theorem 3.3 *Let $\{\xi_0, \xi_1, \dots, \xi_{d-1}\}$ be the set of all d-th roots of unity (ordered arbitrarily). Let $\mathcal{T}$ be a standard d-tableau of shape $\boldsymbol{\lambda} \in \mathcal{P}(d, n)$. For brevity, we set $\mathrm{p}_i := \mathrm{p}(\mathcal{T}|i)$ and $\mathrm{c}_i := \mathrm{c}(\mathcal{T}|i)$ for all $i = 1, \dots, n$. The vector space $V_{\boldsymbol{\lambda}}$ is a representation of $\mathbb{C}(q)\mathrm{Y}_{d,n}(q)$ with the action of the generators on the basis element $\mathbf{v}_{\mathcal{T}}$ defined as follows: for $j = 1, \dots, n$,*

$$t_j(\mathbf{v}_{\mathcal{T}}) = \xi_{\mathrm{p}_j} \mathbf{v}_{\mathcal{T}} \; ; \tag{3.3.8}$$

for $i = 1, \dots, n-1$, if $\mathrm{p}_i > \mathrm{p}_{i+1}$ then

$$g_i(\mathbf{v}_{\mathcal{T}}) = \mathbf{v}_{\mathcal{T}^{s_i}} \; , \tag{3.3.9}$$

if $\mathrm{p}_i < \mathrm{p}_{i+1}$ then

$$g_i(\mathbf{v}_{\mathcal{T}}) = q\, \mathbf{v}_{\mathcal{T}^{s_i}} \; , \tag{3.3.10}$$

and if $\mathrm{p}_i = \mathrm{p}_{i+1}$ then

$$g_i(\mathbf{v}_{\mathcal{T}}) = \frac{q\mathrm{c}_{i+1} - \mathrm{c}_{i+1}}{\mathrm{c}_{i+1} - \mathrm{c}_i}\, \mathbf{v}_{\mathcal{T}} + \frac{q\mathrm{c}_{i+1} - \mathrm{c}_i}{\mathrm{c}_{i+1} - \mathrm{c}_i}\, \mathbf{v}_{\mathcal{T}^{s_i}} \; , \tag{3.3.11}$$

where s_i is the transposition $(i, i+1)$. Further, the set $\{V_{\boldsymbol{\lambda}} \mid \boldsymbol{\lambda} \in \mathcal{P}(d, n)\}$ is a complete set of pairwise non-isomorphic irreducible representations of $\mathbb{C}(q)\mathrm{Y}_{d,n}(q)$.

The above theorem implies that the algebra $\mathbb{C}(q)\mathrm{Y}_{d,n}(q)$ is split. As we have already mentioned, when $q \mapsto 1$, the algebra $\mathbb{C}(q)\mathrm{Y}_{d,n}(q)$ specialises to the group algebra $\mathbb{C}[G(d, 1, n)]$, which is semisimple. By Tits's deformation theorem, we obtain that the algebra $\mathbb{C}(q)\mathrm{Y}_{d,n}(q)$ is also semisimple.

Let now $\theta : \mathbb{C}[q, q^{-1}] \to \mathbb{C}$ be a ring homomorphism such that $\theta(q) = \eta \in \mathbb{C} \setminus \{0\}$. Using the representation theory of $\mathbb{C}(q)\mathrm{Y}_{d,n}(q)$, we have proved the following semisimplicity criterion for $\mathbb{C}\mathrm{Y}_{d,n}(\eta)$ [6, Proposition 9]:

Proposition 3.1 *The specialised Yokonuma–Hecke algebra $\mathbb{C}\mathrm{Y}_{d,n}(\eta)$ is (split) semisimple if and only if $\theta(P(q)) \neq 0$, where*

$$P(q) = \prod_{1 \leqslant i \leqslant n} (1 + q + \cdots + q^{i-1}).$$

Note that following Ariki's semisimplicity criterion [28] for Ariki–Koike algebras (and so, in particular, for Iwahori–Hecke algebras of type A), the algebra $\mathbb{C}\mathrm{Y}_{d,n}(\eta)$ is semisimple if and only if the specialised Iwahori–Hecke algebra $\mathbb{C}\mathcal{H}_n(\eta)$ is semisimple.

Another way to obtain the above result is through our definition of a canonical symmetrising form τ on $\mathrm{Y}_{d,n}(q)$ [6, Proposition 10]. Having calculated the Schur elements of $\mathrm{Y}_{d,n}(q)$ with respect to τ [6, Proposition 11], we can deduce the above semisimplicity criterion with the use of Theorem 3.1. More precisely, we have the following:

Theorem 3.4 *We define the linear map $\tau : \mathrm{Y}_{d,n}(q) \to \mathbb{C}[q, q^{-1}]$ by*

$$\tau(t_1^{r_1} \dots t_n^{r_n} g_w) = \begin{cases} 1 \text{ if } w = 1 \text{ and } r_j = 0 \text{ for all } j = 1, 2, \dots, n, \\ 0 \text{ otherwise,} \end{cases} \tag{3.3.12}$$

where $w \in \mathfrak{S}_n$ and $0 \leqslant r_j \leqslant d-1$ for all $j = 1, 2, \dots, n$. Then τ is a symmetrising form on $\mathrm{Y}_{d,n}(q)$, called the canonical symmetrising form. If $\boldsymbol{\lambda} = (\lambda^{(0)}, \dots, \lambda^{(d-1)}) \in \mathcal{P}(d, n)$, then the Schur element of $V_{\boldsymbol{\lambda}}$ with respect to τ is

$$s_{\boldsymbol{\lambda}} = d^n \, s_{\lambda^{(0)}} s_{\lambda^{(1)}} \dots s_{\lambda^{(d-1)}}, \tag{3.3.13}$$

where $s_{\lambda^{(i)}}$ is the Schur element of the Iwahori–Hecke algebra $\mathcal{H}_{|\lambda^{(i)}|}(q)$ corresponding to $\lambda^{(i)}$ for all $i = 0, 1, \dots, d-1$ (we take $s_\emptyset := 1$).

The Schur elements of Iwahori–Hecke algebras of type A have been calculated by Steinberg [29]. A simple formula for them is given by Jacon and the author in [30].

The connection between the representation theory of the Yokonuma–Hecke algebra and that of Iwahori–Hecke algebras of type A implied by (3.3.13) is explained by a result of Lusztig [31, Sect. 34], who has proved that Yokonuma–Hecke algebras, in general, are isomorphic to direct sums of matrix algebras over certain subalgebras of classical Iwahori–Hecke algebras. Using the new presentation for $\mathrm{Y}_{d,n}(q)$, Jacon and Poulain d'Andecy [32] have explicitly described this isomorphism between the Yokonuma–Hecke algebra of type A and a direct sum of matrix algebras over tensor products of Iwahori–Hecke algebras of type A. Another proof of this result has been given recently in [33], where Espinoza and Ryom-Hansen have constructed a

concrete isomorphism between $\mathrm{Y}_{d,n}(q)$ and Shoji's modified Ariki–Koike algebra. Note that in all cases the result has been obtained over the ring $\mathbb{C}[q^{1/2}, q^{-1/2}]$ (using the generators $\widetilde{g}_i$ defined in Remark 3.1). We have managed to show that it is still valid over the smaller ring $\mathbb{C}[q, q^{-1}]$. We have [1, Theorem 4.3]:

$$\mathrm{Y}_{d,n}(q) \cong \bigoplus_{\mu \in \mathrm{Comp}_d(n)} \mathrm{Mat}_{m_\mu}(\mathcal{H}_{\mu_0}(q) \otimes \mathcal{H}_{\mu_1}(q) \otimes \cdots \otimes \mathcal{H}_{\mu_{d-1}}(q)), \tag{3.3.14}$$

where

$$\mathrm{Comp}_d(n) = \{\mu = (\mu_0, \mu_1, \ldots, \mu_{d-1}) \in \mathbb{N}^d \mid \mu_0 + \mu_1 + \cdots + \mu_{d-1} = n\} \tag{3.3.15}$$

and

$$m_\mu = \frac{n!}{\mu_0! \mu_1! \ldots \mu_{d-1}!} \tag{3.3.16}$$

3.4 Affine and Cyclotomic Yokonuma–Hecke Algebras

In [16], we introduced the affine and cyclotomic Yokonuma–Hecke algebras, which give rise to both Ariki–Koike algebras and Yokonuma–Hecke algebras of type A as quotients and as special cases. Let $n \in \mathbb{N}$, $d \in \mathbb{N}^*$ and $l \in \mathbb{N}^* \cup \{\infty\}$. Let q and $(Q_i)_{i \in \mathbb{N}}$ be indeterminates, and set $\mathcal{R}_l := \mathbb{C}[q^{\pm 1}, Q_0^{\pm 1}, Q_1^{\pm 1}, \ldots, Q_{l-1}^{\pm 1}]$ if $l < \infty$, and $\mathcal{R}_\infty := \mathbb{C}[q^{\pm 1}]$. We define the algebra $\mathrm{Y}(d, l, n)$ to be the associative $\mathcal{R}_l$-algebra generated by the elements

$$g_1, \ldots, g_{n-1}, t_1, \ldots, t_n, X_1, X_1^{-1}$$

subject to the relations (3.3.1)–(3.3.2), together with the following relations concerning the generator X_1:

$$\begin{array}{rll} X_1 X_1^{-1} = X_1^{-1} X_1 = 1 & & \\ X_1 g_1 X_1 g_1 = g_1 X_1 g_1 X_1 & & \\ X_1 g_i = g_i X_1 & \text{for all } i = 2, \ldots, n-1, & \\ X_1 t_j = t_j X_1 & \text{for all } j = 1, \ldots, n, & \end{array} \tag{3.4.1}$$

and if $l < \infty$,

$$(X_1 - Q_0)(X_1 - Q_1) \cdots (X_1 - Q_{l-1}) = 0. \tag{3.4.2}$$

The algebra $\mathrm{Y}(d, \infty, n)$ is called the *affine Yokonuma–Hecke algebra*. For $l < \infty$, the algebra $\mathrm{Y}(d, l, n)$ is called the *cyclotomic Yokonuma–Hecke algebra*. These algebras are isomorphic to the modular framisations of, respectively, the affine Hecke algebra ($l = \infty$) and the Ariki–Koike algebra ($l < \infty$); see definitions in [4, Sect. 6] and [6, Remark 1].

The cyclotomic Yokonuma–Hecke algebra is a quotient of the affine Yokonuma–Hecke algebra by the relation (3.4.2). If we map $X_1 \mapsto Q_0$ for $l < \infty$ or $X_1 \mapsto 1$ for $l = \infty$, we obtain a surjection of $\mathrm{Y}(d, l, n)$ onto $\mathrm{Y}_{d,n}(q)$. If we map $t_j \mapsto 1$ for all $j = 1, \ldots, n$, then we obtain a surjection of $\mathrm{Y}(d, l, n)$ onto $\mathcal{H}(l, n)$, where $\mathcal{H}(l, n)$ denotes the Ariki–Koike algebra associated to $G(l, 1, n)$ for $l < \infty$ and $\mathcal{H}(\infty, n)$ denotes the affine Hecke algebra of type A. Moreover, we have $\mathrm{Y}(d, 1, n) \cong \mathrm{Y}_{d,n}(q)$ and $\mathrm{Y}(1, l, n) \cong \mathcal{H}(l, n)$. In particular, we have $\mathrm{Y}(1, 1, n) \cong \mathcal{H}_n(q)$.

Remark 3.2 Let p be a prime number. In a recent series of papers [34–36], Vignéras introduced and studied a large family of algebras, called *pro-p-Iwahori–Hecke algebras*. They generalise convolution algebras of compactly supported functions on a p-adic connected reductive group that are bi-invariant under the pro-p-radical of an Iwahori subgroup, which play an important role in the p-modular representation theory of p-adic reductive groups. In [17] we have shown that the affine Yokonuma–Hecke algebra $\mathrm{Y}(d, \infty, n)$ is a pro-p-Iwahori–Hecke algebra. Thus, the affine Yokonuma–Hecke algebra generalises the affine Hecke algebra of type A in a similar way that the Yokonuma–Hecke algebra generalises the Iwahori–Hecke algebra of type A. In particular, for $q = p^m$ and $d = p^m - 1$, where m is a positive integer, $\mathrm{Y}(d, \infty, n)$ is isomorphic to the convolution algebra of complex valued and compactly supported functions on the group $\mathrm{GL}_n(F)$, with F a suitable p-adic field, that are bi-invariant under the pro-p-radical of an Iwahori subgroup.

Remark 3.3 Following Lusztig's approach in [31], Cui [37] has established an explicit algebra isomorphism between the affine Yokonuma–Hecke algebra $\mathrm{Y}(d, \infty, n)$ and a direct sum of matrix algebras over tensor products of affine Hecke algebras of type A, similar to (3.3.14). More recently, Poulain d'Andecy [38] obtained the same result, as well as an isomorphism between the cyclotomic Yokonuma–Hecke algebra $\mathrm{Y}(d, l, n)$, where $l < \infty$, and a direct sum of matrix algebras over tensor products of Ariki–Koike algebras, using the same approach as in [32]. The isomorphism theorem for cyclotomic Yokonuma–Hecke algebras has been subsequently re-obtained by Rostam [39] using his result that cyclotomic Yokonuma–Hecke algebras are cyclotomic quiver Hecke algebras.

In [16] we have constructed several bases for the algebra $\mathrm{Y}(d, l, n)$. In order to describe them here, we introduce the following notation: Let $Z_l := \{0, \ldots, l-1\}$ for $l < \infty$ and $Z_\infty := \mathbb{Z}$. We define inductively elements $X_2, \ldots, X_n$ of $\mathrm{Y}(d, l, n)$ by setting

$$X_{i+1} := q^{-1} g_i X_i g_i \quad \text{for all } i = 1, \ldots, n-1.$$

Let $\mathcal{B}_{d,n}$ be a basis of the Yokonuma–Hecke algebra $\mathrm{Y}_{d,n}(q) \cong \mathrm{Y}(d, 1, n)$ over $\mathcal{R}_l$ (we can take, for example, $\mathcal{B}^{\mathrm{H}}_{d,n}$ defined in (3.3.7)). We denote by $\mathcal{B}^{\mathrm{AK}}_{d,l,n}$ the following set of elements of $\mathrm{Y}(d, l, n)$:

$$X_1^{a_1} \ldots X_n^{a_n} \cdot \omega, \quad a_k \in Z_l \text{ and } \omega \in \mathcal{B}_{d,n}.$$

Now, for $k = 1, \ldots, n$, we set

$$
\begin{aligned}
W^{(k)}_{J,a,b} &:= g_J^{-1} \dots g_2^{-1} g_1^{-1} X_1^a t_1^b g_1 g_2 \dots g_{k-1}\,,\\
W^{(k)-}_{J,a,b} &:= g_J \dots g_2 g_1 X_1^a t_1^b g_1^{-1} g_2^{-1} \dots g_{k-1}^{-1}\,,\\
\widetilde{W}^{(k)}_{J,a,b} &:= g_J \dots g_2 g_1 X_1^a t_1^b g_1 g_2 \dots g_{k-1}\,,\\
\widetilde{W}^{(k)-}_{J,a,b} &:= g_J^{-1} \dots g_2^{-1} g_1^{-1} X_1^a t_1^b g_1^{-1} g_2^{-1} \dots g_{k-1}^{-1}\,,
\end{aligned}
$$

where $J \in \{0, \dots, k-1\}$ and $a, b \in \mathbb{Z}$. We use the following standard conventions: for $\epsilon = \pm 1$, $g_J^\epsilon \dots g_2^\epsilon g_1^\epsilon := 1$ and $g_{k-J}^\epsilon \dots g_{k-2}^\epsilon g_{k-1}^\epsilon := 1$ if $J = 0$. Then we denote, respectively, by $\mathcal{B}^{\mathrm{Ind}}_{d,l,n}$, $\mathcal{B}^{\mathrm{Ind}-}_{d,l,n}$, $\widetilde{\mathcal{B}}^{\mathrm{Ind}}_{d,l,n}$ and $\widetilde{\mathcal{B}}^{\mathrm{Ind}-}_{d,l,n}$ the following sets of elements of $\mathrm{Y}(d,l,n)$:

$$W^{(n)}_{J_n,a_n,b_n} \dots W^{(2)}_{J_2,a_2,b_2} W^{(1)}_{J_1,a_1,b_1}\,, \quad J_k \in \{0,\dots,k-1\}, a_k \in Z_l \text{ and } b_k \in \{0,\dots,d-1\}.$$

$$W^{(n)-}_{J_n,a_n,b_n} \dots W^{(2)-}_{J_2,a_2,b_2} W^{(1)-}_{J_1,a_1,b_1}\,, \quad J_k \in \{0,\dots,k-1\}, a_k \in Z_l \text{ and } b_k \in \{0,\dots,d-1\}.$$

$$\widetilde{W}^{(n)}_{J_n,a_n,b_n} \dots \widetilde{W}^{(2)}_{J_2,a_2,b_2} \widetilde{W}^{(1)}_{J_1,a_1,b_1}\,, \quad J_k \in \{0,\dots,k-1\}, a_k \in Z_l \text{ and } b_k \in \{0,\dots,d-1\}.$$

$$\widetilde{W}^{(n)-}_{J_n,a_n,b_n} \dots \widetilde{W}^{(2)-}_{J_2,a_2,b_2} \widetilde{W}^{(1)-}_{J_1,a_1,b_1}\,, \quad J_k \in \{0,\dots,k-1\}, a_k \in Z_l \text{ and } b_k \in \{0,\dots,d-1\}.$$

We then have the following [16, Theorem 4.4]:

Theorem 3.5 *Each set $\mathcal{B}^{\mathrm{AK}}_{d,l,n}$, $\mathcal{B}^{Ind}_{d,l,n}$, $\mathcal{B}^{Ind-}_{d,l,n}$, $\widetilde{\mathcal{B}}^{Ind}_{d,l,n}$ and $\widetilde{\mathcal{B}}^{Ind-}_{d,l,n}$ is an $\mathcal{R}_l$-basis of $\mathrm{Y}(d,l,n)$. In particular, $\mathrm{Y}(d,l,n)$ is a free $\mathcal{R}_l$-module and, if $l < \infty$, its rank is equal to $(dl)^n n!$.*

Remark 3.4 The set $\mathcal{B}^{\mathrm{AK}}_{d,l,n}$ is the analogue of the Ariki–Koike basis of the Ariki–Koike algebra $\mathcal{H}(l,n)$ for $l < \infty$, and the standard Bernstein basis of the affine Hecke algebra of type A for $l = \infty$. The four other sets are inductive sets with respect to n, which are analogous to the inductive bases of $\mathcal{H}(l,n)$ studied in [23, 40].

Furthermore, in [16] we have studied the representation theory of the cyclotomic Yokonuma–Hecke algebra $\mathrm{Y}(d,l,n)$, which is quite similar to the representation theory of the Yokonuma–Hecke algebra $\mathrm{Y}_{d,n}(q)$. From now on, we only consider the case $l < \infty$.

Let $\mathcal{K}_l$ denote the field of fractions of $\mathcal{R}_l$. We will see that the irreducible representations of the algebra $\mathcal{K}_l \mathrm{Y}(d,l,n)$ are parametrised by the dl-partitions of n. Instead of looking though at dl-partitions as dl-tuples of partitions, we look at them as d-tuples of l-partitions, and we call them (d,l)-*partitions* when seen as such. We denote by $\mathcal{P}(d,l,n)$ the set of (d,l)-partitions of n. If $\boldsymbol{\lambda} \in \mathcal{P}(d,l,n)$, then $\boldsymbol{\lambda} = (\boldsymbol{\lambda}^{(0)}, \boldsymbol{\lambda}^{(1)}, \dots, \boldsymbol{\lambda}^{(d-1)})$, where $\boldsymbol{\lambda}^{(i)}$ is an l-partition for all $i = 0, 1, \dots, d-1$, and $\sum_{i=0}^{d-1} |\boldsymbol{\lambda}^{(i)}| = n$. We thus have $\boldsymbol{\lambda}^{(i)} = (\lambda^{(i,0)}, \lambda^{(i,1)}, \dots, \lambda^{(i,l-1)})$, where $\lambda^{(i,j)}$ is a partition for all $i = 0, 1, \dots, d-1$ and $j = 0, 1, \dots, l-1$, and $\sum_{i=0}^{d-1} \sum_{j=0}^{l-1} |\lambda^{(i,j)}| = n$.

A node $\boldsymbol{\theta}$ of $\boldsymbol{\lambda}$ is a 4-tuple (x, y, i, j), where $0 \leqslant i \leqslant d-1$, $0 \leqslant j \leqslant l-1$ and (x, y) is a node of the partition $\lambda^{(i,j)}$. We define $\mathrm{p}(\boldsymbol{\theta}) := i$ to be the *d-position* of $\boldsymbol{\theta}$ and $\mathrm{c}(\boldsymbol{\theta}) := Q_j q^{y-x}$ to be the *l-quantum content* of $\boldsymbol{\theta}$.

Following the definitions in Sect. 3.3.1, a (d, l)-*tableau* is simply a dl-tableau and a *standard* (d, l)-*tableau* is simply a standard dl-tableau. For a (d, l)-tableau $\mathcal{T}$ and for $i = 1, \ldots, n$, we denote respectively by $\mathrm{p}(\mathcal{T}|i)$ and $\mathrm{c}(\mathcal{T}|i)$ the d-position and the l-quantum content of the node with the number i in it.

Now, let $\boldsymbol{\lambda} \in \mathcal{P}(d, l, n)$, and let $V_{\boldsymbol{\lambda}}$ be a $\mathbb{C}(q)$-vector space with a basis $\{\mathbf{v}_{\mathcal{T}}\}$ indexed by the standard (d, l)-tableaux of shape $\boldsymbol{\lambda}$. We set $\mathbf{v}_{\mathcal{T}} := 0$ for any non-standard (d, l)-tableau $\mathcal{T}$ of shape $\boldsymbol{\lambda}$. By [16, Propositions 3.2 and 3.4], the vector space $V_{\boldsymbol{\lambda}}$ is a representation of $\mathcal{K}_l \mathrm{Y}(d, l, n)$, with the action of the generators $g_1, \ldots, g_{n-1}, t_1, \ldots, t_n$ on the basis element $\mathbf{v}_{\mathcal{T}}$ defined exactly as in Theorem 3.3, and the action of the generator X_1 given by:

$$X_1(\mathbf{v}_{\mathcal{T}}) = \mathrm{c}(\mathcal{T}|1)\, \mathbf{v}_{\mathcal{T}}. \tag{3.4.3}$$

Further, the set $\{V_{\boldsymbol{\lambda}} \mid \boldsymbol{\lambda} \in \mathcal{P}(d, l, n)\}$ is a complete set of pairwise non-isomorphic irreducible representations of $\mathcal{K}_l \mathrm{Y}(d, l, n)$.

Remark 3.5 We can easily show, by induction on i, that [16, Lemma 3.3]:

$$X_i(\mathbf{v}_{\mathcal{T}}) = \mathrm{c}(\mathcal{T}|i)\, \mathbf{v}_{\mathcal{T}} \quad \text{for all } i = 1, \ldots, n. \tag{3.4.4}$$

We also have a semisimplicity criterion for cyclotomic Yokonuma–Hecke algebras, which is exactly the same as Ariki's semisimplicity criterion [28] for Ariki–Koike algebras [16, Proposition 4.7]:

Proposition 3.2 *Let $\theta : \mathcal{R}_l \to \mathbb{C}$ be a ring homomorphism such that $\theta(q) \prod_{j=0}^{l-1} \theta(Q_j) \neq 0$. The specialised cyclotomic Yokonuma–Hecke algebra $\mathbb{C}\mathrm{Y}(d, l, n)_\theta$, defined via θ, is (split) semisimple if and only if $\theta(P) \neq 0$, where*

$$P = \prod_{1 \leqslant i \leqslant n} (1 + q + \cdots + q^{i-1}) \prod_{0 \leqslant s < t \leqslant l-1} \prod_{-n < k < n} (q^k Q_s - Q_t).$$

We deduce that the algebra $\mathbb{C}\mathrm{Y}(d, l, n)_\theta$ is semisimple if and only if the specialised Ariki–Koike algebra $\mathbb{C}\mathcal{H}(l, n)_\theta$ is semisimple.

Finally, we have proved the existence of a "canonical" symmetrising form on $\mathcal{K}_l \mathrm{Y}(d, l, n)$ and calculated the Schur elements with respect to it [16, Sect. 7]:

Theorem 3.6 *We define the linear map $\tau : \mathrm{Y}(d, l, n) \to \mathcal{R}_l$ by*

$$\tau(X_1^{a_1} \ldots X_n^{a_n} t_1^{b_1} \ldots t_n^{b_n} g_w) = \begin{cases} 1 \text{ if } w = 1 \text{ and } a_j = b_j = 0 \text{ for all } j = 1, 2, \ldots, n, \\ 0 \text{ otherwise,} \end{cases} \tag{3.4.5}$$

where $w \in \mathfrak{S}_n$, $a_j \in Z_l$ and $0 \leqslant b_j \leqslant d-1$ for all $j = 1, 2, \ldots, n$. Then τ (extended linearly) is a symmetrising form on $\mathcal{K}_l \mathrm{Y}(d, l, n)$. If $\boldsymbol{\lambda} = (\boldsymbol{\lambda}^{(0)}, \ldots, \boldsymbol{\lambda}^{(d-1)}) \in \mathcal{P}(d, l, n)$, then the Schur element of $V_{\boldsymbol{\lambda}}$ with respect to τ is

$$s_{\boldsymbol{\lambda}} = d^n \, s_{\boldsymbol{\lambda}^{(0)}} s_{\boldsymbol{\lambda}^{(1)}} \ldots s_{\boldsymbol{\lambda}^{(d-1)}}, \tag{3.4.6}$$

where $s_{\boldsymbol{\lambda}^{(i)}}$ is the Schur element of the Ariki–Koike algebra $\mathcal{H}(l, |\boldsymbol{\lambda}^{(i)}|)$ corresponding to $\boldsymbol{\lambda}^{(i)}$ for all $i = 0, 1, \ldots, d-1$ (we take $s_{\emptyset} := 1$).

The Schur elements of Ariki–Koike algebras have been calculated independently by Geck–Iancu–Malle [41] and Mathas [42]. A simple formula for them is given by Jacon and the author in [30].

Remark 3.6 The map τ is known to be a symmetrising form on $\mathrm{Y}(d, l, n)$ (defined over $\mathcal{R}_l$) in cases $d = 1$ [43] and $l = 1$ [6]. In these cases, τ is called the *canonical symmetrising form* on $\mathrm{Y}(d, l, n)$.

Remark 3.7 Equation (3.4.6) hints towards an isomorphism between the cyclotomic Yokonuma–Hecke algebra $\mathrm{Y}(d, l, n)$ and a direct sum of matrix algebras over tensor products of Ariki–Koike algebras; this isomorphism was recently described by Poulain d'Andecy in [38] and by Rostam in [39].

3.5 Temperley–Lieb Quotients of Yokonuma–Hecke Algebras

The Temperley–Lieb algebra was introduced by Temperley and Lieb in [44] for its applications in statistical mechanics. Jones [2, 3, 45] later showed that it can be obtained as a quotient of the Iwahori–Hecke algebra $\mathcal{H}_n(q)$ of type A by a two-sided ideal, and used it for the construction of the knot invariant known as the Jones polynomial.

As explained in the introduction, we have three possible analogues of the Temperley–Lieb algebra in the Yokonuma–Hecke algebra setting: the Yokonuma–Temperley–Lieb algebra [18], the Framisation of the Temperley–Lieb algebra [19] and the Complex Reflection Temperley–Lieb algebra [19]. All three are defined as quotients of the Yokonuma–Hecke algebra $\mathrm{Y}_{d,n}(q)$ of type A by a suitable two-sided ideal, and they specialise to the classical Temperley–Lieb algebra for $d = 1$.

In this section, we will determine the irreducible representations of the three algebras by showing which representations of $\mathrm{Y}_{d,n}(q)$ pass to each quotient. We will compute their dimensions and construct bases for them. At the end of this section, it will be clear that the most natural analogue of the Temperley–Lieb algebra in this setting is the Framisation of the Temperley–Lieb algebra.

First, let us recall some information about the classical setting. Let $n \geqslant 3$. The *Temperley–Lieb algebra* $\mathrm{TL}_n(q)$ is defined as the quotient of the Iwahori–Hecke algebra $\mathcal{H}_n(q) \cong \mathrm{Y}_{1,n}(q)$ by the ideal I_n generated by the elements

$$g_{i,i+1} := 1 + g_i + g_{i+1} + g_i g_{i+1} + g_{i+1} g_i + g_i g_{i+1} g_i = \sum_{w \in \langle s_i, s_{i+1} \rangle} g_w$$

for all $i = 1, \dots, n-2$. It turns out that this ideal is principal, and we have $I_n = \langle g_{1,2} \rangle$.

Since the algebra $\mathbb{C}(q)\mathcal{H}_n(q)$ is semisimple, the algebra $\mathbb{C}(q)\mathrm{TL}_n(q)$ is also semisimple and its irreducible representations are precisely the irreducible representations of $\mathbb{C}(q)\mathcal{H}_n(q)$ that pass to the quotient. That is, for $\lambda \in \mathcal{P}(n)$, V_λ is an irreducible representation of $\mathbb{C}(q)\mathrm{TL}_n(q)$ if and only if $g_{1,2}(\mathrm{v}_\mathcal{T}) = 0$ for every standard tableau $\mathcal{T}$ of shape λ. It is easy to see that the latter is equivalent to the trivial representation not being a direct summand of the restriction $\mathrm{Res}^{\mathfrak{S}_n}_{\langle s_1, s_2 \rangle}(E^\lambda)$, where E^λ is the irreducible representation of the symmetric group $\mathfrak{S}_n$ labelled by λ. Since the restriction from $\mathfrak{S}_n$ to $\mathfrak{S}_3 \cong \langle s_1, s_2 \rangle$ corresponds to the simple removal of boxes from the Young diagram of λ, and the trivial representation of $\mathfrak{S}_3$ is labelled by the partition (3), we obtain the following description of the irreducible representations of $\mathbb{C}(q)\mathrm{TL}_n(q)$:

Theorem 3.7 *Let $\lambda \in \mathcal{P}(n)$. We have that V_λ is an irreducible representation of $\mathbb{C}(q)\mathrm{TL}_n(q)$ if and only if the Young diagram of λ has at most two columns.*

Now, let $n \in \mathbb{N}$, and let $\underline{i} = (i_1, \dots, i_p)$ and $\underline{k} = (k_1, \dots k_p)$ be two p-tuples of non-negative integers, with $0 \leqslant p \leqslant n-1$. We denote by $\mathfrak{H}_n$ the set of pairs $(\underline{i}, \underline{k})$ such that

$$1 \leqslant i_1 < i_2 < \cdots < i_p \leqslant n-1 \quad \text{and} \quad i_j - k_j \geqslant 1 \quad \forall\, j = 1, \dots, p.$$

For $(\underline{i}, \underline{k}) \in \mathfrak{H}_n$, we set

$$g_{\underline{i},\underline{k}} := (g_{i_1} g_{i_1-1} \dots g_{i_1-k_1})(g_{i_2} g_{i_2-1} \dots g_{i_2-k_2}) \dots (g_{i_p} g_{i_p-1} \dots g_{i_p-k_p}) \in \mathcal{H}_n(q).$$

We take $g_{\emptyset,\emptyset}$ to be equal to 1. We have that the set

$$\mathcal{B}^{\mathrm{H}}_{1,n} = \{g_w \mid w \in \mathfrak{S}_n\} = \{g_{\underline{i},\underline{k}} \mid (\underline{i}, \underline{k}) \in \mathfrak{H}_n\}$$

is the standard basis of $\mathcal{H}_n(q)$ as a $\mathbb{C}[q, q^{-1}]$-module.

Further, let us denote by $\mathfrak{T}_n$ the subset of $\mathfrak{H}_n$ consisting of the pairs $(\underline{i}, \underline{k})$ such that

$$1 \leqslant i_1 < i_2 < \cdots < i_p \leqslant n-1 \quad \text{and} \quad 1 \leqslant i_1 - k_1 < i_2 - k_2 < \cdots < i_p - k_p \leqslant n-1.$$

Jones [45] has shown that the set

$$\mathcal{B}^{\mathrm{TL}}_{1,n} := \{g_{\underline{i},\underline{k}} \mid (\underline{i}, \underline{k}) \in \mathfrak{T}_n\}$$

is a basis of $\mathrm{TL}_n(q)$ as a $\mathbb{C}[q, q^{-1}]$-module. We have $|\mathcal{B}^{\mathrm{TL}}_n| = C_n$, where C_n is the n-th Catalan number, *i.e.,*

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{1}{n+1}\sum_{k=0}^{n}\binom{n}{k}^2.$$

3.5.1 The Yokonuma–Temperley–Lieb Algebra

Let $d \in \mathbb{N}^*$ and let $n \in \mathbb{N}$ with $n \geqslant 3$. The *Yokonuma–Temperley–Lieb algebra* $\mathrm{YTL}_{d,n}(q)$ is defined as the quotient of the Yokonuma–Hecke algebra $\mathrm{Y}_{d,n}(q)$ by the ideal $I_{d,n} := \langle\, g_{1,2} \,\rangle$.

Since the algebra $\mathbb{C}(q)\mathrm{Y}_{d,n}(q)$ is semisimple, the algebra $\mathbb{C}(q)\mathrm{YTL}_{d,n}(q)$ is also semisimple and its irreducible representations are precisely the irreducible representations of $\mathbb{C}(q)\mathrm{Y}_{d,n}(q)$ that pass to the quotient. That is, for $\boldsymbol{\lambda} \in \mathcal{P}(d,n)$, $V_{\boldsymbol{\lambda}}$ is an irreducible representation of $\mathbb{C}(q)\mathrm{YTL}_{d,n}(q)$ if and only if $g_{1,2}(\mathbf{v}_{\mathcal{T}}) = 0$ for every standard d-tableau $\mathcal{T}$ of shape $\boldsymbol{\lambda}$. It is easy to see that the latter is equivalent to the trivial representation not being a direct summand of the restriction $\mathrm{Res}^{G(d,1,n)}_{\langle s_1, s_2\rangle}(E^{\boldsymbol{\lambda}})$, where $E^{\boldsymbol{\lambda}}$ is the irreducible representation of the complex reflection group $G(d,1,n)$ labelled by $\boldsymbol{\lambda}$. Unfortunately, this restriction for $d > 1$ does not correspond to the simple removal of boxes from the Young diagram of $\boldsymbol{\lambda}$ (as in the symmetric group case), but it is controlled by the so-called *Littlewood–Richardson coefficients*. Using algebraic combinatorics, we obtain the following description of the irreducible representations of $\mathbb{C}(q)\mathrm{YTL}_{d,n}(q)$ [20, Theorem 3]:

Theorem 3.8 *Let $\boldsymbol{\lambda} = (\lambda^{(0)}, \ldots, \lambda^{(d-1)}) \in \mathcal{P}(d,n)$. We have that $V_{\boldsymbol{\lambda}}$ is an irreducible representation of $\mathbb{C}(q)\mathrm{YTL}_{d,n}(q)$ if and only if the Young diagram of $\boldsymbol{\lambda}$ has at most two columns in total, that is, $\sum_{i=0}^{d-1} \lambda_1^{(i)} \leqslant 2$.*

Using the fact that the algebra $\mathbb{C}(q)\mathrm{YTL}_{d,n}(q)$ is semisimple and the above description of its irreducible representations, we have been able to calculate the dimension of the Yokonuma–Temperley–Lieb algebra [20, Proposition 4]. We have

$$\dim_{\mathbb{C}(q)}(\mathbb{C}(q)\mathrm{YTL}_{d,n}(q)) = \frac{n(d^2-d)+d^2+d}{2}\, C_n - (d^2-d).$$

What is more, we have shown in [20] that $\mathrm{YTL}_{d,n}(q)$ is a free $\mathbb{C}[q,q^{-1}]$-module of rank equal to the dimension above. However, note that, even though the set

$$\mathcal{B}^{\mathrm{H}}_{d,n} = \left\{ t_1^{r_1} \ldots t_n^{r_n} g_{\underline{i},\underline{k}} \;\middle|\; (\underline{i},\underline{k}) \in \mathfrak{H}_n,\ 0 \leqslant r_j \leqslant d-1 \text{ for all } j = 1,2,\ldots,n \right\}$$

is a basis of $\mathrm{Y}_{d,n}(q)$ as a $\mathbb{C}[q,q^{-1}]$-module, the set

$$\mathcal{B}^{\mathrm{TL}}_{d,n} = \left\{ t_1^{r_1} \ldots t_n^{r_n} g_{\underline{i},\underline{k}} \;\middle|\; (\underline{i},\underline{k}) \in \mathfrak{T}_n,\ 0 \leqslant r_j \leqslant d-1 \text{ for all } j = 1,2,\ldots,n \right\}$$

is not a basis of $\mathrm{YTL}_{d,n}(q)$ as a $\mathbb{C}[q,q^{-1}]$-module, since $|\mathcal{B}^{\mathrm{TL}}_{d,n}| = d^n C_n$. The set $\mathcal{B}^{\mathrm{TL}}_{d,n}$ is simply a generating set for $\mathrm{YTL}_{d,n}(q)$, and we have managed to find a subset $\mathcal{B}^{\mathrm{YTL}}_{d,n}$ of $\mathcal{B}^{\mathrm{TL}}_{d,n}$ that is a basis of $\mathrm{YTL}_{d,n}(q)$ by proving the following remarkable property: Let $(\underline{i},\underline{k}) \in \mathfrak{T}_n$. We denote by $\mathcal{I}(g_{\underline{i},\underline{k}})$ the set (without repetition) of all indices of the g_j's appearing in $g_{\underline{i},\underline{k}}$, *i.e.,*

$$\mathcal{I}(g_{\underline{i},\underline{k}}) = \{i_1, i_1-1, \ldots, i_1-k_1, i_2, i_2-1, \ldots, i_2-k_2, \ldots, i_p, i_p-1, \ldots, i_p-k_p\}.$$

We define the *weight* of $g_{\underline{i},\underline{k}}$ to be $w_{\underline{i},\underline{k}} := |\mathcal{I}(g_{\underline{i},\underline{k}})|$. We then have [20, Propositions 9, 11, 12]:

$$|\{(r_1,\dots,r_n)\in\{0,\dots,d-1\}^n \mid t_1^{r_1}\dots t_n^{r_n} g_{\underline{i},\underline{k}} \in \mathcal{B}_{d,n}^{\mathrm{YTL}}\}| = 2^{n-w_{\underline{i},\underline{k}}-1}(d^2-d)+d-\delta_{w_{\underline{i},\underline{k}},0}(d^2-d),$$

where $\delta_{i,j}$ stands for Kronecker's delta (note that we have $w_{\underline{i},\underline{k}}=0$ if and only if $g_{\underline{i},\underline{k}}=1$). Thanks to this property, an explicit basis for $\mathrm{YTL}_{d,n}(q)$ as a $\mathbb{C}[q,q^{-1}]$-module is described in [20].

3.5.2 The Framisation of the Temperley–Lieb Algebra

Let $d\in\mathbb{N}^*$ and let $n\in\mathbb{N}$ with $n\geqslant 3$. The *Framisation of the Temperley–Lieb algebra* $\mathrm{FTL}_{d,n}(q)$ is defined as the quotient of the Yokonuma–Hecke algebra $\mathrm{Y}_{d,n}(q)$ by the ideal $J_{d,n} := \langle\, e_1e_2g_{1,2}\,\rangle$. We remark that $J_{d,n}$ can be also defined as the ideal generated by the element $\sum_{0\leqslant a,b\leqslant d-1} t_1^a t_2^b t_3^{-a-b}\, g_{1,2}$.

Again, since the algebra $\mathbb{C}(q)\mathrm{Y}_{d,n}(q)$ is semisimple, the algebra $\mathbb{C}(q)\mathrm{FTL}_{d,n}(q)$ is also semisimple and its irreducible representations are precisely the irreducible representations of $\mathbb{C}(q)\mathrm{Y}_{d,n}(q)$ that pass to the quotient. That is, for $\boldsymbol{\lambda}\in\mathcal{P}(d,n)$, $V_{\boldsymbol{\lambda}}$ is an irreducible representation of $\mathbb{C}(q)\mathrm{FTL}_{d,n}(q)$ if and only if $e_1e_2g_{1,2}(\mathbf{v}_{\mathcal{T}})=0$ for every standard d-tableau $\mathcal{T}$ of shape $\boldsymbol{\lambda}$. Using the formulas for the irreducible representations of $\mathbb{C}(q)\mathrm{Y}_{d,n}(q)$ given by Theorem 3.3, we obtain the following description of the irreducible representations of $\mathbb{C}(q)\mathrm{FTL}_{d,n}(q)$ [1, Theorem 3.10]:

Theorem 3.9 *Let $\boldsymbol{\lambda}=(\lambda^{(0)},\dots,\lambda^{(d-1)})\in\mathcal{P}(d,n)$. We have that $V_{\boldsymbol{\lambda}}$ is an irreducible representation of $\mathbb{C}(q)\mathrm{FTL}_{d,n}(q)$ if and only if the Young diagram of $\lambda^{(i)}$ has at most two columns for all $i=0,\dots,d-1$.*

Following the recipe of [32, Sect. 3], we have proved the following isomorphism theorem for $\mathrm{FTL}_{d,n}(q)$ [1, Theorem 4.3]:

Theorem 3.10 *There exists a $\mathbb{C}[q,q^{-1}]$ algebra isomorphism*

$$\psi_n : \mathrm{FTL}_{d,n}(q) \to \bigoplus_{\mu\in\mathrm{Comp}_d(n)} \mathrm{Mat}_{m_\mu}(\mathrm{TL}_{\mu_0}(q)\otimes\mathrm{TL}_{\mu_1}(q)\otimes\cdots\otimes\mathrm{TL}_{\mu_{d-1}}(q)),$$

where $\mathrm{Comp}_d(n)$ and m_μ are as defined in (3.3.15) and (3.3.16), and we take $\mathrm{TL}_n(q)\cong\mathcal{H}_n(q)$ for $n<3$.

We deduce that the following set is a basis of $\mathrm{FTL}_{d,n}(q)$ as a $\mathbb{C}[q,q^{-1}]$-module [1, Proposition 4.4]:

$$\left\{\psi_n^{-1}(b_0^\mu b_1^\mu\dots b_{d-1}^\mu M_{k,l}^\mu) \mid \mu\in\mathrm{Comp}_d(n),\, b_i^\mu\in\mathcal{B}_{1,\mu_i}^{\mathrm{TL}} \text{ for all } i=0,\dots,d-1,\, 1\leqslant k,l\leqslant m_\mu\right\},$$

where $M_{k,l}^{\mu}$ denotes the elementary $m_\mu \times m_\mu$ matrix with 1 in position (k,l). In particular, $\mathrm{FTL}_{d,n}(q)$ is a free $\mathbb{C}[q,q^{-1}]$-module of rank

$$\sum_{\mu \in \mathrm{Comp}_d(n)} m_\mu^2 \, C_{\mu_0} C_{\mu_1} \cdots C_{\mu_{d-1}}.$$

3.5.3 The Complex Reflection Temperley–Lieb Algebra

Let $d \in \mathbb{N}^*$ and let $n \in \mathbb{N}$ with $n \geqslant 3$. The *Complex Reflection Temperley–Lieb algebra* $\mathrm{CTL}_{d,n}(q)$ is defined as the quotient of the Yokonuma–Hecke algebra $\mathrm{Y}_{d,n}(q)$ by the ideal $K_{d,n} := \langle \sum_{s=0}^{d-1} t_1^s e_1 e_2 g_{1,2} \rangle$. We remark that $K_{d,n}$ can be also viewed as the ideal generated by the element $\sum_{0 \leqslant a,b,c \leqslant d-1} t_1^a t_2^b t_3^c \, g_{1,2}$.

Once more, the algebra $\mathbb{C}(q)\mathrm{CTL}_{d,n}(q)$ is semisimple and, for $\boldsymbol{\lambda} \in \mathcal{P}(d,n)$, $V_{\boldsymbol{\lambda}}$ is an irreducible representation of $\mathbb{C}(q)\mathrm{CTL}_{d,n}(q)$ if and only if $\sum_{s=0}^{d-1} t_1^s e_1 e_2 g_{1,2}(\mathbf{v}_{\mathcal{T}}) = 0$ for every standard d-tableau $\mathcal{T}$ of shape $\boldsymbol{\lambda}$. Using the formulas for the irreducible representations of $\mathbb{C}(q)\mathrm{Y}_{d,n}(q)$ given by Theorem 3.3, we obtain the following description of the irreducible representations of $\mathbb{C}(q)\mathrm{CTL}_{d,n}(q)$ [1, Theorem 5.3]:

Theorem 3.11 *Let $\{\xi_0, \xi_1, \ldots, \xi_{d-1}\}$ be the set of all d-th roots of unity (ordered arbitrarily) as in Theorem 3.3. Let $i_0 \in \{0, \ldots, d-1\}$ be such that $\xi_{i_0} = 1$, and let $\boldsymbol{\lambda} = (\lambda^{(0)}, \ldots, \lambda^{(d-1)}) \in \mathcal{P}(d,n)$. We have that $V_{\boldsymbol{\lambda}}$ is an irreducible representation of $\mathbb{C}(q)\mathrm{CTL}_{d,n}(q)$ if and only if the Young diagram of $\lambda^{(i_0)}$ has at most two columns.*

Following the recipe of [32, Sect. 3], we have proved the following isomorphism theorem for $\mathrm{CTL}_{d,n}(q)$ [1, Theorem 5.8]:

Theorem 3.12 *There exists a $\mathbb{C}[q,q^{-1}]$ algebra isomorphism*

$$\overline{\psi}_n : \mathrm{CTL}_{d,n}(q) \to \bigoplus_{\mu \in \mathrm{Comp}_d(n)} \mathrm{Mat}_{m_\mu}(\mathrm{TL}_{\mu_0}(q) \otimes \mathcal{H}_{\mu_1}(q) \otimes \mathcal{H}_{\mu_2}(q) \otimes \cdots \otimes \mathcal{H}_{\mu_{d-1}}(q)),$$

where $\mathrm{Comp}_d(n)$ *and* m_μ *are as defined in* (3.3.15) *and* (3.3.16)*, and we take* $\mathrm{TL}_n(q) \cong \mathcal{H}_n(q)$ *for* $n < 3$.

We deduce that the following set is a basis of $\mathrm{CTL}_{d,n}(q)$ as a $\mathbb{C}[q,q^{-1}]$-module [1, Proposition 5.9]:

$$\left\{ \overline{\psi}_n^{-1}(b_0^\mu b_1^\mu \ldots b_{d-1}^\mu M_{k,l}^\mu) \mid \mu \in \mathrm{Comp}_d(n), b_0^\mu \in \mathcal{B}_{1,\mu_0}^{\mathrm{TL}}, b_i^\mu \in \mathcal{B}_{1,\mu_i}^{\mathrm{H}} \text{ for all } i = 1, \ldots, d-1, 1 \leqslant k,l \leqslant m_\mu \right\},$$

where $M_{k,l}^{\mu}$ denotes the elementary $m_\mu \times m_\mu$ matrix with 1 in position (k,l). In particular, $\mathrm{CTL}_{d,n}(q)$ is a free $\mathbb{C}[q,q^{-1}]$-module of rank

$$\sum_{\mu \in \mathrm{Comp}_d(n)} m_\mu^2 \, C_{\mu_0} \mu_1! \ldots \mu_{d-1}!.$$

References

1. Chlouveraki, M., Pouchin, G.: Representation theory and an isomorphism theorem for the Framisation of the Temperley-Lieb algebra. Math. Z. **285**(3), 1357–1380 (2017)
2. Jones, V.F.R.: A polynomial invariant for knots via von Neumann algebra. Bull. Am. Math. Soc. (N.S.) **12** (1985), 103–111
3. Jones, V.F.R.: Hecke algebra representations of braid groups and link polynomials. Annals Math. **126**(2), 335–388 (1987)
4. Juyumaya, J., Lambropoulou, S.: On the framization of knot algebras, to appear in New Ideas in Low-dimensional Topology. In Kauffman, L., Manturov, V. (eds) Series of Knots and Everything. WS
5. Yokonuma, T.: Sur la structure des anneaux de Hecke d'un groupe de Chevalley fini. Comptes. R. Acad. Sci. Paris Ser. I Math. **264** (1967) 344–347
6. Chlouveraki, M., Poulain d'Andecy, L.: Representation theory of the Yokonuma-Hecke algebra. Adv. Math. **259**, 134–172 (2014)
7. Juyumaya, J.: Sur les nouveaux générateurs de l'algèbre de Hecke H(G, U,1). J. Algebr. **204**, 49–68 (1998)
8. Juyumaya, J.: Markov trace on the Yokonuma-Hecke algebra. J. Knot Theory Ramif. **13**, 25–39 (2004)
9. Juyumaya, J., Kannan, S.: Braid relations in the Yokonuma-Hecke algebra. J. Algebr. **239**, 272–297 (2001)
10. Ariki, S., Koike, K.: A Hecke algebra of $(\mathbb{Z}/r\mathbb{Z}) \wr S_n$ and construction of its irreducible representations. Adv. Math. **106**, 216–243 (1994)
11. Juyumaya, J., Lambropoulou, S.: p -adic framed braids. Topol. Appl. **154**, 1804–1826 (2007)
12. Juyumaya, J., Lambropoulou, S.: p -adic framed braids II. Adv. Math. **234**, 149–191 (2013)
13. Juyumaya, J., Lambropoulou, S.: An adelic extension of the Jones polynomial. In: Banagl, M., Vogel, D. (eds.) The Mathematics Of Knots, Contributions in the Mathematical and Computational Sciences, vol. 1, Springer, Berlin
14. Juyumaya, J., Lambropoulou, S.: An invariant for singular knots. J. Knot Theory Ramif. **18**(6), 825–840 (2009)
15. Chlouveraki, M., Juyumaya, J., Karvounis, K., Lambropoulou, S.: Identifying the invariants for classical knots and links from the Yokonuma–Hecke algebras, preprint. arXiv:1505.06666
16. Chlouveraki, M., Poulain d'Andecy, L.: Markov traces on affine and cyclotomic Yokonuma–Hecke algebras, Int. Math. Res. Notices (2015). https://doi.org/10.1093/imrn/rnv257
17. Chlouveraki, M., Sécherre, V.: The affine Yokonuma-Hecke algebra and the pro-p-Iwahori Hecke algebra. Math. Res. Lett. **23**(3), 707–718 (2016)
18. Goundaroulis, D., Juyumaya, J., Kontogeorgis, A., Lambropoulou, S.: The Yokonuma–Temperley–Lieb algebra. Banach Cent. Pub. **103** (2014)
19. Goundaroulis, D., Juyumaya, J., Kontogeorgis, A., Lambropoulou, S.: Framization of the Temperley–Lieb algebra, to appear in Mathematical Research Letters. arXiv:1304.7440
20. Chlouveraki, M., Pouchin, G.: Determination of the representations and a basis for the Yokonuma-Temperley-Lieb algebra. Algebr. Represent. Theory **18**(2), 421–447 (2015)
21. Geck, M., Lambropoulou, S.: Markov traces and knot invariants related to Iwahori-Hecke algebras of type B. J. Reine Angew. Math. **482**, 191–213 (1997)
22. Lambropoulou, S.: Solid torus links and Hecke algebras of B-type. In: Yetter, D. N. (ed.) Proceedings of the Conference on Quantum Topology, pp. 225–245, World Scientific Press (1994)
23. Lambropoulou, S.: Knot theory related to generalized and cyclotomic Hecke algebras of type B. J. Knot Theory Ramif. **8**(5), 621–658 (1999)
24. Geck, M., Pfeiffer, G.: Characters of finite Coxeter groups and Iwahori–Hecke algebras, London Mathematical Society Monographs, New Series 21, Oxford University Press, New York (2000)
25. Thiem, N.: Unipotent Hecke algebras: the structure, representation theory, and combinatorics, Ph.D. Thesis, University of Wisconsin (2004)

26. Thiem, N.: Unipotent Hecke algebras of $GL_n(\mathbb{F}_q)$. J. Algebr. **284**, 559–577 (2005)
27. Thiem, N.: A skein-like multiplication algorithm for unipotent Hecke algebras. Trans. Am. Math. Soc. **359**(4), 1685–1724 (2007)
28. Ariki, S.: On the semi-simplicity of the Hecke algebra of $(\mathbb{Z}/r\mathbb{Z}) \wr \mathfrak{S}_n$. J. Algebr. **169**, 216–225 (1994)
29. Steinberg, R.: A geometric approach to the representations of the full linear group over a Galois field. Trans. Amer. Math. Soc. **71**, 274–282 (1951)
30. Chlouveraki, M., Jacon, N.: Schur elements for the Ariki-Koike algebra and applications. J. Algebr. Comb. **35**(2), 291–311 (2012)
31. Lusztig, G.: Character sheaves on disconnected groups VII. Represent. Theory **9**, 209–266 (2005)
32. Jacon, N., Poulain d'Andecy, L.: An isomorphism Theorem for Yokonuma–Hecke algebras and applications to link invariants, to appear in Math. Z
33. Espinoza, J., Ryom-Hansen, S.: Cell structures for the Yokonuma–Hecke algebra and the algebra of braids and ties, preprint. arXiv: 1506.00715
34. Vignéras, M.-F.: The pro-p-Iwahori Hecke algebra of a reductive p-adic group I, to appear in Compositio mathematica (2015)
35. Vignéras, M.-F.: The pro-p-Iwahori Hecke algebra of a reductive p-adic group II. Münster J. Math. **7**, 363–379 (2014)
36. Vignéras, M.-F.: The pro-p-Iwahori Hecke algebra of a reductive p-adic group III, to appear in Journal of the Institute of Mathematics of Jussieu (2015)
37. Cui, W.: Affine cellularity of affine Yokonuma–Hecke algebras, preprint. arXiv:1510.02647
38. Poulain d'Andecy, L.: Invariants for links from classical and affine Yokonuma–Hecke algebras, preprint. arXiv:1602.05429
39. Rostam, S.: Cyclotomic Yokonuma–Hecke algebras are cyclotomic quiver Hecke algebras. Adv. Math. **311**, 662–729 (2017)
40. Ogievetsky, O., Poulain d'Andecy, L.: Induced representations and traces for chains of affine and cyclotomic Hecke algebras. J. Geom. Phys. **87**, 354–372 (2015)
41. Geck, M., Iancu, L., Malle, G.: Weights of Markov traces and generic degrees, Indag. Math. (N.S.) **11**(3) (2000), 379–397
42. Mathas, A.: Matrix units and generic degrees for the Ariki-Koike algebras. J. Algebr. **281**, 695–730 (2004)
43. Malle, G., Mathas, A.: Symmetric cyclotomic Hecke algebras. J. Algebr. **205**(1), 275–293 (1998)
44. Temperley, N., Lieb, E.: Relations between the 'Percolation' and 'Colouring' Problem and other Graph-Theoretical Problems Associated with Regular Planar Lattices: Some Exact Results for the 'Percolation' Problem. Proc. Royal Soc. Ser. A **322**, 251–280 (1971)
45. Jones, V.F.R.: Index for subfactors. Invent. Math. **72**, 1–25 (1983)

Chapter 4
Invariants for Links from Classical and Affine Yokonuma–Hecke Algebras

Loic Poulain d'Andecy

Abstract We present a construction of invariants for links using an isomorphism theorem for affine Yokonuma–Hecke algebras. The isomorphism relates affine Yokonuma–Hecke algebras with usual affine Hecke algebras. We use it to construct a large class of Markov traces on affine Yokonuma–Hecke algebras, and in turn, to produce invariants for links in the solid torus. By restriction, this construction contains the construction of invariants for classical links from classical Yokonuma–Hecke algebras. In general, the obtained invariants form an infinite family of 3-variable polynomials. As a consequence of the construction via the isomorphism, we reduce the number of invariants to study, given the number of connected components of a link. In particular, if the link is a classical link with N components, we show that N invariants generate the whole family.

4.1 Introduction

1. The Yokonuma–Hecke algebras (of type GL), denoted $Y_{d,n}$, have been used by J. Juyumaya and S. Lambropoulou to construct invariants for various types of links, in the same spirit as the construction of the HOMFLYPT polynomial from usual Hecke algebras. We refer to [1] and references therein. In particular, the algebras $Y_{d,n}$ provide invariants for classical links and the natural question was to decide if these invariants were equivalent, or not, to the HOMFLYPT polynomial. This study culminated in the recent discovery [1] that these invariants are actually topologically stronger than the HOMFLYPT polynomial (i.e. they distinguish more links). We refer also to [1, Appendix B] and [2] for a description of the invariants in terms of HOMFLYPT polynomials and linking numbers of sublinks.

L. Poulain d'Andecy (✉)
UFR Sciences exactes et naturelles, Laboratoire de Mathématiques,
Université de Reims Champagne-Ardenne,
EA 4535 Moulin de la Housse, BP 1039, 51100 Reims, France
e-mail: loic.poulain-dandecy@univ-reims.fr

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_4

In [3], another approach to study invariants coming from Yokonuma–Hecke algebras was developed. The starting point was the fact that the algebra $Y_{d,n}$ is isomorphic to a direct sum of matrix algebras with coefficients in tensor products of usual Hecke algebras. This allowed an explicit construction of Markov traces on $\{Y_{d,n}\}_{n\geq 1}$ from the known Markov trace on Hecke algebras (on Hecke algebras, there is a unique Markov trace up to normalisation, and it gives the HOMFLYPT polynomial). In addition to its usefulness for the construction of Markov traces, the approach via the isomorphism also helps to study the resulting invariants. Indeed some properties of the invariants follow quite immediately from a precise understanding of the isomorphism (see paragraph **4** below).

Independently of which approach is used, another ingredient was added in [3]: a third parameter in the invariants. While the first two parameters come from the algebra $Y_{d,n}$, this third parameter γ has its origin in the framed braid group, and corresponds to a certain degree of freedom one has when going from the framed braid group to the algebra $Y_{d,n}$. More precisely, we can deform the standard surjective morphism from the framed braid group algebra to its quotient $Y_{d,n}$ into a family of morphisms (depending on γ) respecting the braid relations and the Markov conditions. Another way of interpreting the parameter γ is that it modifies the quadratic relation satisfied by the generators of $Y_{d,n}$. Its existence explains (or is reflected in) the fact that different presentations for $Y_{d,n}$ were used before. Juyumaya–Lambropoulou invariants correspond to certain specialisations of this parameter γ, depending on the chosen presentation. So the parameter γ unifies every possible choices and yields more general invariants. It is indicated in [1, Remark 8.5] that changing the presentation seems to give a non-equivalent topological invariant.

2. In this paper, we consider the affine Yokonuma–Hecke algebras (of type GL), denoted $\widehat{Y}_{d,n}$. They were introduced in [4] in connections with the representation theory and the Jucys–Murphy elements of the classical Yokonuma–Hecke algebras. Our main goal here is to generalise for $\widehat{Y}_{d,n}$ the whole approach to link invariants via the isomorphism theorem. The invariants are in general for links in the solid torus. The classical links are naturally contained in the solid torus links and, restricted to them, the obtained invariants correspond to the invariants obtained in [3] from $Y_{d,n}$ (naturally seen as a subalgebra of $\widehat{Y}_{d,n}$). Specialising the parameter γ, we identify the Juyumaya–Lambropoulou invariants among them. For those invariants, we emphasize that we recover some known results [1] by a different method and furthermore obtain some new results already in this particular case.

We start with an isomorphism between the algebra $\widehat{Y}_{d,n}$ and a direct sum of matrix algebras with coefficients in tensor products of affine Hecke algebras. As done in [5], the isomorphism can be proved repeating the same arguments as for $Y_{d,n}$ (see [3] where the proof for $Y_{d,n}$ is presented, as a particular case of a more general result by G. Lusztig [6, Sect. 34]). Here we sketch a short different proof for $\widehat{Y}_{d,n}$ using the known result for $Y_{d,n}$. We also prove the analogous theorem for the cyclotomic quotients of $\widehat{Y}_{d,n}$ (with Ariki–Koike algebras replacing affine Hecke algebras). Useful for concrete use, the formulas for the generators are simple and given explicitly.

Concerning links, S. Lambropoulou constructed invariants, analogues of the HOMFLYPT polynomial, for links in the solid torus from affine Hecke algebras

[7]. Then, it was explained in [8] how to obtain invariants for those links from the algebras $\widehat{Y}_{d,n}$, unifying the methods of J. Juyumaya and S. Lambropoulou for $Y_{d,n}$ and the construction of S. Lambropoulou for affine Hecke algebras. Due to the recent results of [1], it is expected that the invariants obtained from $\widehat{Y}_{d,n}$ are stronger than the ones obtained from affine Hecke algebras.

Here we follow the alternative approach which uses the isomorphism to construct Markov traces on the family of algebras $\{\widehat{Y}_{d,n}\}_{n\geq 1}$. To sum up, the Markov traces are constructed and can be calculated with the following steps: for an element of $\widehat{Y}_{d,n}$, apply first the isomorphic map to obtain an element of the direct sum of matrix algebras; then, for each matrix, apply the usual trace which results in an element of a tensor product of affine Hecke algebras; finally apply a tensor product of Markov traces on affine Hecke algebras. Our result consists in obtaining the compatibility conditions relating the Markov traces appearing in different matrix algebras so that the preceding procedure eventually results in a genuine Markov trace on $\{\widehat{Y}_{d,n}\}_{n\geq 1}$.

With the definition used here, for a given $d > 0$, the set of Markov traces on $\{\widehat{Y}_{d,n}\}_{n\geq 1}$ forms a vector space. From the isomorphism, a set of distinguished Markov traces appears naturally, which spans the set of all Markov traces constructed here. Thus, our study of Markov traces (and of invariants) is reduced to the study of these "basic" Markov traces (and of the corresponding "basic" invariants). It turns out that these basic Markov traces are indexed, for a given $d > 0$, by the non-empty subsets $S \subset \{1, \ldots, d\}$ together with a choice, denoted formally by $\boldsymbol{\tau}$, of $|S|$ arbitrary Markov traces on affine Hecke algebras. We note that if we restrict to $Y_{d,n}$, the parameter $\boldsymbol{\tau}$ disappears and the basic Markov traces on $\{Y_{d,n}\}_{n\geq 1}$ are indexed, for a given $d > 0$, only by the non-empty subsets $S \subset \{1, \ldots, d\}$. This recovers a result of [3].

3. Throughout the paper, we intended to give in details the connections between the two approaches, so that one would be able to pass easily from one to the other. This will allow in particular to specialise and translate all our results on the invariants to Juyumaya–Lambropoulou invariants as well.

Roughly speaking, J. Juyumaya and S. Lambropoulou constructed invariants from $Y_{d,n}$ in two steps [9]. The same approach was followed in [8] for $\widehat{Y}_{d,n}$. First a certain trace map, analogous to the Ocneanu trace and satisfying a certain positive Markov condition, was constructed. Then a rescaling procedure was implemented, in order to produce genuine invariants. The rescaling procedure amounts to two things: a renormalisation of the generators and a renormalisation, depending on n, of the trace. In the approach presented here, the first step is included from the beginning in a more general quadratic relation for the generators. The second step is already included in the definition of a Markov trace, namely that it is a family, on n, of trace maps satisfying the two Markov conditions. As a consequence, to obtain invariants here, one directly applies the Markov trace and no rescaling procedure is needed.

For the comparison, our first task is to explain that Juyumaya–Lambropoulou approach is equivalent to considering certain Markov traces (with the definition used here) and to relate their variables with the parameters considered here. Then we need to identify these Markov traces in terms of the ones constructed via the isomorphism theorem. We obtain finally the explicit decomposition of these Markov traces in terms of the basic Markov traces indexed by $S \subset \{1, \ldots, d\}$ and $\boldsymbol{\tau}$ as above.

In particular, for $Y_{d,n}$, this results in an explicit formula for the Juyumaya–Lambropoulou invariants, as studied in [1], in terms of the basic invariants constructed here. We note that, in this case, the parameter $\boldsymbol{\tau}$ is not present, and that Juyumaya–Lambropoulou invariants are also parametrised, for a given $d > 0$, by non-empty subsets of $\{1, \dots, d\}$. Nevertheless, they do not coincide with the basic invariants and the comparison formula is not trivial (see Formulas (4.27) in Sect. 1.5). In general, for $\widehat{Y}_{d,n}$, we obtain the expression of the invariants constructed in [8] in terms of the basic invariants constructed here.

Concerning the third parameter γ, we recall that it was not present in the previous approach. Actually, one need to specialise it to a certain value in our invariants to recover the Juyumaya–Lambropoulou invariants. The two different presentations of $Y_{d,n}$ that were used, as in [1], correspond to two different values of γ that we give explicitly. Similarly, for $\widehat{Y}_{d,n}$, the invariants constructed in [8] correspond to a certain specialisation of γ.

4. We conclude this introduction by describing the main properties obtained for the invariants. As explained before, they follow quite directly from a precise understanding of the isomorphism, and are expressed easily in terms of the basic invariants defined here. The main results are:

- for $d > 0$ and a non-empty subset $S \subset \{1, \dots, d\}$, the corresponding invariants coincide with invariants corresponding to $d' = |S|$ and the full set $\{1, \dots, d'\}$. Therefore, we only have to consider the full sets $\{1, \dots, d\}$ for different $d > 0$.
- further, given a number N of connected components of a link, the invariants corresponding to $\{1, \dots, d\}$ are zero if $d > N$. So, given N, we only have to consider $d = 1, \dots, N$.

Moreover, with the comparison results explained in paragraph 3, it is easy to deduce the similar properties for invariants obtained via Juyumaya–Lambropoulou approach. The first item remains true as it is. The second item results in an explicit formula expressing, if $d > N$, the invariants corresponding to $\{1, \dots, d\}$ in terms of the invariants corresponding to $\{1, \dots, d'\}$ with $d' \leq N$. Specialising γ to the appropriate values and restricting to classical links, we recover with the first item a result of [1]. The second item in this case was proved only for $N \leq 2$ also in [1].

4.2 Affine Yokonuma–Hecke Algebras

Let $d, n \in \mathbb{Z}_{>0}$ and u and v be indeterminates. We work over the ring $\mathbb{C}[u^{\pm 1}, v]$. The properties of the affine Yokonuma–Hecke algebras recalled here can be found in [8].

4.2.1 Definitions

We use $\mathfrak{S}_n$ to denote the symmetric group on n elements, and s_i to denote the transposition $(i, i+1)$. The affine Yokonuma–Hecke algebra $\widehat{Y}_{d,n}$ is generated by elements

$$g_1, \ldots, g_{n-1}, X_1^{\pm 1}, t_1, \ldots, t_n,$$

subject to the following defining relations (4.1)–(4.3):

$$\begin{aligned} g_i g_j &= g_j g_i && \text{for } i, j = 1, \ldots, n-1 \text{ such that } |i-j| > 1, \\ g_i g_{i+1} g_i &= g_{i+1} g_i g_{i+1} && \text{for } i = 1, \ldots, n-2, \\ X_1 g_1 X_1 g_1 &= g_1 X_1 g_1 X_1 & & \\ X_1 g_i &= g_i X_1 && \text{for } i = 2, \ldots, n-1, \end{aligned} \tag{4.1}$$

$$\begin{aligned} t_i t_j &= t_j t_i && \text{for } i, j = 1, \ldots, n, \\ g_i t_j &= t_{s_i(j)} g_i && \text{for } i = 1, \ldots, n-1 \text{ and } j = 1, \ldots, n, \\ t_j^d &= 1 && \text{for } j = 1, \ldots, n, \\ X_1 t_j &= t_j X_1 && \text{for } j = 1, \ldots, n, \end{aligned} \tag{4.2}$$

$$g_i^2 = u^2 + v\, e_i\, g_i \quad \text{for } i = 1, \ldots, n-1, \tag{4.3}$$

where $e_i := \dfrac{1}{d} \displaystyle\sum_{1 \le s \le d} t_i^s t_{i+1}^{-s}$. The elements e_i are idempotents and we have:

$$g_i^{-1} = u^{-2} g_i - u^{-2} v\, e_i \qquad \text{for all } i = 1, \ldots, n-1. \tag{4.4}$$

Let $w \in \mathfrak{S}_n$ and let $w = s_{i_1} s_{i_2} \ldots s_{i_r}$ be a reduced expression for w. Since the generators g_i of $\widehat{Y}_{d,n}$ satisfy the same braid relations as the generators s_i of $\mathfrak{S}_n$, Matsumoto's lemma implies that the following element does not depend on the reduced expression of w:

$$g_w := g_{i_1} g_{i_2} \cdots g_{i_r}\ . \tag{4.5}$$

Elements $X_2, \ldots, X_n$ of $\widehat{Y}_{d,n}$ are defined inductively by

$$X_{i+1} := u^{-2} g_i X_i g_i \quad \text{for } i = 1, \ldots, n-1. \tag{4.6}$$

The elements $X_1, \ldots, X_n$ commute with each other. They also commute with the generators $t_1, \ldots, t_n$ and they satisfy $g_j X_i = X_i g_j$ if $i \neq j, j+1$.

For $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{Z}^n$, we set $X^\lambda := X_1^{\lambda_1} \ldots X_n^{\lambda_n}$. The following set of elements forms a basis of $\widehat{Y}_{d,n}$:

$$\{\, t_1^{a_1} \ldots t_n^{a_n} X^\lambda g_w \mid a_1, \ldots, a_n \in \{1, \ldots, d\},\ \lambda \in \mathbb{Z}^n,\ w \in \mathfrak{S}_n \,\}\,. \tag{4.7}$$

This fact has the following consequences:

- Recall that the Yokonuma–Hecke algebra $Y_{d,n}$ is presented by generators $g_1, \dots, g_{n-1}, t_1, \dots, t_n$ and defining relations those in (4.1)–(4.3) which do not involve the generator X_1. We have that $Y_{d,n}$ is isomorphic to the subalgebra of $\widehat{Y}_{d,n}$ generated by $g_1, \dots, g_{n-1}, t_1, \dots, t_n$ (hence the common names for the generators).
- In particular, the commutative subalgebra $\mathscr{T}_{d,n} := \langle t_1, \dots, t_n \rangle$ of $\widehat{Y}_{d,n}$ generated by $t_1, \dots, t_n$ is isomorphic to the group algebra of $(\mathbb{Z}/d\mathbb{Z})^n$.
- By definition, the affine Hecke algebra (of type GL) is $\widehat{H}_n := \widehat{Y}_{1,n}$. We have, for any $d > 0$, that the quotient of $\widehat{Y}_{d,n}$ by the relations $t_j = 1$, $j = 1, \dots, n$, is isomorphic to $\widehat{H}_n$. We denote by $\overline{\cdot}$ the corresponding surjective morphism from $\widehat{Y}_{d,n}$ to $\widehat{H}_n$, and the generators of $\widehat{H}_n$ are denoted $\overline{g}_1, \dots, \overline{g}_{n-1}, \overline{X}_1^{\pm 1}$.
- The subalgebra of $\widehat{H}_n$ generated by $\overline{g}_1, \dots, \overline{g}_{n-1}$ is the usual Hecke algebra, denoted H_n. We also have $H_n = Y_{1,n}$.

4.2.2 Compositions of n

Let $\mathrm{Comp}_d(n)$ be the set of *d-compositions* of n, that is the set of d-tuples $\mu = (\mu_1, \dots, \mu_d) \in \mathbb{Z}_{\geq 0}^d$ such that $\sum_{1 \leq a \leq d} \mu_a = n$. We denote $\mu \vDash_d n$.

For $\mu \vDash_d n$, the Young subgroup $\mathfrak{S}^\mu$ is the subgroup $\mathfrak{S}_{\mu_1} \times \cdots \times \mathfrak{S}_{\mu_d}$ of $\mathfrak{S}_n$, where $\mathfrak{S}_{\mu_1}$ acts on the letters $\{1, \dots, \mu_1\}$, $\mathfrak{S}_{\mu_2}$ acts on the letters $\{\mu_1 + 1, \dots, \mu_2\}$, and so on. The subgroup $\mathfrak{S}^\mu$ is generated by the transpositions s_i with $i \in I_\mu := \{1, \dots, n-1\} \setminus \{\mu_1, \mu_1 + \mu_2, \dots, \mu_1 + \dots + \mu_{d-1}\}$.

We denote by $\widehat{H}^\mu$ the algebra $\widehat{H}_{\mu_1} \otimes \dots \otimes \widehat{H}_{\mu_d}$ (by convention $\widehat{H}_0 := \mathbb{C}[u^{\pm 1}, v]$). It is isomorphic to the subalgebra of $\widehat{H}_n$ generated by $\overline{X}_1^{\pm 1}, \dots, \overline{X}_n^{\pm 1}$ and $\overline{g}_i$, with $i \in I_\mu$, and is a free submodule with basis $\{\overline{X}^\lambda \overline{g}_w \mid \lambda \in \mathbb{Z}^n,\ w \in \mathfrak{S}^\mu\}$.

Similarly, we have a subalgebra $H^\mu \cong H_{\mu_1} \otimes \dots \otimes H_{\mu_d}$ of the Hecke algebra H_n. It is naturally a subalgebra of $\widehat{H}^\mu$ (generated only by $\overline{g}_i$, with $i \in I_\mu$).

For $\mu \vDash_d n$, let m_μ be the index of the Young subgroup $\mathfrak{S}^\mu$ in $\mathfrak{S}_n$, that is,

$$m_\mu := \frac{n!}{\mu_1! \mu_2! \dots \mu_d!}\,. \tag{4.8}$$

We define the *socle* $\overline{\mu}$ of a d-composition μ by

$$\overline{\mu}_a = \begin{cases} 1 \text{ if } \mu_a \geq 1, \\ 0 \text{ if } \mu_a = 0, \end{cases} \qquad \text{for } a = 1, \dots, d. \tag{4.9}$$

The composition $\overline{\mu}$ belongs to $\mathrm{Comp}_d(N)$ where N is the number of non-zero parts in μ. We denote by Soc_d the set of all socles of d-compositions, or in other words, Soc_d is the set of d-compositions whose parts belong to $\{0, 1\}$. We note that there is a one-to-one correspondence between the set Soc_d and the set of non-empty subsets of $\{1, \dots, d\}$, given by

$$\{1,\dots,d\} \supset S \iff \mu^S \in \mathrm{Soc}_d\,, \quad \text{where } \mu_a^S = \begin{cases} 1 \text{ if } a \in S, \\ 0 \text{ if } a \notin S. \end{cases} \tag{4.10}$$

4.2.3 Characters of $\mathscr{T}_{d,n}$

Let $\{\xi_1,\dots,\xi_d\}$ be the set of roots of unity of order d. A complex character χ of the group $(\mathbb{Z}/d\mathbb{Z})^n$ is characterised by the choice of $\chi(t_j) \in \{\xi_1,\dots,\xi_d\}$ for each $j = 1,\dots,n$. We denote by $\mathrm{Irr}(\mathscr{T}_{d,n})$ the set of complex characters of $(\mathbb{Z}/d\mathbb{Z})^n$, extended to the subalgebra $\mathscr{T}_{d,n} = \langle t_1,\dots,t_n\rangle$ of $\widehat{Y}_{d,n}$.

For each $\chi \in \mathrm{Irr}\left(\mathscr{T}_{d,n}\right)$, we denote by E_χ the primitive idempotent of $\mathscr{T}_{d,n}$ associated to χ. Then the set $\{E_\chi \mid \chi \in \mathrm{Irr}(\mathscr{T}_{d,n})\}$ is a basis of $\mathscr{T}_{d,n}$. Therefore, from the basis (4.7) of $\widehat{Y}_{d,n}$, we obtain the following other basis of $\widehat{Y}_{d,n}$:

$$\{E_\chi X^\lambda g_w \mid \chi \in \mathrm{Irr}(\mathscr{T}_{d,n})\,,\ \lambda \in \mathbb{Z}^n\,,\ w \in \mathfrak{S}_n\}\,. \tag{4.11}$$

Permutations π_χ.

Let $\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})$. For $a \in \{1,\dots,d\}$, we let μ_a be the number of elements $j \in \{1,\dots,n\}$ such that $\chi(t_j) = \xi_a$. Then the sequence $(\mu_1,\dots,\mu_d)$ is a d-composition of n which we denote by $\mathrm{Comp}(\chi)$.

For a given $\mu \vDash_d n$, we consider a particular character $\chi_0^\mu \in \mathrm{Irr}(\mathscr{T}_{d,n})$ such that $\mathrm{Comp}(\chi_0^\mu) = \mu$. The character χ_0^μ is defined by

$$\begin{cases} \chi_0^\mu(t_1) & = \dots = & \chi_0^\mu(t_{\mu_1}) & = \xi_1\,, \\ \chi_0^\mu(t_{\mu_1+1}) & = \dots = & \chi_0^\mu(t_{\mu_1+\mu_2}) & = \xi_2\,, \\ \vdots & \vdots\ \vdots\ \vdots & \vdots & \vdots\ \vdots \\ \chi_0^\mu(t_{\mu_1+\cdots+\mu_{d-1}+1}) & = \dots = & \chi_0^\mu(t_n) & = \xi_d\,. \end{cases} \tag{4.12}$$

The symmetric group $\mathfrak{S}_n$ acts on the set $\mathrm{Irr}(\mathscr{T}_{d,n})$ by the formula $w(\chi)(t_i) = \chi(t_{w^{-1}(i)})$. The stabilizer of χ_0^μ under the action of $\mathfrak{S}_n$ is the Young subgroup $\mathfrak{S}^\mu$. In each left coset in $\mathfrak{S}_n/\mathfrak{S}^\mu$, there is a unique representative of minimal length. So, for any $\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})$ such that $\mathrm{Comp}(\chi) = \mu$, we define a permutation $\pi_\chi \in \mathfrak{S}_n$ by requiring that π_χ is the element of minimal length such that:

$$\pi_\chi(\chi_0^\mu) = \chi\,. \tag{4.13}$$

4.3 Isomorphism Theorems

We present isomorphism theorems for the algebras $\widehat{Y}_{d,n}$ and their cyclotomic quotients. We sketch a short proof, which uses the corresponding result for $Y_{d,n}$ (see [3, Sect. 3.1]). We are still working over $\mathbb{C}[u^{\pm1}, v]$.

4.3.1 Isomorphism Theorem for Affine Yokonuma–Hecke Algebras

For $\mu \vDash_d n$, we consider the algebra $\mathrm{Mat}_{m_\mu}(\widehat{H}^\mu)$ of matrices of size m_μ with coefficients in $\widehat{H}^\mu$. We recall that m_μ, given by (4.8), is the number of characters $\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})$ such that $\mathrm{Comp}(\chi) = \mu$. So we index the rows and columns of a matrix in $\mathrm{Mat}_{m_\mu}(\widehat{H}^\mu)$ by such characters. Moreover, for two characters χ, χ' such that $\mathrm{Comp}(\chi) = \mathrm{Comp}(\chi') = \mu$, we denote by $\mathbf{1}_{\chi,\chi'}$ the matrix in $\mathrm{Mat}_{m_\mu}(\widehat{H}^\mu)$ with 1 in line χ and column χ', and 0 everywhere else.

Theorem 4.1 *The affine Yokonuma–Hecke algebra $\widehat{Y}_{d,n}$ is isomorphic to $\bigoplus_{\mu \vDash_d n} \mathrm{Mat}_{m_\mu}(\widehat{H}^\mu)$, the isomorphism being given on the elements of the basis (4.11) by*

$$\Psi_{d,n} \ : \ E_\chi X^\lambda g_{w^{-1}} \longmapsto u^{\ell(w^{-1})-\ell(\pi_\chi^{-1}w^{-1}\pi_{w(\chi)})}\, \mathbf{1}_{\chi,w(\chi)}\, \overline{X}^{\pi_\chi^{-1}(\lambda)} \overline{g}_{\pi_\chi^{-1}w^{-1}\pi_{w(\chi)}}\,, \quad (4.14)$$

where $\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})$, $\lambda \in \mathbb{Z}^n$ and $w \in \mathfrak{S}_n$ (ℓ is the length function on $\mathfrak{S}_n$).

Proof (Sketch of a proof) We start with explicit formulas for the images of the generators of $\widehat{Y}_{d,n}$ given below in (4.15)–(4.17), and we check that the images of the generators satisfy all the defining relations (4.1)–(4.3) of $\widehat{Y}_{d,n}$. For the relations not involving the generator X_1, this is already known from the isomorphism theorem for $Y_{d,n}$. We omit the remaining straightforward verifications.

Thus, Formulas (4.15)–(4.17) induce a morphism of algebras, and we check that it coincides with $\Psi_{d,n}$ given by (4.14). Again, for the images of the elements of the basis of the form $E_\chi g_{w^{-1}}$, this is already known from the $Y_{d,n}$ situation. The multiplication by X^λ is straightforward.

It remains to check that $\Psi_{d,n}$ is bijective. The surjectivity follows from a direct inspection of Formula (4.16) together with the already known fact that every $\mathrm{Mat}_{m_\mu}(H^\mu)$ is in the image of $\Psi_{d,n}$. The injectivity can be checked directly. Indeed, assume that a certain linear combination $\sum_{\chi,\lambda,w} c_{\chi,\lambda,w} E_\chi X^\lambda g_{w^{-1}}$ is in the kernel of $\Psi_{d,n}$. Then for every χ, λ, w, we obtain that

$$\sum_{w'} u^{\ell(w'^{-1})-\ell(\pi_\chi^{-1}w'^{-1}\pi_{w'(\chi)})} c_{\chi,\lambda,w'}\, \overline{g}_{\pi_\chi^{-1}w'^{-1}\pi_{w'(\chi)}} = 0\,,$$

where the sum is over $w' \in \mathfrak{S}_n$ such that $w'(\chi) = w(\chi)$. For w', w'' satisfying this condition, we have $\pi_\chi^{-1}w'^{-1}\pi_{w'(\chi)} = \pi_\chi^{-1}w''^{-1}\pi_{w''(\chi)}$ if and only if $w' = w''$, and therefore every coefficients in the above sum are 0.

Formulas for the generators.

Here, we give the images under the isomorphism $\Psi_{d,n}$ of the generators of $\widehat{Y}_{d,n}$. We recall that $\sum E_\chi = 1$ in $\widehat{Y}_{d,n}$, the sum being over $\mathrm{Irr}(\mathscr{T}_{d,n})$. Let $\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})$ and set $\mu = \mathrm{Comp}(\chi)$. Then, by definition of π_χ, it is straightforward to see that $\pi_\chi^{-1}(j) = \mu_1 + \cdots + \mu_{a-1} + \alpha$, where $\chi(t_j) = \xi_a$ and $\alpha = \sharp\{k \leq j \mid \chi(t_k) = \xi_a\}$.

- Let $j \in \{1, \dots, n\}$. We have:

$$t_j = \sum_{\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})} E_\chi t_j = \sum_{\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})} E_\chi \chi(t_j) \longmapsto \sum_{\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})} \mathbf{1}_{\chi,\chi}\, \chi(t_j)\,. \quad (4.15)$$

  It follows that the image of e_i, $i = 1, \dots, n-1$, is a sum of diagonal matrices; the coefficient in position χ is 1 if $\chi(t_i) = \chi(t_{i+1})$, and 0 otherwise.
- Let $j \in \{1, \dots, n\}$. We have:

$$X_j = \sum_{\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})} E_\chi X_j \longmapsto \sum_{\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})} \mathbf{1}_{\chi,\chi}\, \overline{X}_{\pi_\chi^{-1}(j)}\,. \quad (4.16)$$

- Let $i \in \{1, \dots, n-1\}$. We have:

$$g_i = \sum_{\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})} E_\chi g_i \quad \text{and} \quad E_\chi g_i \longmapsto \begin{cases} u\,\mathbf{1}_{\chi, s_i(\chi)} & \text{if } s_i(\chi) \neq \chi, \\ \mathbf{1}_{\chi,\chi}\, \overline{g}_{\pi_\chi^{-1}(i)} & \text{if } s_i(\chi) = \chi\,. \end{cases} \quad (4.17)$$

  The first line follows from $\pi_{s_i(\chi)} = s_i \pi_\chi$ if $s_i(\chi) \neq \chi$. The second line follows from $\pi_\chi^{-1}(i+1) = \pi_\chi^{-1}(i) + 1$ if $s_i(\chi) = \chi$.

4.3.2 Isomorphism Theorem for Cyclotomic Yokonuma–Hecke Algebras

Let $\boldsymbol{v} = (v_1, \dots, v_m) \subset \mathbb{C}[u^{\pm 1}, v] \backslash \{0\}$ be an m-tuple of non-zero parameters for a certain $m \in \mathbb{Z}_{>0}$ (equivalently, one could consider $v_1, \dots, v_m$ as indeterminates and work over the extended ring $\mathbb{C}[u^{\pm 1}, v, v_1^{\pm 1}, \dots, v_m^{\pm 1}]$). The cyclotomic Yokonuma–Hecke algebra $Y_{d,n}(\boldsymbol{v})$ is the quotient of the affine Yokonuma–Hecke algebra $\widehat{Y}_{d,n}$ by the relation

$$(X_1 - v_1) \dots (X_1 - v_m) = 0\,. \quad (4.18)$$

It is shown in [8] that the algebra $Y_{d,n}(\boldsymbol{v})$ is a free $\mathbb{C}[u^{\pm 1}, v]$-module with basis

$$\{t_1^{a_1} \dots t_n^{a_n} X^\lambda g_w \mid a_1, \dots, a_n \in \{1, \dots, d\},\ \lambda \in \{0, \dots, m-1\}^n,\ w \in \mathfrak{S}_n\}\,.$$

In particular, if $m = 1$, $Y_{d,n}(\boldsymbol{v})$ is isomorphic to the Yokonuma–Hecke algebra $Y_{d,n}$.

Similarly, the cyclotomic Hecke algebra $H_n(\boldsymbol{v})$ (or the Ariki–Koike algebra) is the quotient of the affine Hecke algebra $\widehat{H}_n$ by the relation $(\overline{X}_1 - v_1) \dots (\overline{X}_1 - v_m) = 0$. Equivalently, it is the quotient of the cyclotomic Yokonuma–Hecke algebra $Y_{d,n}(\boldsymbol{v})$ by the relations $t_j = 1$, $j = 1, \dots, n$. It is a free $\mathbb{C}[u^{\pm 1}, v]$-module with basis $\{\overline{X}^\lambda \overline{g}_w \mid \lambda \in \{0, \dots, m-1\}^n,\ w \in \mathfrak{S}_n\}$.

For $\mu \models_d n$, we set $H(\boldsymbol{v})^\mu := H_{\mu_1}(\boldsymbol{v}) \otimes \dots \otimes H_{\mu_d}(\boldsymbol{v})$. By definition, $H(\boldsymbol{v})^\mu$ is the quotient of the algebra $\widehat{H}^\mu$ by the relations

$$(\overline{X}_{\mu_1+\cdots+\mu_{a-1}+1} - v_1)\dots(\overline{X}_{\mu_1+\cdots+\mu_{a-1}+1} - v_m) = 0\,, \qquad a = 1,\dots,d\,. \tag{4.19}$$

Corollary 4.1 *The cyclotomic Yokonuma–Hecke algebra $Y_{d,n}(\boldsymbol{v})$ is isomorphic to the direct sum $\bigoplus_{\mu \vDash_d n} \mathrm{Mat}_{m_\mu}\big(H(\boldsymbol{v})^\mu\big)$.*

Proof Let $I_{\boldsymbol{v}}$ be the (two-sided) ideal of $\widehat{Y}_{d,n}$ generated by the left hand side of the relation (4.18). For $\mu \vDash_d n$, let $\overline{I}^{\mu}_{\boldsymbol{v}}$ be the ideal of $\widehat{H}^\mu$ generated by the left hand sides of the relations (4.19). The corollary follows from Theorem 4.1 together with the fact that $\Psi_{d,n}(I_{\boldsymbol{v}}) = \bigoplus_{\mu \vDash_d n} \mathrm{Mat}_{m_\mu}\big(\overline{I}^{\mu}_{\boldsymbol{v}}\big)$. It remains to check this fact.

The inclusion "$\subset$" follows at once from Formula (4.16) for $j = 1$. For the other inclusion, let $\mu \vDash_d n$. Let $a \in \{1,\dots,d\}$ such that $\mu_a \neq 0$, so that there is a character χ with $\mathrm{Comp}(\chi) = \mu$ and $\chi(t_1) = \xi_a$. Again, Formula (4.16) for $j = 1$ gives $\Psi_{d,n}(E_\chi X_1) = \mathbf{1}_{\chi,\chi}\overline{X}_{\mu_1+\cdots+\mu_{a-1}+1}$. Therefore, for every generators of $\overline{I}^{\mu}_{\boldsymbol{v}}$, we have in $\Psi_{d,n}(I_{\boldsymbol{v}})$ a matrix in $\mathrm{Mat}_{m_\mu}\big(\widehat{H}^\mu\big)$ with the generator as one diagonal element and 0 everywhere else. As $\Psi_{d,n}(I_{\boldsymbol{v}})$ is an ideal, this shows that $\mathrm{Mat}_{m_\mu}\big(\overline{I}^{\mu}_{\boldsymbol{v}}\big)$ is included in $\Psi_{d,n}(I_{\boldsymbol{v}})$.

4.4 Markov Traces on Affine Yokonuma–Hecke Algebras

From now on, we extend the ground ring $\mathbb{C}[u^{\pm 1}, v]$ to $\mathbb{C}[u^{\pm 1}, v^{\pm 1}]$, and we consider our algebras over this extended ring.

4.4.1 Definition of Markov Traces on $\{\widehat{Y}_{d,n}\}_{n\geq 1}$ and $\{\widehat{H}_n\}_{n\geq 1}$

A Markov trace on the family of algebras $\{\widehat{Y}_{d,n}\}_{n\geq 1}$ is a family of linear functions $\{\rho_{d,n} : \widehat{Y}_{d,n} \to \mathbb{C}[u^{\pm 1}, v^{\pm 1}]\}_{n\geq 1}$ satisfying:

$$\begin{aligned} &\rho_{d,n}(xy) = \rho_{d,n}(yx)\,, && n \geq 1 \text{ and } x, y \in \widehat{Y}_{d,n};\\ &\rho_{d,n+1}(xg_n) = \rho_{d,n+1}(xg_n^{-1}) = \rho_{d,n}(x)\,, && n \geq 1 \text{ and } x \in \widehat{Y}_{d,n}. \end{aligned} \tag{4.20}$$

A Markov trace on the family of algebras $\{\widehat{H}_n\}_{n\geq 1}$ is a family of linear functions $\{\tau_n : \widehat{H}_n \to \mathbb{C}[u^{\pm 1}, v^{\pm 1}]\}_{n\geq 1}$ satisfying:

$$\begin{aligned} &\tau_n(xy) = \tau_n(yx)\,, && n \geq 1 \text{ and } x, y \in \widehat{H}_n;\\ &\tau_{n+1}(x\overline{g}_n) = \tau_{n+1}(x\overline{g}_n^{-1}) = \tau_n(x)\,, && n \geq 1 \text{ and } x \in \widehat{H}_n. \end{aligned} \tag{4.21}$$

Recall the definition of Soc_d from Sect. 4.2. For each $\overline{\mu} \in \mathrm{Soc}_d$ and each $a \in \{1,\dots,d\}$ such that $\overline{\mu}_a \neq 0$, we choose a Markov trace $\{\tau_n^{\overline{\mu},a}\}_{n\geq 1}$ on $\{\widehat{H}_n\}_{n\geq 1}$. By convention, $\widehat{H}_0 := \mathbb{C}[u^{\pm 1}, v^{\pm 1}]$ and maps of the form $\tau_0^{\overline{\mu},a}$ are identities on $\widehat{H}_0$. Below in (4.22), each term in the sum over $\mu \vDash_d n$ acts on $\mathrm{Mat}_{m_\mu}(\widehat{H}^\mu)$. We skip the proof of the following theorem. It can be done exactly as in [3, Lemma 5.4].

Theorem 4.2 *The following maps form a Markov trace on* $\{\widehat{Y}_{d,n}\}_{n\geq 1}$*:*

$$\Big(\sum_{\mu\vdash_d n}(\tau_{\mu_1}^{\overline{\mu},1}\otimes\cdots\otimes\tau_{\mu_d}^{\overline{\mu},d})\circ \mathrm{Tr}_{\mathrm{Mat}_{m_\mu}}\Big)\circ\Psi_{d,n}\,,\qquad n\geq 1\,. \tag{4.22}$$

Roughly speaking, to construct a Markov trace on $\widehat{Y}_{d,n}$, after having applied the isomorphism $\Psi_{d,n}$ and the usual trace of a matrix, we must choose and apply a Markov trace on each component of $\widehat{H}^\mu=\widehat{H}_{\mu_1}\otimes\cdots\otimes\widehat{H}_{\mu_d}$ for each μ. This choice of Markov traces is restricted: if μ' has the same socle as μ (that is, $\mu'_a=0\ \Leftrightarrow\ \mu_a=0$), then the chosen Markov traces on each component of $\widehat{H}^{\mu'}$ must be the same as for $\widehat{H}^\mu$; otherwise, they can be chosen independently.

Basic Markov traces.

Recall the bijection (4.10) between Soc_d and non-empty subsets of $\{1\ldots,d\}$. Following the theorem, we define some distinguished Markov traces as follows:

- Choose a non-empty $S\subset\{1,\ldots,d\}$ and consider the associated $\mu^S\in\mathrm{Soc}_d$. Choose a Markov trace $\{\tau_n^a\}_{n\geq1}$ on $\{\widehat{H}_n\}_{n\geq1}$ for each $a\in S$, and set $\tau_n^{\mu^S,a}=\tau_n^a$, $n\geq1$, in (4.22).
- Then, in (4.22), set all other Markov traces $\{\tau_n^{\overline{\mu},a}\}_{n\geq1}$ with $\overline{\mu}\neq\mu^S$ to be 0.

We denote formally the choice of Markov traces in the first item by $\boldsymbol{\tau}$ and denote by $\{\rho_{d,n}^{S,\boldsymbol{\tau}}\}_{n\geq1}$ the resulting Markov trace on $\{\widehat{Y}_{d,n}\}_{n\geq1}$. We call it a *basic* Markov trace. Every Markov trace constructed in the preceding theorem is a linear combination of basic Markov traces $\{\rho_{d,n}^{S,\boldsymbol{\tau}}\}_{n\geq1}$, where S and $\boldsymbol{\tau}$ vary.

Example 4.1 A map on $\widehat{Y}_{d,n}$ can be seen, up to $\Psi_{d,n}$, as acting on the direct sum of matrix algebras. This way, for a given S, the maps $\rho_{d,n}^{S,\boldsymbol{\tau}}$ are non-zero only on the summands $\mathrm{Mat}_{m_\mu}(\widehat{H}^\mu)$ such that $\overline{\mu}=\mu^S$, that is, such that $\mu_a\neq0$ if and only if $a\in S$. As examples:

- if $S=\{k\}$ then $\rho_{d,n}^{S,\boldsymbol{\tau}}$ is non-zero only on $\mathrm{Mat}_{m_\mu}(\widehat{H}^\mu)$, for $\mu=(0,\ldots,0,n,0,\ldots,0)$ with n in position k. In this case, $\widehat{H}^\mu=\widehat{H}_n$;
- we will see that it is enough to consider the situation $S=\{1,\ldots,d\}$. In this case, $\rho_{d,n}^{S,\boldsymbol{\tau}}$ is non-zero only on $\mathrm{Mat}_{m_\mu}(\widehat{H}^\mu)$, for μ with all parts different from 0.

Remark 4.1 (i) By restriction to the subalgebra $Y_{d,n}$ of $\widehat{Y}_{d,n}$, a Markov trace on $\{\widehat{Y}_{d,n}\}_{n\geq1}$ reduces to a Markov trace on $\{Y_{d,n}\}_{n\geq1}$ (and similarly for $\widehat{H}_n$ and H_n). On $\{H_n\}_{n\geq1}$, there is a unique Markov trace up to a normalisation factor. Therefore, the choice of the Markov traces $\tau_n^{\overline{\mu},a}$ in the theorem above reduces, for $Y_{d,n}$, to a choice of an overall factor $\alpha_{\overline{\mu}}$ for each $\overline{\mu}\in\mathrm{Soc}_d$. This is the result proved in [3]. In other words, for $Y_{d,n}$, the basic Markov traces are parametrised by Soc_d, or equivalently, by the non-empty subsets of $\{1,\ldots,d\}$.

(ii) Let $\mathfrak{Mark}(\widehat{H}_n)$ be the space of Markov traces on $\{\widehat{H}_n\}_{n\geq1}$. The space spanned by the basic Markov traces $\rho_{d,n}^{S,\boldsymbol{\tau}}$ is isomorphic to

$$\bigoplus_{1\leq k\leq d}\binom{d}{k}\mathfrak{Mark}(\widehat{H}_n)^{\otimes k}\ .$$

If we restrict to a subspace of $\mathfrak{Mark}(\widehat{H}_n)$ of dimension D, we obtain a space of Markov traces on $\{\widehat{Y}_{d,n}\}_{n\geq 1}$ of dimension $(D+1)^d-1$. In particular, for $Y_{d,n}$, the dimension is 2^d-1.

4.5 Invariants for Links

Let γ be another indeterminate. We work from now on over the ring $R := \mathbb{C}[u^{\pm 1}, v^{\pm 1}, \gamma^{\pm 1}]$ and we consider now all algebras over this extended ring R.

We sketch a construction of invariants with values in R for $\mathbb{Z}/d\mathbb{Z}$-framed solid torus links. We call *$\mathbb{Z}/d\mathbb{Z}$-framed link* a usual link together with a number in $\mathbb{Z}/d\mathbb{Z}$ (the *framing*) on each connected component. We refer to [7, 8] for definitions and fundamental results (as the analogues of Alexander and Markov theorems) concerning solid torus links and their $\mathbb{Z}/d\mathbb{Z}$-framed versions. Note that any invariant for $\mathbb{Z}/d\mathbb{Z}$-framed links is also an invariant of non-framed links, simply by considering links with all framings equal to 0.

The set of classical links is naturally included in the set of solid torus links (in other words, the braid group is naturally a subgroup of the affine braid group). The construction here includes, by restriction to the subalgebras $Y_{d,n}$, the construction for $\mathbb{Z}/d\mathbb{Z}$-framed classical links explained in [3, Sect. 6]. As the construction and the results equally apply to the classical and the solid torus situations, we will simply use the word *link* to refer to both types of links.

4.5.1 Definition of the Invariants

As in [8], we denote by ${B^{\mathrm{aff}}}_n$ the affine braid group on n strands, and by $\mathbb{Z}/d\mathbb{Z}\wr {B^{\mathrm{aff}}}_n$ the $\mathbb{Z}/d\mathbb{Z}$-framed affine braid group. The generators of $\mathbb{Z}/d\mathbb{Z}\wr {B^{\mathrm{aff}}}_n$ are denoted $\sigma_1,\dots,\sigma_{n-1},\sigma_0,t_1,\dots,t_n$. The defining relations are (4.1)–(4.2) with g_i replaced by σ_i and X_1 by σ_0. The algebra $\widehat{Y}_{d,n}$ is thus a quotient of the group algebra of $\mathbb{Z}/d\mathbb{Z}\wr {B^{\mathrm{aff}}}_n$ by the relation (4.3).

The subgroup of $\mathbb{Z}/d\mathbb{Z}\wr {B^{\mathrm{aff}}}_n$ generated by $\sigma_1,\dots,\sigma_{n-1},\sigma_0$ is ${B^{\mathrm{aff}}}_n$. The algebra $\widehat{H}_n$ is a quotient of the group algebra of ${B^{\mathrm{aff}}}_n$ by the relation $\sigma_i^2 = u^2 + v\sigma_i$, $i = 1, ..., n-1$. Finally, the subgroup $\sigma_1,\dots,\sigma_{n-1}$ of ${B^{\mathrm{aff}}}_n$ is the classical braid group.

Invariants $P_L^\tau(u,v)$ from $\widehat{H}_n$.

Let $\{\tau_n\}_{n\geq 1}$ be a Markov trace on $\{\widehat{H}_n\}_{n\geq 1}$. From the Alexander and Markov theorems for non-framed links (see [7]), we construct the invariant $P_L^\tau(u,v)$, for a link L, as follows:

$$L \longmapsto \beta_L \in B^{\mathrm{aff}}{}_n \longmapsto \pi_n(\beta_L) \longmapsto \tau_n\big(\pi_n(\beta_L)\big) =: P_L^{\tau}(u,v) \in \mathbb{C}[u^{\pm 1}, v^{\pm 1}]\,,$$

where β_L is a braid closing to L and π_n is the natural morphism from $RB^{\mathrm{aff}}{}_n$ to $\widehat{H}_n$, given on the generators by $\sigma_i \mapsto \overline{g}_i$, $i = 1, ..., n-1$ and $\sigma_0 \mapsto \overline{X}_1$.

From the fact that there is a unique Markov trace (up to normalisation) on the usual Hecke algebras $\{H_n\}_{\geq 1}$, all the invariants P_L^{τ}, restricted to the set of classical links, reduce (up to normalisation) to the unique invariant coming from H_n, the HOMFLYPT polynomial.

Invariants $P_L^{d,S,\boldsymbol{\tau}}(u,v,\gamma)$ from $\widehat{Y}_{d,n}$.

We consider the following map from $R[\mathbb{Z}/d\mathbb{Z} \wr B^{\mathrm{aff}}{}_n]$ to $\widehat{Y}_{d,n}$ given on the generators by:

$$\delta_{d,n} \;:\; t_j \mapsto t_j\,, \qquad \sigma_0 \mapsto X_1\,, \qquad \sigma_i \mapsto \big(\gamma + (1-\gamma)e_i\big)g_i\,.$$

One proves as in [3, Sect. 6] that, first, $\delta_{d,n}$ extends to a morphism of algebras, and moreover, that the following procedure defines invariants for $\mathbb{Z}/d\mathbb{Z}$-framed links.

Definition 4.1 Let $\{\rho_{d,n}\}_{n\geq 1}$ be a Markov trace on $\{\widehat{Y}_{d,n}\}_{n\geq 1}$. For a $\mathbb{Z}/d\mathbb{Z}$-framed link L, the invariant $P_L^{\rho_d}(u,v,\gamma)$ is defined as follows

$$L \longmapsto \beta_L \in \mathbb{Z}/d\mathbb{Z} \wr B^{\mathrm{aff}}{}_n \longmapsto \delta_{d,n}(\beta_L) \longmapsto \rho_{d,n}\big(\delta_{d,n}(\beta_L)\big) =: P_L^{\rho_d}(u,v,\gamma) \in R\,,$$

where β_L is a $\mathbb{Z}/d\mathbb{Z}$-framed braid closing to L.

From the preceding section, it is enough to consider the basic Markov traces $\{\rho_{d,n}^{S,\boldsymbol{\tau}}\}_{n\geq 1}$. We denote by $P_L^{d,S,\boldsymbol{\tau}}$ the corresponding invariant and refer to it as a *basic* invariant.

If we restrict a basic invariant to the set of classical $\mathbb{Z}/d\mathbb{Z}$-framed links, it does not depend on $\boldsymbol{\tau}$, and coincides with the invariants constructed in [3].

Remark 4.2 In the definition of the maps $\delta_{d,n}$, a rule $\sigma_i \mapsto (\alpha + \beta e_i)g_i$ would be enough to give a morphism of algebras. The condition $\alpha + \beta = 1$ is necessary for the construction of invariants. We note that, considering the map $\delta_{d,n}$ is equivalent to changing the quadratic relations $g_i^2 = u^2 + ve_ig_i$ to $g_i^2 = u^2\gamma^2 + u^2(1-\gamma^2)e_i + ve_ig_i$. The role of γ is therefore to interpolate between different presentations of $\widehat{Y}_{d,n}$.

4.5.2 Comparison with Other Approaches

For non-framed links from $\widehat{H}_n$ [7].

Define $\widetilde{X}_i \in \widehat{Y}_{d,n}$, $i = 1, \ldots, n$, by the following formulas:

$$\widetilde{X}_1 := X_1 \qquad \text{and} \qquad \widetilde{X}_{i+1} := g_i^{-1}\widetilde{X}_i g_i\,, \quad i = 1, \ldots, n-1\,.$$

Similarly, we have the images $\overline{\widetilde{X}}_1, \ldots, \overline{\widetilde{X}}_n$ of $\widetilde{X}_1, \ldots, \widetilde{X}_n$ in $\widehat{H}_n$.

Let $\boldsymbol{x} := \{x_a\}_{a\in\mathbb{Z}} \subset \mathbb{C}[u^{\pm 1}, v^{\pm 1}]$ be a set of parameters with $x_0 := 1$. From the results in [7] (see Remark 4.3 below), we have a unique Markov trace on $\{\widehat{H}_n\}_{n\geq 1}$, which satisfies in addition

$$\tau_n^{\boldsymbol{x}}\big(\overline{\widetilde{X}}_n^a h\big) = x_a \tau_n^{\boldsymbol{x}}(h)\,, \qquad \forall n \geq 1\,,\ \ \forall a \in \mathbb{Z}\,,\ \ \forall h \in \widehat{H}_{n-1}\,.$$

The corresponding invariant of non-framed links is denoted $P_L^{\boldsymbol{x}}(u, v)$.

Remark 4.3 In [7], the quadratic relation of $\widehat{H}_n$ was $\overline{g}_i^2 = q + (q-1)\overline{g}_i$ and a certain tr, depending on another parameter z was constructed. Setting $\lambda := \frac{z+1-q}{qz}$, an invariant $\mathscr{X}_L(\sqrt{q}, \sqrt{\lambda})$ was obtained, by rescaling the generators, $\overline{g}_i \rightsquigarrow \sqrt{\lambda}\overline{g}_i$, and then rescaling the trace. This is equivalent, in our approach, to setting

$$u = \sqrt{\lambda q}\,, \quad v = \sqrt{\lambda}(q-1) \quad \text{and} \quad \tau_n^{\boldsymbol{x}} := (\sqrt{\lambda}z)^{1-n}\mathrm{tr} = \big(v^{-1}(1-u^2)\big)^{n-1}\mathrm{tr}\,,\ \ \forall n \geq 1\,.$$

In conclusion, we have $\mathscr{X}_L(\sqrt{q}, \sqrt{\lambda}) = P_L^{\boldsymbol{x}}(u, v)$.

For $\mathbb{Z}/d\mathbb{Z}$-framed links from $\widehat{Y}_{d,n}$ [1, 8, 9].

Recall that $\{\xi_1, \dots, \xi_d\}$ is the set of d-th roots of unity. Fix a non-empty subset $D \subset \{1, \dots, d\}$ and, for each $k \in D$, a set $\boldsymbol{x}^{(k)} := \{x_a^{(k)}\}_{a\in\mathbb{Z}}$ of parameters in $\mathbb{C}[u^{\pm 1}, v^{\pm 1}]$ with $x_0^{(k)} = 1$. Denote formally $\boldsymbol{x}$ the set $\boldsymbol{x}^{(k)}$ with $k \in D$.

From the results of [8] (and [9] in the non-affine case), we have a unique Markov trace, denoted $\{\widetilde{\rho}_{d,n}^{D,\boldsymbol{x}}\}_{n\geq 1}$, on $\{\widehat{Y}_{d,n}\}_{n\geq 1}$ which satisfies in addition, for all $n \geq 1$,

$$\widetilde{\rho}_{d,n}^{D,\boldsymbol{x}}\big(\widetilde{X}_n^a t_n^b h\big) = x_{a,b}\,\widetilde{\rho}_{d,n}^{D,\boldsymbol{x}}(h)\,, \quad \forall a \in \mathbb{Z}\,,\ \forall b \in \{1, \dots, d\}\,,\ \forall h \in \widehat{Y}_{d,n-1}\,, \tag{4.23}$$

where the parameters $x_{a,b}$ are given by

$$x_{a,b} = \frac{1}{|D|}\sum_{k\in D} x_a^{(k)}\xi_k^b\,, \qquad \text{for all } a \in \mathbb{Z} \text{and } b \in \{1, \dots, d\}. \tag{4.24}$$

The corresponding invariant of $\mathbb{Z}/d\mathbb{Z}$-framed links, we denote $\widetilde{P}_L^{d,D,\boldsymbol{x}}(u, v, \gamma)$. By restriction to classical $\mathbb{Z}/d\mathbb{Z}$-framed links, the parameters $\boldsymbol{x}$ do not appear, and we obtain invariants labelled by d and D, denoted $\widetilde{P}_L^{d,D}(u, v, \gamma)$ (see Remarks below).

Remark 4.4 (i) In [8], the quadratic relation of $\widehat{Y}_{d,n}$ was $g_i^2 = 1 + (q - q^{-1})e_i g_i$ and a certain trace tr, depending on another parameter z was constructed. An invariant $\Phi_L^{d,D,\boldsymbol{x}}(q, z)$ was obtained. With $\lambda_D := \frac{|D|z-(q-q^{-1})}{|D|z}$, this is equivalent, in our approach, to setting

$$u = \sqrt{\lambda_D}\,,\ v = \sqrt{\lambda_D}(q - q^{-1})\,,\ \gamma = 1 \ \text{ and } \ \widetilde{\rho}_{d,n}^{D,\boldsymbol{x}} := \big(|D|v^{-1}(1-u^2)\big)^{n-1}\mathrm{tr}\,,\ \forall n \geq 1\,.$$

In conclusion, we have $\Phi_L^{d,D,\boldsymbol{x}}(q,z) = \widetilde{P}_L^{d,D,\boldsymbol{x}}(u,v,1)$. Note that the construction in [8] corresponds to the particular case $\gamma = 1$ here.

(ii) The restriction of the above procedure to the subalgebra $Y_{d,n}$ (the non-affine case) gives the comparison of invariants constructed here with the invariants studied in [1]. The invariants only depend on d and D (not on $\boldsymbol{x}$), and were denoted in [1] by $\Phi_{d,D}(q,z)$.

(iii) Originally, in [9], invariants (in the non-affine case) were constructed using $Y_{d,n}$ with a different quadratic equation, namely with $g_i^2 = 1 + (q^2-1)e_i + (q^2-1)e_i g_i$. A certain trace $\tilde{\mathrm{tr}}$, depending on a parameter $\tilde{z} = qz$ was constructed and invariants, denoted $\Gamma_{d,S}$ in [1], were obtained. With the same λ_D as above, the rescaling of the generators is now $g_i \rightsquigarrow \sqrt{\lambda_D} q^{-1} g_i$. A short calculation and comparison with the formula in Remark 4.2 shows that u and v are as in item (i), while now $\gamma = q^{-1}$. As a conclusion, we have, for any classical $\mathbb{Z}/d\mathbb{Z}$-framed link L,

$$\Phi_{d,D}(q,z)(L) = \widetilde{P}_L^{d,D}(u,v,1) \quad \text{and} \quad \Gamma_{d,D}(q,z)(L) = \widetilde{P}_L^{d,D}(u,v,q^{-1})\,. \tag{4.25}$$

4.5.3 Comparison in Terms of the Basic Invariants $P_L^{d,S,\boldsymbol{\tau}}(u,v,\gamma)$

We fix as above D and $\boldsymbol{x}$. It remains to identify the Markov trace $\{\widetilde{\rho}_{d,n}^{D,\boldsymbol{x}}\}_{n\geq 1}$ in terms of the basic Markov traces $\{\rho_{d,n}^{S,\boldsymbol{\tau}}\}_{n\geq 1}$ constructed after Theorem 4.2.

Proposition 4.1 *For each non-empty $S \subseteq D$, let $\boldsymbol{\tau}$ be obtained by taking, for each $k \in S$, the Markov trace $\{\tau^{\boldsymbol{x}^{(k)}}\}_{n\geq 1}$ in position k. Then we have*

$$\{\widetilde{\rho}_{d,n}^{D,\boldsymbol{x}}\}_{n\geq 1} = \frac{1}{|D|} \sum_{S\subseteq D} \left(v^{-1}(1-u^2)\right)^{|S|-1} \{\rho_{d,n}^{S,\boldsymbol{\tau}}\}_{n\geq 1}\,. \tag{4.26}$$

Proof (Sketch of proof) Denote $\{\rho_n\}_{n\geq 1}$ the Markov trace on $\{\widehat{Y}_{d,n}\}_{n\geq 1}$ defined by the right hand side of (4.26). To prove the proposition, we need to check that Condition (4.23) is satisfied.

For $n = 1$, it is a straightforward verification. For $n > 1$, one may check the equivalent condition $\rho_n(\widetilde{X}_n^a t_n^b\, h) = |D| v^{-1}(1-u^2) x_{a,b}\, \rho_{n-1}(h)$. We note that it is enough to take $h = E_\chi X^\lambda g_{w^{-1}}$, where $\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})$ is such that $\mathrm{Comp}(\chi) = \mu$ where $\overline{\mu} = \mu^S$ for some $S \subset D$, and such that $w(\chi) = \chi$ (otherwise the condition is $0 = 0$). Then the condition can be checked by a straightforward calculation of both sides. One may use: an explicit description of the embedding $\widehat{Y}_{d,n} \subset \widehat{Y}_{d,n+1}$ on the matrix algebras side (see [3, Sect. 3.4]); and the fact that $\Psi_{d,n}(\widetilde{X}_j) = \sum_\chi \mathbf{1}_{\chi,\chi} \overline{\widetilde{X}}_{\pi_\chi^{-1}(j)}$ for $j = 1,\ldots,n$ (induction on j).

We note that Formula (4.26) is a triangular change of basis, with inverse

$$\{\rho_{d,n}^{S,\boldsymbol{\tau}}\}_{n\geq 1} = \left(v^{-1}(1-u^2)\right)^{1-|S|} \sum_{D\subseteq S} (-1)^{|S|-|D|} |D| \, \{\widetilde{\rho}_{d,n}^{D,\boldsymbol{x}}\}_{n\geq 1} \, .$$

In particular, restricting to $Y_{d,n}$, we can forget the parameters $\boldsymbol{x}$ on one hand, and the choice $\boldsymbol{\tau}$ on the other. The proposition expresses the Juyumaya–Lambropoulou invariant associated to $D \subset \{1, \dots, d\}$ in terms of our basic invariants associated to $S \subset \{1, \dots, d\}$. The formulas are (with notations as in Remark 4.4 and coefficients α_S as in (4.26)):

$$\Phi_{d,D}(q,z)(L) = \sum_{S\subseteq D} \alpha_S \, P_L^{d,S}(u,v,1) \quad \text{and} \quad \Gamma_{d,D}(q,z)(L) = \sum_{S\subseteq D} \alpha_S \, P_L^{d,S}(u,v,q^{-1}) \, . \tag{4.27}$$

4.6 Properties of Invariants

As consequences of the construction using the isomorphism theorem, we prove several properties of the constructed invariants, focusing essentially on the non-framed links. We emphasize that these properties are valid for all non-framed links (classical and solid torus).

4.6.1 Comparison of Invariants with Different d for Non-framed Links

Let $d > 0$ and S a non-empty subset of $\{1, \dots, d\}$. We denote $d' := |S|$. Let $\boldsymbol{\tau}$ be any choice of d' Markov traces on $\{\widehat{H}_n\}_{n\geq 1}$. The following result says that, for non-framed links, it is enough to consider the situation $S = \{1, \dots, d\}$ for each $d > 0$.

Proposition 4.2 *For any non-framed link L, we have $P_L^{d,S,\boldsymbol{\tau}} = P_L^{d',\{1,\dots,d'\},\boldsymbol{\tau}}$.*

In particular, if $|S| = 1$, the proposition asserts that $P_L^{d,S,\tau}(u,v,\gamma) = P_L^{\tau}(u,v)$, where $P_L^{\tau}(u,v)$ is the invariant of the non-framed link obtained from $\widehat{H}_n$.

Proof Let $\{\xi_1^{(d)}, \dots, \xi_d^{(d)}\}$ denote the d-th roots of unity.

Let $\mu \Vdash_d n$ be such that $\overline{\mu} = \mu^S$, that is, such that $\mu_a \neq 0$ if and only if $a \in S$. To μ, we associate the composition $\mu' = (\mu_{i_1}, \dots, \mu_{i_{d'}}) \Vdash_{d'} n$, where $\mu_{i_1}, \dots, \mu_{i_{d'}}$ are the non-zero parts of μ and $i_1 < \cdots < i_{d'}$. We have $S^{\mu} = S^{\mu'}$ and, in turn, $\widehat{H}^{\mu} = \widehat{H}^{\mu'}$ and $m_{\mu} = m_{\mu'}$.

Let $\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})$ with $\mathrm{Comp}(\chi) = \mu$. For every $j = 1, \dots, n$, by hypothesis on μ and χ, there exists $a \in \{1, \dots, d'\}$ such that $\chi(t_j) = \xi_{i_a}^{(d)}$. Then we set $\chi'(t_j) =$

$\xi_a^{(d')}$. This defines a bijection between the characters $\chi \in \mathrm{Irr}(\mathscr{T}_{d,n})$ with $\mathrm{Comp}(\chi) = \mu$ and the characters $\chi' \in \mathrm{Irr}(\mathscr{T}_{d',n})$ with $\mathrm{Comp}(\chi') = \mu'$. This bijection allows to identify the spaces $\mathrm{Mat}_{m_\mu}(\widehat{H}^\mu)$ and $\mathrm{Mat}_{m_{\mu'}}(\widehat{H}^{\mu'})$. We have moreover $\pi_\chi = \pi_{\chi'}$.

Recall the formulas (4.16)–(4.17) giving the images of the generators $g_1, \dots, g_{n-1}$, X_1 under $\Psi_{d,n}$. It is then immediate to see that $\Psi_{d,n}(x)$ and $\Psi_{d',n}(x)$ coincide in the summand $\mathrm{Mat}_{m_\mu}(\widehat{H}^\mu) = \mathrm{Mat}_{m_{\mu'}}(\widehat{H}^{\mu'})$, for any x in the subalgebra of $\widehat{Y}_{d,n}$ generated by $g_1, \dots, g_{n-1}, X_1$.

Remark 4.5 Here we restrict to a classical non-framed link L. Note that the restriction of the invariants Φ and Γ on the set of non-framed links were called respectively Θ and Δ in [1]. With the comparison formulas (4.27) of the preceding section, a straightforward consequence of Proposition 4.2 is the corresponding result for the Juyumaya–Lambropoulou invariants. Namely, we have

$$\Theta_{d,D}(q,z)(L) = \Theta_{d',\{1,\dots,d'\}}(q,z)(L) \quad \text{and} \quad \Delta_{d,D}(q,z)(L) = \Delta_{d',\{1,\dots,d'\}}(q,z)(L)\,,$$

where $d' = |D|$. In this case, this was proved in [1, Proposition 4.6] by a different approach. In particular if $|D| = 1$, we recover the HOMFLYPT polynomial.

4.6.2 Links with a Fixed Number of Connected Components

Let $\mu \vDash_d n$ and $\lambda \vDash_k n$ for some $d, k > 0$. We will say that λ is a *refinement* of μ if $\{1, \dots, k\}$ can be partitioned into d disjoint subsets (possibly empty): $\{1, \dots, k\} = I_1 \sqcup \cdots \sqcup I_d$, such that $\mu_a = \sum_{i \in I_a} \lambda_i$ for all $a = 1, \dots, d$. As examples, every composition $\lambda \vDash n$ is a refinement of the composition (n), while the composition $(1, \dots, 1) \vDash n$ is a refinement of every composition $\mu \vDash n$.

For a permutation $\pi \in S_n$, we denote $\mathrm{cyc}(\pi)$ the collection of lengths of the cycles of π and we consider it as a composition of n (the order is not relevant here). Then, it is immediate that a permutation $\pi \in S_n$ is conjugate to an element of S^μ if and only if $\mathrm{cyc}(\pi)$ is a refinement of μ.

For a $\mathbb{Z}/d\mathbb{Z}$-framed affine braid $\beta \in \mathbb{Z}/d\mathbb{Z} \wr B^{\mathrm{aff}}{}_n$, we define its *underlying permutation* p_β as the image of β by the natural group homomorphism from $\mathbb{Z}/d\mathbb{Z} \wr B^{\mathrm{aff}}{}_n$ to S_n (defined by $t_j \mapsto 1$, $\sigma_0 \mapsto 1$ and $\sigma_i \mapsto s_i$). Note that $\beta t_j = t_{p_\beta(j)}\beta$, for $j = 1, \dots, n$.

Proposition 4.3 *Let $\beta \in \mathbb{Z}/d\mathbb{Z} \wr B^{aff}{}_n$. In $\Psi_{d,n}\big(\delta_{d,n}(\beta)\big)$, the matrix corresponding to $\mu \vDash_d n$ has all its diagonal elements equal to 0 if $\mathrm{cyc}(\pi)$ is not a refinement of μ.*

Proof Let $x := \delta_{d,n}(\beta) \in \widehat{Y}_{d,n}$. As $\delta_{d,n}$ is a group homomorphism, we have $xt_j = t_{p_\beta(j)}x$. In $\Psi_{d,n}(x)$, in the matrix corresponding to $\mu \vDash_d n$, the coefficient on the diagonal in position χ is $\Psi_{d,n}(E_\chi x E_\chi)$. And we have $E_\chi x E_\chi = E_\chi E_{p_\beta(\chi)} x$. This is equal to 0 if $p_\beta(\chi) \neq \chi$. Now $p_\beta(\chi) = \chi$ if and only if $\pi_\chi^{-1} p_\beta \pi_\chi \in S^\mu$, and this is impossible if $\mathrm{cyc}(p_\beta)$ is not a refinement of μ.

Now we will combine this general result on the isomorphism $\Psi_{d,n}$ with two elementary facts. First, if a composition μ has strictly more non-zero parts than another composition λ, then λ can not be a refinement of μ. Second, the closure of a ($\mathbb{Z}/d\mathbb{Z}$-framed, affine) braid β is a ($\mathbb{Z}/d\mathbb{Z}$-framed) link with N connected components if and only if $\mathrm{cyc}(p_\beta)$ has exactly N non-zero parts. Thus we have obtained the following result.

Corollary 4.2 *Let L be a $\mathbb{Z}/d\mathbb{Z}$-framed link with N connected components. We have $P_L^{d,S,\boldsymbol{\tau}} = 0$ if $|S| > N$.*

Finally, combining this corollary with Proposition 4.2 for a non-framed link L, we conclude our study by determining which basic invariants it is enough to consider given the number of connected components of L.

Corollary 4.3 *Let L be a non-framed link with N connected components. Every invariant for L obtained here from (affine) Yokonuma–Hecke algebras is a combination of the following basic invariants:*

$$P_L^{1,\{1\},\boldsymbol{\tau}}\ ,\ P_L^{2,\{1,2\},\boldsymbol{\tau}}\ ,\ \dots\ ,\ P_L^{N,\{1,\dots,N\},\boldsymbol{\tau}}\ . \tag{4.28}$$

In particular, for a non-framed knot ($N = 1$), it is enough to consider the algebras $\widehat{H}_n = \widehat{Y}_{1,n}$.

For a classical non-framed link, we can forget the parameters $\boldsymbol{\tau}$, and it is enough to calculate N distinct invariants, the first one being the HOMFLYPT polynomial.

Remark 4.6 Here we restrict to a classical $\mathbb{Z}/d\mathbb{Z}$-framed link L with N connected components. In this particular case, we give the translation of Corollary 4.2 in terms of Juyumaya–Lambropoulou invariants, following Formula 4.27 (we write the formulas for the invariants $\Phi_{d,D}$; the same formulas hold for $\Gamma_{d,D}$). A straightforward analysis leads to

$$\Phi_{d,D}(q,z)(L) = \frac{N}{|D|} \sum_{D'\subset D,\ |D'|=N} \Phi_{d,D'}(q, \frac{|D|}{N} z)(L)\,, \qquad \text{if } |D| > N\,.$$

Note that the rescaling of the variable z comes from the fact that the expressions relating (u,v) with (q,z) depend on $|D|$.

If moreover L is a classical non-framed link with N connected components, with Remark 4.5, it is enough to consider $D = \{1,\dots,d\}$, and we obtain

$$\Theta_{d,\{1,\dots,d\}}(q,z)(L) = \frac{N}{d}\binom{d}{N}\Theta_{N,\{1,\dots,N\}}(q, \frac{d}{N} z)(L)\,, \qquad \text{if } d > N\,. \tag{4.29}$$

This formula is the generalisation, for $N > 1$, of [1, Theorem 5.8]. As a consequence, the analogue of Corollary 4.3 holds as well for these invariants: it is enough to consider $\Theta_{1,\{1\}}, \dots, \Theta_{N,\{1,\dots,N\}}$. This generalises [1, Theorem 7.1] for $N > 2$, obtained by a different method.

Acknowledgements It is a pleasure to thank the organisers of the Thales workshop held in Athens in July 2015 and especially Sofia Lambropoulou for her interest in this work.

References

1. Chlouveraki, M., Juyumaya, J., Karvounis, K., Lambropoulou, S.: Identifying the invariants for classical knots and links from the Yokonuma–Hecke algebra. arXiv:1505.06666
2. Poulain d'Andecy, L., Wagner, E.: The HOMFLYPT polynomials of sublinks and the Yokonuma-Hecke algebras. to appear in Proc. Amer. Math. Soc. arXiv:1606.00237
3. Jacon, N., Poulain d'Andecy, L.: An isomorphism theorem for Yokonuma-Hecke algebras and applications to link invariants. Math. Z. **283**(1–2), 301–338 (2016)
4. Chlouveraki, M., Poulain d'Andecy, L.: Representation theory of the Yokonuma-Hecke algebra. Adv. Math. **259**, 134–172 (2014)
5. Cui, W.: Affine cellularity of affine Yokonuma–Hecke algebras. arXiv:1510.02647
6. Lusztig, G.: Character sheaves on disconnected groups VII. Represent. Theory **9**, 209–266 (2005)
7. Lambropoulou, S.: Knot theory related to generalized and cyclotomic Hecke algebras of type B. J. Knot Theory Ramif. **8**(5), 621–658 (1999)
8. Chlouveraki, M., Poulain d'Andecy, L.: Markov traces on affine and cyclotomic Yokonuma–Hecke algebras. Int. Math. Res. **14**, 4167–4228 (2016)
9. Juyumaya, J., Lambropoulou, S.: p-adic framed braids II. Adv. Math. **234**, 149–191 (2013)

Chapter 5
On the Framization of the Hecke Algebra of Type B

Marcelo Flores

Abstract We give a cross look to two framizations of the Hecke algebra of type B. One of these is a particular case of the cyclotomic Yokonuma–Hecke algebra. The other one was recently introduced by the author, J. Juyumaya and S. Lambropoulou. The purpose of this paper is to show the main concepts and results of both framizations, giving emphasis to the second one, and to provide a preliminary comparison of the invariants constructed from both framizations.

Introduction

The idea of framization of a knot algebra was introduced by J. Juyumaya and S. Lambropoulou in [14], and it consists in adding certain new generators, called framing generators, to the original presentation of a knot algebra together with certain relations among the original generators and these new generators. One does this procedure, with the aim of constructing new invariants for framed links, and consequently for classical links, see [11, 13]. It is important to mention that the framization procedure doesn't have a structured recipe, whence it is possible to find more than one framization for the same algebra. However, since the motivation behind the procedure of framization is to obtain new polynomial invariants for (framed) knots and links, focus is always given to those framizations that produce such new invariants. Then, when we handle multiple possible framizations of the same knot algebra, we will choose the one that is more natural from a topological viewpoint, cf. [7].

The Yokonuma–Hecke algebra is the first example of framization, since it is considered as a framization of the Hecke algebra of type A. The last 10 years the Yokonuma–Hecke algebra has earned importance in knot theory, since in [10] it was proved that such an algebra supports a Markov trace, therefore, by using the Jones's recipe, invariants for: framed links [13], classical links [11] and singular links [12] were constructed. It is worth to say, that recently it was proved that the invariants for

M. Flores (✉)
Instituto de Matemáticas, Universidad de Valparaíso,
Gran Bretaña 1091, Valparaíso, Chile
e-mail: marcelo.flores@uv.cl

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_5

classical links constructed in [11] are not topologically equivalent either to the Homflypt polynomial or to the Kauffman polynomial, see [3].

On the other hand, Jones suggested that his recipe for the construction of the Homflypt polynomial might be used for Hecke algebras of other types than A, cf. [9, p. 336]. Then, S. Lambropoulou used the Jones's recipe for the B–Hecke algebra $\mathrm{H}_n(\mathsf{u}, \mathsf{v})$; in fact in [15, 16] she constructed all the possible analogues of the Homflypt polynomial for oriented knots and links inside the solid torus, see also [6].

In [2] taking as model the Yokonuma–Hecke algebra, M. Chlouveraki and L. Poulain d'Andecy introduced the cyclotomic Yokonuma–Hecke algebra, denoted by $\mathrm{Y}(d, m, n)$. These algebras generalize to the Ariki-Koike algebra, and the Yokonuma–Hecke algebra. In particular, the cyclotomic Yokonuma–Hecke algebra provides a framization of the Hecke algebra of type B, since the Ariki-Koike algebra generalizes $\mathrm{H}_n(\mathsf{u}, \mathsf{v})$. Moreover, in [2] it was also proved that $\mathrm{Y}(d, m, n)$ supports a Markov trace, and therefore, using Jones's recipe, an invariant for framed links in the solid torus was constructed.

Recently in [5] we introduced a new framization of the Hecke algebra of type B, denoted by $\mathrm{Y}^{\mathrm{B}}_{d,n} := \mathrm{Y}^{\mathrm{B}}_{d,n}(\mathsf{u}, \mathsf{v})$, with the principal objective to explore their usefulness in knot theory. More precisely, in this article we constructed two linear bases, a faithful tensorial representation of Jimbo type for $\mathrm{Y}^{\mathrm{B}}_{d,n}(\mathsf{u}, \mathsf{v})$, and we proved that $\mathrm{Y}^{\mathrm{B}}_{d,n}$ supports a Markov trace. Finally we defined, by using Jones's recipe, a new invariant for framed and classical links in the solid torus.

The article is organized as follows. In Sect. 5.1 we introduce the notations and background used in the paper. In Sect. 5.2 we review the main results about the cyclotomic Yokonuma–Hecke algebra given in [2], aiming to provide a preliminary comparison of the invariants constructed from both framizations. The following sections keep the order given in [5], and have as objective to show the main results obtained in that work, and also remark some differences between the framizations $\mathrm{Y}^{\mathrm{B}}_{d,n}$ and $\mathrm{Y}(d, 2, n)$.

5.1 Preliminaries

In this section we review known results, necessary for the sequel, and we also fix the terminology and notations that will be used along the article:

- The letters $\mathsf{u}, \mathsf{v}, \mathsf{v}_1, \ldots, \mathsf{v}_m$ denote indeterminates. Consider $\mathbb{K} := \mathbb{C}(\mathsf{u}, \mathsf{v})$, $\mathscr{R}_m := \mathbb{C}[\mathsf{u}^{\pm 1}, \mathsf{v}_1^{\pm 1}, \ldots, \mathsf{v}_m^{\pm 1}]$, and $\mathscr{F}_m$ the field of fractions of $\mathscr{R}_m$.
- The term *algebra* means unital associative algebra over $\mathbb{K}$
- For a finite group G, $\mathbb{K}[G]$ denotes the group algebra of G
- The letters n and d denote two fixed positive integers
- We denote by ω a fixed primitive d–th root of unity
- We denote by $\mathbb{Z}/d\mathbb{Z}$ the group of integers modulo d, $\{0, 1, \ldots, d-1\}$.
- As usual, we denote by ℓ the length function associated to the Coxeter groups.

5.1.1 Braids Groups of Type B

Set $n \geq 1$. Let us denote by W_n the Coxeter group of type $\mathtt{B}_n$. This is the finite Coxeter group associated to the following Dynkin diagram

Define $\mathtt{r}_k = \mathtt{s}_{k-1} \ldots \mathtt{s}_1 \mathtt{r}_1 \mathtt{s}_1 \ldots \mathtt{s}_{k-1}$ for $2 \leq k \leq n$. It is known, see [6], that every element $w \in W_n$ can be written uniquely as $w = w_1 \ldots w_n$ with $w_k \in \mathtt{N}_k$, $1 \leq k \leq n$, where

$$\mathtt{N}_1 = \{1, \mathtt{r}\}, \quad \mathtt{N}_k = \{1, \mathtt{r}_k, \mathtt{s}_{k-1} \ldots \mathtt{s}_i, \mathtt{s}_{k-1} \ldots \mathtt{s}_i \mathtt{r}_i;\ 1 \leq i \leq k-1\}. \tag{5.1}$$

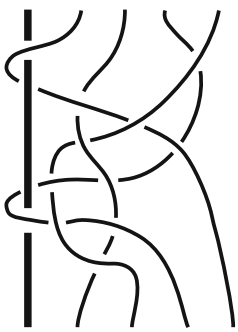
Furthermore, this expression for w is reduced. Hence, we have $\ell(w) = \ell(w_1) + \cdots + \ell(w_n)$.

The corresponding *braid group of type* $\mathtt{B}_n$ associated to W_n, is defined as the group $\widetilde{W}_n$ generated by $\rho_1, \sigma_1, \ldots, \sigma_{n-1}$ subject to the following relations

$$\begin{array}{rll} \sigma_i \sigma_j = \sigma_j \sigma_i & \text{for} & |i-j| > 1, \\ \sigma_i \sigma_j \sigma_i = \sigma_j \sigma_i \sigma_j & \text{for} & |i-j| = 1, \\ \rho_1 \sigma_i = \sigma_i \rho_1 & \text{for} & i > 1, \\ \rho_1 \sigma_1 \rho_1 \sigma_1 = \sigma_1 \rho_1 \sigma_1 \rho_1. & & \end{array} \tag{5.2}$$

Geometrically, braids of type $\mathtt{B}_n$ can be viewed as classical braids of type $\mathtt{A}_n$ with $n+1$ strands, such that the first strand is identically fixed. This is called 'the fixed strand'. The 2nd,..., $(n+1)$st strands are renamed from 1 to n and they are called 'the moving strands'. The 'loop' generator ρ_1 stands for the looping of the first moving strand around the fixed strand in the right-handed sense, see [15, 16]. In Fig. 5.1 we illustrate a braid of type $\mathtt{B}_4$. Another way of visualizing B–type braids geometrically is via symmetric braids, see [18].

Fig. 5.1 A braid of type $\mathtt{B}_4$



Remark 5.1 The group W_n can be realized as a subgroup of the permutation group of the set $X_n := \{-n, \ldots, -2, -1, 1, 2, \ldots, n\}$. More precisely, the elements of W_n are the permutations w such that $w(-m) = -w(m)$, for all $m \in X_n$. Further the elements of W_n can be parameterized by the elements of $X_n^n := \{(m_1, \ldots, m_n) \mid m_i \in X_n \text{ for all } i\}$ (see [8, Lemma 1.2.1]). More precisely, the element $w \in W_n$ corresponds to the element $(m_1, \ldots, m_n) \in X_n^n$ such that $m_i = w(i)$.

For example, $\mathtt{s}_i$ is parameterized by $(1, 2, \ldots, i+1, i, \ldots, n)$ and $\mathtt{r}_1$ is parameterized by $(-1, 2, \ldots, n)$. More generally, if $w \in W_n$ is parameterized by $(m_1, \ldots, m_n) \in X_n^n$, then

$$\begin{array}{ll} w\mathtt{r}_1 & \text{is parameterized by} \quad (-m_1, m_2, \ldots, m_n) \\ w\mathtt{s}_i & \text{is parameterized by} \quad (m_1, \ldots, m_{i+1}, m_i, \ldots, m_n). \end{array} \tag{5.3}$$

Finally we recall [8, Lemma 1.2.2], which is crucial to prove Proposition 5.1

Lemma 5.1 [8, Lemma 1.2.2] *Let $w \in W_n$ parameterized by $(m_1, \ldots, m_n) \in X_n^n$. Then $\ell(w\mathtt{s}_i) = \ell(w) + 1$ if and only if $m_i < m_{i+1}$ and $\ell(w\mathtt{r}_1) = \ell(w) + 1$ if and only if $m_1 > 0$.*

5.1.2 Framed Braid Groups of Type B

We start with the definition of a d–framed version of W_n.

Definition 5.1 The *d–modular framed Coxeter group of type* $\mathtt{B}_n$, $W_{d,n}$, is defined as the group generated by $\mathtt{r}_1, \mathtt{s}_1, \ldots, \mathtt{s}_{n-1}$ and $t_1, \ldots, t_n$ satisfying the Coxeter relations of type B among $\mathtt{r}_1$ and the $\mathtt{s}_i$'s, together with the following relations:

$$\begin{array}{rl} t_i t_j = t_j t_i & \text{for all } i, j, \\ t_i^d = 1 & \text{for all } i, \\ t_j \mathtt{r}_1 = \mathtt{r}_1 t_j & \text{for all } j, \\ t_j \mathtt{s}_i = \mathtt{s}_i t_{s_i(j)} & \text{where } s_i \text{ is the transposition } (i, i+1). \end{array} \tag{5.4}$$

The analogous group defined by the same presentation, where only relations $t_j^d = 1$ are omitted, shall be called *framed Coxeter group of type* $\mathtt{B}_n$ and will be denoted as $W_{\infty,n}$.

Definition 5.2 The *framed braid group of type* $\mathtt{B}_n$, denoted $\mathscr{F}_n^{\mathtt{B}}$, is the group presented by generators $\rho_1, \sigma_1, \ldots, \sigma_{n-1}$, $t_1, \ldots, t_n$ subject to the relations (5.2), together with the following relations:

$$\begin{array}{rl} t_i t_j = t_j t_i & \text{for all } i, j, \\ t_j \rho_1 = \rho_1 t_j & \text{for all } j, \\ t_j \sigma_i = \sigma_i t_{s_i(j)}. & \end{array} \tag{5.5}$$

The *d–modular framed braid group*, denoted $\mathscr{F}^{\mathtt{B}}_{d,n}$, is defined as the group obtained by adding the relations $t_i^d = 1$, for all i, to the above defining presentation of $\mathscr{F}^{\mathtt{B}}_n$.

The mapping that acts as the identity on the generators $\mathtt{r}_1$ and the $\mathtt{s}_i$'s and maps the t_j's to 1 defines a morphism from $W_{d,n}$ onto W_n. Also, we have the natural epimorphism from $\mathscr{F}^{\mathtt{B}}_{d,n}$ onto $W_{d,n}$ defined as the identity on the t_j's and mapping ρ_1 to $\mathtt{r}_1$ and σ_i to $\mathtt{s}_i$, for all i. Thus, we have the following sequence of epimorphisms.

$$\mathscr{F}^{\mathtt{B}}_n \longrightarrow \mathscr{F}^{\mathtt{B}}_{d,n} \longrightarrow W_{d,n} \longrightarrow W_n$$

where the first arrow is the natural projection of $\mathscr{F}^{\mathtt{B}}_n$ to $\mathscr{F}^{\mathtt{B}}_{d,n}$.

5.2 The Cyclotomic Yokonuma–Hecke Algebras

In this section we will review the main results obtained in [2]. First we recall the definition of cyclotomic Yokonuma–Hecke algebra

Definition 5.3 Let d, m and n positive integers. We denote by $\mathrm{Y}(d, m, n)$ to the associative algebra over $\mathscr{R}_m$ generated by framing generators $t_1, \ldots, t_n$, braiding generators $g_1, \ldots, g_{n-1}$ and the cyclotomic generator b_1 subject to the following relations

$$g_i g_j = g_j g_i \quad \text{for} \quad |i-j| > 1, \tag{5.6}$$
$$g_i g_j g_i = g_j g_i g_j \quad \text{for} \quad |i-j| = 1, \tag{5.7}$$
$$b_1 g_i = g_i b_1 \quad \text{for all} \quad i \neq 1, \tag{5.8}$$
$$b_1 g_1 b_1 g_1 = g_1 b_1 g_1 b_1, \tag{5.9}$$
$$t_i t_j = t_j t_i \quad \text{for all} \quad i, j, \tag{5.10}$$
$$t_j g_i = g_i t_{s_i(j)} \quad \text{for all} \quad i, j, \tag{5.11}$$
$$t_i b_1 = b_1 t_i \quad \text{for all} \, i, \tag{5.12}$$
$$t_i^d = 1 \quad \text{for all} \quad i, \tag{5.13}$$
$$g_i^2 = 1 + (\mathsf{u} - \mathsf{u}^{-1}) e_i g_i \quad \text{for all} \quad 1 \leq i \leq n-1, \tag{5.14}$$
$$(b_1 - \mathsf{v}_1) \ldots (b_1 - \mathsf{v}_m) = 0 \tag{5.15}$$

where the e_i's are the elements introduced in [10], that is

$$e_i := \frac{1}{d} \sum_{s=0}^{d-1} t_i^s t_{i+1}^{-s}, \quad 1 \leq i \leq n-1 \tag{5.16}$$

Four linear bases are given for this algebra in [2]. We recall only one of them, which is used by Chlouveraki and Poulain D'Andecy to define a Markov trace in $\mathrm{Y}(d, m, n)$.

For $k = 1, \ldots, n$, we set

$$W_{j,a,b}^{(k)} := g_j^{-1} \ldots g_1^{-1} b_1^a t_1^b g_1 \ldots g_{k-1},$$

where $j = 0, \ldots, k-1$, and $a \in \mathbb{Z}/m\mathbb{Z}$, $b \in \mathbb{Z}/d\mathbb{Z}$. Then, we have the following result

$$B_{d,m,n} := \{ W_{j_n,a_n,b_n}^{(n)} \ldots W_{j_2,a_2,b_2}^{(2)} W_{j_1,a_1,b_1}^{(1)} \mid j_k = 0, \ldots, k-1,\ a_k \in \mathbb{Z}/m\mathbb{Z}, \text{ and } b_k \in \mathbb{Z}/d\mathbb{Z} \}$$

is a basis for $\mathrm{Y}(d, m, n)$, see [2, Sect. 4].

Remark 5.2 To prove the above result, the authors proved first that $B_{d,m,n}$ spans the algebra $\mathrm{Y}(d, m, n)$, and then that its dimension is $(dm)^n n!$. The last result is obtained using tools of representation theory, specifically they constructed a set $\{V_\lambda\}_{\lambda \in \mathscr{P}(d,m,n)}$ of pairwise irreducible non-isomorphic representations of $\mathscr{F}_m \otimes_{\mathscr{R}_m} \mathrm{Y}(d, m, n)$ satisfying the following equation

$$\sum_{\lambda \in \mathscr{P}(d,m,n)} (\dim(V_\lambda))^2 = (dm)^n n!,$$

hence they concluded that $B_{d,m,n}$ is indeed a basis of $\mathrm{Y}_{(d,m,n)}$. Then, in particular, $\mathscr{F}_m \otimes_{\mathscr{R}_m} \mathrm{Y}_{(d,m,n)}$ is a semisimple algebra, for details see [2, Proposition 3.4] and [2, Proposition 4.6]

Using the method of relative traces (see e.g. [2]) it is also proved that the algebra $\mathrm{Y}(d, m, n)$ supports a Markov trace, which we denote by $\mathtt{Tr}_n$. More precisely, this trace is constructed from certain relative traces as follows.

Let z and $x_{a,b}$ with $a \in \{0, \ldots, m-1\}$ and $b \in \{0, \ldots, d-1\}$, be parameters in $\mathscr{R}_m$ such that $x_{0,0} = 1$. The relative traces $\mathtt{tr}_k : \mathrm{Y}(d, m, k) \to \mathrm{Y}(d, m, k-1)$ are given, for any $k \geq 1$, by.

$$\mathtt{tr}_k(W_{j,a,b}^{(k)} w) = z W_{j,a,b}^{(k-1)} \quad \text{if } 0 \leq j \leq k-1 \tag{5.17}$$

$$\mathtt{tr}_k(W_{j,a,b}^{(k)} w) = x_{a,b} w \quad \text{if } j = k-1 \tag{5.18}$$

where $w \in \mathrm{Y}(d, m, k-1)$.

We define

$$\mathtt{Tr}_n := \mathtt{tr}_1 \circ \cdots \circ \mathtt{tr}_n.$$

Then the family $\{\mathtt{Tr}_n\}_{n \geq 1}$ is a Markov trace, for details see [2, Sect. 5].

Finally, as it is usual, using the Jones's recipe, new invariants for framed links in the solid torus are constructed, which are denoted by Γ_m, for details see [2, Sect. 6.3].

Remark 5.3 As we see in [13], to be able define the invariant Γ_m, it is necessary that the trace parameters satisfy a non-linear system of equations. In this case the system is the following

$$\frac{1}{d}\sum_{s=0}^{d-1} x_{0,-s}x_{a,b+s} = x_{a,b}E \qquad \text{for all } a \in \{0,\dots,m-1\} \text{ and } b \in \{0,\dots,d-1\},$$

where $E = \mathtt{Tr}_{i+1}(e_i)$, and it is called the affine E–system, also any solution of this system is referred to by saying that it satisfies the affine E–condition. This system is solved in [2, Sect. 6.5] using only standard tools of linear algebra.

5.3 The Algebra $\mathrm{Y}^{\mathtt{B}}_{d,n}$

We begin this section giving the definition of the framization of the Hecke algebra of type B introduced in [5], denoted by $\mathrm{Y}^{\mathrm{B}}_{d,n}$, which will be the main object of study from now on.

Definition 5.4 Let $n \geq 2$. The algebra $\mathrm{Y}^{\mathrm{B}}_{d,n} := \mathrm{Y}^{\mathrm{B}}_{d,n}(\mathsf{u},\mathsf{v})$, is defined as the algebra over $\mathbb{K} := \mathbb{C}(\mathsf{u},\mathsf{v})$ generated by framing generators $t_1,\dots,t_n$, braiding generators $g_1,\dots,g_{n-1}$ and the loop generator b_1, subject to the following relations

$$g_i g_j = g_j g_i \quad \text{for} \quad |i-j|>1, \tag{5.19}$$
$$g_i g_j g_i = g_j g_i g_j \quad \text{for} \quad |i-j|=1, \tag{5.20}$$
$$b_1 g_i = g_i b_1 \quad \text{for all} \quad i \neq 1, \tag{5.21}$$
$$b_1 g_1 b_1 g_1 = g_1 b_1 g_1 b_1, \tag{5.22}$$
$$t_i t_j = t_j t_i \quad \text{for all} \quad i,j, \tag{5.23}$$
$$t_j g_i = g_i t_{s_i(j)} \quad \text{for all} \quad i,j, \tag{5.24}$$
$$t_i b_1 = b_1 t_i \quad \text{for all i}, \tag{5.25}$$
$$t_i^d = 1 \quad \text{for all} \quad i, \tag{5.26}$$
$$g_i^2 = 1 + (\mathsf{u}-\mathsf{u}^{-1})e_i g_i \quad \text{for all } i, \tag{5.27}$$
$$b_1^2 = 1 + (\mathsf{v}-\mathsf{v}^{-1}) f_1 b_1., \tag{5.28}$$

where the e_i's are the elements defined in (5.16) and

$$f_1 := \frac{1}{d}\sum_{s=0}^{d-1} t_1^s.$$

For $n = 1$, we define $\mathrm{Y}^{\mathrm{B}}_{d,1}$ as the algebra generated by 1, b_1 and t_1 satisfying the relations (5.25), (5.26) and (5.28).

Notice that the elements f_1 and e_i's are idempotents. In Fig. 5.2 we illustrate the generators of the algebra $\mathrm{Y}^{\mathrm{B}}_{d,n}$.

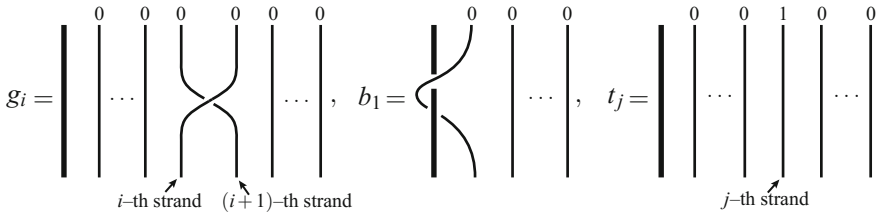


Fig. 5.2 The generators of $\mathrm{Y}^{\mathrm{B}}_{d,n}$ geometrically

Note. By taking $d = 1$, the algebra $\mathrm{Y}^{\mathrm{B}}_{1,n}$ becomes $\mathrm{H}_n(\mathsf{u}, \mathsf{v})$. Further, by mapping $g_i \mapsto h_i$ and $t_i \mapsto 1$, we obtain an epimorphism from $\mathrm{Y}^{\mathrm{B}}_{d,n}$ to $\mathrm{H}_n(\mathsf{u}, \mathsf{v})$. Moreover, if we map the t_i's to a fixed non–trivial d–th root of the unity, we have an epimorphism from $\mathrm{Y}^{\mathrm{B}}_{d,n}$ to $\mathrm{H}_n(\mathsf{u}, 1)$.

Remark 5.4 As we noted previously, the cyclotomic Yokonuma–Hecke algebra also provides a framization of the Hecke algebra of type B. Specifically, if we put $m = 2$, $\mathsf{v}_1 = \mathsf{v}$ and $\mathsf{v}_2 = -\mathsf{v}^{-1}$, then $\mathrm{Y}(d, 2, n)$ is a framization of $\mathrm{H}_n(\mathsf{u}, \mathsf{v})$. But, also we can note that the relation (5.15) doesn't involve framing elements like the defining relation (5.28) of $\mathrm{Y}^{\mathrm{B}}_{d,n}$. In fact $\mathrm{Y}(d, m, n)$ is essentially the Yokonuma–Hecke algebra of type A with the cyclotomic generator and relation attached. This fact makes us think that $\mathrm{Y}^{\mathrm{B}}_{d,n}$, at least algebraically, is a more natural framization for $\mathrm{H}_n(\mathsf{u}, \mathsf{v})$ than $\mathrm{Y}(d, 2, n)$, since the quadratic relation for the loop generator involves the idempotent element f_1, which plays the analogous role as e_i in the quadratic relation of the braiding generators. Then, in some way the braiding generators and the loop generator interact with the framing generators from a more homogeneous way in $\mathrm{Y}^{\mathrm{B}}_{d,n}$.

5.4 A Tensorial Representation of $\mathrm{Y}^{\mathrm{B}}_{d,n}$

In this section we define a tensorial representation of $\mathrm{Y}^{\mathrm{B}}_{d,n}$, the definition of this representation is based on the tensorial representation constructed by Green in [8] for the Hecke algebra of type B and following the idea of an extension of the Jimbo representation of the Hecke algebra of type A to the Yokonuma–Hecke algebra proposed by Espinoza and Ryom–Hansen in [4].

Let V be a $\mathbb{K}$–vector space with basis $\mathscr{B} = \{v_i^r \,;\, i \in X_n,\ 0 \leq r \leq d-1\}$. As usual we denote by $\mathscr{B}^{\otimes k}$ the natural basis of $V^{\otimes k}$ associated to $\mathscr{B}$.

We define the endomorphisms $T, B : V \to V$ by: $(v_i^r)T = \omega^r v_i^r$, and

$$(v_i^r)B = \begin{cases} v_{-i}^r & \text{for } \ i > 0 \text{ and } r = 0, \\ v_{-i}^r + (\mathsf{v} - \mathsf{v}^{-1})v_i^r & \text{for } \ i < 0 \text{ and } r = 0, \\ v_{-i}^r & \text{for } \ r \neq 0. \end{cases}$$

On the other hand we define $G : V \otimes V \to V \otimes V$ by

$$(v_i^r \otimes v_j^s)G = \begin{cases} \mathsf{u} v_j^s \otimes v_i^r & \text{for } i = j \text{ and } r = s, \\ v_j^s \otimes v_i^r & \text{for } i < j \text{ and } r = s, \\ v_j^s \otimes v_i^r + (\mathsf{u} - \mathsf{u}^{-1}) v_i^r \otimes v_j^s & \text{for } i > j \text{ and } r = s, \\ v_j^s \otimes v_i^r & \text{for } r \neq s. \end{cases}$$

For all $1 \leq i \leq n-1$ and $1 \leq j \leq n$, we extend these endomorphisms to the endomorphisms T_i, G_i and B_1 of the n–th tensor power $V^{\otimes n}$ of V, as follows:

$$T_j := 1_V^{\otimes(j-1)} \otimes T \otimes 1_V^{\otimes(n-j)} \quad \text{and} \quad G_i := 1_V^{\otimes(i-1)} \otimes G \otimes 1_V^{\otimes(n-i-1)} \quad \text{and} \quad B_1 := B \otimes 1_V^{\otimes(n-1)},$$

where $1_V^{\otimes k}$ denotes the endomorphism identity of $V^{\otimes k}$.

Theorem 5.1 (See [5, Theorem 1]) *The mapping $b_1 \mapsto B_1$, $g_i \mapsto G_i$ and $t_i \mapsto T_i$ defines a representation Φ of $\mathrm{Y}_{d,n}^{\mathrm{B}}$ in* $\mathrm{End}(V^{\otimes n})$.

We shall finish the section enunciating Proposition 5.1, which is an analogue of [8, Lemma 3.1.4]. This proposition is used in the proof of Theorem 5.2 and describes, through Φ, the action of W_n on the basis $\mathscr{B}^{\otimes n}$.

The defining generators b_1 and g_i of the algebra $\mathrm{Y}_{d,n}^{\mathrm{B}}$ satisfy the same braid relations as the Coxeter generators $\mathtt{r}$ and $\mathtt{s}_i$ of the group W_n. Thus, the well–known Matsumoto's Lemma implies that if $w_1 \dots w_m$ is a reduced expression of $w \in W_n$, with $w_i \in \{\mathtt{r}, \mathtt{s}_1, \dots, \mathtt{s}_{n-1}\}$, then the following element g_w is well–defined:

$$g_w := g_{w_1} \cdots g_{w_m}, \tag{5.29}$$

where $g_{w_i} = b_1$, if $w_i = \mathtt{r}$ and $g_{w_i} = g_j$, if $w_i = \mathtt{s}_j$.

The notation Φ_w stands for the image by Φ of $g_w \in \mathrm{Y}_{d,n}^{\mathrm{B}}$.

Proposition 5.1 (See [5, Proposition 3]) Let $w \in W_n$ parameterized by $(m_1, \dots, m_n) \in X_n^n$. Then

$$(v_1^{r_1} \otimes \cdots \otimes v_n^{r_n})\Phi_w = v_{m_1}^{r_{|m_1|}} \otimes \cdots \otimes v_{m_n}^{r_{|m_n|}}.$$

5.5 Linear Bases for $\mathrm{Y}_{d,n}^{\mathrm{B}}$

In this section we construct two linear bases for the algebra $\mathrm{Y}_{d,n}^{\mathrm{B}}$, which will be denote by C_n and D_n. The first one is used for defining a Markov trace on $\mathrm{Y}_{d,n}^{\mathrm{B}}$ (as we will see in the next section), and the second one plays a technical role for proving that C_n is a linearly independent set.

Set $\overline{b}_1 := b_1$, $\overline{b}_k := g_{k-1} \dots g_1 b_1 g_1 \dots g_{k-1}$, and $b_k := g_{k-1} \dots g_1 b_1 g_1^{-1} \dots g_{k-1}^{-1}$ for all $2 \leq k \leq n$. For all $1 \leq k \leq n$, let us define inductively the sets $N_{d,k}$ by

$$N_{d,1} := \{t_1^m, \overline{b}_1 t_1^m \,;\, 0 \leq m \leq d-1\} \quad \text{and}$$
$$N_{d,k} := \{t_k^m, \overline{b}_k t_k^m, g_{k-1} x \,;\, x \in N_{d,k-1},\ 0 \leq m \leq d-1\} \quad \text{for all } 2 \leq k \leq n.$$

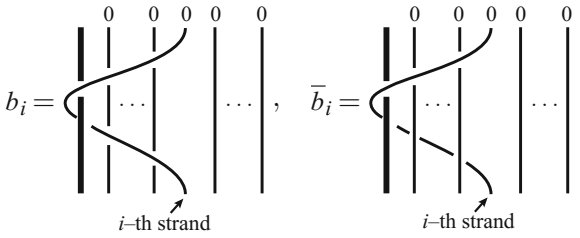


Fig. 5.3 Elements b_i and $\overline{b}_i$ geometrically

Analogously, for all $1 \leq k \leq n$ we define inductively the sets $M_{d,k}$ exactly like $N_{d,k}$'s but exchanging $\overline{b}_k$ by b_k in each case (Fig. 5.3)

Now notice that every element of $M_{d,k}$ has the form $\mathfrak{m}^{+}_{k,j,m}$ or $\mathfrak{m}^{-}_{k,j,m}$ with $j \leq k$ and $0 \leq m \leq d-1$, where

$$\mathfrak{m}^{+}_{k,k,m} := t_k^m, \qquad \mathfrak{m}^{+}_{k,j,m} := g_{k-1} \cdots g_j t_j^m \quad \text{for } j < k,$$

and

$$\mathfrak{m}^{-}_{k,k,m} := t_k^m b_k, \qquad \mathfrak{m}^{-}_{k,j,m} := g_{k-1} \cdots g_j b_j t_j^m \quad \text{for } j < k.$$

Similar expressions exist for elements in $N_{d,k}$ exchanging b_k by $\overline{b}_k$ as well, see [5, Sect. 4.1].

Remark 5.5 Observe that the above elements are the natural analogues of the elements $W^{(k)}_{j,a,b}$ given in Sect. 5.2. Indeed, $\mathfrak{m}^{+}_{k,j,m}$ (respectively $\mathfrak{m}^{-}_{k,j,m}$) correspond to $W^{(k)}_{j-1,0,m}$ (respectively $W^{(k)}_{j-1,1,m}$).

Further, we consider the set $\mathsf{D}_n = \{\mathfrak{n}_1\mathfrak{n}_2 \cdots \mathfrak{n}_n \mid \mathfrak{n}_i \in N_{d,i}\}$, which is a spanning set of $\mathrm{Y}^{\mathrm{B}}_{d,n}$ [5, Proposition 4], moreover using the representation given in the previous section, we can prove that D_n is also a linearly independent set, then we have

Theorem 5.2 *D_n is a linear basis for $\mathrm{Y}^{\mathrm{B}}_{d,n}$. Hence the dimension of $\mathrm{Y}^{\mathrm{B}}_{d,n}$ is $2^n d^n n!$.*

Sketch of the Proof of Theorem 5.2

Firstly, taking into account the structure properties of W_n given in Sect. 5.1.1, it is easy to see that we can write the basis D_n as follows

$$\mathsf{D}_n = \{g_w t_1^{m_1} \cdots t_n^{m_n};\ w \in W_n,\ (m_1, \ldots, m_n) \in (\mathbb{Z}/d\mathbb{Z})^n\}.$$

for details see [5, Proposition 1].

Secondly, we shall use a certain basis $\mathscr{D}$ of V introduced by Espinoza and Ryom–Hansen in [4]. More precisely, $\mathscr{D}$ consist of the following elements:

$$u_k^r = \sum_{i=0}^{d-1} \omega^{ir} v_k^i \tag{5.30}$$

where k is running X_n and $0 \leq r \leq d-1$.

Moreover, it is not difficult to prove that:

(i) $(u_k^r)T = u_k^{r+1}$.
(ii) For a $w \in W_n$ parameterized by $(i_1, \ldots, i_n)$, we have

$$(u_1^0 \otimes \cdots \otimes u_n^0)\Phi_w = u_{i_1}^0 \otimes \cdots \otimes u_{i_n}^0.$$

Note that (ii) follows by Proposition 5.1 and (5.30).

Now, suppose that

$$\sum_{c \in \mathsf{D}_n} \lambda_c c = \sum_{w \in W_n;\ m \in (\mathbb{Z}/d\mathbb{Z})^n} \lambda_{w,m} g_w t_1^{m_1} \ldots t_n^{m_n} = 0,$$

where $m = (m_1, \ldots, m_n)$. Then applying Φ and evaluating in the element $u_1^0 \otimes \cdots \otimes u_n^0$, we have

$$\begin{aligned} \sum \lambda_{w,m}(u_1^0 \otimes \cdots \otimes u_n^0)\Phi(g_w t_1^{m_1} \ldots t_n^{m_n}) &= 0 \\ \sum \lambda_{w,m}(u_1^0 \otimes \cdots \otimes u_n^0)\Phi_w T_1^{m_1} \cdots T_n^{m_n} &= 0 \end{aligned}$$

and by (ii) we obtain

$$\sum \lambda_{i,m} u_{i_1}^0 \otimes \cdots \otimes u_{i_n}^0 T_1^{m_1} \cdots T_n^{m_n} = 0,$$

where $i := (i_1, \ldots, i_n)$ runs in X_n^n and $m := (m_1, \ldots, m_n)$ runs in $(\mathbb{Z}/d\mathbb{Z})^n$. Finally, by using i) the result follows. Further we have the following corollary.

Corollary 5.1 *The representation Φ is faithful.*

Finally, we consider $\mathsf{C}_n = \{\mathfrak{m}_1 \mathfrak{m}_2 \cdots \mathfrak{m}_n \mid \mathfrak{m}_i \in M_{d,i}\}$, which also is a spanning set for $\mathrm{Y}_{d,n}^{\mathrm{B}}$, this fact is proved using the computations listed in [5, Lemmas 5, 6 and 7], then as D_n and C_n have the same cardinality we deduce the following result.

Proposition 5.2 *The set C_n is a basis for $\mathrm{Y}_{d,n}^{\mathrm{B}}$.*

5.6 A Markov Trace on $\mathrm{Y}_{d,n}^{\mathrm{B}}$

In this section we show that the algebra $\mathrm{Y}_{d,n}^{\mathrm{B}}$ supports a Markov trace. This fact was proved by using the method of relative traces, cf. [1, 2]. In few words, the method consists in constructing a certain family of linear maps $\mathrm{tr}_n : \mathrm{Y}_{d,n}^{\mathrm{B}} \longrightarrow \mathrm{Y}_{d,n-1}^{\mathrm{B}}$, called *relative traces*, which builds step by step the desired Markov properties. Finally, the Markov trace on $\mathrm{Y}_{d,n}^{\mathrm{B}}$ is defined by

$$\mathrm{Tr}_n := \mathrm{tr}_1 \circ \cdots \circ \mathrm{tr}_n.$$

Let z be an indeterminate and denote by $\mathbb{L}$ the field of rational functions $\mathbb{K}(z) = \mathbb{C}(\mathsf{u}, \mathsf{v}, z)$. We set $x_0 := 1$ and from now on we fix non–zero parameters $x_1, \ldots, x_{d-1}$ and $y_0, \ldots, y_{d-1}$ in $\mathbb{L}$.

Definition 5.5 For $n \geq 1$, we define the linear functions $\mathrm{tr}_n : \mathrm{Y}^{\mathrm{B}}_{d,n} \longrightarrow \mathrm{Y}^{\mathrm{B}}_{d,n-1}$ as follows. For $n = 1$, $\mathrm{tr}_1(t_1^{a_1}) = x_{a_1}$ and $\mathrm{tr}_1(b_1 t_1^{a_1}) = y_{a_1}$. For $n \geq 2$, we define tr_n on the basis C_n of $\mathrm{Y}^{\mathrm{B}}_{d,n}$ by:

$$\mathrm{tr}_n(w\mathfrak{m}_n) = \begin{cases} x_m w & \text{for } \mathfrak{m}_n = t_n^m \\ y_m w & \text{for } \mathfrak{m}_n = b_n t_n^m \\ z w \mathfrak{m}^{\pm}_{n-1,k,m} & \text{for } \mathfrak{m}_n = \mathfrak{m}^{\pm}_{n,k,m} \end{cases} \tag{5.31}$$

where $w := \mathfrak{m}_1 \cdots \mathfrak{m}_{n-1} \in \mathsf{C}_{n-1}$. Note that (5.31) also holds for $w \in \mathrm{Y}^{\mathrm{B}}_{d,n-1}$.

Using the previous definition and the relations on $\mathrm{Y}^{\mathrm{B}}_{d,n}$, it is not difficult to prove that tr_n has the following properties:

- $\mathrm{tr}_n(XYZ) = X\mathrm{tr}_n(Y)Z$, for all $X, Z \in \mathrm{Y}^{\mathrm{B}}_{d,n-1}$ and $Y \in \mathrm{Y}^{\mathrm{B}}_{d,n}$. (5.32)
- $\mathrm{tr}_n(Xt_n) = \mathrm{tr}_n(t_n X)$, for all $X \in \mathrm{Y}^{\mathrm{B}}_{d,n}$. (5.33)
- $\mathrm{tr}_{n-1}(\mathrm{tr}_n(Xg_{n-1})) = \mathrm{tr}_{n-1}(\mathrm{tr}_n(g_{n-1}X))$, for all $X \in \mathrm{Y}^{\mathrm{B}}_{d,n}$. (5.34)

for details see [5, Lemmas 9, 10 and 13] respectively.

We define $\mathrm{Tr}_n : \mathrm{Y}^{\mathrm{B}}_{d,n} \to \mathbb{L}$ inductively by:

$$\mathrm{Tr}_1 := \mathrm{tr}_1 \quad \text{and} \quad \mathrm{Tr}_n := \mathrm{Tr}_{n-1} \circ \mathrm{tr}_n.$$

Thus, we obtain directly

- $\mathrm{Tr}_n(1) = 1$
- $\mathrm{Tr}_n(x) = \mathrm{Tr}_k(x)$, for $x \in \mathrm{Y}^{\mathrm{B}}_{d,k}$ and $n \geq k$.

Let us denote Tr the family $\{\mathrm{Tr}_n\}_{n\geq 1}$. The following theorem is one of the main results of [5].

Theorem 5.3 Tr *is a Markov trace on* $\{\mathrm{Y}^{\mathrm{B}}_{d,n}\}_{n\geq 1}$. *That is, for every* $n \geq 1$, *the linear map* $\mathrm{Tr}_n : \mathrm{Y}^{\mathrm{B}}_{d,n} \longrightarrow \mathbb{L}$ *satisfies the following rules:*

(i) $\mathrm{Tr}_n(1) = 1$,
(ii) $\mathrm{Tr}_{n+1}(Xg_n) = z\mathrm{Tr}_n(X)$,
(iii) $\mathrm{Tr}_{n+1}(Xb_{n+1}t^m_{n+1}) = y_m \mathrm{Tr}_n(X)$,
(iv) $\mathrm{Tr}_{n+1}(Xt^m_{n+1}) = x_m \mathrm{Tr}_n(X)$,
(v) $\mathrm{Tr}_n(XY) = \mathrm{Tr}_n(YX)$,

where $X, Y \in \mathrm{Y}^{\mathrm{B}}_{d,n}$.

Proof Rules (ii)–(iv) are a direct consequences of (5.32). We prove rule (v) by induction on n. For $n = 1$, the rule holds since $\mathrm{Y}^{\mathrm{B}}_{d,1}$ is commutative. Suppose now that (v) is true for all k less than n. We prove it first for $Y \in \mathrm{Y}^{B}_{d,n-1}$ and $X \in \mathrm{Y}^{\mathrm{B}}_{d,n}$. We have

$$\begin{aligned}\mathrm{Tr}_n(XY) &= \mathrm{Tr}_{n-1}(\mathrm{tr}_n(XY)) \overset{(32)}{=} \mathrm{Tr}_{n-1}(\mathrm{tr}_n(X)Y) \\ &\overset{\text{(induction)}}{=} \mathrm{Tr}_{n-1}(Y\mathrm{tr}_n(X)) \overset{(32)}{=} \mathrm{Tr}_{n-1}(\mathrm{tr}_n(YX)).\end{aligned}$$

Hence, $\mathrm{Tr}_n(XY) = \mathrm{Tr}_n(YX)$ for all $X \in \mathrm{Y}^{\mathrm{B}}_{d,n}$ and $Y \in \mathrm{Y}^{\mathrm{B}}_{d,n-1}$. Now, we prove the rule for $Y \in \{g_{n-1}, t_n\}$. By using (5.33) and (5.34), we get

$$\mathrm{Tr}_n(XY) = \mathrm{Tr}_{n-2}(\mathrm{tr}_{n-1}(\mathrm{tr}_n(XY))) = \mathrm{Tr}_{n-2}(\mathrm{tr}_{n-1}(\mathrm{tr}_n(YX))).$$

In summary, we have

$$\mathrm{Tr}_n(XY) = \mathrm{Tr}_n(YX)$$

for all $X \in \mathrm{Y}^{\mathrm{B}}_{d,n}$ and $Y \in \mathrm{Y}^{\mathrm{B}}_{d,n-1} \cup \{g_{n-1}, t_n\}$. Clearly, having in mind the linearity of Tr_n, this last equality implies that rule (v) holds.

Note 5.1 As we could see in the proof of Theorem 5.3, (5.32), (5.33) and (5.34) give step by step the desired Markov properties for Tr. This fact is the principal advantage of use of the relative traces technique.

Remark 5.6 Let $\mathtt{Tr}$ the Markov trace of $\mathrm{Y}(d, 2, n)$ recalled in Sect. 5.2. Then, considering Remark 5.5, and ignoring the fact that $\mathtt{Tr}$ has a different domain, we can say that the Markov trace defined in this section "coincides" with $\mathtt{Tr}$, by taking $x_{0,m} = x_m$ and $x_{1,m} = y_m$, for all $m \in \{0, \ldots, d-1\}$.

5.7 The E–condition and the F–condition

We want to construct a new invariant for applying Jones's recipe to the pair $(\mathrm{Y}^{\mathrm{B}}_{d,n}$, Tr). For that, as it was seen in [13], we need that the following equation holds

$$\mathrm{Tr}_{n+1}(we_n) = \mathrm{Tr}_n(w)\mathrm{Tr}_{n+1}(e_n) \qquad \text{for all } w \in \mathrm{Y}^{\mathrm{B}}_{d,n}. \tag{5.35}$$

Then we must establish sufficient conditions over the parameters $x_1, \ldots, x_{d-1}$, $y_0, \ldots, y_{d-1} \in \mathbb{L}$, such that (5.35) be satisfied.

With this goal in mind, we define the elements $E^{(k)}$ and $F^{(k)}$ as follows

$$E^{(k)} := \frac{1}{d}\sum_m x_{k+m}x_{d-m} \qquad \text{for} \qquad 0 \le k \le d-1 \tag{5.36}$$

$$F^{(k)} := \frac{1}{d}\sum_{m} x_{d-m}y_{k+m} \qquad \text{for} \qquad 0 \leq k \leq d-1. \tag{5.37}$$

where the summations over m's are regarded modulo d. Note that $E^{(0)} = \mathrm{Tr}_n(e_n)$.

By using the above definitions and the trace rules, we obtain the following results.

Lemma 5.2 *Let $w = w't_n^k$, where $w' \in \mathrm{Y}^{\mathrm{B}}_{d,n-1}$. Then*

$$\mathrm{Tr}_{n+1}(we_n^{(m)}) = \frac{E^{(k+m)}}{x_k}\mathrm{Tr}_n(w).$$

Hence, $\mathrm{Tr}_{n+1}(we_n) = \frac{E^{(k)}}{x_k}\mathrm{Tr}_n(w)$.

Lemma 5.3 *Let $w = w'b_nt_n^k$, where $w' \in \mathrm{Y}^{\mathrm{B}}_{d,n-1}$. Then*

$$\mathrm{Tr}_{n+1}(we_n^{(m)}) = \frac{F^{(k+m)}}{y_k}\mathrm{Tr}_n(w).$$

In particular, we have $\mathrm{Tr}_{n+1}(we_n) = \frac{F^{(k)}}{y_k}\mathrm{Tr}_n(w)$.

Lemma 5.4 *Let $w = w'\mathfrak{m}^{\pm}_{n,k,\alpha}$, with $w' \in \mathrm{Y}^{\mathrm{B}}_{d,n-1}$. Then* $\mathrm{Tr}_{n+1}(we_n) = z\mathrm{Tr}_n(xe_{n-1})$, *where* $x = \mathfrak{m}^{\pm}_{n-1,k,\alpha}w'$.

Considering the previous lemmas, the following definition becomes natural.

Definition 5.6 The E–system is the non–linear system formed by the following $d-1$ equations:

$$E^{(m)} = x_m E^{(0)} \qquad (0 < m \leq d-1)$$

Any solution $(\mathrm{x}_1, \ldots, \mathrm{x}_n)$ of the E–system is referred to by saying that it satisfies the E–condition.

The elements $E^{(k)}$ and the E–system were originally introduced in [13] in order to define new invariants. Specifically, whenever the trace parameters of the Markov trace on the Yokonuma–Hecke algebra satisfy the E–system we have an invariant for framed and classical knots and links in the 3-sphere. Further, in [13, Appendix] P. Gérardin showed that the solutions of the E–system are parameterized by the non-empty subsets of $\mathbb{Z}/d\mathbb{Z}$. Now, we introduce the F–system

Definition 5.7 Assume now that $(\mathrm{x}_1, \ldots, \mathrm{x}_n)$ a solution of the E–system parameterized by the set $S \subseteq \mathbb{Z}/d\mathbb{Z}$. The F–system is the following homogeneous linear system of d equations in $y_0, \ldots, y_{d-1}$:

$$\mathrm{F}^{(m)} = y_m \mathrm{E}^{(0)} \qquad (0 \leq m \leq d-1)$$

where $\mathrm{E}^{(0)}$ and $\mathrm{F}^{(m)}$ are the elements that result from replacing x_i by x_i in (5.36) and (5.37) respectively, that is:

$$\mathrm{E}^{(0)} := \frac{1}{d}\sum_m \mathrm{x}_m \mathrm{x}_{d-m} \quad \text{and} \quad \mathrm{F}^{(m)} := \frac{1}{d}\sum_m \mathrm{x}_{d-m}\mathrm{y}_{k+m}.$$

Also we have that $\mathrm{E}^{(0)} = \frac{1}{|S|}$, see [13, Sect. 4.3]. Thus the F–system is formed by the following equations:

$$\sum_m \mathrm{x}_{d-m}\mathrm{y}_{k+m} - \frac{d}{|S|}\mathrm{y}_m = 0 \qquad (0 \le m \le d-1). \tag{5.38}$$

For any solution $(\mathrm{y}_0, \ldots, \mathrm{y}_n)$ of the F–system we say that it satisfies the F–condition.

Remark 5.7 Note that the F–system is a particular case of the affine E–system recalled in Remark 5.3. More precisely, when the trace parameters are specialized to complex numbers $(\mathrm{x}_1, \ldots, \mathrm{x}_n)$ and $(\mathrm{y}_0, \ldots, \mathrm{y}_n)$ that satisfy the E–condition and the F–condition respectively, then these also satisfy the affine E–system for $m = 2$.

Finally using Lemmas 5.2, 5.3 and 5.4 the following theorem is obtained easily.

Theorem 5.4 *We assume that the trace parameters are specialized to complex numbers* $(\mathrm{x}_1, \ldots, \mathrm{x}_n)$ *and* $(\mathrm{y}_0, \ldots, \mathrm{y}_n)$ *that satisfy the* E*–condition and the* F*–condition respectively. Then*

$$\mathrm{Tr}_{n+1}(we_n) = \mathrm{Tr}_n(w)\mathrm{Tr}_{n+1}(e_n) \qquad \textit{for all} \qquad w \in \mathrm{Y}^{\mathrm{B}}_{d,n}. \tag{5.39}$$

5.7.1 *Solving the* F*–system*

The affine E–system is solved in [2] using only standard tools of linear algebra, then by Remark 5.7, in particular, it provides a solution for the F–system. Now, we give an alternative approach to solve the F–system, following the method of resolution of the E–system done by P. Gérardin, that is, by using some tools from the complex harmonic analysis on finite groups, see [13, Appendix]. We shall introduce first some notations and definitions, necessary for solving the F–system by the method of Gérardin.

We shall regard the group algebra $\Lambda := \mathbb{L}[\mathbb{Z}/d\mathbb{Z}]$, as the algebra formed by all complex functions on $\mathbb{Z}/d\mathbb{Z}$, where the product is the convolution product, that is:

$$(f * g)(x) = \sum_{y \in \mathbb{Z}/d\mathbb{Z}} f(y)g(x-y) \qquad \text{where} \quad f, g \in \Lambda.$$

As usual, we denote by $\delta_a \in \Lambda$ the function with support $\{a\}$.

Also we denote by $\mathbf{e}_a$'s the characters of $\mathbb{Z}/d\mathbb{Z}$, that is:

$$\mathbf{e}_a : b \mapsto \cos\left(\frac{2\pi ab}{d}\right) + i \sin\left(\frac{2\pi ab}{d}\right).$$

The Fourier transform $\mathscr{F}$ on Λ is the automorphism defined by $f \mapsto \widehat{f}$, where

$$\widehat{f}(x) := (f * \mathbf{e}_x)(0) = \sum_{y \in \mathbb{Z}/d\mathbb{Z}} f(y)\mathbf{e}_x(-y).$$

Recall that $(\mathscr{F}^{-1} f)(x) = d^{-1}\widehat{f}(-u)$, where $\widehat{f}(v) = \sum_{u \in G} f(u)\mathbf{e}_v(-u)$. For more properties of Fourier transform over finite groups see [17].

To solve the E–system, Gérardin considered the elements $x \in \Lambda$, defined by $x(k) = x_k$. Then, he interpreted the E–system as the functional equation $x * x = (x * x)(0)x$ with the initial condition $x(0) = 1$. Now, by applying the Fourier transform on this functional equation we obtain $\widehat{x}^2 = (x * x)(0)\widehat{x}$. These last equations imply that $\widehat{x}$ is constant on its support S, where it takes the values $(x * x)(0)$. Thus, we have

$$\widehat{x} = (x * x)(0) \sum_{s \in S} \delta_s.$$

By applying $\mathscr{F}^{-1}$ and the properties listed in the proposition above, Gérardin showed that the solutions of the E–system are parameterized by the non–empty subsets of $\mathbb{Z}/d\mathbb{Z}$. More precisely, for such a subset S, the solution x_S is given as follows.

$$x_S = \frac{1}{|S|} \sum_{s \in S} \mathbf{e}_s.$$

Now, in order to solve the F–system with respect to x_S, we define $y \in \Lambda$ by $y(k) = y_k$. Then we have $F^{(k)} = d^{-1}(x * y)(k)$. So, to solve the F–system is equivalent to solving the following functional equation:

$$x * y = (x * x)(0)y.$$

which, applying the Fourier transform, is equivalent to:

$$\widehat{x}\widehat{y} = (x * x)(0)\widehat{y}.$$

This equation implies that the support of $\widehat{y}$ is contained in the support of $\widehat{x}$. Now, set S the support of $\widehat{x}$. Then we can write $\widehat{y} = \sum_{s \in S} \lambda_s \delta_s$. Finally applying $\mathscr{F}^{-1}$ to the last equation, we get:

$$y = \frac{1}{d} \sum_{s \in S} \lambda_s \mathbf{e}_s.$$

Thus, we have proved the following proposition.

Proposition 5.3 *The solution of the* F*-system with respect to the solution* x_S *of the* E*–system is in the form:*

$$y_S = \sum_{s \in S} \alpha_s \mathbf{e}_s,$$

where the α_s*'s are complex numbers.*

5.8 Knot and Link Invariants from $\mathrm{Y}^{\mathtt{B}}_{d,n}$

In this section we define invariants for knots and links in the solid torus, by using the Jones's recipe applied to the pairs $(\mathrm{Y}^{\mathtt{B}}_{d,n}, \mathrm{Tr}_n)$ where $n \geq 1$. To do that, we fix a subset S of $\mathbb{Z}/d\mathbb{Z}$ from now, and we will consider the trace parameters x_k's and y_k's as the solutions given in the previous section associated to set S. The invariants constructed here will take values in $\mathbb{L}$.

As in the classical case, the closure of a framed braid α of type B (recall Sect. 5.1.2) is defined by joining with simple (unknotted and unlinked) arcs its corresponding endpoints and is denoted by $\widehat{\alpha}$. The result of closure, $\widehat{\alpha}$, is a framed link in the solid torus, denoted ST. This can be understood by viewing the closure of the fixed strand as the complementary solid torus. For an example of a framed link in the solid torus see Fig. 5.4.

By the analogue of the Markov theorem for ST (cf. for example [15, 16]), isotopy classes of oriented links in ST are in bijection with equivalence classes of braids of type B and this bijection carries through to the class of framed links of type B.

We set

$$\lambda_S := \frac{z - (\mathrm{u} - \mathrm{u}^{-1})E_S}{z} \quad \text{and} \quad \Lambda_S := \frac{1}{z\sqrt{\lambda_S}}, \tag{5.40}$$

where $E_S = \mathrm{Tr}(e_i) = 1/|S|$. We are now in the position to define link invariants in the solid torus.

Definition 5.8 For α in $\mathscr{F}^{\mathtt{B}}_n$, the Markov trace Tr with the trace parameters specialized to solutions of the E–system and the F–system, and π the natural epimorphism of $\mathscr{F}^{\mathtt{B}}_n$ onto $\mathrm{Y}^{\mathtt{B}}_{d,n}$ we define

$$\mathscr{X}^{\mathtt{B}}_S(\widehat{\alpha}) := \Lambda_S^{n-1}(\sqrt{\lambda_S})^e \,\mathrm{Tr}(\pi(\alpha)),$$

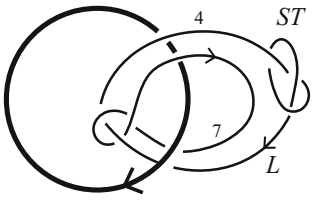


Fig. 5.4 A framed link in the solid torus

where e is the exponent sum of the σ_i's that appear in α. Then $\mathscr{X}_S^{\mathrm{B}}$ is a Laurent polynomial in u, v and z and it depends only on the isotopy class of the framed link $\widehat{\alpha}$, which represents an oriented framed link in ST.

Remark 5.8 The invariants $\mathscr{X}_S^{\mathrm{B}}$, when restricted to framed links with all framings equal to 0, give rise to invariants of oriented classical links in ST. By the results in [3] and since classical knot theory embeds in the knot theory of the solid torus, these invariants are distinguished from the Lambropoulou invariants [15]. More precisely, they are not topologically equivalent to these invariants on *links*.

Remark 5.9 As we said in the Remark 5.7, when we focus in the algebra $Y(d, 2, n)$, the affine E–condition coincide with our conditions (E– and F– condition). Then, we consider the polynomial Γ_m defined in [2, Sect. 6.3] for $m = 2$, which is given by

$$\Gamma_2(\widehat{\alpha}) := \Lambda_S^{n-1}(\sqrt{\lambda_S})^e \, \mathtt{Tr}(\overline{\pi}(\alpha)),$$

where $\overline{\pi} : \mathscr{F}_n^{\mathrm{B}} \to \mathrm{Y}(d, 2, n)$ is the natural algebra epimorphism given by

$$\rho_1 \mapsto b_1, \quad \sigma_i \mapsto g_i, \quad i = 1, \ldots, n-1, \quad \text{and} \quad t_j \mapsto t_j$$

At first sight the invariants look similar, but the structural differences between $\mathrm{Y}_{d,n}^{\mathrm{B}}$ and $\mathrm{Y}(d, 2, n)$ commented in Remark 5.4 make them differ. For example, for the loop generator twice, we have the following

In $\mathrm{Y}_{d,n}^{\mathrm{B}}$	In $\mathrm{Y}(d, 2, n)$
$\mathrm{Tr}(\pi(b_1^2)) = \mathrm{Tr}(1 + (\mathsf{v} - \mathsf{v}^{-1})b_1 f_1)$	$\mathtt{Tr}(\overline{\pi}(b_1^2)) = \mathtt{Tr}(1 + (\mathsf{v} - \mathsf{v}^{-1})b_1)$
$= 1 + \frac{(\mathsf{v}-\mathsf{v}^{-1})}{d} \sum_s \mathrm{Tr}(b_1 t_1^s)$	$= 1 + (\mathsf{v} - \mathsf{v}^{-1})y_0$
$= 1 + \frac{(\mathsf{v}-\mathsf{v}^{-1})}{d} \sum_s y_s$	

Therefore

$$\mathscr{X}_S^{\mathrm{B}}(\widehat{b_1^2}) = 1 + \frac{(\mathsf{v} - \mathsf{v}^{-1})}{d} \sum_s y_s \quad \text{and} \quad \Gamma_2(\widehat{b_1^2}) = 1 + (\mathsf{v} - \mathsf{v}^{-1})y_0.$$

Then clearly for the framed link $\widehat{b_1^2}$, the two invariants have different values, nevertheless in order to do a proper comparison of these invariants is necessary a deeper study.

References

1. Aicardi, F., Juyumaya, J.: Markov trace on the algebra of braids and ties. Mosc. Math. J. **16**(3), 397–431 (2016)
2. Chlouveraki, M., Poulain d'Andecy, L.: Markov trace on affine and cyclotomic Yokonuma–Hecke algebras. Int. Math. Res. Not. **2016**, 4167–4228 (2016)

3. Chlouveraki, M., Juyumaya, J., Karvounis, K., and Lambropoulou, S. Identifying the invariants for classical knots and links from the Yokonuma–Hecke algebras (2015). arXiv:1505.06666
4. Espinoza, J., Ryom-Hansen, S. Cell structures for the Yokonuma–Hecke algebra and the algebra of braids and ties (2016). arXiv:1506.00715
5. Flores, M., Juyumaya, J., Lambropoulou, S. A framization of the Hecke algebra of type B J. Pure Appl. Algebra (2017). https://dx.doi.org/10.1016/j.jpaa.2017.05.006
6. Geck, M., Lambropoulou, S.: Markov traces and knot invariants related to Iwahori–Hecke algebras of type B. J. Reine. Angew. Math. **482**, 191–213 (1997)
7. Goundaroulis, D., Juyumaya, J., Kontogeorgis, A., and Lambropoulou, S. Framization of the Temperley-Lieb Algebra. to appear in Mathematical Research Letters (2016). arXiv:1304.7440v3
8. Green, R.M.: Hyperoctahedral schur algebras. J. Algebr. **192**, 418–438 (1997)
9. Jones, V.: Hecke algebra representations of braid groups and link polynomials. Ann. Math. **126**, 335–388 (1987)
10. Juyumaya, J.: Markov trace on the Yokonuma-Hecke algebra. J. Knot Theory Ramif. **13**, 25–39 (2004)
11. Juyumaya, J., Lambropoulou, S.: An adelic extension of the jones polynomial. In: Banagl, M., Vogel, D. (eds.) Mathematics of Knots. Contributions in the Mathematical and Computational Sciences, vol. 1, pp. 825–840. Springer, Berlin (2009)
12. Juyumaya, J., Lambropoulou, S.: An invariant for singular knots. J. Knot Theory Ramif. **18**(6), 825–840 (2009)
13. Juyumaya, J., Lambropoulou, S.: p-adic framed braids II. Adv. Math. **234**, 149–191 (2013)
14. Juyumaya, J., Lambropoulou, S.: On the framization of knot algebras. In: Kauffman, L., Manturov, V. (eds.) New Ideas in Low-dimensional Topology. Series on Knots and Everything. World Scientific, Singapore (2014)
15. Lambropoulou, S.: Solid torus links and Hecke algebras of B–type. In: Yetter, D.N. (ed.) Proceeding of the Conference on Quantum Topology, pp. 225–245. World Scientific Press, Singapore (1994)
16. Lambropoulou, S.: Knot theory related to generalized and cyclotomic hecke algebras of type b. J. Knot Theory Ramif. **8**(5), 621–658 (1999)
17. Terras, A.: Fourier Analysis of Finite Groups and Applications. Lond. Math. Soc. Stud. Text **43**, (1999)
18. Tom Dieck, T.: Symmetriche Brcken und Knotentheorie zu den Dynkin-Diagrammen vom Typ B. J. reine und angewandte Math **451**, 71–88 (1994)

Chapter 6
Link Diagrams in Seifert Manifolds and Applications to Skein Modules

Boštjan Gabrovšek and Maciej Mroczkowski

Abstract In this survey paper we present results about link diagrams in Seifert manifolds using arrow diagrams, starting with link diagrams in $F \times S^1$ and $N \hat{\times} S^1$, where F is an orientable and N an unorientable surface. Reidemeister moves for such arrow diagrams make the study of link invariants possible. Transitions between arrow diagrams and alternative diagrams are presented. We recall results about the Kauffman bracket and HOMFLYPT skein modules of some Seifert manifolds using arrow diagrams, namely lens spaces, a product of a disk with two holes times S^1, $\mathbb{R}P^3 \# \mathbb{R}P^3$, and prism manifolds. We also present new bases of the Kauffman bracket and HOMFLYPT skein modules of the solid torus and lens spaces.

6.1 Arrow Diagrams of Links in Products and Twisted Products of S^1 and a Surface

Let F be an orientable surface and N an unorientable surface. In this section we recall the construction of arrow diagrams for links in $F \times S^1$, introduced in [13], and $N \hat{\times} S^1$, introducted in [11]. These diagrams are very similar to gleams introduced in [16].

B. Gabrovšek (✉)
Faculty of Mathematics and Physics, University of Ljubljana,
Jadranska 21, 1000 Ljubljana, Slovenia
e-mail: bostjan.gabrovsek@fmf.uni-lj.si

M. Mroczkowski
Institute of Mathematics, Faculty of Mathematics, Physics and Informatics,
University of Gdansk, 80-308 Gdansk, Poland
e-mail: mmroczko@mat.ug.edu.pl

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_6

6.1.1 Arrow Diagrams of Links in $F \times S^1$

Let L be a link in $M = F \times S^1$. We cut M along $F_0 = F \times \{1\}$, $1 \in S^1$, to get $M' = F \times [0, 1]$. By a general position argument we may assume that L intersects F_0 transversally in a finite number of points. In M' the link L becomes L' - a collection of circles and arcs with endpoints coming in pairs $(x, 0)$ and $(x, 1)$, $x \in F$. Let π be the vertical projection from M' onto F. Then $\pi(L')$ is a collection of closed curves. Again, by a general position argument, we may assume that there are only transversal double points in $\pi(L')$ and the endpoints of arcs are projected onto points distinct from these double points. An arrow diagram D of the link L is $\pi(L')$ with some extra information: for double points P, the usual information of over- and undercrossing is encoded depending on the relative height of the two points $\pi^{-1}(P)$ in $F \times [0, 1]$; for points Q that are projections of endpoints of arcs $(x, 0)$ and $(x, 1)$ in L', orient L in such a way that the height drops by 1 in L' when the first coordinate crosses x, and put on Q an arrow indicating this orientation.

Thus, an arrow diagram D is a collection of immersed curves in F, with under- and overcrossing information for double points and some arrows on these curves. For an example see the diagram on Fig. 6.1.

We call an arrow diagram regular, if none of the following forbidden positions appear on the diagram:

(1)
(i) cusps (Fig. 6.2a),
(ii) self-tangency points (Fig. 6.2b),
(iii) triple points (Fig. 6.2c),
(iv) two arrows coincide (Fig. 6.2d),
(v) arrows and crossings coincide (Fig. 6.2e).

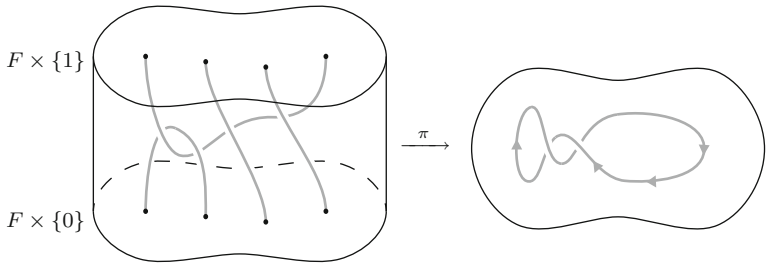


Fig. 6.1 A link in $F \times S^1$ and its diagram

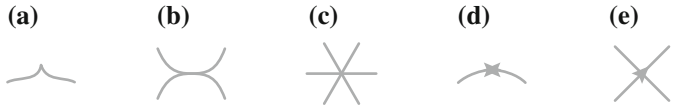


Fig. 6.2 Forbidden positions of regular diagrams in $F \times S^1$

With standard arguments of general position we may assume that every link admits a regular diagram.

We complete this section by providing a list of generalized Reidemeister moves associated with the arrow diagrams. As usual, Reidemeister moves coincide with the change of diagram that occurs when an isotopy of the link is performed in such a way that the configuration of arcs (and arrows) passes through a forbidden position in the projection. We therefore assign each forbidden position an associated Reidemeister move. By general position theory, the ambient isotopy bringing one link diagram to another one passes through only a finite number of such forbidden positions, hence a finite number of Reidemeister-type moves.

As in the classical case, Reidemeister moves Ω_1, Ω_2, and Ω_3 (Fig. 6.3) arise from the forbidden positions (i), (ii), and (iii), respectively. Positions (iv) and (v) generate Reidemeister moves Ω_4 ("arrow annihilation") and Ω_5 ("arrow push"), respectively. Graphical interpretation of moves Ω_4 and Ω_5 are presented in Figs. 6.4 and 6.5, respectively.

We conclude this section with the following Reidemeister-type theorem.

Theorem 6.1 *A link L_1 is ambient isotopic to a link L_2 if and only if an arrow diagram D_1 of L_1 can be obtained from an arrow diagram D_2 of L_2 by a finite series of Reidemeister moves Ω_1 to Ω_5.*

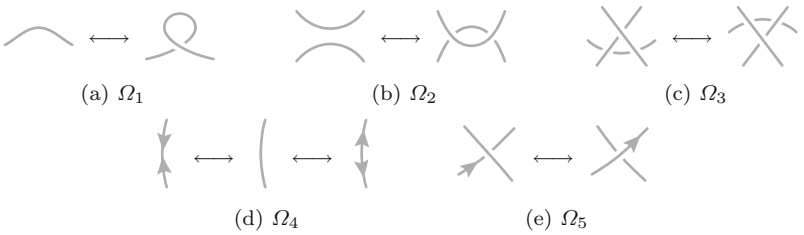


Fig. 6.3 Classical Reidemeister moves Ω_1 – Ω_3 and two "arrow" moves Ω_4 and Ω_5

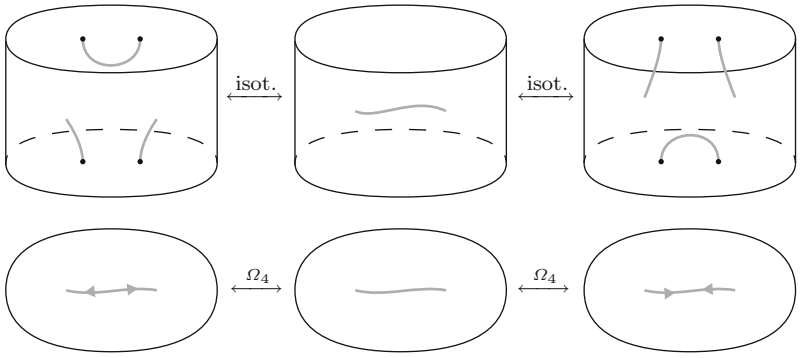


Fig. 6.4 Interpretation of the move Ω_4

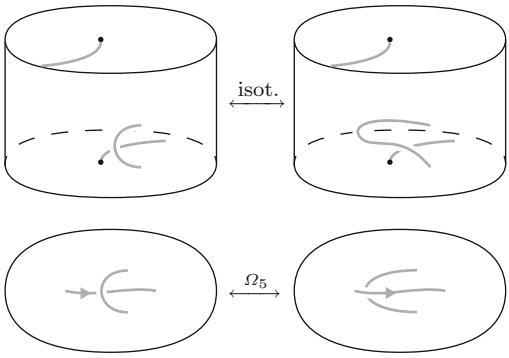


Fig. 6.5 Interpretation of the move Ω_5

6.1.2 Arrow Diagrams of Links in $N\hat{\times}S^1$

An unorientable surface N is obtained from a sphere with n holes S_n by glueing k of the boundary S^1's with antipodal maps (which is equivalent to glueing Mobius bands to these holes). Denote the k boundary S^1's by C. Let $M = N\hat{\times}S^1$ be obtained from $M' = S_n \times S^1$ by glueing $(x, y) \in C \times S^1$ to $(-x, r(y))$, where r is a reflection of S^1 (one may take the complex conjugation). Let L be a link in M. In M', L becomes L', a collection of circles and arcs with endpoints coming in antipodal pairs (x, y) and $(-x, r(y))$ in $C \times S^1$.

For L' in $S_n \times S^1$ one constructs an arrow diagram as in the previous subsection, the only difference being that there are now some arcs with endpoints coming in antipodal pairs. For an example of a diagram see Fig. 6.6.

We call a diagram regular, if, in addition to the list (1), none of the following forbidden positions appear on the diagram:

(vi) tangency points with the boundary (Fig. 6.7a),
(vii) crossing coincides with the boundary (Fig. 6.7b),
(viii) arrow coincides with the boundary (Fig. 6.7c).

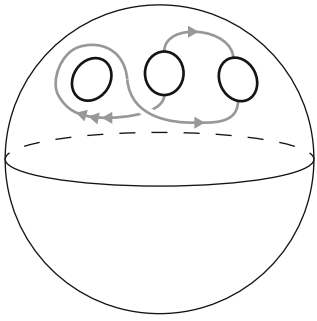


Fig. 6.6 A diagram of a link in $N\hat{\times}S^1$

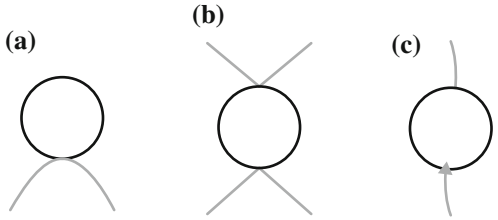


Fig. 6.7 Additional forbidden positions of regular diagrams in $N\hat{\times}S^1$

To these additional forbidden positions (vi), (vii), and (viii), we associate Reidemeister moves Ω_6, Ω_7, and Ω_8 (Fig. 6.8), respectively.

Theorem 6.2 *A link L_1 is ambient isotopic to a link L_2 in $N\hat{\times}S^1$ if and only if an arrow diagram D_1 of L_1 can be obtained from an arrow diagram D_2 of L_2 by a series of Reidemeister moves Ω_1 to Ω_8.*

Example 6.1 The connected sum of two projective spaces $\mathbb{R}P^3\sharp\mathbb{R}P^3$ is also a twisted S^1 over $\mathbb{R}P^2$. Thus, diagrams consists of closed curves and arcs in a disk with endpoints of arcs coming in antipodal pairs on the boundary of the disk. An example of Reidemeister moves between diagrams is presented in Fig. 6.9.

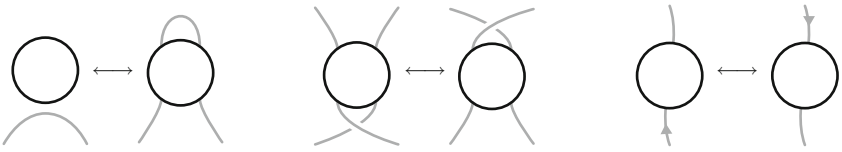


Fig. 6.8 Additional Reidemeister moves

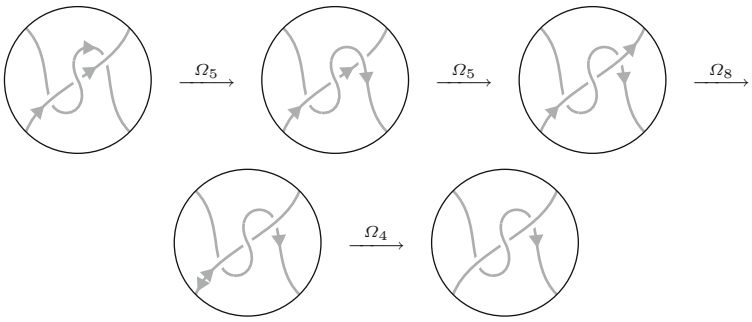


Fig. 6.9 Reidemeister moves on diagrams of a link in $\mathbb{R}P^3\sharp\mathbb{R}P^3$

6.2 Arrow Diagrams for Links in Seifert Manifolds

Definition 6.1 A standard fibered torus corresponding to a pair of coprime integers (q, p), $q > 0$, or $\frac{p}{q}$, is the solid cylinder $D^2 \times I$, where we identify the ends of the solid cylinder with a $2\pi p/q$ twist. Each S^1 fiber comes from q vertical segments in the cylinder, except for the core fiber which comes from the central vertical segment. We call this core fiber *exceptional* if $q > 1$.

Definition 6.2 A Seifert manifold (also a Seifert fibered space) is a closed 3-manifold which can be decomposed into a disjoint union of S^1's (called fibers), such that each tubular neighbourhood of a fiber is a standard fibered torus.

Any orientable Seifert manifold M can be obtained from $F \times S^1$ or $N \hat{\times} S^1$ through a finite number of surgeries (q_i, p_i) on vertical S^1 fibers (see [14]).

To perform such a surgery one removes a vertical solid torus T_1 with longitude l_1 and meridian m_1 on ∂T_1 corresponding to a vertical and horizontal S^1's in the product $F \times S^1$ or $S_n \times S^1$ in the case of $N \hat{\times} S^1$ (see the previous section). Then, another solid torus T_2 with fixed longitude l_2 and meridian m_2 is glued to ∂T_1, so that m_2 is glued to the curve $q_i m_1 + p_i l_1$, see Fig. 6.10. After glueing the meridional disk along this curve, the remaining ball of T_1 is glued to finish the surgery.

The diagram of a link L in the Seifert manifold M is constructed as before, assuming that L misses the exceptional fibers of the surgeries (which can be done by general position). These exceptional fibers project to points in F or N, disjoint from the curves of the diagram. These points appear in the diagrams, together with the type of surgery (q_i, p_i) next to them. See Fig. 6.11 as an example.

If $q_i = 1$, the fiber is not exceptional and, as a shorthand, next to the point onto which it is projected, we put (p_i) instead of $(1, p_i)$. For instance, if M is an S^1-bundle over F, then there is a unique (p) fiber in the diagrams, $p \in \mathbb{Z}$. If $p = 0$ one gets just $S_1 \times F$.

With the added surgery fibers we get an additional forbidden position:

(ix) the surgery point and strand coincide (Fig. 6.12a), (2)

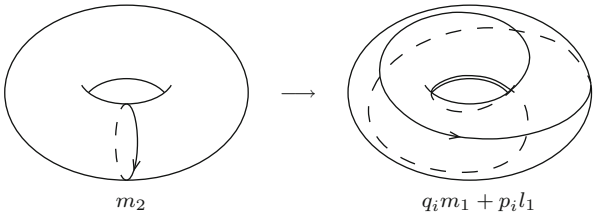


Fig. 6.10 Glueing map of the surgery

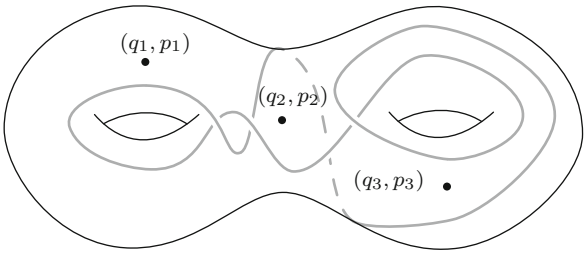
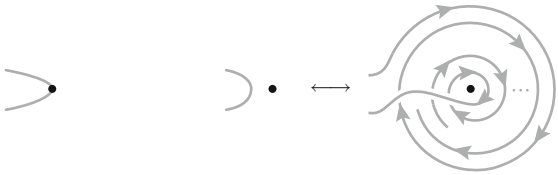


Fig. 6.11 A diagram of a link in a Seifert manifold with three surgeries



(a) Forbidden position (b) $\Omega_{(q,p)}$

Fig. 6.12 A forbidden position and the corresponding slide move $\Omega_{(q,p)}$

which gives rise to the Reidemeister move $\Omega_{(q_i,p_i)}$, corresponding to sliding an arc of the link L through the meridional disk of T_2. The $\Omega_{(q,p)}$ move is shown in Fig. 6.12, it consists of going q times around the exceptional point and adding p arrows uniformly on every $2\pi q/p$ angle.

Theorem 6.3 *A link L_1 is ambient isotopic to a link L_2 in an orientable Seifert manifold M if and only if an arrow diagram D_1 of L_1 can be obtained from an arrow diagram D_2 of L_2 by a series of Reidemeister moves Ω_1 to Ω_8 and $\Omega_{(q_i,p_i)}$.*

Let D be an arrow diagram of a link L. We call a component of D an *oval* if it is a component without crossing with possible arrows on it. We call an oval *nested* if it lies in the interior of a disk bound by another oval.

If n arrows lie consecutively we simplify the diagram by placing only one arrow with an integer n next to it. We interpret a negative integer above an arrow as $|n|$ reversed arrows, see Fig. 6.13. In lens spaces we can alternatively think of $L(p, q)$ as a manifold with Heegaard genus 1 decomposition of two solid tori. We isotope L into the first solid torus and project it to annulus as before and an arrow diagram of a link in $L(p, q)$ can thus be viewed as a diagram on a disk. For $L(p, 1)$, the move $\Omega_{(1,p)}$, which we will denote by $\Omega_{(p)}$ (see [11]), is a winding around the boundary of the disk with p arrows added, see Fig. 6.14.

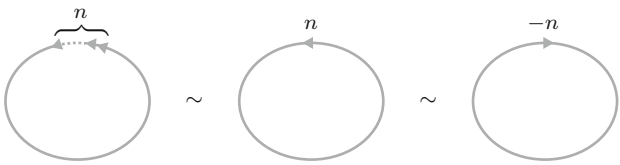


Fig. 6.13 Oval notation

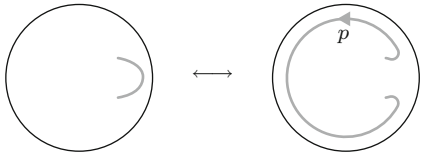
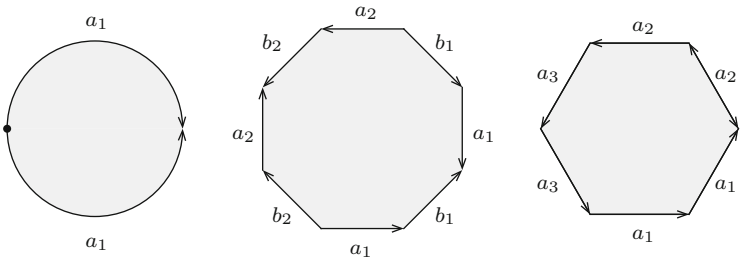


Fig. 6.14 The $\Omega_{(p)}$ move



(a) S^2 (b) $T^2\#T^2$ (c) $K\#\mathbb{R}P^2$

Fig. 6.15 Fundamental polygons, where K is the Klein bottle

6.2.1 Alternative Diagrams for Links in Seifert Manifolds

On some occasions it may be convenient to cut the base surface to its fundamental polygon and get diagrams on a regular n-gon. Such diagrams can be expanded to all Seifert manifolds, orientable and non-orientable, since any Seifert manifold can be obtained from an S^1-bundle over F or N through a finite number of surgeries (q_i, p_i) on vertical S^1 fibers. These diagrams were introduced in [12], also see [5].

We start by taking the fundamental polygon G of the surface, with the standard identification of the edges of G. We distinguish between three cases: either the surface is S^2, a genus $g > 0$ surface F, or a non-orientable surface N, (see Fig. 6.15).

Take $G \times [0, 1]$. By glueing $\{x\} \times \{0\}$ to $\{x\} \times \{1\}$ for each $x \in G$, we get the trivial circle bundle $G \times S^1$. Since G is a disk, we can orient all the fibers $\{x\} \times S^1$ coherently. If two oriented edges a_i and a_i' are identified in G, in order to get F, we can identify the cylinders $a_i \times S^1$ and $a_i' \times S^1$ in two essentially different ways: $a_i \times S^1$ can be glued to $a_i' \times S^1$ by identity or by a reflection on the S^1 component.

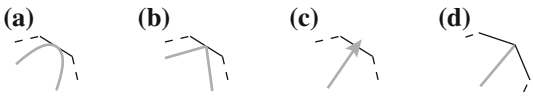


Fig. 6.16 Forbidden positions of regular diagrams

We assign to each edge sign ± 1, which takes the value $+1$ if the identification was made by identity and -1 if the identification was made by reflection.

After the above identifications the resulting space is an S^1-bundle over F. Any Seifert manifold can be obtained from this S^1-bundle by performing a finite number of (q_i, p_i)-surgeries on vertical fibers.

We remark that for an orientable Seifert fibered space M, the base space does not need to be oriented, but the edge signs are determined.

Since the vertical projection maps, as before, an exceptional fiber to a point in the base space, it is enough to specify the image of each exceptional fiber in G, which is done by placing a point on G decorated by the surgery coefficient (q_i, p_i) of the fiber.

We call a diagram regular if, in addition to forbidden positions (1) and (2), none of the following situations occur on the diagram:

(x) border tangency (Fig. 6.16a),
(xi) crossing lies on the border (Fig. 6.16b),
(xii) arrow lies on the border (Fig. 6.16c),
(xiii) arc goes through the basepoint (the preimage of the 0-cell of F) (Fig. 6.16d).

Positions (x), (xi), and (xii) generate the Reidemeister moves Ω_9, $\Omega_{10}^{O/N\pm}$, and $\Omega_{11}^{\pm}$, that act across edges in G (Fig. 6.17). The move Ω_{10} comes in four flavours: the base surface is orientable (O) or non-orientable (N) and the sign of the edge is positive or negative. Similarly, the sign of the move $\Omega_{11}^{\pm}$ corresponds to the sign of the edge we are pushing the arrow through.

Position (xiii) generates the Reidemeister move Ω_{12} that tells us what happens when we push an arc over the basepoint. The move comes in three flavours: if G is a orientable genus $g > 0$ surface we have Ω_{12}^{O}, if G is the 2-sphere we have Ω_{12}^{S}, and if G is a non-orientable surface we have Ω_{12}^{N}. Figure 6.18 shows the geometrical interpretation of Ω_{12} in the case of a double torus.

Considering the arguments above, we can now formulate the following Reidemeister theorem for links in Seifert manifolds.

Theorem 6.4 *Two arrow diagrams for links in a Seifert manifold M represent the same link up to ambient isotopy if and only if they are connected through a finite series of Reidemeister moves $\Omega_1 - \Omega_5$, $\Omega_9 - \Omega_{12}$, and $\Omega_{(q_i, p_i)}$.*

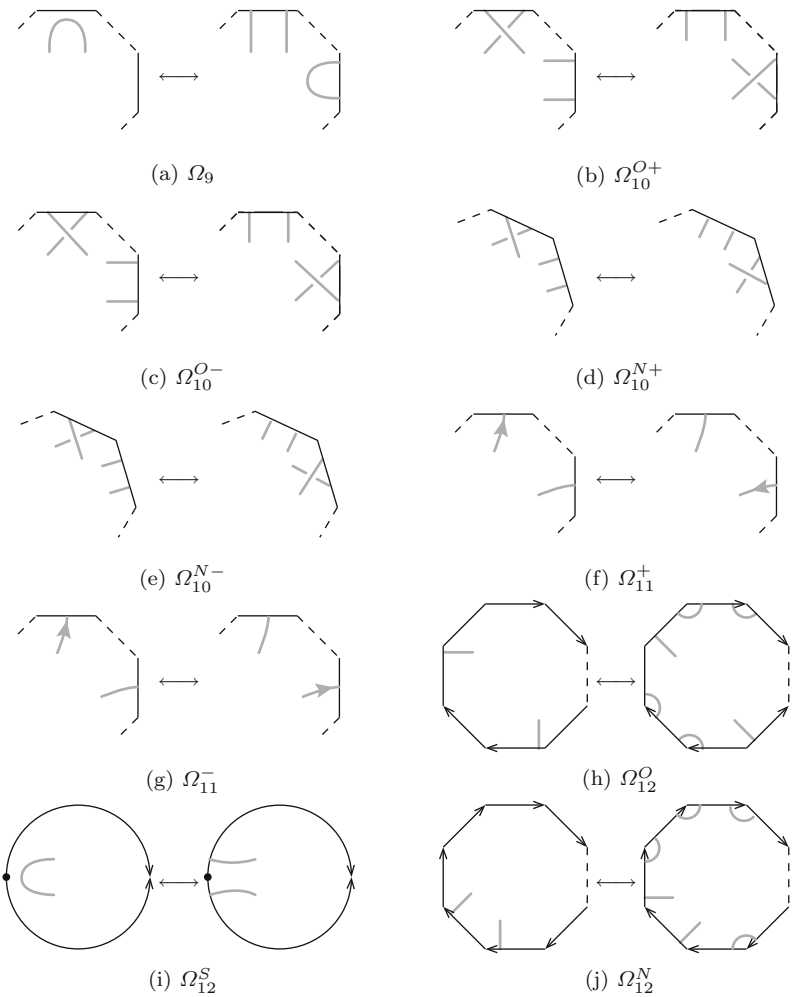


Fig. 6.17 Additional Reidemeister moves

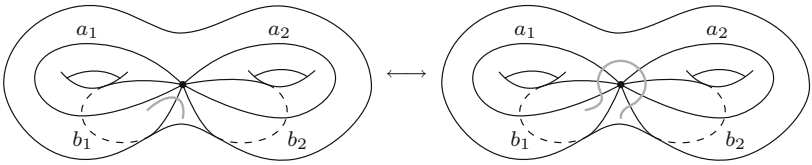


Fig. 6.18 Visualization of the move Ω_{12}^{O}

6.3 Diagram Conversions for Links in the Solid Torus

In this section we define two widely used diagrams of links in the solid torus and links in lens spaces, namely the classical diagram (e.g. [7, 15]) and the mixed link diagram (e.g. [2, 8, 9]). In addition we show how to pass between these diagrams.

6.3.1 Classical Diagrams

Let $T = D^2 \times S^1$ be the solid torus and let $L \subset T$ be a link in T. A classical link diagram of L is the projection of L to the annulus in T that lies in the plane spanned by the longitude l of T, see Fig. 6.19.

As in the previous section, we can think of $L(p, q)$ as glueing of two solid tori. A classical diagram of a link in $L(p, q)$ can again be viewed as a diagram of a link in a solid torus. Alternatively, one can think of $L(p, q)$ as the result of a p/q rational surgery performed on a unknot C in S^3, where again we can project the knot to the annulus $S^3 \setminus C$.

6.3.2 Passing from Classical to Arrow Diagrams

Consider a classical diagram of a link in $D \times S^1$. Each such diagram can be obtained by closing a (n, n)-tangle T with n strands parallel to $\{P\} \times S^1$ for any point $P \in D$. The construction of the corresponding arrow diagram is presented in Fig. 6.20. By rotating the tangle it can be made horizontal, so that in the arrow diagram it will look also like T. In the arrow diagram, the strands become arcs starting at the upper endpoints of the tangle, with arrows on them, then going under the tangle and joining the lower endpoints of the tangle. Applying some Ω_5 moves all arrows can be moved to the upper endpoints of the tangle. Then one notices that the strands go from the upper to the lower endpoints of T with a full negative twist.

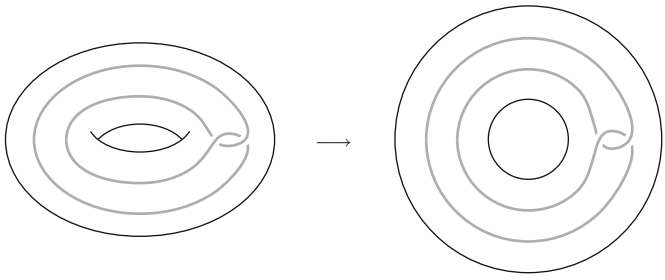


Fig. 6.19 A link in the solid torus (left) and its classical diagram (right)

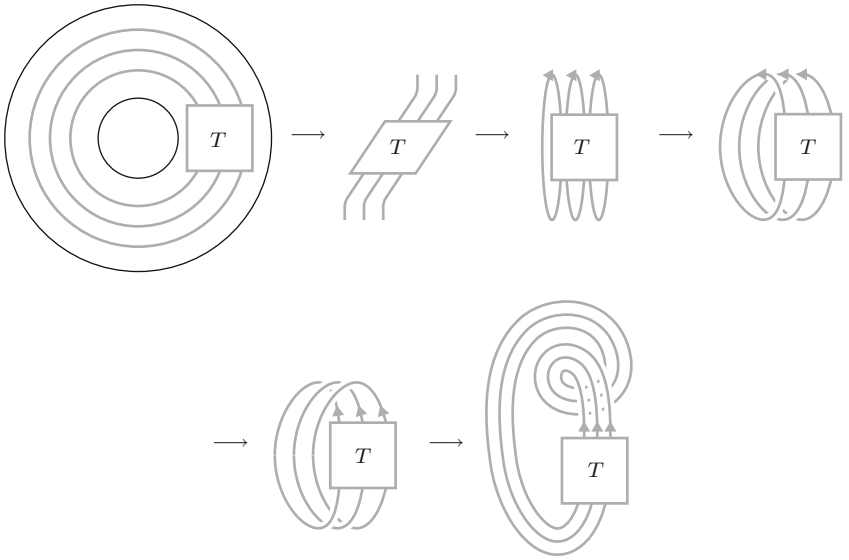


Fig. 6.20 Passing from classical diagrams to arrow diagrams

Summing up, if one has a classical diagram of a link in the solid torus, presented as a closure of a tangle, an arrow diagram for this link is obtained from the same tangle, by adding arrows going up at the upper endpoints of the tangle, making a full negative twist of the strands and closing the tangle *on the left*. By a similar construction, one can add a positive twist and close the tangle on the right.

6.3.3 Passing from Arrow to Classical Diagrams

An arrow diagram can be viewed as an almost flat diagram outside small neighborhoods of the arrows (i.e. it lies in a thickened $D \times \{1\}$ in $D \times S^1$). The neighborhoods of arrows correspond to vertical strands parallel to $\{P\} \times S^1$, $P \in D$. We choose an arbitrary direction in $\mathbb{R}^2$ of the arrow diagram (for instance the vertical one), and rotate the diagram around the axis orthogonal to this direction. One may assume, by general position, that the arrows do not point in the chosen direction. Then the arrows become vertical strands: just before the arrow a vertical strand goes up above other strands, and just after the arrow a vertical strand goes from below under other strands. Closing these vertical strands in an annulus gives a classical diagram from the original arrow diagram. An example is presented in Fig. 6.21.

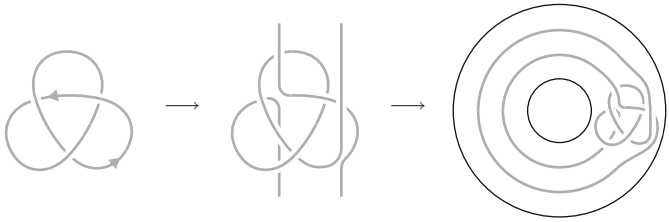


Fig. 6.21 Passing from arrow diagrams to classical diagrams

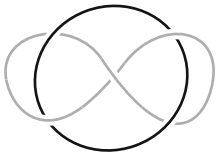


Fig. 6.22 A mixed link diagram for the solid torus

Fig. 6.23 A mixed link diagram for $L(p, q)$

6.3.4 Mixed Link Diagrams

Let $N(U)$ be a the thickened unknot U in S^3. Since $T = S^3 \setminus N(U)$ is a solid torus, we can represent any link L in T with a diagram of $L \cup U$ in the plane. We call L the moving component and U the fixed component. We also keep track of the two types of component, by coloring them with two different colors. Such a diagram is called a mixed link diagram for T (Fig. 6.22).

Every closed oriented 3-manifold M can be constructed from a link $A \subset S^3$ on which we perform (integer or rational) Dehn surgeries on its components [10, 17]. Furthermore, each component of the link A can be assumed to be unknotted. In this way, we can represent a link L in M by a diagram $A \cup L$ in the plane, but again, we keep track of the fixed and moving parts by coloring them with two distinct colors; in addition, we equip each component of A with the surgery coefficient.

For example, the lens space $L_{p,q}$ is the result of a (q, p) or $\frac{p}{q}$ rational surgery on the unknot in S^3. Figure 6.23 shows an example of a knot in $L(p, q)$ (see [1, 8, 9]).

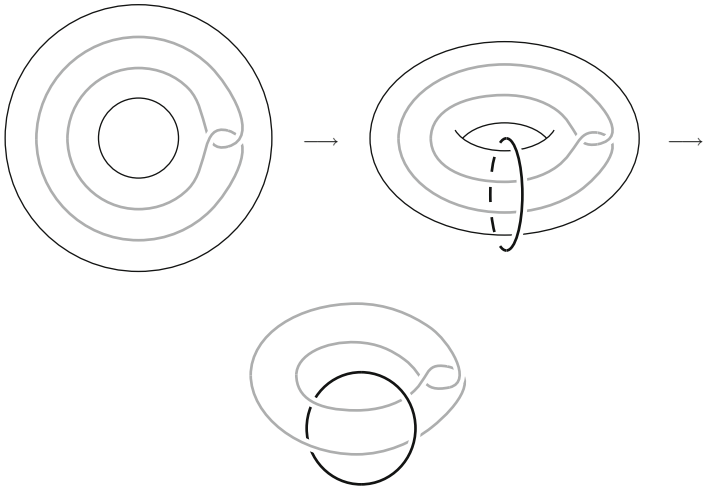


Fig. 6.24 Passing from classical diagrams to mixed link diagrams

6.3.5 Passing from Classical Diagrams to Mixed Link Diagrams for the Solid Torus

Passing from classical diagrams to mixed link diagrams and back is easy: the complement of the solid torus is the thickened $N(U)$, thus U represents the fixed component, see Fig. 6.24.

6.3.6 Passing from Mixed Link Diagrams to Classical Diagrams for the Solid Torus

To pass from mixed link diagrams to classical diagrams, we isotope the moving components of mixed link in such a way, that overcrossings with the fixed component U lie on one side, say on the left and undercrossings lie on the other side. Strands connecting undercrossings and overcrossings should also connect on one side, say the top, see Fig. 6.25.

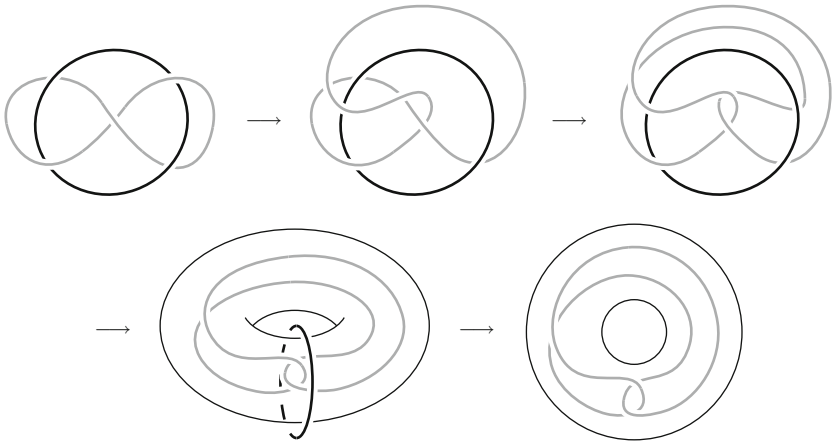
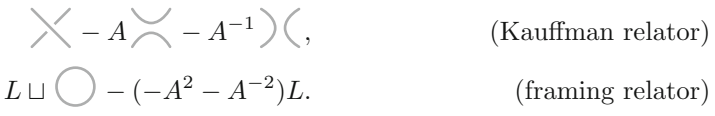


Fig. 6.25 Passing from mixed link diagrams to classical diagrams

6.4 The Kauffman Bracket and HOMFLYPT Skein Modules

Let M be an orientable 3-manifold. Take a coefficient ring R and a unit $A \in R$ (an element with a multiplicative inverse). And let $\mathcal{L}_{\text{fr}}(M)$ be the set of isotopy classes of framed links in M, including the class of the empty link $[\emptyset]$. Let $R\mathcal{L}_{\text{fr}}(M)$ be the free R-module spanned by $\mathcal{L}_{\text{fr}}(M)$.

We would like to impose the Kauffman relation and the framing relation in $R\mathcal{L}_{\text{fr}}(M)$. We therefore take the submodule $\mathcal{S}_{\text{fr}}(M)$ of $R\mathcal{L}_{\text{fr}}(M)$ generated by



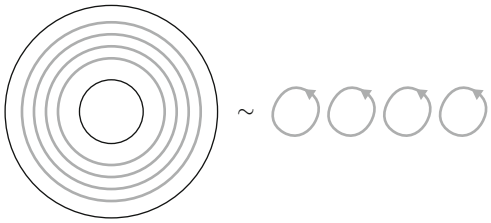
The Kauffman bracket skein module $\mathcal{S}_{2,\infty}(M)$ is defined as $R\mathcal{L}_{\text{fr}}(M)$ modulo these two relations:

$$\mathcal{S}_{2,\infty}(M) = R\mathcal{L}_{\text{fr}}(M)/\mathcal{S}(M).$$

Example 6.2 For the 3-sphere, $\mathcal{S}_{2,\infty}(S^3)$ is a free R-module with a basis consisting of a single element, the equivalence class of the unknot (here, we exclude the empty link). Expressing a link in this basis and, for the unknot, evaluating $[O] = 1$, we get exactly the Kauffman bracket.

Theorem 6.5 ([15]) *Let T be the solid torus, $\mathcal{S}_{2,\infty}(T)$ is freely generated by an infinite set of generators $\{x^n\}_{n=0}^{\infty}$, where x^n, $n > 0$, is a parallel copy of n longitudes of T (see Fig. 6.26) and x^0 is the empty link.*

Fig. 6.26 Generator x^4



Theorem 6.6 ([7]) *$\mathcal{S}_{2,\infty}(L(p,q))$ is freely generated by the set of generators $\{x^n\}_{n=0}^{\lfloor p/2 \rfloor}$, where x^n, $n > 0$, is a parallel copy of n longitudes of $T \subset L(p,q)$ (see Fig. 6.26) and x^0 is the empty link.*

For the HOMFLYPT skein module we take oriented (unframed) links and impose on them the HOMFLYPT skein relation.

Let the ring R have two units $v, z \in R$. Let $\mathcal{L}_{\text{or}}(M)$ be the set of isotopy classes of oriented links in M, including the class of the empty link $[\emptyset]$ and let $R\mathcal{L}_{\text{or}}(M)$ be the free R-module spanned by $\mathcal{L}_{\text{or}}(M)$.

We take the submodule $\mathcal{S}(M)$ of $R\mathcal{L}_{\text{or}}(M)$ generated by the expressions

(HOMFLYPT relator)

We also add to $\mathcal{S}(M)$ the HOMFLYPT relation involving the empty knot:

(empty knot relator)

The HOMFLYPT skein module $\mathcal{S}_3(M)$ of M is $R\mathcal{L}_{\text{or}}(M)$ modulo the above relations:

$$\mathcal{S}_3(M) = R\mathcal{L}(M)/\mathcal{S}(M).$$

Let $t_n, n \in \mathbb{Z}$ be the knot in Fig. 6.27a, b, note that for $n < 0$, t_n is $t_{|n|}$ with reversed orientation.

Theorem 6.7 ([6, 15]) *The HOMFLYPT skein module of the solid torus T is a free R-module, generated by the infinite set*

$$\mathcal{B} = \{t_{k_1} \cdots t_{k_s} \mid k_i \in \mathbb{Z} \setminus \{0\},\ k_1 \leq \cdots \leq k_s\} \cup \{\emptyset\}.$$

Fig. 6.27 HOMFLYPT skein module generators

(a) t_n (b) t_{-n} (c) $t_4^2 t_{-3}$

6.4.1 Results for Kauffman Bracket and HOMFLYPT Skein Modules Using Arrow Diagrams

The first result, using arrow diagrams, was to compute the Kauffman bracket skein module of a product of S^1 with a disk with two holes. There is an incompressible, non-boundary parallel torus immersed (not embedded) in such a manifold. In several examples of manifolds, such surfaces, when embedded, yield torsion, so it was interesting to consider a case with an immersion instead of an embedding. There was no torsion in this case:

Theorem 6.8 ([13]) *Let M be the product of a disk with two holes and S^1. Then $\mathcal{S}_{2,\infty}(M)$ is freely generated by an infinite set of generators.*

Now, consider links in the manifold $\mathbb{R}P^3 \sharp \mathbb{R}P^3$ (see Example 6.1). Let $t = -A^{-3}x$ (see Fig. 6.26). We can view t as an oval with one arrow, obtained from x by adding a negative kink. The multiplication of two arrow diagrams of links in the solid torus consists in putting them in two disjoint disks. Thus, for example, $x^3 = x^2x = xxx$ is the diagram were the three x's are in three disjoint disks.

Let Q_n, $n \in \mathbb{N} \cup \{0\}$ be defined by:

$$Q_0 = 1, \ Q_1 = t \text{ and } Q_n = tQ_{n-1} - Q_{n-2}$$

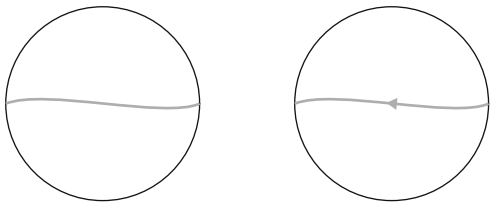
Let E be the knot with diagram in Fig. 6.28a: an arc with two antipodal endpoints and no crossings. Let E' be the knot with diagram in Fig. 6.28b: an arc as in E with an added arrow.

Theorem 6.9 ([11]) $\mathcal{S}_{2,\infty}(\mathbb{R}P^3 \sharp \mathbb{R}P^3) = R \oplus R \oplus R[t]/S$, *where $R = \mathbb{Z}[A, A^{-1}]$ and S is the submodule of $R[t]$ generated by:*

$$(A^{n+1} + A^{n-1})(Q_n - 1) - 2(A + A^{-1}) \sum_{k=1}^{\frac{n}{2}} A^{n+2-4k}, \text{ for } n \geq 2 \text{ even},$$

$$(A^{n+1} + A^{n-1})(Q_n - t) - 2t \sum_{k=1}^{\frac{n-1}{2}} A^{n+1-4k}, \text{ for } n \geq 3 \text{ odd}.$$

The generators of the two R's are E and E'.



(a) E (b) E'

Fig. 6.28 Generators E and E'

By adding (q, p) fibers to $\mathbb{R}P^3 \sharp \mathbb{R}P^3$ one gets the prism manifolds. If $q = 1$, the prism manifold is denoted by M_p (it has no exceptional fibers, just as $L(p, 1)$).

Theorem 6.10 ([12]) *$\mathcal{S}_{2,\infty}(M_p)$ is a free R-module generated by $\emptyset$, x, x^2,...,$x^{1+\lfloor \frac{p}{2} \rfloor}$, E and (if p is even) E'. Thus, it has $3 + \lfloor \frac{p}{2} \rfloor$ generators, if p is odd, and $4 + \frac{p}{2}$ generators, if p is even.*

Let us now turn to HOMFLYPT skein modules. Consider now links in the lens space $L_{p,1}$. Recall, from Sect. 6.2 that arrow diagrams of such links lie in the disk and there is an additional slide move $\Omega_{(p)}$ (Fig. 6.14).

Let

$$\mathcal{B}_p = \{t_{k_1} \cdots t_{k_s} \mid k_i \in \mathbb{Z} \setminus \{0\},\ -\frac{p}{2} < k_1 \leq \cdots \leq k_s \leq \frac{p}{2}\} \cup \{\emptyset\}.$$

Using the move $\Omega_{(p)}$, elements in $\mathcal{B}$ can be expressed with elements in $\mathcal{B}_p$. In fact, one gets more:

Theorem 6.11 ([4]) *$\mathcal{S}_3(L_{p,1})$ is free with basis $\mathcal{B}_p$.*

See also [3] for the braid approach to the HOMFLYPT skein module of $L(p, 1)$.

6.5 Alternative Bases for the Kauffman Bracket Skein Modules

We denote by P_n an oval with n counterclockwise arrows ($|n|$ clockwise arrows if $n < 0$) and by y_n a nested system of n ovals with one counterclockwise arrow on each if $n > 0$ or clockwise arrow if $n < 0$. By convention y_0 is the empty link. See Fig. 6.29.

We exhibit some alternative bases for the Kauffman bracket skein modules of the solid torus and lens spaces. The results follow easily from the following lemmas.

Lemma 6.1 ([13]) *In $\mathcal{S}_{2,\infty}(T)$ one can revert x in the sense that*

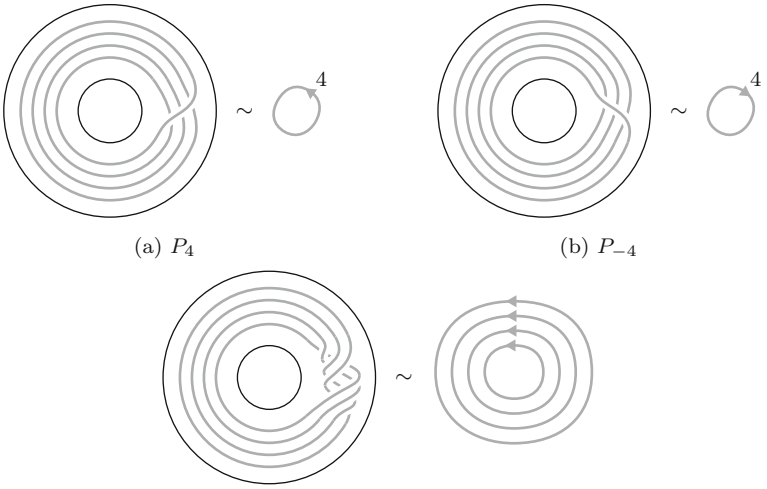
$$x = P_1 = A^6\, P_{-1}.$$

Lemma 6.2 ([13]) *From an oval P_n, $n > 0$, we can push off an arrow using Reidemeister moves and skein relations in the sense that*

$$P_n = -A^{-2} P_{n-1} x - A^2 P_{n-2}.$$

Similarly for P_n, $n < 0$, it holds

$$P_n = -A^2 P_{n+1} P_{-1} - A^{-2} P_{n+2}.$$



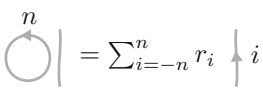
(a) P_4 (b) P_{-4}

(c) y_4

Fig. 6.29 Diagrams of P_4, P_{-4}, and y_4

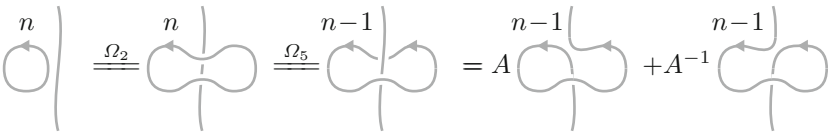
The following lemma illustrates the methods used when doing calculations in $\mathcal{S}_{2,\infty}(T)$.

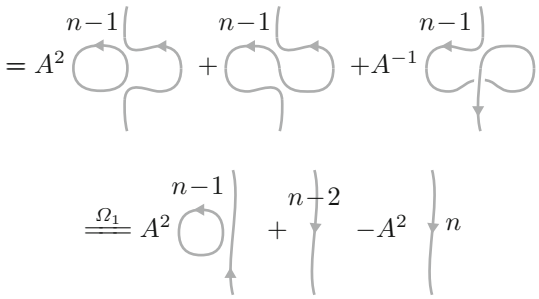
Lemma 6.3 *An oval can be pushed to an adjacent arc in $\mathcal{S}_{2,\infty}(T)$ using skein relations in the sense that*



for $n > 0$ and $r_i \in R$, furthermore it holds that $r_n = -A^{2n+2}$ and $r_{-n} = -A^2$.

Proof First we make an Ω_2 move through the arc and push an arrow through, then we resolve the newly changed crossing by the Kauffman skein relation. For the first term we resolve the remaining crossing and for the second term we push the arrow through the remaining crossing and perform an Ω_1:





If $n = 1$, the first two lines of the equations above show that $r_1 = -A^4$ and $r_{-1} = -A^2$. For $n > 1$, we repeat this process with the oval with $n - 1$ arrows to the left of the strand with one arrow up. Then, we iterate this process with ovals with less and less arrows until there is only one arrow left. At each step an arrow will be transferred from the oval to the strand upwards with coefficient A^2. The coefficient of r_n is thus $(A^2)^{n-1}(-A^4) = -A^{2n+2}$. It follows from the equations above that in the procedure we get n arrows pointing downwards only at the first step, so r_{-n} is $-A^2$.

In addition to Turaev's basis of $\mathcal{S}_{2,\infty}(T)$ in [15] and Hoste/Przytycki's basis of $\mathcal{S}_{2,\infty}(L(p,q))$ in [7], we show in the next propositions some alternative choices for bases of these skein modules.

Proposition 6.1 *The set* $\{P_n\}_{n=1}^{\infty} \cup \{\emptyset\}$ *forms a free basis of* $\mathcal{S}_{2,\infty}(T)$.

Proof From Lemma 6.2 it follows that, for $n > 0$, P_n is a polynomial of degree n with leading invertible coefficient $(-1)^{n+1}A^{-2n+2}$. Thus, the P_n's can be expressed with the x^n's with an upper triangular matrix with invertible coefficients on the diagonal. It follows that $\{P_n\}_{n=1}^{\infty} \cup \{\emptyset\}$ is a basis of $\mathcal{S}_{2,\infty}(T)$.

Proposition 6.2 *The set* $\{P_{-n}\}_{n=1}^{\infty} \cup \{\emptyset\}$ *forms a free basis of* $\mathcal{S}_{2,\infty}(T)$.

Proof From Lemma 6.1, $P_{-1} = A^{-6}P_1 = A^{-6}x$. From Lemma 6.2 it follows that, for $n < 0$, P_n is a polynomial of degree $|n|$ with leading invertible coefficient $(-1)^{n+1}A^{-2n-8}$. The rest of the proof is the same as in the preceding proposition.

Proposition 6.3 *The set* $\{y_n\}_{n=0}^{\infty}$ *forms a free basis of* $\mathcal{S}_{2,\infty}(T)$.

Proof We will show that y_n is a polynomial of degree n in x with an invertible leading coefficient. Then we will be done, just as in the proofs of the preceding two propositions.

The proof is by induction on n. Obviously it holds for $n = 1$ because $y_1 = x$. It will be useful to have this more general induction hypothesis, for $k, l \geq 0, k + l \leq n$:

$(H_{k,l})$: x^k nested inside y_l is a polynomial of degree $k + l$ with an invertible leading coefficient.

For instance x nested inside y_l is just y_{l+1}. We perform an inner induction on l. The hypothesis is true for $l = 0$ because x^k nested in y_0 is just x^k. Suppose $(H_{k',l'})$ is true for all $k' + l' < n$ and that it is also true for all $x^{k'}$ nested in $y_{l'}$ where $l' < l$, $k' + l' = n$.

Now consider x^k nested in y_l, $k + l = n$. Use Lemma 6.3 to push the k x's into the most nested oval of y_l. Pushing one such x, the only nonzero coefficients will be $r_1 = -A^4$ and $r_{-1} = -A^2$ corresponding to adding a counterclockwise or clockwise arrow respectively to this most nested oval of y_l. As originally there is one countercklockwise arrow on this oval, this oval will become P_m when the x's are pushed into it, with $-k + 1 \leq m \leq k + 1$. We know that P_m is a polynomial in x of degree $|m|$ with an invertible leading coefficient. By induction on n, the terms with $m < k + 1$ will have degree less than n (some arrows are cancelled). The remaining term, $m = k + 1$, corresponds to all x's being pushed as counterclockwise arrows, yielding $(-A^4)^k$ times P_{k+1} nested inside y_{l-1}. Now P_{k+1} is of degree $k + 1$ with an invertible leading coefficient. The terms of P_{k+1} of degree less than $k + 1$ nested in y_{l-1} will have degree less than n by induction. Finally, the only term remaining is an invertible coefficient times x^{k+1} nested in y_{l-1}, which is of degree n times an invertible coefficient by induction on l. Thus $(H_{k,l})$ is true.

Proposition 6.4 *The set* $\{y_{-n}\}_{n=0}^{\infty}$ *forms a free basis of* $\mathcal{S}_{2,\infty}(T)$.

Proof The proof mirrors that of the previous proposition, using $P_{-1} = A^{-6}x$ instead of x.

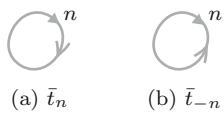
As, for $n > 0$, P_n, P_{-n}, y_n and y_{-n} are all polynomials of degree n in x with an invertible leading coefficient, the following proposition follows from Theorem 6.6.

Proposition 6.5 *The sets* $\{P_n\}_{n=1}^{\lfloor p/2 \rfloor} \cup \{\emptyset\}$, $\{P_{-n}\}_{n=1}^{\lfloor p/2 \rfloor} \cup \{\emptyset\}$, $\{y_n\}_{n=0}^{\lfloor p/2 \rfloor}$, *and* $\{y_{-n}\}_{n=0}^{\lfloor p/2 \rfloor}$ *are all free bases of* $\mathcal{S}_{2,\infty}(L(p, q))$.

6.6 Alternative Bases for the HOMFLYPT Skein Modules

The following two lemmas are from [4]. Using them, we will exhibit new bases for $\mathcal{S}_3(L_{p,1})$. Recall that t_n, $n \in \mathbb{Z} \setminus \{0\}$, stands for an oval with $|n|$ counter clockwise arrows on it, which is oriented in a counterclockwise way if $n > 0$ and in a clockwise way otherwise.

Denote by $\bar{t}_n$, $n \in \mathbb{Z} \setminus \{0\}$, the oval obtained from t_n by reversing all arrows and the orientation (Fig. 6.30).



(a) $\bar{t}_n$ (b) $\bar{t}_{-n}$

Fig. 6.30 Knots $\bar{t}_n$ and $\bar{t}_{-n}$ for $n > 0$

Lemma 6.4 *In $\mathcal{S}_3(T)$ we can revert clockwise arrows on an oval in the sense that for $n > 0$*

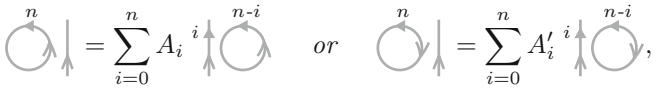
$$\bar{t}_n = \sum_i A_i T_i \text{ and } \bar{t}_{-n} = \sum_i A'_i T'_i$$

where $A_i, A'_i \in R$, $T_i, T'_i \in \mathcal{B}$. More precisely, any $T_i = t_{k_1} \cdots t_{k_s}$, where all $k_i > 0$ and $k_1 + \cdots + k_s = n$. Similarly, any $T'_i = t_{k_1} \cdots t_{k_s}$, where all $k_i < 0$ and $k_1 + \cdots + k_s = -n$.

The following lemma is a reformulation of Lemma 6.2 in [4], emphasizing the orientations and making the coefficients A_0 and A'_0 explicit.

Lemma 6.5 *Let D be a diagram of a link L with an oval containing n arrows, $n \in \mathbb{Z}$, and a strand adjacent to it, that may contain arrows outside the drawn region. We say that the orientations of the oval and the strand agree if the oval has a counterclokwise orientation and the strand right to it is oriented upwards or if both have opposite orientations. Otherwise we say that their orientations disagree.*

The oval can be pushed through the strand. There are four possible configurations of orientations, but we will explicitly point out two, with $n > 0$ and the orientations agreeing or disagreeing. The following formulas hold in $\mathcal{S}_3(T)$:



where $A_i, A'_i \in R$. Furthermore, we can keep track of the coefficients A_0 and A'_0 getting $A_0 = v^{2n}$ and $A'_0 = v^{-2n}$. For the remaining two configurations the transfer of arrows is as in the first formula when the orientations agree and as in the second formula when the orientations disagree.

In [2] a new basis for $\mathcal{S}_3(T)$ is presented. Translating this basis into arrow diagrams gives a basis $\mathcal{B}'$. Its elements have diagrams consisting of $s \in \mathbb{N}$ concentric ovals with $k_i \in \mathbb{Z} \setminus \{0\}$ arrows on each oval, denoted $t_{k_1,k_2,\ldots,k_s}$, satisfying $k_1 \geq k_2 \geq \ldots \geq k_s$, with k_1 arrows on the most nested oval, k_2 arrows on the next one and, so on, with k_s arrows on the oval containing all other ovals, see Fig. 6.31. With this notation

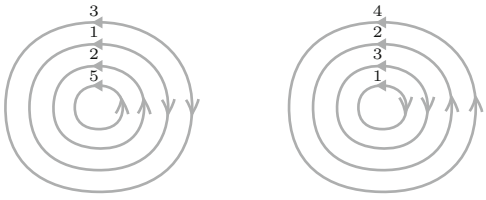


Fig. 6.31 An element from $\mathcal{B}'$ and $\mathcal{B}''$

(a) $t_{5,2,-1,-3} \in \mathcal{B}'$ (b) $t_{-1,-3,2,4} \in \mathcal{B}''$

$$\mathcal{B}' = \{t_{k_1,\ldots,k_s} \mid k_i \in \mathbb{Z} \setminus \{0\},\ s \in \mathbb{N},\ k_1 \geq k_2 \geq \cdots \geq k_s\}.$$

We will exhibit a similar basis of $\mathcal{S}_3(T)$, using an order relation similar to the one used in [2]:

$$\mathcal{B}'' = \{t_{k_1,\ldots,k_s} \mid k_i \in \mathbb{Z} \setminus \{0\}, s \in \mathbb{N}, 0 > k_1 \geq \ldots \geq k_l, 0 < k_{l+1} \leq \ldots \leq k_s\}.$$

We introduce an order relation on configurations of ovals with arrows, i.e. on diagrams with no crossings, and such that there are no ovals with zero arrows. This relation is defined in a lexicographical way by considering in this order:

1. The number of all arrows counted as positive for all ovals independently of orientations.
2. The number of ovals.
3. A lexicographical ordering of ovals with positive arrows: from ovals with the smallest number of arrows to ovals with the largest number of arrows.
4. A lexicographical ordering of ovals with negative arrows in the same way as in (3), but taking absolute values (i.e. the number of arrows) into account.

To illustrate 3, in $\mathcal{B}$ one has, $t_1^2 t_3^2 > t_1^2 t_2 t_4$, because $(1, 1, 3, 3) > (1, 1, 2, 4)$ (lexicographically). To illustrate 4, $t_{-2}^2 t_{-4}^2 > t_{-2}^2 t_{-3} t_{-5}$ because $(|-2|, |-2|, |-4|, |-4|) > (|-2|, |-2|, |-3|, |-5|)$.

The order defined above becomes a total order when restricted to $\mathcal{B}$ or $\mathcal{B}''$. Indeed, if two configurations of ovals have the same order, they will have the same series of arrows on ovals, for example $(-3, -3, -1, 1, 4, 4)$. Restricting to $\mathcal{B}$ or $\mathcal{B}''$ this determines completely the diagram.

Lemma 6.5 can be refined to take into account this ordering, in the following sense:

Lemma 6.6 *Let D be a diagram with no crossings, with an oval t_i containing no other ovals, next to an oval t_j (it can be nested in t_j or not). Suppose that $0 < |i| \leq |j|$ if i and j have the same sign, or $i < 0 < j$ otherwise. Then one can push t_i through t_j getting $v^{\pm 2i}$ times the diagram in which the whole t_i is pushed plus terms of lower order.*

Proof If $i < 0 < j$, then it follows from Lemma 6.5 that all terms will have less arrows than D, except for the term corresponding to the whole t_i being pushed, which comes with a factor v^{-2i}. If $0 < i \leq j$, then, from the same lemma, it follows that we will have a term as before, this time with coefficient v^{2i}, plus terms for which some $a > 0$ arrows will be transferred from t_i to t_j. In the lexicographical order the change will be from $(\ldots, i, \ldots, j, \ldots)$ to $(\ldots, i - a, \ldots, j + a, \ldots)$ which is of lower order. Similarly for $j < i < 0$ there will be a change from $(\ldots, j, \ldots, i, \ldots)$ to $(\ldots, j - a, \ldots, i + a, \ldots)$ which is again of lower order.

Theorem 6.12 *$\mathcal{B}''$ is a basis of the free skein module $\mathcal{S}_3(T)$.*

Proof Let $T = t_{k_1} \cdots t_{k_s} \in \mathcal{B}$, $k_1 \leq k_2 \leq \cdots \leq k_s$, $k_i \in \mathbb{Z} \setminus \{0\}$. We construct first a function $F : \mathcal{B} \to R\mathcal{B}''$. If $k_l < 0$ and $k_{l+1} > 0$, $F(T)$ will be equal to $t_{k_l,k_{l-1},\cdots,k_1,k_{l+1},\cdots,k_s}$ (which has the same order as T) times an invertible coefficient plus terms of lower order. F is defined by induction on the order. F is the identity if the number of arrows is 0 (empty diagrams) or if the number of ovals is 1.

If $k_s > 0$ push all other ovals into t_{k_s}, using Lemma 6.6. We get v to some power times t_{k_s} around all other ovals plus terms of lower order. Reexpress these terms of lower order in $\mathcal{B}$ by pushing ovals out of t_{k_s} again. As we push through the oval with a maximum of arrows, Lemma 6.6 guarantees that the order cannot increase. Thus F of the terms that were reexpressed in $\mathcal{B}$ will be of lower order than the term with all ovals pushed into t_{k_s} and the induction can be applied.

If $k_s < 0$ push all ovals into t_{k_1} to get again v to some power times t_{k_1} around all other ovals plus terms of lower order which, again, are reexpressed with terms in $\mathcal{B}$ of lower order.

Now repeat this process for ovals inside k_s or k_1, pushing ovals into the oval with maximum positive arrows (or maximum negative if there are no ovals with positive arrows). For terms with lower order reexpress them in $\mathcal{B}$: if some arrows were killed we will obviously get terms with lower order also in $\mathcal{B}$; if some arrows were transferred (by construction from ovals with less arrows to ovals with more arrows) we will get terms with lower lexicographical order. It is clear that the order cannot increase back to the order of T, when pushing the ovals which have lost some arrows out to get elements in $\mathcal{B}$.

Continue until all ovals are nested. We get at the end v to some power times $t_{k_l,k_{l-1},\cdots,k_1,k_{l+1},\cdots,k_s} \in \mathcal{B}''$ plus terms of lower order.

Now extend F linearly from $R\mathcal{B}$ to $R\mathcal{B}''$. The matrix of F with respect to the ordered $\mathcal{B}$ and $\mathcal{B}''$ will be upper triangular with invertible elements (powers of v) on the diagonal. This shows that $\mathcal{B}''$ is a basis of $\mathcal{S}_3(T)$.

Recall that the basis $\mathcal{B}_p$ of $\mathcal{S}_3(L(p, 1))$ consists of diagrams with non-nested ovals and the number of arrows k_i on each of them satisfying $-\frac{p}{2} < k_i \leq \frac{p}{2}$.

We want to exhibit a new basis of this skein module, $\mathcal{B}''_p$, using a proof similar to that of Theorem 6.12. Let:

$$\mathcal{B}''_p = \{t_{k_1,\ldots,k_s} \mid k_i \in \mathbb{Z} \setminus \{0\},\ s \in \mathbb{N},$$

$$0 > k_1 \geq \ldots \geq k_l > -\frac{p}{2},\ 0 < k_{l+1} \leq \ldots \leq k_s \leq \frac{p}{2}\}.$$

Thus $\mathcal{B}''_p$ is $\mathcal{B}''$ with arrows on ovals restricted to the interval $(-\frac{p}{2}, \frac{p}{2}]$. The order on $\mathcal{B}''$ used in the proof of the preceding theorem restricts to an order on $\mathcal{B}''_p$. We use it in the proof of the next theorem.

Theorem 6.13 *$\mathcal{B}''_p$ is a basis of the free skein module $\mathcal{S}_3(L(p, 1))$.*

Proof We construct a function $F : \mathcal{B}_p \to R\mathcal{B}''_p$ in the same way as it was done in the proof of Theorem 6.12, having a similar property, namely that F of an element

T in $\mathcal{B}_p$ will be equal to the element of $\mathcal{B}''_p$ consisting of the same ovals as in T but nested, times an invertible coefficient plus terms of lower order. When reexpressing an element in $\mathcal{B}_p$ it may happen that the number of arrows is reduced with $\Omega_{(p)}$ moves but this lowers the order.

Extending F linearly to $R\mathcal{B}_p$, its matrix is again upper triangular with invertible elements on the diagonal, from which it follows that $\mathcal{B}''_p$ is a basis of $\mathcal{S}_3(L(p, 1))$.

Acknowledgements The first author was supported by the Slovenian Research Agency grants J1-8131, J1-7025 and N1-0064.

References

1. Diamantis, I., Lambropoulou, S.: Braid equivalences in 3-manifolds with rational surgery description. Topol. Appl. **194**, 269–295 (2015)
2. Diamantis, I., Lambropoulou, S.: A new basis for the Homflypt skein module of the solid torus. J. Pure Appl. Algebr. **220**(2), 577–605 (2016)
3. Diamantis, I., Lambropoulou, S., Przytycki, J.: Topological steps toward the Homypt skein module of the lens spaces $L(p, 1)$ via braids. J. Knot Theory Ramif. **25**(14) (2016)
4. Gabrovšek, B., Mroczkowski, M.: The HOMFLYPT skein module of the lens spaces. Topol. Appl. **175**(9), 72–80 (2014)
5. Gabrovsek, B., Manfredi, E.: On the Seifert fibered space link group. Topol. Appl. **206**, 255–275 (2016)
6. Hoste, J., Kidwell, M.: Dichromatic link invariants. Trans. Amer. Math. Soc. **321**(1), 197–229 (1990)
7. Hoste, J., Przytycki, J.H.: The (2;1)-skein module of lens spaces; a generalization of the Jones polynomial. J. Knot Theory Ramif. **2**(3), 321–333 (1993)
8. Lambropoulou, S., Rourke, C.P.: Markov's theorem in 3-manifolds. Topol. Appl. **78**, 95–122 (1997)
9. Lambropoulou, S., Rourke, C.P.: Algebraic Markov equivalence for links in 3-manifolds. Compositio Math. **142**, 1039–1062 (2006)
10. Lickorish, W.B.R.: A representation of orientable combinatorial 3-manifolds. Ann. Math. **76**, 531–540 (1962)
11. Mroczkowski, M.: KBSM of the connected sum of two projective spaces. J. Knot Theory Ramif. **20**(5), 651–675 (2011)
12. Mroczkowski, M.: KBSM of a family of prism manifolds. J. Knot Theory Ramif. **20**(1), 159–170 (2011)
13. Mroczkowski, M., Dabkowski, M.: KBSM of the product of a disk with two holes and S^1. Topol. Appl. **156**, 1831–1489 (2009)
14. Orlik, P.: Seifert Manifolds. Lecture Notes in Mathematics, vol. 291. Springer, Berlin (1972)
15. Turaev, V.G.: The Conway and Kauffman modules of a solid torus. J. Sov. Math. **52**, 2799–2805 (1990)
16. Turaev, V.G.: Shadow links and face models of statistical mechanics. J. Differ. Geom. **36**(1), 35–74 (1992)
17. Wallace, A.H.: Modifications and cobounding manifolds. Can. J. Math. **12**, 503–528 (1960)

Chapter 7
The Braid Approach to the HOMFLYPT Skein Module of the Lens Spaces $L(p, 1)$

Ioannis Diamantis and Sofia Lambropoulou

Abstract In this paper we present recent results toward the computation of the HOMFLYPT skein module of the lens spaces $L(p, 1)$, $\mathcal{S}(L(p, 1))$, via braids. Our starting point is the knot theory of the solid torus ST and the Lambropoulou invariant, X, for knots and links in ST, the universal analogue of the HOMFLYPT polynomial in ST. The relation between $\mathcal{S}(L(p, 1))$ and $\mathcal{S}(\mathrm{ST})$ is established in Diamantis et al. (J Knot Theory Ramif, 25:13, 2016, [5]) and it is shown that in order to compute $\mathcal{S}(L(p, 1))$, it suffices to solve an infinite system of equations obtained by performing all possible braid band moves on elements in the basis of $\mathcal{S}(\mathrm{ST})$, Λ, presented in Diamantis and Lambropoulou (J Pure Appl Algebra, 220(2):577–605, 2016, [4]). The solution of this infinite system of equations is very technical and is the subject of a sequel work (Diamantis and Lambropoulou, The HOMFLYPT skein module of the lens spaces $L(p, 1)$ via braids, in preparation, [2]).

7.1 Introduction

Skein modules were introduced by Przytycki [20] and Turaev [23]. They generalize knot polynomials in S^3 to knot polynomials in arbitrary 3-manifolds. The essence is that skein modules are quotients of free modules over ambient isotopy classes of links in 3-manifolds by properly chosen local (skein) relations.

Let M be an oriented 3-manifold, $R = \mathbb{Z}[u^{\pm 1}, z^{\pm 1}]$, $\mathcal{L}$ the set of all oriented links in M up to ambient isotopy in M and let S be the submodule of $R\mathcal{L}$ generated by the

I. Diamantis
International College Beijing, China Agricultural University,
No.17 Qinghua East Road, Haidian District, Beijing
100083, People's Republic of China
e-mail: ioannis.diamantis@hotmail.com

S. Lambropoulou (✉)
Departament of Applied Mathematics, National Technical University of Athens,
Zografou Campus, 15780 Athens, Greece
e-mail: sofia@math.ntua.gr

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_7

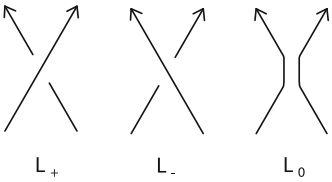


Fig. 7.1 The links L_+, L_-, L_0 locally

Fig. 7.2 An element in the basis of $\mathcal{S}(\mathbb{R}P^2 \hat{\times} I)$

skein expressions $u^{-1}L_+ - uL_- - zL_0$, where L_+, L_- and L_0 comprise a Conway triple represented schematically by the illustrations in Fig. 7.1.

For convenience we allow the empty knot, $\emptyset$, and add the relation $u^{-1}\emptyset - u\emptyset = zT_1$, where T_1 denotes the trivial knot. Then the *HOMFLYPT skein module* of M is defined to be:

$$\mathcal{S}(M) = \mathcal{S}\left(M; \mathbb{Z}\left[u^{\pm 1}, z^{\pm 1}\right], u^{-1}L_+ - uL_- - zL_0\right) = {}^{R\mathcal{L}}/_{S}.$$

Skein modules of 3-manifolds have become very important algebraic tools in the study of 3-manifolds, since their properties renders topological information about the 3-manifolds. Unlike the Kauffman bracket skein module, the HOMFLYPT skein module of a 3-manifold, also known as *Conway skein module* and as *third skein module*, is very hard to compute and very little is known so far. More precisely, $\mathcal{S}(S^3) = \mathbb{Z}[v^{\pm 1}, z^{\pm 1}]$, where the empty link is a generator of the module [6, 22]. Also, for the solid torus ST, $\mathcal{S}(\mathrm{ST})$ is a free, infinitely generated $\mathbb{Z}[u^{\pm 1}, z^{\pm 1}]$-module isomorphic to the symmetric tensor algebra $SR\widehat{\pi}^0$, where $\widehat{\pi}^0$ denotes the conjugacy classes of non trivial elements of $\pi_1(\mathrm{ST})$ [4, 11, 14, 23]. Further, let F denote a surface. Then $\mathcal{S}(F \times I)$ is an algebra which, as an R module, is a free module isomorphic to the symmetric tensor algebra, $SR\pi^o$, where π^o denotes the conjugacy classes of nontrivial elements of $\pi_1(F)$ [21]. Moreover, $\mathcal{S}(\mathbb{R}P^2 \hat{\times} I)$ is freely generated by standard oriented unlinks as presented in Fig. 7.2 [19] and $\mathcal{S}(S^1 \times S^2)$ is freely generated by the empty link over a properly chosen ring [8]. Finally, $\mathcal{S}(M_1 \# M_2)$ is isomorphic to $\mathcal{S}(M_1) \otimes \mathcal{S}(M_2)$ modulo torsion, where M_1, M_2 are oriented 3-manifolds and $M_1 \# M_2$ their connected sum [9].

In [7] the HOMFLYPT skein module of the lens spaces $L(p, 1)$ is computed using diagrammatic method. The diagrammatic method could in theory be generalized to the case of $L(p, q), q > 1$, but the diagrams become even more complex to analyze and several induction arguments fail.

In [14] the most generic analogue of the HOMFLYPT polynomial, X, for links in the solid torus ST has been derived from the generalized Hecke algebras of type B, $\mathrm{H}_{1,n}$, via a unique Markov trace constructed on them. This algebra was defined by Lambropoulou in [14] and is related to the knot theory of the solid torus, the Artin group of Coxeter group of type B, $B_{1,n}$, and to the affine Hecke algebra of type A. The Lambropoulou invariant X recovers the HOMFLYPT skein module of ST, $\mathcal{S}(\mathrm{ST})$, and is appropriate for extending the results to the lens spaces $L(p,q)$, since the combinatorial setting is the same as for ST, only the braid equivalence includes the braid band moves (shorthanded to *bbm*), which reflect the surgery description of $L(p,q)$. For the case of $L(p,1)$, in order to extend X to an invariant of links in $L(p,1)$ we need to solve an infinite system of equations resulting from the braid band moves. Namely we force:

$$X_{\widehat{\alpha}} = X_{\widehat{bbm(\alpha)}}, \tag{7.1}$$

for all $\alpha \in \bigcup_{\infty} B_{1,n}$ and for all possible slidings of α.

The above equations have particularly simple formulations with the use of a new basis, Λ, for the HOMFLYPT skein module of ST, that we give in [1, 4]. This basis was predicted by Przytycki and is crucial in this paper, since *bbm*'s are naturally described by elements in this basis.

In order to show that the set Λ is a basis for $\mathcal{S}(\mathrm{ST})$, we started in [4] with the well-known basis of $\mathcal{S}(\mathrm{ST})$, Λ', discovered independently in [11, 23] with diagrammatic methods, and a basis Σ_n of the algebra $\mathrm{H}_{1,n}$ and we followed the steps below:

- An ordering relation in Λ' is defined and it is shown that the set is totally ordered.
- Elements in Λ' are converted to linear combinations of elements in the new set Λ as follows:
- Elements in Λ' are first converted to elements in the linear basis of $\mathrm{H}_{1,n}(q)$, Σ_n.
- Using conjugation, the gaps appearing in the indices of the looping generators in the monomials in $\bigcup_n \Sigma_n$ are managed.
- Using conjugation, the exponents of the looping generators are ordered.
- Using conjugation and stabilization moves, the 'braiding tails' are removed from the above monomials and thus, the initial elements in $\bigcup_n \Sigma_n$ are converted to linear combination of elements in Λ.
- Finally, the sets Λ' and Λ are related via a block diagonal matrix, where each block is an infinite lower triangular matrix.
- The diagonal elements in the above matrix are invertible, making the matrix invertible and thus, the set Λ is a basis for $\mathcal{S}(\mathrm{ST})$.

The new basis is appropriate for computing the HOMFLYPT skein module of the lens spaces $L(p,q)$ in general. Note that $\mathcal{S}(\mathrm{ST})$ plays an important role in the study of HOMFLYPT skein modules of arbitrary c.c.o. 3-manifolds, since every c.c.o. 3-manifold can be obtained by surgery along a framed link in S^3 with unknotted components. The family of the lens spaces, $L(p,q)$, comprises the simplest example, since they are obtained by rational surgery on the unknot.

Equations (7.1) are very controlled in the algebraic setting, because, as shown in [5], they can be performed only on elements in Λ. This is shown by following the technique developed in [4]. The difference lies in the fact that here we deal with elements in Λ and at the same time with their result after the performance of a *bbm* and we keep track of how *bbm*'s affect the steps described above. More precisely, in [5] we followed the steps below:

- Equations (7.1) boil down by linearity to considering only words in the canonical basis Σ'_n of the algebra $\mathrm{H}_{1,n}(q)$.
- For words in $\bigcup_n \Sigma'_n$ equations of the form $X_{\widehat{\alpha'}} = X_{\widehat{bbm_{\pm 1}(\alpha')}}$ are obtained, where $\widehat{bbm_{\pm 1}(\alpha')}$ is the result of the performance of a braid band move on the *first* moving strand of the closed braid $\widehat{\alpha'}$, and $\alpha' \in \bigcup_n \Sigma'_n$.
- Then, elements in $\bigcup_n \Sigma'_n$ are expressed to elements in the linear basis Σ_n of $\mathrm{H}_{1,n}(q)$ and it is shown that the equations for words in $\bigcup_n \Sigma'_n$ are equivalent to equations of the form $X_{\widehat{\alpha}} = X_{\widehat{bbm_{\pm 1}(\alpha)}}$, where $\alpha \in \bigcup_n \Sigma_n$.
- A set Λ^{aug} is then introduced, consisting of monomials in the looping generators t_i's with no gaps in the indices (as in Λ) but not necessarily ordered exponents.
- Equations for words in $\bigcup_n \Sigma_n$ are now reduced to equations obtained from elements in the $\mathrm{H}_{1,n}(q)$-module Λ^{aug}, where the braid band moves are performed on *any moving strand*.
- Equations of the form $X_{\widehat{\beta}} = X_{\widehat{bbm_{\pm i}(\beta)}}$ are now obtained, where $\widehat{bbm_{\pm i}(\beta)}$ is the result of the performance of a braid band move on the i^{th} moving strand of the closed braid $\widehat{\beta}$, and β an element in the augmented set Λ^{aug} followed by a 'braiding tail'.
- Using conjugation, the exponents then become in decreasing order and equations obtained from elements in the $\mathrm{H}_{1,n}(q)$-module Λ^{aug} by performing *bbm*'s on all moving strands are reduced to equations for words in $\mathrm{H}_{1,n}(q)$-module Λ by performing *bbm*'s on all moving strands.
- The 'braiding tails' from elements in the $\mathrm{H}_{1,n}(q)$-module Λ are now eliminated and it is shown that equations for words in the $\mathrm{H}_{1,n}(q)$-module Λ by performing braid band moves on any strand, are now reduced to equations obtained from elements in the basis Λ of $\mathcal{S}(\mathrm{ST})$ by performing braid band moves on every moving strand:

$$\mathcal{S}\left(L(p,1)\right) \;=\; \frac{\mathcal{S}(\mathrm{ST})}{< a - bbm_i(a) >}, \text{ for all } i \text{ and for all } a \in \Lambda. \qquad (7.2)$$

In [2] we elaborate on the infinite system.

The importance of our approach is that it can shed light on the problem of computing skein modules of arbitrary c.c.o. 3-manifolds, since any 3-manifold can be obtained by surgery on S^3 along unknotted closed curves. Indeed, one can use our results in order to apply a braid approach to the skein module of an arbitrary c.c.o. 3-manifold. The main difficulty of the problem lies in selecting from the infinitum of band moves (or handle slide moves) some basic ones, solving the infinite system of equations and proving that there are no dependencies in the solutions. Note that the

computation of $\mathcal{S}(L(p,1))$ is equivalent to constructing all possible analogues of the HOMFLYPT or 2-variable Jones polynomial for knots and links in $L(p,1)$, since the linear dimension of $\mathcal{S}(L(p,1))$ means the number of independent HOMFLYPT-type invariants defined on knots and links in $L(p,1)$.

The paper is organized as follows: In Sect. 7.2 we recall the setting and the essential techniques and results from [3, 15–18]. More precisely, we describe braid equivalence for knots and links in $L(p,1)$ and we present a sharpened version of the Reidemeister theorem for links in $L(p,1)$. We also provide geometric formulations of the braid equivalence via mixed braids in S^3 using the L-moves and the braid band moves and give algebraic formulations in terms of the mixed braid groups $B_{1,n}$. In Sect. 7.3 we present results from [4, 14]. More precisely, we recover the HOMFLYPT skein module of the solid torus ST, $\mathcal{S}(\mathrm{ST})$, via algebraic techniques and we present a new basis for $\mathcal{S}(\mathrm{ST})$, Λ, from which the braid band moves are naturally described. The aim of this section is to set a homogeneous ground in computing skein modules of c.c.o. 3-manifolds in general via algebraic means. In Sect. 7.4 we derive the relation between $\mathcal{S}(\mathrm{ST})$ and $\mathcal{S}(L(p,1))$ and show that in order to compute $\mathcal{S}(L(p,1))$ we only need to consider elements in the basis Λ of $\mathcal{S}(\mathrm{ST})$ and impose on the Lambropoulou invariant X relations coming by performing all possible braid band moves on elements in Λ.

7.2 Topological and Algebraic Tools

7.2.1 Mixed Links and Isotopy in $L(p,1)$

We consider ST to be the complement of a solid torus in S^3. Then, an oriented link L in ST can be represented by an oriented *mixed link* in S^3, that is, a link in S^3 consisting of the unknotted *fixed part* $\widehat{I}$ representing the complementary solid torus in S^3 and the *moving part* L that links with $\widehat{I}$ (Fig. 7.3). A *mixed link diagram* is a diagram $\widehat{I} \cup \widetilde{L}$ of $\widehat{I} \cup L$ on the plane of $\widehat{I}$, where this plane is equipped with the top-to-bottom direction of I (see [13, 17]).

It is well–known that the lens spaces $L(p,1)$ can be obtained from S^3 by surgery on the unknot with surgery coefficient p. Surgery along the unknot can be realized

Fig. 7.3 A mixed link in S^3

by considering first the complementary solid torus and then attaching to it a solid torus according to some homeomorphism on the boundary. Thus, isotopy in $L(p, 1)$ can be viewed as isotopy in ST together with the *band moves* in S^3, which reflect the surgery description of the manifold, see Fig. 7.4 (see [17]). In [3] we show that in order to describe isotopy for knots and links in a c.c.o. 3-manifold, it suffices to consider only the type α band moves (see Fig. 7.5) and thus, isotopy between oriented links in $L(p, 1)$ is reflected in S^3 by means of the following result (cf. [17, Theorem 7.8], [3, Theorem 6]):

Two oriented links in $L(p, 1)$ *are isotopic if and only if two corresponding mixed link diagrams of theirs differ by isotopy in* ST *together with a finite sequence of the type* α band moves.

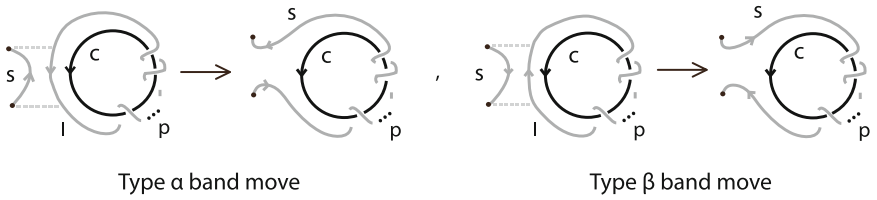


Fig. 7.4 The two types of band moves

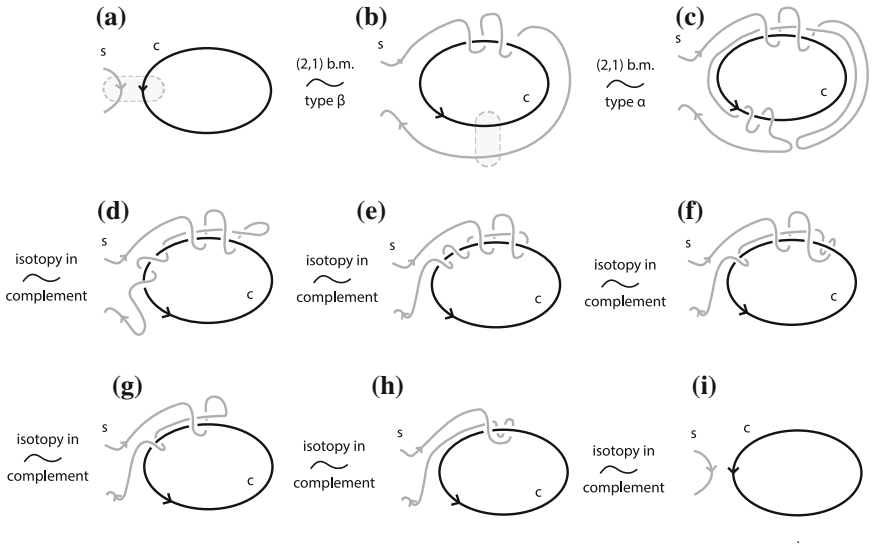


Fig. 7.5 A type-β band move follows from a type-α band move in the case of integral surgery coefficient

7.2.2 Mixed Braids for Knots and Links in $L(p, 1)$

By the Alexander theorem for knots in solid torus ([13, Theorem 1]), a mixed link diagram $\widehat{I} \cup \widetilde{L}$ of $\widehat{I} \cup L$ may be turned into a *mixed braid* $I \cup \beta$ with isotopic closure. This is a braid in S^3 where, without loss of generality, its first strand represents $\widehat{I}$, the fixed part, and the other strands, β, represent the moving part L. The subbraid β is called the *moving part* of $I \cup \beta$ (see Fig. 7.6).

The sets of mixed braids related to ST form groups, which are in fact the Artin braid groups of type B. The mixed braid group on n moving strands is denoted as $B_{1,n}$ and it admits the following presentation:

$$B_{1,n} = \left\langle t, \sigma_1, \ldots, \sigma_{n-1} \;\middle|\; \begin{array}{l} \sigma_1 t \sigma_1 t = t \sigma_1 t \sigma_1 \\ t\sigma_i = \sigma_i t, \quad i > 1 \\ \sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}, \quad 1 \leq i \leq n-2 \\ \sigma_i \sigma_j = \sigma_j \sigma_i, \quad |i-j| > 1 \end{array} \right\rangle, \tag{7.3}$$

where the 'braiding' generators σ_i and the 'looping' generator t are illustrated in Fig. 7.7 [13–15].

In $B_{1,n}$ we also define the elements:

$$t_i' := \sigma_i \sigma_{i-1} \ldots \sigma_1 t \sigma_1^{-1} \ldots \sigma_{i-1}^{-1} \sigma_i^{-1} \text{ and } t_i := \sigma_i \sigma_{i-1} \ldots \sigma_1 t \sigma_1 \ldots \sigma_{i-1} \sigma_i, \tag{7.4}$$

for $i = 0, 1, \ldots, n-1$, where $t_0 = t = t_0'$. For illustrations see Fig. 7.8.

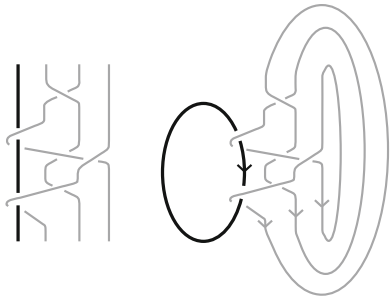


Fig. 7.6 The closure of a mixed braid to a mixed link

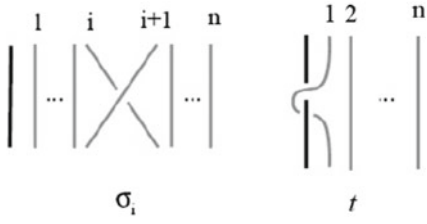


Fig. 7.7 The generators of $B_{1,n}$

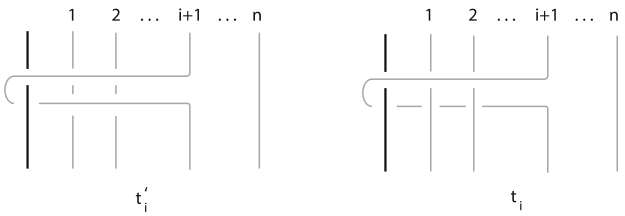


Fig. 7.8 The elements t_i' and t_i

Remark 7.1 In [14, Proposition 1] it is shown using the Artin combing, that every element in $B_{1,n}$ can be written in the form $\tau' \cdot w$, where τ' is a word in the looping elements t_i' and $w \in B_n$ is the 'braiding tail'. Furthermore, the looping elements t_i' are conjugates. On the other hand, the elements t_i commute. Moreover, the t_i's are consistent with the braid band move used in the link isotopy in $L(p, 1)$, in the sense that a braid band move can be described naturally with the use of the t_i's.

7.2.3 Braid Equivalence for Knots and Links in $L(p, 1)$

In [17] the authors give a sharpened version of the classical Markov theorem for braid equivalence in S^3 based only on one type of moves, the L-moves: An *L-move* on a mixed braid $I \bigcup \beta$, consists in cutting an arc of the moving sub-braid open and pulling the upper cut point downward and the lower upward, so as to create a new pair of braid strands with corresponding endpoints, and such that both strands cross entirely *over* or *under* with the rest of the braid. Stretching the new strands over will give rise to an L_o*-move* and under to an L_u*-move*. For an illustration see Fig. 7.9. An algebraic L-move has the following algebraic expression for an L_o-move and an L_u-move respectively:

$$\begin{aligned} \alpha = \alpha_1\alpha_2 &\overset{L_o}{\sim} \sigma_i^{-1} \dots \sigma_n^{-1}\alpha_1'\sigma_{i-1}^{-1} \dots \sigma_{n-1}^{-1}\sigma_n^{\pm 1}\sigma_{n-1} \dots \sigma_i\alpha_2'\sigma_n \dots \sigma_i \\ \alpha = \alpha_1\alpha_2 &\overset{L_u}{\sim} \sigma_i \dots \sigma_n\alpha_1'\sigma_{i-1} \dots \sigma_{n-1}\sigma_n^{\pm 1}\sigma_{n-1}^{-1} \dots \sigma_i^{-1}\alpha_2'\sigma_n^{-1} \dots \sigma_i^{-1} \end{aligned} \tag{7.5}$$

where α_1, α_2 are elements of $B_{1,n}$ and α_1', $\alpha_2' \in B_{1,n+1}$ are obtained from α_1, α_2 by replacing each σ_j by σ_{j+1} for $j = i, \dots, n-1$.

In order to translate isotopy for links in $L(p, 1)$ into mixed braid equivalence, we first define a *braid band move* to be a move between mixed braids, which is a type α band move between their closures. It starts with a little band oriented downward, which, before sliding along a surgery strand, gets one twist *positive* or *negative* (see Fig. 7.10).

Then, isotopy in $L(p, 1)$ is translated on the level of mixed braids by means of the following theorem:

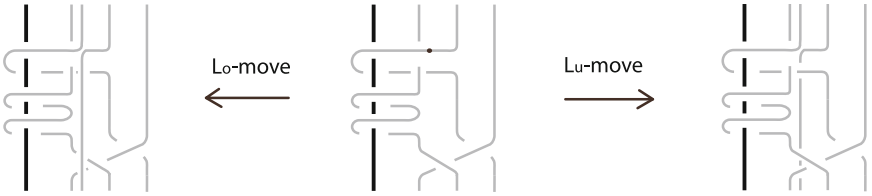


Fig. 7.9 A mixed braid and the two types of L-moves

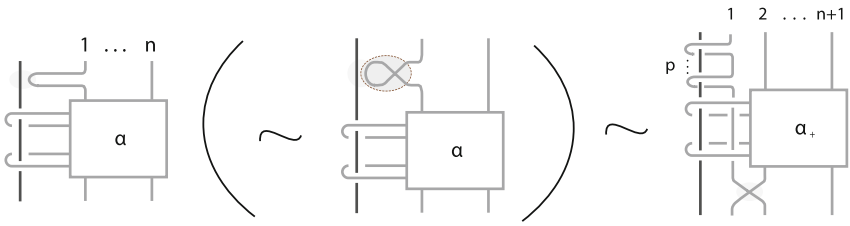


Fig. 7.10 Braid band move performed on the first moving strand

Theorem 7.1 ([18, Theorem 5]) *Let L_1, L_2 be two oriented links in $L(p,1)$ and let $I \cup \beta_1$, $I \cup \beta_2$ be two corresponding mixed braids in S^3. Then L_1 is isotopic to L_2 in $L(p,1)$ if and only if $I \cup \beta_1$ is equivalent to $I \cup \beta_2$ in $\bigcup_{n=1}^{\infty} B_{1,n}$ by the following moves:*

$$\begin{array}{lll}
(i)\ \textit{Conjugation}: & \alpha \sim \beta^{-1}\alpha\beta, & \alpha, \beta \in B_{1,n} \\
(ii)\ \textit{Stabilization moves}: & \alpha \sim \alpha\sigma_n^{\pm 1} \in B_{1,n+1}, & \alpha \in B_{1,n} \\
(iii)\ \textit{Loop conjugation}: & \alpha \sim t^{\pm 1}\alpha t^{\mp 1}, & \alpha \in B_{1,n} \\
(iv)\ \textit{Braid band moves}: & \alpha \sim t^p \alpha_+ \cdot \sigma_1^{\pm 1}, & \alpha_+ \in B_{1,n+1}
\end{array}$$

where α_+ is the word α with all indices shifted by 1 (see Fig. 7.10). Equivalently, L_1 is isotopic to L_2 if and only if $I \cup \beta_1$ and $I \cup \beta_2$ differ by moves (iii) and (iv) as above, while moves (i) and (ii) are replaced by the two types of L-moves, see Eq. (7.5).

Notation 7.1 We denote a braid band move by *bbm* and, specifically, the result of a positive or negative braid band move performed on the ith-moving strand of a mixed braid β by $bbm_{\pm i}(\beta)$.

Note that in the statement of Theorem 7.1 in [18] the braid band moves take place on the last strand of a mixed braid. Clearly, this is equivalent to performing the braid band moves on the first moving strand or, in fact, on any specified moving strand of the mixed braid. Indeed, in a mixed braid $\beta \in B_{1,n}$ consider the last moving strand of β approaching the surgery strand I from the right. Before performing a *bbm* we apply conjugation (isotopy in ST) and we obtain an equivalent mixed braid where the first strand is now approaching I (see Fig. 7.11). In terms of diagrams we have the following ([5, Lemma 1]):

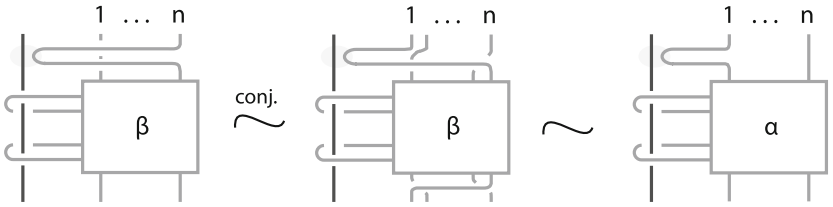


Fig. 7.11 A *bbm* may always be assumed to be performed on the first moving strand of a mixed braid

$$\begin{array}{lll} \beta & \sim (\sigma_{i-1}\ldots\sigma_1\sigma_1^{-1}\ldots\sigma_{i-1}^{-1})\cdot\beta \sim \underbrace{(\sigma_1^{-1}\ldots\sigma_{i-1}^{-1})\cdot\beta\cdot(\sigma_{i-1}\ldots\sigma_1)}_{\alpha} \\ \downarrow & & \downarrow \\ bbm_{\pm n}(\beta) & = & bbm_{\pm 1}(\alpha) \end{array}$$

7.2.4 The Generalized Iwahori–Hecke Algebra of Type B

It is well-known that $B_{1,n}$ is the Artin group of the Coxeter group of type B, which is related to the Hecke algebra of type B, $\mathrm{H}_n(q, Q)$ and to the cyclotomic Hecke algebras of type B. In [14] it has been established that all these algebras are related to the knot theory of ST. The basic one is $\mathrm{H}_n(q, Q)$, a presentation of which is obtained from the presentation (7.3) of the Artin group $B_{1,n}$ by corresponding the braiding generator σ_i to g_i and by adding the quadratic relations:

$$g_i^2 = (q-1)g_i + q \tag{7.6}$$

and the relation $t^2 = (Q-1)\,t + Q$. The cyclotomic Hecke algebras of type B are denoted $\mathrm{H}_n(q, d)$, $d \in \mathbb{N}$, and they admit presentations that are obtained by the quadratic relation (7.6) and the relation $t^d = (t-u_1)(t-u_2)\ldots(t-u_d)$. In [14] also the *generalized Iwahori–Hecke algebra of type B*, $\mathrm{H}_{1,n}(q)$, is introduced, as the quotient of $\mathbb{C}\left[q^{\pm 1}\right] B_{1,n}$ over the quadratic relations (7.6). Namely:

$$\mathrm{H}_{1,n}(q) = \frac{\mathbb{C}\left[q^{\pm 1}\right] B_{1,n}}{\langle \sigma_i^2 - (q-1)\,\sigma_i - q\rangle}.$$

Note that in $\mathrm{H}_{1,n}(q)$ the generator t satisfies no polynomial relation, making the algebra $\mathrm{H}_{1,n}(q)$ infinite dimensional, and as observed by T. tom Dieck, $\mathrm{H}_{1,n}$ is closely related to the affine Hecke algebra of type A, $\widetilde{\mathrm{H}}_n(q)$. In [14] the algebra $\mathrm{H}_{1,n}(q)$ is denoted as $\mathrm{H}_n(q, \infty)$.

In [12] V.F.R. Jones gives the following linear basis for the Iwahori–Hecke algebra of type A, $\mathrm{H}_n(q)$:

$$S = \left\{(g_{i_1} g_{i_1-1} \dots g_{i_1-k_1})(g_{i_2} g_{i_2-1} \dots g_{i_2-k_2}) \dots (g_{i_p} g_{i_p-1} \dots g_{i_p-k_p})\right\},$$

for $1 \leq i_1 < \dots < i_p \leq n-1$.

The basis S yields directly an inductive basis for $\mathrm{H}_n(q)$, which is used in the construction of the Ocneanu trace, leading to the HOMFLYPT or 2-variable Jones polynomial.

We also introduce in $\mathrm{H}_{1,n}(q)$ the 'looping elements'

$$t_0' = t_0 := t, \quad t_i' = g_i \dots g_1 t g_1^{-1} \dots g_i^{-1} \quad \text{and} \quad t_i = g_i \dots g_1 t g_1 \dots g_i \tag{7.7}$$

that correspond to the elements of Eq. 7.4 of $B_{1,n}$.

From [14] we have:

Theorem 7.2 ([14, Proposition 1 and Theorem 1]) *The following sets form linear bases for* $\mathrm{H}_{1,n}(q)$*:*

$$\begin{aligned} &(i) \quad \Sigma_n = \{t_{i_1}^{k_1} \dots t_{i_r}^{k_r} \cdot \sigma\}, \text{ where } 0 \leq i_1 < \dots < i_r \leq n-1, \\ &(ii) \quad \Sigma_n' = \{t_{i_1}'^{\,k_1} \dots t_{i_r}'^{\,k_r} \cdot \sigma\}, \text{ where } 0 \leq i_1 < \dots < i_r \leq n-1, \end{aligned}$$

where $k_1, \dots, k_r \in \mathbb{Z}$ *and* σ *a basic element in* $\mathrm{H}_n(q)$.

For an illustration of the t_i's and the t_i''s recall Fig. 7.8.

Remark 7.2 Note that the form of the elements in the set Σ_n' is consistent with the Artin combing in the mixed braid group $B_{1,n}$ (recall Remark 7.1). However the looping part of a mixed braid after the combing contains elements of a free group on n generators, so the indices as well as the exponents have no restrictions. In Theorem 7.2 though, the indices of the t_i's and the t_i''s in the above sets are ordered but are not necessarily consecutive, neither do they need to start from t. Also, the exponents are arbitrary.

Notation 7.2 We shall denote

$$\Sigma := \bigcup_n \Sigma_n \quad \text{and similarly} \quad \Sigma' := \bigcup_n \Sigma_n'. \tag{7.8}$$

7.3 The HOMFLYPT Skein Module of the Solid Torus

In [11, 23] the HOMFLYPT skein module of the solid torus ST has been computed using diagrammatic methods by means of the following theorem:

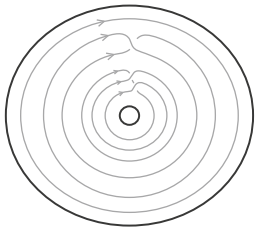


Fig. 7.12 A basic element of $\mathcal{S}(\mathrm{ST})$

Theorem 7.3 (Turaev, Kidwell–Hoste) *The skein module* $\mathcal{S}(\mathrm{ST})$ *is a free, infinitely generated* $\mathbb{Z}[u^{\pm 1}, z^{\pm 1}]$*-module isomorphic to the symmetric tensor algebra* $SR\widehat{\pi}^0$*, where* $\widehat{\pi}^0$ *denotes the conjugacy classes of non trivial elements of* $\pi_1(\mathrm{ST})$.

In the diagrammatic setting of [11, 23], ST is considered as Annulus $\times$ Interval. A basic element of $\mathcal{S}(\mathrm{ST})$ in the context of [11, 23], is illustrated in Fig. 7.12. The HOMFLYPT skein module of ST is particularly important, because any closed, connected, oriented (c.c.o.) 3-manifold can be obtained by surgery along a framed link in S^3 with unknotted components. Such component can be viewed as a solid torus complementing ST in S^3.

7.3.1 *Recovering* $\mathcal{S}$ (ST) *Using Algebraic Means*

In [14] the bases Σ'_n are used for constructing a Markov trace on $\bigcup_{n=1}^{\infty} \mathrm{H}_{1,n}(q)$.

Theorem 7.4 ([14, Theorem 6]) *Given* z, s_k*, with* $k \in \mathbb{Z}$ *specified elements in* $R = \mathbb{Z}\left[q^{\pm 1}\right]$*, there exists a unique linear Markov trace function*

$$\mathrm{tr} : \bigcup_{n=1}^{\infty} \mathrm{H}_{1,n}(q) \to R\,(z, s_k)\,, k \in \mathbb{Z}$$

determined by the rules:

$$\begin{array}{lll} (1)\ \mathrm{tr}(ab) & = \mathrm{tr}(ba) & \text{for } a, b \in \mathrm{H}_{1,n}(q) \\ (2)\ \mathrm{tr}(1) & = 1 & \text{for all } \mathrm{H}_{1,n}(q) \\ (3)\ \mathrm{tr}(ag_n) & = z\mathrm{tr}(a) & \text{for } a \in \mathrm{H}_{1,n}(q) \\ (4)\ \mathrm{tr}(at_n'^{k}) & = s_k \mathrm{tr}(a) & \text{for } a \in \mathrm{H}_{1,n}(q),\ k \in \mathbb{Z}. \end{array}$$

Note that the use of the looping elements t'_i enable the trace tr to be defined by just extending by rule (4) the three rules of the Ocneanu trace on the algebras $\mathrm{H}_n(q)$ [12], recall Remark 7.1. Using tr the second author constructed a universal HOMFLYPT-type invariant for oriented links in ST. Namely, let $\mathcal{L}$ denote the set of oriented links in ST. Then:

Theorem 7.5 ([14, Definition 1]) *The function* $X : \mathcal{L} \to R(z, s_k)$

$$X_{\widehat{\alpha}} = \Delta^{n-1} \cdot \left(\sqrt{\lambda}\right)^{e} \mathrm{tr}\left(\pi\left(\alpha\right)\right),$$

where $\Delta := -\frac{1-\lambda q}{\sqrt{\lambda}(1-q)}$, $\lambda := \frac{z+1-q}{qz}$, $\alpha \in B_{1,n}$ *is a word in the* σ_i*'s and* t_i'*'s,* $\widehat{\alpha}$ *is the closure of* α*, e is the exponent sum of the* σ_i*'s in* α*, and* π *the canonical map of* $B_{1,n}$ *on* $\mathrm{H}_{1,n}(q)$*, such that* $t \mapsto t$ *and* $\sigma_i \mapsto g_i$*, is an invariant of oriented links in* ST.

In the braid setting of [14], the elements of $\mathcal{S}(\mathrm{ST})$ correspond bijectively to the elements of the following set Λ':

$$\Lambda' = \{t^{k_0} t_1'^{k_1} \dots t_n'^{k_n},\ k_i \in \mathbb{Z} \setminus \{0\},\ k_i \geq k_{i+1}\ \forall i,\ n \in \mathbb{N}\}. \tag{7.9}$$

So, we have that Λ' is a basis of $\mathcal{S}(\mathrm{ST})$ in terms of braids. Note that Λ' is a subset of $\bigcup_n \mathrm{H}_{1,n}$ and, in particular, Λ' is a subset of Σ'. Note also that in contrast to elements in Σ', the elements in Λ' have no gaps in the indices, the exponents are ordered and there are no 'braiding tails'.

Remark 7.3 The Lambropoulou invariant X recovers $\mathcal{S}(\mathrm{ST})$, because it gives distinct values to distinct elements of Λ', since $tr(t^{k_0} t_1'^{k_1} \dots t_n'^{k_n}) = s_{k_n} \dots s_{k_1} s_{k_0}$.

Note also that the invariant X is defined by the skein relation:

$$\frac{1}{\sqrt{q}\sqrt{\lambda}} X_{L_+} - \sqrt{q}\sqrt{\lambda} X_{L_-} = \left(\sqrt{q} - \frac{1}{\sqrt{q}}\right) X_{L_o}$$

and an infinitum of initial conditions, one for each element in $\mathcal{S}(\mathrm{ST})$ as shown in Fig. 7.12. Namely, if U denotes the unknot, then $X_U = 1$ and $X_{t^{k_0}\widehat{t_1'^{k_1} \dots t_n'^{k_n}}} = \Delta^{n-1} \cdot s_{k_n} \dots s_{k_1} s_{k_0}$.

7.3.2 The New Basis, Λ, of $\mathcal{S}$(ST)

In [4] we give a different basis Λ for $\mathcal{S}(\mathrm{ST})$, which was predicted by J. Przytycki and which is described in Eq. 7.10 in open braid form (Figs. 7.13 and 7.14).

Theorem 7.6 ([4, Theorem 2]) *The following set is a* $\mathbb{Z}[q^{\pm 1}, z^{\pm 1}]$*-basis for* $\mathcal{S}(\mathrm{ST})$*:*

$$\Lambda = \{t^{k_0} t_1^{k_1} \dots t_n^{k_n},\ k_i \in \mathbb{Z} \setminus \{0\},\ k_i \geq k_{i+1}\ \forall i,\ n \in \mathbb{N}\}. \tag{7.10}$$

The importance of the new basis Λ of $\mathcal{S}(\mathrm{ST})$ lies in the simplicity of the algebraic expression of a braid band move, which extends the link isotopy in ST to link isotopy in $L(p, 1)$ and this fact was our motivation for establishing this new basis Λ (recall Theorem 7.1(iv), Remark 7.1 and Fig. 7.10).

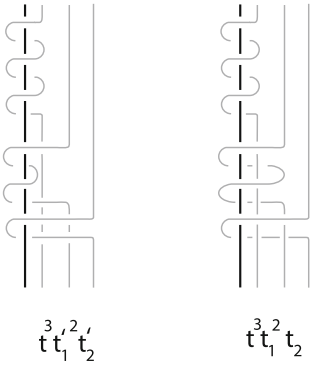


Fig. 7.13 Elements in two different bases of $\mathcal{S}(\mathrm{ST})$

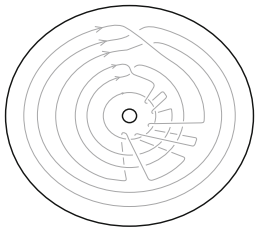


Fig. 7.14 An element of the new basis Λ

Notice that comparing the set Λ with the set Σ, we observe that there are no gaps in the indices of the t_i's and the exponents are in decreasing order. Also, there are no 'braiding tails' in the words in Λ.

Our method for proving Theorem 7.6 is the following:

- We define total orderings in the sets Λ' and Λ,
- we show that the two ordered sets are related via a lower triangular infinite matrix with invertible elements on the diagonal, and
- using this matrix, we show that the set Λ is linearly independent.

In order to relate the two sets via a lower triangular infinite matrix we start with elements in the basic set Λ' and we first convert them into sums of elements in Σ, containing the linear bases of the algebras $\mathrm{H}_{1,n}(q)$. These elements consist of two parts: arbitrary monomials in the t_i's followed by 'braiding tails' in the bases of the algebras $\mathrm{H}_n(q)$. Then, these elements are converted into elements in the set Λ by:

- managing the gaps in the indices,
- by ordering the exponents of the t_i's and
- by eliminating the 'braiding tails'.

It is worth mentioning that these procedures are not independent in the sense that when, for example one manages the gaps in the indices of the looping generators t_i's, 'braiding tails' may occur and also the exponents of the t_i's may alter. Similarly, when the 'braiding tails' are eliminated, gaps in the indices of the t_i's might occur. This is a long procedure that eventually stops and only elements in the set Λ remain.

7.3.3 An Ordering for the Sets Σ', Σ, Λ' and Λ

For defining orderings in the sets Σ, Σ', Λ and Λ' we need the notion of the *index* of a word w in any of these sets, denoted $ind(w)$. In Λ' or Λ $ind(w)$ is defined to be the highest index of the t'_i's, resp. of the t_i's in w. Similarly, in Σ' or Σ, $ind(w)$ is defined as above by ignoring possible gaps in the indices of the looping generators and by ignoring the braiding parts in the algebras $\mathrm{H}_n(q)$. Moreover, the index of a monomial in $\mathrm{H}_n(q)$ is equal to 0.

Example 7.1

$$\begin{array}{lll} i. & ind(t'^{k_0}{t'_1}^{k_1}\dots{t'_n}^{k_n}) = n = ind(t^{k_0}{t_1}^{k_1}\dots{t_n}^{k_n}) & \\ ii. & ind(t^{k_0}t_2^{k_2}\cdot\sigma) = 2 & , \ \sigma \in \mathrm{H}_n(q) \\ iii. & ind(\sigma) = 0 & , \ \sigma \in \mathrm{H}_n(q) \end{array}$$

We now proceed with defining an ordering relation in the sets Σ and Σ', which passes to their respective subsets Λ and Λ':

Definition 7.1 ([4, Definition 2]) Let $w = {t'_{i_1}}^{k_1}\dots{t'_{i_\mu}}^{k_\mu}\cdot\beta_1$ and $u = {t'_{j_1}}^{\lambda_1}\dots{t'_{j_\nu}}^{\lambda_\nu}\cdot\beta_2$ in Σ', where $k_t, \lambda_s \in \mathbb{Z}$ for all t, s and $\beta_1, \beta_2 \in H_n(q)$. Then, we define the following ordering in Σ':

(a) If $\sum_{i=0}^{\mu} k_i < \sum_{i=0}^{\nu}\lambda_i$, then $w < u$.
(b) If $\sum_{i=0}^{\mu} k_i = \sum_{i=0}^{\nu}\lambda_i$, then:
(i) if $ind(w) < ind(u)$, then $w < u$,
(ii) if $ind(w) = ind(u)$, then:
(α) if $i_1 = j_1, \dots, i_{s-1} = j_{s-1}, i_s < j_s$, then $w > u$,
(β) if $i_t = j_t$ for all t and $k_\mu = \lambda_\mu, k_{\mu-1} = \lambda_{\mu-1}, \dots, k_{i+1} = \lambda_{i+1}, |k_i| < |\lambda_i|$, then $w < u$,
(γ) if $i_t = j_t$ for all t and $k_\mu = \lambda_\mu, k_{\mu-1} = \lambda_{\mu-1}, \dots, k_{i+1} = \lambda_{i+1}, |k_i| = |\lambda_i|$ and $k_i > \lambda_i$, then $w < u$,
(δ) if $i_t = j_t\ \forall t$ and $k_i = \lambda_i, \forall i$, then $w = u$.

The ordering in the set Σ is defined as in Σ', where t'_i's are replaced by t_i's.

Example 7.2

$$\begin{array}{lllll} \text{i. } t^2{t'_1}^8 & < t^6{t'_1}^{10} & \text{since } 2+8 < 6+10 & , & \text{(Def. 7.1a)} \\ \text{ii. } t^2{t'_1}^8 & < t^3{t'_1}^4{t'_2}^3 & \text{since } ind(t^2{t'_1}^8) < ind(t^3{t'_1}^4{t'_2}^3) & , & \text{(Def. 7.1b(i))} \\ \text{iii. } t^2{t'_1}^3{t'_3}^4{t'_4}^8 & < t^8t'_1t'_2{t'_4}^7 & \text{since } 3 = i_3 > j_3 = 2 & , & \text{(Def. 7.1b(ii)}(\alpha)\text{)} \\ \text{iv. } t^2{t'_1}^4t'_3{t'_4}^3 & < t^{14}{t'_1}^{-8}t'_3{t'_4}^3 & \text{since } |4| < |-8| & , & \text{(Def. 7.1b(ii)}(\beta)\text{)} \\ \text{v. } t{t'_1}^4t'_3{t'_4}^3 & < t^{10}{t'_1}^{-4}t'_3{t'_4}^3 & \text{since } -4 < 4 & , & \text{(Def. 7.1b(ii)}(\gamma)\text{)} \end{array}$$

In order to eventually get to the infinite matrix relating the two basic sets Λ' and Λ we need to define the *subsets of level k*, $\Lambda_{(k)}$ and $\Lambda'_{(k)}$, of Λ and Λ' respectively [4, Definition 3], to be the sets:

$$\begin{aligned} \Lambda_{(k)} &:= \{t_0^{k_0} t_1^{k_1} \dots t_m^{k_m} \,|\, \textstyle\sum_{i=0}^{m} k_i = k,\ k_i \in \mathbb{Z} \setminus \{0\},\ k_i \geq k_{i+1}\ \forall i\} \\ \Lambda'_{(k)} &:= \{t_0'^{k_0} t_1'^{k_1} \dots t_m'^{k_m} \,|\, \textstyle\sum_{i=0}^{m} k_i = k,\ k_i \in \mathbb{Z} \setminus \{0\},\ k_i \geq k_{i+1}\ \forall i\} \end{aligned} \tag{7.11}$$

In [4] it was shown that the sets $\Lambda_{(k)}$ and $\Lambda'_{(k)}$ are totally ordered and well ordered for all k ([4, Propositions 1 and 2]). Note that the sets Λ and Λ' admit a natural grading: $\Lambda = \bigoplus_k \Lambda_{(k)}$ and $\Lambda' = \bigoplus_k \Lambda'_{(k)}$.

In the rest of the section we will be using the ordering in the transitions from Σ' to Σ and from Σ to Λ.

7.3.4 From Λ' to Σ

In this subsection we convert monomials in the t_i''s to expressions containing the t_i's. Full details and related technical lemmas are provided in [4]. In order to simplify the expressions in this step we first introduce the following notation.

Notation 7.3 We set $\tau_{i,i+m}^{k_{i,i+m}} := t_i^{k_i} \dots t_{i+m}^{k_{i+m}}$ and $\tau'^{k_{i,i+m}}_{i,i+m} := t_i'^{k_i} \dots t_{i+m}'^{k_{i+m}}$, for $m \in \mathbb{N}$, $k_j \neq 0$ for all j.

We now introduce the notion of *homologous words*, which is crucial for relating the sets Λ' and Λ via a triangular matrix.

Definition 7.2 ([4, Definition 4]) We say that two words $w' \in \Lambda'$ and $w \in \Lambda$ are *homologous*, denoted $w' \sim w$, if w is obtained from w' by changing t_i' into t_i for all i.

Example 7.3 The words $t^2 {t_1'}^{-1} {t_2'}^3$ and $t^2 {t_1}^{-1} {t_2}^3$ are homologous. Note also that $\Lambda' \ni t \sim t \in \Lambda$.

In [4, Lemma 11] it is shown that the following relations hold in $\mathrm{H}_{1,n}(q)$ for $k \in \mathbb{Z}\setminus\{0\}$:

$$t_m'^{\,k} = q^{-mk} t_m^k + \textstyle\sum_i f_i(q) t_m^k w_i + \sum_i g_i(q) t^{\lambda_0} t_1^{\lambda_1} \dots t_m^{\lambda_m} u_i$$

where $w_i, u_i \in \mathrm{H}_{m+1}(q)$, $\forall i$, $\sum_{i=0}^{m} \lambda_i = k$ and $\lambda_i \geq 0$, $\forall i$, if $k > 0$ and $\lambda_i \leq 0$, $\forall i$, if $k < 0$.

Using now these relations, we have that every element in Λ' can be expressed as a linear sum of each homologous word, the homologous word with 'braiding tails' and elements in Σ of lower order, as illustrated abstractly in Fig. 7.15. More precisely:

Theorem 7.7 ([4, Theorem 7]) *The following relations hold in* $\mathrm{H}_{1,n}(q)$ *for* $k_r \in \mathbb{Z}$, $r = 0, \dots, m$*:*

$$\tau'^{k_{0,m}}_{0,m} = q^{-\sum_{n=1}^{m} n k_n} \cdot \tau_{0,m}^{k_{0,m}} + \sum_i f_i(q) \cdot \tau_{0,m}^{k_{0,m}} \cdot w_i + \sum_j g_j(q) \cdot \tau_j \cdot u_j,$$

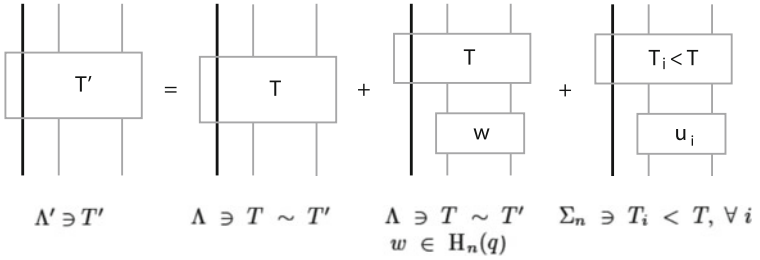


Fig. 7.15 Illustrating Theorem 7.7

where $w_i, u_j \in \mathrm{H}_{m+1}(q)$*, for all* i, j*,* τ_j *a monomial of the* t_i*'s such that* $\tau_j < \tau_{0,m}^{k_{0,m}}$ *for all* j *and* $f_i, g_j \in \mathbb{C}$*, for all* i, j*.*

Example 7.4 We shall now give examples of monomials in Λ' converted into sums of elements in Σ using technical lemmas from [4] ([4, Lemmas 3, 4, 5, 6, 9 and 11]). The examples provide at the same time the motivation for the subsections that follow.

i. Consider the monomial $tt_1't_2'^{-2} \in \Lambda'$. We have that:

$$\begin{aligned} tt_1't_2'^{-2} &= q^3 \cdot tt_1t_2^{-2} + q^4(q^{-1}-1) \cdot tt_1t_2^{-2} \cdot g_1^{-1} + \\ &+ 1 \cdot \left[(q-1)^2 g_1^{-1}g_2^{-1} - (q-1)^3(g_1^{-2} - q^{-1})g_2^{-1} - q^{-1}(q-1)^3 g_2 g_1^{-1} g_2^{-1}\right] + \\ &+ tt_1^{-1} \cdot \left[(q-1)(q^2-q+1) \cdot g_2^{-1} - (q-1)^2 \cdot g_1 g_2 g_1^{-1} g_2^{-1}\right] + \\ &+ tt_2^{-1} \cdot \left[q^2(q-1) \cdot g_2^{-1} + q(q-1)^3 \cdot g_2^{-1} - q(q-1)^2 \cdot g_2 g_1^{-1} g_2^{-1}\right] + \\ &+ t_1t_2^{-1} \cdot \left[q(q-1) \cdot g_2 g_1^{-1} g_2^{-1} - q(q-1)^2 \cdot g_1^{-1} g_2^{-1}\right] + \\ &+ t^{-1}t_1 \cdot \left[-(q-1) \cdot g_2 g_1^{-1} g_2^{-1} - q^{-1}(q-1)^2 \cdot g_1^{-1} g_2^{-1}\right] \end{aligned}$$

So, we obtain the homologous word $w = tt_1t_2^{-2}$, the word w again followed by the braiding element g_1^{-1} and terms in Σ of order less than w: since either their index is less that $ind(w)$ (the terms 1, tt_1^{-1} and $t^{-1}t_1$), or they contain gaps in the indices (the terms tt_2^{-1} and $t_1t_2^{-1}$).

ii. A simpler example is the following:

$$tt_1'^{-2} = tt_1{}^{-2} - (q^{-1}-1)\underline{t_1{}^{-1}} \cdot g_1^{-1} + (q^{-1}-1)t^{-1} \cdot \underline{g_1}$$

We obtained the homologous word $tt_1{}^{-2}$, the element $t^{-1} \cdot g_1$ in Σ comprising the monomial t^{-1} with gaps in the indices, followed by 'braiding tail' g_1^{-1} and also the lower order term $t^{-1}g_1$ in the $\mathrm{H}_n(q)$-module Λ.

In Theorem 7.7 note that in the right hand side there are terms which do not belong to the set Λ. The point now is that these terms are elements in the set of bases Σ on the Hecke algebras $\mathrm{H}_{1,n}(q)$, but, when we are working in $\mathcal{S}(\mathrm{ST})$, which is the knots and links level, such elements must be considered up to conjugation by any generator of the algebra and up to stabilization moves (recall Theorem 7.1). Topologically, conjugation corresponds to closing a mixed braid.

7.3.5 *From Σ to Λ: Managing the Gaps*

In this subsection we show how to deal with monomials in Σ where the looping elements do not have consecutive indices. We call *gaps* in monomials of the t_i's any gaps occurring in the indices. After managing, that is eliminating, the gaps we pass to the augmented $\mathrm{H}_n(q)$-module Λ^{aug}, which consists of monomials in the t_i's with consecutive indices but not necessarily ordered exponents.

Definition 7.3 ([5, Definition 3]) We define the sets:

$$\Lambda_n^{aug} := \{t_0^{k_0} t_1^{k_1} \dots t_n^{k_n},\ k_i \in \mathbb{Z}^*\}, \quad \Lambda^{aug} := \bigcup_{n\in\mathbb{N}} \Lambda_n^{aug},$$

and the *subset of level k* of Λ^{aug} $\Lambda_{(k)}^{aug}$:

$$\Lambda_{(k)}^{aug} := \left\{ t_0^{k_0} t_1^{k_1} \dots t_m^{k_m} \,|\, \sum_{i=0}^{m} k_i = k,\ k_i \in \mathbb{Z}^* \right\}.$$

Note that in [5] the set Λ^{aug} is denoted by L. Obviously the set Λ (Eq. 7.10) is a subset of Λ^{aug}.

Notation 7.4 Whenever we talk about a module with coefficients in $\bigcup_{n\in\mathbb{N}} \mathrm{H}_n(q)$ we shall be denoting it by $\mathrm{H}_n(q)$-module.

In what follows for the expressions that we obtain after appropriate conjugations we shall use the notation $\widehat{=}$.

Theorem 7.8 ([4, Theorem 8]) *Let τ be a monomial in the t_i's with gaps in the indices. Then we have that:*

$$\tau \,\widehat{=}\, \sum_i \tau_i \cdot \beta_i,$$

where $\tau_i \in \Lambda^{aug}$ and $\beta_i \in \bigcup_{n\in\mathbb{N}} \mathrm{H}_n(q)$ for some $n \in \mathbb{N}$ and for all i.

Theorem 7.8 is best demonstrated in the following example on a word with two gaps. Note that we underline expressions which are crucial for the next step.

Example 7.5 For the 2-gap word $t^{k_0}t_1^{k_1}t_3t_5^2t_6^{-1} \in \Sigma$ we have:

$$
\begin{aligned}
t^{k_0}\underline{t_1^{k_1}t_3t_5^2t_6^{-1}} &= t^{k_0}t_1^{k_1}\underline{g_3t_2g_3}t_5^2t_6^{-1} && = g_3t^{k_0}t_1^{k_1}t_2t_5^2t_6^{-1}g_3 && \widehat{=}\\
&\widehat{=} t^{k_0}t_1^{k_1}t_2\underline{t_5^2}t_6^{-1}g_3^2 && = t^{k_0}t_1^{k_1}t_2\underline{t_5}t_5t_6^{-1}g_3^2 && =\\
&= t^{k_0}t_1^{k_1}t_2\underline{g_5g_4}t_3g_4g_5t_5t_6^{-1}g_3^2 && = \underline{g_5g_4}t^{k_0}t_1^{k_1}t_2t_3g_4g_5t_5t_6^{-1}g_3^2 && \widehat{=}\\
& && \widehat{=} t^{k_0}t_1^{k_1}t_2t_3\underline{g_4g_5t_5}t_6^{-1}g_3^2g_5g_4 && =\\
&= t^{k_0}t_1^{k_1}t_2t_3\big[\,q^2t_3g_4g_5 && + q(q-1)t_4g_5 + (q-1)t_5g_4\,\big]t_6^{-1}g_3^2g_5g_4 && =\\
&= q^2t^{k_0}t_1^{k_1}t_2t_3^2\underline{g_4g_5t_6^{-1}}g_3^2g_5g_4 && + q(q-1)t^{k_0}t_1^{k_1}t_2t_3t_4\underline{g_5t_6^{-1}}g_3^2g_5g_4 && +\\
& && + (q-1)t^{k_0}t_1^{k_1}t_2t_3t_5\underline{g_4t_6^{-1}}g_3^2g_5g_4 && =\\
&= q^2t^{k_0}t_1^{k_1}t_2t_3^2\underline{t_6^{-1}}g_4g_5g_3^2g_5g_4 && + (q-1)t^{k_0}t_1^{k_1}t_2t_3\underline{t_5}t_6^{-1}g_4g_3^2g_5g_4 && +\\
& && + q(q-1)t^{k_0}t_1^{k_1}t_2t_3t_4\underline{t_6^{-1}}g_5g_3^2g_5g_4 && \widehat{=}
\end{aligned}
$$

$$
\begin{aligned}
\widehat{=} q^2t^{k_0}t_1^{k_1}t_2t_3^2\underline{g_6^{-1}g_5^{-1}}t_4^{-1}g_5^{-1}g_6^{-1}g_4g_5g_3^2g_5g_4 &+ q(q-1)t^{k_0}t_1^{k_1}t_2t_3t_4\underline{g_6^{-1}}t_5^{-1}g_6^{-1}g_5g_3^2g_5g_4 &+\\
&+ (q-1)t^{k_0}t_1^{k_1}t_2t_3\underline{g_5}t_4\underline{g_5}t_6^{-1}\cdot(g_4g_3^2g_5g_4) &=\\
= q^2g_6^{-1}g_5^{-1}t^{k_0}t_1^{k_1}t_2t_3^2t_4^{-1}g_5^{-1}g_6^{-1}g_4g_5g_3^2g_5g_4 &+ q(q-1)g_6^{-1}t^{k_0}t_1^{k_1}t_2t_3t_4t_5^{-1}g_6^{-1}g_5g_3^2g_5g_4 &+\\
&+ (q-1)g_5t^{k_0}t_1^{k_1}t_2t_3t_4t_6^{-1}g_5g_4g_3^2g_5g_4 &\widehat{=}\\
\widehat{=} q^2t^{k_0}t_1^{k_1}t_2t_3^2t_4^{-1}g_5^{-1}g_6^{-1}g_4g_5g_3^2g_5g_4g_6^{-1}g_5^{-1} &+ q(q-1)t^{k_0}t_1^{k_1}t_2t_3t_4t_5^{-1}g_6^{-1}g_5g_3^2g_5g_4g_6^{-1} &+\\
&+ (q-1)t^{k_0}t_1^{k_1}t_2t_3t_4\underline{t_6^{-1}}g_5\cdot(g_4g_3^2g_5g_4g_5) &=\\
= q^2t^{k_0}t_1^{k_1}t_2t_3^2t_4^{-1}g_5^{-1}g_6^{-1}g_4g_5g_3^2g_5g_4g_6^{-1}g_5^{-1} &+ q(q-1)t^{k_0}t_1^{k_1}t_2t_3t_4t_5^{-1}g_6^{-1}g_5g_3^2g_5g_4g_6^{-1} &+\\
&+ (q-1)t^{k_0}t_1^{k_1}t_2t_3t_4\underline{g_6^{-1}}t_5^{-1}g_6^{-1}g_5g_4g_3^2g_5g_4g_5 &\widehat{=}\\
\widehat{=} q^2t^{k_0}t_1^{k_1}t_2t_3^2t_4^{-1}g_5^{-1}g_6^{-1}g_4g_5g_3^2g_5g_4g_6^{-1}g_5^{-1} &+ q(q-1)t^{k_0}t_1^{k_1}t_2t_3t_4t_5^{-1}g_6^{-1}g_5g_3^2g_5g_4g_6^{-1} &+\\
&+ (q-1)t^{k_0}t_1^{k_1}t_2t_3t_4t_5^{-1}g_6^{-1}g_5g_4g_3^2g_5g_4g_5g_6^{-1}
\end{aligned}
$$

7.3.6 From the $\mathbf{H}_n(q)$-Module Σ to the $\mathbf{H}_n(q)$-Module Λ: Ordering the Exponents

By Theorem 7.8 we have to deal with elements in Λ^{aug}, where the looping generators have consecutive indices but their exponents are not in decreasing order, followed by 'braiding tails'. We show that these elements are conjugate equivalent to sums of elements in the $\mathrm{H}_n(q)$-module Λ, namely, elements in Λ followed by 'braiding tails'.

Theorem 7.9 ([4, Theorem 9]) *For an element in the* $\mathrm{H}_n(q)$*-module* Λ^{aug} *we have that:*

$$\tau_{0,m}^{k_{0,m}} \cdot w \widehat{=} \sum_j \tau_{0,j}^{\lambda_{0,j}} \cdot w_j,$$

where $\tau_{0,j}^{\lambda_{0,j}} \in \Lambda$ *and* $w, w_j \in \bigcup_{n\in\mathbb{N}} \mathrm{H}_n(q)$ *for all* j.

Example 7.6 Consider the element $tt_1^2t_2^3 \in \Lambda^{aug}$ and apply Theorem 7.9 on the first 'bad' exponent occurring in the word, starting from right to left. In that way we obtain a word with one less 'bad' exponent, so applying Theorem 7.9 again we obtain elements in the set Λ followed by braiding tails. More precisely:

$$tt_1^2t_2^3 \widehat{=} a \cdot t^3t_1^2t_2 \cdot u_1 \; + \; b \cdot t^2t_1^2t_2^2 \cdot u_2 \; + \; c \cdot t^4t_1t_2 \cdot u_3$$

where $u_1, u_2, u_3 \in \mathrm{H}_3(q)$, for all i and $a, b, c \in \mathbb{C}[q^{\pm 1}]$.

7.3.7 From the $\mathbf{H}_n(q)$-Module Λ to Λ: Eliminating the Tails

So far we have seen how to convert elements in the basis Λ' to sums of elements in Σ and then, using conjugation, how these elements are expressed as sums of elements in the $\mathrm{H}_n(q)$-module Λ. We now present results on how using conjugation and stabilization moves all these elements in the $\mathrm{H}_n(q)$-module Λ are expressed as sums of elements in the set Λ with scalars in the field $\mathbb{C}$. We will use the symbol $\simeq$ when a stabilization move is performed and $\widehat{\simeq}$ when both stabilization moves and conjugation are used. More precisely, in [4] we prove the following:

Theorem 7.10 ([4, Theorem 10]) *For a word in the* $\mathrm{H}_n(q)$*-module,* Λ *we have:*

$$\tau_{0,m}^{k_{0,m}} \cdot w_n \widehat{\simeq} \sum_j f_j(q,z) \cdot \tau_{0,u_j}^{v_{0,u_j}},$$

such that $\sum v_{0,u_j} = \sum k_{0,m}$, $\tau_{0,u_j}^{v_{0,u_j}} \in \Lambda^{aug}$ *and* $\tau_{0,u_j}^{v_{0,u_j}} < \tau_{0,m}^{k_{0,m}}$ *for all* j.

Example 7.7 In this example we demonstrate how to eliminate the 'braiding tail' in a word.

$$\begin{aligned} t^3\underline{t_1^2}t_2^{-1}g_1^{-1} &= t^3t_1t_2^{-1}\underline{t_1g_1^{-1}} \;=\; t^3t_1t_2^{-1}g_1\underline{t} \,\widehat{=}\, t^4\underline{t_1}t_2^{-1}g_1 \;=\; t^4t_2^{-1}\underline{t_1g_1} \;= \\ &= (q-1)t^4\underline{t_1}t_2^{-1} \;+\; qt^4t_2^{-1}g_1\underline{t} \,\widehat{=}\, (q-1)t^5\underline{t_2^{-1}}g_1^2 \;+\; qt^5\underline{t_2^{-1}}g_1 \;= \\ &= (q-1)t^5t_1^{-1}g_2^{-1}g_1^2g_2^{-1} \;+\; qt^5t_1^{-1}g_2^{-1}g_1g_2^{-1}. \end{aligned}$$

We have that:

$$\begin{aligned} g_2^{-1}g_1^2g_2^{-1} &= q^{-2}(q-1)g_1g_2g_1 \;-\; (q^{-1}-1)^2g_2g_1 \;-\; (q^{-1}-1)^2g_1g_2 \;+ \\ &\quad + (q-1)(q^{-1}-1)^2g_1 \;+\; q(q^{-1}-1)g_2^{-1} \;+\; 1, \\ g_2^{-1}g_1g_2^{-1} &= q^{-2}g_1g_2g_1 \;+\; q^{-1}(q^{-1}-1)g_2g_1 \;+\; q^{-1}(q^{-1}-1)g_1g_2 \;+ \\ &\quad + (q^{-1}-1)^2g_1, \end{aligned}$$

and so

$$\begin{aligned} (q-1)\cdot t^5t_1^{-1}g_2^{-1}g_1^2g_2^{-1} &\,\widehat{\approx}\, \left((q-1)+q^{-1}(q-1)^3\right)\cdot t^5t_1^{-1} - q^{-3}(q^{-1}-1)^3z^2\cdot t^4 + \\ &\quad + 3q^{-3}(q-1)^4z\cdot t^4 - q^{-1}(q-1)^2z\cdot t^4 - q^{-3}(q-1)^5\cdot t^4, \\ q\cdot t^5t_1^{-1}g_2^{-1}g_1g_2^{-1} &\,\widehat{\approx}\, z\cdot t^5t_1^{-1} + q^{-1}(q^{-1}-1)z^2\cdot t^4 + 2(q^{-1}-1)^2z\cdot t^4 + \\ &\quad + q(q^{-1}-1)^3\cdot t^4. \end{aligned}$$

Remark 7.4 This is a long procedure, since eliminating the tails will give rise to gaps in the indices again. Recall however that when managing gaps in the indices and ordering the exponents of the t_i's we also obtain 'braiding tails'. This is a long procedure which, as shown in [4] using complex and technical inductions, this procedure eventually terminates and only elements in Λ remain. This means that Λ is a generating set for $\mathcal{S}(\mathrm{ST})$. This procedure is abstractly demonstrated in Fig. 7.16. In the figure, we start with a word τ' in the old basis of $\mathcal{S}(\mathrm{ST})$, Λ', and applying Theorems 7.8–7.10 we end up with a linear combination of the homologous word $\tau \in \Lambda$ and of elements $\lambda_i \in \Lambda$ smaller than τ.

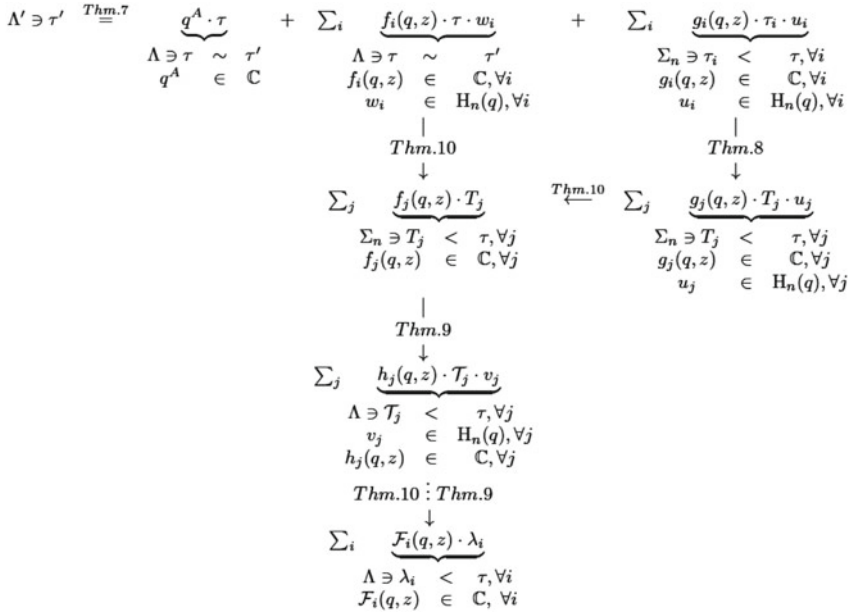


Fig. 7.16 From Λ' to Λ

7.3.8 The Infinite Matrix

With the orderings given in Definition 7.1, in [4] we showed that the infinite matrix converting elements of the basis Λ' of $\mathcal{S}(\mathrm{ST})$ to elements of the set Λ is a block diagonal matrix, where each block corresponds to a subset of Λ' of level k and it is an infinite lower triangular matrix with invertible elements in the diagonal. This constitutes our strategy for proving Theorem 7.6. More precisely, fixing the level k of a subset of Λ', the proof of Theorem 7.6 is based on the following:

(1) A monomial $w' \in \Lambda'_{(k)} \subseteq \Lambda'$ can be expressed as linear combinations of elements in $\Lambda_{(k)} \subseteq \Lambda$, v_i, followed by monomials in $\mathrm{H}_n(q)$, with scalars in $\mathbb{C}$ such that there exists $j : v_j = w \sim w'$.
(2) Applying conjugation and stabilization moves on all v_i's results in elements u_i in $\Lambda_{(k)}$, such that $u_i < v_i$ for all i.
(3) The coefficient of w is an invertible element in $\mathbb{C}$.
(4) $\Lambda_{(k)} \ni w < u \in \Lambda_{(k+1)}$.
(5) Using this infinite diagonal matrix, in [4, Theorem 11] we showed that the set Λ is linearly independent. Hence, using the above and Remark 7.4, Λ forms a basis for the HOMFLYPT skein module of ST.

7.4 Topological Steps Toward $\mathcal{S}(L(p,1))$

We now return to our initial aim, that is, the computation of the HOMFLYPT skein module of a lens space $L(p,1)$. As explained in the Introduction, in order to compute $\mathcal{S}(L(p,1))$ we need to normalize the invariant X (recall Theorem 7.5) by forcing it to satisfy all possible braid band moves (*bbm*), recall Eq. 7.1. At this point the reader should recall the discussion in the Introduction culminating to Eq. 7.2. In order to simplify this system of equations, in [5] we first show that performing a *bbm* on a mixed braid in $B_{1,n}$ reduces to performing *bbm*'s on elements in the canonical basis, Σ'_n, of the algebra $\mathrm{H}_{1,n}(q)$ and, in fact, on their first moving strand. We then reduce the equations obtained from elements in Σ' to equations obtained from elements in Σ. In order now to reduce further the computation to elements in the basis Λ of $\mathcal{S}(\mathrm{ST})$, we first recall that elements in Σ consist in two parts: a monomial in t_i's with possible gaps in the indices and unordered exponents, followed by a 'braiding tail' in the basis of $\mathrm{H}_n(q)$. So, we first manage the gaps in the indices of the looping generators of elements in Σ, obtaining elements in the augmented $\mathrm{H}_n(q)$-module Λ^{aug} (followed by 'braiding tails'). Note that the performance of *bbm*'s is now considered to take place on any moving strand. We then show that the equations obtained from elements in the $\mathrm{H}_n(q)$-module Λ^{aug} by performing *bbm*'s on any strand are equivalent to equations obtained from elements in the $\mathrm{H}_n(q)$-module Λ by performing *bbm*'s on any strand (ordering the exponents in the t_i's). We finally eliminate the 'braiding tails' from elements in the $\mathrm{H}_n(q)$-module Λ and reduce the computations to the set Λ, where the *bbm*'s are performed on any moving strand (see [5]). Thus, in order to compute $\mathcal{S}(L(p,1))$, it suffices to solve the infinite system of equations obtained by performing *bbm*'s on any moving strand of elements in the set Λ.

In this section we present the above steps in more details. The procedure is similar to the one described in [4] (see Sect. 7.2 in this paper), but now we do it simultaneously before and after the performance of a braid band move.

7.4.1 Reducing Computations from $B_{1,n}$ to Σ'_n

We now show that it suffices to perform *bbm*'s on elements in the linear basis of $\mathrm{H}_{1,n}(q)$, Σ'. As already mentioned, this is the first step in order to restrict the performance of *bbm*'s only on elements in the basis Λ. We first recall that by the Artin combing we can write words in $B_{1,n}$ in the form $\tau' \cdot w$, where τ' is a monomial in the t'_i's and $w \in B_n$ (Remark 7.1). We then note the following:

Lemma 7.1 *Braid band moves are interchangeable with the Artin combing.*

Proof Let $d \in B_{1,n}$. Then, the proof is clear from the following diagram:

$$\begin{array}{ccc} B_{1,n} \ni \ d & \xrightarrow{(\pm)(p,1)bbm} & t^p d' \cdot \sigma_1^{\pm 1} \\ \Big\downarrow \textit{Artin combing} & & \Big\downarrow \textit{Artin combing} \\ \tau_1' \cdot w & \xrightarrow{(\pm)(p,1)bbm} & t^p \tau_2' \cdot \sigma_1^{\pm 1} \end{array}$$

□

Lemma 7.2 ([5, Lemma 2]) *Braid band moves and the quadratic relation (skein relation) are interchangeable.*

Proof By Lemma 7.1, a word in $B_{1,n}$ can be assumed in the form $\tau_1' \cdot w$, where τ_1' is a monomial in t_i''s and $w \in B_n$. Seen as a monomial in $\mathrm{H}_n(q)$ and applying the quadratic relation, the element w can be written as a sum: $w = \sum_{i=1}^{n} f_i(q) w_i$, where the w_i's are words in $\mathrm{H}_n(q)$ in canonical form and the $f_i(q)$ are expressions in $\mathbb{C}$ for all i. We perform a braid band move on the element $\tau_1' \cdot w \in B_{1,n}$ and we obtain:

$$\tau_1' \cdot w \xrightarrow[bbm]{(\pm)(p,1)} t^p \tau_2' \cdot w_+ \sigma_1^{\pm 1},$$

where $w_+ \in B_{n+1}$ is the same word as w but with all indices shifted by one. Hence, on the algebra level we have $w_i \in \mathrm{H}_n(q)$ and $w_+ = \sum_{i=1}^{n} f_i(q) w_{i_+} \in \mathrm{H}_{n+1}(q)$. So:

$$\tau_1' \cdot w \ = \ \tau_1' \cdot \sum_{i=1}^{n} f_i(q) w_i \ \xrightarrow[bbm]{(\pm)(p,1)} \ t^p \tau_2' \cdot \sum_{i=1}^{n} f_i(q) w_{i_+} g_1^{\pm 1}.$$

On the other hand, for each mixed braid $\tau_1' \cdot w_i$ we also have:

$$\tau_1' \cdot w_i \xrightarrow[bbm]{(\pm)(p,1)} t^p \tau_2' \cdot w_{i_+} \sigma_1^{\pm 1} \ \forall \ i,$$

and thus, on the algebra level we finally obtain:

$$\tau_2' \cdot \sum_{i=1}^{n} f_i(q) w_i g_1^{\pm 1} \ = \ t^p \tau_2' \cdot w_+ g_1^{\pm 1}.$$

That is, the following diagram commutes:

$$\begin{array}{ccc} \tau_1' \cdot w & \xrightarrow{(\pm)(p,1)bbm} & t^p \tau_2' \cdot w_+ \sigma_1^{\pm 1} \\ \Big\downarrow \textit{quadratic} & & \Big\downarrow \textit{quadratic} \\ \sum_i f_i(q) \tau_1' \cdot w_i & \xrightarrow{(\pm)(p,1)bbm} & \sum_i f_i(q) t^p \tau_2' \cdot w_{i_+} g_1^{\pm 1} \end{array}$$

See also Fig. 7.17. So the proof is concluded.

□

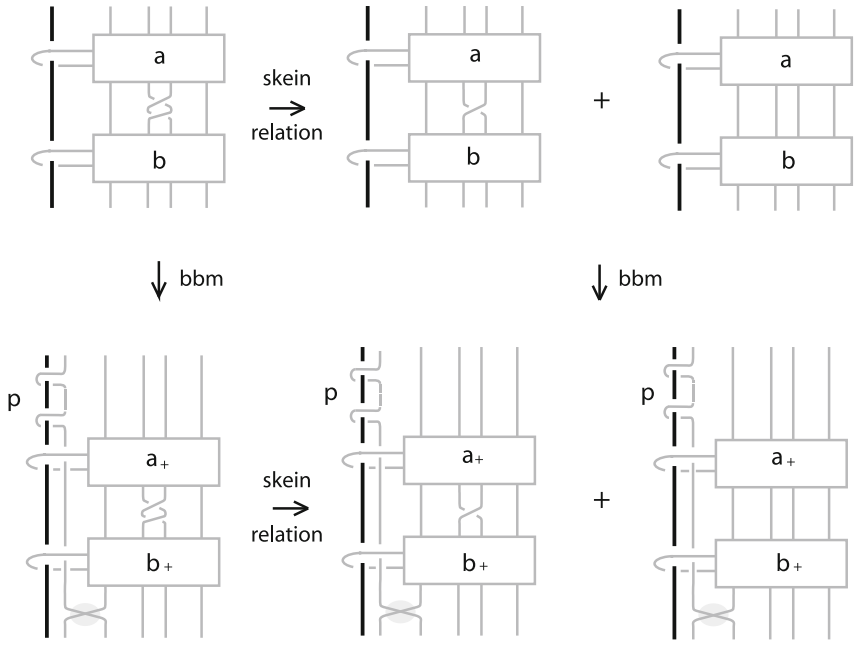


Fig. 7.17 Proof of Lemma 7.2

Furthermore we have:

Lemma 7.3 *The procedure of bringing a looping monomial in the t_i''s in the form of elements in the sets Σ_n' of Theorem 7.2 is consistent with the braid band moves.*

Proof Let $\tau_1' \cdot w$ be an element in $B_{1,n}$ and $t^p \tau_2' \cdot w_+ \cdot \sigma_1^{\pm 1}$ the result of a performance of a *bbm* on $\tau_1' \cdot w$. We now follow [14, Proposition 2 and Theorem 1] so as to order the indices of the monomials τ_1' and τ_2' in the t_i''s (before and after the performance of the *bbm*). Cabling the first moving strand coming from the performance of the *bbm* with the fixed strand of the mixed braid, and viewing this cable as one thickened strand, we have that the procedure we follow to order the indices of the t_i''s is identical before and after the performance of the *bbm* and this concludes the proof. □

Using Lemmas 7.1, 7.2 and 7.3 we now have the following:

Proposition 7.1 ([5, Proposition 1]) *It suffices to consider the performance of braid band moves on the first strand of only elements in the set Σ'.*

Proof By the Artin combing, any $d \in B_{1,n}$ can be written in the form $\tau' \cdot w$, where τ' is a monomial in t_i''s and $w \in B_n$. By Lemma 7.1 we have that:

$$X_{\widehat{\tau' \cdot w}} = X_{t^p \widehat{\tau'' \cdot \sigma_1^{\pm 1}} \cdot w_+} \quad \overset{(Lemma\ 7.2)}{\Rightarrow}$$

$$\sum_i A_i \cdot X_{\widehat{\tau' \cdot w_i}} = \sum_i A_i \cdot X_{t^p \widehat{\tau'' \cdot \sigma_1^{\pm 1}} \cdot w_{i_+}},$$

where w_i are words in reduced form in $\mathrm{H}_n(q)$, $\forall i$ and $A_i \in \mathbb{C}$. Then, we order the indices of the monomials in the t_i''s before and after the performance of the *bbm*. The result follows from Lemma 7.3. □

7.4.2 *From the Set Σ' to the Set Σ*

We shall now show that it suffices to perform *bbm*'s on elements in the linear bases Σ of the algebras $\mathrm{H}_{1,n}(q)$, which includes as a proper subset the basis Λ of $\mathcal{S}(\mathrm{ST})$, described in Sect. 7.3.2. Indeed, let $\tau' \cdot w \in \Sigma'$ as above. Then, by (7.7) we have:

$$\tau' \cdot w = (t^{k_0} t_1'^{k_1} \dots t_m'^{k_m}) \cdot w = \underbrace{t^{k_0} (t_1 g_1^{-2})^{k_1} \dots (t_m g_m^{-1} \dots g_2^{-1} g_1^{-2} g_2^{-1} \dots g_m^{-1})^{k_m}}_{\tau} \cdot w =$$

$$= \quad \tau \cdot w,$$

see top row of Fig. 7.18.

We then perform a *bbm* on the first moving strand of both $\tau' \cdot w$ and $\tau \cdot w$ (see bottom row of Fig. 7.18) and we cable the new parallel strand together with the surgery strand. Denote the result of cabling the new strand appearing after the performance of the *bbm* with the fixed strand as $cbl(ps)$. Then:

$$\begin{array}{ccc} \tau' \cdot w & \overset{bbm}{\rightarrow} & cbl(ps) \cdot \tau' \cdot w \cdot g_1^{\pm 1} \\ \| & & \| \\ \tau \cdot w & \overset{bbm}{\rightarrow} & cbl(ps) \cdot \tau \cdot w \cdot g_1^{\pm 1} \end{array}$$

So: $X_{\widehat{\tau' \cdot w}} = X_{\widehat{bbm(\tau' \cdot w)}} \Leftrightarrow X_{\widehat{\tau \cdot w}} = X_{\widehat{bbm(\tau \cdot w)}}$. But since $\tau \cdot w \in \mathrm{H}_{1,n}(q)$ for some $n \in \mathbb{N}$, we can express $\tau \cdot w$ as a sum of elements in the linear basis Σ_n of $\mathrm{H}_{1,n}(q)$, that is, $\tau \cdot w = \sum_i a_i T_i \cdot w_i$, where $a_i \in \mathbb{C}$ and $T_i \cdot w_i \in \Sigma_n$, for all i and T_i is a monomial in the t_j's with possible gaps in the indices and unordered exponents. Then, by Theorem 7.5:

$$\begin{aligned} X_{\widehat{\tau \cdot w}} = X_{\widehat{bbm(\tau \cdot w)}} \Leftrightarrow tr(\tau \cdot w) &= \Delta \cdot tr\left(cbl(ps)\tau \cdot w \cdot g_1^{\pm 1}\right) \\ \Leftrightarrow \sum_i a_i\, tr(T_i \cdot w_i) &= \Delta \cdot \sum_i a_i\, tr\left(\cdot cbl(ps) T_i \cdot w_i \cdot g_1^{\pm 1}\right). \end{aligned}$$

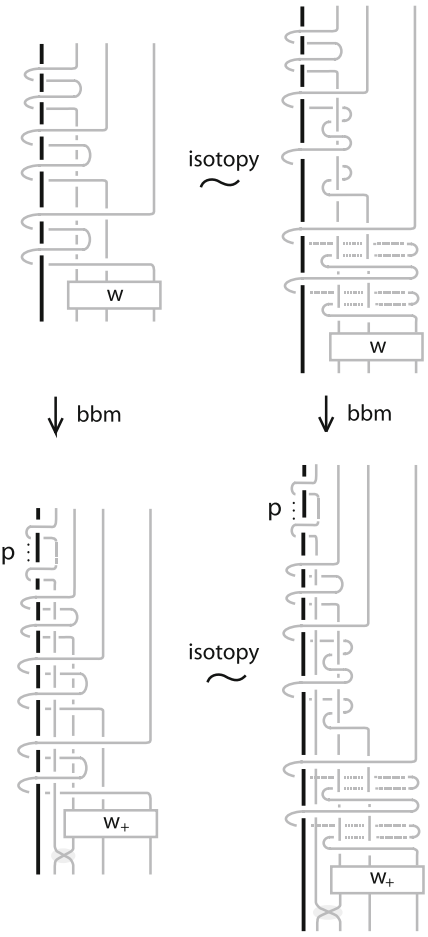


Fig. 7.18 Proof of Proposition 7.2

We conclude that:

$$\begin{array}{ccc} \tau' \cdot w & \overset{bbm}{\rightarrow} & cbl(ps) \cdot \tau' \cdot w \cdot g_1^{\pm 1} \quad (*) \\ \| & & \| \\ \tau \cdot w & \overset{bbm}{\rightarrow} & cbl(ps) \cdot \tau \cdot w \cdot g_1^{\pm 1} \\ \| & & \| \\ \sum_i a_i \cdot T_i \cdot w_i & \overset{bbm}{\rightarrow} & \sum_i a_i \cdot t^p T_{i+} \cdot w_{i+} g_1^{\pm 1} \quad (**) \end{array}$$

The main points in the above procedure is that when performing a *bbm*, the looping generators t_i (in some $\mathrm{H}_{1,n}$) remain formally the same (in some $\mathrm{H}_{1,n+1}$). This implies that after the *bbm* and the cabling operation, the words $\tau' \cdot w$ and $\tau \cdot w$ remain formally the same. Furthermore, when converting the words $\tau \cdot w$ and $bbm(\tau \cdot w)$ into linear sums of elements in the bases Σ_n and Σ_{n+1} respectively, the coefficients $a_i \in \mathbb{C}$ in $(**)$ remain the same.

The above are summarized in the following proposition:

Proposition 7.2 ([5, Proposition 2]) *The equations*

$$X_{\widehat{T' \cdot w}} = X_{\widehat{bbm_1(T' \cdot w)}} \tag{7.12}$$

result from equations of the form

$$X_{\widehat{T \cdot w}} = X_{\widehat{t^p T_+ \cdot w_+ \cdot g_1^{\pm 1}}}, \tag{7.13}$$

where $T' \cdot w \in \Sigma'$ *and* $T \cdot w \in \Sigma$.

7.4.3 From the Set Σ to the $\mathbf{H}_n(q)$-Module Λ^{aug}: Managing the Gaps

As mentioned in Sect. 7.3, a word in Σ is a monomial in the t_i's followed by a 'braiding tail', a monomial in the g_i's. This 'braiding tail' is a word in the algebra $\mathrm{H}_n(q)$ and the monomial in the t_i's may have gaps in the indices. Using the ordering relation given in Definition 7.1 and conjugation these gaps are managed by showing that a monomial in the t_i's can be expressed as a sum of monomials in the t_i's with consecutive indices, which are of less order than the initial word and which are followed by 'braiding tails' (Theorem 7.8). Note that the exponents of the t_i's are in general not ordered, so these end monomials do not necessarily belong to the basis Λ of $\mathcal{S}(\mathrm{ST})$. In order to restrict the *bbm*'s only on elements in Λ, we need first to augment the set Λ. So, as a first step we consider the augmented set Λ^{aug} in $\mathcal{S}(\mathrm{ST})$ that contains monomials in the t_i's with consecutive indices but arbitrary exponents. Note that, due to the presence of 'braiding tails', the set Λ^{aug} is considered as an $\mathrm{H}_n(q)$-module.

We now proceed with showing that Eq. (7.13) (Proposition 7.2) reduce to equations of the same type, but with elements in the set Λ^{aug}. For that, we need the following lemma about the monomial $t_1^k \in \Sigma \setminus \Lambda^{aug}$, which serves as the basis of the induction applied for proving the main result of this section, Proposition 7.3.

Lemma 7.4 ([5, Lemma 3]) *The equations* $X_{\widehat{t_1^k}} = X_{\widehat{t^p t_2^k \sigma_1^{\pm 1}}}$ *are equivalent to the equations*

$$\begin{aligned} X_{\widehat{t^{u_0} t_1^{u_1}}} &= X_{\widehat{t^p t_1^{u_0} t_2^{u_1} \sigma_1^{\pm 1}}}, \ \forall\, u_0, u_1 < k \ : \ u_0 + u_1 = k, \\ X_{\widehat{t^k}} &= X_{\widehat{t^p t_1^k \sigma_1^{\pm 1}}}, \quad \textit{bbm on 1st strand}, \ t^k \in \Sigma_2, \\ X_{\widehat{t^k}} &= X_{\widehat{t^p t_1^k \sigma_2 \sigma_1^{\pm 1} \sigma_2^{-1}}} \ \textit{bbm on 2nd strand}, \ t^k \in \Sigma_2. \end{aligned}$$

Indeed, in [5] we prove that:

$$\begin{array}{lcl} t_1^k & \xrightarrow[1^{st}\,str.]{bbm} & t^p t_2^k \sigma_1^{\pm 1} \\ \widehat{=} & & \widehat{=} \\ (q-1)^2 \sum_{j=0}^{k-2} \sum_{\phi=0}^{k-2-j} q^{j+\phi} t^{j+1+\phi} t_1^{k-1-j-\phi} & \xrightarrow[1^{st}\,str.]{bbm} & (q-1)^2 \sum_{j=0}^{k-2} \sum_{\phi=0}^{k-2-j} q^{j+\phi} t^p t_1^{j+1+\phi} t_2^{k-1-j-\phi} \sigma_1^{\pm 1} \\ (q-1)(k-1) q^{k-1} z t^k & \xrightarrow[1^{st}\,str.]{bbm} & (q-1)(k-1) q^{k-1} z t^p t_1^k \sigma_1^{\pm 1} \\ (q-1) q^{k-1} z t^k & \xrightarrow[1^{st}\,str.]{bbm} & (q-1) q^{k-1} z t^p t_1^k \sigma_1^{\pm 1} \\ q^k t^k & \xrightarrow[2^{nd}\,str.]{bbm} & q^k z t^p t_1^k \sigma_2 \sigma_1^{\pm 1} \sigma_2^{-1}. \end{array}$$

which captures the main idea of the proof of Lemma 7.4.

Let now τ_{gaps} denote a word containing gaps in the (ordered) indices but not starting with a gap. When managing the gaps, the first part of the word (before the first gap) remains intact after managing the gaps and the same carries through after the performance of a *bbm* on the first moving strand. That is, the following diagram commutes:

$$\begin{array}{ccc} \tau_{gaps} \cdot w & \xrightarrow[bbm]{1^{st}\,str.} & t^p \tau_{gaps_+} \cdot w_+ g_1^{\pm 1} \\ \Big| & & \Big| \\ man.\ gaps & & man.\ gaps \\ \downarrow & & \downarrow \\ \sum_i A_i \tau_i \cdot w_i & \xrightarrow[bbm]{1^{st}\,str.} & \sum_i A_i t^p \tau_{i_+} \cdot w_{i_+} g_1^{\pm 1} \end{array}$$

where $\tau_{gaps} \cdot w \in \Sigma$ and $\tau_i \in \Lambda^{aug}$, for all i.

In the case where the word $\tau_{gaps} \cdot w \in \Sigma$ starts with a gap, in [5] it is shown that equations obtained from $\tau \cdot w$ are equivalent to equations obtained from elements $\tau_i \cdot w_i \in \Sigma$, where τ_i are monomials in the t_i's *not* starting with a gap, *but with the bbm performed on any strand* (see Fig. 7.19).

The above are summarized in the following proposition:

Proposition 7.3 ([5, Proposition 3]) *In order to obtain an equivalent infinite system to the one obtained from elements in* Σ *by performing braid band moves on the first moving strand, it suffices to consider monomials in* Λ^{aug} *followed by braiding tails in* $\mathrm{H}_n(q)$ *and perform braid band moves on any moving strand.*

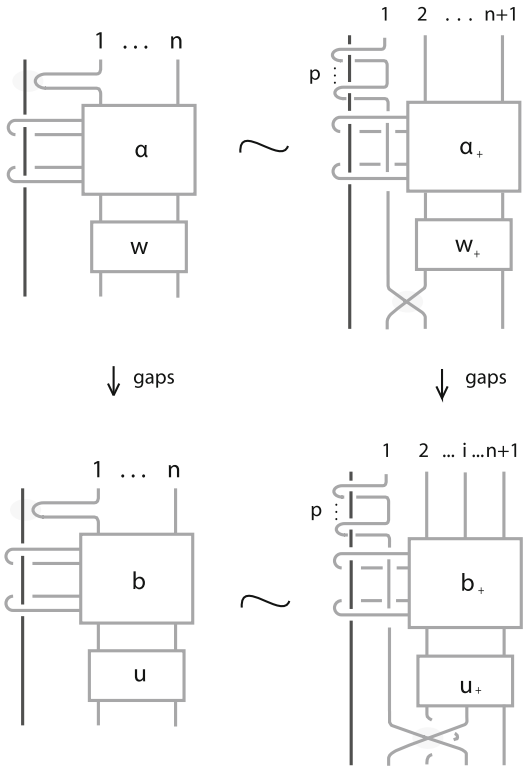


Fig. 7.19 *Bbm*'s before and after managing the gaps

7.4.4 *From the* $\mathbf{H}_n(q)$*-Module* Λ^{aug} *to the* $\mathbf{H}_n(q)$*-Module* Λ*: Ordering the Exponents*

The monomials in the t_i's that we obtain after managing the gaps are not elements in the set Λ, since the exponents of the loop generators are not necessarily ordered. We now order the exponents of the t_i's and we show that equations obtained from elements in the $\mathrm{H}_n(q)$-module Λ^{aug} reduce to equations obtained from elements in the $\mathrm{H}_n(q)$-module Λ.

The procedure we follow is similar to the one described in [4], but, as we mentioned earlier, in this case we do it simultaneously before and after the performance of a *bbm*.

Proposition 7.4 ([5, Proposition 4]) *Equations of the infinite system obtained from elements in* Λ^{aug} *followed by braiding tails in* $\mathrm{H}_n(q)$ *are equivalent to equations obtained from elements in* Λ *followed by braiding tails, where a braid band move can be performed on any moving strand.*

Proof It follows from Theorem 7.9, since all steps followed so as to order the exponents in a monomial in the t_i's remain the same after the performance of a *bbm*, ignoring the t^p appearing after the *bbm*. □

7.4.5 *From the* $\mathbf{H}_n(q)$*-Module* Λ *to* Λ*: Eliminating the Tails*

We now deal with the 'braiding tails'. Applying the same technique as in Theorem 7.10 before and after the performance of a *bbm*, we first prove that equations obtained by performing *bbm*'s on any moving strand on elements in Λ followed by words in $\mathrm{H}_n(q)$, reduce to equations obtained by performing *bbm*'s on any moving strand from elements in Λ^{aug} (with no 'braiding tails').

Proposition 7.5 ([5, Proposition 5]) *Equations obtained from bbm's on elements in* Λ *followed by words in* $\mathrm{H}_n(q)$ *are equivalent to equations obtained by performing a braid band move on any moving strand on elements in* Λ^{aug}.

Proof We perform a *bbm* on an element $a \cdot w$ in the $\mathrm{H}_n(q)$-module Λ and we cable the parallel strand with the surgery strand, see Fig. 7.20. We then apply Theorem 7.10 before and after the performance of the *bbm* and uncable the parallel strand. The proof is illustrated in Fig. 7.20. □

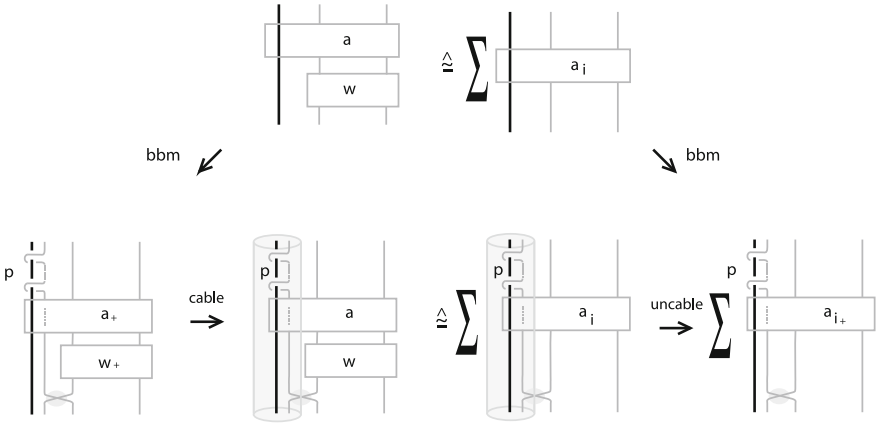


Fig. 7.20 The proof of Proposition 7.5

$$
\begin{array}{ccc}
tt_1t_2 \cdot g_1g_2g_1 & \xrightarrow[bbm]{1^{st}str.} & t^p t_1t_2t_3 \cdot g_2g_3g_2g_1^{\pm 1} \\
| & & | \\
elim.tails & & elim.tails \\
\downarrow & & \downarrow \\
(q-1)(q^2-q+1) \cdot tt_1t_2 & \xrightarrow[bbm]{1^{st}str.} & (q-1)(q^2-q+1) \cdot t^p t_1t_2t_3g_1^{\pm 1} \\
+ & & + \\
q(q-1)^2z \cdot tt_1^2 & \xrightarrow[bbm]{1^{st}str.} & q(q-1)^2z \cdot t^p t_1t_2^2g_1^{\pm 1} \\
+ & & + \\
\left[q^2(q-1)(q^2-q+1)z^2\right] \cdot t^3 & \xrightarrow[bbm]{1^{st}str.} & \left[q^2(q-1)(q^2-q+1)z^2\right] \cdot t^p t_1^3g_1^{\pm 1} \\
+ & & + \\
a \cdot t^2t_1 & \xrightarrow[bbm]{1^{st}str.} & a \cdot t^p t_1^2t_2g_1^{\pm 1}
\end{array}
$$

Example 7.8 In this example we demonstrate Proposition 7.5.
where $a = q^3z + q^2(q-1)^2 + 2q^2(q-1)^2z + q(q-1)^4z$.

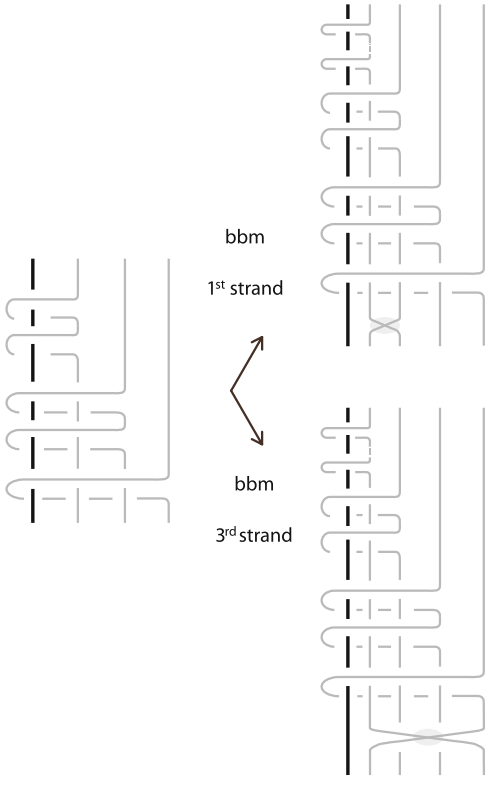


Fig. 7.21 The performance of a *bbm* on the $1st$ and on the $3rd$ strand of an element in Λ

In the process of eliminating the 'braiding tails' we have so far reached *bbm*'s in the set Λ^{aug} (with no 'braiding tails'). Following the same procedure and applying the same techniques as in [4], we finally reach the set Λ by means of the following:

Theorem 7.11 ([5, Theorem 8]) *In order to obtain the bbm equations needed for computing the HOMFLYPT skein module of the lens spaces $L(p,1)$, it suffices to perform braid band moves on any strand on elements in the basis Λ of $\mathcal{S}$(ST).*

Proof The proof is based on Theorems 7.8, 7.9 and 7.10 and the fact that the *bbm*'s commute with the stabilization moves (Lemma 7.1) and the skein (quadratic) relation (Lemma 7.2). □

Remark 7.5 The fact that the *bbm*'s and conjugation do not commute, results in the need of performing braid band moves on all moving strands of elements in Λ (Fig. 7.21).

7.5 Conclusions

In this paper we present recent results toward the computation of the HOMFLYPT skein module of the lens spaces $L(p,1)$, $\mathcal{S}(L(p,1))$, via braids. We first presented a new basis Λ of $\mathcal{S}$(ST) in braid form, which is crucial for the braid approach to $\mathcal{S}(L(p,1))$ and we then related $\mathcal{S}(L(p,1))$ to $\mathcal{S}$(ST) by means of equations resulting from *bbm*'s. In particular, we showed that by considering elements in the basis Λ and imposing on values on them of the Lambropoulou invariant X for knots and links in ST relations coming from the performance of *bbm*'s on all their moving strands, we arrive at an infinite system of equations, the solution of which coincides with the computation of $\mathcal{S}(L(p,1))$. Our results are summarized to the following:
In order to compute $\mathcal{S}(L(p,1))$ it suffices to solve the infinite system of equations:

$$X_{\widehat{\tau}} = X_{\widehat{bbm_i(\tau)}},$$

where $bbm_i(\tau)$ is the result of the performance of bbm on the ith-moving strand of $\tau \in \Lambda$, for all $\tau \in \Lambda$ and for all i.

In [2] we elaborate on the system. This is a very technical and difficult task. These results will serve as our main tool for computing $\mathcal{S}(L(p,q))$ in general.

References

1. Diamantis, I.: The HOMFLYPT skein module of the lens spaces $L(p,1)$ via braids. Ph.D. thesis, National Technical University of Athens, 2015
2. Diamantis, I., Lambropoulou, S.: The HOMFLYPT skein module of the lens spaces $L(p,1)$ via braids, in preparation
3. Diamantis, I., Lambropoulou, S.: Braid equivalences in 3-manifolds with rational surgery description. Topol. Appl. **194**, 269–295 (2015)

4. Diamantis, I., Lambropoulou, S.: A new basis for the HOMFLYPT skein module of the solid torus. J. Pure Appl. Algebra **220**(2) 577–605 (2016)
5. Diamantis, I., Lambropoulou, S., Przytycki, J.: Topological steps toward the HOMFLYPT skein module of the lens spaces $L(p, 1)$ via braids. J. Knot Theory Ramif. **25**, 13 (2016)
6. Freyd, P., Yetter, D., Hoste, J., Lickorish, W.B.R., Millett, K., Ocneanu, A.: A new polynomial invariant of knots and links. Bull. Am. Math. Soc. **12**, 239–249 (1985)
7. Gabrovšek, B., Mroczkowski, M.: The Homlypt skein module of the lens spaces $L(p, 1)$. Topol. Appl. **175**, 72–80 (2014)
8. Gilmer, P., Zhong, J.: On the HOMFLYPT skein module of $S^1 \times S^2$. Mathematische Zeitschrift **237**, 769–814 (2001)
9. Gilmer, P., Zhong, J.: The HOMFLYPT skein module of a connected sum of 3-manifolds. Algebr. Geom. Topol. **1**, 605–625 (2001)
10. Häring-Oldenburg, R., Lambropoulou, S.: Knot theory in handlebodies. J. Knot Theory Ramif. **11**(6), 921–943 (2002)
11. Hoste, J., Kidwell, M.: Dichromatic link invariants. Trans. Am. Math. Soc. **321**(1), 197–229 (1990)
12. Jones, V.F.R.: A polynomial invariant for links via Neumann algebras. Bull. Am. Math. Soc. **129**, 103–112 (1985)
13. Lambropoulou, S.: Solid torus links and Hecke algebras of B-type. In: Yetter, D.N. (ed.) Quantum Topology, pp. 225–245. World Scientific Press, Singapore (1994)
14. Lambropoulou, S.: Knot theory related to generalized and cyclotomic Hecke algebras of type B. J. Knot Theory Ramif. **8**(5), 621–658 (1999)
15. Lambropoulou, S.: Braid structures in handlebodies, knot complements and 3-manifolds. Proceedings of Knots in Hellas '98. World Scientific Press, Series of Knots and Everything, vol. 24, pp. 274–289 (2000)
16. Lambropoulou, S.: L-moves and Markov theorems. J. Knot Theory Ramif. **16**(10), 1–10 (2007)
17. Lambropoulou, S., Rourke, C.P.: Markov's theorem in 3-manifolds. Topol. Appl. **78**, 95–122 (1997)
18. Lambropoulou, S., Rourke, C.P.: Algebraic Markov equivalence for links in 3-manifolds. Compos. Math. **142**, 1039–1062 (2006)
19. Mroczkowski, M.: Polynomial Invariants of links in the projective space. Fund. Math. **184**, 223–267 (2004) (Proceedings of Knots in Poland 2003)
20. Przytycki, J.: Skein modules of 3-manifolds. Bull. Pol. Acad. Sci.: Math. **39, 1–2**, 91–100 (1991)
21. Przytycki, J.: Skein module of links in a handlebody. In: Apanasov, B., Neumann, W.D., Reid, A.W., Siebenmann, L. (eds.)Topology 90, Proceedings of the Research Semester in Low Dimensional Topology at OSU, pp. 315–342. De Gruyter Verlag (1992)
22. Przytycki, J.H., Traczyk, P.: Invariants of links of Conway type. Kobe J. Math. **4**, 115–139 (1987)
23. Turaev, V.G.: The Conway and Kauffman modules of the solid torus. Zap. Nauchn. Sem. Lomi **167**, 79–89 (1988). English translation: J. Soviet Math. 2799–2805 (1990)

Chapter 8
Some Hecke-Type Algebras Derived from the Braid Group with Two Fixed Strands

Dimitrios Kodokostas and Sofia Lambropoulou

Abstract We construct some Hecke-type algebras, and most notably the quotient algebra $\mathrm{H}_{2,n}(q)$ of the group-algebra $\mathbb{Z}[q^{\pm 1}]\, B_{2,n}$ of the mixed braid group $B_{2,n}$ with two identity strands and n moving ones, over the quadratic relations of the classical Hecke algebra for the braiding generators. The groups $B_{2,n}$ are known to be related to the knot theory of certain families of 3-manifolds, and the algebras $\mathrm{H}_{2,n}(q)$ are aimed for the construction of invariants of oriented knots and links in these manifolds. To this end, one needs a suitable basis of $\mathrm{H}_{2,n}(q)$, and we have singled out a subset Λ_n of this algebra for which we proved it is a spanning set, whereas ongoing research aims at proving it to be a basis.

Keywords Mixed braid group on two fixed strands · Mixed Hecke algebra Quadratic relation · Hecke-type algebras

Introduction

It is established that knots and links in arbitrary knot complements, in compact, connected, oriented (c.c.o.) 3-manifolds and in handlebodies may be represented by mixed links and mixed braids in S^3 [4, 9, 13]. The braid structures related to knots and links in the above spaces are the mixed braid groups $B_{m,n}$ and appropriate cosets of theirs [15]. An element in $B_{m,n}$ is a classical braid in S^3 on $m+n$ strands with the first m strands forming the identity braid. The mixed braid groups enable the algebraic formulation of the geometric braid equivalences in the above spaces [4, 9, 16].

D. Kodokostas · S. Lambropoulou (✉)
Department of Applied Mathematics, National Technical University of Athens, Zografou Campus, 15780 Athens, Greece
e-mail: sofia@math.ntua.gr
URL: http://www.math.ntua.gr/ sofia

D. Kodokostas
e-mail: dkodokostas@math.ntua.gr

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_8

In this paper we focus on the mixed braid groups $B_{2,n}$, which are related to knots and links in certain families of 3-manifolds like, for example, the handlebody of genus two, the complement of the 2-unlink in S^3 and the connected sums of two lens spaces, which are of interest also to some biological applications [3]. We define the quotient algebras $\mathrm{H}_{2,n}(q)$, $\mathrm{H}_{2,n}(q, u_1, v_1)$ and $\mathrm{H}_{2,n}(q, u_1, \ldots, u_{d_1}, v_1, \ldots, v_{d_2})$ of $B_{2,n}$ over the quadratic relations of the classical Iwahori–Hecke algebra for the braiding generators, and polynomial relations for the looping generators. We then focus on $\mathrm{H}_{2,n}(q)$ and present a subset Λ_n of it, indicating the reason it has to be a spanning set for its additive structure. The set Λ_n potentially constitutes a linear basis of $\mathrm{H}_{2,n}(q)$, a fact whose proof is the object of ongoing research.

The sets Λ_n for $n \in \mathbb{N}$ are destined to provide an appropriate inductive basis for the sequence of algebras $\mathrm{H}_{2,n}(q)$, $n \in \mathbb{N}$, in order to construct Homflypt-type invariants for oriented links in 3-manifolds whose braid structure is encoded by the groups $B_{2,n}$. It is known that the mixed braid groups $B_{1,n}$ have been utilized for constructing Homflypt-type invariants for oriented links in the solid torus [8, 12, 14] and the lens spaces $L(p, 1)$ [6], following the original Jones construction of the classical Homflypt polynomial for oriented links in S^3 using the Iwahori–Hecke algebra of type A and the Ocneanu trace [10]. For our purposes we first need to construct appropriate algebras related to the mixed braid groups $B_{2,n}$, and then to chose an appropriate inductive bases on them, so that the construction of Oceanu-type Markov traces on these algebras would be possible, which subsequently can be used for the construction of link invariants.

The paper is organized as follows: in Sect. 8.1.1 we recall the definition and a presentation of the mixed braid group $B_{2,n}$ and we define some important elements of it which we call loopings. In Sect. 8.1.2 we define our quotient algebras $\mathrm{H}_{2,n}(q)$, $\mathrm{H}_{2,n}(q, u_1, v_1)$ and $\mathrm{H}_{2,n}(q, u_1, \ldots, u_{d_1}, v_1, \ldots, v_{d_2})$. In Sect. 8.2 we provide a potential basis Λ_n for the algebra $\mathrm{H}_{2,n}(q)$, and we give the necessary lemmata for proving it to be spanning set of the algebra.

8.1 The Mixed Braid Groups $B_{2,n}$ and Related Hecke-Type Algebras

8.1.1 The Mixed Braid Group $B_{2,n}$ On Two Mixed Strands and Other Related Groups

For each $n \in \mathbb{N}$, the elements of the *mixed braid group $B_{2,n}$ on two fixed strands* are defined to be the braids with $n + 2$ strands where the first two of them are straight, and the group operation is by definition the usual braid concatenation. A description of $B_{2,n}$ in terms of generators and relations is the following [15]:

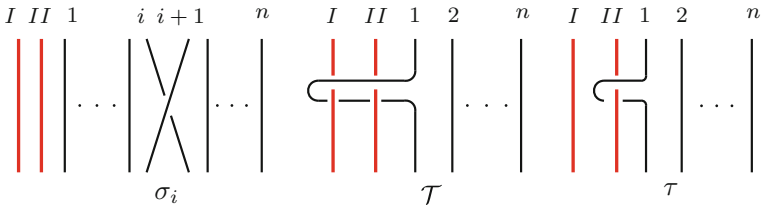


Fig. 8.1 The generators of $B_{2,n}$

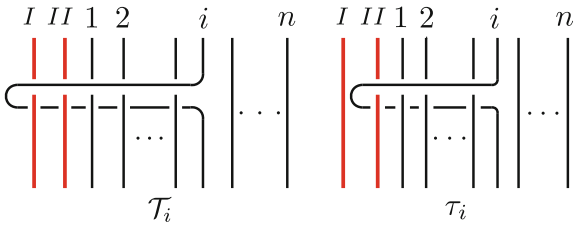


Fig. 8.2 The looping elements $\mathcal{T}_i$, τ_i

$$B_{2,n} = \left\langle \tau, \mathcal{T}, \sigma_1, \ldots, \sigma_{n-1} \;\middle|\; \begin{array}{l} \sigma_k\sigma_j = \sigma_j\sigma_k \ (|k-j|>1), \ \ \sigma_k\sigma_{k+1}\sigma_k = \sigma_{k+1}\sigma_k\sigma_{k+1} \ (1 \le k \le n-1) \\ \mathcal{T}\,\sigma_k = \sigma_k\,\mathcal{T} \ (k \ge 2), \quad \mathcal{T}\,\sigma_1\,\mathcal{T}\,\sigma_1 = \sigma_1\,\mathcal{T}\,\sigma_1\,\mathcal{T} \\ \tau\,\sigma_k = \sigma_k\,\tau \ (k \ge 2), \ \ \tau\,\sigma_1\,\tau\,\sigma_1 = \sigma_1\,\tau\,\sigma_1\,\tau, \\ \tau(\sigma_1\mathcal{T}\sigma_1) = (\sigma_1\mathcal{T}\sigma_1)\tau \end{array} \right\rangle$$

where σ_i, τ, $\mathcal{T}$ are shown in Fig. 8.1; I, II indicate the two *fixed strands* as they are called, whereas $1, 2, \ldots, n$ indicate the *moving strands*. The braids $\tau, \mathcal{T}$ and their inverses are called the *looping generators*, whereas σ_i and its inverse are called the i-th *braiding generators* for $i = 1, 2, \ldots, n-1$, whereas i is called the *index* of the i-th braiding generators.

Below we define the elements $\mathcal{T}_i$, τ_i, $i = 1, \ldots, n$ of $B_{2,n}$ which will be of central importance to us in what follows.

Definition 8.1 The *looping elements* or just *loopings* $\mathcal{T}_i$, τ_i of $B_{2,n}$ are those braids in which all strands are straight except for the i-th moving strand that loops once around the first or the second fixed strand respectively, first going over and then (after the looping) under the rest of the strands to its left (see Fig. 8.2). We use the name *looping* for the inverses of these elements as well, and we say that each one of the loopings $\mathcal{T}_i^{\pm 1}$, $\tau_i^{\pm 1}$ has *index* i. Formally:

$$\mathcal{T}_1 := \mathcal{T}, \quad \tau_1 := \tau \ \ \textit{and}\ \mathcal{T}_i := \sigma_{i-1}\ldots\sigma_1\mathcal{T}\sigma_1\ldots\sigma_{i-1}, \quad \tau_i := \sigma_{i-1}\ldots\sigma_1\tau\sigma_1\ldots\sigma_{i-1}\ \textit{for}\ i > 1.$$

As is shown in Fig. 8.3, the defining relation $\tau\sigma_1\mathcal{T}\sigma_1 = \sigma_1\mathcal{T}\sigma_1\tau$ of $B_{2,n}$ which is now written as $\tau_1\sigma_1\mathcal{T}_1\sigma_1 = \sigma_1\mathcal{T}_1\sigma_1\tau_1$, holds in general for all $i = 1, 2 \ldots, n-1$ (just slide the τ_i looping to pass through the $\mathcal{T}_i$ looping):

$$\tau_i\sigma_i\mathcal{T}_i\sigma_i = \sigma_i\mathcal{T}_i\sigma_i\tau_i.$$

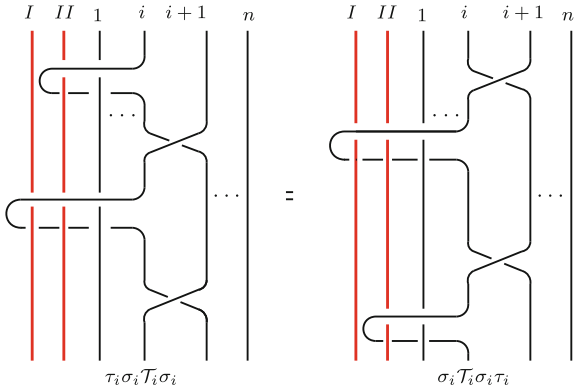


Fig. 8.3 The relation $\tau_i\sigma_i\mathcal{T}_i\sigma_i = \sigma_i\mathcal{T}_i\sigma_i\tau_i$ in $B_{2,n}$

Clearly, $B_{2,n}$ is generated by the set $\{\mathcal{T}_i, \tau_i, \sigma_i | i = 1, \ldots, n\}$ as well. Also, clearly, the group $B_{2,n}$ is a subgroup of the usual braid group B_{2+n} in $2+n$ strands [15]. In turn, $B_{2,n}$ contains some important subgroups. One of them is the *pure mixed braid group on two fixed strands* $P_{2,n}$ that consists of all pure braids in $B_{2,n}$. For a further study of the structure of the groups $B_{2,n}$, $P_{2,n}$ and their generalizations $B_{m,n}$, $P_{m,n}$ as well as the underlying Coxeter-type groups see [2, 15].

Other important subgroups of $B_{2,n}$ are those generated by $\mathcal{T}, \sigma_1, \ldots, \sigma_{n-1}$ and by $\tau, \sigma_1, \ldots, \sigma_{n-1}$, which are isomorphic to the *mixed braid group on one fixed strand* $B_{1,n}$ defined in terms of generators and relations in an analogous manner as $B_{2,n}$ and which, in fact, is the Artin braid group of type B. Indeed, the defining relations $\mathcal{T}\sigma_1\mathcal{T}\sigma_1 = \sigma_1\mathcal{T}\sigma_1\mathcal{T}$ and $\tau\sigma_1\tau\sigma_1 = \sigma_1\tau\sigma_1\tau$ of $B_{2,n}$ are of the same type as the four-term defining relations of the Artin braid group of type B.

8.1.2 *The Algebra* $\mathbf{H_{2,n}(q)}$ *and Other Related Algebras*

We define the algebra $\mathrm{H}_{2,n}(q)$ for each $n \in \mathbb{N}$ as a quotient of an appropriate group-algebra of $B_{2,n}$ over appropriate quadratic relations. Namely:

Definition 8.2 The *mixed Hecke algebra on two fixed strands* $\mathrm{H}_{2,n}(q)$ is defined as the unital associative algebra:

$$\mathrm{H}_{2,n}(q) := \frac{\mathbb{Z}[q^{\pm 1}]\, B_{2,n}}{\langle {\sigma_i}^2 - (q-1)\,\sigma_i - q \cdot 1,\ i = 1, 2, \ldots, n-1\rangle},$$

where q is a variable.

In general we use the same notation for the elements of $B_{2,n}$ when considered as elements of $\mathrm{H}_{2,n}(q)$, except for σ_i which we denote g_i, $i = 1, \ldots, n$. $\mathrm{H}_{2,n}(q)$ has equivalently a presentation with generators $\tau, \mathcal{T}, g_1, \ldots, g_{n-1}$ and relations:

$$\begin{array}{lrll}
(b_1) & g_k g_{k+1} g_k &= g_{k+1} g_k g_{k+1} & \text{for } 1 \le k \le n-1 \\
(b_2) & g_k g_j &= g_j g_k & \text{for } |k-j| > 1 \\
(\mathcal{T}_1) & \mathcal{T}\, g_k &= g_k\, \mathcal{T} & \text{for } k \ge 2 \\
(\mathcal{T}_2) & \mathcal{T}\, g_1\, \mathcal{T}\, g_1 &= g_1\, \mathcal{T}\, g_1\, \mathcal{T} & \\
(\tau_1) & \tau\, g_k &= g_k\, \tau & \text{for } k \ge 2 \\
(\tau_2) & \tau\, g_1\, \tau\, g_1 &= g_1\, \tau\, g_1\, \tau & \\
(m) & \tau(g_1 \mathcal{T} g_1) &= (g_1 \mathcal{T} g_1)\tau & \\
(q) & g_i^2 &= (q-1)\, g_i + q \cdot 1 & \text{for } i = 1, 2, \ldots, n-1
\end{array} \tag{8.1}$$

(here (b) stands for "braid", (m) for "mixed" and (q) for "quadratic"). The elements $\tau, \mathcal{T}$ and their inverses will be called the *looping generators* of the algebra. The elements $g_1, \ldots, g_{n-1}$ and their inverses will be called the *braiding generators* of the algebra, whereas i will be the *index* of g_i, g_i^{-1}.

Since the classical Artin braid group B_n embeds naturally in $B_{2,n}$, we have that the classical Iwahori–Hecke Algebra $H_n(q)$ is a subalgebra of $\mathrm{H}_{2,n}(q)$ in a natural way as well. Furthermore, note that the relations $(\mathcal{T}_2)$ and (τ_2) are of the same type as the well-known four-term defining relation of the Artin braid group of type B which is realized here by the mixed braid group $B_{1,n}$ with one fixed strand, hence it embeds naturally in $B_{2,n}$. So, the algebra $\mathrm{H}_{2,n}(q)$ extends naturally the mixed Hecke algebra $\mathrm{H}_{1,n}(q)$ introduced in [14] as "generalized Hecke algebra" of type B. The algebra $\mathrm{H}_{2,n}(q)$ clearly contains two subalgebras isomorphic to $\mathrm{H}_{1,n}(q)$.

We can define a lot of other interesting related algebras, a few of which as follows:

Definition 8.3 The algebra $\mathrm{H}_{2,n}(q, u_1, v_1)$ is defined as:

$$\mathrm{H}_{2,n}(q, u_1, v_1) := \frac{\mathbb{Z}\,[q^{\pm 1}, u_1^{\pm 1}, v_1^{\pm 1}]\, B_{2,n}}{\langle (q),\ \mathcal{T}^2 = (u_1 - 1)\mathcal{T} + u_1 \cdot 1,\ \tau^2 = (v_1 - 1)\tau + v_1 \cdot 1 \rangle},$$

where distinct variables u_1, v_1 are associated to $\mathcal{T}, \tau$.

Note that the relations $(\mathcal{T}_2)$ and (τ_2) are of the same type as the defining relations of the Hecke algebra of type B [7]. Furthermore, it is clear from the quadratic relations for the looping generators in Definition 8.3 that the algebra $\mathrm{H}_{2,n}(q, u_1, v_1)$ extends the classical Hecke algebra of type B. In fact $\mathrm{H}_{2,n}(q, u_1, v_1)$ contains two subalgebras isomorphic to the Hecke algebra of type B.

Definition 8.4 The *cyclotomic algebra* $\mathrm{H}_{2,n}(q, u_1, \ldots, u_{d_1}, v_1, \ldots, v_{d_2})$ is defined as:

$$\mathrm{H}_{2,n}(q, u_1, \ldots, u_{d_1}, v_1, \ldots, v_{d_2}) := \frac{\mathbb{Z}[q^{\pm 1}, u_1^{\pm 1}, \ldots, u_{d_1}^{\pm 1}, v_1^{\pm 1}, \ldots, v_{d_2}^{\pm 1}]\, B_{2,n}}{\langle (q),\ (\mathcal{T} - u_1) \ldots (\mathcal{T} - u_{d_1}) = 0,\ (\tau - v_1) \ldots (\tau - v_{d_1}) = 0 \rangle},$$

where $q, u_1, \ldots, u_{d_1}, v_1, \ldots, v_{d_2}$ are variables and the last two relations are called *cyclotomic relations* for $\mathcal{T}$ and τ respectively (see Fig. 8.4).

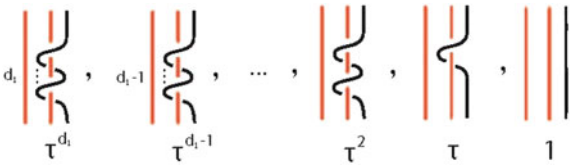


Fig. 8.4 The mixed braids involved in the cyclotomic relation of the loop generator τ

Analogously to the algebras defined above, by relations for $(\mathcal{T}_2)$ and (τ_2) and the defining relation for $\mathcal{T}, \tau$ in Definition 4 it follows that the algebra $\mathrm{H}_{2,n}(q, u_1, \ldots, u_{d_1}, v_1, \ldots, v_{d_2})$ extends naturally the Ariki–Koike algebra of type B [1], and in fact it contains two subalgebras isomorphic to the Ariki–Koike algebra.

The mixed Hecke algebra $\mathrm{H}_{1,n}(q)$, the Iwahori–Hecke algebra of type B, and the Ariki–Koike algebra of type B are all related to the knot theory of the solid torus and the lens spaces [5, 6, 8, 12, 14]. Note that each one of the three types of algebras that we define here surjects naturally onto its corresponging B-type algebra, for example via the following mappings respectively:

- $\mathcal{T} \mapsto 1, \tau \mapsto \tau, g_i \mapsto g_i$ surjects $\mathrm{H}_{2,n}(q)$ onto $\mathrm{H}_{1,n}(q)$.
- $\mathcal{T} \mapsto 1, \tau \mapsto \tau, g_i \mapsto g_i$ and specializing u_1 to 1, surjects $\mathrm{H}_{2,n}(q, u_1, v_1)$ onto the Hecke algebra of type B.
- $\mathcal{T} \mapsto 1, \tau \mapsto \tau, g_i \mapsto g_i$ and specializing u_i to 1 for $i = 1, \ldots, d_1$, surjects $\mathrm{H}_{2,n}(q, u_1, \ldots, u_{d_1}, v_1, \ldots, v_{d_2})$ onto the Ariki–Koike algebra of type B.

Finally, let us note that the algebras $\mathrm{H}_{2,n}(q, u_1, v_1)$ and $\mathrm{H}_{2,n}(q, u_1, \ldots, u_{d_1}, v_1, \ldots, v_{d_2})$ can be viewed as quotient algebras of $\mathrm{H}_{2,n}(q)$, if in Definitions 8.2–8.4 we use $\mathbb{Z}[q^{\pm 1}, u_1^{\pm 1}, \ldots, u_{d_1}^{\pm 1}, v_1^{\pm 1}, \ldots, v_{d_2}^{\pm 1}]$ as a common ring of coefficients for all three algebras.

8.2 A Spanning Set and Potential Basis for the Algebra $\mathrm{H}_{2,n}(q)$.

We still call $\mathcal{T}_i, \tau_i$ and their inverses as *looping elements* or *loopings* when we cosider them as elements of $\mathrm{H}_{2,n}(q)$, and similarly we call i as their *index*. For $\mathcal{T}_i, \tau_i$ as elements of the algebra we have

$$\mathcal{T}_i = g_{i-1} \ldots g_1 \mathcal{T} g_1 \ldots g_{i-1} \text{ and } \tau_i = g_{i-1} \ldots g_1 \tau g_1 \ldots g_{i-1}.$$

Our aim is to provide a "nice" form for any element w of $\mathrm{H}_{2,n}(q)$ using these looping elements and the g_i's, so that a possible spanning set of the algebra reveals itself. Since w is a $\mathbb{Z}[q^{\pm 1}]$-linear combination of images in the algebra $\mathrm{H}_{2,n}(q)$ of braids in $B_{2,n}$, one has to think about only the case of putting an image of a braid w in a "nice" form.

Now let us recall that the set $\{\mathcal{T}_i, \tau_i, \sigma_i | i = 1, \ldots, n\}$ generates $B_{2,n}$, thus an arbitrary braid w of $B_{2,n}$ is written as a finite product of elements of this set and their inverses; as a matter of fact, it can be written so in many ways. Previous work done on specific subsets of $B_{2,n}$, shows that considering their elements as belonging to appropriate related algebras similar to $\mathrm{H}_{2,n}(q)$, we can put them in canonical forms useful for constructing Markov traces over these algebras. For example whenever w is a product of only the g_i's (considered as an element of the algebra) then w actually belongs to $\mathrm{H}_n(q)$ and as such it is subjected to the canonical form of the classical Hecke algebra $\mathrm{H}_n(q)$ of type A, given by V.F.R. Jones [10]. Also, whenever w is a product of only τ_i, g_i's (thus containing no $\mathcal{T}_i$'s), it actually belongs to $\mathrm{H}_{1,n}(q)$ (as mentioned in the previous section, this is the generalized Hecke algebra of type B), and therefore it is subjected to the canonical form given in [14]. Such a w is written as a finite $\mathbb{Z}[q^{\pm 1}]$-linear combination of products of τ_i's and g_i's with the τ_i's appearing first, and moreover with the indices of the τ_i's in increasing order from left to right.

Theorem 8.1 below tells us how to bring any w of $B_{2,n}$ to a similar "nice" form when considered as an element of the algebra $\mathrm{H}_{2,n}(q)$. At the same time we get a spanning set Λ_n for the additive structure of $\mathrm{H}_{2,n}(q)$ as a $\mathbb{Z}[q^{\pm 1}]$-module. What the theorem actually says is that every element of the algebra $\mathrm{H}_{2,n}(q)$ is written as a finite $\mathbb{Z}[q^{\pm 1}]$-linear combination of products of $\mathcal{T}_i$, τ_i's and g_i's with the $\mathcal{T}_i$, τ_i's appearing first, and moreover with the indices of the τ_i's in increasing order from left to right. To achieve this for the image of a braid given in a product form we can try first to push all g_i's (i.e. the images of the σ_i's) at the end using braid isotopies (at the braid level) together with the quadratic relations in the algebra $\mathrm{H}_{2,n}(q)$ (see Lemma 8.1). And then we can similarly try to push all loopings with big indices after those with smaller ones (see Lemma 8.2). Working out specific examples one soon realizes that pushing the g_i's is always possible, and that pushing the loopings with big indices after those with small ones can be almost always achieved, except that in the process some new g_i's might be created, and pushing them anew to the end might increase the indices of the loopings from which it passes, leaving quite open the question of whether the indices of the loopings can indeed be ordered. We deal with this issue in Lemma 8.3.

Theorem 8.1 *Any element in* $\mathrm{H}_{2,n}(q)$ *can be written as a finite* $\mathbb{Z}[q^{\pm 1}]$*-linear combination of the form (suppressing the coefficient in* $\mathbb{Z}[q^{\pm 1}]$ *of each term):*

$$\sum (\Pi_1 \Pi_2 \cdots \Pi_n) G$$

where G *is a finite product of braiding generators, and* Π_i *is a finite product of only the loopings* $\mathcal{T}_i, \tau_i, \mathcal{T}_i^{-1}, \tau_i^{-1}$ *for all* i*. Thus the following is a spanning set of the algebra* $\mathrm{H}_{2,n}(q)$*:*
$\Lambda_n := \{\Pi_1 \Pi_2 \cdots \Pi_n G \mid G =$ *an element in some basis of* $\mathrm{H}_n(q)$ *and* $\Pi_i =$ *finite product of only the loopings* $\mathcal{T}_i, \tau_i, \mathcal{T}_i^{-1}, \tau_i^{-1}, \forall i$ *}.*

The definition of the looping elements and braiding generators can be repeated for the other algebras which we defined in Sect. 8.1.2, and the proof of Theorem 8.1 can be repeated unaltered step by step to get:

Theorem 8.2 *Let* $A = \mathrm{H}_{2,n}(q, u_1, v_1)$, $R = \mathbb{Z}[q^{\pm 1}, u_1^{\pm 1}, v_1^{\pm 1}]$ *or* $A = \mathrm{H}_{2,n}(q, u_1, \ldots, u_{d_1}, v_1, \ldots, v_{d_2})$, $R = \mathbb{Z}[q^{\pm 1}, u_1^{\pm 1}, \ldots, u_{d_1}^{\pm 1}, v_1^{\pm 1}, \ldots, v_{d_2}^{\pm 1}]$. *Then any element in A can be written as a finite R-linear combination of the form (suppressing the coefficient in R of each term):*

$$\sum (\Pi_1 \Pi_2 \cdots \Pi_n) G$$

where G is a finite product of braiding generators, and Π_i *is a finite product of only the loopings* $\mathcal{T}_i, \tau_i, \mathcal{T}_i^{-1}, \tau_i^{-1}$ *for all i. Thus the following is a spanning set of the algebra R:*
$\Lambda_n := \{\Pi_1 \Pi_2 \cdots \Pi_n G \mid G =$ *an element in some basis of* $\mathrm{H}_n(q)$ *and* $\Pi_i =$ *finite product of only the loopings* $\mathcal{T}_i, \tau_i, \mathcal{T}_i^{-1}, \tau_i^{-1}, \forall i\ \}$.

Below we list the necessary lemmata for the proof of Theorem 8.1 which is quite technical since it has to deal carefully with the indices appearing in a given product of loopings as well as with the possible recursion phenomena that might arise during the process. We provide the actual proof in [11]. The lemmata equip us with specific formulas for pushing braiding generators to the right of a product of loopings, and also for pushing loopings with high indices to the right of loopings with lower indices in a product of loopings. The lemmata also explain how we can deal with recursion phenomena.

Lemma 8.1 *Let us call* $A = q^{-1} - 1$, $B = q - 1$. *And let us denote the identity element of* $\mathrm{H}_{2,n}(q)$ *by* 1. *Then the following hold in* $\mathrm{H}_{2,n}(q)$*:*

(1) $g_i^{-1} = q^{-1} g_i + A \cdot 1$

(2) $g_i \mathcal{T}_j^{\pm 1} = \mathcal{T}_j^{\pm 1} g_i$	$g_i \tau_j^{\pm 1} = \tau_j^{\pm 1} g_i$	*whenever* $j \neq i, i+1$
(3) $g_i \mathcal{T}_i = q^{-1} \mathcal{T}_{i+1} g_i + A \mathcal{T}_{i+1}$	$g_i \tau_i = q^{-1} \tau_{i+1} g_i + A \tau_{i+1}$	
(4) $g_i \mathcal{T}_i^{-1} = \mathcal{T}_{i+1}^{-1} - A \mathcal{T}_{i+1}^{-1} + A \mathcal{T}_i^{-1}$	$g_i \tau_i^{-1} = \tau_{i+1}^{-1} - A \tau_{i+1}^{-1} + A \tau_i^{-1}$	
(5) $g_i \mathcal{T}_{i+1} = q \mathcal{T}_i g_i + B \mathcal{T}_{i+1}$	$g_i \tau_{i+1} = q \tau_i g_i + B \tau_{i+1}$	
(6) $g_i \mathcal{T}_{i+1}^{-1} = q^{-1} \mathcal{T}_i^{-1} g_i + A \mathcal{T}_i^{-1}$	$g_i \tau_{i+1}^{-1} = q^{-1} \tau_i^{-1} g_i + A \tau_i^{-1}$.	

(7) (The passage property) Any product $g_k^{\epsilon} t_l^{\zeta}$ ($\epsilon, \zeta \in \{1, -1\}$, t_l *a looping) can be written as a finite linear combination of the form (suppressing the coefficient in* $\mathbb{Z}[q^{\pm 1}]$ *of each term on the right-hand side):*

$$g_i^{\epsilon} t_l^{\zeta} = \sum t_l^{\zeta} g_i^{\epsilon} + \sum t_i^{\zeta} + \sum t_i^{\zeta} g_i^{\epsilon} + \sum t_{i+1}^{\zeta}.$$

(where possibly some of the terms are missing).

(8) (The big passage property) Let Π *be a finite product of k in number loopings with indices in the interval* $[m, M]$, *and let* $i \in [m, M-1]$. *Then* $g_i^{\pm 1} \Pi$ *can be written as a finite linear combination of the form (suppressing the coefficient in* $\mathbb{Z}[q^{\pm 1}]$ *of each term on the right-hand side):*

$$g_i^{\pm 1} \Pi = \sum \Pi_1 g_i^{\pm 1} + \sum \Pi_2$$

(where possibly some terms are missing) with each Π_1, Π_2 a product of k in number loopings with indices in $[m, M]$.

Part (2) of the lemma can be seen in the braid level via trivial braid isotopies, and parts (2)–(6) can also be seen pictorially after at most two applications of the quadratic relation to the braids of the left-hand side. These are summarized in part (7). Since on both sides of parts (2)–(6) each term contains a single looping and in part (7) the index of the looping either does not change at all or if it does, it decreases by 1 but then never below the index i of the braiding generator, or else it increases by 1 but then never by 1 above the index i of the braiding generator, we get part (8) at once.

The following lemma describes how we can push a looping with high index to the right of a looping with smaller index.

Lemma 8.2 *For $j < i$ and $\epsilon, \zeta \in \{1, -1\}$ each one of the words $\mathcal{T}_i^{\epsilon}\mathcal{T}_j^{\zeta}$, $\mathcal{T}_i^{\epsilon}\tau_j^{\zeta}$, $\tau_i^{\epsilon}\mathcal{T}_j^{\zeta}$, $\tau_i^{\epsilon}\tau_j^{\zeta}$ can be written as a linear combination of the form (suppressing the coefficient in $\mathbb{Z}[q^{\pm 1}]$ of each term on the right-hand side):*

1. $\mathcal{T}_i^{\epsilon}\mathcal{T}_j^{\zeta} = \mathcal{T}_j^{\zeta}\mathcal{T}_i^{\epsilon}$, $\tau_i^{\epsilon}\tau_j^{\zeta} = \tau_j^{\zeta}\tau_i^{\epsilon}$, $\mathcal{T}_i^{\epsilon}\tau_j^{\zeta} = \tau_j^{\zeta}\mathcal{T}_i^{\epsilon}$
2. $\tau_i^{\epsilon}\mathcal{T}_j^{\epsilon} = \mathcal{T}_j^{\epsilon}\tau_i^{\epsilon} + \mathcal{T}_j^{\epsilon}\tau_i^{\epsilon}G^{\epsilon} + \tau_j^{\epsilon}\mathcal{T}_i^{\epsilon}G^{\epsilon}$, *where* $G = g_j g_{j+1} \dots g_{i-2} g_{i-1}^{-1} g_{i-2}^{-1} \dots g_{j+1}^{-1} g_j^{-1}$.
3. $\tau_i^{\epsilon}\mathcal{T}_j^{-\epsilon} = \mathcal{T}_j^{-\epsilon}\tau_i^{\epsilon} + \mathcal{T}_j^{-\epsilon}\tau_j^{\epsilon}G^{\epsilon} + \tau_i^{\epsilon}\mathcal{T}_i^{-\epsilon}G^{\epsilon}$, *where* $G = g_j g_{j+1} \dots g_{i-2} g_{i-1} g_{i-2} \dots g_{j+1} g_j$.

The proof of this lemma is easy, as part (1) can be seen at the braid level via braid isotopies, and the last two parts can also be seen at the braid level as a double application of the quadratic relation at the obvious crossings so that the i-looping can be moved above the j-looping.

The lemma that follows is the last one that we need for the proof of Theorem 8.1, and it says that a certain class of words actually satisfies the theorem. These words have the odd property that whenever we apply all the previous formulas in order to write them as sums of monomials in the way demanded by the theorem, they are written so except from the fact that one of the monomials is the word itself. Fortunately, the coefficients appearing in these equalities are well behaved and we can solve for the given word so that it is indeed expressed in way described in the theorem. This recursion phenomenon is possible only because one of the monomials on the right-hand side in case (3) of Lemma 8.2 still starts with an i-looping instead of a j-looping.

In the statement of the lemma it is convenient to write $[i, j]$ in the bottom of a product of looping or braiding generators to indicate that their indices lie in the interval $[i, j]$, and to write $< i, j >$ to indicate that these indices are also in increasing order (from left to right).

Lemma 8.3 *Let us denote elements in $\{\mathcal{T}_i^{\pm 1}, \tau_i^{\pm 1}\}$ indiscreetly by t_i. Then each one of the words $\tau_M^{\epsilon}\mathcal{T}_M^{-\epsilon}t_m^{\zeta}$ with $m < M$ and $\epsilon, \zeta \in \{-1, 1\}$ can be written as a finite*

linear combination of the form (suppressing the coefficient in $\mathbb{Z}[q^{\pm 1}]$ of each term on the right-hand side):

$$\tau_M^{\epsilon} \mathcal{T}_M^{-\epsilon} t_m^{\zeta} = \sum \left(t_m t_{m_1} t_{m_2} \right)_{<m,M>} G_{[m,M-1]}$$

where each G is a finite product of $g_i^{\pm 1}$'s (notice the crucial fact that every term of the last sum starts with an m-looping).

The proof of this lemma is not as immediate as the proofs in the previous lemmata. We have to examine all possible cases separately, applying the quadratic relations appropriately and using isotopies at the braid level. The reader is referred to [11] for full details of the proof, as well as for the proof of Theorem 8.1 which is a consequence of Lemmata 8.1–8.3.

Remark In [11], we also conjecture that the set Λ_n is a linear basis for the algebra $\mathrm{H}_{2,n}(q)$. This is not straightforward to prove, as the algebra $\mathrm{H}_{2,n}(q)$ is infinite dimensional. Nevertheless we can get some insight of how Λ_n behaves, by examining its counterparts in the other algebras $\mathrm{H}_{2,n}(q, u_1, v_1)$ and $\mathrm{H}_{2,n}(q, u_1, \ldots, u_{d_1}, v_1, \ldots, v_{d_2})$, defined in this paper, and for which these counterparts also constitute spanning sets (Theorem 8.2). Although these algebras are infinite dimensional too, the exponents of the loopings in the elements of the above spanning sets are bounded, a fact that makes these algebras easier to study.

8.3 Conclusion and Further Research

In this paper we have defined some Hecke-type algebras related to the mixed braid group $B_{2,n}$ on two fixed strands, and we have focused on one of them, namely on the mixed Hecke algebra $\mathrm{H}_{2,n}(q)$ which is defined as the quotient of the group-algebra $\mathbb{Z}[q^{\pm 1}]B_{2,n}$ over the quadratic relations of the usual Hecke algebra. These algebras are related to the knot theory of various 3-manifolds whose knot structure is encoded by the mixed braid groups $B_{2,n}$, such as handelebodies of genus two, and connected sums of lens spaces. We have given here a subset Λ_n of $\mathrm{H}_{2,n}(q)$ and provided the necessary lemmata along with hints for their truth, for proving that Λ_n is a spanning set for the additive structure of the algebra [11]. We conjecture that Λ_n is actually a basis for $\mathrm{H}_{2,n}(q)$ and this the subject of current research. Then, based on previous work done on similar Hecke-type algebras, we expect that we can use Λ_n for the construction of knot invariants in the above 3-manifolds.

Acknowledgements This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES: Reinforcement of the interdisciplinary and/or inter-institutional research and innovation.

References

1. Ariki, S., Koike, K.: A Hecke algebra of $\mathbb{Z}/r\mathbb{Z} \wr S_n$ and construction of its irreducible representations. Adv. Math. **106**, 216–243 (1994)
2. Bardakov, V: Braid groups in handlebodies and corresponding Hecke algebras,. In: Lambropoulou, S., Stefaneas, P., Theodorou, D., Kauffman, L.(eds) Algebraic Modeling of Topological and Computational Structures and Applications, Springer Proceedings in Mathematics and Statistics (PROMS) (2017)
3. Buck, D., Mauricio, M.: Connect sum of lens spaces surgeries: application to hin recombination. Math. Proc. Camb. Phil. Soc. **150**(3), 505–525 (2011). https://doi.org/10.1017/S0305004111000090
4. Diamantis, I., Lambropoulou, S.: Braid equivalences in 3-manifolds with rational surgery description. Topol. Appl. **194**(2), 269–295 (2015). https://doi.org/10.1016/j.topol.2015.08.009
5. Diamantis, I., Lambropoulou, S.: A new basis for the Homflypt skein module of the solid torus. J. Pure Appl. Algebra **220**(2), 557–605 (2015). http://doi.org/10.1016/j.jpaa.2015.06.014
6. Diamantis, I., Lambropoulou, S., Przytycki, J.H.: Topological steps toward the HOMFLYPT skein module of the lens spaces L(p,1) via braids. J. Knot Theory Ramif. **25**(14), 1650084 (2016)
7. Dipper, R., James, G.D.: Representations of Hecke algebras of type B_n. J. Algebra **146**, 454–481 (1992)
8. Geck, M., Lambropoulou, S.: Markov traces and knot invariants related to Iwahori-Hecke algebras of type B. J. für die reine und angewandte Mathematik **482**, 191–213 (1997)
9. Häring-Oldenburg, R., Lambropoulou, S.: Knot theory in handlebodies. J. Knot Theory Ramif. **11**(6), 921–943 (2002)
10. Jones, V.F.R.: Hecke algebra representations of braid groups and link polynomials. Ann. Math. **126**, 335–388 (1987)
11. D. Kodokostas, S. Lambropoulou A spanning set and potential basis of the mixed Hecke algebra on two fixed strands, (submitted for publication)
12. Lambropoulou, S.: Solid torus links and Hecke algebras of B-type. In: Yetter, D.N. (ed.) Quantum Topology, pp. 225–245. World Scientific Press, USA (1994)
13. Lambropoulou, S., Rourke, C.P.: Markov's theorem in 3-manifolds. Topol. Appl. **78**, 95–122 (1997)
14. Lambropoulou, S.: Knot theory related to generalized and cyclotomic Hecke algebras of type B. J. Knot Theory Ramif. **8**(5), 621–658 (1999)
15. S. Lambropoulou, Braid structures in knot complements, handlebodies and 3–manifolds. In: Proceedings of Knots in Hellas '98. Series of Knots and Everything, vol. 24, pp. 274–289. World Scientific Press, USA (2000)
16. Lambropoulou, S., Rourke, C.P.: Algebraic Markov equivalence for links in 3-manifolds. Compos. Math. **142**, 1039–1062 (2006)

Chapter 9
Braid Groups in Handlebodies and Corresponding Hecke Algebras

Valeriy G. Bardakov

Abstract In this paper we study the kernel of the homomorphism $B_{g,n} \to B_n$ of the braid group $B_{g,n}$ in the handlebody $\mathscr{H}_g$ to the braid group B_n. We prove that this kernel is semi-direct product of free groups. Also, we introduce an algebra $H_{g,n}(q)$, which is some analog of the Hecke algebra $H_n(q)$, constructed by the braid group B_n.

9.1 Introduction

Let $\mathscr{H}_g$ be a handlebody of genus g. The braid group $B_{g,n}$ on n stings in the handlebody $\mathscr{H}_g$ was introduced by A.B. Sossinsky [1] and independently S. Lambropoulou [2]. Properties of this group are studied by V.V. Vershinin [3, 4] and by S. Lambropolou [2, 5].

The motivation for studying braids from $B_{g,n}$ comes from studying oriented knots and links in knot complements in compact connected oriented 3-manifolds and in handlebodies, since these spaces may be represented by a fixed braid or a fixed integer-framed braid in $\mathbb{S}^3$ [2, 6–8]. Then knots and links in these spaces may be represented by elements of the braid group $B_{g,n}$ or an appropriate cosets of these group [2]. In particular, if M denotes the complements of the g-unlink or a connected sum of g lens spaces of type $L(p, 1)$ or a handlebody of genus g, then knots and links in these spaces may be represented precisely by the braids in $B_{g,n}$ for $n \in \mathbb{N}$. In the case $g = 1$, $B_{1,n}$ is the Artin group of type $\mathscr{B}$.

The group $B_{g,n}$ can be considered as a subgroup of the braid group B_{g+n} on $g + n$ strings such that the braids in $B_{g,n}$ leave the first g strings identically fixed. Using this fact V.V. Vershinin [3] and S. Lambropoulou [2] defined a epimorphism of $B_{g,n}$ onto the symmetric group S_n and prove that the kernel of this epimorphism: $P_{g,n}$ is a subgroup of the pure braid group P_{g+n} and is semi-direct products of free

V.G. Bardakov (✉)
Sobolev Institute of Mathematics, Novosibirsk State University,
Novosibirsk State Agrarian University,
Novosibirsk 630090, Russia
e-mail: bardakov@math.nsc.ru

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_9

groups. On the other side, V.V. Vershinin [3] noted that there is a decomposition $B_{g,n} = R_{g,n} \leftthreetimes B_n$ for some group $R_{g,n}$ and the braid group B_n. S. Lambropoulou proved in [5] that $R_{1,n}$ is isomorphic to a free group F_n of rank n.

The Hecke algebra of type A, $H_n(q)$ was used by V.F.R. Jones [9] for the construction of polynomial invariant for classical links, the well known HOMFLYPT polynomial. S. Lambropoulou [5] defined a generalization of the Hecke algebra of type B to construct a HOMFLYPT-type polynomial invariant for links in the solid torus, which was then used in [10] for extending the study to the lens spaces $L(p, 1)$.

In the present paper we define some decomposition of $P_{g,n}$ as semi-direct product of free groups, which is different from the decomposition defined in [2, 3]. Also, we study the group $R_{g,n}$ and prove that this group is semi-direct products of free groups. Using this decomposition and the decomposition $B_{g,n} = R_{g,n} \leftthreetimes B_n$ we define some algebra, which is a generalization of the Hecke algebra $H_n(q)$.

The paper is organized as follows. In Sect. 2, we remind some facts on the braid group B_n. In particular, we define vertical and horizontal decompositions of the pure braid group P_n. In Sect. 3, we recall some facts on the group $B_{g,n}$, we describe the vertical decomposition of $P_{g,n}$ and we define the horizontal decomposition of $P_{g,n}$. In Sect. 4, we study $R_{g,n}$ and we construct vertical and horizontal decompositions for this group. In Sect. 5 we introduce some algebra $H_{g,n}(q)$ which is a generalization of the Hecke algebra $H_n(q)$ and we find the quotient of $B_{g,n}$ by the relations $\sigma_i^2 = 1$.

9.2 Braid Group

In this section we recall some facts on the braid groups (see [11, 12]).

The braid group B_m, $m \geq 2$, on m strings is generated by elements

$$\sigma_1, \sigma_2, \ldots, \sigma_{m-1},$$

and is defined by relations

$$\begin{aligned} &\sigma_i\sigma_j = \sigma_j\sigma_i, && \text{for } |i-j| > 1, \\ &\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}, && \text{for } i = 1, 2, \ldots, m-2. \end{aligned}$$

A subgroup of B_m which is generated by elements

$$a_{ij} = \sigma_{j-1}\sigma_{j-2}\ldots\sigma_{i+1}\sigma_i^2\sigma_{i+1}^{-1}\ldots\sigma_{j-2}^{-1}\sigma_{j-1}^{-1}, \quad 1 \leq i < j \leq m,$$

is called the *pure braid group* and is denoted P_m. This group is defined by the relations

$$a_{ik}a_{ij}a_{kj} = a_{kj}a_{ik}a_{ij}, \tag{9.1}$$

$$a_{nj}a_{kn}a_{kj} = a_{kj}a_{nj}a_{kn}, \text{ for } n < j, \tag{9.2}$$

$$(a_{kn}a_{kj}a_{kn}^{-1})a_{in} = a_{in}(a_{kn}a_{kj}a_{kn}^{-1}), \text{ for } i < k < n < j, \tag{9.3}$$

$$a_{kj}a_{in} = a_{in}a_{kj}, \text{ for } k < i < n < j \text{ or } n < k. \tag{9.4}$$

The subgroup P_m is normal in B_m, and the quotient B_m/P_m is the symmetric group S_m. The generators of B_m act on the generator $a_{ij} \in P_m$ by the rules:

$$\sigma_k^{-1} a_{ij} \sigma_k = a_{ij}, \text{ for } k \neq i-1, i, j-1, j, \tag{9.5}$$
$$\sigma_i^{-1} a_{i,i+1} \sigma_i = a_{i,i+1}, \tag{9.6}$$
$$\sigma_{i-1}^{-1} a_{ij} \sigma_{i-1} = a_{i-1,j}, \tag{9.7}$$
$$\sigma_i^{-1} a_{ij} \sigma_i = a_{i+1,j}[a_{i,i+1}^{-1}, a_{ij}^{-1}], \text{ for } j \neq i+1 \tag{9.8}$$
$$\sigma_{j-1}^{-1} a_{ij} \sigma_{j-1} = a_{i,j-1}, \tag{9.9}$$
$$\sigma_j^{-1} a_{ij} \sigma_j = a_{ij} a_{i,j+1} a_{ij}^{-1}, \tag{9.10}$$

where $[a, b] = a^{-1}b^{-1}ab$.

Denote by

$$U_i = \langle a_{1i}, a_{2i}, \ldots, a_{i-1,i} \rangle, \quad i = 2, \ldots, m,$$

a subgroup of P_m. It is known that U_i is a free group of rank $i-1$. One can rewrite the relations of P_m as the following conjugation rules (for $\varepsilon = \pm 1$):

$$a_{ik}^{-\varepsilon} a_{kj} a_{ik}^{\varepsilon} = (a_{ij} a_{kj})^{\varepsilon} a_{kj} (a_{ij} a_{kj})^{-\varepsilon}, \tag{9.11}$$
$$a_{kn}^{-\varepsilon} a_{kj} a_{kn}^{\varepsilon} = (a_{kj} a_{nj})^{\varepsilon} a_{kj} (a_{kj} a_{nj})^{-\varepsilon}, \text{ for } n < j, \tag{9.12}$$
$$a_{in}^{-\varepsilon} a_{kj} a_{in}^{\varepsilon} = [a_{ij}^{-\varepsilon}, a_{nj}^{-\varepsilon}]^{\varepsilon} a_{kj} [a_{ij}^{-\varepsilon}, a_{nj}^{-\varepsilon}]^{-\varepsilon}, \text{ for } i < k < n, \tag{9.13}$$
$$a_{in}^{-\varepsilon} a_{kj} a_{in}^{\varepsilon} = a_{kj}, \text{ for } k < i < n < j \text{ or } n < k. \tag{9.14}$$

From these rules it follows that U_m is normal in P_m and hence P_m has the following decomposition: $P_m = U_m \rtimes P_{m-1}$, where the action of P_{m-1} on U_m is define by the rules (9.11)–(9.14). By induction on m, P_m is the semi-direct product of free groups:

$$P_m = U_m \rtimes (U_{m-1} \rtimes (\ldots \rtimes (U_3 \rtimes U_2) \ldots)).$$

We will call this decomposition *vertical decomposition*.

Let $U_m^{(k)}$, $k = 1, 2, \ldots, m$, be the subgroup of P_m which is generated by a_{ij}, where $k < j$. Then $U_m^{(k)} = U_m \rtimes (U_{m-1} \rtimes (\ldots \rtimes (U_{k+2} \rtimes U_{k+1}) \ldots))$. By definition $U_m^{(1)} = P_m$ and these groups form the normal series

$$1 = U_m^{(m)} \leq U_m^{(m-1)} \leq \ldots \leq U_m^{(2)} \leq U_m^{(1)} = P_m,$$

where

$$U_m^{(r)} / U_m^{(r+1)} \cong F_r, \quad r = 1, 2, \ldots, m-1.$$

On the other side, define the following subgroups of P_m:

$$V_k = \langle a_{k,k+1}, a_{k,k+2}, \ldots, a_{k,m} \rangle, \quad k = 1, 2, \ldots, m-1.$$

This group is free of rank $m-k$. Using the defining relation of P_m, it is not difficult to prove the following:

Lemma 9.1 *In P_n hold the following conjugation rules (for $\varepsilon = \pm 1$):*

(1) $a_{kj}^{-\varepsilon} a_{ik} a_{kj}^{\varepsilon} = (a_{ik}a_{ij})^{\varepsilon} a_{ik} (a_{ik}a_{ij})^{-\varepsilon}$, *where* $i < k < j$;
(2) $a_{jk}^{-\varepsilon} a_{ik} a_{jk}^{\varepsilon} = (a_{ij}a_{ik})^{\varepsilon} a_{ik} (a_{ij}a_{ik})^{-\varepsilon}$, *where* $i < j < k$;
(3) $a_{kn}^{-\varepsilon} a_{ij} a_{kn}^{\varepsilon} = [a_{ik}^{-\varepsilon}, a_{in}^{-\varepsilon}]^{\varepsilon} a_{ij} [a_{ik}^{-\varepsilon}, a_{in}^{-\varepsilon}]^{-\varepsilon}$, *where* $i < k < j < n$;
(4) $a_{in}^{-\varepsilon} a_{kj} a_{in}^{\varepsilon} = a_{kj}$, *where* $k < i$ *and* $n < j$;
(5) $a_{kj}^{-\varepsilon} a_{in} a_{kj}^{\varepsilon} = a_{in}$, *where* $n < k$.

From this lema it follows that V_1 is normal in P_m and we have decomposition $P_m = V_1 \leftthreetimes P_{m-1}$. By induction on m, P_m is the semi-direct products of free groups:

$$P_m = V_1 \leftthreetimes (V_2 \leftthreetimes (\ldots \leftthreetimes (V_{m-2} \leftthreetimes V_{m-1}) \ldots)).$$

We will call this decomposition *horizontal decomposition*. Let $V_m^{(k)}$ be a subgroup of P_m which is generated by a_{ij} for $i < k$. Then we have the normal series

$$1 = V_m^{(1)} \leq V_m^{(2)} \leq \cdots \leq V_m^{(m-1)} \leq V_m^{(m)} = P_m,$$

where

$$V_m^{(r)} / V_m^{(r-1)} \cong F_{m-r+1}, \quad r = 2, 3, \ldots, m.$$

A motivation for the terms vertical and horizontal is as follows. If we put the generators of P_m in the following table

$$\begin{array}{cccccc}
a_{12}, & a_{13}, & \ldots & a_{1,m-1}, & a_{1,m}, \\
 & a_{23}, & \ldots & a_{2,m-1}, & a_{2,m}, \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
 & & & a_{m-2,m-1}, & a_{m-2,m}, \\
 & & & & a_{m-1,m},
\end{array}$$

then the generators from the kth row generate V_k and the generators from the kth column generate U_{k+1}. The group $U_m^{(r)}$ is generated by the last $m-r$ columns of this table and the group $V_m^{(r)}$ is generated by the first $r-1$ rows of this table.

9.3 Braid Groups in Handlebodies

Recall some facts on the braid group $B_{g,n}$ on n strings in the handlebody $\mathscr{H}_g$ (see [1–3]). The group $B_{g,n}$ is generated by elements

$$\tau_1, \tau_2, \ldots \tau_g, \sigma_{g+1}, \sigma_{g+2}, \ldots, \sigma_{g+n-1},$$

and is defined by the following list of relations

$$\begin{array}{ll}
\sigma_i\sigma_j = \sigma_j\sigma_i, & \text{for } |i-j|>1,\\
\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}, & \text{for } i = g+1,\dots,g+n-2,\\
\tau_k\sigma_i = \sigma_i\tau_k, & \text{for } k\geq 1,\ \ i\geq g+2,\\
\tau_k(\sigma_{g+1}\tau_k\sigma_{g+1}) = (\sigma_{g+1}\tau_k\sigma_{g+1})\tau_k, & \text{for } k = 1,\dots,g,\\
\tau_k(\sigma_{g+1}^{-1}\tau_{k+l}\sigma_{g+1}) = (\sigma_{g+1}^{-1}\tau_{k+l}\sigma_{g+1})\tau_k, & \text{for } k = 1,\dots,g-1,\ \ l = 1,\dots,g-k.
\end{array}$$

The group $B_{g,n}$ can be considered as a subgroup of the classical braid group B_{g+n} on $n+g$ strings such that the braids from $B_{g,n}$ leave the first g strings unbraided. Then $\tau_k = a_{k,g+1}$, in B_{g+n}, i.e.

$$\tau_k = \sigma_g\sigma_{g-1}\dots\sigma_{k+1}\sigma_k^2\sigma_{k+1}^{-1}\dots\sigma_{g-1}^{-1}\sigma_g^{-1},\quad k = 1,2,\dots,g.$$

The elements τ_k, $k = 1,2,\dots,g$, generate a free group of rank g which is isomorphic to $U_{g+1} = \langle a_{1,g+1}, a_{2,g+1},\dots,a_{g,g+1}\rangle$ in B_{g+n}. Also, we see that some other generators of P_{g+n} lie in $B_{g,n}$. Put them in the table bellow:

$$\begin{array}{ccccc}
a_{1,g+1}, & a_{1,g+2}, & \dots & a_{1,g+n-1}, & a_{1,g+n},\\
a_{2,g+1}, & a_{2,g+2}, & \dots & a_{2,g+n-1}, & a_{2,g+n},\\
\dots & \dots & \dots & \dots & \dots\\
a_{g-1,g+1}, & a_{g-1,g+2}, & \dots & a_{g-1,g+n-1}, & a_{g-1,g+n},\\
a_{g,g+1}, & a_{g,g+2}, & \dots & a_{g,g+n-1}, & a_{g,g+n},\\
\hline
 & a_{g+1,g+2}, & \dots & a_{g+1,g+n-1}, & a_{g+1,g+n},\\
 & \dots & \dots & \dots & \dots\\
 & & & a_{g+n-2,g+n-1}, & a_{g+n-2,g+n},\\
 & & & & a_{g+n-1,g+n}.
\end{array}$$

We will denote by $\widetilde{B}_n$ the subgroup of $B_{g,n}$ which is generated by σ_{g+1}, σ_{g+2}, $\dots$, σ_{g+n-1}. It is evident that $\widetilde{B}_n$ is isomorphic to B_n. The corresponding pure braid group of $\widetilde{B}_n$ will be denote $\widetilde{P}_n$. This group is isomorphic to P_n and is generated by elements from the above table which lie under the horizontal line. The group $\widetilde{P}_n$ has the following vertical decomposition

$$\widetilde{P}_n = \widetilde{U}_n \leftthreetimes (\widetilde{U}_{n-1} \leftthreetimes (\dots \leftthreetimes (\widetilde{U}_3 \leftthreetimes \widetilde{U}_2)\dots)),$$

where

$$\widetilde{U}_i = \langle a_{g+1,g+i}, a_{g+2,g+i},\dots,a_{g+i-1,g+i}\rangle,\quad i = 2,3,\dots,n.$$

There is an epimorphism:

$$\psi_n : B_{g,n} \longrightarrow S_n,$$

which is defined by the rule

$$\psi_n(\tau_k) = 1,\ k = 1,2,\dots,g,\quad \psi_n(\sigma_i) = (i,i+1),\ i = g+1,g+2,\dots,g+n-1.$$

This epimorphism is induced by the standard epimorphism $B_{g+n} \longrightarrow S_{g+n}$. Let $P_{g,n} = \mathrm{Ker}(\psi_n)$. Then $P_{g,n}$ is generated by the element from the table above. In [2, 3] it was proved that there exists the following short exact sequence:

$$1 \longrightarrow P_{g,n} \longrightarrow B_{g,n} \longrightarrow S_n \longrightarrow 1,$$

and was found the vertical decomposition of $P_{g,n}$:

$$P_{g,n} = U_{g+n} \leftthreetimes (U_{g+n-1} \leftthreetimes (\ldots \leftthreetimes (U_{g+2} \leftthreetimes U_{g+1}) \ldots)).$$

If we let $U_{g+n}^{(g+n-i)} = U_{g+n} \leftthreetimes (U_{g+n-1} \leftthreetimes (\ldots \leftthreetimes (U_{g+n-i+2} \leftthreetimes U_{g+n-i+1}) \ldots))$, then we get the normal series:

$$1 = U_{g+n}^{(g+n)} \leq U_{g+n}^{(g+n-1)} \leq \ldots \leq U_{g+n}^{(g)} = P_{g,n},$$

where

$$U_{g+n}^{(r)} / U_{g+n}^{(r+1)} \cong F_r, \quad r = g, g+1, \ldots, g+n-1.$$

The homomorphism which is induced by the embedding $B_{g,n} \longrightarrow B_{g+n}$ sends this normal series to the corresponding normal series for P_{g+n}.

Let us construct the horizontal decomposition for $P_{g,n}$. To do this, define the following subgroups in $P_{g,n}$:

$$\begin{aligned}
&V_{g,1} = \langle a_{1,g+1}, a_{1,g+2}, \ldots, a_{1,g+n} \rangle, \\
&V_{g,2} = \langle a_{2,g+1}, a_{2,g+2}, \ldots, a_{2,g+n} \rangle, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&V_{g,g} = \langle a_{g,g+1}, a_{g,g+2}, \ldots, a_{g,g+n} \rangle, \\
&V_{g,g+1} = \langle a_{g+1,g+2}, a_{g+1,g+3}, \ldots, a_{g+1,g+n} \rangle, \\
&V_{g,g+2} = \langle a_{g+2,g+3}, a_{g+2,g+4}, \ldots, a_{g+2,g+n} \rangle, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&V_{g,g+n-1} = \langle a_{g+n-1,g+n} \rangle.
\end{aligned}$$

We see that $V_{g,g+i} = V_{g+i}$ for all $i = 1, 2, \ldots, n-1$, these subgroups lie in $\widetilde{P}_n$ and as we know

$$\widetilde{P}_n = V_{g,g+1} \leftthreetimes (V_{g,g+2} \leftthreetimes (\ldots \leftthreetimes (V_{g,g+n-2} \leftthreetimes V_{g,g+n-1}) \ldots))$$

is the horizontal decomposition of $\widetilde{P}_n$.

We see that the vertical decomposition of $P_{g,n}$ is a part of the vertical decomposition for P_{g+n}. For the horizontal decomposition situation is more complicated.

Example 9.1 The horizontal decomposition of P_4 has the form

$$P_4 = V_1 \leftthreetimes (V_2 \leftthreetimes V_3),$$

where

$$V_1 = \langle a_{12}, a_{13}, a_{14}\rangle, \quad V_2 = \langle a_{23}, a_{24}\rangle, \quad V_3 = \langle a_{34}\rangle,$$

and V_1 is normal in P_4, V_2 is normal in $V_2 \leftthreetimes V_3$. The group $P_{2,2}$ contains subgroups

$$V_{2,1} = \langle a_{13}, a_{14}\rangle, \quad V_{2,2} = V_2 = \langle a_{23}, a_{24}\rangle, \quad V_{2,3} = V_3 = \langle a_{34}\rangle,$$

but in this case $V_{2,1}$ is not normal in $P_{2,2}$, Indeed, from Lemma 9.1 we have the following relations

$$a_{23}^{-\varepsilon} a_{13} a_{23}^{\varepsilon} = (a_{12}a_{13})^{\varepsilon} a_{13} (a_{12}a_{13})^{-\varepsilon}, \quad a_{24}^{-\varepsilon} a_{14} a_{24}^{\varepsilon} = (a_{12}a_{14})^{\varepsilon} a_{14} (a_{12}a_{14})^{-\varepsilon}.$$

Hence, $P_{2,2}$ contains not only $V_{2,1}$ but its normal closure in $P_{2,2}$ and we get the horizontal decomposition

$$P_{2,2} = \overline{V}_{2,1} \leftthreetimes (V_2 \leftthreetimes V_3),$$

where $\overline{V}_{2,1} = \langle V_{2,1}\rangle^{P_{2,2}}$ is the normal closure of $V_{2,1}$ in $P_{2,2}$.

In the general case, let $\overline{V}_{g,i}$ be the normal closure of $V_{g,i}$ in the subgroup $\langle V_{g,i}, V_{g,i+1}, \ldots, V_{g,g}, \widetilde{P}_n\rangle$, i.e.

$$\overline{V}_{g,i} = \langle V_{g,i}\rangle^{\langle V_{g,i}, V_{g,i+1}, \ldots, V_{g,g}, \widetilde{P}_n\rangle}.$$

Lemma 9.2 *(1)* $\overline{V}_{g,g} = V_{g,g} \cong F_n$.

(2) $\overline{V}_{g,i}$ *is a subgroup of* V_i *for every* $i = 1, 2, \ldots, g$ *and, in particular, is a free group.*

Proof (1) We see that $V_{g,g} = V_g$ and from Lemma 9.1 it follows that V_g is normal in $\langle V_{g,g}, \widetilde{P}_n\rangle$.

(2) The fact that $\overline{V}_{g,i}$ is a subgroup of V_i follows from the conjugation rules of Lemma 9.1. The fact that $\overline{V}_{g,i}$ is free follows from the fact that V_i is free.

Since $\widetilde{P}_n \cong P_n$, it has the horizontal decomposition:

$$\widetilde{P}_n = V_{g+1} \leftthreetimes (V_{g+2} \leftthreetimes (\ldots \leftthreetimes (V_{g+n-2} \leftthreetimes V_{g+n-1}) \ldots)).$$

Using Lemma 9.2, one can construct the horizontal decomposition of $P_{g,n}$.

Theorem 9.1 *The group* $P_{g,n}$ *is the semi-direct products of groups:*

$$P_{g,n} = \overline{V}_{g,1} \leftthreetimes (\overline{V}_{g,2} \leftthreetimes (\ldots \leftthreetimes (\overline{V}_{g,g} \leftthreetimes \widetilde{P}_n) \ldots)).$$

Proof Use induction on g. If $g = 1$, then $P_{1,n} = P_{n+1}$ and the horizontal decomposition for P_{n+1} gives the horizontal decomposition: $P_{1,n} = V_g \leftthreetimes \widetilde{P}_n$. As follows from Lemma 9.2, $V_g = \overline{V}_{g,g}$.

Let $g > 1$. Define a homomorphism of $P_{g,n}$ onto the group $P_{g-1,n}$ which sends all generators of $V_{g,1}$ to the unit and keeps all other generators. The kernel of this homomorphism is the normal closure of $V_{g,1}$ in $P_{g,n}$. Denote this kernel by $\overline{V}_{g,1} = \langle V_{g,1} \rangle^{P_{g,n}}$. Since, $V_{g,1}$ is a subgroup in V_1 and V_1 is normal in P_{g+n} we get that $\overline{V}_{g,1}$ lies in V_1 and hence is a free. We have the decomposition $P_{g,n} = \overline{V}_{g,1} \leftthreetimes P_{g-1,n}$. Using the induction hypothesis we get the required decomposition.

9.4 The Kernel of the Epimorphism $B_{g,n} \to \widetilde{B}_n$

As was noted in [3] there is an epimorphism

$$\varphi_n : B_{g,n} \longrightarrow \widetilde{B}_n,$$

where the subgroup $\widetilde{B}_n = \langle \sigma_{g+1}, \sigma_{g+2}, \ldots, \sigma_{g+n-1} \rangle$ is isomorphic to B_n. This endomorphism is defined by the rule

$$\varphi_n(\tau_k) = 1,\ k = 1, 2, \ldots, g, \quad \varphi_n(\sigma_i) = \sigma_i,\ i = g+1, g+2, \ldots, g+n-1.$$

If we denote $R_{g,n} = \mathrm{Ker}(\varphi_n)$, then $B_{g,n} = R_{g,n} \leftthreetimes \widetilde{B}_n$. The purpose of this section is the description of the group $R_{g,n}$.

Considering the table with the generators of $P_{g,n}$ (see Sect. 3), we see that all generators, which lie in the fist g rows of this table, are elements of $R_{g,n}$. Denote by $Q_{g,n}$ the subgroup of $R_{g,n}$ that is generated by these elements, i.e. $Q_{g,n} = \langle V_{g,1}, V_{g,2}, \ldots, V_{g,g} \rangle$. Then $R_{g,n} = \langle Q_{g,n} \rangle^{B_{g,n}}$ is the normal closure of $Q_{g,n}$ in $B_{g,n}$.

On the other side, if we denote

$$U_{g,g+i} = \langle a_{1,g+i}, a_{2,g+i}, \ldots, a_{g,g+i} \rangle \leq U_{g+i}, \quad i = 1, 2, \ldots, n,$$

then $Q_{g,n} = \langle U_{g,g+1}, U_{g,g+2}, \ldots, U_{g,g+n} \rangle$. Note that $U_{g,g+1} = U_{g+1}$. Let $\overline{U}_{g,g+i}$ be the normal closure of $U_{g,g+i}$ in U_{g+i}, $i = 1, 2, \ldots, n$. In this notations it holds:

Theorem 9.2 *The group $R_{g,n}$ has the following decompositions:*

(1) $R_{g,n} = \overline{U}_{g,g+n} \leftthreetimes (\overline{U}_{g,g+n-1} \leftthreetimes (\ldots \leftthreetimes (\overline{U}_{g,g+2} \leftthreetimes \overline{U}_{g,g+1}) \ldots))$,
where $\overline{U}_{g,g+1} = U_{g+1}$.

(2) $R_{g,n} = \overline{V}_{g,1} \leftthreetimes (\overline{V}_{g,2} \leftthreetimes (\ldots \leftthreetimes (\overline{V}_{g,g-1} \leftthreetimes \overline{V}_{g,g}) \ldots))$,
where $\overline{V}_{g,g} = V_g$.

Proof As we know, $R_{g,n}$ is the normal closure of $Q_{g,n}$ in $B_{g,n}$. To find this closure, at first, consider conjugations of the generators of $Q_{g,n}$ by σ_{g+k}, $k = 1, 2, \ldots, n-1$.

Conjugating generators from the kth column of our table by σ_{g+k}, we have

$$\sigma_{g+k}^{-1} a_{i,g+k} \sigma_{g+k} = a_{i,g+k} a_{i,g+k+1} a_{i,g+k}^{-1}, \quad i = 1, 2, \ldots, g.$$

Conjugating generators from the $(k+1)$st column of our table by σ_{g+k}, we have

$$\sigma_{g+k}^{-1} a_{i,g+k+1} \sigma_{g+k} = a_{i,g+k}, \quad i = 1, 2, \ldots, g.$$

Generators from all other columns commute with σ_{g+k}.

Hence, for every generator a_{ij} of $Q_{g,n}$ and every σ_k, $k = g+1, g+2, \ldots, g+n-1$, the element $\sigma_k^{-1} a_{ij} \sigma_k$ lies in $Q_{g,n}$.

For the group $\widetilde{B}_n$ there exists the following exact sequence

$$1 \longrightarrow \widetilde{P}_n \longrightarrow \widetilde{B}_n \longrightarrow S_n \longrightarrow 1.$$

Let $m_{kl} = \sigma_{k-1}\sigma_{k-2}\ldots\sigma_l$ for $l < k$ and $m_{kl} = 1$ in other cases. Then the set

$$\Lambda_n = \left\{ \prod_{k=g+2}^{g+n} m_{k,j_k} \mid 1 \le j_k \le k \right\}$$

is a Schreier set of coset representatives of $\widetilde{P}_n$ in $\widetilde{B}_n$.

From the previous observations it follows that if $\alpha \in \Lambda$, then for every generator a_{ij} of $Q_{g,n}$ the element $\alpha^{-1} a_{ij} \alpha$ lies in $Q_{g,n}$.

Now we will consider the conjugations of the generators of $Q_{g,n}$ by the generators of $P_{g,n}$.

(1) To prove the first decomposition, take some element $h \in B_{g,n}$ and using the vertical decomposition, write it in the normal form:

$$h = u_{g+1} u_{g+2} \ldots u_{g+n} \alpha, \text{ where } u_k \in U_k, \quad \alpha \in \Lambda_n.$$

In every group U_k, $k = g+1, g+2, \ldots, g+n$, we have two subgroups:

$$U_{g,k} = \langle a_{1k}, a_{2k}, \ldots, a_{gk} \rangle, \quad \widetilde{U}_k = \langle a_{g+1,k}, a_{g+2,k}, \ldots, a_{k-1,k} \rangle \le \widetilde{P}_n.$$

We see that $U_{g,k}$ is a subgroup of $P_{g,n}$, $\widetilde{U}_k$ is a subgroup of $\widetilde{P}_n$ and $U_k = \langle U_{g,k}, \widetilde{U}_k \rangle$ is a free group. Define the projection $\pi_k : U_k \longrightarrow \widetilde{U}_k$ by the rules

$$\pi_k(a_{ik}) = 1, \quad i = 1, 2, \ldots, g;$$

$$\pi_k(a_{jk}) = a_{jk}, \quad j = g+1, g+2, \ldots, k-1.$$

The kernel $\mathrm{Ker}(\pi_k) = \overline{U}_{g,k}$ is the normal closure $\langle U_{g,k} \rangle^{U_k}$ of $U_{g,k}$ into U_k. Hence, element u_k can be written in the form $u_k = \overline{u}_k \pi_k(u_k)$ for some $\overline{u}_k \in \overline{U}_{g,k}$. Denote for simplicity $w_{g+i} = \pi_{g+1}(u_{g+i})$, we can rewrite h in the form

$$h = (\overline{u}_{g+1} w_{g+1})(\overline{u}_{g+2} w_{g+2}) \ldots (\overline{u}_{g+n} w_{g+n}) \alpha.$$

Shifting all w_{g+i} to the right we get

$$h = \overline{u}_{g+1}\overline{u}_{g+2}^{w_{g+1}^{-1}} \dots \overline{u}_{g+n}^{w_{g+n-1}^{-1}w_{g+n-2}^{-1}\dots w_{g+1}^{-1}} w_{g+1}w_{g+2}\dots w_{g+n}\alpha.$$

The image $\varphi_n(h) = w_{g+1}w_{g+2}\dots w_{g+n}\alpha$. Hence, the element

$$\overline{u}_{g+1}\,\overline{u}_{g+2}^{w_{g+1}^{-1}} \;\dots\; \overline{u}_{g+n}^{w_{g+n-1}^{-1}w_{g+n-2}^{-1}\dots w_{g+1}^{-1}}$$

lies in the kernel $\mathrm{Ker}(\varphi_n) = R_{g,n}$. We proved that any element from $R_{g,n}$ lies in the product

$$\overline{U}_{g,g+n} \leftthreetimes (\overline{U}_{g,g+n-1} \leftthreetimes (\dots \leftthreetimes (\overline{U}_{g,g+2} \leftthreetimes \overline{U}_{g,g+1})\dots)).$$

On the other side, this product contains $Q_{g,n}$ and is normal in $B_{g,n}$. Hence, $R_{g,n}$ has the required decomposition.

(2) Prove the second decomposition. Denote $G = \overline{V}_{g,1} \leftthreetimes (\overline{V}_{g,2} \leftthreetimes (\dots \leftthreetimes (\overline{V}_{g,g-1} \leftthreetimes \overline{V}_{g,g})\dots))$ the group from the right side of the decomposition of $R_{g,n}$. Take an arbitrary element $h \in B_{g,n}$. By the theorem on the horizontal decomposition of $B_{g,n}$ it has the following normal form:

$$h = v_1v_2\dots v_g b, \text{ where } v_i \in V_{g,i}, b \in \widetilde{B}_n$$

and $v_1v_2\dots v_g$ lies in $R_{g,n}$. Under the endomorphism $\varphi_n : B_{g,n} \longrightarrow \widetilde{B}_n$ element h goes to the element b. Hence, G lies in $R_{g,n}$.

On the other side we must prove that $R_{g,n}$ lies in G. We know that $R_{g,n}$ is the normal closure of $Q_{g,n}$ in $B_{g,n}$. As was shown before, if $\alpha \in \Lambda_n$ is a coset representative of $\widetilde{P}_n$ into $\widetilde{B}_n$ and a_{ij} is some generator of $Q_{g,n}$, then a_{ij}^{α} lies in $Q_{g,n}$. Considering the conjugation of $a_{ij} \in Q_{g,n}$ by generators of $P_{g,n}$ and using the Lemma 9.1 we get an element which lies in G.

In the case $g = 1$ we have $P_{1,n} = P_{n+1}$ and $R_{1,n} = \langle a_{12}, a_{13}, \dots, a_{1,n+1}\rangle \cong F_n$ and we get the decomposition which was found in [5].

Corollary 9.1 $B_{1,n} \cong F_n \leftthreetimes B_n$.

9.5 Some Analog of the Hecke Algebra for the Braid Group in the Handlebody

Let q be some complex number. Recall that the Hecke algebra $H_n(q)$ is an associative $\mathbb{C}$-algebra with unit, which is generated by elements

$$s_1, s_2, \dots, s_{n-1},$$

and is defined by the relations

$$\begin{array}{ll} s_i s_j = s_j s_i, & \text{for } |i-j|>1, \\ s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1}, & \text{for } 1 \le i \le n-2, \\ s_i^2 = (q-1)s_i + q, & \text{for } i = 1, \ldots, n-1. \end{array}$$

The algebra $H_n(q)$ has the following linear basis:

$$S = \{(s_{i_1} s_{i_1-1} \ldots s_{i_1-k_1})(s_{i_2} s_{i_2-1} \ldots s_{i_2-k_2}) \ldots (s_{i_p} s_{i_p-1} \ldots s_{i_p-k_p})\}$$

for $1 \le i_1 < \cdots < i_p \le n-1$. The basis S is used in the construction of the Markov trace, leading to the HOMFLYPT or 2-variable Jones polynomial (see [9]).

The braid group in the solid torus $B_{1,n}$ is the Artin group of the Coxeter group of type $\mathscr{B}$, which is related to the Hecke algebra of type $\mathscr{B}$. The generalized Hecke algebra of type $\mathscr{B}$, $H_{1,n}(q)$ is defined by S. Lambropoulou in [13]. $H_{1,n}(q)$ is isomorphic to the affine Hecke algebra of type $\mathscr{A}$, $\widetilde{H}_n(q)$. A unique Markov trace is constructed on the algebras $H_{1,n}(q)$ that leads to an invariant for links in the solid torus, the universal analogue of the HOMFLYPT polynomial for the solid torus.

The algebra $H_{1,n}(q)$ is generated by elements

$$t, t^{-1}, s_1, s_2, \ldots, s_{n-1},$$

and is defined by the relations

$$\begin{array}{ll} s_i s_j = s_j s_i, & \text{for } |i-j|>1, \\ s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1}, & \text{for } 1 \le i \le n-2, \\ s_i^2 = (q-1)s_i + q, & \text{for } i = 1, 2, \ldots, n-1, \\ s_1 t s_1 t = t s_1 t s_1, & \\ t s_i = s_i t, & \text{for } i > 1. \end{array}$$

Hence,

$$H_{1,n}(q) = \frac{\mathbb{C}[B_{1,n}]}{\langle \sigma_i^2 - (q-1)\sigma_i - q \rangle}.$$

Note that in $H_{1,n}(q)$ the generator t satisfies no polynomial relation, making the algebra $H_{1,n}(q)$ infinite dimensional. If we set $t = 0$ in $H_{1,n}(q)$, we obtain the Hecke algebra $H_n(q)$.

In $H_{1,n}(q)$ are defined in [5] the elements

$$t_i = s_i s_{i-1} \ldots s_1 t s_1 \ldots s_{i-1} s_i, \quad t_i' = s_i s_{i-1} \ldots s_1 t s_1^{-1} \ldots s_{i-1}^{-1} s_i^{-1}.$$

It was then proved that the following sets form linear bases for $H_{1,n}(q)$:

$$\Sigma_n = t_{i_1}^{k_1} t_{i_2}^{k_2} \ldots t_{i_r}^{k_r} \sigma, \text{ where } 1 \le i_1 < \ldots < i_r \le n-1,$$

$$\Sigma'_n = (t'_{i_1})^{k_1}(t'_{i_2})^{k_2}\dots(t'_{i_r})^{k_r}\sigma, \text{ where } 1 \le i_1 < \dots < i_r \le n-1,$$

where $k_1, k_2, \dots, k_r \in \mathbb{Z}$ and σ a basis element in $H_n(q)$. The basis Σ'_n is used in [5, 13] for constructing a Markov trace on $\bigcup_{n=1}^{\infty} H_{1,n}(q)$ and a universal HOMFLYPT-type invariant for oriented links in the solid torus.

Definition 9.1 Let $q \in \mathbb{C}$. The algebra $H_{g,n}(q)$ is an associative algebra over $\mathbb{C}$ with unit that is generated by

$$t_1^{\pm 1}, t_2^{\pm 1}, \dots, t_g^{\pm 1}, s_{g+1}, s_{g+2}, \dots s_{g+n-1}$$

and is defined by the following relations

$$\begin{array}{ll}
s_i s_j = s_j s_i, & \text{for } |i-j| > 1,\\
s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1}, & \text{for } i = g+1, g+2, \dots, g+n-2,\\
s_i^2 = (q-1)s_i + q, & \text{for } i = g+1, g+2, \dots, g+n-1,\\
t_k s_i = s_i t_k, & \text{for } k \ge 1, i \ge g+2,\\
t_k(s_{g+1} t_k s_{g+1}) = (s_{g+1} t_k s_{g+1}) t_k, & \text{for } k = 1, 2, \dots, g,\\
t_k(s_{g+1} t_{k+l} s_{g+1}) = (s_{g+1} t_{k+l} s_{g+1}) t_k, & \text{for } k = 1, 2, \dots, g-1; l = 1, 2, \dots, g-k.
\end{array}$$

Remark 9.1 More natural to consider the generators $s_1, s_2, \dots s_{n-1}$ instead $s_{g+1}, s_{g+2}, \dots s_{g+n-1}$, but for technical reasons we will use our notation.

We see that

$$H_{g,n}(q) = \frac{\mathbb{C}[B_{g,n}]}{\langle \sigma_i^2 - (q-1)\sigma_i - q \rangle}.$$

If we consider the algebra $H_n(q)$ as a vector space over $\mathbb{C}$, then it is isomorphic to the vector space $\mathbb{C}[S_n]$. Thus to study $H_{g,n}(q)$, we define a group $G_{g,n}$ as the quotient of $B_{g,n}$ by the relations

$$\sigma_i^2 = 1, \quad i = 1, 2, \dots, n-1.$$

Denote the natural homomorphism $B_{g,n} \longrightarrow G_{g,n}$ by ψ and let $\psi(a_{ij}) = b_{ij}$.

Theorem 9.3 *The group $G_{g,n}$ is the semi-direct product $G_{g,n} = F_g^n \leftthreetimes S_n$ of the direct product of n copies of free group F_g of rank g and the symmetric group S_n.*

Proof We know that $B_{g,n} = R_{g,n} \leftthreetimes \widetilde{B}_n$ and

$$R_{g,n} = \overline{U}_{g,g+n} \leftthreetimes (\overline{U}_{g,g+n-1} \leftthreetimes (\dots \leftthreetimes (\overline{U}_{g,g+2} \leftthreetimes \overline{U}_{g,g+1}) \dots)).$$

On the other side, U_k, $k = g+1, g+2, \dots, g+n$, is generated by two subgroups:

$$U_{g,k} = \langle a_{1k}, a_{2k}, \dots, a_{gk} \rangle \text{ and } \widetilde{U}_k = \langle a_{g+1,k}, a_{g+2,k}, \dots, a_{k-1,k} \rangle \le \widetilde{P}_n.$$

Note that under the homomorphism ψ the subgroups $\widetilde{U}_k$ go to 1. Hence, $\psi(\overline{U}_{g,g+k}) = \psi(U_{g,g+k}) \cong F_g$ and

$$\psi(R_{g,n}) = \psi(U_{g,g+n}) \leftthreetimes (\psi(U_{g,g+n-1}) \leftthreetimes (\ldots \leftthreetimes (\psi(U_{g,g+2}) \leftthreetimes \psi(U_{g,g+1})) \ldots))$$

is the semi-direct product of n copies of free group F_g. Also, $G_{g,n} = \psi(R_{g,n}) \leftthreetimes \psi(\widetilde{B}_n) \cong \psi(R_{g,n}) \leftthreetimes S_n$.

Let us show that in fact $\psi(R_{g,n})$ is the direct product of n copies of free group F_g. To do it it is enough to prove that

$$\psi(R_{g,n}) = \psi(U_{g,g+n}) \times \psi(R_{g,n-1}).$$

Denote $b_{ij} = \psi(a_{ij})$, consider the relations (9.11)–(9.14) and find their images under the homomorphism ψ. The relation (9.11) goes to the relation

$$b_{i,g+j}^{-\varepsilon} b_{g+j,g+n} b_{i,g+j}^{\varepsilon} = (b_{i,g+n} b_{g+j,g+n})^{\varepsilon} b_{g+j,g+n} (b_{i,g+n} b_{g+j,g+n})^{-\varepsilon},$$

but $b_{g+j,g+n} = 1$ and we have the trivial relations.

The relation (9.12) goes to the relation

$$b_{l,g+j}^{-\varepsilon} b_{l,g+n} b_{l,g+j}^{\varepsilon} = (b_{l,g+n} b_{g+j,g+n})^{\varepsilon} b_{l,g+n} (b_{l,g+n} b_{g+j,g+n})^{-\varepsilon},$$

but $b_{g+j,g+n} = 1$ and we have the relation

$$b_{l,g+j}^{-\varepsilon} b_{l,g+n} b_{l,g+j}^{\varepsilon} = b_{l,g+n}.$$

The relation (9.13) goes to the relation

$$b_{i,g+j}^{-\varepsilon} b_{l,g+n} b_{i,g+j}^{\varepsilon} = [b_{i,g+n}^{-\varepsilon}, b_{g+j,g+n}^{-\varepsilon}]^{\varepsilon} b_{l,g+n} [b_{i,g+n}^{-\varepsilon}, b_{g+j,g+n}^{-\varepsilon}]^{-\varepsilon}, \quad i < l < g + j,$$

but $b_{g+j,g+n} = 1$ and we have the relation

$$b_{i,g+j}^{-\varepsilon} b_{l,g+n} b_{i,g+j}^{\varepsilon} = b_{l,g+n}, \quad i < l < g + j.$$

The relation (9.14) goes to the relation

$$b_{i,g+j}^{-\varepsilon} b_{l,g+n} b_{i,g+j}^{\varepsilon} = b_{l,g+n}, \quad l < i < g + j < g + n.$$

The image of $\widetilde{B}_n \leq B_{g,n}$ is isomorphic to S_n. Thus from the decomposition $B_{g,n} = R_{g,n} \leftthreetimes \widetilde{B}_n$ we get the required decomposition for $G_{g,n}$.

From this theorem we have

Corollary 9.2 *Every element $h \in G_{g,n}$ has the unique normal form*

$$h = h_1 h_2 \ldots h_n \alpha,$$

where h_i is a reduced word in the free group

$$\psi(U_{g,g+i}) = \langle b_{1,g+i}, b_{2,g+i}, \ldots, b_{g,g+i} \rangle \cong F_g, \quad i = 1, 2, \ldots, n,$$

$\alpha \in \Lambda_n$ is a coset representative of $\psi(R_{g,n})$ in $G_{g,n}$.

Using this normal form and the fact that these normal forms form a linear basis of $\mathbb{C}[G_{g,n}]$, we can try to define a basis of $H_{g,n}(q)$. In the algebra $H_{g,n}(q)$ define the following elements

$$\begin{array}{llll} t_{1,g+1} = t_1, & t_{1,g+2}, & \ldots, & t_{1,g+n}, \\ t_{2,g+1} = t_2, & t_{2,g+2}, & \ldots, & t_{2,g+n}, \\ \ldots & \ldots & \ldots & \ldots \\ t_{g,g+1} = t_g, & t_{g,g+2}, & \ldots, & t_{g,g+n}, \end{array}$$

where

$$t_{ij} = s_{j-1}s_{j-2}\ldots s_1 t_{i,g+1} s_1 \ldots s_{j-2}s_{j-1}, \quad 1 \le i \le g, \quad g+2 \le j \le g+n.$$

These elements correspond to the elements b_{ij} from $G_{g,n}$, which were defined in Theorem 9.3 as the images of the elements a_{ij} under the map ψ. It is not difficult to see that the elements

$$t_{1,g+i}, t_{2,g+i}, \ldots, t_{g,g+i}$$

generate a free group of rank g.

The algebra $H_{g,n}(q)$ contains the subalgebra with the set of generators s_{g+1}, s_{g+2}, $\ldots s_{g+n-1}$, which is isomorphic to $H_n(q)$. Let $\Sigma_{g,n}$ be the following set in $H_{g,n}(q)$:

$$u_1 u_2 \ldots u_n \sigma,$$

where u_i is a reduced word in the free group

$$\langle t_{1,g+i}, t_{2,g+i}, \ldots, t_{g,g+i} \rangle \cong F_g, \quad i = 1, 2, \ldots, n,$$

σ is a basis element in $H_n(q)$.

Conjecture 9.1 *There is an isomorphism $H_{g,n}(q) \cong \mathbb{C}[G_{g,n}]$ as $\mathbb{C}$-modules. In particular, the set $\Sigma_{g,n}$ is a basis of the algebra $H_{g,n}(q)$.*

The algebra $H_{2,n}(q)$ is the subject of study in [14], where one set of elements in $H_{2,n}(q)$, different from our $\Sigma_{2,n}$, is proved to be a spanning set for the algebra.

At the end, we formulate the following question for further investigation.

Question 9.1 *Is it possible to define a Markov trace on the algebra $\bigcup_{n=1}^{\infty} H_{g,n}(q)$ and construct some analogue of the HOMLYPT polynomial that is an invariant of links in the handlebody $\mathscr{H}_g$?*

Acknowledgements The author gratefully acknowledges Prof. Lambropoulou and her students and colleagues: Neslihan Gügümcü, Stathis Antoniou, Dimos Goundaroulis, Ioannis Diamantis, Dimitrios Kodokostas for the kind invitation to the Athens, where this paper was written, for conversation and interesting discussions.
This work was supported by the Russian Foundation for Basic Research (project 16-01-00414).

References

1. Sossinsky, A.B.: Preparation theorem for isotopy invariants of links in 3-manifold. Quantum Groups. In: Proceedings of the Conference on Quantum Groups. Lecture Notes in Mathematics N 1510, pp. 354–362. Springer, Berlin (1992)
2. Lambropoulou, S.: Braid structure in knot complements, handlebodies and 3-manifold. Knots In: Hellas (Delphi) Proceedings of International Conference Knot Theory and its Ramifications, WS, 274–289 (1998)
3. Vershinin, V.V.: On braid groups in handlebodies. Sib. Math. J. **39**, N 4, 645–654 (1998). Translation from Sibirisk. Mat. Zh. **39**, N 4, 755–764 (1998)
4. Vershinin, V.V.: Generalization of braids from a homological point of view. Sib. Adv. Math. **9**(2), 109–139 (1999)
5. Lambropoulou, S.: Knot theory related to generalized and cyclotomic Hecke algebras of type B. J. Knot Theory Ramif. **8**(5): 621–658 (1999)
6. Haring-Oldenburg, R., Lambropoulou, S.: Knot theory in handlebodies. J. Knot Theory Ramif. **11**(6), 921–943 (2002)
7. Lambropoulou, S., Rourke, C.P.: Markov's theorem in 3-manifolds. Topol. Appl. **78**, 95–122 (1997)
8. Lambropoulou, S., Rourke, C.P.: Algebraic Markov equivalence for links in 3-manifolds. Compos. Math. **142**, 1039–1062 (2006)
9. Jones, V.F.R.: Hecke algebra representations of braid groups and links polynomials. Ann. Math. **126**, 335–388 (1984)
10. Diamantis, J., Lambropoulou, S., Przytycki, J.: Topological steps toward the HOMFLYPT skein module of the lens spaces $L(p, 1)$ via braids. J. Knot Theory Ramif. **25**(14), 1650084 (2016)
11. Birman, J.S.: Braids, Links, and Mapping Class Groups. Ann. Math. Studies, vol. 82, Princeton University Press, USA(1974)
12. Markov, A.A.: Foundations of the algebraic theory of braids. Tr. Mat. Inst. Steklov [Proc. Steklov Inst. Math.]. **16**, 1–54 (1945)
13. Lambropoulou, S.: Solid torus links and Hecke algebras of B-type. In: Yetter D.N. (ed.) Proceedings of the Conference Quantum Topology World Scientific Press (1994)
14. Kodokostas, D., Lambropoulou, S.: A spanning set and potential basis of the mixed Hecke algebra on two fixed strands. arXiv:1704.03676

Chapter 10
Infinite Loop Spaces, Dyer–Lashof Algebra, Cohomology of the Infinite Symmetric Group and Modular Invariants

Nondas E. Kechagias

Abstract In this lecture note we survey results obtained under the research program Thalis (Kechagias, J. Homotopy Relat. Struct. 8(2), 201–229, (2013), [24], J. Pure Appl. Algebra 219(4), 839–863, (2015), [25]) and place them in the context of algebraic topology. It is divided into two parts. In the first part, we survey infinite loop spaces, Ω-spectra, and their relation with the symmetric groups. In the second part, we express the component Dyer–Lashof coalgebras as subalgebras of a cofree unstable coalgebra on two cogenerators using an extension of the Peterson conjecture. We also compare and approximate $H^*\left(Q_0S^0;\mathbb{Z}/p\mathbb{Z}\right)$ with certain free objects using modular invariants. A new basis for $H^*\left(B\Sigma_\infty;\mathbb{Z}/p\mathbb{Z}\right)$ is provided.

10.1 Introduction

In this lecture note we survey results obtained under the research program Thalis [24, 25] and place them in the context of algebraic topology. Moreover, we correct some statements appearing in [24]. We begin by surveying infinite loop spaces, Ω-spectra, generalized cohomology theories and how these concepts relate to each other. Section one is intended to be understood by a general audience. The infinite loop space structure is reflected by certain invariants; e.g., the Dyer–Lashof algebra. We recall the Dyer–Lashof algebra and its algebraic structure following May [10] in section two. Its connection with modular invariants is discussed in section three. Modular invariant theory plays an important role in *modp* cohomology of groups and representation theory. Its application in algebraic topology was realized by Madsen and since then many people have demonstrated its use. We provide an incomplete list of references which is not intended to be up to date.

The action of the Steenrod algebra plays an important role in studying homotopy groups of spheres. It is also of great use in the study of cohomology of infinite

N.E. Kechagias (✉)
Department of Mathematics, University of Ioannina,
45500 Ioannina, Greece
e-mail: nkechag@uoi.gr

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_10

loop spaces. We closely investigate its action on both the Dickson and Dyer–Lashof algebras.

The Dyer–Lashof algebra R is a component coalgebra $R = \bigoplus_{n\geq 0} R[n]$. We consider the category of connected cocommutative positively graded coalgebras. Let $\mathscr{A}$ stand for the Steenrod algebra. The advantage of using modular invariant theory is the complete knowledge of the action on Dickson algebra generators instead of the standard approach as in [10]. The hom dual of $R[n]$ is isomorphic to the classical Dickson algebra D_n, for $p = 2$, as $\mathscr{A}$-algebras [44]. Here the hom dual stands for the $\mathbb{F}_p$ dual. The classical Dickson algebra of length n, $D_n = P[y_1, \ldots, y_n]^{GL_n}$, is a polynomial algebra. The Peterson conjecture is about the global structure of the classical Dickson algebra as an unstable algebra over the Steenrod algebra. Pengelley, Peterson and Williams, for $p = 2$ [48], proved that the classical Dickson algebra is a free unstable algebra on a certain cyclic module, modulo one additional relation. Here and bellow the term "free unstable algebra" is used according to Steenrod in [56] p. 29, see Definition 10.9. Because of Mui's isomorphism between $(R[n])^*$ and D_n for $p = 2$ [44], it follows that $R[n]$ is isomorphic to a subcoalgebra of a cofree unstable coalgebra on one cogenerator.

We consider the odd primary case. The Bockstein operations introduce amusing complications. We prove that $R[n]$ is isomorphic to a subcoalgebra of a cofree unstable $\mathscr{A}$-coalgebra on two cogenerators in [25]. Dually, $R[n]^*$ is isomorphic to a free unstable $\mathscr{A}$-algebra on a module generated by two elements μ and u modulo certain relations (Definition 10.9). This is one of our main theorems. As a corollary we obtain an isomorphism $Hom_{\mathscr{C}\mathscr{A}}(R[n], R[n]) \cong \mathbb{Z}/p\mathbb{Z}$. We discuss this extension in section four.

We close this survey by giving an alternative description of the *modp* cohomology of the infinite symmetric group using modular invariants as it appeared in [24]. Barratt–Priddy [4] and Quillen [50] independently established the connection between the classifying space of the infinite symmetric group and the stable range of spheres. We recall that the **k**-stem or **k**th stable homotopy group of spheres is $\pi_{n+k}(S^n)$ for $n > k + 1$. Our motivation is a better understanding of the action of the Steenrod algebra on the cohomology of infinite loop space of the zero sphere $H^*\left(Q_0S^0; \mathbb{F}_p\right)$. An explicit basis for the cohomology of the infinite symmetric group is constructed in terms of Dickson invariants, Theorem 12 in [24].

Let $ED_n := (E(x_1, \ldots, x_n) \otimes P[y_1, \ldots, y_n])^{GL_n}$ stand for the extended Dickson algebra of length n. Let $ED = \bigoplus_{n\geq 1} ED_n^+$ be the Dickson algebra monomials of positive degree of any length and SED a certain submodule that will be defined later. Let $SED^{\#}$ be the corresponding abelian restricted Lie algebras and $V\left(SED^{\#}\right)$ the corresponding universal enveloping algebra. We prove that $V\left(SED^{\#}\right)$ and $H^*\left(Q_0S^0; \mathbb{F}_p\right)$ are isomorphic algebras but not as Steenrod algebras. In [24] we stated that they are not isomorphic as Steenrod modules and this is not what we proved. The map we used to define the isomorphism provides no information as a Steenrod module isomorphism. We take the opportunity to correct some mathematical statements pointed out by the referee. Wellington also proved that $H^*\left(Q_0S^0; \mathbb{F}_p\right)$ is not a free unstable

Steenrod algebra in his memoirs A.M.S., [59] p. 35. We approximate the difference defining certain subalgebras C^k of $V\left(SED^{\#}\right)$. Their images in $H^*\left(Q_0S^0;\mathbb{F}_p\right)$ under the above isomorphism filter $H^*\left(Q_0S^0;\mathbb{F}_p\right)$. Certain quotients of those $\mathscr{A}$-algebras are isomorphic with free Steenrod algebras generated by certain Dickson submodules, Theorem 15 in [24].

We thank the referee very much for his numerous valuable suggestions-corrections regarding the exposition of this work and the correction regarding the misleading statement in [24] as mentioned above.

10.2 A Brief Survey

Topology is the study of topological spaces and continuous maps. A "nice" topological space is a CW complex; i.e., a space built up from cells. By attaching cells on a space X one can built a large class of topological spaces which have a certain amount of geometric intuition behind them. A major goal of algebraic topology is to study topological spaces by means of geometric invariants such as homotopy or (co)homology. There are far too many topological spaces and maps to deal with effectively. Thus a homotopy theorist studies homotopy classes of maps from X to Y, $[X, Y]$, and classes of spaces under homotopy equivalences. Two fundamental problems in homotopy theory are to determine if X and Y are of the same homotopy type and to compute $[X, Y]$. If X is a sphere and maps are respecting the base points, $[S^n, Y]$ is just the nth homotopy group of Y: $\pi_n(Y)$.

Homotopy theory has a strong interplay with many other areas of mathematics, particular geometry and algebra. Many geometric properties can be classified using homotopy theory. Just to mention a few with great impact in mathematics.

- The **Hopf invariant** problem introduced by Hopf related to many questions. *For which n does there exist a map $f : S^{2n-1} \to S^n$ with Hopf invariant one? For which n, $\mathbb{R}^n$ is a division algebra with a norm preserving multiplication?*
- The **vector field** problem. *What is the maximum number of linearly independent vector fields on a sphere?*
- The **Kervaire invariant** problem of differential topology. *In which dimensions n does there exist a smooth framed manifold with Kervaire invariant one?*

Since manifolds are of a great importance in mathematics we discuss a well known geometric application and its relation with homotopy theory [61].

Given a closed n-manifold M, under what conditions is it like the boundary of an $(n+1)$-manifold W? Two manifolds are called "bordant" if and only if their disjoint union is the boundary of an $(n+1)$-manifold. This defines an equivalence relation. Thom [57] showed that the bordism classes form a graded ring called cobordism ring (group). He reduced the study of cobordism groups to the study of stable homotopy groups:

$$\lim_m \pi_{n+m}\left(MO(n)\right).$$

Here the spaces $MO(n)$ constructed by Thom are called Thom complexes. The limit is formed by the maps $\Sigma MO(n) \to \Sigma MO(n+1)$ and Σ is the reduced suspension functor. He constructed a spectrum **MO** which is approximated by the spaces $MO(n)$. We shall introduce two notions, "stable" homotopy theory and "spectra". It was stated by Milnor [39] that one should deal with single objects like *spectra* **E** rather than the spaces $E(n)$ which approximate it.

In algebraic topology, it is rather important to distinguish between unstable problems, which arise in some definite dimension, and stable problems, which arise in any sufficiently large dimension. To that direction we have the suspension homomorphism

$$\Sigma : \pi_{n+k}\left(S^n\right) \to \pi_{n+k+1}\left(S^{n+1}\right) \text{ or } [S^{n+k}, S^n] \to [\Sigma S^{n+k}, \Sigma S^n]$$

and the Freudenthal Suspension Theorem says that this homomorphism is an isomorphism for $n > k+1$. Hence the groups $\pi_{n+k}(S^n)$ are independent of dimension for $n > k+1$ and called the k-stem. As long as the dimension of the sphere is large enough the homotopy group depends only on the relative dimension k. This is encoded by an object called "*spectrum*". Spectra are objects that capture the properties of topological spaces which are independent of the dimension. For example cohomology functors are themselves spectra. Spectra have extremely nice formal properties, and they can be used to give information about certain homotopy groups of spaces, known as the "*stable*" homotopy groups. Before commenting further on this notion, we recall some classical constructions.

Serre [52] called a space X equipped with a product $\mu : X \times X \to X$ which satisfies appropriate axioms an "H-space" in honor of Hopf's work on the topology of Lie groups. Any topological monoid is an H-space but a general H-space need not be associative, not even up to homotopy.

If X is an H-space, μ and the Künneth formula are used to define a "Pontryagin product" on $H_*(X)$. An important example of an H-space is a "loop space"

$$\Omega X = \left\{\left(S^1, 1\right) \to (X, x_0)\right\} \text{ with the compact open topology.}$$

The H-space structure does not depend on μ, up to isomorphism of H-spaces. ΩX is the fiber space of the Serre fibration $\Omega X \to PX \to X$. Here $PX = \{([0,1], 0) \to (X, x_0)\}$ is the contractible path space of X. Serre used the fibration above and spectral sequences to come up with very important results in homotopy theory [52, 53].

One can iterate the loop space construction

$$\Omega^n X = \Omega\left(\Omega^{n-1} X\right) = \cdots =: \left\{\left(S^n, *\right) \to (X, x_0)\right\}.$$

A space Y is called an "infinite loop space", if there is a sequence of spaces $\{X_n\}$ with $Y = X_0$ and weak homotopy equivalences

$$X_n \overset{\simeq}{\to} \Omega X_{n+1}.$$

We recall that a map $g : X \to Y$ between connected spaces is a weak equivalence if

$$g_n : \pi_n(X) \to \pi_n(Y)$$

is an isomorphism for all n. This implies that $g_* : [W, X] \to [W, Y]$ is an isomorphism for all CW-complexes W. Weak equivalences are used instead of homotopy equivalences to avoid technical nuisance.

One may ask: "how do infinite loop spaces arise in nature"? We comment on this question bellow.

In homology theory we have:

$$\overline{H}_n(X, \pi) \cong \overline{H}_{n+1}(\Sigma X, \pi).$$

Here π is an abelian group. Although the ordinary reduced cohomology of a sphere is very simple,

$$\overline{H}^*\left(S^n\right) = \begin{cases} \mathbb{Z}, \text{ if } * = n \\ 0, \text{ if } * \neq n \end{cases}$$

the cohomology of ΩS^n is a free algebra on one generator of degree $n-1$.

We recall that the suspension functor Σ and the loop functor Ω are adjoint (Kahn [20]):

$$[X, \Omega Y] \leftrightarrow [\Sigma X, Y]. \tag{10.1}$$

This bijection implies $\pi_n(\Omega X) \cong \pi_{n+1}(X)$.

A sequence of spaces $\{X_n\}$ and maps $\{\Sigma X_n \to X_{n+1}\}$ is an example of a "spectrum". For a complete account see Adams [1]. The notion of a spectrum is due to Lima [28].

A "spectrum" is a sequence of spaces and maps $\{X_n, f_n\}$ together with structure maps $f_n : \Sigma X_n \to X_{n+1}$ or $\overline{f}_n : X_n \to \Omega X_{n+1}$. These maps do not need to be defined until some arbitrarily high suspension.

Two basic examples are in order. There is a functor from spaces to spectra called Σ^∞, the suspension spectrum, $\Sigma^\infty X := \{X_n = \Sigma^n X,\ f_n = 1\}$. And the loop spectrum defined bellow.

Let us note that there are various other definitions of spectra, together with definitions for morphisms of spectra, fibrations, and cofibrations such that there are a lot of model categories of spectra. Most of them are Quillen adjoint and yield isomorphic homotopy categories. For model categories see [16].

The story started with Hurewicz (1939) who studied aspherical spaces with trivial higher homotopy groups. For example closed orientable surfaces of genus greater than zero are such examples. Eilenberg and MacLane [13] took a step further. They considered spaces whose homotopy groups vanish in all but one dimension n. These spaces, called Eilenberg–MacLane spaces, were to assume enormous importance in homotopy theory

$$\pi_* (K(\pi, n)) = \begin{cases} \pi, \text{ if } * = n \\ 0, \text{ if } * \neq n \end{cases}.$$

Here π is an abelian group if $n \geq 1$. Because of (10.1), $K(\pi, n) \simeq \Omega K(\pi, n+1)$ and $\{K(\pi, n)\}$ is an example of an Ω-spectrum defined bellow and any Eilenberg–MacLane space is an infinite loop space.

Let us return to sequences of spaces $X_0, X_1, \ldots$. Each weak equivalence $X_n \xrightarrow{\simeq} \Omega X_{n+1}$ can of course be transformed into a map $\Sigma X_n \to X_{n+1}$. So such a sequence is a spectrum. It requires a special name. Let $\mathbf{E}$ be a spectrum, $\left\{E_n, \overline{f}_n : E_n \to \Omega E_{n+1}\right\}$; we call $\mathbf{E}$ an "Ω-spectrum" if the maps $\overline{f}_n$ are weak homotopy equivalences. If $\mathbf{E}$ is an Ω-spectrum we have the following spacial property

$$\left[\Sigma^\infty X, \mathbf{E}\right]_n = [X, E_n].$$

We can replace the spaces E_n by weakly equivalent ones so that we actually get homeomorphisms $E_n \cong \Omega E_{n+1}$ [32]. This is very appropriate in the theory of infinite loop spaces, where it is necessary to keep the geometry under control.

Using the suspension spectrum one can define "stable" homotopy groups. The stable homotopy groups of a complex X are defined by

$$\pi_n^s (X) = \left\{S^n, X\right\} = \lim_{m \to \infty} [S^{n+m}, \Sigma^m X] = [\Sigma^\infty S^0, \Sigma^\infty X]_n.$$

Here S^0 stands for the zero sphere.

Complexes X and Y are of the same stable homotopy type, if there exists an integer m such that $\Sigma^m X$ and $\Sigma^m X$ are homotopy equivalent. The set of stable homotopy classes of maps from X to Y shall be

$$\lim_{m \to \infty} [\Sigma^m X, \Sigma^m Y].$$

This bicovariant functor (covariant in Y and contravariant in X) is "stable homotopy" and is closely related with the suspension spectrum. It would be nice to work in a category where stable maps were simply maps. Stable homotopy in the category of spaces is in fact a rather awkward interplay. One constantly suspends maps to stay in the stable range. So people work in a "stable homotopy category". A map of Ω-spectra $f : E \to F$ is a set of morphisms $\{f_n : E_n \to F_n\}$ such that the following diagram commutes

$$\begin{array}{ccc} E_n & \xrightarrow{\simeq} & \Omega E_{n+1} \\ f_n \downarrow & & \downarrow \Omega f_{n+1} \\ F_n & \xrightarrow{\simeq} & \Omega F_{n+1} \end{array}$$

The category of Ω-spectra is defined and its homotopy category, given by inverting maps of Ω-spectra which are homotopy equivalences, is called the **stable homotopy category**.

Ordinary cohomology is related with an Ω-spectrum, namely an Eilenberg–MacLane spectrum as follows. By obstruction theory, if X is any complex, then $[X, K(\pi, n)]$ is in one to one correspondence with ordinary cohomology $H^n(X, \pi)$. So cohomology groups are representable functors by Brown's Representability Theorem. This remarkable theorem [8] states that if you assign some algebraic object $\mathbf{F}(X)$ to a complex X and it has certain properties, then there is a spectrum $\mathbf{F}$ such that

$$\mathbf{F}(X) = [X, \mathbf{F}] \text{ or } F^n(X) = \left[X, F^n\right] \text{ for } n \geq 0.$$

This only happens if $F^n \xrightarrow{\simeq} \Omega F^{n+1}$ and it is related with generalized cohomology theories (Whitehead [60]). The category of Ω-spectra is equivalent to the category of cohomology theories on based spaces. Axioms for (co)homology theories were announced by Eilenberg and Steenrod in 1945. It was very impressive that a complicated subject as cohomology theory could be characterized by simple axioms. All but one of their axioms have a generalized character, the Dimension axiom is quite specific. When one needs it, one may throw in the wedge axiom of Milnor [40].

Definition 10.1 Let $F^n : \mathbf{CW} \to \mathbf{Ab}$ be a sequence of functors from the category of CW complexes to abelian groups, indexed with $n \in \mathbb{Z}$. The sequence $\{F^n\}$ is called a generalized reduced cohomology theory on CW complexes, if the following axioms hold for each n:

1. F^n is invariant under basepoint-preserving weak homotopy equivalence, and as such factors through the pointed homotopy category of CW-complexes.
2. $F^n\left(\bigvee_\alpha X_\alpha\right) \simeq \bigsqcup_\alpha F^n(X_\alpha)$.
3. Let $X = A \cup B$ be the union of subcomplexes and each contain the base point. If $a \in F^n(A)$ and $b \in F^n(B)$ restrict to the same element in $F^n(A \cap B)$, there exists an element $x \in F^n(X)$ which restricts over A and B, respectively.
4. For each CW pair (X, A) there is a long exact sequence

$$\cdots \to F^n(X) \to F^n(A) \to F^n(X, A) \to F^{n+1}(X) \to \cdots$$

where we write $F^n(X, A)$ for $F^n(X/A)$.

From the knowledge of all $F^n\left(S^0\right)$ (the coefficient ring) one can derive $F^n\left(S^k\right)$ and from the knowledge of the attaching maps of a CW complex X one can derive $F^n(X)$ up to extension for all X. F^n should convert homotopy colimits into limits. We just note that an analogue of Brown representability holds for simplicial presheaf categories over pointed complete cofibrantly generated simplicial model categories that satisfy certain properties.

So any generalized cohomology theory yields an Ω-spectrum and vice versa. Moreover, all algebraic properties of $\mathbf{F}(X)$ are reflected in $\mathbf{F}$ up to homotopy. For example, if $\mathbf{F}(X) \cong [X, \mathbf{F}]$ is a ring, then $\mathbf{F}$ is a ring up to homotopy.

If one starts from a cohomology theory, then one constructs a representing spectrum. If one takes the corresponding cohomology theory, then one will recover the

original cohomology theory up to isomorphism. In a celebrated paper G.W. Whitehead showed that if we start from the spectrum $\mathbf{F}$, we can define a generalized (co)-homology theory [60]. As we mentioned at the beginning of this survey, this is done in such a way that the Thom spectrum, although it is not an Ω-spectrum, gives rise to real, unoriented bordism and cobordism. Conversely, if one starts from a spectrum $\mathbf{F}$, then he will construct the corresponding generalized cohomology theory $\mathbf{F}^*$. Thus a representing Ω-spectrum $\mathbf{E}$. Then $\mathbf{E}$ and $\mathbf{F}$ have the same homotopy groups (the coefficient groups for $\mathbf{F}^*$). Any spectrum $\mathbf{F}$ can be replaced with an equivalent Ω-spectrum $\mathbf{E}$:

$$E_n = \lim_m \Omega^m F^{n+m}.$$

$\Omega^\infty F$ is defined to be E_0. Thus Ω^∞ is a functor from spectra to spaces and Σ^∞ from spaces to spectra. The values of the functor Ω^∞ are infinite loop spaces. An example: $QX := \Omega^\infty \Sigma^\infty X = \lim_m \Omega^m \Sigma^m X$. One may think of infinite loop spaces as the abelian groups up to homotopy in the strongest sense.

May and Thomason [34] proved that if $\mathscr{C}$ is a "strict symmetric monoidal category" then the classifying space $B\mathscr{C} = |N\mathscr{C}|$, the geometric realization of the nerve, is an infinite loop space. All infinite loop spaces arise from this "construction" and all such constructions are equivalent in some sense.

One may ask: "how much of Brown's representability theorem carries over to a general model category?" Jardine [19] proved an analogue of Brown representability theorem for simplicial presheaf categories. Of course there are examples of model categories for which Brown's representability fails.

It is evident now that a category is required which allows one to study both homotopy theory and cohomology theories. Roughly speaking, this category should provide the following:

- We should have CW-spectra, so we should be able to carry over CW approximation from the classical setting.
- It should be additive.
- It should give us a faithful embedding of stable homotopy category of spaces into that of spectra: its suspension spectrum should yield stable homotopy groups.
- The suspension functor should be an equivalence.
- We should have the Brown representability analogue.
- We should have certain maps of spectra $\mathbf{E} \times \mathbf{E} \to \mathbf{E}$ so that extra structure will be provided (like the cup product in cohomology).

Let us return to the specific theme expressed by the title of this work. All (co)-homologies are understood to be mod p in this work, p is any prime number.

A natural question to ask is *what do the spaces* $\Omega^n X$ *look like*? ΩX is an H-space and $\Omega^2 X$ is more than an H-space (Stasheff [54]). So $\Omega^\infty X$ admits extra structure. Their invariants reflecting the infinite loop space structure such as the homology operations were defined and studied by Kudo and Araki [26, 27] for $p = 2$, Browder [7] and Dyer and Lashof [12] for p an odd prime. They were motivated by Steenrod's

[55] construction of cohomology operations, and studied the *modp* homology of $\Omega^{\infty}\Sigma^{\infty}X$ for a prime p.

Boardman and Vogt [5] (inspired by Milgram [35]) observed that the product in an iterated loop space is not only homotopy commutative and associative, but it is subject to a whole hierarchy of higher commutativity and associativity relations. The work of May [33] and Segal [51] gave a good understanding of these spaces. May constructed approximation models $C_m X$ for $\Omega^m \Sigma^m X$ such that they are weak homotopy equivalent. These spaces play an important role in the subject.

Let Σ_n be the symmetric group of degree n, i.e., Σ_n is the group of all permutations of n elements and $B\Sigma_n$ its classifying space (Milnor [37]). The infinite symmetric group Σ_∞ is defined to be the direct limit of the symmetric groups under the natural inclusions:

$$i_n^m : \Sigma_n \to \Sigma_m \text{ by } i_n^m(\alpha)(i) = \begin{cases} \alpha(i), \text{ if } 1 \le i \le n \\ i, \text{ if } n \le i \le m \end{cases}.$$

Here $\alpha \in \Sigma_n$. There are also homomorphisms $\mu_{n,m} : \Sigma_n \times \Sigma_m \to \Sigma_{n+m}$ by

$$\mu_{n,m}(\alpha, \beta)(i) \begin{cases} \alpha(i), \text{ if } 1 \le i \le n \\ n + \beta(i-n), \text{ if } n < i \le n+m \end{cases}.$$

The maps $\mu_{n,m}$ induce a multiplication in the homology group $H_*(B\Sigma_\infty)$ calculated by Nakaoka [45] (see also [15]).

When one constructs a model for $\Omega^{\infty}\Sigma^{\infty}X$, by the method of J.P. May [33] or any other method, one uses the spaces $B\Sigma_n$:

$$C(X) = \bigsqcup_{n \ge 1} E\Sigma_n \times_{\Sigma_n} X^n \swarrow\approx .$$

Here $E\Sigma_n$ is the universal covering space for $B\Sigma_n$, Σ_n acts on X^n by permutation of coordinates and $\approx$ stands for certain relations [33].

For the zero sphere S^0, $C\left(S^0\right) = \bigsqcup_{n \ge 1} B\Sigma_n$ has a natural associative H-space structure with unit induced from $\mu_{n,m}$ upon applying the functor B and using $B(\Sigma_n \times \Sigma_m) = B\Sigma_n \times B\Sigma_m$.

Barratt and Priddy [4] and Quillen [50] independently proved the following important theorem in homotopy theory. For details, please see [1].

Theorem 10.1 *(i)* $Q\left(S^0\right)$ *is the group completion of* $C\left(S^0\right)$.

(ii) There are natural inclusions $i_n : B\Sigma_n \to Q\left(S^0\right)$ *inducing an* H*-map* $i : C\left(S^0\right) \to Q\left(S^0\right)$ *with* $Bi : BC\left(S^0\right) \to Q\left(S^1\right)$ *a homotopy equivalence. Moreover,* $\Omega BC\left(S^0\right) \simeq Q\left(S^0\right)$.

(iii) $H_*\left(\Omega BC\left(S^0\right)\right) \cong H_*(\mathbb{Z} \times B\Sigma_\infty)$.

$Q\left(S^0\right)$ is naturally equipped with two products [30]. One is the "loop sum" which is given by $\Omega^{n-1}(*) : (\Omega^n S^n) \times (\Omega^n S^n) \to (\Omega^n S^n)$ where $* : (\Omega S^n) \times (\Omega S^n) \to$

(ΩS^n) is the usual loop sum. This product is reflected in $C\left(S^0\right)$ by the family of homomorphisms

$$\phi_{n,m} : \Sigma_n \times \Sigma_m \to \Sigma_{n+m} \text{ given by}$$

$$\phi_{n,m}(\alpha, \beta)(i) = \begin{cases} \alpha(i), 1 \leq i \leq n; \\ n + \beta(i-n), n+1 \leq i \leq n+m. \end{cases}$$

The second one is the composition product which is reflected in $C\left(S^0\right)$ by the family of homomorphisms

$$\psi_{n,m} : \Sigma_n \times \Sigma_m \to \Sigma_{nm} \text{ given by}$$

$$\psi_{n,m}(\alpha, \beta)(i, j) = (\alpha(i), \beta(j)), 1 \leq i \leq n \text{ and } 1 \leq j \leq m.$$

Here (i, j) is given the lexicographical ordering [35].

$H_*\left(Q\left(S^0\right)\right)$ is a coalgebra as it is with all spaces but also an algebra. Since the induced product in homology is a map of coalgebras, it makes $H_*\left(Q\left(S^0\right)\right)$ a Hopf algebra [41]. Topologists call a Hopf algebra a "coalgebraic group" because it is an abelian group object in the category of coalgebras. The second multiplication induces another map. Altogether this total structure is called a "Hopf ring" or a "coalgebraic ring" ([49]). The coalgebraic ring structure of $H_*\left(Q\left(S^0\right)\right)$ is of interest in the study of homology of infinite loop spaces [9, 15, 58].

It is well known that all spectra are module spectra over the sphere spectrum. The *modp* homology of any infinite loop space becomes an $H_*\left(QS^0\right)$-module in the category of coalgebras. $H_*\left(QS^0\right)$ is our object of study.

Before commenting further on the homology of infinite loop spaces, we shall recall the Steenrod operations. Steenrod introduced a family of operations

$$Sq^i : H^n(-; \mathbb{F}_2) \to H^{n+i}(-; \mathbb{F}_2)$$

which behave well with respect to suspension. They were extended to any odd prime number p. For all integers i and n there is a natural transformation of functors which is a homomorphism

$$P^i : H^n\left(-; \mathbb{F}_p\right) \to H^{n+2i(p-1)}\left(-; \mathbb{F}_p\right)$$

and the Bockstein operator

$$\beta : H^n\left(; \mathbb{F}_p\right) \to H^{n+1}\left(; \mathbb{F}_p\right)$$

associated to the short exact sequence $0 \to \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p^2\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z} \to 0$.

The stable operations in *modp* cohomology form an algebra $\mathscr{A}$ called the Steenrod algebra [56]. The *modp* Steenrod algebra $\mathscr{A}$ is the quotient of the "free" associative unital graded $\mathbb{F}_p$-algebra generated by the elements:

$$Sq^i, i > 0, \text{ if } p = 2,$$

$$\beta \text{ and } P^i, i > 0 \text{ and } \beta^2 = 0, \text{ if } p > 2;$$

by the ideal generated by the elements, known as the **Adem relations**

$$Sq^i Sq^j - \sum_0^{[i/2]} \binom{j-k-1}{i-2k} Sq^{i+j-k} Sq^k \text{ , } i, j > 0 \text{ such that } i < 2j, p = 2;$$

$$P^i P^j - \sum_0^{[i/p]} \binom{(p-1)(j-t)-1}{i-pk} P^{i+j-t} P^t, i, j > 0 \text{ such that } i < pj \text{ and}$$

$$P^i \beta P^j - \sum_0^{[i/p]} (-1)^{i+t} \binom{(p-1)(j-t)}{i-pt} \beta P^{i+j-t} P^t$$

$$- \sum_0^{[(i-1)/p]} (-1)^{i+t-1} \binom{(p-1)(j-t)-1}{i-pt-1} P^{i+j-t} \beta P^t, i, j > 0 \text{ such that } i \leq pj, p > 2.$$

Sq^0 for $p = 2$ and P^0 for $p > 2$ are understood to be the unit.

Steenrod and Adem proved that for any space (or spectrum) X, $H^*(X)$ is in a natural way a graded $\mathscr{A}$-module. Moreover, the Cartan formula holds for cohomology classes

$$P^n(xy) = \sum_{i+j=n} P^i(x) P^j(y).$$

The cohomology operations can be used to study homotopy groups of spheres (for an introduction see Mosher-Tangora [42]).

Let X be an infinite loop space. There is another algebra R acting on homology of infinite loop spaces generated by

$$Q^i : H_n\left(X; \mathbb{F}_p\right) \to H_{n+2(p-1)i}\left(X; \mathbb{F}_p\right)$$

and

$$\beta : H_n\left(X; \mathbb{F}_p\right) \to H_{n-1}\left(X; \mathbb{F}_p\right)$$

which satisfy certain properties [10]. They can also be used to study stable homotopy groups of spheres (for an introduction see Wellington [59]).

$H_*(QX)$ admits an $\mathscr{A}-R$ allowable Hopf algebra structure (May [10]); i.e., it is a Hopf algebra on which both the Steenrod algebra and the Dyer–Lashof algebra act and the two actions satisfy the Nishida relations [46]. We shall study their structure in the next sections.

10.3 The Component Dyer–Lashof Coalgebras

The structure of $H_*(Q_0S^0)$ has been described in terms of Dyer–Lashof operations. These family of homology operations were defined by Kudo and Araki [26, 27] for $p = 2$, Browder [7] and Dyer and Lashof [12] for p an odd prime.

$$Q^i : H_n\left(QX, \mathbb{F}_p\right) \to H_{n+2(p-1)i}\left(QX, \mathbb{F}_p\right)$$

and

$$\beta : H_n\left(QX, \mathbb{F}_p\right) \to H_{n+1}\left(QX, \mathbb{F}_p\right).$$

$Q^s y = 0$ if $2s < |y|$ while $Q^s y = y^p$ if $2s = |y|$ and $Q^s[0] = 0$ for $s > 0$. For $k \in \mathbb{Z}$ there is a component in QS^0. Here Q_kS^0 stands for the kth component. All components are homotopically equivalent. Let $[k]$ be the image of the generator of $H_0\left(Q_kS^0\right)$ in $H_0\left(QS^0\right)$. The element $[0] \in H_0(QX)$ is the identity for the Pontryagin algebra $H_*(QX)$ induced by the loop product. Let Q_0X denote the base point component. In degree zero $H_*(QS^0)$, $[k] \cdot [l] = [k + l]$ and $Q^0[k] = [pk]$. The operation Q^0 acts on the homology of the base point component $H_*(Q_0S^0)$ (Madsen [29], Cohen [10]).

Iterates of the Dyer–Lashof operations are of the form

$$Q^{(I,\varepsilon)} = \beta^{\varepsilon_k} Q^{i_k} \dots \beta^{\varepsilon_1} Q^{i_1}$$

where $(I, \varepsilon) = ((i_k, \dots, i_1), (\varepsilon_k, \dots, \varepsilon_1))$ with $\varepsilon_j = 0$ or 1 and i_j a non-negative integer for $j = 1, \dots, k$. For $p = 2$, $\varepsilon_j = 0$ for all j. $(I, \varepsilon) \in \mathbb{N}^k \times \mathbb{F}_2^k$. For $p = 2$, $I \in \mathbb{N}^k$. The excess of (I, ε) or $Q^{(I,\varepsilon)}$, denoted $e(I, \varepsilon)$, is defined by

$$e(I) = i_k - \left(\sum_1^{k-1} i_t\right) \text{ for } p = 2;$$

$$e(I, \varepsilon) = i_k - 2(p-1)\left(\sum_1^{k-1} i_t\right) - \varepsilon_k + \sum_1^{k-1} \varepsilon_t \text{ for } p \text{ an odd prime.}$$

There are relations among the iterated operations called Adem relations, so that an operation can be reduced to a sum of *admissible* operations after applying Adem relations. The Steenrod algebra $\mathscr{A}$ acts on the right of R via Nishida relations (see [10] p. 6).

The Dyer–Lashof algebra R is given as the quotient of a free associative algebra generated by $\left\{Q^i \mid i \geq 0\right\} \cup \left\{\beta Q^i \mid i \geq 1\right\}$ modulo the Adem relations, the relation $\beta^2 = 0$ and negative excess, see the details in [10].

It is a non-negatively graded Hopf algebra and also a component coalgebra $R = \bigoplus_{k \geq 0} R[k]$ with respect to the length of the operations.

Let $k \geq 1$, $R[k]$ is the sub $\mathscr{A}$-coalgebra spanned by admissible elements of fixed length k and non-negative excess.

$$\{Q^{(I,\varepsilon)} \mid (I,\varepsilon) \text{ admissible}, l(I) = k \text{ and } e(I,\varepsilon) \geq 0\}.$$

The coproduct is given by

$$\psi Q^{(I,\varepsilon)} = \sum Q^{(K,\varepsilon')} \otimes Q^{(I-K,\varepsilon'')}.$$

The sum and difference of two sequences of the same length is defined termwise under the conventions $\beta^2 = 0 = \beta^{-1}$ and $Q^{-i} = 0$. We note that the sum of two admissible sequence is also admissible, but the difference need not be. For admissible sequences (I,ε) we consider increasing sequences in $\mathbb{N}^k$, i.e. $(I,\varepsilon) = ((i_k,\ldots,i_1),(\varepsilon_k,\ldots,\varepsilon_1))$ such that $i_t + \varepsilon_{t-1} > pi_{t-1}$ and $i_1 \geq \varepsilon_1 \geq 0$ for $2 \leq t \leq k$.

We define an ordering on the set of sequences (I,ε) following May [10]. This ordering will be of special interest in the last section.

Definition 10.2 (a) For a sequence (I,ε), define $I_j = ((i_j,\ldots,i_1),(\varepsilon_j,\ldots,\varepsilon_1))$, $1 \leq j \leq k$, and similarly for $p = 2$. Note that $e(I_j) = e(J_j)$ for all j implies $(I,\varepsilon) = (J,\varepsilon')$, and define a total ordering of the sequences of length k by $(I,\varepsilon) < (J,\varepsilon')$ if $e(I_j) < e(J_j)$ for the largest j such that $e(I_j) \neq e(J_j)$. The weight of (I,ε) or $Q^{(I,\varepsilon)}$ is defined to be p^k, if the length of (I,ε) is k.

(b) A natural order is defined on the set of all sequences (I,ε) as follows: $(I,\varepsilon) < (J,\varepsilon')$, if

(i) $l(I,\varepsilon) < l(J,\varepsilon')$ or

(ii) $l(I,\varepsilon) = l(J,\varepsilon')$ and $(I,\varepsilon) < (J,\varepsilon')$.

Note that $i_k = e(I_{k-1})$ is equivalent to $e(I) = 0$.

Next we proceed to the definition of a vector space which plays an important role on defining the homology of Q_0S^0.

Definition 10.3 Let MS^0 be the $\mathbb{F}_p$-vector space on $BMS^0 = \{Q^{(I,\varepsilon)}x \mid (I,\varepsilon)$ is admissible of non-negative excess and positive degree$\} \cup \{[0]\}$. Here x represents the image of the generator of $\bar{H}_0(S^0)$ in $H_0(QS^0)$ and $Q^{(I,\varepsilon)}x$ corresponds to $Q^{(I,\varepsilon)}x \cdot \left\langle Q^{(I,\varepsilon)}x\right\rangle^{-1} = Q^{(I,\varepsilon)}x \cdot \left\langle -p^{l(I)}\langle x\rangle\right\rangle$ translating to the base point component (May [10] p. 55). Here $\langle x\rangle$ denotes the component of x.

We denote $(BMS^0)^+ = BMS^0 - \{[0]\}$.

Using the loop product we can discard the classes with $i_k = e(I_{k-1})$, i.e.,

$$Q^{(I,\varepsilon)}x \equiv Q^{I_{k-1}}x * \cdots Q^{I_{k-1}}x = \left(Q^{I_{k-1}}x\right)^p.$$

Now we state an important theorem.

Theorem 10.2 (Madsen, May) *$H_*(Q_0S^0)$ is the free commutative algebra generated by certain elements in $\left(BMS^0\right)^+$ modulo the ideal generated by*

$$\{Q^s y - y^p \mid |y| = 2s \text{ and } y \in \left(BMS^0\right)^+\} \text{ and } y^2 = 0, \text{ if } |y| = \text{odd and } p > 2.$$

See Wellington [59] or Theorem 1 and 2 in Kechagias [24] for details. Next we discuss the coalgebra structure for $R[n]$.

A **cofree** coalgebra is completely understood if a basis for the space of the primitive monomials elements is known. We recall that a primitive element of a co-algebra C is an element x that satisfies

$$\Delta x = x \otimes 1 + 1 \otimes x.$$

Madsen (for $p = 2$) and May (for p odd) explicitly described the primitive elements for each n [10, 29]. The primitive monomials are described in the next definition.

To simplify notation we merge the two sequences $(I, \varepsilon) = ((i_k, \ldots, i_1), (\varepsilon_k, \ldots, \varepsilon_1))$ into one $((\varepsilon_k, i_k), \ldots, (\varepsilon_1, i_1))$.

Definition 10.4 Let t be a natural number such that $0 \leq t \leq n-1$ and the term p^{-1} is omitted in the following expressions:
$I_{n,t} = \left(\left(0, p^{n-1} - p^{t-1}\right), \ldots, \left(0, p^{n-t} - 1\right), \left(0, p^{n-t-1}\right), \ldots, (0, 1)\right)$;
$J_{n,t} = \left(\left(0, p^{n-1} - p^{t-1}\right), \ldots, \left(0, p^{n-t} - 1\right), \left(1, p^{n-t-1}\right), \left(0, p^{n-t-2}\right), \ldots,$
$(0, 1)\big)$.

Let t and s be natural numbers such that $0 \leq s < t \leq n-1$ and the term p^{-1} is omitted in the following expression:
$K_{n;s,t} = \big(\left(0, p^{n-1} - p^{t-1} - p^{s-1}\right), \ldots, \left(0, p^{n-t} - p^{t-s} - 1\right), \left(1, p^{n-t-1} - p^{t-s-1}\right),$
$\left(0, p^{n-s-2} - p^{t-s-2}\right), \ldots, \left(0, p^{n-s} - 1\right), \left(1, p^{n-s-1}\right), \left(0, p^{n-s-2}\right), \ldots, (0, 1)\big)$.

The elements described above are divided in to three classes according to the number of Bockstein operations involved: non, one, and two. $\left\{I_{n,t}, J_{n,t} \text{ and } K_{n;s,t}\right\}$ for certain indices. Sequences of operations involving Bockstein operations correspond to particular classes involving x_i's in $H^*\left(\left(B\mathbb{F}_p\right)^n; \mathbb{F}_p\right)$ under the appropriate monomorphism.

Madsen [29] and May [10] computed the hom-duals of $R[n]$ in analogy with Milnor's computation of the hom-dual of the Steenrod algebra [38].

Theorem 10.3 (Madsen $p = 2$) *The* hom*-dual of $R[n]$, $(R[n])^*$, is a graded polynomial algebra generated by $\{\left(Q^{I_{n,t}}\right)^* \mid 0 \leq t \leq n-1\}$.*

Theorem 10.4 (May p odd) *The* hom*-dual of $R[n]$, $(R[n])^*$ is a graded algebra generated by certain elements*

$$\left\{\left(Q^{I_{n,t}}\right)^*,\ \left(Q^{J_{n,t}}\right)^*\ and\ \left(Q^{K_{n;s,t}}\right)^*\ |\ 0 \leq s < t \leq n-1\right\}$$

subject to particular relations.

For details see Theorem 4, p. 844, in [24].

In the next section the relation between $R[n]$ and modular invariants is discussed.

10.4 The Dickson Algebras as $\mathscr{A}$-Modules

Let G be a finite group, then there is a canonical embedding $G \hookrightarrow \Sigma_{|G|}$ induced by the regular representation. If $H \leq G$, we have a commutative diagram

$$\begin{array}{ccc} G & \overset{reg}{\rightarrow} & \Sigma_{|G|} \\ j \uparrow & & \uparrow \Delta \\ H & \overset{reg^{|G:H|}}{\rightarrow} & \left(\Sigma_{|H|}\right)^{|G:H|} \end{array}$$

Here we write $G = Hg_1 \bigsqcup \cdots \bigsqcup Hg_{|G:H|}$, where the g_i are representatives of the right coset of H in G. These maps provide canonical characteristic classes for group cohomology. For example, if $G = \left(\mathbb{F}_p\right)^n$, then

$$Imreg^* = H^*(B\mathbb{Z}/p\mathbb{Z} \times \cdots \times B\mathbb{Z}/p\mathbb{Z}; \mathbb{F}_p)^{GL\left(n,\mathbb{F}_p\right)}$$

which is the extended Dickson algebra defined bellow. This may partially explain why the Dickson algebra plays such an important role. Invariant theory was used to detect unexpected multiplicative relations in $H^*\left(B\Sigma_n; \mathbb{F}_p\right)$ [3, 14, 15]. The ring of invariants are at the core of any computations of $H^*\left(B\Sigma_n; \mathbb{F}_p\right)$. They build up successively, yielding relations rich in symmetry but of a high convoluted type. It is known that $H^*\left(B\Sigma_n; \mathbb{F}_p\right)$ is detected by elementary abelian subgroups [50]. Moreover, the restriction from the cohomology of a higher symmetric group to a lower one is surjective. Generators appearing at a finite stage (in $H^*\left(B\Sigma_n; \mathbb{F}_p\right)$ for some n) will be nontrivial images under restriction. On the other hand, the precise multiplicative relations among them change at each level, as the lattice of detecting subgroups becomes larger and more complicated. For example, the interested reader might examine $H^*\left(B\Sigma_8; \mathbb{F}_2\right)$ and $H^*\left(B\Sigma_{12}; \mathbb{F}_2\right)$ [14] or [3]. For notation and details please see [3].

It is well known that

$$H^*(B\mathbb{Z}/p\mathbb{Z} \times \cdots \times B\mathbb{Z}/p\mathbb{Z}; \mathbb{F}_p) \cong \begin{cases} P\left[y_1, \ldots, y_n\right], \text{for } p = 2 \\ E\left(x_1, \ldots, x_n\right) \otimes P\left[y_1, \ldots, y_n\right] \end{cases}$$

as $\mathscr{A}$-algebras. Here the *$\mathscr{A}$-algebra* structure on the right hand side is given by:

$$P^i(ab) = \sum_{0 \le j \le i} \left(P^j a\right)\left(P^{i-j} b\right) \text{ and } \beta x_i = y_i.$$

Let $g \in GL\left(n, \mathbb{F}_p\right)$ and $g = \left(g_{i,j}\right)$, then

$$\left(g_{i,j}\right) x_s = \sum_{i=1}^{n} g_{i,s} x_s \text{ for } 1 \le s \le n.$$

This action is extended to the y_i's via the Bockstein homomorphism:

$$\left(g_{i,j}\right) y_s = \sum_{i=1}^{n} g_{i,s} y_s \text{ for } 1 \le s \le n.$$

It follows that the group action commutes with the Steenrod algebra action and $(E(x_1, \ldots, x_n) \otimes \mathbb{F}_p[y_1, \ldots, y_n])^G$ is an $\mathscr{A}$-module for any subgroup G of $GL\left(n, \mathbb{F}_p\right)$.

The extended Dickson algebra ED_n and a certain subalgebra SED_n related with the Dyer–Lashof coalgebra $R[n]$ along with the Mui generators are defined.

Theorem 10.5 ([11]) *The classical Dickson algebra is a polynomial algebra on n generators*

$$D_n := \mathbb{F}_p[y_1, \ldots, y_n]^{GL\left(n, \mathbb{F}_p\right)} = \mathbb{F}_p\left[d_{n,n-1}, \ldots, d_{n,1}, d_{n,0}\right].$$

Let L_n and $L_{n,k+1}$ denote the following determinants:

$$L_n = \begin{vmatrix} y_1 & \cdots & y_n \\ y_1^p & \cdots & y_n^p \\ \vdots & & \vdots \\ y_1^{p^{n-1}} & \cdots & y_n^{p^{n-1}} \end{vmatrix} \text{ and } L_{n,k+1} = \begin{vmatrix} y_1 & \cdots & y_n \\ y_1^p & \cdots & y_n^p \\ \vdots & & \vdots \\ y_1^{p^n} & \cdots & y_n^{p^n} \end{vmatrix}$$

where the row $\left(y_1^{p^{k+1}}, \ldots, y_n^{p^{k+1}}\right)$ is missing in $L_{n,k+1}$. Here $0 \le k+1 \le n$. $L_{n,n}$ is denoted by L_n. Moreover, $L_{n,0} = L_n^p$. A Dickson algebra generator was defined as follows:

$$d_{n,k+1} = \frac{L_{n,k+1}}{L_n}.$$

Mui (see [43]) gave an invariant theoretic description of the cohomology algebra of the symmetric group and calculated the associated rings of invariants involving the exterior subalgebra $E(x_1, \ldots, x_n)$ of $H^*\left((B\mathbb{Z}/p\mathbb{Z})^n; \mathbb{F}_p\right)$.

Certain elements $M_{n;s_1,\ldots,s_k}$ have been defined by Mui in [43] as follows:

$$M_{n;s_1,\ldots,s_k} = \frac{1}{k!}\begin{vmatrix} x_1 & \cdots & x_1 \\ \vdots & & \vdots \\ x_1 & \cdots & x_n \\ y_1 & \cdots & y_n \\ \vdots & & \vdots \\ y_1^{p^{n-1}} & \cdots & y_n^{p^{n-1}} \end{vmatrix}.$$

Here there are k rows of x_i's and the s_ith's powers are omitted, where $0 \le s_1 < \cdots < s_k \le n-1$ in the determinant. We call these elements Mui generators.

Here we use the notation introduced by Mui in [43] for the Dickson algebra generators. The symbol $d_{n,k}$ corresponds to $d_{n,n-k}$ in [24, 25].

Theorem 10.6 ([43]) *The extended Dickson algebra*

$$ED_n := (E(x_1, \ldots, x_n) \otimes \mathbb{F}_p[y_1, \ldots, y_n])^{GL(n,\mathbb{F}_p)}$$

is a free D_n-module with basis given by certain elements called Mui generators along with 1*:*

$$M_{n;s_1,\ldots,s_k} L_n^{p-2}, \text{ for } 1 \le k \le n, \text{ and } 0 \le s_1 < \cdots < s_k \le n-1.$$

Its algebra structure is given by certain relations.

Definition 10.5 Let SED_n be the subalgebra of ED_n generated by

$$\left\{d_{n,t},\, M_{n;t} L_n^{p-2},\, M_{n;s,t} L_n^{p-2} \mid 0 \le s < t \le n-1\right\}.$$

The generators $d_{n,t}$, $M_{n;t} L_n^{p-2}$ and $M_{n;s,t} L_n^{p-2}$ above correspond to the primitive elements $Q^{I_{n,t}}$, $Q^{J_{n,t}}$ and $Q^{K_{n;s,t}}$ in $R[n]$ respectively, see Theorem 10.8.

Proposition 10.1 ([22]) *SED_n is an $\mathscr{A}$-algebra which is a free D_n-module with basis given by elements consisting of the following monomials along with* 1*:*

$$\prod_{1\le r\le k} M_{n;s_{2r-1},s_{2r}} L_n^{p-2} \text{ and } \left(\prod_{1\le r\le q} M_{n;s_{2r-1},s_{2r}} L_n^{p-2}\right) M_{n;s_{2q+1}} L_n^{p-2}.$$

Here, $1 \le k \le \left[\frac{n}{2}\right]$, $0 \le s_1 < s_2 < \cdots < s_{2k} \le n-1$ *and* $1 \le q \le \left[\frac{n}{2}\right]$, $2q+1 \le n$, $0 \le s_1 < s_2 < \cdots < s_{2q} < s_{2q+1} \le n-1$.

The action of the Steenrod algebra on the cohomology of infinite loop spaces plays an important role in computing the Adams spectral sequence [59]. The key

step for calculations is to express the Mui generators involved in SED_n as Steenrod admissible operations on the generator with the least degree. Here comes an idea of Frank Peterson.

Peterson conjectured that the classical Dickson algebra admits a certain global structure as an unstable algebra over the Steenrod algebra. This conjecture was solved by Pengelley, Peterson and Williams for $p = 2$ [48]. They proved that the classical Dickson algebra is a free unstable algebra on a certain cyclic module, modulo one additional relation. Their approach leads to a characterization of other unstable cyclic modules.

Pengelley and Williams considered the odd primary case [47]. They proved that the classical Dickson algebra, D_n, is a free unstable algebra on a certain cyclic module, modulo one additional relation. Frank Peterson conjectured that the Steenrod algebra action on D_n characterizes its algebra structure.

It is well known that the operations Sq^{2^n}, $n \geq 0$, for $p = 2$ constitute a system of multiplicative generators for $\mathscr{A}$; so do the operations β and P^{p^n}, $n \geq 0$ for $p > 2$.

Definition 10.6 Let $p = 2$. For a sequence of integers $I = (i_1, \ldots, i_n)$, let Sq^I denote $Sq^{i_1} \ldots Sq^{i_n}$. The sequence I is said admissible if $i_t \geq 2i_{t+1}$ for all $t \geq 1$ $(i_{n+1} = 0)$.

Let $p > 2$. For a sequence of integers $I = (\varepsilon_0, i_1, \varepsilon_1, \ldots, i_n, \varepsilon_n)$, where the ε_k are 0 or 1, let P^I denote $\beta^{\varepsilon_0} P^{i_1} \beta^{\varepsilon_1} \cdots P^{i_n} \beta^{\varepsilon_n}$. The sequence I is said admissible if $i_t \geq pi_{t+1} + \varepsilon_t$ for all $t \geq 1$ $(i_{n+1} = 0)$. The operations Sq^I (P^I) with I admissible are called admissible monomials.

The next theorem provides the action of the Steenrod algebra generators on the generators of D_n. This remarkable property comes from the simple action of the Steenrod algebra on the lowest degree generator of D_n, $d_{n,n-1}$:

Theorem 10.7 (Theorem 30, p. 169, [21])

$$P^{p^k} d_{n,i}^{p^j} = \begin{cases} d_{n,i-1}^{p^j}, & \text{if } k = j + i - 1 \text{ and } 0 \leq i < n; \\ -d_{n,i}^{p^j} d_{n,n-1}^{p^j}, & \text{if } k = j + n - 1; \\ 0, & \text{otherwise.} \end{cases}$$

The following corollary expresses the module relation used by Pengelley and Williams.

Corollary 10.1 $P^{p^{n-2}} P^{p^{n-1}} d_{n,n-1} = 2P^{p^{n-1}} P^{p^{n-2}} d_{n,n-1}$.

Now we describe the Dickson algebra generators as Steenrod algebra actions on the lowest degree generator.

Corollary 10.2 *Let* $0 \leq m \leq n - 2$, *then*

$$P^{p^m + \cdots + p^{n-2}} d_{n,n-1} = P^{p^m} \ldots P^{p^{n-2}} d_{n,n-1} = d_{n,m}\,.$$

The remarkable property of the classical Dickson algebra is expressed in the next proposition.

Proposition 10.2 *(a) Let* $0 \leq k \leq n-2$ *and* $2 \leq t \leq p-1$, *then* $P^{tp^k} d_{n,k+1} = 0$.
(b) Let $p^{n-2} < \kappa < p^{n-1}$ *and* $\kappa \neq p^t + \cdots + p^{n-2}$ *with* $0 \leq t \leq n-2$, *then* $P^{\kappa} d_{n,n-1} = 0$.

Proof We shall provide with a simple proof following a referee's suggestion. The Steenrod algebra action commutes with the group action on $\mathbb{F}_p[y_1, \ldots, y_n]$. Hence if d is an invariant polynomial with respect to some subgroup $G \leq GL\left(n, \mathbb{F}_p\right)$, then $P^m d$ is also invariant under the same subgroup G.

(a) Let $2a$ denote the degree of $P^{tp^k} d_{n,k+1}$ which is also an element of D_n.

$$2a = 2\left(p^n - p^{k+1} + tp^k\,(p-1)\right) \text{ and}$$

$$2\sum_{0}^{n-1} a_i \left(p^n - p^i\right) = 2a \text{ for some natural numbers } a_i.$$

$$\left(p^n - p^{k+1} + 2\,(p-1)\right) \leq a \leq \left(p^n - p^{k+1} + p^{n-2}\,(p-1)^2\right).$$

Because of the restrictions $a = p^n - p^i + p^n - p^j$ with $0 \leq i, j \leq n-1$ and

$$2p^{n-1} + p^{k+1} \leq p^i + p^{n-2} + p^j.$$

If $n > 3$, then $k = n-2$ implies no solution. So let $k < n-2$, then $i = j = n-1$. But then

$$a = p^n - p^{k+1} + tp^k\,(p-1) = 2\left(p^n - p^{n-1}\right)$$

which implies that p divides t. Finally $P^{tp^k} d_{n,k+1} = 0$.

For $n = 3$, $0 \leq k \leq 1$ and $2 \leq t \leq p-1$. Then $P^t d_{3,1} = 0$ and also $P^{tp} d_{3,2} = 0$.

It follows that $P^{tp^k} d_{n,k+1} = 0$.

(b) As above $P^{\kappa} d_{n,n-1} \in D_n$ and $2a = |P^{\kappa} d_{n,n-1}| = 2\left(p^n - p^{n-1} + k\,(p-1)\right)$ such that

$$2\left(p^n - p^{n-1} + p^{n-2}\,(p-1)\right) < 2a < 2\left(p^n - p^{n-1} + p^{n-1}\,(p-1)\right) \text{ or}$$

$$\left(p^n - p^{n-2}\right) < a < 2\left(p^n - p^{n-1}\right) \text{ and}$$

$$a \neq p^n - p^t \text{ for } 0 \leq t \leq n-2.$$

Since $a = \sum_{0}^{n-1} a_i \left(p^n - p^i\right)$ for some natural numbers a_i, the claim follows.

The following theorem relates to the hom-dual of $R[n]$ with the Dickson algebras.

Theorem 10.8 ([29, 44] $p = 2$, [22] $p =$ odd) *There exists an* $\mathscr{A}$*-algebra isomorphism*

$$T_n : SED_n \to R[n]^*$$

given by $T_n\left(d_{n,i+1}\right) = \left(Q^{I_{n,i}}\right)^*$, $T_n\left(M_{n;i}L_n^{p-2}\right) = \left(Q^{J_{n,i}}\right)^*$, *and* $T_n\left(M_{n;s,i}L_n^{p-2}\right) = \left(Q^{K_{n;s,i}}\right)^*$.

The action of the Steenrod algebra on Mui generators is given in Proposition 23, p. 850 in [25].

Theorem 10.9 *(a) The value of* P^{P^i} *on* $M_{n;s_1,\ldots,s_k}L_n^{p-2}$ *is as follows*

$$\begin{cases} M_{n;s_1,\ldots,s_{j-1},\ldots,s_k}L_n^{p-2}, & \text{if } i = s_j - 1 \text{ and } s_{j-1} \neq s_j - 1; \\ 0, & \text{if } i = s_j - 1 = s_j - 1 \text{ or } i - 1 \neq s_j, n; \\ -L_n^{p-2}\left(M_{n;s_1,\ldots,s_k}d_{n,n-1} + \sum_{i=1}^{k}(-1)^{k+i}M_{n;s_1,\ldots,\hat{s_i},\ldots,s_k,n-1}d_{n,s_i}\right), & \\ & \text{if } i = n-1 \text{ and } s_k < n-1; \\ (p-2)M_{n;s_1,\ldots,s_k}L_n^{p-2}d_{n,n-1}, & \text{if } i = n-1 \text{ and } s_k = n-1. \end{cases}$$

(b) The value of β *on* $M_{n;s_1,\ldots,s_k}L_n^{p-2}$ *is as follows*

$$\begin{cases} (-1)^{k-1}M_{n;s_2,\ldots,s_k}L_n^{p-2}, & \text{if } s_1 = 0; \\ 0, & \text{otherwise.} \end{cases}$$

We express the relations we are about to define on our free unstable module in the next corollary. The proof consists of applications of Theorems 10.9 and 10.7.

Corollary 10.3 *Let* $M = M_{n;n-2,n-1}L_n^{(p-2)}$.

1. $\beta M = 0 = P^{p^k}M$, *if* $k \neq n-3$ *and* $k < n-1$.
2. $P^{p^{n-3}}P^{p^{n-3}}M = 0$.
3. $P^{p^{n-3}}P^{p^{n-1}}M = P^{p^{n-1}}P^{p^{n-3}}M$.
4. $P^{p^{n-2}}P^{p^{n-2}}P^{p^{n-3}}M = 0 = P^{p^{n-3}}P^{p^{n-2}}P^{p^{n-3}}M$.
5. $\left(\beta P^{(0,\ldots,n-2)}\right)\left(\beta P^{(0,\ldots,n-3)}\right)M = (-1)\,P^{(0,\ldots,n-2)}d_{n,n-1}$.

Here $P^{(0,\ldots,k)}$ *stands for* $P^1\ldots P^{p^k}$.

10.5 SED_n as a Quotient of a Free Unstable Algebra

In this section we shall define an unstable $\mathscr{A}$-module $\mathscr{M}(\mu, u)$ and from it an unstable $\mathscr{A}$-algebra $\mathscr{Q}(\mu, u)$. Finally an isomorphism between $\mathscr{Q}(\mu, u)$ and SED_n will be defined. We extract results from [24].

Let us recall the definition of a free unstable $\mathscr{A}$-algebra from [56]. We need some definitions. Recall an admissible monomial P^I in the Steenrod algebra from Definition 10.6.

Definition 10.7 The excess of an admissible sequence I is defined to be

$$(i_1 - 2i_2) + (i_2 - 2i_3) + \cdots + (i_{n-1} - 2i_n) + i_n \text{ if } p = 2;$$

$$2(i_1 - pi_2) + 2(i_2 - pi_3) + \cdots + 2i_n + \varepsilon_0 - \varepsilon_1 - \cdots \varepsilon_n \text{ if } p > 2.$$

and it is denoted by $e(I)$.

An $\mathscr{A}$-module M is called **unstable**, if $\beta^\varepsilon P^i x = 0$ for $2i + \varepsilon > |x|$ and $\varepsilon \in \{0, 1\}$. Recall the definition of the suspension functor $\Sigma : \mathscr{M} \to \mathscr{M}$ in the category of $\mathbb{Z}$-graded $\mathscr{A}$-modules and morphisms being $\mathscr{A}$-linear maps of degree zero. Given an $\mathscr{A}$-module M, ΣM is defined by $(\Sigma M)^m \cong M^{m-1}$ and the $\mathscr{A}$-action is given by $\theta(\Sigma m) = (-1)^{|\theta|} \Sigma \theta m$ for $m \in M$ and $\theta \in \mathscr{A}$.

Definition 10.8 $F(n)$ is called a free unstable $\mathscr{A}$-module on one n-dimensional generator, if $F(n) \cong \Sigma^n(\mathscr{A}/B(n))$. Here $B(n)$ denotes the subspace generated by all admissible monomials of excess greater than n in $\mathscr{A}$. $B(n)$ is a left ideal.

A free unstable $\mathscr{A}$-module is the direct sum of free unstable $\mathscr{A}$-modules on one generator.

Following Steenrod [56] we recall the definition of a free unstable A-algebra.

Definition 10.9 Let M be an $\mathscr{A}$-module. A free $\mathscr{A}$-algebra generated by M, denoted by VM, is the quotient of the tensor algebra $T(M)$ on M by the ideal of $T(M)$ generated by all elements of the form

$$x \otimes y - (-1)^{|x||y|} y \otimes x \text{ and } P^n x = x^{\otimes p} \text{ for } 2n = |x|.$$

If M is a free unstable $\mathscr{A}$-module, then VM is called the completely free unstable $\mathscr{A}$-algebra generated by M.

Now we are ready to proceed to our object of study.

Definition 10.10 The module $\mathscr{M}(\mu, u)$ has two generators μ and u of degrees $2\left(p^n - p^{n-1} - p^{n-2}\right)$ and $2\left(p^n - p^{n-1}\right)$ respectively and relations

$$P^{p^k}\mu = 0 = P^{p^l}u, \tag{10.2}$$

$$\text{for } -1 \le k \le n-4, k = n-2 \text{ and } -1 \le l \le n-3;$$

$$P^{p^{n-3}} P^{p^{n-3}} \mu = 0 = P^{p^{n-2}} P^{p^{n-2}} u; \tag{10.3}$$

$$P^{p^{n-3}} P^{p^{n-2}} P^{p^{n-3}} \mu = 0 = P^{p^{n-2}} P^{p^{n-2}} P^{p^{n-3}} \mu; \tag{10.4}$$

$$P^{p^{n-1}} P^{p^{n-3}} \mu = P^{p^{n-3}} P^{p^{n-1}} \mu \text{ and } P^{p^{n-2}} P^{p^{n-1}} u = 2P^{p^{n-1}} P^{p^{n-2}} u. \tag{10.5}$$

$$P^{(-1,\ldots,n-2)}P^{(-1,\ldots,n-3)}\mu = P^{(0,\ldots,n-2)}u. \tag{10.6}$$

Here $P^{p^{-1}}$ or P^{-1} stands for β and $P^{(-1,\ldots,k)}$ for $\beta P^1 \cdots P^{p^k}$.

Using the definitions above we define the main object of study in this paper.

Definition 10.11 Let $\mathscr{Q}(\mu, u)$ be the free unstable $\mathscr{A}$-algebra on the module $\mathscr{M}(\mu, u)$ subject to the following relations:

$$\mu^2 = 0 \text{ and } P^{p^{n-1}}u = (p-1)u^2. \tag{10.7}$$

$$P^{p^{n-1}}\mu = (p-2)\mu u \text{ and} \tag{10.8}$$

$$P^{p^{n-1}}P^{p^{n-2}}P^{p^{n-3}}\mu = -P^{p^{n-2}}P^{p^{n-3}}\mu u + \mu P^{p^{n-3}}P^{p^{n-2}}u - P^{p^{n-3}}\mu P^{p^{n-2}}u.$$

We will now relate $\mathscr{Q}(\mu, u)$ and SED_n. We check that the action on the monomials $M_{n;n-2,n-1}L_n^{p-2}$ and $d_{n,n-1} \in SED_n$ obey the defining $\mathscr{A}$-action relations on the generators μ and $u \in \mathscr{Q}(\mu, u)$. Using Corollaries 10.2 and 10.3 we show that the map from $\mathscr{Q}(\mu, u)$ to SED_n is an epimorphism. It remains to show that it is also one to one. This is done by showing that $\mathscr{Q}(\mu, u)$ has no more algebra generators than SED_n and they satisfy the same relations. Thus we have the following.

Theorem 10.10 (Theorem 26, [25]) *The algebra $\mathscr{Q}(\mu, u)$ is isomorphic as an $\mathscr{A}$-algebra to SED_n.*

For the details see [25].

A cofree coalgebra on an algebraic object is characterized by its universal property. We consider the category of connected cocommutative positively graded coalgebras. A cofree unstable $\mathscr{A}$-coalgebra of finite type is isomorphic to the dual of a free unstable $\mathscr{A}$-algebra (Steenrod [56]).

Let $\mathscr{CA}$ be the category of unstable coalgebras i.e. an object of $\mathscr{CA}$ has both an unstable right $\mathscr{A}$-module structure and a connected cocommutative F_p-coalgebra structure subject to compatibility conditions [6]. Let $\mathscr{MA}$ be the category of connected unstable right $\mathscr{A}$-modules. We recall that an unstable right $\mathscr{A}$-module consists of a positively graded F_p-module M and a graded module map

$$\mathscr{A}^i \otimes M_n \to M_{n-2(p-1)i} \text{ with the property}$$

$$P^k m = 0, \text{ if } |m| < 2pk; \text{ and } \beta P^k m = 0, \text{ if } |m| = 2pk+1.$$

The comultiplication map in $\mathscr{CA}$ is an unstable $\mathscr{A}$-module map and the pth root map $\xi : M_{pk} \to M_k$, dual to the pth power map, satisfies

$$\xi(m) = P^k(m).$$

Here $M = \bigoplus_i M_i$ is an unstable $\mathscr{A}$-module. For example $H_*\left(X, F_p\right)$ is an object in $\mathscr{C}\mathscr{A}$ for a connected space X and its comultiplication is induced by the diagonal.

For a connected unstable $\mathscr{A}$-module M, the cofree unstable $\mathscr{A}$-coalgebra cogenerated by M, UM, has the following universal property: UM comes with an $\mathscr{A}$-module map $i : UM \to M$ and if C is an unstable $\mathscr{A}$-coalgebra and $f : C \to M$ an $\mathscr{A}$-module map, there exists a unique $\mathscr{A}$-coalgebra map $\overline{f} : C \to UM$ such that $f = i\overline{f}$. If M is of finite type, then UM is dual to the free unstable $\mathscr{A}$-algebra VM^* generated by the dual $\mathscr{A}$-module M^*. Moreover, U is a functor from the category $\mathscr{M}\mathscr{A}$ to $\mathscr{C}\mathscr{A}$ right adjoint to the forgetful functor.

We compared SED_n with a free unstable algebra on $\mathscr{M}(\mu, u)$. Now we attempt the dual comparison. Since SED_n is of finite type, one would compare $(SED_n)^*$ with a cofree unstable coalgebra on $(\mathscr{M}(\mu, u))^*$. Using the isomorphism $R[n]^* \cong SED_n$ as $\mathscr{A}$-algebras, the next corollary follows.

Corollary 10.4 (Corollary 41, [25]) *$R[n]$ is isomorphic to a subcoalgebra of a cofree unstable coalgebra on two cogenerators.*

Corollary 10.5 (Corollary 42, [25]) *Let $1 \leq n$, then $Hom_{\mathscr{C}\mathscr{A}}(R[n], R[n]) \cong \mathbb{F}_p$.*

Proof It suffices to prove the dual version of the claim:

$$Hom_{\mathscr{U}\mathscr{A}}(\mathscr{Q}(\mu, u), \mathscr{Q}(\mu, u)) \cong \mathbb{F}_p.$$

We have to consider the degree of the lowest degree μ and the degree u. Because of Theorem 10.10 and Corollary 10.3 (e), we only have to examine the image of μ. Now the claim follows.

Kechagias (unpublished [23]) and Hung [18] obtained a similar version of the last result using different approaches.

10.6 *modp* Cohomology of the Infinite Symmetric Group

There are intimate connections between the cohomology of the symmetric groups and the structure of cohomology operations. These connections were exploited and developed from about 1952–1964 in work of J. Adem, N. Steenrod, A. Dold, N. Nakaoka and others. Meanwhile, in the period from 1959–1961 E. Dyer and R. Lashof discovered a fundamental relationship between $H_*\left(B\Sigma_n; \mathbb{F}_p\right)$ and the structure of the homology of infinite loop spaces.

The original calculation of $H_*(B\Sigma_\infty)$ is due to Nakaoka [45]. A different approach due to Milgram [36], Barratt–Priddy [4], and Quillen [50] yields the homology of the infinite symmetric group as a corollary of fundamental results in homotopy theory.

The theme in this section is to provide a set of generators for the *modp* cohomology of the classifying space of the infinite symmetric group. We obtain this set by comparing $H^*(Q_0S^0)$ with a free unstable algebra on a module of Dickson invariants. We show that $H^*(Q_0S^0)$ is isomorphic to a free unstable algebra as algebras. However, this isomorphism can not preserve the action of the Steenrod algebra. Nevertheless, the description of $H^*(Q_0S^0)$ in terms of the free unstable algebra provides an explicit basis for the cohomology algebra. This is the basis we use for $H^*(B\Sigma_\infty, \mathbb{F}_p)$.

We recall that an Abelian restricted Lie algebra is an $\mathbb{F}_p$-module L together with a restriction (pth power operation) $\xi : L_n \to L_{pn}$, for pn even, such that $\xi(ad) = a^p\xi(d)$ and $\xi\left(d + d'\right) = \xi(d) + \xi\left(d'\right)$ for d, d' of even degree and $a \in \mathbb{F}_p$.

Using ED and SED (10.9) we define the corresponding Lie algebras. We follow May [31].

Definition 10.12 Let $D^\#$ be the Lie algebra with underlying graded module D and its Lie product zero. Let a restriction (according to Milnor–Moore) ξ be defined on the components $D_k^\#$ of $D^\#$ and extended naturally:

$$\xi : D_k^\# \to D_k^\#$$

by $\xi(d) = P^m d$ for $2m = |d|$.

Let $V\left(D^\#\right)$ be the universal enveloping algebra on $D^\#$ given by

$$V\left(D^\#\right) = \mathbf{A}\left(D^\#\right)/I.$$

Here I is the ideal generated by

$$\xi(d) - (d)^p \text{ for } |d| = \text{even or } p = 2.$$

Here $\mathbf{A}$ stands for the free commutative algebra on $D^\#$, $d \in D^\#$ and $(d)^p \in \mathbf{A}\left(D^\#\right)$. The $\mathscr{A}$-module structure is determined by the given action on D and the Cartan formula.

Let ED be the Dickson algebra monomials of positive degree of any length and SED its obvious submodule respectively:

$$ED = \bigoplus_{k\geq 1} ED_k^+,\ SED = \bigoplus_{k\geq 1} SED_k^+. \tag{10.9}$$

For $p = 2$, they coincide.

We consider $SED^\#$ as a sub-restricted Lie algebra of $ED^\#$ and $V\left(SED^\#\right)$ the universal enveloping algebra on $SED^\#$.

We must note that there is no relation between the original product structure in SED_n expect for the pth power and the one imposed by the definition above. Let us also note that $V\left(SED^\#\right)$ is primitively generated commutative Hopf algebra with $P\left(V\left(SED^\#\right)\right) = SED^\#$.

We note that the order defined in Definition 10.2 can be transposed to SED. We also use Definition 10.4.

Definition 10.13 Let χ be the function defined between the module basis for SED_k and the monoid $\mathbb{N}^k \times \mathbb{F}_2^k$ as follows

$$\chi\left(d_{k,i}\right) = I_{k,i}, \chi\left(M_{k;i}L_k^{p-2}\right) = J_{k,i} \text{ and } \chi\left(M_{k;s,m}L_k^{p-2}\right) = K_{k;s,m}$$

and the rule $\chi(uv) = \chi(u) + \chi(v)$. Here u and v are monomials. Via χ and the Definition 10.2 monomials in SED_k are also ordered. Moreover, using the weight (Definition 10.2) and the order defined above, monomials in SED are ordered as well.

Using the defined ordering on SED above, a basis for $V\left(SED^{\#}\right)$ is obtained as follows:

$$BV\left(SED^{\#}\right) = \{1\} \cup$$

$$\left\{(d_1)^{m_1}\ldots(d_k)^{m_k} \;\middle|\; \begin{array}{l} k \geq 1, d_i \in B(SED), d_1 < \cdots < d_k, m_i = 1 \\ \text{if } |d_i| = \text{odd and } 0 < m_i < p \text{ if } |d_i| = \text{even} \end{array}\right\}.$$

Here $B(SED)$ stands for the basis of monomials of SED.

Theorem 10.11 (Theorem 9, [24]) *There is an isomorphism of algebras*

$$e : V\left(SED^{\#}\right) \to H^*(Q_0S^0).$$

For a natural number $n \in \mathbb{N} \cong \pi_0\left(C\left(S^0\right)\right)$, $B\Sigma_n$ is the corresponding path component of $C\left(S^0\right)$. We may form $BC\left(S^0\right)$, the classifying space of $C\left(S^0\right)$, and $\Omega BC\left(S^0\right)$ the loop space of $BC\left(S^0\right)$. So there is a map

$$C\left(S^0\right) \to \Omega BC\left(S^0\right)$$

and this is a map of H-spaces. Now $\pi_0\left(\Omega BC\left(S^0\right)\right) \cong \mathbb{Z}$ is the universal group associated to monoid $\pi_0(C\left(S^0\right))$.

In general, it is well known that if X is a connected associative H-space, then X has the structure of an H-group, i.e., $X \simeq \Omega BX$. For X not connected (like $C\left(S^0\right)$) the result fails. However, one can still "adjoint" inverses to X and inquire about the homology algebra of the resulting space and its relation to the homology algebra of X.

Since $\Omega BC\left(S^0\right)$ is a grouplike H-space, its path components are all homotopy equivalent. So

$$\Omega BC\left(S^0\right) \simeq \pi_0\left(\Omega BC\left(S^0\right)\right) \times \left(\Omega BC\left(S^0\right)\right)_0.$$

Since $\pi_1(B\Sigma_\infty) \cong \Sigma_\infty$ is not abelian, $B\Sigma_\infty$ is not an H-space. Nevertheless, $H_*(B\Sigma_\infty)$ has a ring structure and that moreover it is isomorphic as a ring to $H_*\left(Q_0\left(S^0\right)\right)$ via an isomorphism induced by a map

$$B\Sigma_\infty \to Q_0 S^0.$$

Now, $H_*\left(\Omega BC\left(S^0\right)\right) \cong H_*\left(\mathbb{Z} \times B\Sigma_\infty\right)$. Moreover, $\Omega BC\left(S^0\right) \simeq Q\left(S^0\right)$ and $Q\left(S^0\right)$ is the group completion of $C\left(S^0\right)$. Please recall Theorem 10.1.

In the early 1970s this isomorphism formed the starting point for much of Quillen's work relating the classifying spaces of finite groups of Lie type to stable homotopy theory.

Let us just note that the homology of $B\Sigma_n$ injects with any untwisted coefficients k, onto a direct summand in the homology of $B\Sigma_\infty$ for each n via the homology map induced from the usual inclusion of groups. In particular

$$H_*\left(B\Sigma_\infty; k\right) \cong \coprod_1^\infty H_*\left(B\Sigma_n, B\Sigma_{n-1}; k\right).$$

Partition $n = p^{j_1} + p^{j_2} + \cdots + p^{j_r}$ with $0 \le j_1 \le j_2 \le \cdots \le j_r$. Then a homology basis for $H_*\left(B\Sigma_n, B\Sigma_{n-1}\right)$ is given by the elements

$$Q^{I(1)} \cdot Q^{I(2)} \cdot \cdots \cdot Q^{I(r)}.$$

Here $l\left(I\left(t\right)\right) = j_t$ and $Q^{I(t)}$ is an admissible element in R. The r-tuple $(j_1, \ldots, j_r)$ ranges over all partitions of n as above. Here the $\mathscr{A}$ action is given by Cartan formula and Nishida relations [4, 12, 45].

A positive integer valued function ϕ is defined on the basis of $H_*\left(B\Sigma_\infty\right)$ as follows

$$\phi\left(Q^{J(1)} \cdot Q^{J(2)} \cdot \cdots \cdot Q^{J(k)}\right) = \sum_t p^{l(J_t)}.$$

Now a filtration of $H_*\left(B\Sigma_\infty\right)$ is defined in terms of ϕ: let Φ_q be generated additively by elements $Q^{J(1)} \cdot Q^{J(2)} \cdot \cdots \cdot Q^{J(k)}$ such that

$$\phi\left(Q^{J(1)} \cdot Q^{J(2)} \cdot \cdots \cdot Q^{J(k)}\right) \ge q. \tag{10.10}$$

Hence $\Phi_{q+t} \le \Phi_q$.

In this framework an explicit basis for $H^*(B\Sigma_\infty, \mathbb{F}_p)$ can be constructed in terms of Dickson invariants. We note that the products that appear in the description of the generators, in the next theorem, are the old ones coming from the original products in the Dickson algebras.

Theorem 10.12 (Theorem 12, [24]) *(a) For $p = 2$, $H^*(B\Sigma_\infty)$ is isomorphic to the polynomial algebra generated by*

$$\left\{ \prod_{1 \le r \le k} d_{k,k-r}^{m_r} \mid (m_1, \ldots, m_k) \notin (2\mathbb{N})^k \text{ for } k \ge 1 \right\}.$$

(b) For $p > 2$, $H^*(B\Sigma_\infty)$ is isomorphic to the tensor product of exterior, polynomial, and truncated polynomial algebras at height p. (i) The exterior generators are

$$\left\{ \prod_{1\le r\le k} d_{k,k-r}^{m_r} M_{k,j} L_k^{p-2} \prod_{t=1}^{\ell} M_{k;s_{2t-1},s_{2t}}^{\varepsilon_t} L_k^{p-2} \mid m_r \in \mathbb{N},\ \varepsilon_t \in \{0,1\},\ 0 \le j < k \right\}.$$

(ii) The polynomial generators are

$$\left\{ \prod_{1\le r\le k} d_{k,k-r}^{m_r} \mid (m_1,\dots,m_k) \notin (p\mathbb{N})^k \text{ for } k \ge 1 \right\}.$$

(iii) The truncated polynomial generators are

$$\left\{ \prod_{1\le r\le k} d_{k,k-r}^{m_r} \prod_{t=1}^{\ell} M_{k;s_{2t-1},s_{2t}}^{\varepsilon_t} L_k^{p-2} \mid m_r \in \mathbb{N},\ (\varepsilon_1,\dots,\varepsilon_l) \ne \mathbf{0} \text{ for } k \ge 1 \right\}.$$

The algebra basis defined in [24] for $H^*(B\Sigma_\infty, \mathbb{F}_p)$ is more explicit than the one given by Nakaoka [45] for $p = 2$ and Adem-Milgram [2]. See also Nguyen Hung [17].

Following Wellington [59], a description of $H^*(Q_0S^0)$ by $\mathscr{A}$-algebras will be given. There exists an increasing sequence of algebras having the property that although none of these algebras is free, their associated algebras are free as unstable $\mathscr{A}$-algebras. So $H^*(Q_0S^0)$ is filtered by an increasing sequence of $\mathscr{A}$-algebras. First we need a definition.

An increasing filtration inspired by May on $B\left(V\left(SED^{\#}\right)\right)$ is given by

$$\sigma : B\left(V\left(SED^{\#}\right)\right) \to \mathbb{N} \text{ by } \sigma\left(\prod_i d_{k,i}^{m_i}\right) = p^k \text{ and extend to } BV\left(SED^{\#}\right)$$

$$\text{by } \sigma\left((d_1)^{m_1}\dots(d_n)^{m_n}\right) = \sum_i m_i \sigma\left(d_i\right).$$

We note that the Steenrod operations respect the filtrations but σ does not respect the restriction on $V\left(SED^{\#}\right)$. We quote from [24]: $Sq^1 d_{1,0} = d_{1,0}^2 \approx \left(d_{1,0}\right)^2$ (the product in $V\left(SED^{\#}\right)$), but $\sigma\left(d_{1,0}^2\right) = 2 \ne \sigma\left(\left(d_{1,0}\right)^2\right) = 2^2$.

We recall that a decreasing filtration $\left\{\Phi_q\right\}$ was defined on $H_*\left(B\Sigma_\infty\right)$ in (10.10). Here σ defines an increasing filtration Φ^q on $V\left(SED^{\#}\right)$: let Φ^q be generated additively by elements d such that $\sigma(d) \le q$. Thus $\Phi^q \le \Phi^{q+1}$ for t a natural number. We may say that $\left\{\Phi_q\right\}$ and $\{\Phi^q\}$ are "dual" to each other.

Using the filtration associated with σ above free objects are defined as follows.

Definition 10.14 Let $SED(s)$ be the $\mathscr{A}$-submodule of SED generated by the basis elements $d \in SED$ such that $\sigma(d) = s$. Thus $SED = \bigoplus SED(s)$. Let $C^k = V(\bigoplus_1^k SED^{\#}(s))$ and A^k its image in $H^*(Q_0S^0)$ under the isomorphism between $V\left(SED^{\#}\right)$ and $H^*(Q_0S^0)$ as in Theorem 10.11.

The free objects defined above can be used to approximate $H^*(Q_0S^0)$. We close this survey with the following theorem. As we mentioned in the introduction, the point of our approach is to incorporate modular invariants as computational tools. We recall that the map e as in Theorem 10.11 is not a map of $\mathscr{A}$-algebras, but it is a map of $\mathscr{A}$-algebras up to filtration as is shown in [24], Proposition 13. This is crucial in the proof of the next theorem.

Theorem 10.13 (Theorem 15, [24]) *The sequence $A^1 \subset A^2 \subset \cdots \subset H^*(Q_0S^0)$ is a complete filtration of $\mathscr{A}$-subalgebras. There exist isomorphisms of algebras $A^{s+1} \cong A^s \otimes V(SED^{\#}(s+1))$ and isomorphisms of $\mathscr{A}$-algebras $A^{s+1}/A^s \cong V(SED^{\#}(s+1))$.*

Proof Firstly, $A^{s+1} = Im\left(C^{s+1} = C^s \otimes V(SED^{\#}(s+1))\right) \to H^*(Q_0S^0)$. Thus

$$A^{s+1} \cong A^s \otimes V(SED^{\#}(s+1)) \text{ as algebras.}$$

Secondly, because each element of A^{s+1} is a product of images of elements of $SED^{\#}(t)$ with $t \leq s+1$ and the Steenrod operations are applied by the Cartan formula A^{s+1} is an $\mathscr{A}$-algebra. For the same reasons $A^{s+1}/A^s \cong V(SED^{\#}(s+1))$ as $\mathscr{A}$-algebras.

References

1. Adams, J.F.: Infinite Loop Spaces. Annals of Mathematics Studies, vol. 90. Princeton University Press, Princeton (1978) (University of Tokyo Press, Tokyo)
2. Adem, A., Milgram, R.J.: Cohomology of Finite Groups. Grundlehren der Mathematischen Wissenschaften, vol. 309, p. viii+327. Springer, Berlin (1994)
3. Adem, A., Maginnis, J.S., Milgram, R.J.: Symmetric invariants and cohomology of groups. Mathematische Annalen **287**, 391–411 (1990)
4. Barratt, M., Priddy, S.: On the homology of non-connected monoids and their associated groups. Comment. Math. Helv. **47**, 1–14 (1972)
5. Boardman, J.M., Vogt, R.M.: Homotopy Invariant Algebraic Structures on Topological Spaces. Lecture Notes in Mathematics, vol. 347. Springer, Berlin (1973), x+257 pp
6. Bousfield, A.K., Kan, D.M.: The homotopy spectral sequence of a space with coefficients in a ring. Topology **11**, 79–106 (1972)
7. Browder, W.: Homology operations and loop spaces. Ill. J. Math. **4**, 347–357 (1960)
8. Brown, E.H.: Cohomology theories. Ann. Math **75**(2), 467–484 (1962)
9. Chon, P.H.: Modular coinvariants and the *mod p* homology of QS^k. Proc. Lond. Math. Soc **112**(3) (2016); no. 2, 351–374
10. Cohen, F., Lada, T., May, J.P.: The Homology of Iterated Loop Spaces. Lecture Notes in Mathematics, vol. 533. Springer, Berlin (1976), vii+490 pp

11. Dickson, L.E.: A fundamental system of invariants of the general modular linear group with a solution of the form problem. Trans. A. M. S. **12**, 75–98 (1911)
12. Dyer, E., Lashof, R.K.: Homology of iterated loop spaces. Am. J. Math. **84**, 35–88 (1962)
13. Eilenberg, S., MacLane, S.: Relations between homology and homotopy groups of spaces. Ann. Math. **46**(2), 480–509 (1945)
14. Feshbach, M.: The $mod 2$ cohomology rings of the symmetric groups and invariants. Topology **41**(1), 57–84 (2002)
15. Giusti, G., Salvatore, P., Sinha, D.: The $mod 2$ cohomology rings of symmetric groups. J. Topol. **5**(1), 169–198 (2012)
16. Hovey, M.: Model Categories. Mathematical Surveys and Monographs, vol. 63, p. 213. A. M. S., Oxford (1999)
17. Hung, N.H.V.: The $mod 2$ cohomology algebras of symmetric groups. Jpn. J. Math. (N.S.) **13**(1), 169–208 (1987)
18. Hung, N.H.V.: The homomorphisms between the Dickson-Mui algebras as modules over the Steenrod algebra. Math. Ann. **353**(3), 827–866 (2012)
19. Jardine, J.F.: Representability theorems for presheaves of spectra. J. Pure Appl. Algebra **215**(1), 77–88 (2011)
20. Kan, D.M.: Adjoint functors. Trans. Am. Math. Soc. **87**, 294–329 (1958)
21. Kechagias, N.E.: The Steenrod algebra action on generators of rings of invariants of subgroups of $GL_n(\mathbb{F}_p)$. Proc. Am. Math. Soc. **118**(3), 943–952 (1993)
22. Kechagias, N.E.: Extended Dyer-Lashof algebras and modular coinvariants. Manuscripta Math. **84**(3–4), 261–290 (1994)
23. Kechagias, N.E.: A Steenrod-Milnor action ordering on Dickson invariants (2008), 16 pp
24. Kechagias, N.E.: Dickson invariants and a new description of $H^*(Q_0S^0;\mathbb{F}_p)$ via $H^*(B\Sigma_\infty;\mathbb{F}_p)$. J. Homotopy Relat. Struct. **8**(2), 201–229 (2013)
25. Kechagias, N.E.: The component Dyer-Lashof coalgebras as subcoalgebras of free unstable coalgebras. J. Pure Appl. Algebra **219**(4), 839–863 (2015)
26. Kudo, T., Araki, S.: On $H_*(\Omega N(S^n);\mathbb{F}_2)$. Proc. Jpn. Acad. **32**, 333–335 (1956)
27. Kudo, T., Araki, S.: Topology of H_n-spaces and H-squaring operations. Mem. Fac. Sci. Kyūsyū Univ. Ser. A. **10**, 85–120 (1956)
28. Lima, E.L.: The Spanier-Whitehead duality in new homotopy categories. Summa Brasil. Math. **4**(1959), 91–148 (1959)
29. Madsen, I.: On the action of the Dyer-Lashof algebra in $H_*(G)$. Pac. J. Math. **60**(1), 235–275 (1975)
30. Madsen, Ib., Milgram, R.J.: The Classifying Spaces for Surgery and Cobordism of Manifolds. Annals of Mathematics Studies, vol. 92. Princeton University Press, Princeton (1979), xii+279 pp (University of Tokyo Press, Tokyo)
31. May, J.P.: Some remarks on the structure of Hopf algebras. Proc. A. M. S. **23**(3), 708–713 (1969)
32. May, J.P.: Categories of Spectra and Infinite Loop Spaces. Lecture Notes in Mathematics, vol. 99, pp. 448–479. Springer, Berlin (1969)
33. May, J.P.: The Geometry of Iterated Loop Spaces. Lectures Notes in Mathematics, vol. 271. Springer, Berlin (1972), viii+175 pp
34. May, J.P., Thomason, R.: The uniqueness of infinite loop space machines. Topology **17**(3), 205–224 (1978)
35. Milgram, R.J.: Iterated loop spaces. Ann. Math. **84**(2), 386–403 (1966)
36. Milgram, R.J.: The $mod 2$ spherical characteristic classes. Ann. Math. **92**(2), 238–261 (1970)
37. Milnor, J.: Construction of universal bundles. II. Ann. Math. **63**(2), 430–436 (1956)
38. Milnor, J.: The Steenrod algebra and its dual. Ann. Math. **67**(2), 150–171 (1958)
39. Milnor, J.: On the cobordism ring Ω^* and a complex analogue. I. Am. J. Math. **82**, 505–521 (1960)
40. Milnor, J.: On axiomatic homology theory. Pac. J. Math. **12**, 337–341 (1962)
41. Milnor, J.W., Moore, J.C.: On the structure of Hopf algebras. Ann. Math. **81**, 211–264 (1965)

42. Mosher, R.E., Tangora, M.C.: Cohomology Operations and Applications in Homotopy Theory. Harper & Row, Publishers, New York (1968), x+214 pp
43. Mùi, H.: Modular invariant theory and the cohomology algebras of the symmetric groups. J. Fac. Sci. Univ. Tokyo **IA**, 319–369 (1975)
44. Mùi, H.: Homology operations derived from modular coinvariants. Algebraic Topology, Gö ttingen 1984. Lecture Notes in Mathematics, vol. 1172, pp. 85–115. Springer, Berlin (1985)
45. Nakaoka, M.: Homology of the infinite symmetric group. Ann. Math. **73**(2), 229–257 (1961)
46. Nishida, G.: Cohomology operations in iterated loop spaces. Proc. Jpn. Acad. **44**, 104–109 (1968)
47. Pengelley, D.J., Williams, F.: The global structure of odd-primary Dickson algebras as algebras over the Steenrod algebra. Math. Proc. Camb. Philos. Soc. **136**(1), 67–73 (2004)
48. Pengelley, D.J., Peterson, F.P., Williams, F.: A global structure theorem for the mod 2 Dickson algebras, and unstable cyclic modules over the Steenrod and Kudo-Araki-May algebras. Math. Proc. Camb. Philos. Soc. **129**(2), 263–275 (2000)
49. Quillen, D.: The spectrum of an equivariant cohomology ring. I, II. Ann. Math. **94**(2), 549–572 (1971); **94**(2), 573–602 (1971)
50. Ravenel, D.C., Wilson, W.S.: The Hopf ring for complex cobordism. J. Pure Appl. Algebra **9**(3), 241–280 (1976/1977)
51. Segal, G.: Configuration-spaces and iterated loop-spaces. Invent. Math. **21**, 213–221 (1973)
52. Serre, J.-P.: Homologie singuli ére des espaces fibrés. Ann. Math. **54**(2), 425–505 (1951)
53. Serre, J.-P.: Groupes d'homotopie et classes de groupes abéliens. Ann. Math. **58**(2), 258–294 (1953)
54. Stasheff, J.: Homotopy associativity of H-spaces. I. Trans. Am. Math. Soc. **108**, 275–292 (1963)
55. Steenrod, N.E.: Products of Cocycles and Extensions of Mappings. Annals of Mathematics. Second Series, vol. 48, pp. 290–320 (1947)
56. Steenrod, N.E., Epstein, D.B.A.: Cohomology Operations, vol. 50. Princeton University Press, Princeton (1962)
57. Thom, R.: Quelques propriét és globales des variétés différentiables. Comment. Math. Helv. **28**, 17–86 (1954)
58. Turner, P.R.: Dickson coinvariants and the homology of QS^0. Math. Z. **224**(2), 209–228 (1997)
59. Wellington, R.J.: The Unstable Adams Spectral Sequence for Free Iterated Loop Spaces. Memoirs A.M.S., vol. 258 (1982)
60. Whitehead, G.W.: Generalized homology theories. Trans. Am. Math. Soc. **102**, 227–283 (1962)
61. Whitehead, G.W.: Fifty years of homotopy theory. Bull. Am. Math. Soc. (N.S.) **8**(1), 1–29 (1983)

Part II
Algebraic Modeling of Applications

Chapter 11
Linking in Systems with One-Dimensional Periodic Boundaries

Kenneth C. Millett and Eleni Panagiotou

Abstract With a focus on one-dimensional periodic boundary systems, we describe the application of extensions of the Gauss linking number of closed rings to open chains and, then, to systems of such chains via the periodic linking and periodic self-linking of chains. These lead to the periodic linking matrix and its associated eigenvalues providing measures of entanglement that can be applied to complex systems. We describe the general one-dimensional case and applications to one-dimensional Olympic gels and to tubular filamental structures.

11.1 Introduction

The objective of this report is to describe the application of the Gauss linking number [1] to collections of open chains in models of filamental systems that employ periodic boundary conditions (PBC). These enable one to define periodic linking and periodic self-linking numbers [2, 3] that quantify the linking between pairs of filaments and, thereby, define the periodic linking matrix. The information they provide has been studied in several one-dimensional PBC models, for example: general systems such as polymer gels [4], Olympic gels [5], and filamental structures in a long tube [6], see Fig. 11.1.

In the next section we describe the Gauss linking and self-linking numbers, one-dimensional periodic boundary condition models, the extension to periodic linking and self linking, and the definition of the periodic linking matrix whose eigenvalues quantify the extent of entanglement in the systems to which they are applied. We

K.C. Millett (✉) · E. Panagiotou
Department of Mathematics, University of California, Santa Barbara, CA, USA
e-mail: millett@math.ucsb.edu

E. Panagiotouat
e-mail: panagiotou@math.ucsb.edu

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_11

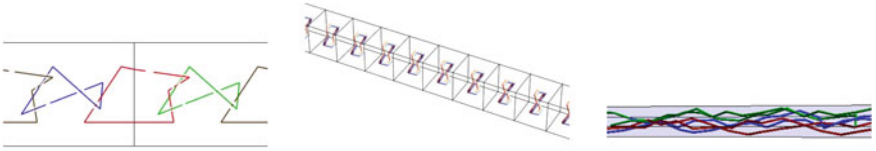


Fig. 11.1 A general 1PBC example, an Olympic system, and filamental structure in a tube

will then describe instances in which the periodic linking matrix provides important information in the study of general systems such as polymer gels, Olympic gels, and filamental structures such as vortex flow lines in tubes.

11.2 Gauss Linking, Periodic Boundary Condition (PBC) Models, Periodic Linking, and the Periodic Linking Matrix

11.2.1 Gauss Linking and Self-linking

The linking number between two oriented chains, l_1 and l_2, is defined using parameterizations of the chains, $\gamma_1(t)$ *and* $\gamma_2(s)$, via the Gauss linking integral:

Definition 11.1 The Gauss *linking number* of two disjoint (closed or open) oriented curves l_1 and l_2, whose arc-length parametrizations are $\gamma_1(t)$, $\gamma_2(s)$ respectively, is defined as a double integral over l_1 and l_2 [1]:

$$L(l_1, l_2) = \frac{1}{4\pi} \int_{[0,1]} \int_{[0,1]} \frac{(\dot{\gamma}_1(t), \dot{\gamma}_2(s), \gamma_1(t) - \gamma_2(s))}{||\gamma_1(t) - \gamma_2(s)||^3} dtds, \tag{11.1}$$

where $(\dot{\gamma}_1(t), \dot{\gamma}_2(s), \gamma_1(t) - \gamma_2(s))$ is the triple product of the derivatives, $\dot{\gamma}_1(t)$ and $\dot{\gamma}_2(s)$, and of the difference $\gamma_1(t) - \gamma_2(s)$.

Definition 11.2 (*Self-linking number*) Let l denote a chain, parameterized by $\gamma(t)$, then the *self-linking number* of l is defined as:

$$\begin{aligned} Sl\,(l) = &\frac{1}{4\pi} \int_{[0,1]^*} \int_{[0,1]^*} \frac{(\dot{\gamma}\,(t)\,, \dot{\gamma}\,(s)\,, \gamma\,(t) - \gamma\,(s))}{||\gamma\,(t) - \gamma\,(s)\,||^3} dtds \\ &+ \frac{1}{2\pi} \int_{[0,1]} \frac{(\dot{\gamma}(t), \ddot{\gamma}(t), \dddot{\gamma}(t))}{||\dot{\gamma}(t) \times \ddot{\gamma}(t)||^2} dt. \end{aligned} \tag{11.2}$$

where $\dot{\gamma}(t)$, $\ddot{\gamma}(t)$, and $\dddot{\gamma}(t)$ are the first, second, and third derivatives of $\gamma(t)$, respectively, and $(\dot{\gamma}(t), \ddot{\gamma}(t), \dddot{\gamma}(t))$ is their triple product.

The self-linking number consists of two terms, the first being the Gauss integral and the second being the total torsion of the curve.

11.2.2 One-Dimensional Periodic Boundary Conditions

The underlying structure of the *Periodic Boundary Condition*, PBC, model employed in this study consists of a cube or a solid right cylinder whose x length is one and whose y and z coordinates lie within the unit square or, respectively, in a disc of radius $a > 0$. The three-dimensional body contains a collection of arcs whose endpoints either lie in the interior or intersect the $x = 0$ or $x = 1$ faces under the constraint that the pattern on both faces is identical, see Fig. 11.1. The latter condition allows one to create and infinite structure by taking the union of integer translates of the cells and taking the unions of the resulting one-chains to define a collection of one dimensional chains. As, in general, these chains may be non-compact, we will require that each chain has precisely the same number of edges, N, thereby imposing one aspect of homogeneity. Due to the PBC structure, there is also a large scale homogeneity in the collection of chains.

11.2.3 Periodic Linking and Self-linking

In a PBC model, each chain is translated to give an infinite collection copies of itself. As a consequence, one is faced with quantifying the linking of one chain, l_0 with infinitely many translation copies of itself, $I_v = I_0 + \mathbf{v}$, or with infinitely many copies of another chain, $J_v = J_0 + \mathbf{v}$. This is achieved by employing Panagiotou's periodic linking and self-linkings numbers described next. In the periodic system we define linking at the level of free chains (i.e. the collection of translation copies of a chain, I_0; see [2] for a discussion of the motivation for this definition). The underlying idea is to calculate the linking between a chain in one free chain with all the chains in the other free chain.

Definition 11.3 (*Periodic linking number*) Let I and J denote two (closed, open or infinite) free chains in a periodic system. Suppose that I_0 is an image of the free chain I in the periodic system. The *periodic linking number*, LK_P, between two free chains I and J is defined as:

$$LK_P(I, J) = \sum_{\mathbf{v} \neq \mathbf{0}} L(I_0, J_0 + \mathbf{v}), \tag{11.3}$$

where the sum is taken over all the images of the free chain J in the periodic system.

The periodic linking number has the following properties with respect to the structure of the cell, see [2], which follow directly by its definition:

(i) The infinite sum defining LK_P converges, i.e. LK_P makes sense mathematically.
(ii) LK_P captures all the linking that all the images of a free chain impose to an image of the other.
(iii) LK_P is independent of the choice of the image I_0 of the free chain I in the periodic system.
(iv) LK_P is independent of the choice, the size and the shape of the generating cell.
(v) LK_P is symmetric.

The quantification of the linking of a free chain with itself is a bit special and requires a bit more care as there are two contributing cases, the linking of a chain with itself and the linking of a chain with translations of itself. As a consequence, one is led to the following definitions [2]:

Definition 11.4 (*Periodic self-linking number*) Let I denote a free chain in a periodic system and let I_0 be an image of I, then the *periodic self-linking number* of I is defined as:

$$SL_P(I) = Sl(I_0) + \sum_{v \neq u} L(I_0, I_v), \tag{11.4}$$

where the index v runs over all the images of I, except I_0, in the periodic system.

As with the periodic linking number, the mathematical proof of the existence of this quantity and its properties are proved in [2].

11.2.4 Periodic Linking Matrix

In order to analyze the linking entanglement present in our PBC system, L, consisting of a finite number of free chains, $l_1, l_2, ..., l_n$, we employ an $n\ x\ n$ real symmetric matrix, LM_C, whose i, jth entry is defined by equation

$$\begin{aligned} LM_{C_{i,i}} &= SL_P(l_i) \\ LM_{C_{i,j}} &= LK_P(l_i, l_j) \end{aligned} \tag{11.5}$$

In the case of a single generating chain, l, the periodic linking matrix consists of a single entry, the periodic self-linking number, $Sl(l)$. From the definition, there are two contributing factors, the self-linking given by the Eq. (11.2) and the linking between distinct copies, reflecting distinct features of periodic self-linking.

For systems with two independent chain types, the periodic linking matrix adds entanglement information due to the linking between the two distinct chains. Associated to the periodic linking matrix are two real eigenvalues, $e_1(L)$ and $e_2(L)$, given in decreasing order. The larger of these, $e_1(L)$ is proposed as the dominant characterization of the linking entanglement of the PBC system. The set of eigenvalues is the *periodic linking spectrum* of the system.

Similarly, for systems with n independent chain types, one defines the periodic linking matrix, LM_C. The associated ordered collection of eigenvalues, $e_1(L), \ldots, e_n(L)$ define the spectrum of the PBC system.

11.3 General Systems, Olympic Systems, Tubular Systems

11.3.1 General Systems

For computational efficiency, only a small portion of the physical system is simulated and periodic boundary constraints are used to avoid boundary effects. The size of the simulation cell may influence the results of a computational experiment. We examine how the periodic linking matrix changes with respect to the size of the simulation cell.

By concatenating m cells we obtain a larger cell that we denote mC, which applies PBC to the chains that touch its faces in the x-direction. We can concatenate cells of the type mC by gluing their x-faces with respect to the PBC, in order to create the same periodic system that is generated by the cell C. In this section we study the periodic linking matrix of a periodic system as the size of the cell used for its simulation, characterized by m, increases. We will see that the linking matrix depends on the size of the cell used for the simulation of a system. Since the periodic system simulated is the same, one would expect the periodic linking matrix to retain certain entanglement information. However, we will see that in a topological sense, these systems are different. With our study we extract entanglement information that is invariant of the cell size as well as information that depends on it.

Let C denote a cell composed by n generating chains, and let LM_C denote the corresponding periodic linking matrix of size $n \times n$. Without loss of generality we will concatenate cells always to the positive direction of the x-axis. Let mC denote the cell that results by gluing m copies of C respecting the PBC. Then mC has more chains. More precisely:

Lemma 11.1 *Let C be a cell with n generating chains. Then the cell mC that results by gluing m copies of C respecting the PBC, has mn generating chains.*

Remark 11.1 The different generating chains in mC generate different free chains in the periodic system. We denote the free chains in mC generated by $i^{(j)}$, $j = 0, \ldots, m-1$, as $I^{(j)} = I^{(0)} + \mathbf{v}_j$.

Thus the corresponding periodic linking matrix, LM_{mC} has size $mn \times mn$. Indeed, the cells C and mC describe different topological objects. If we identify the faces of the cell, then we will get an n-component link in the solid torus in the first case and a mn-component link in the second case. The 3-manifolds are the same in both cases even though the links that they contain are different, related by an m-fold covering space of the second manifold over the first. So, we notice that the linking matrices

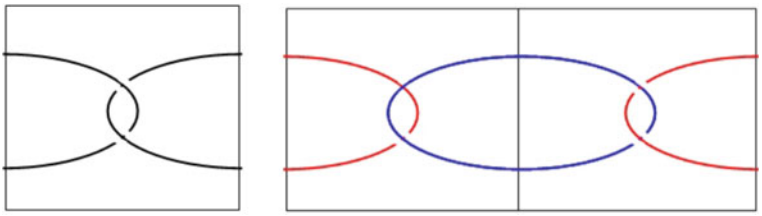


Fig. 11.2 One chain in a system with one PBC. Left: the original cell C contains 1 generating chain. Right: the cell $2C$ contains 2 generating chains

LM_C and LM_{mC} are different, but the periodic system that the cells generate and whose entanglement we wish to measure, is the same. For this purpose, we will study the dependence of the periodic linking matrix on the cell size and we will look for quantities that remain invariant of cell size.

In the following we will prove that some of the eigenvalues of the periodic linking matrix are independent of cell size. First we will study the simplest case of the periodic linking matrix of a single chain in a cell with one PBC. Next, we will generalize this to the case of n chains in a cell with one PBC. This case will facilitate the understanding of the general case of systems employing one PBC. The methods presented here can also be used to obtain similar results in 2 and 3 PBC.

We will next study the case of a cell with one PBC that contains one generating chain that unfolds in k cells. The periodic linking matrix of that system has size 1×1, $LM_C = SL_P(I) = Sl(I_0) + \sum_i L(I_0, I_i)$.

If we concatenate m cells to create a larger cell mC, then by Lemma 11.1 there are m generating chains in $k_1 C$, we denote $I^{(0)}$, $I^{(1)} = I^{(0)} + (1, 0, 0), \ldots, c$, $I^{(m)} = I^{(0)} + (m, 0, 0)$ (see Fig. 11.2). The linking matrix for this cell has size $m \times m$ and has the following form:

$$LM_{mC} = \begin{bmatrix} SL_P(I^{(0)}) & LK_P(I^{(0)}, I^{(1)}) & \ldots & LK_P(I^{(0)}, I^{(m-1)}) \\ LK_P(I^{(0)}, I^{(1)}) & SL_P(I^{(1)}) & \ldots & LK_P(I^{(1)}, I^{(m-1)}) \\ \ldots & \ldots & \ldots & \ldots \\ LK_P(I^{(0)}, I^{(m-1)}) & LK_P(I^{(1)}, I^{(m-1)}) & \ldots & SL_P(I^{(m-1)}) \end{bmatrix} \tag{11.6}$$

Lemma 11.2 *Let C denote a cell with one PBC that consists of only one chain, I. Let mC denote the cell that results after gluing m copies of C, then LM_{mC} is a symmetric centrosymmetric matrix.*

Proposition 11.1 *Let I denote a chain in a cell C with one PBC. Let mC denote the cell that results after gluing m copies of C. Then the jth eigenvalue of LM_{mC} is given by:*

$$\lambda_j = SL_P(I^{(0)}) + 2 \sum_{k=1}^{\frac{m-1}{2}} LK_P(I^{(0)}, I^{(k)}) \cos\left(\frac{2\pi}{m} k(j-1)\right) \tag{11.7}$$

for m odd and

$$\begin{aligned}\lambda_j &= SL_P(I^{(0)}) + (-1)^{(j-1)} LK_P(I^{(0)}, I^{(\lfloor \frac{m-1}{2} \rfloor + 1)}) \\ &+ 2 \sum_{k=1}^{\lfloor \frac{m-1}{2} \rfloor} LK_P(I^{(0)}, I^{(k)}) \cos\left(\frac{2\pi}{m} k(j-1)\right)\end{aligned} \tag{11.8}$$

for m even.

Remark 11.2 (i) λ_1 is independent of cell-size, m and $\lambda_1 = SL_P(I)$ for all m.

(ii) There are at most $1 + \lfloor \frac{m-1}{2} \rfloor$ distinct eigenvalues, as expected for real circulant matrices [7]. Indeed, notice that $\sin^2\left(\frac{2\pi(j-1)}{m}\right) = \sin^2\left(\frac{2\pi(m-(j-1))}{m}\right)$ and $\cos\left(\frac{2\pi(j-1)}{m}\right) = \cos\left(\frac{2\pi(m-(j-1))}{m}\right)$. Therefore, $\lambda_j = \lambda_{m-j+2}$ for all $j > 1$.

(iii) For closed chains and for $m > 2|mu(I_0)|$, the jth eigenvalue of the linking matrix has a simpler formula, namely:

$$\lambda_j = Sl(I_0) + 2 \sum_{k=1}^{\frac{m-1}{2}} L(I_0, I_0 + (k, 0, 0)) \cos\left(\frac{2\pi}{m} k(j-1)\right) \tag{11.9}$$

for m odd and

$$\begin{aligned}\lambda_j &= Sl(I_0) + (-1)^{(j-1)} L(I_0, I_0 + \left(\left(\lfloor \frac{m-1}{2} \rfloor + 1\right), 0, 0\right) \\ &+ 2 \sum_{k=1}^{\lfloor \frac{m-1}{2} \rfloor} L(I_0, I_0 + (k, 0, 0)) \cos\left(\frac{2\pi}{m} k(j-1)\right)\end{aligned} \tag{11.10}$$

for m even.

Remark 11.3 The difference between the first two eigenvalues of LM_{mC} is:

$$\begin{aligned}\lambda_1 - \lambda_2 &= SL_P(I^{(0)}) + 2 \sum_{k=1}^{\frac{m-1}{2}} LK_P(I^{(0)}, I^{(k)}) - SL_P(I^{(0)}) \\ &- 2 \sum_{k=1}^{\frac{m-1}{2}} LK_P(I^{(0)}, I^{(k)}) \cos\left(\frac{2\pi}{m} k\right) \\ &= 2 \sum_{k=1}^{\frac{m-1}{2}} LK_P(I^{(0)}, I^{(k)}) \left(1 - \cos\left(\frac{2\pi}{m} k\right)\right)\end{aligned} \tag{11.11}$$

for m odd and

$$\lambda_1 - \lambda_2 = SL_P(I^{(0)}) + LK_P(I^{(0)}, I^{(\lfloor \frac{m-1}{2} \rfloor + 1)}) + 2 \sum_{k=1}^{\lfloor \frac{m-1}{2} \rfloor} LK_P(I^{(0)}, I^{(k)}) - SL_P(I^{(0)})$$
$$- LK_P(I^{(0)}, I^{(\lfloor \frac{m-1}{2} \rfloor + 1)}) - 2 \sum_{k=1}^{\lfloor \frac{m-1}{2} \rfloor} LK_P(I^{(0)}, I^{(k)}) \cos\left(\frac{2\pi}{m} k\right)$$
$$= 2 \sum_{k=1}^{\lfloor \frac{m-1}{2} \rfloor} LK_P(I^{(0)}, I^{(k)}) \left(1 - \cos\left(\frac{2\pi}{m} k\right)\right) \tag{11.12}$$

for m even.

The above formula shows that the difference between the first eigenvalues does not depend on the self-linking number of the chain. The formula indicates that the difference, which is a measure of the homogeneity of the entanglement, is a weighted function of the linking numbers of the chain with its images. Interestingly, for large m, the linking with the nearest images contributes less than the linking with further images.

Remark 11.4 Often in applications one is interested in the average properties of filaments. Cancellations may occur when using the Gauss and periodic linking number. For this reason, one may want to use the absolute values of all the entries of the periodic linking matrix, we call the *absolute periodic linking matrix*. The absolute periodic linking matrix is also symmetric centrosymmetric. Lower bounds on the maximum eigenvalue of nonnegative real symmetric centrosymmetric matrices can be found in [7].

Next, we will extend our previous results to the case of n chains in a system with one PBC.

Let us consider n chains, say $H1, H2, \ldots, Hn$ in a system with one PBC that unfold in $k_i, i = 1, \ldots, n$ cells each. The periodic linking matrix of that system has size $n \times n$,

$$LM_C = \begin{bmatrix} SL_P(H1) & LK_P(H1, H2) & \ldots & LK_P(H1, Hn) \\ LK_P(H1, H2) & SL_P(H2) & \ldots & LK_P(H2, Hn) \\ \ldots & & & \\ LK_P(H1, Hn) & LK_P(H2, Hn) & \ldots & SL_P(Hn) \end{bmatrix} \tag{11.13}$$

Then the matrix LM_{mC} has size $mn \times mn$, since to each free chain, Hj, of the cell C, correspond m free chains, $Hj^{(i)}$, $i = 0, \ldots, m-1$, in the cell mC (see Lemma 11.1) (see Fig. 11.3). We make the convention that the uth row of LM_{mC}, where $u = rm + l$ corresponds to the free chain $H(r+1)^{(l-1)}$. Therefore, the (q, w)th element of LM_{mC}, where $q = q_1 m + q_2$, $w = w_1 m + w_2$, is: $LK_P(H(q_1 + 1)^{(q_2-1)}, H(w_1 + 1)^{(w_2-1)})$.

Proposition 11.2 *Let C denote a cell with one PBC that consists of n chains. Let mC denote the cell that results after gluing m copies of C, then LM_{mC} can be expressed*

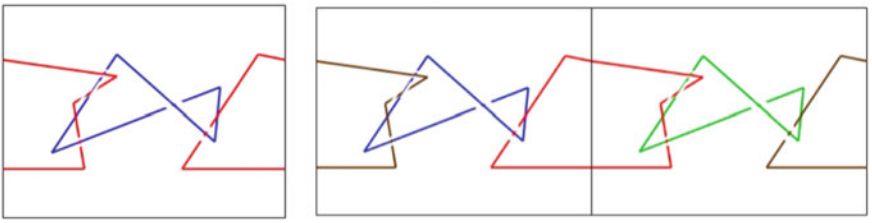


Fig. 11.3 2 chains in a system with one PBC. Left: the original cell C contains 2 generating chains. Right: the cell $2C$ contains 4 generating chains

as an $n \times n$ block matrix of $m \times m$ symmetric circulant matrices. Moreover, the diagonal block matrices are symmetric centrosymmetric matrices. The eigenvalues of the (i,i)th block of LM_{mC}, $i = 1, \ldots, n$, are:

$$\lambda_s = SL_P(Hi^{(0)}) + 2\sum_{k=1}^{\frac{m-1}{2}} LK_P(Hi^{(0)}, Hi^{(k)}) \cos\left(\frac{2\pi}{m}k(s-1)\right) \tag{11.14}$$

for m odd and

$$\begin{aligned}\lambda_s = {} & SL_P(Hi^{(0)}) + (-1)^{(s-1)} LK_P(Hi^{(0)}, Hi^{(\lfloor\frac{m-1}{2}\rfloor+1)}) \\ & + 2\sum_{k=1}^{\lfloor\frac{m-1}{2}\rfloor} LK_P(Hi^{(0)}, Hi^{(k)}) \cos\left(\frac{2\pi}{m}k(s-1)\right)\end{aligned} \tag{11.15}$$

for m even, $s = 1, \ldots, m$.

The eigenvalues of the (i,j)th block of LM_{mC}, $1 \le i < j \le n$, are:

$$\lambda_s = LK_P(Hi^{(0)}, Hj^{(0)}) + 2\sum_{k=1}^{\frac{m-1}{2}} LK_P(Hi^{(0)}, Hj^{(k)}) \cos\left(\frac{2\pi}{m}k(s-1)\right), \tag{11.16}$$

for m odd and

$$\begin{aligned}\lambda_s = {} & LK_P(Hi^{(0)}, Hj^{(0)}) + (-1)^{(s-1)} LK_P(Hi^{(0)}, Hj^{(\lfloor\frac{m-1}{2}\rfloor+1)}) \\ & + 2\sum_{k=1}^{\lfloor\frac{m-1}{2}\rfloor} LK_P(Hi^{(0)}, Hj^{(k)}) \cos\left(\frac{2\pi}{m}k(s-1)\right)\end{aligned} \tag{11.17}$$

for m even, $s = 1, \ldots, m$.

Notice that in the case of n chains in a system with 1 PBC the periodic linking matrix is no longer a circulant matrix and its eigenvalues are not known. However, the eigenvalues of its block matrices are known. More precisely, LM_{mC} can be expressed as

$$LM_{mC} = \Sigma M + \Lambda M \tag{11.18}$$

where ΣM, ΛM are $m \times m$ block matrices. ΣM is a diagonal block matrix, whose blocks represent the linking of a chain Hi with its own images and are symmetric centrosymmetric. ΛM is a block matrix whose diagonal matrices are zero and its off-diagonal matrices represent the linking between different generating chains, and are symmetric circulant matrices

The following proposition shows that some of the eigenvalues of the periodic linking matrix are invariant of cell-size, m.

Proposition 11.3 *Let LM_C be the periodic linking matrix of a periodic system generated by the cell C with one PBC, which contains n chains. Then any other periodic linking matrix LM_{mC} of the same periodic system generated by the cell mC is of the form*

$$LM_{mC} = \begin{bmatrix} LM_C & E \\ 0 & F \end{bmatrix} \tag{11.19}$$

where E has size $1 \times (m-1)$ and F has size $(m-1) \times (m-1)$.

Remark 11.5 From Proposition 11.3 it follows that the eigenvalues of LM_C are among the eigenvalues of LM_{mC}, for all m.

11.3.2 Olympic Systems

Olympic systems are collections of small ring polymers whose aggregate properties are largely characterized by the extent (or absence) of topological linking in contrast with the topological entanglement arising from physical movement constraints associated with excluded volume contacts or arising from chemical bonds. These were first discussed by de Gennes [8] and have been of interest ever since due to their particular properties and their occurrence in natural organisms, for example as intermediates in the replication of circular DNA in the mitochondria of malignant cells or in the kinetoplast DNA networks of trypanosomes [9–15]. In this project, we studied systems that have an intrinsic one, two, or three dimensional character and consist of large collections of ring polymers modeled using periodic boundary conditions. In this report we will focus on the one-dimensional facets of these structures, see Fig. 11.4. We identified and discussed the evolution of the dimensional character of the large scale topological linking as a function of density. We identified the critical

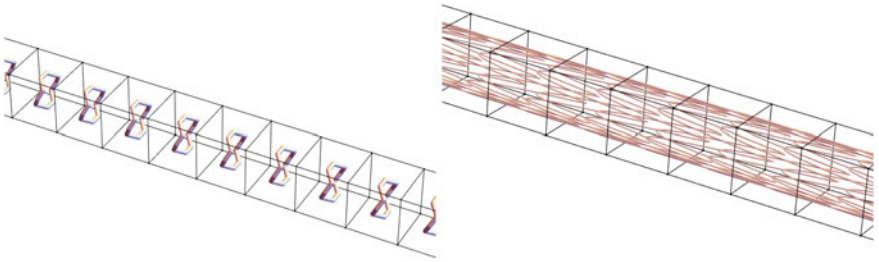


Fig. 11.4 1D PBC systems: unlinked and saturated

densities at which infinite linked subsystems arise, the onset of percolation, in the periodic boundary condition systems. We showed that, with increasing density, the topological entanglement of these systems increases in complexity, dimension, and probability.

11.3.2.1 Analysis of One-Dimensional PBC Olympic Systems

The mean absolute linking number in saturated systems in 1 PBC becomes greater than zero at density $\rho \approx 0.08$, when the mean valence and the probability of percolation become non-zero, see Fig. 11.5. It is interesting to notice that the mean absolute linking number exceeds one, showing that, even though the polygons are not knotted and are just close enough to link, there exist polygons with absolute linking greater than one. At the critical density the mean absolute linking becomes 1.3, indicating the presence of many pairs of polygons with absolute linking number greater than one. The mean absolute linking number continues to increase with density, approaching the value 2. This suggests that at high densities unknotted polygons can have high linking numbers, a conclusion supported by the growth of the total absolute linking as a function of the density.

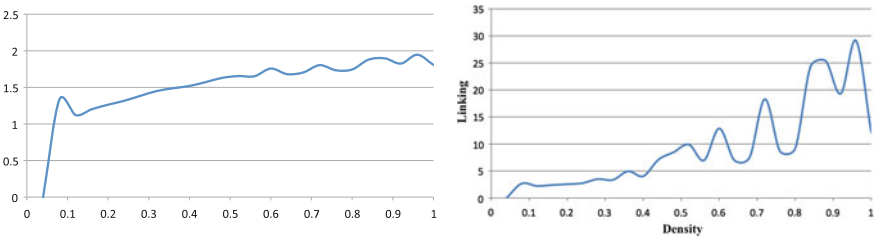


Fig. 11.5 The mean absolute linking of two chains and the mean total absolute linking per chain as a function of density for 1PBC saturated systems

11.3.2.2 Percolation Analysis

For PBC Olympic systems, we proposed the following relationship between the probability of total saturation as a function of the density of the system:

$$p(\rho) = \frac{1}{1 + \rho^{-\alpha} e^{-k\rho}}$$

as inspired by the logistic equation where ρ is the density, $\alpha = n + \frac{1}{2}$, n is the dimension of the PBC system, and k is a constant.

Note that a polygon can link with one of its translations when the size of the simulation box is similar to the size of the chain. On average, a polygon will link with its own image when the length of the simulation cell is $l \approx 2\langle R_g^2 \rangle^{1/2} \approx 2 \cdot 3.23 = 9.12$, which corresponds to a density of $\rho = \frac{1}{6.45} = 0.15$ for rings. More precisely, in a system with one, two, or three PBC, when the length of the cell is $l < 2\lambda_n$.This implies that we should expect linking to occur when $l \leq 2\lambda_1$ giving a critical density of $\rho_{C1} > \frac{1}{11.94} = 0.0837521$ for a one-dimensional system.

In a one dimensional PBC system, see Fig. 11.6, we notice that the probability of saturation becomes greater than 0 at $\rho \approx 0.12$, in agreement with our analysis. In a one dimensional PBC system, more than half of the conformations are fully saturated once the density has exceeded 0.28. Using *Matlab*'s non-linear fitting, we find $k = 7.032$ with an $R^2 = 0.9918$, see Fig. 11.6. This suggests that the probability of linking between two translations of a polygon as a function of density is:

$$p(\rho) = \frac{1}{1 + \rho^{-1.5} e^{-7.032\rho}}$$

to be compared with the probability of linking between two random unknotted polygons provided in [16].

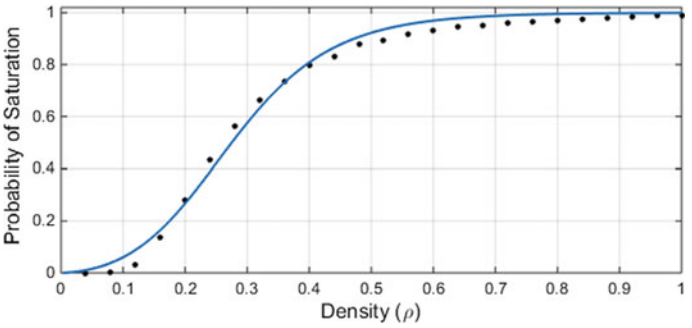


Fig. 11.6 The observed probability of saturation in a one dimensional PBC system as a function of density plotted against its fitting curve $p(\rho)$

11.3.2.3 Analysis of Valence, $|V|$

In this section we discuss the mean valence, $|V|$, of a polygon in a percolated system, i.e. the average number of polygons with which an individual polygon may link. Recalling that the typical enveloping ellipsoid has characteristic radii of $\lambda_1 = 5.97$, $\lambda_2 = 4.09$, and $\lambda_3 = 2.9$ [17], let us denote by I_0 a random polygon of N unit length edges in a system with n =1, 2, or 3 PBC and defined by its generating cell of dimensions l_x, l_y, l_z. We propose that I_0 may link any of its own translations if the enveloping ellipsoid of the translation intersects the enveloping ellipsoid of I_0. To obtain an estimate of the valence, we first consider the enveloping ellipsoid of I_0 and form a shell around it by adding a thickness proportional to the characteristic radii in the nearest direction ie. $\lambda_n \frac{\langle Rg^2\rangle^{1/2}}{\lambda_1}$. The respective radii defining this shell are then $(\lambda_1 + \lambda_1 \frac{\langle Rg^2\rangle^{1/2}}{\lambda_1})$, $(\lambda_2 + \lambda_2 \frac{\langle Rg^2\rangle^{1/2}}{\lambda_1})$, $(\lambda_3 + \lambda_3 \frac{\langle Rg^2\rangle^{1/2}}{\lambda_1})$. We estimate the valence by counting the number of images whose center of gravity are contained within this volume by dividing by the volume of the generating cell. Setting the static dimensions of the generating cell equal to $2 * \lambda_n$, we find the following estimates for mean valence in an n-PBC system:

$$\langle |V| \rangle_{n-PBC} \quad \leq \quad \frac{\frac{4}{3}\pi\lambda_1\lambda_2\lambda_3\left(1 + \frac{\langle Rg^2\rangle^{1/2}}{\lambda_1}\right)\rho^n}{(2\lambda_{n+1})^{(3-n)}}.$$

Using the appropriate values of n, we bound the mean valence in each system of PBC's by:

$$\langle |V| \rangle_{1PBC} \quad \leq \quad 16.28\rho$$

$$\langle |V| \rangle_{2PBC} \quad \leq \quad 296.61\rho^2$$

$$\langle |V| \rangle_{3PBC} \quad \leq \quad 1089.03\rho^3$$

We expect these to be upper bound estimates, especially at higher densities, due to the over counting of some portion of cells whose centers are not within the shell. However, the combined volume of those on the boundary of the shell could add to this count.

The mean valence of a saturated system in 1 PBC becomes non-zero at 2 for $\rho \geq 0.12$ corresponding with the critical density for filamental percolation $\rho_{C1} > 0.0836$. The mean valence continues to increase non-monotonically thereafter, see Fig. 11.7. Notice that, at the saturation density of $\rho = 0.28$, we have a mean valence of approximately 2.53, indicating that at least a fourth of the linked polygons link with their second order neighbors. This saturation density corresponds to an edge length of $l \approx 3.57 < 2\lambda_1$, which explains why the mean valence becomes greater than two.

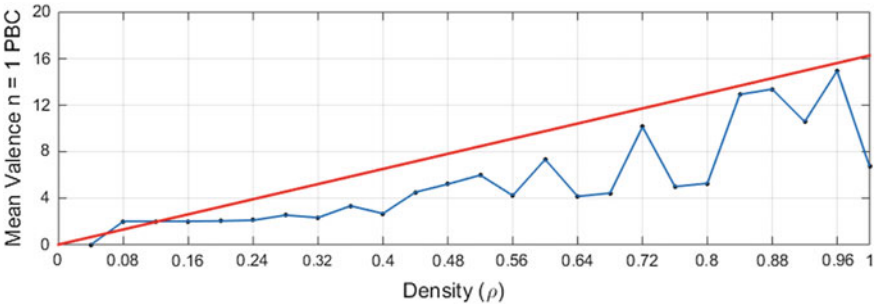


Fig. 11.7 Mean valence of the total 1PBC system superimposed with the analytical model. We notice that $< |V| >_{1PBC} \leq 16.28\rho$, as expected

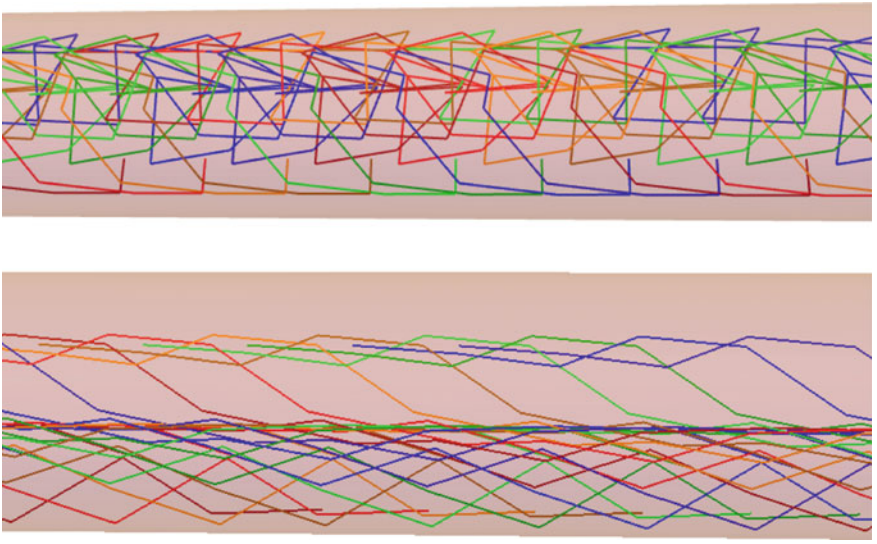


Fig. 11.8 PBC examples: unit radius disc cross-section, single chain of length 25. Left, 0.00 alignment constraint (random) and, second, 0.50 alignment constraint

11.3.3 Vortex Flow in Tubular Systems

In Fig. 11.8 we show a simple example of a length 25 chains contained in a tube whose cross-sectional disc of radius 1 under an alignment constraint scaled to 0.00 and 0.50. In the context of this study, the alignment constraint is scaled from 0 to 1 by a continuous control, c, of the proposed step angle in reference to the tube axis and the radius of the tube, r. If a vertex of the chain were to lie on the boundary of the tube, the magnitude of the azimuthal angle, θ, is limited, for $0 < r \leq \frac{1}{2}$, by $s(r) = ArcSin(2r)$ in the implementation of the generation of the chain. Furthermore, for each step in the generation of the chain, a random number, δ, in [0,1] is selected:

If $\delta \geq \frac{5\,c}{4}$ then either, equally and uniformly,

$$(\pi - (1-c)s(r)) \leq \theta \leq (\pi + (1-c)s(r))$$

or

$$-(1-c)s(r) \leq \theta \leq ((1-c)s(r).$$

If $\frac{5\,c}{4} > \delta \geq 0$ then, uniformly,

$$-(1-c)s(r) < \theta < (1-c)s(r).$$

We interpret these chains as representing short vortex flow lines in a tubular system and will focus on the case in which there are three independent PBC families of chains generating the structure representing the flow lines, see Fig. 11.8. They may, however, be understood as representing polymeric chains subject to tubular confinement and an alignment constraint.

For these filamental structures, we systematically studied a range of lengths, tube cross-sections, and scaled alignment conditions in order to estimate how these fundamental parameters influenced the shape and entanglement that depend upon them. For example, for a sample size of 500, the random chains in the unit radius tube have an average tube length of 7.86 units with an average radius of gyration of 3.57, see Fig. 11.9. The average absolute self-linking is 1.09, see Fig. 11.10. As a measure of entanglement, we find the average maximal, medial, minimal absolute eigenvalues of the linking matrix to be 1.18, 1.01 and, 0.46 respectively. In contrast, for an alignment condition of 0.50, one has an average tube length of 18, 35 with a radius of gyration of 22.77 and an average absolute self-linking of 1.34. The absolute eigenvalues have averages of 2.20, 1.27, and 0.57 respectively. The increased level of entanglement found in aligned systems compared to a fully random system is a key result of our analysis of the filamental structure's dependence upon cross-sectional constraints and alignment.

In Fig. 11.11 we show the evolution of the absolute values of the three eigenvalues of a PBC system generated by three independent filaments of length 25. Here, one observes a decreasing tendency in the magnitude of the eigenvalues with increasing tube radius and an increasing tendency with increasing alignment constraint. While one may expect a random system to exhibit a stronger degree of entanglement, we have seen that these filaments have smaller diameter (or squared radius of gyration) thereby offering them a significantly smaller opportunity to entangle with nearby filaments whereas filaments subject to an alignment constraint have a significantly larger number of adjacent filaments with which they may entangle. Thus, we see that the magnitude of the eigenvalues increase with increasing alignment. In addition, for a fixed alignment constraint, the magnitude of the eigenvalues decreases with increasing tube radius across the range of radii presented here, i.e. from 0.1 through 5.0 showing that the filamental structure widely explores the cylindrical tube leading to a decreasing density leading to decreasing entanglement.

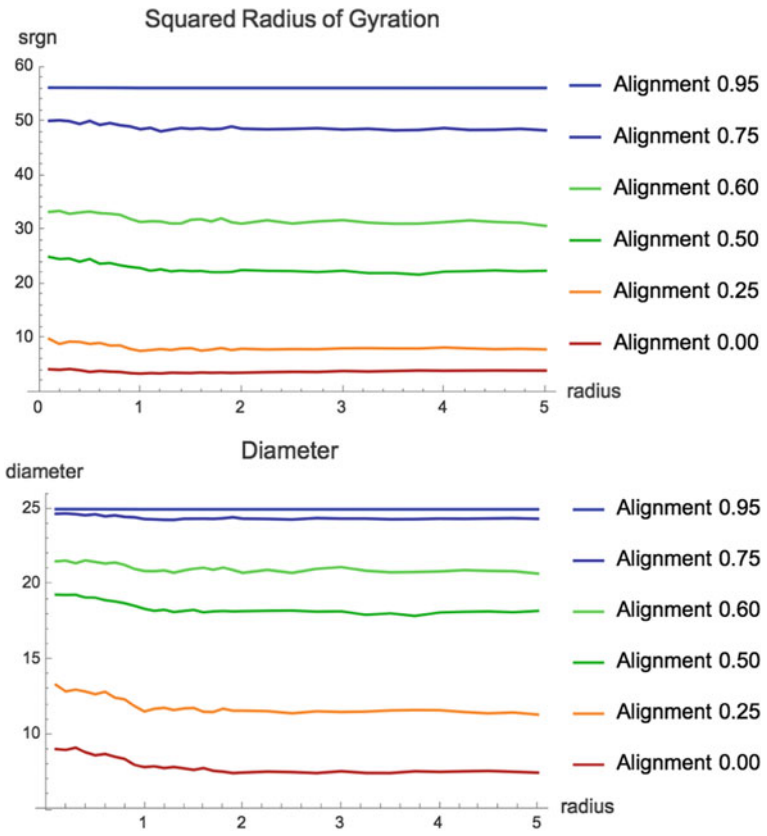


Fig. 11.9 The mean squared radius of gyration and diameter of 25 step chains as a function of the tube radius and alignment constraint

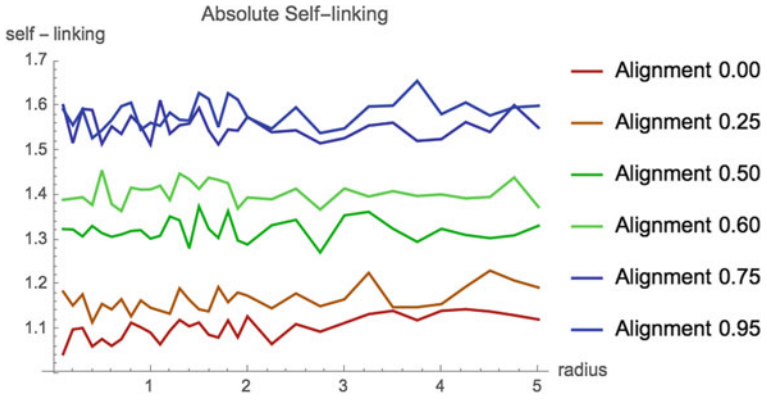


Fig. 11.10 Absolute self-linking of single 25 step chains as a function of alignment constraint and tubular radius

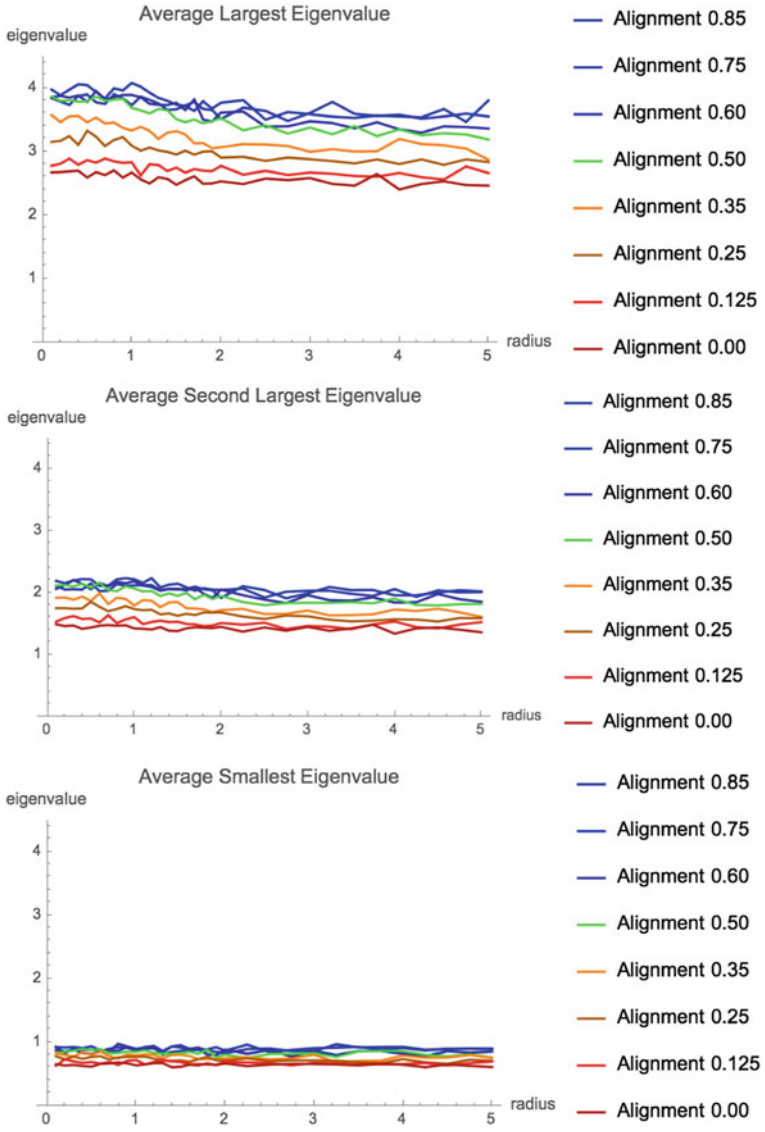


Fig. 11.11 Effect of tube radius and alignment constraint on the mean absolute eigenvalues of a PBC system generated by three independent filaments of length 25

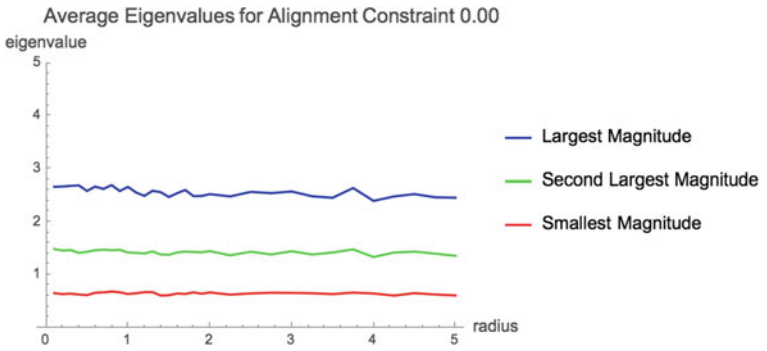


Fig. 11.12 Effect of random alignment, fixed at 0.00, on length 25 filament; radial scale 0.10–5.00

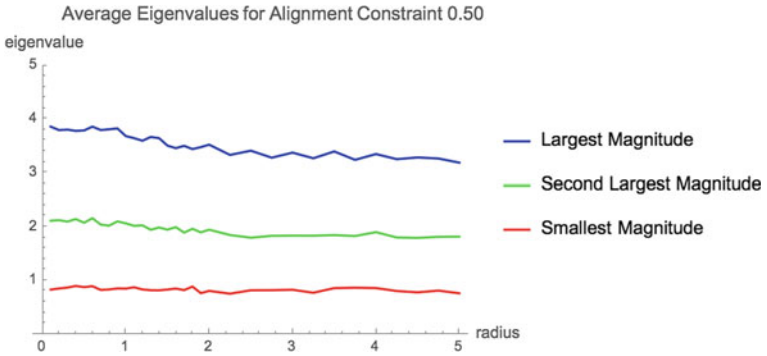


Fig. 11.13 Effect of alignment constraint, fixed at 0.50, on length 25 filament; radial scale 0.10–5.00

Consistent with our earlier analysis, we find that a random system displays the smallest entanglement as measured by the magnitude of the eigenvalues, see Fig. 11.12. In Fig. 11.13 we see that the two eigenvalues of larger magnitudes are rather larger than the random system but tend to get smaller with increasing tube radius while the smallest is relatively stable in magnitude. Since the character of this decrease in the magnitude of the eigenvalues holds across the scale of the alignment constraint, we expect that it is an artifact of the decrease in density of the filaments with increasing tube radius. Considering, in Fig. 11.14, the change in magnitude of the eigenvalues for alignment constraint of 0.85, we do not see any meaningful change in magnitude with increasing tube radius as the magnitudes remain roughly constant at the largest eigenvalue measures of entanglement.

We now wish to characterize the consequences of increasing the alignment constraint for a fixed tube radius. For a tube of radius equal to 0.10, 1.00 or 5.00, in Fig. 11.15 we see that there is a visible increase in the magnitude of the largest eigenvalue as the alignment constraint increases independent of the radius of the

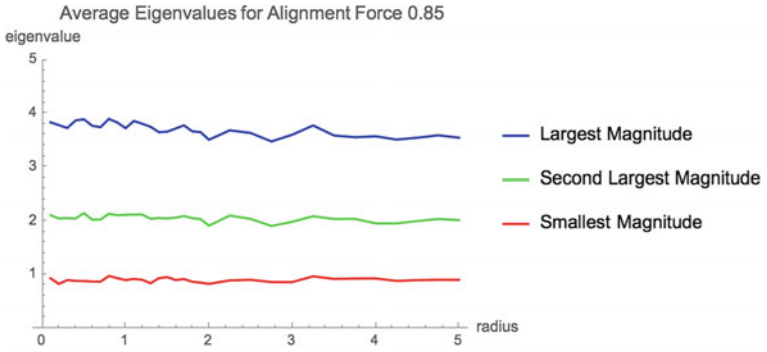


Fig. 11.14 Effect of alignment constraint, fixed at 0.85, on length 25 filament; radial scale 0.10–5.00

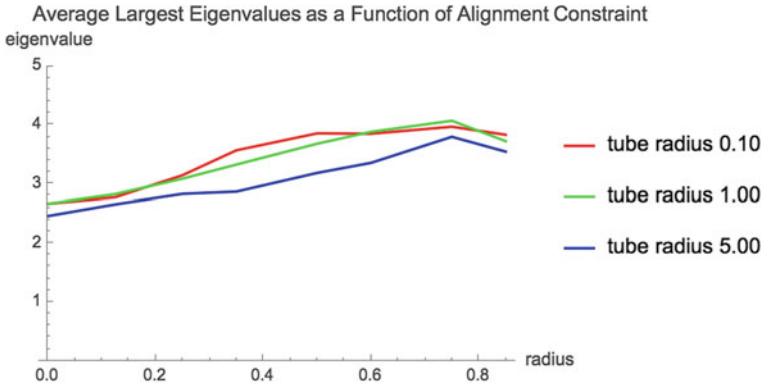


Fig. 11.15 Effect of alignment constraint on the mean largest magnitude eigenvalue for length 25 filaments for tube radii 0.10, 1.00, and 5.00

tube. Indeed, the actual values of the magnitude of the largest eigenvalue are quite similar, independent of the radius of the tube though a bit lower for the very largest tube.

11.4 Conclusions

The entanglement in polymer melts, Olympic gels, and systems of vortex flow lines is fundamentally a many body problem. Our goal is to describe its principal properties with a measure of entanglement that takes into consideration the overall conformation of the system. For this purpose we defined the linking matrix. For systems employing PBC, we then defined the periodic linking matrix using the periodic linking and self-linking measures. In the simulation of a polymer system, the size of the cell may vary.

It is necessary to know how the data obtained from different cell sizes are related. By focusing on an arbitrary fixed periodic system simulated by a varying cell-size simulation box with one PBC, we proved that some of the eigenvalues of the periodic linking matrix are invariant of cell size. This information can be used to characterize such a one dimensional periodic system. This led to results concerning the evolution of these structural measures as the size of the basic cell changes [4]. The analysis of one dimensional PBC systems was applied to Olympic systems and to tubular systems to give the results reported here [5, 6] giving new insight into the structure of these systems.

Acknowledgements This research has been cofinanced by the European Union (European Social Fund — ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) — Research Funding Program: THALIS.

References

1. Gauss, K.F.: Zur mathematischen theorie der electrodynamischen wirkungen. Werke Konigl. Ges. Wiss. Gottingen, vol. 5, p. 605 (1877)
2. Panagiotou, E.: The linking number in systems with periodic boundary conditions. J. Comput. Phys. **300**, 533–573 (2015)
3. Panagiotou, E., Millett, K.C., Lambropoulou, S.: The linking number and the writhe of uniform random walks and polygons in confined spaces. J. Phys. A: Math. Theor. **43**(4), 045208 (2010)
4. Panagiotou, E., Millett, K.C.: Linking matrices in systems with periodic boundary conditions (2016)
5. Igram, S., Millett, K.C., Panagiotou, E.: Resolving critical degrees of entanglement in olympic ring systems. J. Knot Theory Ramif. **25**(14), 1650081 (2016)
6. Millett, K.C., Panagiotou, E.: Entanglement transitions in one-dimensional confined fluid flow. Fluid Dyn. Res. **tbd**(tbd), tbd (2016)
7. Rojo, O., Rojo, H.: Some results on symmetric circulant matrices and on symmetric centrosymmetric matrices. Linear Algebra Appl. **392**, 211–233 (2004)
8. De Gennes, P.-G.: Scaling Concepts in Polymer Physics. Cornell University Press, Ithaca (1979)
9. Chen, J., Rauch, C.A., White, J.H., Englund, P.T., Cozarelli, N.R.: The topology of kinetoplast DNA network. Cell **80**, 61–69 (1995)
10. Lukes, J., Guilbride, D.L., Votypka, J., Zikova, A., Benne, R., Englund, P.T.: Kinetoplast DNA network: evolution of an improbable structure. Eukaryot. Cell **1**(4), 495–502 (2002)
11. Micheletti, C., Marenduzzo, D., Orlandini, E.: Polymers with spatial or topological constraints: theoretical and computational results. Phys. Rep. **504**, 1–73 (2011)
12. Diao, Y., Hinson, K., Arsuaga, J.: The growth of minicircle networks on regular lattices. J. Phys. A: Math. Theor. **45**(3), 035004 (2012)
13. Diao, Y., Hinson, K., Kaplan, R., Vazquez, M., Arsuaga, J.: The effects of density on the topological structure of the mitochondrial DNA from trypanosomes. J. Math. Biol. **64**(6), 1087–1108 (2012)
14. Arsuaga, J., Diao, Y., Hinson, K.: The effect of angle restriction on the topological characteristics of minicircle networks. J. Stat. Phys. **146**(2), 434–445 (2012)

15. Diao, Y., Hinson, K., Sun, Y., Arsuaga, J.: The effect of volume exclusion on the formation of DNA mini circle networks: implications to kinetoplast DNA. J. Phys. A: Math. Theor. **48**, 1–11 (2015)
16. Hirayama, N., Tsurusaki, K., Deguchi, T.: Linking probabilities of off-lattice self-avoiding polygons and the effects of excluded volume. J. Phys. A: Math. Theor. **42**, 105001 (2009)
17. Millett, K.C., Plunkett, P., Piatek, M., Rawdon, E.J., Stasiak, A.: Effect of knotting on polymer shapes and their enveloping ellipsoids. J. Chem. Phys. **130**(16), 165104 (2009)

Chapter 12
On the Height of Knotoids

Neslihan Gügümcü and Louis H. Kauffman

Abstract Knotoid diagrams are defined in analogy to open ended knot diagrams with two distinct endpoints that can be located in any region of the diagram. The height of a knotoid is the minimal crossing distance between the endpoints taken over all equivalent knotoid diagrams. We define two knotoid invariants; the affine index polynomial and the arrow polynomial that were originally defined as virtual knot invariants given in (Kauffman, J Knot Theory Ramif 21(3), 37, 2012) [6], (Kauffman, J Knot Theory Ramif 22(4), 30, 2013) [8], respectively, but here are described entirely in terms of knotoids in S^2. We reprise here our results given in (Gügümcü, Kauffman, Eur J Combin 65C, 186–229, 2017) [3] that show that both polynomials give a lower bound for the height of knotoids.

12.1 Introduction

The theory of knotoids was introduced by V. Turaev [17] in 2012. A knotoid diagram [17] is an open ended knot diagram with two endpoints that can be located in any region of the diagram. The theory of knotoids forms a new diagrammatic theory that is an extension of the classical knot theory. In this paper, we give an exposition of two new polynomial knotoid invariants that were constructed in [3].

It is natural to examine knotoids in the context of virtual knot theory [5, 6]. Virtual knots are knots in thickened surfaces (or knot diagrams in surfaces) taken up to handle stabilization. There is a diagrammatic theory for virtual knots, as we explain briefly in this paper. The endpoints of a knotoid diagram can be connected to form what we call the virtual closure of the diagram. This way of connecting the endpoints of

N. Gügümcü (✉)
Department of Applied Mathematics, National Technical University of Athens,
Zografou Campus, 15780 Athens, Greece
e-mail: nesli@central.ntua.gr

L.H. Kauffman
Department of Mathematics, University of Illinois at Chicago,
Chicago, IL, USA
e-mail: kauffman@uic.edu

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_12

a knotoid diagram forms a well-defined map from the set of knotoids to the set of virtual knots. Virtual knot invariants can be then applied to extract knotoid invariants by using the virtual closure map.

Section 12.2 recollects the fundamental concepts of knotoids. Knotoid diagrams can be defined both in S^2 and $\mathbb{R}^2$. There is an inclusion map between two sets of knotoids; knotoids in $\mathbb{R}^2$ and knotoids in S^2, induced by the inclusion $\mathbb{R}^2 \hookrightarrow S^2$. Knotoids in $\mathbb{R}^2$ are a part of geometric 3-dimensional knot theory, and, as such, are related to open-ended embeddings of intervals in three-dimensional space. We discuss this point of view in Sect. 2.4.

Given a knotoid K in S^2, we can ask how far apart the endpoints need to be in all instances of diagrams for the equivalence class of K. The smallest distance between two endpoints of K (in terms of number of classical crossings created while connecting the endpoints with an embedded arc that goes under) is called the *height* of the knotoid. The height is an invariant of knotoids in S^2 [17]. In Sect. 12.3 we define the height of a knotoid as in [17]. We then make a remark on virtual knot theory and discuss the virtual closure map. In Sect. 3.2, we mention two conjectures from [3]. One of the conjectures asserts that there are virtual knots of genus 1 which are not in the image of the virtual closure map. We have proved this conjecture by a use of surface bracket polynomial of virtual knots/links and we will present the prof in a subsequent paper. The other conjecture asserts that the normalized bracket polynomial (the Jones polynomial) detects the trivial knotoid. This conjecture extends the well-known conjecture on the Jones polynomial of knots.

Section 12.4 is devoted to two polynomial invariants of knotoids; the affine index polynomial and the arrow polynomial. Both of these polynomials are examined by the authors in full detail in [3]. The affine index polynomial and the arrow polynomial were originally defined as virtual knot invariants [2, 8, 14]. We observe in [3] that they can be defined as invariants of knotoids by considering only knotoid diagrams. In this paper we present our main results given in [3], that consist in lower bound estimations for the height of knotoids via these two polynomials. We give sketches of the proofs for our results. We end the section with examples for the use of these estimations.

12.2 About Knotoids

A knotoid diagram K in S^2 or in $\mathbb{R}^2$ is defined as a generic immersion,

$$K : [0, 1] \to S^2 \text{ or } \mathbb{R}^2 \text{ such that}$$

- K has finitely many transversal double points. These points are endowed with over/under-crossing data and called the *crossings* of K.
- The images of 0 and 1 are two points distinct from each other, and from any of crossings of K. These two points are the *endpoints* of a knotoid diagram and called the *tail* and the *head* of K, respectively. A knotoid diagram is always oriented from its tail to its head.

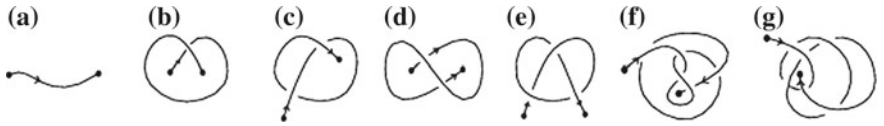


Fig. 12.1 Some examples of knotoid diagrams

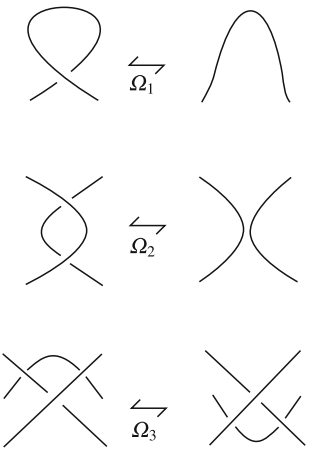


Fig. 12.2 Ω-moves

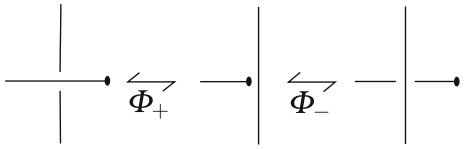


Fig. 12.3 Forbidden knotoid moves

The trivial knotoid diagram is an embedding of the unit interval into S^2 (or in $\mathbb{R}^2$). It is depicted by an arc without any crossings as shown in Fig. 12.1a.

The three Reidemeister moves shown in Fig. 12.2, are defined on knotoid diagrams, and denoted by Ω_1, Ω_2, Ω_3, respectively. These moves modify a knotoid diagram within a small disk as shown in the figure without utilizing the endpoints. The moves in Fig. 12.3 that consist of pulling an endpoint over or under a strand, are the *forbidden over and under moves*, and denoted by Φ_+ and Φ_-, respectively. It is clear that if both Φ_+ and Φ_-- moves were allowed, any knotoid diagram could be turned into the trivial knotoid diagram.

The $\Omega_{i=1,2,3}$-moves plus isotopy of S^2 generate an equivalence relation on knotoid diagrams in S^2 (for knotoid diagrams in the plane we consider the isotopy of the plane). A *knotoid* is defined to be an equivalence class of all equivalent knotoid diagrams up to this equivalence relation. The set of all knotoid classes in S^2 and in $\mathbb{R}^2$ are denoted by $\mathscr{K}(S^2)$ and $\mathscr{K}(\mathbb{R}^2)$, respectively.

There is a well-defined map between the two knotoid sets,

$$\iota : \mathscr{K}(\mathbb{R}^2) \to \mathscr{K}(S^2),$$

that is induced by the inclusion $\mathbb{R}^2 \hookrightarrow S^2 = \mathbb{R}^2 \cup \{\infty\}$. Any knotoid in S^2 can be represented by a knotoid diagram in $\mathbb{R}^2$ by pushing a representative diagram in S^2 away from the $\infty \in S^2$. Considering the equivalence class of this planar representation in $\mathscr{K}(\mathbb{R}^2)$, there is a well-defined map $\rho : \mathscr{K}(S^2) \to \mathscr{K}(\mathbb{R}^2)$. It is clear that $\iota \circ \rho = id$ so that ι is surjective. However, there are examples of nontrivial knotoids in $\mathbb{R}^2$ which are trivial in $\mathscr{K}(S^2)$ showing that the map ι is not injective. An example for this is given in Fig. 12.1b. The knotoid diagram in the figure which represents a nontrivial planar knotoid [17] although it represents the trivial knotoid in S^2. In this paper, we study the knotoids in S^2 and we mean knotoids in S^2 unless otherwise stated.

Definition 12.1 Let $\mathscr{M}$ be a category of mathematical structures (e.g. polynomials, Laurent polynomials, the integers modulo five, commutative rings, groups, $\cdots$). An *invariant* of knotoids is a mapping $\mathrm{I} : \mathscr{K}(S^2) \to \mathscr{M}$ such that equivalent knotoid diagrams map to equivalent structures in $\mathscr{M}$.

12.2.1 Knots via Knotoid Diagrams

In [17] the study of knotoid diagrams is suggested as a new diagrammatic approach to the study of knots in three dimensional space $\mathbb{R}^3$ in the following way. The endpoints of a knotoid diagram can be connected with an embedded arc in S^2 that is declared to go under each strand it crosses. In this way we obtain an oriented classical knot diagram in S^2 representing a knot in $\mathbb{R}^3$. The resulting knot diagram is called the *underpass closure* of the knotoid diagram. Note that the arc connecting the endpoints is unique up to isotopy. We say that a knotoid diagram K *represents* a classical knot κ if κ is represented by the underpass closure of K.

Alternatively, the endpoints of a knotoid diagram in S^2 can be connected with an embedded arc in S^2 which is declared to go over every strand it crosses. The resulting oriented knot diagram is called the *overpass closure* of the knotoid diagram. Note that this type of connection arc is also unique up to isotopy. The underpass closure and the overpass closure of a knotoid diagram may represent inequivalent knots. For instance, the knotoid diagram given in Fig. 12.4 represents a trefoil via the underpass closure and represents the trivial knot via the overpass closure. In order to have a well-defined representation of knots via knotoid diagrams, we should fix the closure type. The closure type is chosen to be the underpass closure and the following proposition follows.

Proposition 12.1 ([17] *Two knotoid diagrams K_1 and K_2 represent the same classical knot if and only if they are related to each other by finitely many $\Omega_{i=1,2,3}$- moves, Φ_{-}-moves (forbidden under moves) and isotopy of S^2.*

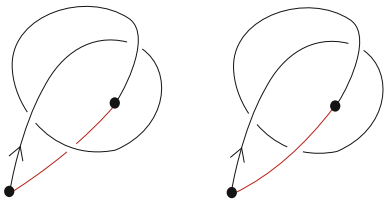


Fig. 12.4 Two types of closures

Given a knot in $\mathbb{R}^3$, we take an oriented diagram of the knot in S^2. Cutting out an underpassing arc which may contain no crossing or one or more crossings from this diagram, results in a knotoid diagram that represents the given knot. In fact, any knot in $\mathbb{R}^3$ can be represented by a knotoid diagram in S^2.

Representing knots in three dimensional space via knotoid diagrams may ease the computation or give finer estimations for knot invariants. See [17] for the knot group computation via knotoid diagrams and for the lower bound estimation for the Seifert genus of knots.

12.2.2 Knotoids as an Extension of Knot Theory

There is a well-defined injective map,

$$\alpha : \{\text{Knot Diagrams in } S^2 / \langle \Omega_1, \Omega_2, \Omega_3 \rangle\} \to \mathscr{K}(S^2),$$

where $\langle \Omega_1, \Omega_2, \Omega_3 \rangle$ denotes the equivalence on knot diagrams given by the three Reidemeister moves defined on knot diagrams. Let D be an oriented knot diagram in S^2. Cutting out an open arc of D which is apart from the crossings of D results in a knotoid diagram with two endpoints in the same local region of the diagram. The map α is induced by assigning D to the resulting knotoid diagram. It is verified in [17] that the knotoid obtained via the map α does not depend on the knot diagram chosen or the arc that is cut out from the knot diagram. Therefore α is a well-defined map. It is also verified in [17] that the map α is injective.

Definition 12.2 Knotoids that are in the image of the α map, are called *knot-type knotoids*.

A knot-type knotoid has at least one diagram in its equivalence class whose endpoints lie in the same local region of the diagram. Such a diagram is called a *knot-type knotoid diagram*. The Fig. 12.1a, b and e set some examples of knot-type knotoid diagrams.

Definition 12.3 The knotoids that are not in the image of α, are called *proper or pure knotoids*.

The endpoints of a proper knotoid can lie in any but different local regions of the diagram for any its representative diagram. The set of knotoids in S^2 can be regarded as a union of these two types of knotoids, that is,

$$\mathscr{K}(S^2) = \{\text{Knot-type knotoids}\} \cup \{\text{Proper knotoids}\}.$$

The map α, being an injective map, gives a one-to-one correspondence between the set of equivalence classes of knots and the set of knot-type knotoids. Thus the theory of knotoids extends the theory of classical knots (knots in $\mathbb{R}^3$). Notice that by the allowance of the Φ_{-}-move on knotoid diagrams, the theory of knotoids becomes an equivalent theory to the theory of classical knots.

12.2.3 The Algebraic Structure on the Set of Knotoids

A multiplication operation is defined on the set of knotoids, $\mathscr{K}(S^2)$ in [17] as follows. Let k_1, k_2 be two knotoids represented by two knotoid diagrams K_1 and K_2. Take 2-disk neighborhoods B_1, B_2 of the head of K_1, and the tail of K_2, respectively, such that both discs intersect the diagrams along a radius. We glue $S^2 - \text{Int}(B_1)$ to $S^2 - \text{Int}(B_2)$ through a homeomorphism taking ∂B_1 to ∂B_2 and carrying the single intersection point of ∂B_1 and K_1 to the single intersection point of ∂B_2 and K_2. Then $K_1 - \text{Int}(B_1)$ meets with $K_2 - \text{Int}(B_2)$ at one point and form a knotoid diagram in S^2 that is denoted by K_1K_2. The knotoid diagram K_1K_2 represents the product knotoid k_1k_2 in S^2.

It can be verified by the reader that the multiplication operation on knotoids is associative and the trivial knotoid is the identity element. The set of knotoids, $\mathscr{K}(S^2)$ endowed with the multiplication forms a semigroup with identity element [17].

12.2.4 A Geometric Interpretation of Knotoids

It is natural to see a knotoid diagram in $\mathbb{R}^2$ as a generic projection of an open-ended, oriented space curve. Given an open-ended, smooth, oriented curve that is embedded in $\mathbb{R}^3$ with a generic projection to a plane. The endpoints of the curve determine two lines that pass through the endpoints and are perpendicular to the plane. The generic projection of the curve to the plane along the lines is a knotoid diagram in that plane when self-crossings of the projection curve are endowed with over and under-crossing data accordingly with the weaving of the space curve. We call an open-ended curve embedded in $\mathbb{R}^3$ that has a generic projection to a plane, a *generic curve with respect to the plane*.

Any knotoid diagram in $\mathbb{R}^2$ determines an open-ended oriented curve embedded in $\mathbb{R}^3$. Let K be a knotoid diagram in $\mathbb{R}^2$. The plane of the diagram is identified with $\mathbb{R}^2 \times \{0\} \subset \mathbb{R}^3$. The overpasses of the diagram are pushed into the upper half-space

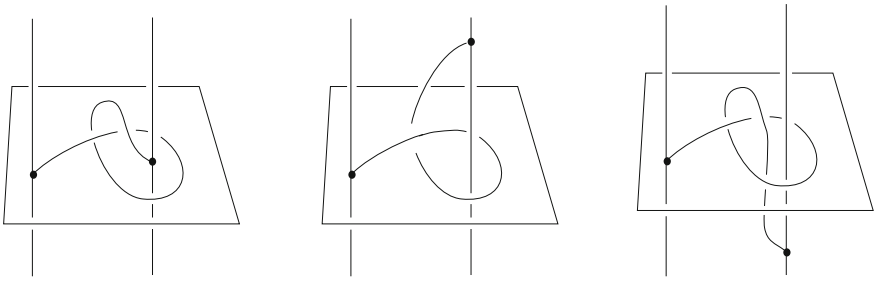


Fig. 12.5 Space curves obtained by the knotoid diagram in Fig. 12.1c

and the underpasses are pushed into the lower half-space in the vertical direction. The tail and the head of the diagram are attached to the two lines, t $\times \mathbb{R}$ and h $\times \mathbb{R}$ that pass through the tail and the head, respectively, and are perpendicular to the plane of the diagram. Moving the endpoints of K along these special lines gives rise to several open-ended oriented curves in $\mathbb{R}^3$ with endpoints lying on these lines. Figure 12.5 illustrates three space curves obtained from the same knotoid diagram in the xy-plane.

Definition 12.4 Let c_1, c_2 be two open-ended, smooth, oriented curves that are embedded in $\mathbb{R}^3$ with a generic projection to the same plane, and t and h denote their endpoints. The curves c_1 and c_2 are said to be *line isotopic* if there is a smooth ambient isotopy of the pair $(\mathbb{R}^3 \setminus \{t \times \mathbb{R}, h \times \mathbb{R}\}, t \times \mathbb{R} \cup h \times \mathbb{R})$, taking one curve to the other curve in the complement of the lines, taking endpoints to endpoints, and taking the lines to the lines; $t \times \mathbb{R}$ to $t \times \mathbb{R}$ and $h \times \mathbb{R}$ to $h \times \mathbb{R}$.

Theorem 12.1 ([3, Theorem 2.2]) *Two oriented curves in $\mathbb{R}^3$ that are both generic with respect to a given plane, are line isotopic (with respect to the lines determined by the endpoints of the curves and the plane) if and only if the projections of the curves to the plane when endowed with over/under data at each self-crossing points accordingly to the weaving of the curves, are equivalent knotoid diagrams.*

Note that the equivalence classes of knotoid diagrams that are the projections of the same open-ended curve embedded in three-dimensional space, may vary with respect to the projection plane. Figure 12.6 depicts a space curve and the projections of the curve to the xy- and the xz-plane. It is clear that the xz-projection gives the trivial knotoid. The projection to the xy- plane, however, gives the knotoid diagram given in Fig. 12.1c and one can show that this diagram represents a nontrivial knotoid in S^2 and in $\mathbb{R}^2$ (see Sect. 1.3.1 and [3] for more details). In [3] we suggest that one can study the set of all knotoids assigned to one space curve for understanding the physical properties of the space curve.

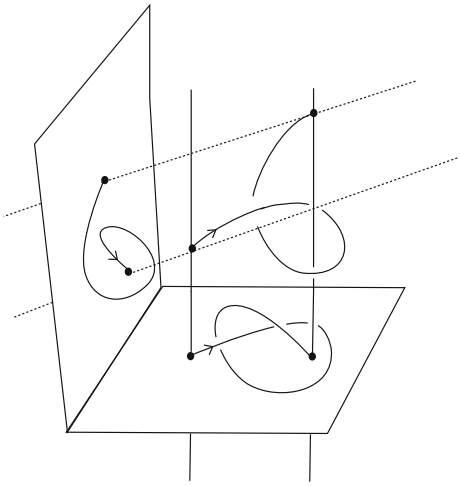


Fig. 12.6 Knotoids in different projection planes

12.3 The Height of a Knotoid

The *height* (or the *complexity* as used in [17]) of a knotoid diagram in S^2 is the minimum number of crossings that a connection arc creates during the underpass closure. The *height of a knotoid* K, $h(K)$ is defined as the minimum of the heights, taken over all knotoid diagrams equivalent to K. The height is a knotoid invariant [17]. We note that we use the term height instead of complexity for knotoids to focus on this concept and allow us to use the word complexity more freely. One often refers to the complexity of a knot or a knotoid in terms of its crossing number or the virtual crossing number of the closure and other measures of how complicated is the topological type of the object. We hope that the reader will agree that this choice of terminology is useful in this case.

The height of a knotoid is preserved under the basic involutions of knotoid diagrams that are the *reversion*, *mirror image*, and the *symmetry* [17]. That is, for a knotoid K,

$$h(K) = h(mir(K)) = h(sym(K)) = h(rev(K)).$$

The height is also invariant under the isotopy of S^2 so that we can consider only planar knotoid diagrams and connection arcs in $\mathbb{R}^2$ for the computation of the height.

Theorem 12.2 ([17, Theorem 4.3]) *The height of a product knotoid* k_1k_2, $h(k_1k_2) = h(k_1) + h(k_2)$ *for any* $k_1, k_2 \in \mathscr{K}(S^2)$.

As pointed out in [17], a knotoid is of knot-type if and only if it has zero height or equivalently, a knotoid is a proper knotoid if and only if it has a nonzero height.

Thus the height is an efficient tool to measure how far a knotoid is from being a knot and for the classification of knotoids. The following conjecture has been made by V. Turaev.

Conjecture 12.1 [17] Minimal diagrams (with respect to the crossing number) of knot-type knotoids have zero height.

We have found a proof of this conjecture and we will give the proof in [4].

12.3.1 The Bracket Polynomial of a Knotoid

The bracket polynomial of a knotoid [17] is defined by extending the state expansion of the bracket polynomial of knots [9, 10]. Each crossing of a knotoid diagram K is smoothed either by A- or *B-type smoothing*, as shown in Fig. 12.7. A smoothing site is labeled by 1 if A-smoothing is applied and labeled by -1 if B-smoothing is applied at a particular crossing. A *state* of the knotoid diagram K is a choice of smoothing each crossing of K with the labels at smoothing sites. Each state of K consists of disjoint embedded circular components and a single long segment component with two endpoints. The initial conditions given in Fig. 12.7 are sufficient for the skein computation of the bracket polynomial of a knotoid.

Definition 12.5 The bracket polynomial of a knotoid diagram K is defined as

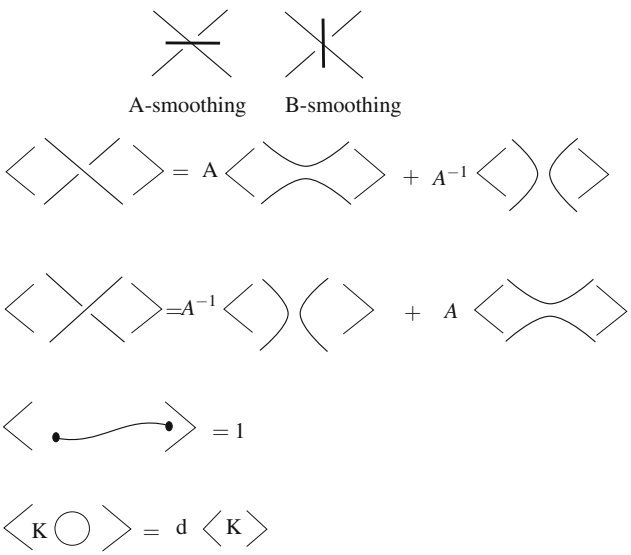


Fig. 12.7 Skein relations of the bracket polynomial

$$< K >= \sum_S A^{\sigma(S)} d^{\|S\|-1},$$

where the sum is taken over all states, $\sigma(S)$ is the sum of the labels of the state S, $\|S\|$ is the number of components of S, and $d = -A^2 - A^{-2}$.

The *writhe* of a knotoid diagram K, $\mathrm{wr}(K)$ is the number of positive crossings (crossings with sign $+1$, see Fig. 12.8) minus the number of negative crossings (crossings with sign -1) of K. The writhe is invariant under the Ω_2 and Ω_3 moves but Ω_1-move changes the writhe by ± 1. The bracket polynomial turns into an invariant for knotoids with a normalization by the writhe. The *normalized bracket polynomial* of a knotoid K, f_K is defined as the multiplication, $f_K = (-A^3)^{-\mathrm{wr}(K)} < K >$ [17].

The normalized bracket polynomial of knotoids in S^2 generalizes the Jones polynomial of knots in $\mathbb{R}^3$ with the substitution $A = t^{-1/4}$. Note that the Jones polynomial of the trivial knotoid is trivial.

Example 12.1 Let K_1 be the knotoid diagram illustrated in Fig. 12.9 with $\mathrm{wr}(K_1) = +2$. As we show in the figure, the bracket polynomial of K_1, $< K_1 >= A^2 + 1 - A^{-4}$ and so $f_{K_1} = A^{-4} + A^{-6} + A^{-10}$. This implies that K_1 is a non-trivial knotoid.

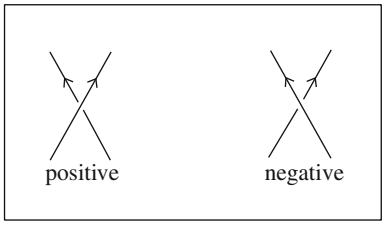


Fig. 12.8 Crossing types

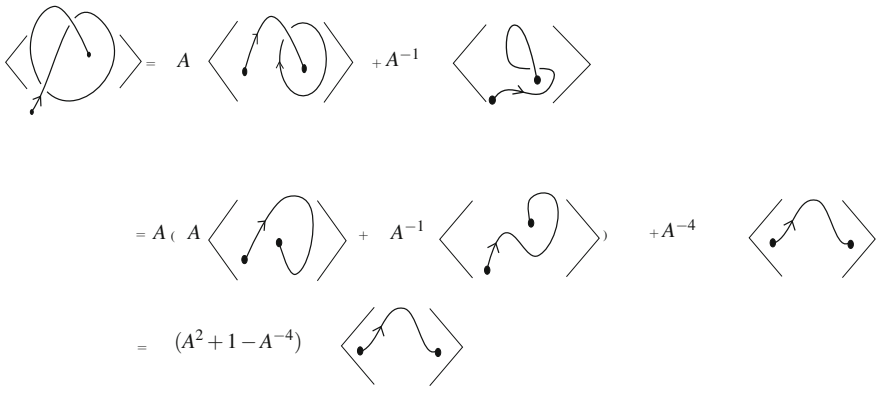


Fig. 12.9 Computation of the bracket polynomial of K_1

The Jones polynomial conjecture asserts that the Jones polynomial detects the trivial knot. This well-known conjecture can be extended to the following conjecture for knotoids in S^2.

Conjecture 12.2 The normalized bracket polynomial of knotoids in S^2 (or the Jones polynomial) detects the trivial knotoid.

12.3.2 Briefly on Virtual Knots

The theory of virtual knots that was introduced by the econd listed author [5, 6] in 1996, studies the embeddings of circles in thickened surfaces modulo isotopies and diffeomorphisms and one-handle stabilization of the surfaces.

Virtual knot theory has a diagrammatic formulation. In the diagrammatic theory, virtual knots and links are represented by diagrams with finitely many *classical crossings* (transversal self-crossings of the underlying curve that are endowed with over/under- data) and *virtual crossings* which are neither over-crossings or under-crossings. A virtual crossing is a combinatorial structure of the diagram that is indicated by two crossing segments with a small circle placed around the crossing point. A knot/link diagram with virtual crossings is called a *virtual knot/link diagram*.

The moves on virtual knot/link diagrams are generated by the classical Reidemeister moves plus the detour move. The detour move allows a segment with a consecutive sequence of virtual crossings to be excised and replaced by any other such a segment with a consecutive virtual crossings, as shown in Fig. 12.10. Local expressions that generate the detour move are illustrated in Figs. 12.11 and 12.12.

Virtual knot and link diagrams that can be connected by a finite sequence of these moves are said to be *equivalent* or *virtually isotopic*. Corresponding equivalence classes of virtual knot/link diagrams with respect to the equivalence relation generated by these moves, are called *virtual knots/links*.

There is a one-to-one correspondence between topological and diagrammatic approach. Further information on virtual knots and their association with thickened surfaces can be found in [5–7, 12, 13, 15, 16]. Here we state the following theorem.

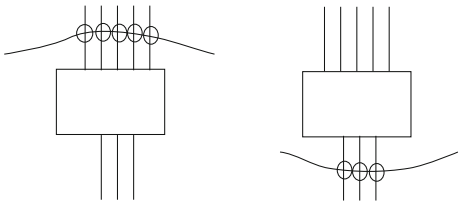


Fig. 12.10 The detour move

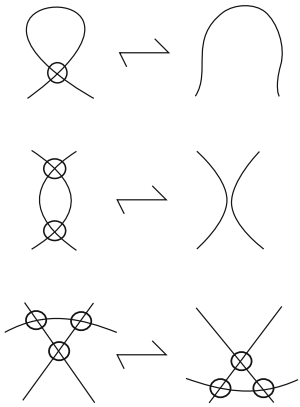


Fig. 12.11 Local detour moves

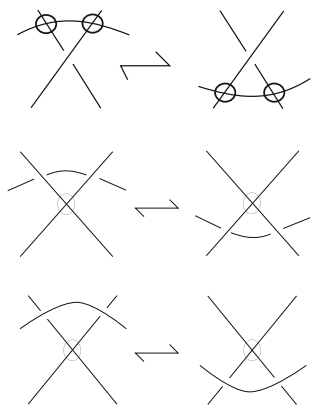


Fig. 12.12 The virtual forbidden moves

Theorem 12.3 ([6, 11, 16, Theorems 1, 3.3, 4.1]) *Two virtual link diagrams are virtually isotopic if and only if their surface embeddings are equivalent up to isotopy and diffeomorphism of the surface, and addition/removal of empty handles.*

12.3.3 A Transition to Virtual Knot Theory

A knotoid diagram in S^2 represents a virtual knot as pointed out in [17]. The endpoints of a knotoid diagram can be connected with an embedded arc in S^2 in the virtual fashion, that is, a virtual crossing is created every time the connection arc crosses through a strand of the diagram. The resulting diagram is a virtual knot diagram that can also be represented in a torus without any virtual crossings by attaching a 1-handle to the sphere of the diagram which holds the connection arc.

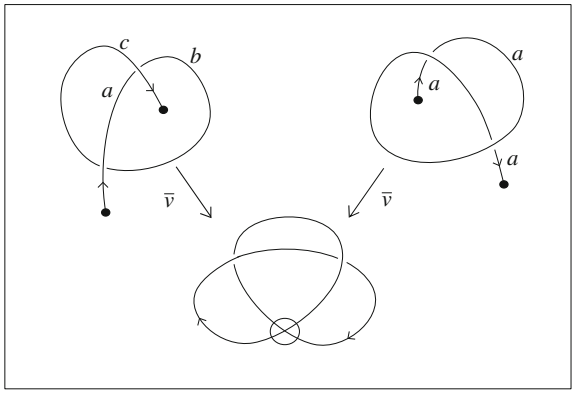


Fig. 12.13 Nonequivalent knotoid diagrams with the same virtual closure

Connecting the endpoints of a knotoid diagram in S^2 in the virtual fashion induces a well-defined map that is called the *virtual closure map* and is denoted by $\overline{v}$,

$$\overline{v}: \{\text{Knotoids in } S^2\} \to \{\text{Virtual knots of genus} \leq 1\}.$$

Figure 12.13 illustrates a pair of knotoid diagrams whose virtual closures are the same virtual knot. It can be shown that the knotoid group [17] of the left hand side knotoid diagram has the presentation $< x, y | x^2 = y^3 >$ (it is isomorphic to the knot group of the trefoil knot). On the other hand, the knotoid group of the right hand side knotoid diagram is isomorphic to $\mathbb{Z}$. Since the knotoid group is a knotoid invariant [17], we conclude that these two diagrams are nonequivalent knotoid diagrams. One other way to distinguish these two knotoid diagrams is the tricolorability of knotoids. The tricolorability of classical knots can be extended to an invariant of knotoids by applying the rules of tricolorability directly to knotoid diagrams. A knotoid diagram K in S^2 is *tricolorable* if each overpassing strand of K (a strand of K that initiates at an undercrossing and terminates at the next undercrossing or terminates at the head of K or initiates at the tail of K and terminates at the next undercrossing) can be colored with one of three colors with respect to the following rules.

- At least two colors must be used.
- At each crossing, the three incident strand should be colored either with the same color or with three different colors.

In Fig. 12.13 shows that the left hand side knotoid diagram is tricolorable with the colors a, b, c and the right hand side knotoid diagram is not tricolorable; the diagram can be colored with only one color. Since tricolorability is a knotoid invariant it follows that they are not equivalent knotoid diagrams. Therefore, the virtual closure map is not an injective map [3]. We will discuss on tricolorability and in general coloring of knotoids in detail in a subsequent paper.

In our paper [3] we conjectured that there are genus 1 virtual knots that do not lie in the image of the virtual closure map. We have proved this conjecture by examining

of the surface states of the surface bracket polynomial of virtual knots which lie in the image of the virtual closure map. Our proof will appear in [4]. In this paper we state the following.

Proposition 12.2 ([4]) *The virtual closure map is not surjective.*

The virtual closure map, being a well-defined map forms a machinery to define knotoid invariants by virtual knot invariants. In fact for any invariant of a virtual knot, denoted by Inv, we can define a knotoid invariant, denoted by I through the following formula.

$$I(K) = \text{Inv}(\overline{v}(K)), \text{ for a knotoid } K \text{ in } S^2.$$

In fact, many invariants obtained from the virtual closure can be defined directly for the knotoids in their own category, without the consideration of the virtual closure map. This is the case for the affine index polynomial and the arrow polynomial as described in the next section.

Remark 12.1 There is a rich subject of *virtual knotoids* where we allow knotoid diagrams with virtual crossings. This subject was firstly proposed in [17] and discussed briefly in [3] and will be the subject of further papers.

12.4 Estimation of Height

12.4.1 Affine Index Polynomial of a Knotoid

The affine index polynomial of a knotoid is constructed in terms of weights assigned to crossings. The *underlying flat diagram* of a knotoid diagram is obtained by omitting the over/under-data at each crossing of the knotoid diagram and turning them into *flat crossings*. An *arc* of a flat knotoid diagram either connects an endpoint of the diagram to a flat crossing or connects a flat crossing to the next flat crossing. There is an integer labeling rule assigned to the arcs of the underlying flat diagram of a knotoid diagram which is set as follows. If an arc crosses a (flat) crossing towards right then its label is decreased by one and if it crosses a node towards left then its label is increased by one. See Fig. 12.14 for the labeling at a flat crossing.

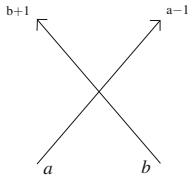


Fig. 12.14 Integer labeling at a flat crossing

Let K be a knotoid diagram. The labeling procedure begins by the arc that is adjacent to the tail of K and it is labeled by 0 conventionally. Note that the adjacent arcs to the tail and the head take the same label.

There are two integer outcomes by the labeling rule at each crossing of K. The *positive* and *negative weights* of a crossing c, denoted by $w_+(c)$ and $w_-(c)$, respectively, are defined as the differences of the labels around c. That is,

$$w_+(c) = a - (b+1)$$
$$w_-(c) = b - (a-1)$$

where a and b are the labels for the left and the right incoming arcs at the corresponding node to c, respectively.
The *weight* of c is defined as

$$w_K(c) = \begin{cases} w_+(c), & \text{if the sign of } c \text{ is positive} \\ w_-(c), & \text{if the sign of } c \text{ is negative.} \end{cases}$$

Definition 12.6 The *affine index polynomial* of a knotoid diagram K is defined by the equation,

$$P_K(t) = \sum_c \text{sign}\,(c)(t^{w_K(c)} - 1) = \sum_c \text{sign}\,(c)t^{w_K(c)} - \text{wr}(K)$$

where $\text{wr}(K)$ is the writhe of K.

Theorem 12.4 ([3, Theorem 5.1]) *The affine index polynomial is an invariant of knotoids in* S^2.

Proof It is sufficient to check the change in the weights under oriented $\Omega_{i=1,2,3}$-moves. The labeling is uniquely inherited by these moves. It is left to the reader the check the crossing added/removed by an Ω_1-move has zero weight. Two crossings added/removed by an Ω_2-move have opposite weights so their contributions cancel each other. The weights of the crossings in an Ω_3-move do not change. Therefore the polynomial remains the same under these moves.

Theorem 12.5 ([3, Theorem 4.12]) *Let K be a knotoid in S^2. The height of K is greater than or equal to the maximum degree of the affine index polynomial of K.*

Proof (Sketch of the proof) The proof of the theorem relies mostly on the following observation. Each crossing of a knotoid diagram determines a unique loop (a continuous path obtained by traversing the diagram starting and ending at a crossing accordingly to the orientation of the diagram) throughout the diagram. A loop that is determined by a crossing is called the *loop at the crossing*. The algebraic intersection number of a loop at a crossing, with other strands of the diagram is equal to either the positive weight or the negative weight of that crossing, depending to the orientation of the loop. On the left hand side of Fig. 12.15, a small portion of the loop $l(C)$ at

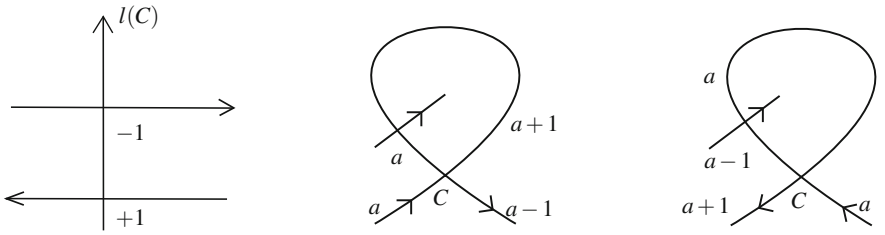


Fig. 12.15 Two possible loops at the crossing C with different orientations

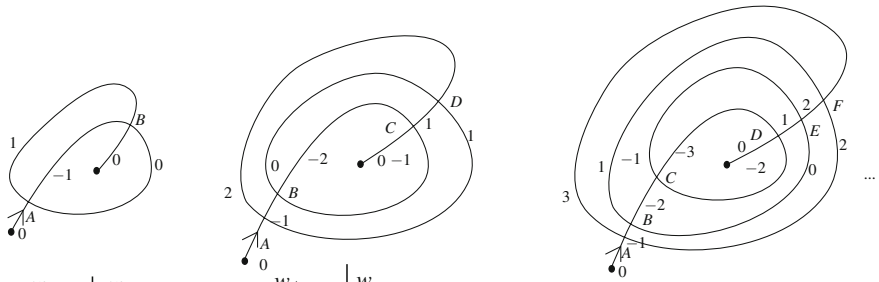
Fig. 12.16 Seifert smoothing of a crossing

a crossing C is illustrated and we see the convention for the algebraic intersection number of the loop with other strands. On the right-hand side of the figure, two possible types of loops at C that may be observed accordingly to the orientation of the diagram, are shown. The algebraic intersection number of the strands shown in the figure with the loops are $+1$ and -1, respectively. By the same figure, it can also be verified that $w_-(C) = +1$ for the first loop that is oriented counterclockwise and $w_+(C) = -1$ for the second loop that is oriented clockwise (See [3] for more details).

Let $\tilde{K}$ be a representative knotoid diagram of K. A crossing of $\tilde{K}$ is a *maximal weight crossing* if $w_{\tilde{K}}(c)$ is maximal among the weights of crossings of $\tilde{K}$. Let c be a maximal weight crossing of $\tilde{K}$. All crossings which are seen twice along the loop at the crossing c are smoothed in the oriented way (*Seifert smoothing*), as shown in Fig. 12.16.

This smoothing results in disjoint oriented embedded circles in S^2 (*Seifert circles*) and a long segment containing the endpoints with an orientation on it from the tail to head. Let $I_{\tilde{K}}$ be the algebraic intersection number of the long segment with the Seifert circles. It is observed that,

- $|I_{\tilde{K}}| \leq$ the number of the resulting Seifert circles enclosing the endpoints,
- $I_{\tilde{K}}$ is equal to the algebraic intersection number of the loop at c with the rest of the diagram since the crossings on the loop which contribute to the intersection number non-trivially are not smoothed. Then,
  $|I_{\tilde{K}}| = w_{\tilde{K}}(c)$,
- The number of the Seifert circles enclosing the endpoints $\leq h(\tilde{K})$, by the Jordan curve theorem.
  Therefore we have the following inequality,
- $w_{\tilde{K}}(c) \leq h(\tilde{K})$.



	w_+	w_-
A	-1	1
B	1	-1

	w_+	w_-
A	-2	2
B	-1	1
C	2	-2
D	1	-1

	w_+	w_-
A	-3	3
B	-2	2
C	-1	1
D	3	-3
E	2	-2
F	1	-1

Fig. 12.17 Flat spiral knotoid diagrams

The maximum degree of the affine index polynomial is an invariant since the polynomial is an invariant. This implies that there exists a crossing at each representative diagram whose weight is equal to $w_{\bar{K}}(c)$. By applying the same procedure given above to each representative diagram, we can conclude that the maximum degree of the affine index polynomial is a lower bound for the height of K.

Example 12.2 The height of a knotoid represented with a standard n-fold spiral diagram can be read by Theorem 12.5. In particular, the affine index polynomials of the knotoid diagrams K_1, K_2 and K_3 that are overlying the flat diagrams given in Fig. 12.17 with all positive crossings, are $P_{K_1}(t) = t + t^{-1} - 2$, $P_{K_2}(t) = t^2 + t + t^{-1} + t^{-2} - 4$ and $P_{K_3}(t) = t^3 + t^2 + t + t^{-1} + t^{-2} + t^{-3} - 6$, respectively. It can be verified by the figure that the heights of the diagrams K_1, K_2, K_3 are 1, 2, 3, respectively. Then by Theorem 12.5 we conclude that the heights of the knotoids represented by K_1, K_2, K_3 are 1, 2 and 3, respectively. This argument is generalized as follows. The affine index polynomial of a knotoid represented by an n-fold spiral knotoid diagram is of the form $t^n + t^{n-1} + \cdots + t + t^{-1} + \cdots + t^{-(n-1)} + t^{-n} - 2n$ if all crossings of the diagram are positive. The maximal degree of the affine index polynomial is n. Then by Theorem 12.5, the height of the knotoid is at least n. The height of the n-fold spiral diagram is n. Therefore, the height of a knotoid represented by a n-fold spiral diagram is equal to n. This implies that we have an infinite set of knotoids whose heights are given by the maximal degrees of their affine index polynomials.

Proposition 12.3 *The height of a product knotoid k_1k_2, $h(k_1k_2) \geq deg\,P(k_1) + deg\,P(k_2) \geq deg\,P(k_1k_2)$, where $deg\,P(k_i)$ is the maximal degree of the affine index polynomial of k_i, $i = 1, 2$.*

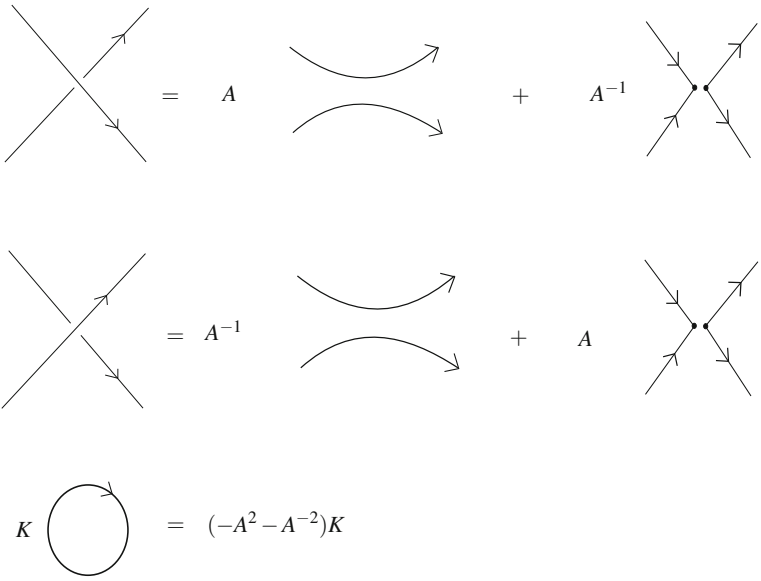


Fig. 12.18 Oriented state expansion

Proof The inequality $deg\,P(k_1k_2) \leq deg\,P(k_1) + deg\,P(k_2)$ can be verified by the reader. By Theorem 12.2 given in Sect. 1.3 we know that the height is an additive invariant. Then it follows that $h(k_1k_2) \geq deg\,P(k_1) + deg\,P(k_2)$.

12.4.2 The Arrow Polynomial of a Knotoid

The arrow polynomial of a knotoid is constructed by the *oriented state expansion* of the bracket polynomial for knotoids; consisting of oriented and disoriented smoothings of crossings, shown in Fig. 12.18.

Smoothing out all the crossings of a knotoid diagram results in *oriented states* containing circular components and a long segment component. There is an extra combinatorial structure on the state components coming out by the disoriented smoothings. This new structure is seen in the form of a pair of cusps, each of cusps has two arcs either going into the cusp or going out from the cusp. Each cusp generates two angles; one is an acute, the other one is an obtuse angle. The local region which is spanned by the acute angle is called the *inside* of the cusp.

There is a list of rules given in Fig. 12.19 to reduce the number of cusps in state components. The rules essentially eliminate two consecutive cusps with insides located on the same side of the segment connecting them. As a consequence of the Jordan curve theorem, all cusps on circular state components of a knotoid diagram are eliminated so that circular components are free of cusps. Each circular and

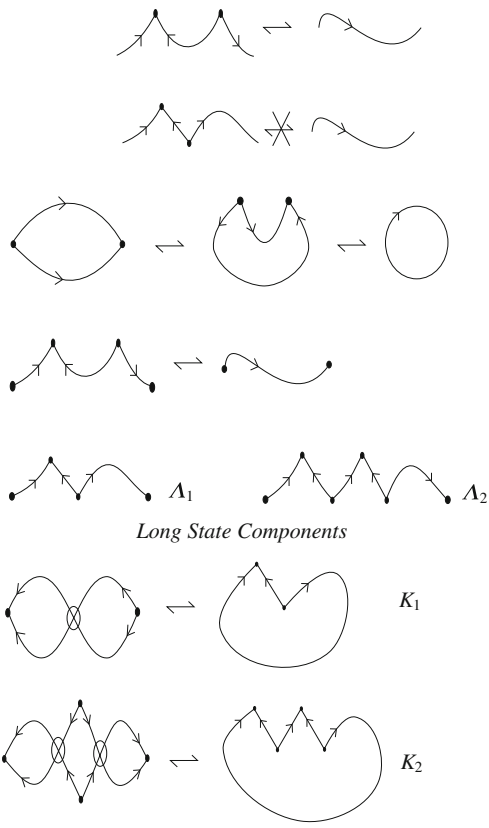


Fig. 12.19 Reduction rules for the arrow polynomial

the long state component without cusps contributes as a $-A^2 - A^{-2}$ factor to the polynomial. The long state components with consecutive cusps whose insides are located at different sides cannot be saved from the cusps and contribute to the arrow polynomial as variables Λ_i.

Definition 12.7 The arrow polynomial of a knotoid diagram K in S^2 is defined as follows.

$$\mathscr{A}[K] = \sum\nolimits_S < K|S > (-A^2 - A^{-2})^{\|S\|-1} \Lambda_i,$$

where the sum runs over the oriented bracket states, $< K|S >$ is the usual vertex weights of the bracket polynomial, $\|S\|$ is the number of components of the state S and Λ_i is the variable associated to the long segment component of S with irreducible cusps.

Theorem 12.6 ([3, Theorem 5.1]) *The normalization of the arrow polynomial by* $(-A^3)^{-\mathrm{wr}(K)}$*, where* K *is a knotoid diagram and* $\mathrm{wr}(K)$ *is its writhe, is a knotoid invariant.*

Proof The proof follows similarly with the proof for the invariance of the normalized arrow polynomial for virtual knots/links [2].

Definition 12.8 The Λ-degree of a summand of the arrow polynomial of a knotoid in S^2 which is of the form, $A^m \Lambda_i$ is equal to i. The *Λ-degree* of the arrow polynomial of a knotoid is defined to be the maximum Λ-degree among the Λ-degrees of the summands.

The arrow polynomial was firstly defined as a virtual knot invariant by H. Dye and the second listed author [2] and independently by Y. Miyazawa [14]. It also utilizes the oriented state expansion and is defined in a similar way as it is defined for knotoids. Oriented state components of a virtual knot diagram consist of circular components. Circular components may have irreducible cusps as illustrated in Fig. 12.19. A circular component with $2i$ irreducible cusps contributes as a K_i-variable to the arrow polynomial. Since oriented states of a classical knot are cusp-free, no K_i-variables occur in the arrow polynomial of classical knots [6]. Similarly, circular components of an oriented state of a knotoid are cusp-free and so no K_i-variables occur in the arrow polynomial of a knotoid in S^2. However, closing the endpoints of a knotoid diagram via the virtual closure map turns Λ_i-variables into K_i-variables that are assigned to the circular components obtained via this closure.

Definition 12.9 The *K-degree* of a summand of the arrow polynomial of a virtual knot which is of the form, $A^m(K_{i_1}{}^{j_1} K_{i_2}{}^{j_2} \dots K_{i_n}{}^{j_n})$ is equal to

$$i_1 \times j_1 + \cdots + i_n \times j_n.$$

The *K-degree* of the arrow polynomial of a virtual knot is the maximum monomial degree in K.

Definition 12.10 The least number of virtual crossings that a virtual knot can have among its virtual equivalence class is the *virtual crossing number*.

Theorem 12.7 ([2, Theorem 2.3]) *The virtual crossing number of a virtual knot/link is greater than or equal to the maximal K-degree of the arrow polynomial of that virtual knot/link.*

Theorem 12.8 ([3, Theorem 5.4]) *The height of a knotoid K in S^2, $h(K)$ is greater than or equal to the Λ-degree of its arrow polynomial.*

Proof Closing a knotoid diagram virtually corresponds, in the states, to closing the endpoints of the long state components in the virtual fashion. Therefore the Λ_i-variables assigned to long state components with surviving cusps transform to K_i-variables assigned to the circular components with surviving cusps in the arrow polynomial of the virtual knot obtained by the virtual closure. By this observation and Theorem 12.7, we have the following inequality for the knotoid k.

$$\text{The } \Lambda\text{-degree of } \mathscr{A}[K] \leq \#\text{ of virtual crossings of the knot } \overline{\nu}(K).$$

The least number of virtual crossings obtained by closing a knotoid diagram virtually, is equal to the height of that diagram. Thus,

$$\#\text{ of virtual crossings of the knot } \overline{v}(K) \leq h(\tilde{K}),$$

for any knotoid diagram $\tilde{K}$ representing K. This gives the following inequality,

$$\text{The } \Lambda - \text{degree of } \mathscr{A}[K] \leq h(\tilde{K}).$$

The inequality above holds for any knotoid diagram in the equivalence class of K since the Λ-degree of the polynomial is invariant under $\Omega_{i=1,2,3}$-moves. Therefore we have,

$$\text{The } \Lambda - \text{degree of } \mathscr{A}[K] \leq h(K).$$

Proposition 12.4 *The height of a product knotoid k_1k_2, $h(k_1k_2) \geq deg_\Lambda \mathscr{A}[k_1] + deg_\Lambda \mathscr{A}[k_2] \geq deg_\Lambda \mathscr{A}[k_1k_2]$, where $deg_\Lambda \mathscr{A}[k_i]$ is the Λ-degree of $\mathscr{A}[k_i]$, $i = 1, 2$.*

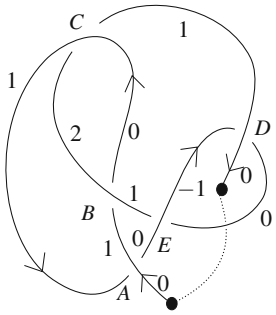
Proof It can be verified by the reader that $deg_\Lambda \mathscr{A}[k_1k_2] \leq deg_\Lambda \mathscr{A}[k_1] + deg_\Lambda \mathscr{A}[k_2]$. By Theorem 12.2 given in Sect. 1.3 we have $h(k_1k_2) = h(k_1) + h(k_2)$. Then by Theorem 12.8 it follows that $h(k_1k_2) \geq deg_\Lambda \mathscr{A}[k_1] + deg_\Lambda \mathscr{A}[k_2]$.

12.4.3 A Discussion on Two Polynomials

The arrow polynomial coincides with the bracket polynomial of knotoids [3, 17] if the Λ_i-variables assigned to long segment components of oriented states are set to be equal to 1. Thus the arrow polynomial is a generalization of the bracket polynomial of knotoids. The affine index polynomial uses the flat biquandle structure [8] of a knotoid diagram. It is quite a different concept than both the arrow and the bracket polynomial.

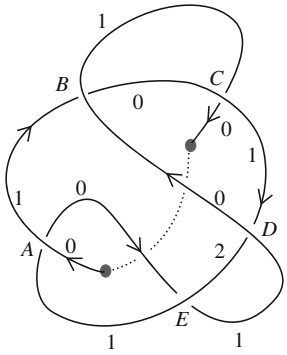
Two estimations of the height given by the affine index and the arrow polynomials are used to determine the height of many knotoids. Here we present one example where the arrow polynomial gives a more accurate estimation of the height than the affine index polynomial and another example where both polynomials are not sufficient to make an exact estimation of the height.

Example 12.3 The knotoid diagram K in Fig. 12.20 represents the knotoid listed as knotoid 5.7 [1]. The affine index polynomial of the knotoid 5.7 is trivial (the reader can verify this by the weight chart of K given in the figure). However, the arrow polynomial of the knotoid is nontrivial. In fact we have, $\mathscr{A}[K] = (-A^{-3} + A - 2A^5 + A^9) + (A^{-9} - 2A^{-5} + 2A^{-1} - 2A^3 + A^7)\Lambda_1$. Thus the Λ-degree of the arrow polynomial of the knotoid 5.7 is 1. Also, it is clear that the height of the diagram K is one. By Theorem 12.8, it follows that the height of the knotoid 5.7 is 1 and this knotoid is in fact a proper knotoid.



	w_+	w_-
A	0	0
B	−1	1
C	1	−1
D	1	−1
E	−1	1

Fig. 12.20 The knotoid diagram K and its weight chart



	w_+	w_-
A	0	0
B	0	0
C	0	0
D	−1	1
E	1	−1

Fig. 12.21 The knotoid diagram K and its weight chart

Example 12.4 The knotoid diagram K given in Fig. 12.21 represents the knotoid listed as knotoid 5.24 [1]. It can be verified that the affine index polynomial of the knotoid is $P_K(t) = t + t^{-1} - 2$. The arrow polynomial of the knotoid is $\mathscr{A}[K]=(-A^{-7} + A^{-3} - A - A^5 + A^9) + (2A^{-1} - 3A^3 + A^7)\Lambda_1$. Both the maximal degree of the affine index polynomial polynomial and the Λ-degree of the arrow polynomial are equal to 1. Thus the height of the knotoid is at least 1, by Theorems 12.5 and 12.8. It is clear that the diagram K has height 2. Thus we have $1 \leq h(K) \leq 2$. We conclude that the knotoid 5.24 is a proper knotoid but both the affine index polynomial and the arrow polynomial cannot give an exact height estimation for this knotoid.

Acknowledgements The first author would like to thank her supervisor Sofia Lambropoulou for several fruitful discussions and for her suggestion of the subject of knotoids for the author's PhD study.
This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES: Reinforcement of the interdisciplinary and/or inter-institutional research and innovation.

References

1. Bartholomew, A.: Andrew Bartholomew's Mathematics Page: Knotoids (2015). http://www.layer8.co.uk/maths/knotoids/index.htm
2. Dye, H.A., Kauffman, L.H.: Virtual crossing number and the arrow polynomial. J. Knot Theory Ramif. **18**(10), 1335–1357 (2009)
3. Gügümcü, N., Kauffman, L.H.: New invariants of knotoids. Eur. J. Combin. **65**C, 186–229 (2017). https://doi.org/10.1016/j.ejc.2017.06.004
4. Gügümcü, N., Kauffman, L.H.: Parity in Knotoids, (in preparation)
5. Kauffman, L.H.: Virtual knot theory. Eur. J. Combin. **20**, 663–690 (1999)
6. Kauffman, L.H.: Introduction to virtual knot theory. J. Knot Theory Ramif. **21**(13), 37 (2012)
7. Kauffman, L.H.: Detecting virtual knots. Atti. Sem. Mat. Fis. Univ. Modena, **49**, (Suppl.), 241–282 (2001)
8. Kauffman, L.H.: An affine index polynomial invariant of virtual knots. J. Knot Theory Ramif. **22**(4), 30 (2013)
9. Kauffman, L.H.: Knots and Physics. Series on Knots and Everything, 4th edn., vol. 53, pp. xviii+846. World Scientific Publishing Co. Pte. Ltd., Hackensack (2013)
10. Kauffman, L.H.: New invariants in the theory of knots. Am. Math. Mon. **95**, 195–242 (1988)
11. Kamada, N., Kamada, S.: Abstract link diagrams and virtual knots. J. Knot Theory Ramif. **9**(1), 93106 (2000)
12. Kuperberg, G.: What is a virtual link? Algebraic Geometric Topol. **3**, 587–591 (2003)
13. Manturov, V.O., Ilyutko, D.P.: The State of Art: Virtual Knots. Series on Knots and Everything, vol. 51. World Scientific Publishing Co.Pte. Ltd., Hackensack (2013)
14. Miyazawa, Y.: A multivariable polynomial invariant for unoriented virtual knots and links. J. Knot Theory Ramif. **17**(11), 1311–1326 (2008)
15. Satoh, S.: Virtual knot presentation of ribbon torus-knots. J. Knot Theory Ramif. **9**(4), 531–542 (2000)
16. Scott Carter, J., Kamada, S., Saito, M.: Stable equivalence of knots and virtual knot Cobordisms, Knots 2000 Korea, Vol. 1 (Yongpyong). J. Knot Theory Ramifications **11**(3), 311–322 (2002)
17. Turaev, V.: Knotoids. Osaka J. Math. **49**(1), 195–223 (2012)

Chapter 13
Fourier Braids

Stephan Klaus

Abstract By the closure operation, knots can be represented by cyclic braids, which can be unfolded as periodic complex valued functions. Their description by Fourier series allows an approximation by finite Laurent polynomials $g(z)$. We define an algebraic discriminant $\Delta_g^n(z)$, such that an n-braid is given by those $g(z)$ satisfying the condition (S) of having all roots not on the unit circle. We study property (S) from the algebraic and topological viewpoint. Using further algebraic conditions for $g(z)$ we obtain algebraic representations of cyclic braids in thickened surfaces, which represent periodic boundary conditions.

Keywords Braid · Knot · Fourier degree · Laurent polynomial · Discriminant Surface knots · Periodic boundary conditions

MSC 2010 12D10 · 42A05 · 42A10 · 57M25 · 57M27

13.1 Cyclic Braids and the Closure Operation

We note that this research is based on earlier results of the author on algebraic constructions of knots given in [2–4, 6]. The author has also given a talk on these results in the Oberwolfach Workshop 2014 on Algebraic Structures in Low-Dimensional Topology [5].

Because of Alexander's Theorem, knots and links can be obtained by closing braids. We recall the basic notions. Let

$$C^n(\mathbb{C}) := \{(z_1, z_2, \ldots, z_n) \mid z_i \in \mathbb{C}, z_i \neq z_j \ \forall i, j\}$$

S. Klaus (✉)
Mathematisches Forschungsinstitut Oberwolfach,
Schwarzwaldstrasse 9-11, 77709 Oberwolfach-Walke, Germany
e-mail: klaus@mfo.de

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_13

be the **ordered configuration space** of n points in the complex plane. There is a free operation of the symmetric group Σ_n on $C^n(\mathbb{C})$ and the quotient gives us the (unordered) **configuration space** $C^n(\mathbb{C})/\Sigma_n$. As a base point in these spaces we can use the configuration $(1, 2, \ldots, n)$ or the configuration $(1, \epsilon, \ldots, \epsilon^{n-1})$ with $\epsilon := e^{2\pi i/n}$.

An n-braid is a closed loop in the configuration space, i.e. a continuous function

$$b : S^1 \to C^n(\mathbb{C})/\Sigma_n.$$

In the following, we often jump between the equivalent descriptions $f : I \to X$ with $f(0) = f(1)$ and $f : S^1 \to X$ for periodic functions with values in some topological space X. Here the identification $I/_{0\sim 1} \cong \mathbb{R}/\mathbb{Z} \cong S^1 \subset \mathbb{C}$ is given by the map $t \mapsto e^{2\pi i t}$.

The set of pointed homotopy classes of n-braids starting and ending in a fixed configuration is just the **braid group**

$$Br_n := \pi_1(C^n(\mathbb{C})/\Sigma_n).$$

The group structure is given by loop sum which we denote by $\perp$, i.e. concatenation of braids. If we do not specify the initial configuration as a base point and consider free homotopy classes, then the set of free homotopy classes of braids is given by the set of conjugacy classes of the group Br_n.

In order to define the **closure** operation we use the open unit disc $\dot{D}^2$ instead of the complex plane. As the strands are images of the closed interval I, they always lie in a bounded region and hence we can shrink them to the unit disc by a suitable contraction factor for any braid. (Alternatively, we could use a fixed diffeomorphism $\mathbb{C} \cong \dot{D}^2$.) Then the n strands of the braid b are embedded in the cylinder $D^2 \times I^1 \subset \mathbb{R}^3$ by $(b(t), t)$ and the closure $\hat{b}$ of b is defined by connecting the initial points and end points of the strands in the bottom and top disc of the cylinder, where the connecting paths lie outside the cylinder and have to be 'parallel'.

Here is an equivalent (isotopic) definition of the closure operation which utilizes the torus parametrization:

$$\tau : D^2 \times I \to \mathbb{R}^3$$

$$\tau(z, t) := \begin{pmatrix} \cos(2\pi t)(2 + Re(z)) \\ \sin(2\pi t)(2 + Re(z)) \\ Im(z) \end{pmatrix}.$$

Then the closure $\hat{b}$ is defined by $\tau(b(t), t)$.

For a general n-braid b the closure does not give a knot but a link in $\mathbb{R}^3$. In fact, the strands of a braid b define a permutation $\rho(b)$ of the set of initial points. Then the closure $\hat{b}$ is a knot if and only if $\rho(b)$ is an n-cycle. We call a braid with this property a **cyclic braid**.

In the general case, the link components of $\hat{b}$ correspond to the cycles in the cycle decomposition of the permutation $\rho(b)$. The other extreme is the case of a **pure braid** which is defined by the condition that $\rho(b)$ is the identity. Because of the covering map

$$C^n(\mathbb{C}) \to C^n(\mathbb{C})/\Sigma_n$$

we get an exact sequence of fundamental groups

$$1 \to PBr_n \to Br_n \to \Sigma_n \to 1$$

where $PBr_n := \pi_1(C^n(\mathbb{C}))$ denotes the **pure braid group**.

It is also well-known that the braid group has a presentation by generators and relations

$$Br_n = \langle \sigma_1, \sigma_2, \ldots, \sigma_{n-1} \mid R1, R2 \rangle$$

$$R1 : \sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1} \ \forall i = 1, \ldots, n-2$$

$$R2 : \sigma_i \sigma_j = \sigma_j \sigma_i \ \forall i, j = 1, \ldots, n-2, |i-j| > 1$$

where the σ_i are the standard braid generators (half-twists of the ith and $(i+1)$th strands). The homomorphism ρ is just given by sending σ_i to the transposition of i and $i+1$.

We note that n, the number of strands of a cyclic n-braid, gives a bound for the bridge number of the closure $\hat{b}$. This fact can be seen from the well-known result that the bridge number of a knot $k : S^1 \to \mathbb{R}^3$ can be defined by the following minimum. (See [8] for background in differential topology.) Let $v \in S^2$ be a direction in $\mathbb{R}^3$ and consider the projection of k on the line spanned by v, i.e. $p_{k,v} : S^1 \to \mathbb{R}$, $p_{k,v} := \langle k(t), v \rangle$ (scalar product). By transversality, it is possible to find a direction v and to change k slightly up to isotopy such that $p_{k,v}$ is a Morse-function which has the property that the finite set of singular points consist of local minima and maxima only (i.e. no saddle points) and $p_{k,v}$ takes different values on them. Then the number of local minima equals the number of local maxima because the Euler characteristics of S^1 vanishes. It is well-known that the minimal possible number $deg^M(k) \in \mathbb{N}$ of local minima (where we allow to change k up to isotopy) gives just the bridge number of the knot. Obviously it holds

$$deg^M(k) = 1 \iff k \text{ is the unknot.}$$

Proposition 13.1 *Let b be a cyclic n-braid, then the bridge number of its closure $\hat{b}$ satisfies $deg^M(\hat{b}) \leq n$.*

Proof Consider the closure as above and chose v in the plane spanned by x and y. After a suitable isotopy (e.g., center the braid along a small disc around 0 such that the closed braid is contained in a small tube around the unit circle), each strand contributes with one local minimum (and one local maximum) to the singular points of $p_{k,v}$. ■

13.2 Unfolding of Cyclic Braids

A general n-braid is given by n functions $I \to \mathbb{C}$ which start and end in the same configuration and whose graphs do not intersect. For a cyclic braid b, we can give a representation by only *one* complex periodic function:

Proposition 13.2 *Concatenation of strands defines a homeomorphism from the set of cyclic n-braids to the space of continuous functions*

$$UCB_n := \{f : S^1 \to \mathbb{C} \mid f(\epsilon^k z) \neq f(z)\ \forall z \in S^1, k = 1, 2, \ldots, \lfloor n/2 \rfloor\}$$

with $\epsilon := e^{2\pi i/n}$ *and* $\lfloor r \rfloor$ *the largest integer smaller than or equal a real number r. In particular, the set of free homotopy classes of cyclic n-braids is given by* $\pi_0(UCB_n)$. *The closure of a cyclic n-braid b (such that the associated function f takes values in the unit disk) is given as*

$$\hat{b}(t) = \begin{pmatrix} \cos(2\pi nt)(2 + Re(f(e^{2\pi it}))) \\ \sin(2\pi nt)(2 + Re(f(e^{2\pi it}))) \\ Im(f(e^{2\pi it})) \end{pmatrix}.$$

Proof For a cyclic braid b, we pick one of the points $u \in \mathbb{C}$ of the initial configuration as a start point with time parameter $t = 0$ and then we concatenate the n strands $b_1, b_2, \ldots, b_n$ which are numerated in the order of the associated n-cycle $\rho(b)$, i.e. b_k starts at $\rho(b)^{k-1}(u)$ and ends in $\rho(b)^k(u)$. This defines a function

$$b_1 \perp b_2 \perp \ldots \perp b_n : [0, n] \to \mathbb{C}$$

which starts and ends in the same point u, see Fig. 13.1 (drawing by the author). Rescaling this function to $[0, 1]$ and using $I/_{0\sim 1} \cong S^1$ we obtain a periodic function

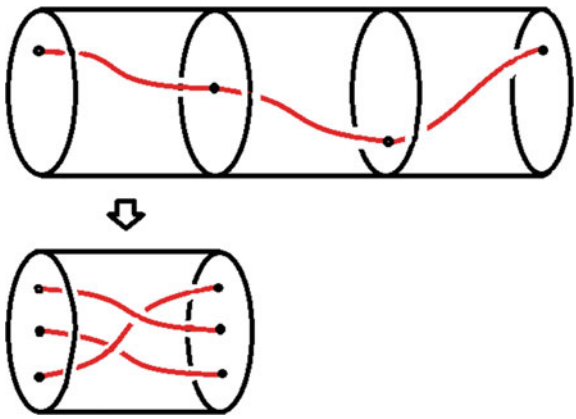


Fig. 13.1 Unfolding of a cyclic three-braid

$f : S^1 \to \mathbb{C}$. Now, multiplication of z by ϵ^k is just a time shift for the function f which changes the start point to the kth strand (according to the cyclic order given by $\rho(b)$). Hence the condition $f(\epsilon^k z) \neq f(\epsilon^l z) \forall z \in S^1$ is equivalent to the condition that the strands b_{k+i} and b_{l+i} do not intersect for all i. Because of $\epsilon^n = 1$ and symmetry, we only need the condition $f(\epsilon^k z) \neq f(z) \forall z \in S^1$, $k = 1, 2, \ldots, \lfloor n/2 \rfloor$ in order to guarantee that the n strands $b_1, b_2, \ldots, b_n$ do not intersect. If a periodic function $f : I \to \mathbb{C}$ satisfies this condition it can in turn be interpreted as a cyclic braid by defining the kth strand as the (rescaled) restriction of f to the segment $[k/n, (k+1)/n]$. This transformation from cyclic n-braids to functions in UCB_n and back is clearly bijective and continuous which proves the statement on π_0. As the closure can be given by $\hat{b}(t) = \tau(b(t), t)$ and as $\hat{b}$ winds n times around the z-axis, the last formula follows. ■

We call the transformation of a cyclic n-braid b to the associated function $f(z) \in UCB_n$ the **unfolding** of the braid (and 'UCB' is the abbreviation for 'unfolded cyclic braids'). Moreover, the defining property of UCB_n can be formulated with the ***n*-th discriminant** of the function $f(z)$

$$\Delta_f^{(n)}(z) := \prod_{k=1}^{\lfloor n/2 \rfloor} (f(\epsilon^k z) - f(z))$$

by the condition $\Delta_f^{(n)}(z) \neq 0$ for all $z \in S^1$.

As an example, we consider the **torus knot**

$$T_{n,m} : \mathbb{R}/\mathbb{Z} \to \mathbb{R}^3$$

$$T_{n,m}(z) := \begin{pmatrix} \cos(2\pi nt)(2 + \cos(2\pi mt)) \\ \sin(2\pi nt)(2 + \cos(2\pi mt)) \\ \sin(2\pi mt) \end{pmatrix}$$

with $(n, m) = 1$. By definition, $T_{n,m}$ is the closure of a cyclic braid $b_{m,n}$ with n strands which 'wind around' m/n times starting from the initial configuration $(1, \epsilon, \ldots, \epsilon^{n-1})$ (see Fig. 13.2 as an example, drawing by the author). In particular, the Morse index satisfies $deg^M(T_{n,m}) \leq n$. As $T_{n,m}$ and $T_{m,n}$ are isotopic in $\mathbb{R}^3$, it holds also $deg^M(T_{n,m}) \leq m$. Hence $deg^M(T_{2,m}) = 2$ for $m > 1$ odd because these torus knots are known to be non-trivial.

Hence the unfolded function has winding number m around the core of the torus and is given by $f(z) = z^m$. Here, the discriminant is given by $\Delta_{z^m}^{(n)}(z) = \prod_{k=1}^{\lfloor n/2 \rfloor} z^m(\epsilon^{km} - 1)$ which is non-zero on S^1.

As another example, we consider the **figure-eight knot** k_4 (see Fig. 13.3, drawing by the author) which can be obtained as the closure of the cyclic 3-braid given by the braid word $\sigma_1\sigma_2^{-1}\sigma_1\sigma_2^{-1}$. In particular $deg^M(k_4) \leq 3$, but it is well-known that actually $deg^M(k_4) = 2$.

Fig. 13.2 Torus knot $T_{3,8}$

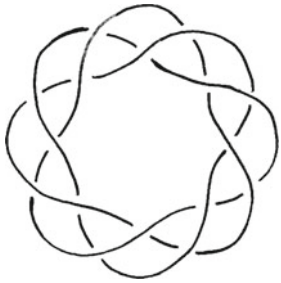


Fig. 13.3 Figure-eight knot

In order to construct a 'nice' explicit unfolding $f(z)$ for the figure-eight knot, we note that we should first construct an unfolding $g(z)$ for the cyclic 3-braid $\sigma_1\sigma_2^{-1}$ (which closes to the unknot!) and then we can set $f(z) := g(z^2)$. The reason is that the braid word for $f(z)$ is the square of the braid word for $g(z)$ and squaring of z has the effect that we run two times through the unfolding function $g(z)$. Now, inspection shows that $g(z) = (z + \bar{z})(1 + z - \bar{z})$ is an element of UCB_3 and serves as an unfolding for $\sigma_1\sigma_2^{-1}$. (It is remarkable that the image of $g(z)$ in $\mathbb{C}$ gives a lemniscate, i.e. the true figure-8 curve!) Hence

$$f(z) = (z^2 + \bar{z}^2)(1 + z^2 - \bar{z}^2) = z^4 - \bar{z}^4 + z^2 + \bar{z}^2$$

gives an unfolding for the figure-eight knot. The discriminant is given by $\Delta_f^{(3)}(z) = (z^4 + \bar{z}^2)(\epsilon - 1) + (-\bar{z}^4 + z^2)(\bar{\epsilon} - 1)$ with $\epsilon = e^{2\pi i/3}$.

The preceding examples show that certain manipulations on unfolding functions have a geometric meaning for the corresponding cyclic braids. Now we list some connections in this direction. Note that $\bar{z} = z^{-1}$ for all $z \in S^1$.

Proposition 13.3 *For cyclic n-braids b and their associated unfoldings $f(z) \in UCB_n$, the following correspondences hold true:*

unfolding:	*cyclic n-braid:*
$a + f(z)$	*translation by a*
$af(z)$	*rotation/dilatation by a*
$\overline{f(z)}$	*mirror braid*
$z^n f(z)$	*Dehn twist*
$f(\bar{z})$	*time reversal, inverse braid*
$f(uz)$	*time/phase shift by t*
$f(z^k)$	*k-fold concatenation*

where $a \in \mathbb{C}$, $u = e^{2\pi it} \in S^1$ *and* $(n, k) = 1$.

Proof The statements on translation, rotation/dilatation and mirror image are clear as they hold in $\mathbb{C}$. If we multiply $f(z)$ by z^n, then the new function is also an element in UCB_n and in the corresponding braid, each strand is multiplied by the function z which just yields a full Dehn twist of the whole braid. We note that a multiplication by a power z^k with $0 < k < n$ would in general not give back a cyclic braid (this can also be checked with the defining property $\Delta_f^{(n)}(z) \neq 0$ which then would not be respected). The statements on $f(\bar{z})$ and $f(uz)$ are also clear as they hold in S^1. The last statement for $f(z^k)$ follows as z^k gives a k-fold covering of S^1. ■

One can ask if the *concatenation of two braids* also corresponds to a similar operation for their unfoldings. Unfortunately, this seems not to be the case. The reason is that the concatenation of cyclic braids in general does not yield again a cyclic braid, as this does not even hold on the level of permutations. In order to obtain from a cyclic braid b again a cyclic braid, we could concatenate b with a pure braid c. Now a pure braid has no unfolding but is given by n loop functions $c_1, c_2, \ldots, c_n$ which do not intersect. Then the concatenation $b \perp c$ is again a cyclic braid which has a (rescaled) unfolding $b_1 \perp c_1 \perp b_2 \perp c_2 \perp \ldots \perp b_n \perp c_n$ and there seems to exist no nice way to express this on the level of unfolded functions.

Also the *Markov stabilization* seems to have no nice description using unfolded functions. Recall that the Markov stabilization of a cyclic n-braid to a cyclic $(n+1)$-braid is defined by concatenation with $\sigma_n^{\pm 1}$. After closure, it corresponds to the first Reidemeister Move.

13.3 The Winding Number of a Cyclic Braid

Now we introduce the winding number of a cyclic n-braid using its unfolding $f(z)$. As the 'kth phase difference'

$$\delta_k(z) := f(\epsilon^k z) - f(z)$$

has to be non-zero on S^1 for all $k = 1, 2, \ldots, \lfloor n/2 \rfloor$, we can define its winding number (induced maps on the fundamental group $\pi_1(S^1) = \mathbb{Z}$)

$$deg(\delta_k : S^1 \to \mathbb{C}-\{0\} \simeq S^1) \in \mathbb{Z}.$$

Proposition 13.4 *The winding numbers $deg(\delta_k)$ all are equal. The winding number of the discriminant $deg(\Delta_f^{(n)} : S^1 \to \mathbb{C}-\{0\})$ is given by $deg(\Delta_f^{(n)}) = deg(\delta_f) \cdot \lfloor n/2 \rfloor$.*

Proof For $\alpha : S^1 \to \mathbb{C}-\{0\}$ note that $deg(\alpha(z)) = deg(\alpha(uz))$ where $u \in S^1$ is any phase shift. Now we have

$$f(\epsilon^k z) - f(z) = (f(\epsilon^k z) - f(\epsilon^{k-1} z)) + (f(\epsilon^{k-1} z) - f(z))$$

where the first bracket on the right side is just δ_1 with a phase shift of ϵ^{k-1} and the second bracket is δ_{k-1}, thus

$$\delta_k(z) = \delta_1(\epsilon^{k-1} z) + \delta_{k-1}(z).$$

We prove the following general result: If $\alpha, \beta : S^1 \to \mathbb{C}-\{0\}$ have the property that also their sum $\alpha + \beta$ takes values in $\mathbb{C}-\{0\}$ then it holds $deg(\alpha) = deg(\beta) = deg(\alpha + \beta)$ for their winding numbers. For the proof, consider the diagram

$$\begin{array}{ccc} \mathbb{C}-\{0\} \times \mathbb{C}-\{0\} & \stackrel{+}{\to} & \mathbb{C} \\ \cup & & \cup \\ (\mathbb{C}-\{0\} \times \mathbb{C}-\{0\}) - D & \stackrel{+}{\to} & \mathbb{C}-\{0\} \end{array}$$

where $D := \{z, -z \mid z \in \mathbb{C}-\{0\}\}$ is the anti-diagonal of $\mathbb{C}-\{0\}$. With $\mathbb{C}-\{0\} \simeq S^1$, it is straightforward to check that $(\mathbb{C}-\{0\} \times \mathbb{C}-\{0\}) - D \simeq S^1$, that the inclusion in $\mathbb{C}-\{0\} \times \mathbb{C}-\{0\}$ induces in π_1 the diagonal $\mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$, and that addition in the lower line of the diagram induces the identity map of $\mathbb{Z}$. The general result follows as (α, β) take values in $(\mathbb{C}-\{0\} \times \mathbb{C}-\{0\}) - D$ by assumption.

Now the statement on $deg(\delta_k)$ follows by induction from the above splitting of δ_k. The statement on the winding number of the discriminant follows as complex multiplication $\mathbb{C}-\{0\} \times \mathbb{C}-\{0\} \to \mathbb{C}-\{0\}$ induces addition in π_1. ■

Definition: This number $deg(\delta_f) \in \mathbb{Z}$ is called the **winding number** of the cyclic n-braid.

13.4 Finite Fourier Approximations

In this section we will consider approximations of the functions in UCB_n by Fourier sums $\sum c_k e^{2\pi i k t}$, i.e. by Laurent polynomials $\sum c_k z^k$. We recall that a continuous periodic function $f : S^1 \to \mathbb{C}$ can be approximated by a finite Fourier sum given any **error bound** $r > 0$, i.e. there exists a Laurent polynomial with

$$|f(z) - \sum_{N^- \leq k \leq N^+} c_k z^k| < r$$

for all $z \in S^1$. We call (N^-, N^+) (with $c_{N^-} \neq 0$ and $c_{N^+} \neq 0$) the **bidegree** of the Laurent polynomial $g(z) := \sum_{N^- \leq k \leq N^+} c_k z^k$.

When approximating an unfolding $f(z) \in UCB_n$ we have to be careful that we stay in the same component of UCB_n.
Definition: For $f(z) \in UCB_n$ we define its **width** by

$$w(f) := \min\{|f(\epsilon^k z) - f(z)| \mid \forall z \in S^1, k = 1, 2, \ldots, \lfloor n/2 \rfloor\},$$

which is a positive real number.

Lemma 13.1 *A continuous function $g : S^1 \to \mathbb{C}$ which approximates $f(z)$ by an error bound smaller than the width $\frac{1}{2}w(f)$ is a function in UCB_n which lies in the same path component as $f(z)$, i.e. $f(z)$ and $g(z)$ yield (freely) isotopic cyclic n-braids.*

Proof $w(f)$ is non-zero as $|f(\epsilon^k z) - f(z)| > 0$ on S^1 and the minimum is realized by some z as S^1 is compact. By the triangle inequality, $g(z)$ also satisfies $g(\epsilon^k z) - g(z) \neq 0$ for all $z \in S^1, k = 1, 2, \ldots, \lfloor n/2 \rfloor$. More general, this is true for all functions h_t in the linear homotopy $h : S^1 \times I \to \mathbb{C}$, which is defined by $h_t(z) := f(z) + t(g(z) - f(z))$. Hence all h_t lie in UCB_n which shows that f and g lie in the same path component of UCB_n and hence are associated to isotopic cyclic n-braids. ■

In particular, this holds for a suitable approximation of $f(z)$ by a Laurent polynomial $g(z) = \sum_{N^- \leq k \leq N^+} c_k z^k$. For $N^-, N^+ \in \mathbb{Z}$, denote by

$$\mathbb{C}[z, z^{-1}]^{(N^-, N^+)}$$

the complex $(N^+ - N^- + 1)$-dimensional vector space of Laurent polynomials $g(z)$ with bidegree (n^-, n^+) such that $N^- \leq n^- \leq n^+ \leq N^+$. We also denote for $N \in \mathbb{N}$ the complex $(2N + 1)$-dimensional vector space of Laurent polynomials of order $\leq N$ by

$$\mathbb{C}[z, z^{-1}]^{\pm N} := \mathbb{C}[z, z^{-1}]^{(-N, +N)}.$$

Now we define:

$$UCB_n^{\pm\infty} := UCB_n \cap \mathbb{C}[z, z^{-1}]$$

$$UCB_n^{(N^-, N^+)} := UCB_n \cap \mathbb{C}[z, z^{-1}]^{(N^-, N^+)}$$

$$UCB_n^{\pm N} := UCB_n \cap \mathbb{C}[z, z^{-1}]^{\pm N}$$

This gives a stratification of the space UCB_n of cyclic n-braids

$$UCB_n^{\pm 1} \subset UCB_n^{\pm 2} \subset UCB_n^{\pm 3} \subset \ldots \subset UCB_n^{\pm\infty} \subset UCB_n.$$

Corollary 13.1 *Finite Fourier approximation defines a sequence of isotopy types of cyclic n-braids*

$$\pi_0 UCB_n^{\pm 1} \to \pi_0 UCB_n^{\pm 2} \to \pi_0 UCB_n^{\pm 3} \to \cdots \to \pi_0 UCB_n^{\pm \infty} \approx \pi_0 UCB_n.$$

Every cyclic n-braid b can be represented up to isotopy by a Laurent polynomial.

Definition: We call the smallest possible order N of a Laurent polynomial representing a cyclic n-braid b up to isotopy its **Fourier degree** $deg^F(b)$.

As an example, the 'canonical' cyclic n-braid $b_{n,m}$ which gives the torus knot $T_{n,m}$ as a closure has $deg^F(b_{n,m}) \leq m$ because z^m serves as an unfolding of b. As the torus knots $T_{n,m}$ and $T_{m,n}$ are isotopic in $\mathbb{R}^3$, we see that deg^F is primarily an invariant of cyclic n-braids (like the number of strands). For the cyclic 3-braid $b = \sigma_1\sigma_2^{-1}\sigma_1\sigma_2^{-1}$ with closure the figure-eight knot, we get $deg^F(b) \leq 4$ by our computation of an unfolding in the preceding section.

In fact, L. Kauffman [1] and A. Trautwein [12] defined the **Fourier degree for knots** in a similar way. A knot can be considered as a periodic function $k : \mathbb{R}^1 \to \mathbb{R}^3$ and it is also possible to approximate such functions by finite Fourier sums, although the details are a little different as k takes values in real 3-dimensional space, not in $\mathbb{C}$ as in our case. Then Kauffman's Fourier degree $Deg^F(k)$ of a knot type is defined as the minimal Fourier order (i.e. largest frequency) that one needs for a Fourier approximation for k (up to isotopy). Here is a connection between our Fourier degree for cyclic braids and that of Kauffman for knots:

Proposition 13.5 *Let b be a cyclic n-braid and* $\hat{b}$ *its knot closure. Then it holds*

$$Deg^F(\hat{b}) \leq n + deg^F(b).$$

Proof Let $d := deg^F(b)$. By definition, there is a unfolding of b given by a Laurent-Polynomial $g(z) = \sum_{-d \leq k \leq d} c_k z^k$. Now we use the explicit formula for the closure map

$$\hat{b}(t) = \begin{pmatrix} \cos(2\pi nt)(2 + Re(\sum_{-d \leq k \leq d} c_k e^{2\pi ikt})) \\ \sin(2\pi nt)(2 + Re(\sum_{-d \leq k \leq d} c_k e^{2\pi ikt})) \\ Im(\sum_{-d \leq k \leq d} c_k e^{2\pi ikt}) \end{pmatrix}.$$

Because of the trigonometric sum and product formulas which can be derived from $e^{2\pi i(r+s)t} = e^{2\pi i(r)t} e^{2\pi i(s)t}$, the largest frequency in a product of two periodic functions is the sum of largest frequencies of the factors. Hence the first two coordinates contain maximal frequencies of order $n + d$, whereas d is the largest frequency in the last coordinate. ■

This result can be interpreted in the following way: The maximal frequency in the unfolded braid is enlarged by n because of the n-fold winding of the braid around the z-axis in order to get the closure.

13.5 Property *S* and the Discriminant Variety

Now we consider the discriminant varieties in the spaces of Laurent polynomials

$$\Delta_n^{\pm\infty} := \{g(z) \in \mathbb{C}[z, z^{-1}] \mid \Delta_g^{(n)}(z) \text{ has a zero on } S^1\},$$

$$\Delta_n^{\pm N} := \{g(z) \in \mathbb{C}[z, z^{-1}]^{\pm N} \mid \Delta_g^{(n)}(z) \text{ has a zero on } S^1\},$$

which are the complements of the subspaces $UCB_n^{\pm\infty}$ and $UCB_n^{\pm N}$.

This leads us to consider the following open subset S of complex polynomials:

$$S := \{p(z) \in \mathbb{C}[z] \mid p(z) \neq 0 \text{ on } S^1\}.$$

By definition, S is the complement in $\mathbb{C}[z]$ of a codimension-1 discriminant variety

$$\Delta := \{p(z) \in \mathbb{C}[z] \mid p(z) \text{ has a zero on } S^1\}.$$

As every Laurent polynomial $g(z)$ can be written as $g(z) = z^{N^-}p(z)$ with a uniquely defined complex polynomial $p(x) := z^{-N^-}g(z)$ (with $p(0) \neq 0$), the condition of having no zeros on S^1 depends only on the polynomial $p(z)$ which can be considered as the essential part of $g(z)$.

Here is a result which connects the winding numbers of $p(z)$ and $g(z)$ (as functions from S^1 to $\mathbb{C}-\{0\}$) with the number of zeros of $p(z)$ in the unit disc.

Proposition 13.6 *The winding numbers of $p(z) \in S$ and $g(z) = z^{N^-}p(z)$ are given by*

$$deg(p) = \textit{number of zeros of } p \textit{ in the open unit disc } \dot{D}^2$$

$$deg(g) = deg(p) + N^-.$$

Proof Let $a(x), b(x) \in \mathbb{C}[z]$ be polynomials which are non-zero on S^1. As complex multiplication $\mathbb{C}-\{0\} \times \mathbb{C}-\{0\} \to \mathbb{C}-\{0\}$ induces addition in π_1, we see that $deg(ab) = deg(a) + deg(b)$. If b has no zeros on the unit disc, we get a homotopy $b : D^2 \to \mathbb{C}-\{0\}$ of $b : S^1 \to \mathbb{C}-\{0\}$ to the constant map $b(0) \in \mathbb{C}-\{0\}$, hence $deg(b) = 0$. Assume that a is normed and has all zeros on the open unit disc, i.e. $a(z) = \prod_{i=1}^{d}(z - z_i)$ with all $z_i \in \dot{D}^2$. Then for $t \in I$ we define

$$a_t(z) := \prod_{i=1}^{d}(z - tz_i) \in \mathbb{C}[z]$$

which gives a homotopy $a_t : S^1 \times I \to \mathbb{C}-\{0\}$ from a to $a_0(z) = z^d$. Hence $deg(a) = deg(z^d) = d$. Now every polynomial $p(z)$ can be split as $p(z) = a(z)b(z)$ with a and b as above and the statement on $deg(p)$ follows. The last statement on $deg(g)$ follows just from $g(z) = z^{N^-}p(z)$. ■

Now we consider S from the algebraic viewpoint. Given a polynomial $p(z) \in \mathbb{C}[z]$, we look for an algebraic invariant δ_p which detects the properties $p \in S$ or $p \in \Delta$. This leads to the following construction. We assume that p is normed with decomposition into linear factors $p(z) = \prod_{i=1}^{d}(z - z_i)$. Then we define

$$\delta_p(z) := \prod_{i=1}^{d}(z - z_i\bar{z}_i).$$

By definition, $\delta_p(z)$ is a polynomial of the same order as $p(z)$ with zeros the real numbers $z_i\bar{z}_i$. Hence $\delta_p(1) = 0$ is equivalent to the condition that $p(z)$ has a zero on S^1, i.e. $p \in \Delta$. Thus

$$\delta_p := \delta_p(1) = \prod_{i=1}^{d}(1 - z_i\bar{z}_i) \in \mathbb{R}$$

serves as an invariant we are looking for. Unfortunately, δ_p cannot be expressed as a polynomial invariant in the coefficients of $p(z)$ and their complex conjugates. It is not possible to apply here the fundamental theorem for symmetric polynomials (in contrast to e.g. $\prod_{i=1}^{d}(z - z_i^2)$) because $\mathbb{C}[z_1, \ldots, z_n, \bar{z}_1, \ldots, \bar{z}_n]^{\Sigma_n}$ is not the polynomial ring in the elementary symmetric functions on the z_i and that on their conjugates $\bar{z}_i$.

As an example, we consider S in low degrees. Clearly, a linear complex polynomial $p(z) = z + a_0$ is in S if and only if $a_0 \notin S^1$. Already the case of a quadratic complex polynomial $p(z) = z^2 + a_1 z + a_0$ demonstrates the difficulty to describe S by conditions on the coefficients. We have $a_1 = -(z_1 + z_2)$ and $a_0 = z_1 z_2$, whereas

$$\delta_p = (1 - z_1\bar{z}_1)(1 - z_2\bar{z}_2) = 1 - (z_1\bar{z}_1 + z_2\bar{z}_2) + z_1\bar{z}_1 z_2\bar{z}_2.$$

Now $z_1\bar{z}_1 z_2\bar{z}_2 = a_0\bar{a}_0$, but $a_1\bar{a}_1 = (z_1\bar{z}_1 + z_2\bar{z}_2) + (z_1\bar{z}_2 + z_2\bar{z}_1)$ and there is no way to express the mixed sum by polynomials in the coefficients a_0, a_1 and their conjugates. Instead it is possible to give δ_p by a complicated real algebraic function of a_0 and a_1.

13.6 Cylic Braids Avoiding Fixed Links and Knots in Spaces with Periodic Boundary Conditions

We have seen that cyclic n-braids are given by unfoldings $f(z)$ which can be chosen as Laurent polynomials

$$UCB_n^{\infty} = \{f(z) \in \mathbb{Z}[z, z^{-1}] \mid \Delta_f^{(n)}(z) \neq 0 \text{ on } S^1\}.$$

Now we consider the additional condition

$$f(z) \neq 0 \text{ on } S^1$$

which means that we exclude the soul $\{0\} \times S^1$ from the solid torus $D^2 \times S^1$. Then the closure operation $b \mapsto \hat{b}$ gives us a knot in $\mathbb{R}^3 - (S^1 \times \{0\})$ which itself is diffeomorphic to an open full torus with an interior point (corresponding to infinity) removed. If we compactify $\mathbb{R}^3$ to S^3, we do not have to remove this inner point, but removing finitely many points from the space where a knot, link or braid is embedded does not change the isotopy classes of embeddings. Hence the algebraic space

$$UCB^{\infty}_{n,1} := \{f(z) \in \mathbb{Z}[z, z^{-1}] \mid f(z)\Delta^{(n)}_f(z) \neq 0 \text{ on } S^1\}.$$

serves as a model for knots in the solid torus. See the work of S. Lambropoulou [7] for more details on the theory of knots in thickened surfaces.

More generally,

$$UCB^{\infty}_{n,m} := \{f(z) \in \mathbb{Z}[z, z^{-1}] \mid \prod_{k=0}^{m-1}(f(z) - k)\Delta^{(n)}_f(z) \neq 0 \text{ on } S^1\}$$

is the algebraic space of unfolded n-braids which avoid in $S^1 \times \mathbb{C}$ the unlink with m components given by $S^1 \times \{k\}$ for $k = 0, 1, \ldots, m-1$. Their homotopy classes just form the subset of cyclic braids in the relative braid group $Br_{n,m}$, see [7]. Hence $UCB^{\infty}_{n,m}$ is a model space for certain knots in the ambient space $\mathbb{R}^3 - (S^1 \times \{0, 1, \ldots, m-1\})$ which is diffeomorphic to the complement of m solid tori in $\mathbb{R}^3$ which are unknotted and unlinked.

In order to obtain knots in the thickened torus $S^1 \times S^1 \times I$, we have to modify our construction by using the Hopf link H in $S^1 \times \mathbb{C}$ instead of the trivial 2-link $(S^1 \times \{0, 1\})$. The reason is that the complement of H is diffeomorphic to the thickened torus (minus the point at infinity). As the Hopf link can be constructed by the embedding $z \mapsto \{0, z\}$ and as each of the n strands has to avoid it, the unfolded cyclic braid has to avoid the n-fold Hopf link. Thus the algebraic space

$$UCB^{\infty}_{n,H} := \{f(z) \in \mathbb{Z}[z, z^{-1}] \mid f(z)(f(z) - z^n)\Delta^{(n)}_f(z) \neq 0 \text{ on } S^1\}$$

is a model for knots in the thickened torus $S^1 \times S^1 \times I$.

In particular, these algebraic spaces approximate knots in the spaces $S^1 \times I \times I$ and $S^1 \times S^1 \times I$ with one- and two-periodic boundary conditions. I.e. these spaces are formed from the cube I^3 by identifying one or two antipodal pairs of faces. It would be interesting to model by this method more general knotted configurations in spaces with periodic boundary conditions. This could produce new connections of Fourier series and Laurent polynomials to applications of knot theory in polymer physics, see [9–11].

Acknowledgements This work was done during research visits with Sofia Lambropoulou at the National Technical University in Athens in 2012 and 2014 and at the Mathematisches Forschungsinstitut Oberwolfach in 2012 and 2013. I would like to thank Sofia Lambropoulou cordially for her hospitality and very helpful discussions on this subject, in particular on knots in thickened surfaces and with periodic boundary conditions.

This research has - by the mentioned visits in Athens - been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES: Reinforcement of the interdisciplinary and/or inter-institutional research and innovation.
I would also like to thank Eleni Panagiotou (U.S.B. California), Gert-Martin Greuel (University of Kaiserslautern), Louis Kauffman (University of Illinois at Chicago) and Hanspeter Kraft (University of Basel) for helpful discussions on certain geometric and algebraic aspects of this theory. Last but not least I would like to thank the unknown referee for his suggestions.

References

1. Kauffman, L.H.: Fourier knots. Ideal Knots, pp. 364–373. Series on Knots and Everything, vol. 19. World Scientific (1997)
2. Klaus, S.: The solid trefoil knot as an algebraic surface, pp. 2–4, featured article in CIM Bulletin (2010) (Coimbra International Center for Mathematics)
3. Klaus, S.: The solid torus knots as algebraic surfaces, preprint, 9 p., Oberwolfach (2013)
4. Klaus, S.: On algebraic, PL and Fourier degrees of knots, preprint, 12 p., Oberwolfach (2013)
5. Klaus, S.: On algebraic, PL and Fourier degrees of knots and braids. In: Oberwolfach Workshop on Algebraic Structures in Low-Dimensional Topology, 25 May - 31 May 2014, organised by Kauffman, L.H., Manturov, V.O., Orr, K.E., Schneiderman, R., Oberwolfach Reports OWR 11.2, Report No. 26, pp. 1434–1438 (2014) (Mathematisches Forschungsinstitut Oberwolfach)
6. Klaus, S.: Möbius strips, knots, pentagons, polyhedra and the SURFER software. In: Decker, W. et al. (eds.) Singularities and Computer Algebra, Festschrift for Gert-Martin Greuel on the Occasion of his 70th Birthday, pp. 161–172. Singularities and Computer Algebra, Springer (2017)
7. Lambropoulou, S.: Braid structures in knot complements, handlebodies and 3-manifolds. In: Series on Knots and Everything: Knots in Hellas -98, Proceedings of the International Conference on Knot Theory and Its Ramifications, vol. 24, pp. 274–289. World Scientific (2000)
8. Milnor, J.: Topology from the Differentiable Viewpoint. Princeton University Press, Princeton (1997)
9. Panagiotou, E.: The linking number in systems with periodic boundary conditions. J. Comput. Phys. **300**, 533–573 (2015)
10. Panagiotou, E., Tzoumanekas, C., Lambropoulou, S., Millett, K.C., Theodorou, D.N.: A Study of the Entanglement in Systems with Periodic Boundary Conditions, Progress of Theoretical Physics, Supplement No. 191, pp. 172–181 (2011)
11. Panagiotou, E., Millett, K.C., Lambropoulou, S.: Quantifying entanglement for collections of chains in periodic boundary conditions models. IUTAM Symposium on Topological Fluid Dynamics: Theory and Applications, Procedia IUTAM, vol. 7, pp. 251–260 (2013)
12. Trautwein, A.K.: An introduction to harmonic knots. Ideal Knots. Series on Knots and Everything, vol. 19, pp. 353–363. World Scientific (1997)

Chapter 14
Molecular Simulation of Ionic Liquids: Complex Dynamics and Structure

Niki Vergadou

Abstract Ionic Liquids (ILs) are organic salts with melting temperatures below 100° C. They are characterized by an exceptional combination of properties that renders them very good candidates for use in many cutting-edge technological applications. The organic and simultaneously ionic nature of the constitutive ions results in diverse interactions that directly affect the microscopic structure and the dynamical behaviour of ILs. Molecular simulation methods using optimized force fields are applied for the study of the complex dynamics and the spatial organization in ILs.

14.1 Introduction to Ionic Liquids

Ionic fluids consist entirely of ions. Within this category of materials, a new class of fluids has emerged in the last few decades under the term "Ionic Liquids" (ILs) [1, 2]. ILs differ from molten salts due to the fact that they are usually composed of one large asymmetric cation and one organic or inorganic anion and the combination of this type of ionic chemical structures leads to salts with lower melting temperatures. Following the description of Paul Walden who was one of the first [3, 4] to observe organic salts in 1914, ILs are considered as the salts that are in the liquid state at room temperature and by convention below 100° C. In Fig. 14.1, some typical anions and cations are shown. There is a vast number of anions and cations that can be combined to form millions of potential ILs [5], revealing an enormous territory that despite having attracted great scientific interest in recent years, still remains largerly unexplored.

ILs are identified as novel designer solvents and advanced materials that can be utilized in a wide range of processes and applications. This fact is predominantly attributed to their great chemical tunability due to the diverse chemical structure of

N. Vergadou (✉)
Molecular Thermodynamics and Modelling of Materials Laboratory, Institute of Nanoscience and Nanotechnology, National Center for Scientific Research "Demokritos", 153 10 Aghia Paraskevi Attikis, Greece
e-mail: n.vergadou@inn.demokritos.gr

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_14

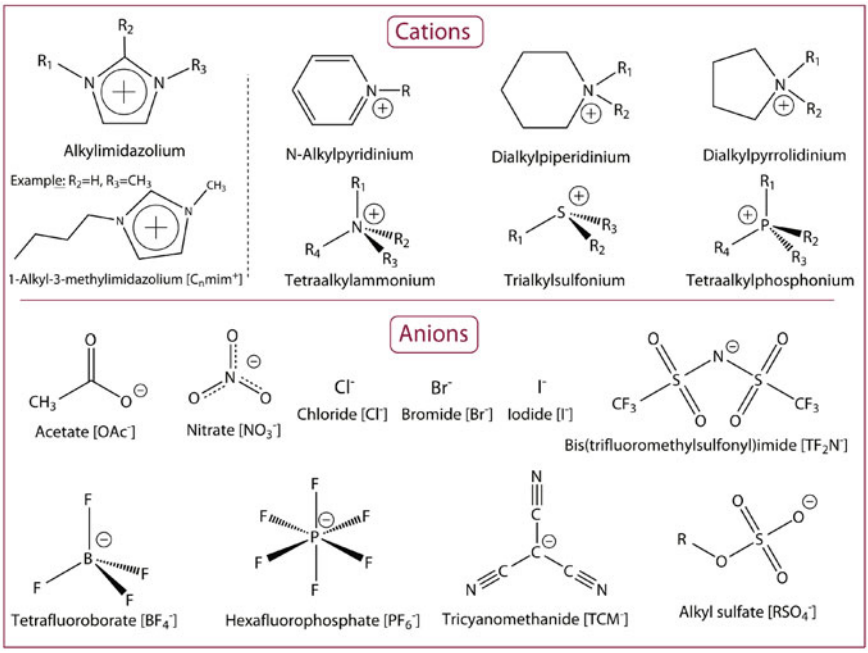


Fig. 14.1 Chemical structure of commonly used cations and anions

the anions and cations involved that enables the design of task specific ILs (TSILs) with controlled macroscopic properties. Apart from this important characteristic, ILs also combine a number of other exceptional properties [6–9] such as negligible vapour pressure, non-flammability, good thermal and electrochemical stability, wide range of temperatures over which they remain in the liquid state as well as very good solvation properties and in many cases high CO_2 absorption and separation capacity [10, 11]. Their range of application constantly grows and extends from green chemistry and electrochemistry to biotechnology, environmental engineering and novel separation processes (Fig. 14.2).

14.2 Molecular Simulation

The plethora of ILs and the multitude of technologies in which ILs can be used, necessitates the unraveling of the underlying mechanisms [12] that are responsible for their macroscopic behaviour. Molecular simulation [13, 14] is based on the fundamental principles of statistical mechanics and provides a unique systematic way in this direction, enabling at the same time the prediction of a number of materials properties. Among the properties that can be calculated from molecular simulations are: (i) ther-

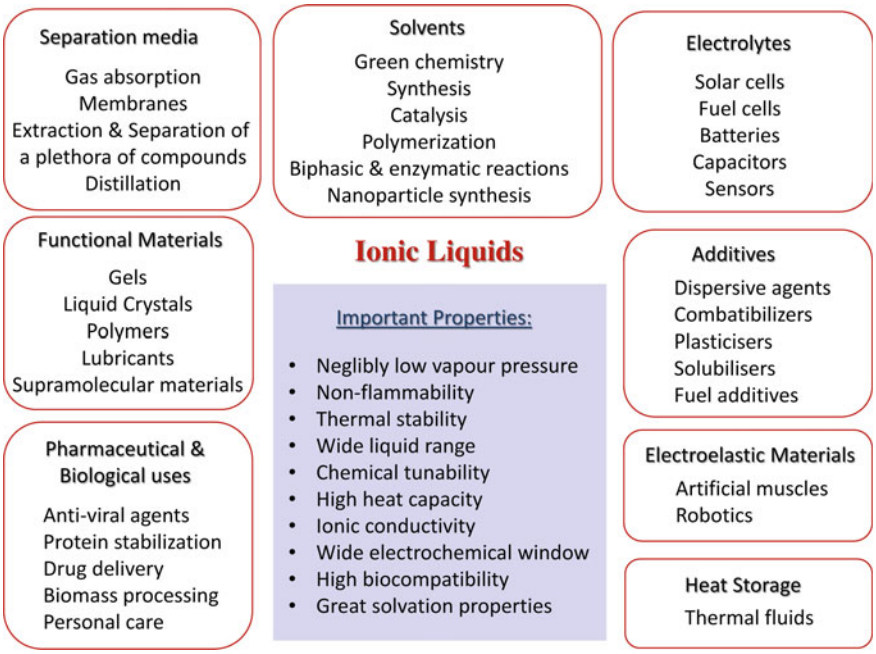


Fig. 14.2 Important properties of ILs and indicative range of applications

modynamic properties such as density, isothermal and isobaric compressibility, heat capacities, Gibbs free energy, Helmholtz free energy, activity coefficients, (ii) structural properties, (iii) dynamical and transport properties including local dynamics, diffusion coefficients, viscosities and ionic conductivities, (iv) surface and interfacial properties such as surface tension, (v) phase equilibria, (vi) mechanical properties and (vi) selectivity and permeability properties.

Various molecular simulation methods can be implemented depending on the materials under study and the time- and length-scales involved in the specific problem at hand, ranging from ab initio quantum mechanical methods, atomistic simulations (Monte Carlo, Molecular Dynamics, Transitions State Theory of Infrequent Events) to mesoscopic methods (Coarse-grained simulations, Kinetic Monte Carlo etc.). This chapter involves the computational study of ILs at the atomistic level using molecular dynamics (MD) simulation [15, 16].

The accuracy in the predictions of molecular simulation relies to a large extend on the force field used for the representation of the inter- and intramolecular interactions in the system. The interaction potential is typically of the form:

$$U = \sum_{bonds} k_b(l - l_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \sum_{n=1}^{4} k_\chi \left[1 + \cos\left(n\chi - \delta\right)\right] + \sum_{impropers} k_\psi \left(\psi - \psi_0\right)^2 + \sum_{\iota=1}^{n-1} \sum_{j>1}^{n} \left\{ 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] + \frac{q_i q_i}{4\pi\varepsilon_0 r_{ij}} \right\} \tag{14.1}$$

where l, θ, χ and ψ denote bond length, bond angle, dihedral and improper angle, respectively, and the subscript "0" refers to the equilibrium values. In the dihedral potential term, parameter n is the multiplicity of the dihedral angle and δ is the phase shift of the dihedral potential over the full range of rotation. Partial charges are denoted by q_i, ε_0 is the vacuum permittivity and ε, σ are the Lennard-Jones (LJ) parameters.

There are several difficulties and challenges in the molecular simulation of ILs [17–19]. The intense chemical diversity that characterizes the ions in ILs, prohibits the use of general force fields and necessitates the development and optimization of system specific interaction potentials. At the same time, complex inter-ionic and strong electrostatic interactions are present in these systems, hence polarizability and charge transfer effects have to be taken into account in order to simulate accurately their behaviour.

14.3 IL Structure and Dynamics

ILs exhibit a wide range of time scales in the relaxation of their various modes of motion. They retain their structural organization at much longer distances compared to ordinary liquids and are characterized by a sluggish dynamical behavior. ILs are glass-forming materials and therefore the prediction of their transport properties, especially at low temperatures, is a very demanding task. The discussion that follows focuses on molecular simulation results on ILs with imidazolium-based cations, specifically the 1-alkyl-3-methylimidazolium ($[C_n mim^+]$) cations, varying the cations alkyl tail, coupled with bis(trifluoromethylsulfonyl)imide ($[TF_2N^-]$) or tricyanomethanide ($[TCM^-]$) anions. MD simulations of several tens of nanoseconds were performed in a wide temperature range, and at atmospheric pressure using classical force fields that have been optimized and extensively validated for these two imidazolium-based IL families [20, 21].

14.3.1 Spatial Organization

ILs are heterogeneous fluids that due to the ionic interactions, form polar and non-polar domains [22]. Structural order is retained at long distances and is clearly

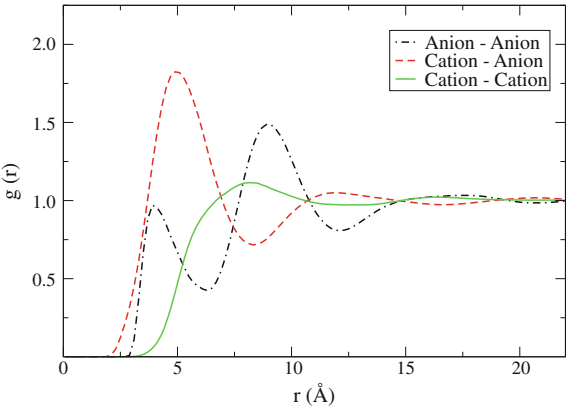


Fig. 14.3 Radial discribution functions (RDF) between the ions centers-of-mass of $[C_8mim^+][TCM^-]$ IL at 298 K [21]

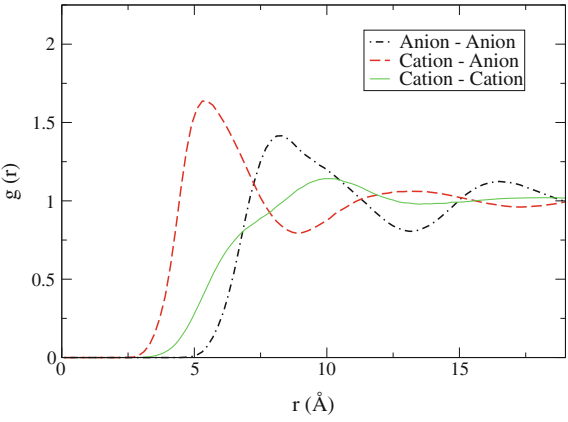


Fig. 14.4 Radial discribution functions (RDF) between the ions centers-of-mass of $[C_8mim^+][TF_2N^-]$ IL at 398 K [23]

depicted by intense oscillations in the radial distribution functions (RDF) [21, 23] of the ions center of mass. The anion-cation RDF reveals the strong ionic coupling of the counterions (Figs. 14.3 and 14.4) and exhibits multiple coordination shells (not in phase with the anion-anion RDF) while cation-cation RDFs show a much broader distribution. The spatial distribution function of the central carbon in $[TCM^-]$ around $[C_4mim^+]$ at 298K (using an iso-surface value equal to 8.1 nm^{-3}) is shown in Fig. 14.5.

The effect of the alkyl tail length on these properties was also investigated and tail aggregation phenomena (Fig. 14.6), which become more evident for the longer alkyl chain lengths, were detected by calculating radial distribution functions between different sites on the ions [20, 21, 23]. The microscopic local structure reflects a spatially heterogeneous environment that evolves from the interplay between short-range collective interactions (non-polar tail groups) and long-range electrostatic interactions (cation's head groups). For a larger number of carbon atoms in the cation's alkyl tail, liquid crystalline-like structures emerge [24].

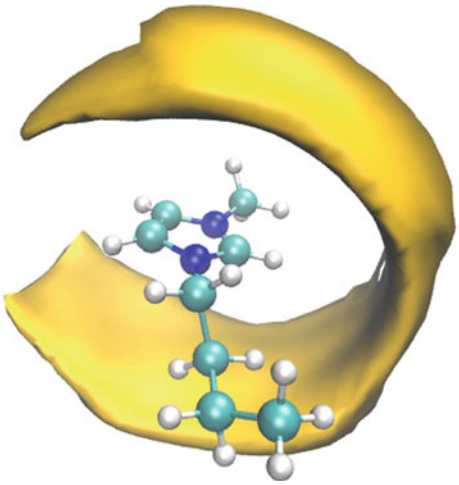


Fig. 14.5 Spatial distribution function of the central carbon in [TCM^-] around [C_4mim^+] (iso-surface value equal to 8.1 nm^{-3})

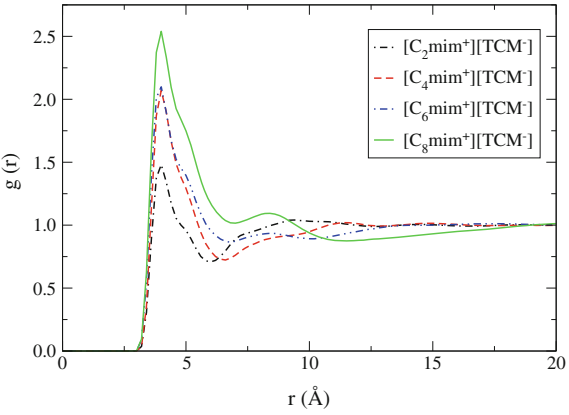


Fig. 14.6 Radial distribution function (RDF) between the terminal carbon atoms in the cation alkyl chain of [C_2mim^+][TCM^-] (black), [C_4mim^+][TCM^-] (red), [C_6mim^+][TCM^-] (blue) and [C_8mim^+][TCM^-] (green) at 298 K [21]

14.3.2 Dynamical Heterogeneity

ILs are viscous liquids and their dynamics often resembles the one of the supercooled liquids. Cooperative motion and caging effects are present in these systems leading to heterogeneity phenomena [25–30] and to a non-Arrhenius behaviour. In glass-forming materials, at short times the particles are trapped in "cages" of adjacent particles, while the escape from the cage takes place at longer time scales as the temperature decreases. Complex and heterogeneous dynamics has been detected in ILs both experimentally [31–35] and computationally [20, 21, 35–43]

An estimation of the time scales at which dynamic heterogeneity appears can be obtained from the non-Gaussian parameters $\alpha_n(t), n = 2, 3, \ldots$ [44], with $\alpha_2(t)$ defined as:

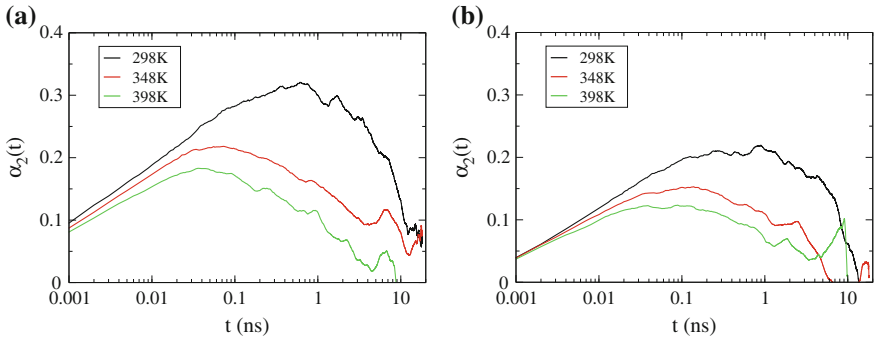


Fig. 14.7 Non-Gaussian parameter $\alpha_2(t)$ at 298, 348 and 398 K for **a** cation and **b** anion of $[C_8mim^+][TF_2N^-]$ IL [20]

$$\alpha_2(t) = \frac{3\langle|\mathbf{r}_i(t) - \mathbf{r}_i(t_0)|^4\rangle}{5\langle|\mathbf{r}_i(t) - \mathbf{r}_i(t_0)|^2\rangle^2} - 1 \tag{14.2}$$

where $|\mathbf{r}_i(t) - \mathbf{r}_i(t_0)|$ represents the displacement of a particle (or an ion's center of mass) at a time interval $t - t_0$ and the brackets denote the mean over all particles. Non-zero values in $\alpha_2(t)$ signify dynamical heterogeneity phenomena with the maximum being only indicative of the time of maximum heterogeneity, as the non-Gaussian behaviour may be preserved at much longer times [45, 46]. The ballistic motion is characterised by zero $\alpha_2(t)$ values and $\alpha_2(t)$ begins to increase at time scales associated with β relaxation, dropping again at the long time limit to zero (α relaxation) [47].

The non-Gaussian parameter is plotted in Fig. 14.7 for the cation and the anion in $[C_8mim^+][TF_2N^-]$ IL at three temperatures [20]. These plots clearly depict a non-Gaussian nature in both ions that is preserved at longer times, with the cation exhibiting a higher a_2 peak than the anion. This is also true for the ILs with the shorter alkyl tails [20] and is indicative of a more pronounced heterogeneous behaviour for the cation. The maxima obtain a higher value as the temperature decreases and at longer times. In case of $[C_nmim^+][TCM^-]$, the anion is the one with the more intense maximum [21, 48] as shown in Fig. 14.8 for $[C_8mim^+][TCM^-]$ cation and anion at 298 K.

A quantification of the dynamic heterogeneity can be achieved by measuring a time-space correlation function corresponding to a classical expression of van Hove function [49]:

$$G(\mathbf{r}, t) = \frac{1}{N}\left\langle \sum_{i=1}^{N}\sum_{j=1}^{N} \delta\left(\mathbf{r} - \mathbf{r}_i(t) + \mathbf{r}_j(0)\right)\right\rangle \tag{14.3}$$

where δ is Dirac delta and the brackets denote time average from an equilibrium trajectory in phase space. Therefore, although $G(\mathbf{r}, t)$ is a dynamic function, it is simultaneously a measure of an equilibrium property. The function $G(\mathbf{r}, t)$ is

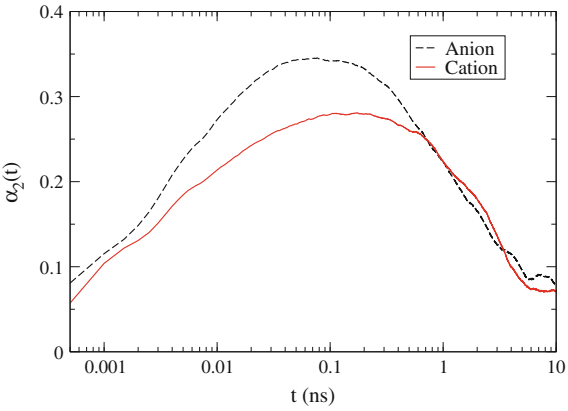


Fig. 14.8 Non-Gaussian parameter $\alpha_2(t)$ for the cation (red) and the anion (black) of [C_8mim^+] [TCM^-] IL at 298K [21]

proportional to the probability that a particle is located at position $\mathbf{r}$ at time t given that a particle was located at $\mathbf{r}(0)$ at $t = 0$ and can be separated in a self (G_s) for $i = j$ and a distinct part (G_d) for $i \neq j$:

$$G(\mathbf{r}, t) = G_s(\mathbf{r}, t) + G_d(\mathbf{r}, t) \tag{14.4}$$

Single-ion dynamics can be quantified by the self-part of the van Hove correlation function [50] that provides the distribution of particle dispacements for different times:

$$G_s(\mathbf{r}, t) = \frac{1}{N}\left\langle \sum_{i=1}^{N}\sum_{j=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i(t) + \mathbf{r}_i(0)) \right\rangle \tag{14.5}$$

At time $t = 0$, $G_s(\mathbf{r}, 0) = \delta$ and for all times, it is normalized by:

$$\int G_s(\mathbf{r}, t)\, d\mathbf{r} = 1 \tag{14.6}$$

At the long time limit, the particle is independent of its initial position:

$$\lim_{r\to\infty} G_s(\mathbf{r}, t) = \lim_{t\to\infty} G_s(\mathbf{r}, t) = \frac{1}{V} \approx 0 \tag{14.7}$$

where V is the system volume. The probability density that a particle displaces by distance r [51] is given by:

$$g_s(r, t) = \int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} G_s(\mathbf{r}, t)\, r^2 \sin\theta d\phi d\theta \tag{14.8}$$

and for an isotropic medium $G_s(\mathbf{r}, t) = G_s(r, t)$ and $g_s(r, t) = 4\pi r^2 G_s(r, t)$.

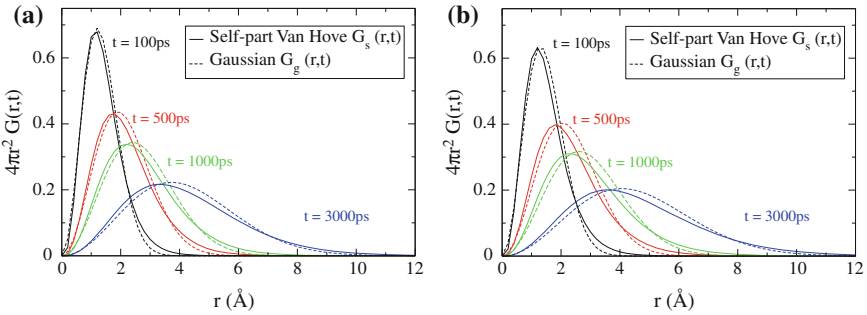


Fig. 14.9 Self-part of the Van Hove function $G_s\,(r, t)$ as a function of distance at 298 K plotted against the expected Gaussian distribution for **a** the cation and **b** the anion of $[C_6mim^+][TCM^-]$

In the case of Fickian diffusion, the self-part of the Van Hove function follows the Gaussian approximation [52]:

$$G_s{}^g\,(r, t) = \left[\frac{3}{2\pi\left\langle \Delta r^2(t)\right\rangle}\right]^{\frac{3}{2}} \exp\left[\frac{-3r^2}{2\left\langle \Delta r^2(t)\right\rangle}\right] \tag{14.9}$$

with $\left\langle \Delta r^2(t)\right\rangle$ being the mean square displacement (MSD) at a time interval t.

In Fig. 14.9, the self-part of the Van Hove function $G_s\,(r, t)$ for (a) the cation and (b) the anion of $[C_6mim^+][TCM^-]$ [21] is plotted against the expected Gaussian distribution at 298 K and for t = 100, 500, 1000, 3000 ps. Deviations of the $G_s\,(r, t)$ from the expected Gaussian are evidence of dynamic heterogeneity, which for the systems under study is present at time intervals typically ranging from a few ps up to several ns. For the longer alkyl tail IL, the non-Gaussian behavior persists at longer times, while these deviations are diminished as the temperature is increased. A multifractal character has been also reported [42] in relation to the heterogeneous nature of ILs.

The divergence between the $G_s\,(r, t)$ and $G_s{}^g\,(r, t)$ curves depicts the existence at intermediate times of ions with "faster" or "slower" mobility than expected based on that distance that each ion has travelled at time t. The crossing points of the two curves at short and long distances are used to identify dynamically distinguishable ions at a specific time interval [47].

The spatial correlation of these subsets of fast and slow ions manifests the occurrence of clustering phenomena between mobile and immobile ions. This is clearly shown in Fig. 14.10, in which the RDFs between the centers of mass of slow anions – slow cations, fast anions – fast cations and fast – slow anions and cations are shown for (a) $[C_4mim^+][TCM^-]$ [21] and (b) $[C_6mim^+][TCM^-]$ at 298 K for t = 500 ps. These RDFs are plotted against the RDF that corresponds to the anion-cation centers of mass as calculated from all ions, exhibiting maxima at the same distances with much higher peaks, though, in the case of ions of same mobility. The evolution of

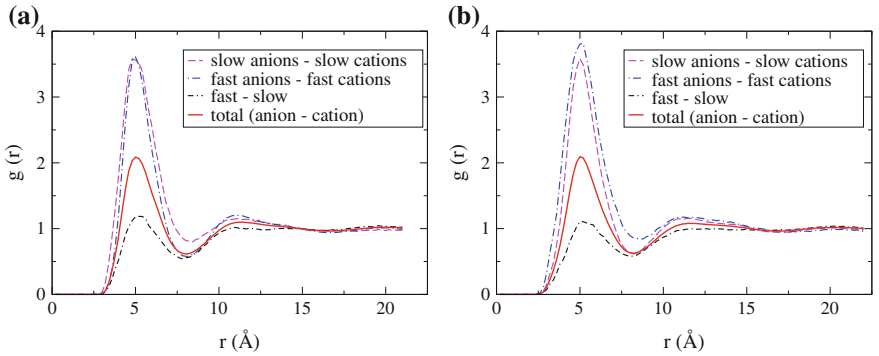


Fig. 14.10 Radial distribution functions (RDF) between the centers of mass of slow anions slow cations (magenta), fast anions fast cations (blue) and fast slow anions and cations (black) plotted against the total RDF (red) calculated for **a** $[C_6mim^+][TCM^-]$ [21] and **b** $[C_6mim^+][TCM^-]$ at 298 K for t = 500 ps. The red line corresponds to the radial distribution functions between the anion-cation centers of mass as calculated from all ions

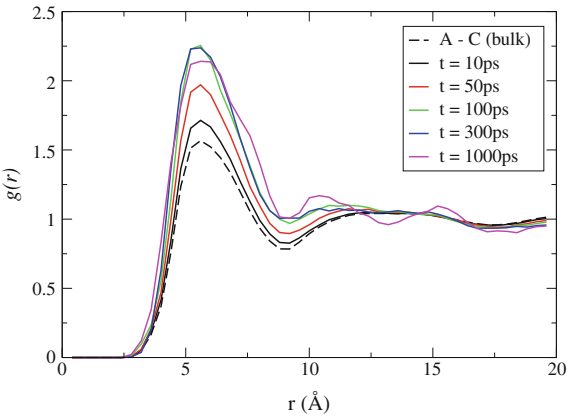


Fig. 14.11 Radial distribution functions between the centers of mass of fast cations and fast anions calculated at 298 K for t = 10, 50, 100, 300 and 1000 ps of $[C_8mim^+][TF_2N^-]$ [20]

the clustering tendency in time is shown in Fig. 14.11 as calculated for fast cations and fast anions of $[C_8mim^+][TF_2N^-]$ at 298 K and t = 10, 50, 100, 300 and 1000 ps [20].

14.3.3 Diffusional Anisotropy

Anisotropy in the ions translational motion is present in some ionic species in ILs [20, 21, 40, 53]. Such phenomena can be traced by examining the ions diffusional motion along specific axes dictated by the geometry of the ions. In Fig. 14.12, four axes are defined for the case of imidazolium-based cations: the vector NN that connects the

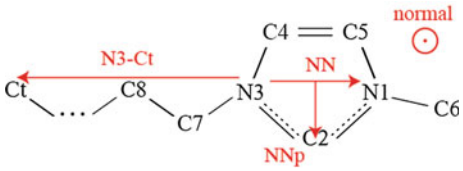


Fig. 14.12 The vectors defined on the imidazolium-based cation along which the translational motion of the center of mass was analyzed

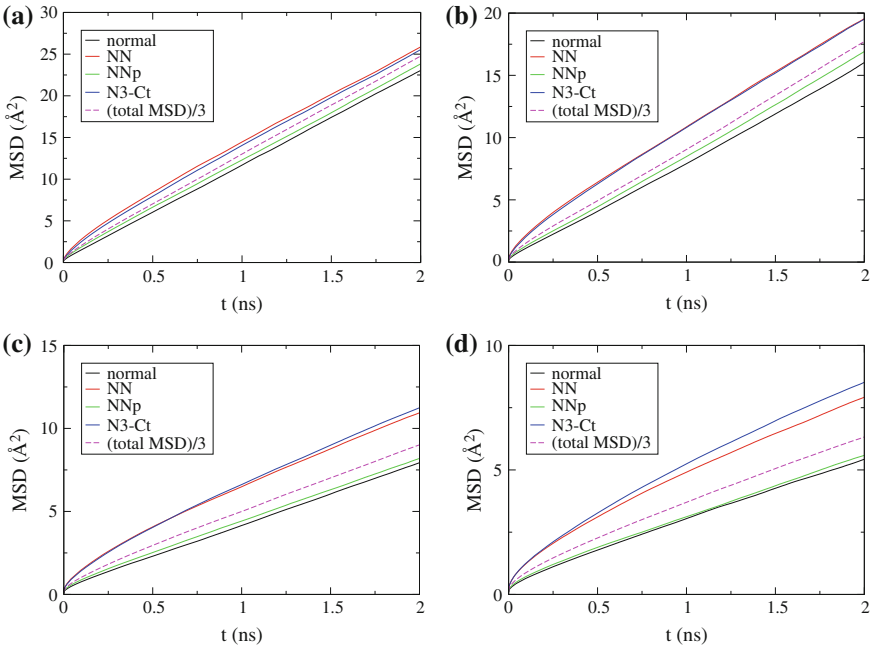


Fig. 14.13 Mean square displacement (MSD) at 298 K calculated along the direction of the vector that is normal to the imidazolium plane (black), the vectors NN (red) and NNp (green), and along the the end-to-end vector N3-Ct of the alkyl tail (blue) of the cation compared to the 1/3 of the total MSD of the center of mass (magenta dashed line) for **a** $[C_2mim^+][TCM^-]$, **b** $[C_4mim^+][TCM^-]$, **c** $[C_6mim^+][TCM^-]$ and **d** $[C_8mim^+][TCM^-]$ [21]

two nitrogen atoms in the imidazolium ring, the normal vector to the imidazolium ring, the vector NNp that is perpendicular to the former two vectors and the end-to-end vector N3-Ct of the alkyl tail.

Diffusional anisotropy can be extracted by comparing the MSD along each axis to the one third of the total MSD, as shown in Fig. 14.13 for the cations in $[C_2mim^+][TCM^-]$, $[C_4mim^+][TCM^-]$, $[C_6mim^+][TCM^-]$ and $[C_8mim^+][TCM^-]$ ILs at 298K. In case of an isotropic diffusional motion the MSD along any axis would coincide with the total MSD/3. For all systems presented, there is a more facile movement along the NN and N3-Ct vectors, which becomes more intense for

the cations with the longer alkyl tails. For $[C_8mim^+][TCM^-]$ IL in particular, the movement along the N3-Ct vector deviates from the one of NN vector, exhibiting at 1.5ns an MSD almost by a factor of two higher than the 1/3 of the total MSD. The movement along the NNp vector and the vector that is normal to the imidazolium plane appears to be rather hindered compared to the other two directions. At 298K, diffusional anisotropy is maintained over long timescales that extend even to the Einstein regime. Similar behaviour is observed for the $[C_nmim^+][TF_2N^-]$ ILs in which, apart from the cation, there is also a preferential movement along the direction of the vector that connects the two sulfurs of the $[TF_2N^-]$ anion [20].

14.3.4 Transport Properties

Long MD simulations in the order of several tens of nanoseconds have been performed for the determination of the transport properties of the ILs under study. The self-diffusion coefficients of the ions were calculated using the Einstein relation [54]:

$$D = \frac{1}{2d} \lim_{t\to\infty} \frac{d}{dt} \left\langle |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2 \right\rangle \tag{14.10}$$

where the brackets indicate the mean square displacement (MSD) over all ions' centers of mass and d is the dimensionality of the conducted diffusivity.

The determination of the diffusion coefficients from an MD simulation presupposes that the system is simulated long enough, so that it has reached the Fickian regime. Normal diffusivity can be identified by a slope equal to unity in the log(MSD) versus log(t) plot. Self-diffusion coefficients were calculated for imidazolium-based ILs under study [20, 21, 48]. The self-diffusion coefficients for the ions in $[C_8mim^+][TCM^-]$ IL are shown in Fig. 14.14 at various temperatures and at atmospheric pressure. The calculated diffusivities for the cation are in excel-

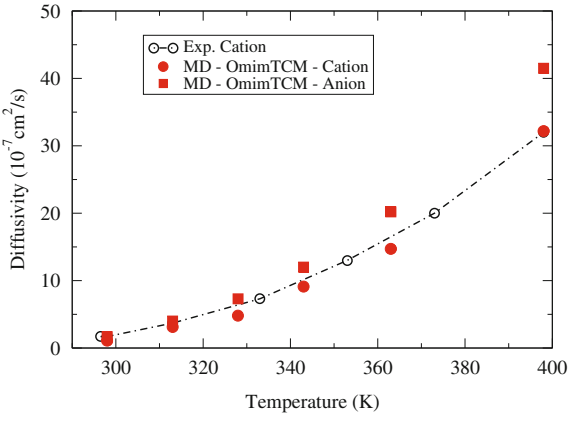


Fig. 14.14 Self-diffusion coefficients [21, 48] of the anion (*squares*) and the cation (*circles*) versus temperature for $[C_8mim^+][TCM^-]$ IL. The open points correspond to NMR experimental measurements for the cation [55]

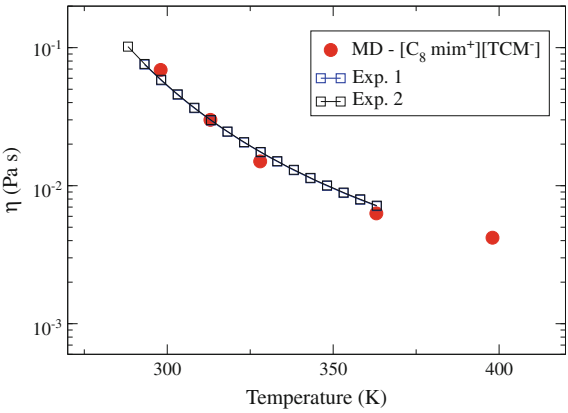


Fig. 14.15 Viscosity calculations from MD simulations (full points) [21, 48] as a function of temperature plotted against experimental data (lines with open points) [48, 56] for $[C_8mim^+][TCM^-]$

lent agreement with NMR experimental measurements [55] of the same IL system, confirming the ability of the recently optimized force field [21] to reproduce the properties of the $[C_nmim^+][TCM^-]$ IL family with great accuracy. No experimental results are currently available for the $[TCM^-]$ anion.

Shear viscosity was extracted from the autocorrelation function of the pressure tensor using the Green-Kubo relation:

$$\eta = \frac{V}{k_B T} \int_0^\infty dt \left\langle P_{\alpha\beta}(t) P_{\alpha\beta}(0) \right\rangle \tag{14.11}$$

where $P_{\alpha\beta}(t)$ is the $\alpha\beta$-element of the pressure tensor at time t ($\alpha \neq \beta$). The temperature dependence of the viscosity is captured very well [20, 21, 48] for all ILs with good agreement with available experimental data. In Fig. 14.15, viscosities calculated from MD in a wide temperature range for $[C_8mim^+][TCM^-]$ are plotted against experimental measurements exhibiting a very good agreement.

14.4 Conclusions

Ionic fluids inherently carry a higher degree of complexity than systems comprised of neutral species. ILs, in particular, exhibit an exceptional combination of properties that originate significantly from their dual organic and ionic nature. A fundamental understanding of the diverse interactions and the microscopic mechanisms that give rise to the non-trivial spatial and dynamical behaviour in ILs is required, especially considering more complex IL systems such as multi-compound IL fluids, ILs as gas separation media [10, 57] and under confinement in solid substrates [58, 59] or IL-IL mixtures [60, 61]. Theoretical and computational studies combined in a synergistic manner with high-resolution experimental techniques should target in elucidating the underlying complex phenomena, extending the fundamental knowledge and reveal-

ing the chemical structure – property relations in this class of advanced materials with enormous applicability in a wide range of state-of-the art technologies.

References

1. Holbrey, J., Seddon, K.: Ionic liquids. Clean Prod. Process. **1**, 223–236 (1999)
2. Wasserscheid, P., Welton, T.: Ionic Liquids in Synthesis, 2nd edn. Wiley-VCH Weinheim, Germany (2008)
3. Gabriel, S., Weiner, J.: Ueber einige Abkömmlinge des Propylamins. Ber. Dtsch. Chem. Ges. **21**, 2669–2679 (1888)
4. Walden, P.: Ueber die Molekulargrösse und elektrische Leitfähigkeit einiger geschmolzener Salze. Bull. Acad. Imp. Sci. (St. Ptersbourg) **8**, 405–422 (1914)
5. Rogers, R.D.: Reflections on ionic liquids. Nature **447**, 917–918 (2007)
6. Kirchner, B. (ed.): Topics in Current Chemistry- Ionic Liquids, vol. 290. Springer, Berlin (2010)
7. Handy, S.T. (ed.): Ionic Liquids - Classes and Properties. InTech, Rijeka, Croatia (2011)
8. Zhang, S., Lu, X., Zhou, Q., Li, X., Zhang, X., Li, S.: Ionic Liquids - Physicochemical Properties. Elsevier, Amsterdam (2009)
9. Visser, A.E., Bridges, N.J., Rogers, R.D. (eds.): Ionic Liquids: Science and Applications. ACS Symposium Series, vol. 1117 (2012)
10. Vergadou, N., Androulaki, E., Economou, I.G.: Molecular simulation methods for CO_2 capture and gas separation with emphasis on ionic liquids. In: Papadopoulos, A.I., Seferlis, P. (eds.) Process Systems and Materials for CO2 Capture: Modelling, Design, Control and Integration, pp. 79–111. Wiley, Chichester (2017)
11. Radmin, M., de Loos, T.W., Vlugt, T.J.H.: State-of-the-Art of CO2 capture with ionic liquids. Ind. Eng. Chem. Res. **51**, 8149–8177 (2012)
12. Weingärtner, H.: Understanding ionic liquids at the molecular level: facts, problems, and controversies. Angew. Chem. - Int. Edition **47**(4), 654–670 (2008)
13. Allen, M.P., Tildesley, D.J.: Computer Simulation of Liquids. Oxford University Press, Oxford (1987)
14. Frenkel, D., Smit, B.: Understanding Molecular Simulation - From Algorithms to Applications. Academic Press, San Diego (2002)
15. Haile, J.M.: Molecular Dynamics Simulation: Elementary Methods. Wiley, New York (1992)
16. Rapaport, D.C.: The Art of Molecular Dynamics Simulation, 2nd edn. Cambridge University Press, Cambridge (2004)
17. Salanne, M.: Simulations of room temperature ionic liquids: from polarizable to coarse-grained force fields. Phys. Chem. Chem. Phys. **17**, 14270–14279 (2015)
18. Batista, M.L.S., Coutinho, J.A.P., Gomes, J.R.B.: Prediction of ionic liquids properties through molecular dynamics simulations. Curr. Phys. Chem. **4**, 151–172 (2014)
19. Dommert, F., Wendler, K., Berger, R., Delle Site, L., Holm, C.: Force fields for studying the structure and dynamics of ionic liquids: a critical review of recent developments. Chem. Phys. Chem. **13**(7), 1625–1637 (2012)
20. Androulaki, E., Vergadou, N., Economou, I.G.: Analysis of the heterogeneous dynamics of imidazolium-based [Tf_2N^-] ionic liquids using molecular simulation. Mol. Phys. **112**, 2694–2706 (2014)
21. Vergadou, N., Androulaki, E., Hill, J.-R., Economou, I.G.: Molecular simulations of imidazolium-based tricyanomethanide ionic liquids using an optimized classical force field. Phys. Chem. Chem. Phys. **18**, 6850–6860 (2016)
22. Wang, Y., Voth, G.A.: Unique spatial heterogeneity in ionic liquids. J. Am. Chem. Soc. **127**, 12192–12193 (2005)
23. Androulaki, E., Vergadou, N., Ramos, J., Economou, I.G.: Structure, thermodynamic and transport properties of imidazolium-based bis(trifluoromethylsulfonyl)imide ionic liquids from molecular dynamics simulations. Mol. Phys. **110**(11–12), 1139–1152 (2012)

24. Ji, Y., Shi, R., Wang, Y., Saielli, G.: Effect of the chain length on the structure of ionic liquids: from spatial heterogeneity to ionic liquid crystals. J. Phys. Chem. B **117**, 1104–1109 (2013)
25. Donati, C., Glotzer, S.C., Poole, P.H., Kob, W., Plimpton, S.J.: Spatial correlations of mobility and immobility in a glass-forming Lennard-Jones liquid. Phys. Rev. E **60**(3), 3107–3119 (1999)
26. Qian, J., Hentschke, R., Heuer, A.: On the origin of dynamic heterogeneities in glass-forming liquids. J. Chem. Phys. **111**(22), 10177–10182 (1999)
27. Ediger, M.D.: Spatially heterogeneous dynamics in supercooled liquids. Annu. Rev. Phys. Chem. **51**, 99–128 (2000)
28. Debenedetti, P.G., Stillinger, F.H.: Supercooled liquids and the glass transition. Nature **410**, 259–267 (2001)
29. Richert, R.: Heterogeneous dynamics in liquids: fluctuations in space and time. J. Phys.: Condens. Matter **14**(23), R703–R738 (2002)
30. Chaudhuri, P., Sastry, S., Kob, W.: Tracking heterogeneous dynamics during the alpha relaxation of a simple glass former. Phys. Rev. Lett. **101**(19), 190601 (2008)
31. Paul, A., Mandal, P.K., Samanta, A.: On the optical properties of the imidazolium ionic liquids. J. Phys. Chem. B **109**(18), 9148–9153 (2005)
32. Aki, S.N.V.K., Brennecke, J.F., Samanta, A.: How polar are room-temperature ionic liquids? Chem. Commun. **5**, 413–414 (2001)
33. Ito, N., Arzhantsev, S., Maroncelli, M.: The probe dependence of solvation dynamics and rotation in the ionic liquid 1-butyl-3-methyl-imidazolium hexafluorophosphate. Chem. Phys. Lett. **396**(1–3), 83–91 (2004)
34. Jin, H., Li, X., Maroncelli, M.: Heterogeneous solute dynamics in room temperature ionic liquids. J. Phys. Chem. B **111**(48), 13473–13478 (2007)
35. Hu, Z., Margulis, C.J.: Heterogeneity in a room-temperature ionic liquid: persistent local environments and the red-edge effect. Proc. Nat. Acad. Sci. USA **103**(4), 831–836 (2006)
36. Del Pópolo, M.G., Voth, G.A.: On the structure and dynamics of ionic liquids. J. Phys. Chem. B **108**(5), 1744–1752 (2004)
37. Ishida, T., Shirota, H.: Dicationic versus monocationic ionic liquids: distinctive ionic dynamics and dynamical heterogeneity. J. Phys. Chem. B **117**(4), 1136–1150 (2013)
38. Habasaki, J., Ngai, K.L.: Molecular dynamics studies of ionically conducting glasses and ionic liquids: wave number dependence of intermediate scattering function. J. Chem. Phys. **133**(12), 124505 (2010)
39. Urahata, S.M., Ribeiro, M.C.C.: Unraveling dynamical heterogeneity in the ionic liquid 1-Butyl-3-methylimidazolium chloride. J. Phys. Chem. Lett. **1**(11), 1738–1742 (2010)
40. Liu, H., Maginn, E.: A molecular dynamics investigation of the structural and dynamic properties of the ionic liquid 1-n-butyl-3-methylimidazolium bis(trifluoromethanesulfonyl)imide. J. Chem. Phys. **135**(12), 124507 (2011)
41. Hu, Z., Margulis, C.J.: Room-temperature ionic liquids: slow dynamics, viscosity and the red edge effect. Acc. Chem. Res. **40**(11), 1097–1105 (2007)
42. Habasaki, J., Ngai, K.L.: Multifractal nature of heterogeneous dynamics and structures in glass forming ionic liquids. J. Non-Cryst. Solids **357**(2), 446–453 (2011)
43. Jeong, D., Choi, M.Y., Kim, H.J., Jung, Y.: Fragility, Stokes-Einstein violation, and correlated local excitations in a coarse-grained model of an ionic liquid. Phys. Chem. Chem. Phys. **12**, 2001–2010 (2010)
44. Rahman, A.: Correlations in the motion of atoms in liquid argon. Phys. Rev. **136**(2A), A405–A411 (1964)
45. Szamel, G., Flenner, E.: Independence of the relaxation of a supercooled fluid from its microscopic dynamics: Need for yet another extension of the mode-coupling theory. Europhys. Lett. **67**, 779–785 (2004)
46. Reichman, D.R., Rabani, E., Geissler, P.L.: Comparison of dynamical heterogeneity in hard-sphere and attractive glass formers. J. Phys. Chem. B **109**, 14654–14658 (2005)
47. Kob, W., Donati, C., Plimpton, S.J., Poole, P.H., Glotzer, S.C.: Dynamical heterogeneities in a supercooled Lennard-Jones liquid. Phys. Rev. Lett. **79**(15), 2827–2830 (1997)

48. Zubeir, L.F., Rocha, M.A.A., Vergadou, N., Weggemans, W.M.A., Peristeras, L.D., Schulz, P.S., Economou, I.G., Kroon, M.C.: Thermophysical properties of imidazolium tricyanomethanide ionic liquids: experiments and molecular simulation. Phys. Chem. Chem. Phys. **18**, 23121–23138 (2016)
49. Van Hove, L.: Correlations in space and time and born approximation scattering in systems of interacting particles. Phys. Rev. **95**(1), 249–262 (1954)
50. Berne, B.J., Harp, G.D.: On the calculation of time correlation functions. In: Prigogine, I., Rice, S.A. (eds.) Advances in Chemical Physics, vol. XVII, pp. 63–227. Wiley, Honoken (1970)
51. Gaub, M., Fritzsche, S., Haberlandt, R., Theodorou, D.N.: Van Hove function for diffusion in zeolites. J. Phys. Chem. B **103**(22), 4721–4729 (1999)
52. Vineyard, G.H.: Scattering of slow neutrons by a liquid. Phys. Rev. **110**(5), 999–1010 (1958)
53. Urahata, S.M., Ribeiro, M.C.: Single particle dynamics in ionic liquids of 1-alkyl-3-methylimidazolium cations. J. Chem. Phys. **122**(2), 024511 (2005)
54. Einstein, A.: Investigations on the Theory of the Brownian Movement. Dover, New York (1956)
55. Papavassiliou, G., Fardis, M.: Nuclear Magnetic Resonance Laboratory, Institute of Nanoscience and Nanotechnology, NCSR "Demokritos", Greece - Unpublished results
56. Romanos, G.E., Zubeir, L.F., Likodimos, V., Falaras, P., Kroon, M.C., Iliev, B., Adamova, G., Schubert, T.J.S.: Enhanced CO2 capture in binary mixtures of 1-alkyl-3-methylimidazolium tricyanomethanide ionic liquids with water. J. Phys. Chem. B **117**, 12234–12251 (2013)
57. Bara, J.E.: Ionic liquids in gas separation membranes. In: Hoek, E.M.V., Tarabara, V.V. (eds.) Encyclopedia of Membrane Science and Technology, pp. 1–23. Wiley, Hoboken (2013)
58. Kritikos, G., Vergadou, N., Economou, I.G.: Molecular dynamics simulation of highly confined glassy ionic liquids. J. Phys. Chem. C **120**(2), 1013–1024 (2016)
59. Sha, M., Dou, Q., Wu, G.: Molecular dynamics simulation of ionic liquids adsorbed onto a solid surface and confined in nanospace. In: Springborg, M. (ed.) Chemical Modelling: Applications and Theory, vol. 9, pp. 186–217. The Royal Society of Chemistry, Cambridge (2012)
60. Chatel, G., Pereira, J.F.B., Debbeti, V., Wang, H., Rogers, R.D.: Mixing ionic liquids - "simple mixtures" or "double salts"? Green Chem. **16**(4), 2051–2083 (2014)
61. Niedermeyer, H., Hallett, J.P., Villar-Garcia, I.J., Hunt, P.A., Welton, T.: Mixtures of ionic liquids. Chem. Soc. Rev. **41**(23), 7780–7802 (2012)

Chapter 15
Topological Surgery in Nature

Stathis Antoniou and Sofia Lambropoulou

Abstract In this paper, we extend the formal definition of topological surgery by introducing new notions in order to model natural phenomena exhibiting it. On the one hand, the common features of the presented natural processes are captured by our schematic models and, on the other hand, our new definitions provide the theoretical setting for examining the topological changes involved in these processes.

Keywords Topological surgery · Gluing homeomorphism · Three-space
Three-sphere · Solid topological surgery · Embedded topological surgery
Mathematical modeling · Natural processes · Joining 'thread'
Topological 'drilling' · Continuous process · Dynamics · Attracting forces
Repelling forces · Recoupling · Reconnection · DNA recombination
Chromosomal crossover · Magnetic reconnection · Cosmic magnetic lines
Falaco solitons · Tornadoes · Whirls · Waterspouts · Mitosis · Meiosis
Gene transfer · Necking · Black holes

2010 Mathematics Subject Classification: 57R65 · 57N12 · 57M25 · 57M99 · 37B99 · 92B99

Introduction

Topological surgery is more than a mathematical definition allowing the creation of new manifolds out of known ones. As we point out in [2], it is a process triggered by forces that appear in both micro and macro scales. For example, in dimension 1 topological surgery can be seen in DNA recombination and during the reconnection of

S. Antoniou · S. Lambropoulou (✉)
Department of Applied Mathematics, National Technical University of Athens,
Zografou Campus, 15780 Athens, Greece
e-mail: sofia@math.ntua.gr

S. Antoniou
e-mail: santoniou@math.ntua.gr

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_15

cosmic magnetic lines, while in dimension 2 it happens when genes are transferred in bacteria and during the formation of black holes. Inspired by these natural processes, in [2, 6] we enhance the formal definition of topological surgery with the observed *forces* and *dynamics*, thus turning it into a continuous process.

Furthermore, in [2] we observe that phenomena like tension on soap films or the merging of oil slicks are undergoing 1-dimensional surgery but they happen on surfaces instead of 1-manifolds. Similarly, moving up one dimension, during the biological process of mitosis and during tornado formation, 2-dimensional surgery is taking place on 3-dimensional manifolds instead of surfaces. Thus, in order to fit natural phenomena where the interior of the initial manifold is filled in, in [2, 6] we extend the formal definition by introducing the notion of *solid topological surgery* in both dimensions 1 and 2.

Finally, in [2] we notice that in some phenomena exhibiting topological surgery, the ambient space is also involved. For example in dimension 1, during DNA recombination the initial DNA molecule which is recombined can also be knotted. In other words, the initial 1-manifold can be a knot (an embedding of the circle) instead of an abstract circle. Similarly in dimension 2, the processes of tornado and black hole formation are not confined to the initial manifold and topological surgery is causing (or is caused by) a change in the whole space. We therefore define the notion of *embedded topological surgery* which allows us to model these kind of phenomena but also to view all natural phenomena exhibiting topological surgery as happening in 3-space instead of abstractly.

The notions and ideas presented in this paper can be found in [2, 6]. However, [2, 6] contain a much deeper analysis of the dynamical system modeling 2-dimensional surgery, which is discussed here very briefly.

The paper is organized as follows: it starts by recalling the formal definitions of surgery in Sect. 15.1. Dynamic topological surgery is then presented in natural processes and introduced as a schematic model in Sect. 15.2. Next, solid surgery is defined in Sect. 15.3 where the dynamical system modeling 2-dimensional surgery is also presented. Finally, embedded surgery and its differences in dimensions 1 and 2 are discussed in Sect. 15.4.

We hope that the presented definitions and phenomena will broaden our understanding of both the topological changes exhibited in nature and topological surgery itself.

15.1 The Formal Definitions of Surgery

We recall the following well-known definition:

Definition 15.1 An *m-dimensional n-surgery* is the topological procedure of creating a new m-manifold M' out of a given m-manifold M by removing a framed n-embedding $h : S^n \times D^{m-n} \hookrightarrow M$ and replacing it with $D^{n+1} \times S^{m-n-1}$, using the 'gluing' homeomorphism h along the common boundary $S^n \times S^{m-n-1}$. Namely, and

denoting surgery by χ:

$$M' = \chi(M) = \overline{M \setminus h(S^n \times D^{m-n})} \cup_{h|_{S^n \times S^{m-n-1}}} (D^{n+1} \times S^{m-n-1}).$$

Further, the *dual m-dimensional* $(m-n-1)$*-surgery* on M' removes a dual framed $(m-n-1)$-embedding $g : D^{n+1} \times S^{m-n-1} \hookrightarrow M'$ such that $g|_{S^n \times S^{m-n-1}} = h^{-1}|_{S^n \times S^{m-n-1}}$, and replaces it with $S^n \times D^{m-n}$, using the 'gluing' homeomorphism g (or h^{-1}) along the common boundary $S^n \times S^{m-n-1}$. That is:

$$M = \chi^{-1}(M') = \overline{M' \setminus g(D^{n+1} \times S^{m-n-1})} \cup_{h^{-1}|_{S^n \times S^{m-n-1}}} (S^n \times D^{m-n}).$$

From the above definition it follows that $M = \chi^{-1}(\chi(M))$ and $n+1 \leq m$. For further reading see [9–11]. We shall now apply the above definition to dimensions 1 and 2.

15.1.1 1-Dimensional 0-Surgery

We only have one kind of surgery on a 1-manifold M, the *1-dimensional 0-surgery*:

$$M' = \chi(M) = \overline{M \setminus h(S^0 \times D^1)} \cup_{h|_{S^0 \times S^0}} (D^1 \times S^0).$$

The above definition means that two segments $S^0 \times D^1$ are removed from M and they are replaced by two different segments $D^1 \times S^0$ by reconnecting the four boundary points $S^0 \times S^0$ in a different way. In Figs. 15.1a and 15.2a, $S^0 \times S^0 = \{1, 2, 3, 4\}$. As one possibility, if we start with $M = S^1$ and use as h the standard (identity) embedding denoted with h_s, we obtain two circles $S^1 \times S^0$. Namely, denoting by 1 the identity homeomorphism, we have $h_s : S^0 \times D^1 = D^1 \amalg D^1 \xrightarrow{1 \amalg 1} S^0 \times D^1 \hookrightarrow M$, see Fig. 15.1a. However, we can also obtain one circle S^1 if h is an embedding h_t that reverses the orientation of one of the two arcs of $S^0 \times D^1$. Then in the substitution, joining endpoints 1 to 3 and 2 to 4, the two new arcs undergo a half-twist, see Fig. 15.2a. More specifically, if we take $D^1 = [-1, +1]$ and define the homeomorphism $\omega : D^1 \to D^1; t \to -t$, the embedding used in Fig. 15.2a is $h_t : S^0 \times D^1 = D^1 \amalg D^1 \xrightarrow{1 \amalg \omega} S^0 \times D^1 \hookrightarrow M$ which rotates one D^1 by 180°. The difference between the embeddings h_s and h_t of $S^0 \times D^1$ can be clearly seen by comparing the four boundary points 1, 2, 3 and 4 in Figs. 15.1a and 15.2a.

Note that in dimension one, the dual case is also an 1-dimensional 0-surgery. For example, looking at the reverse process of Fig. 15.1a, we start with two circles $M' = S^1 \amalg S^1$ and, if each segment of $D^1 \times S^0$ is embedded in a different circle, the result of the (dual) 1-dimensional 0-surgery is one circle: $\chi^{-1}(M') = M = S^1$.

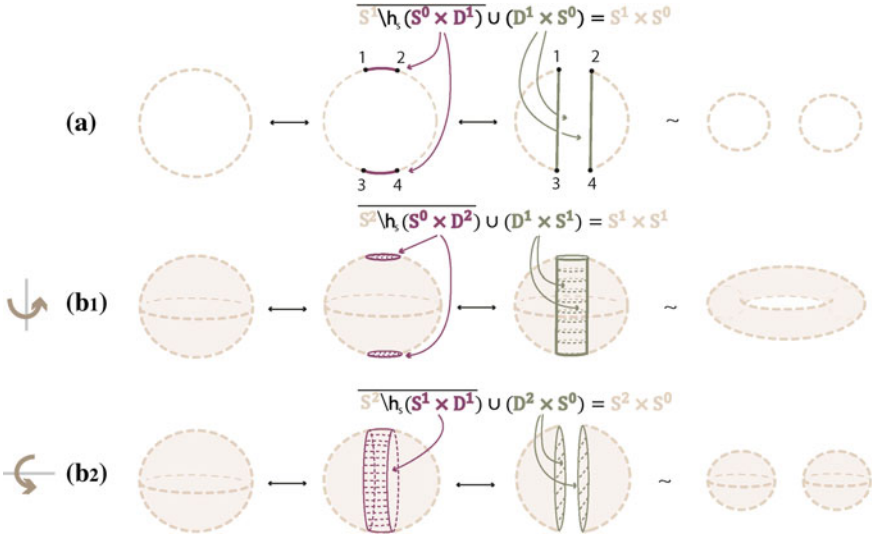


Fig. 15.1 Formal **(a)** 1-dimensional 0-surgery **(b_1)** 2-dimensional 0-surgery and **(b_2)** 2-dimensional 1-surgery using the standard embedding h_s

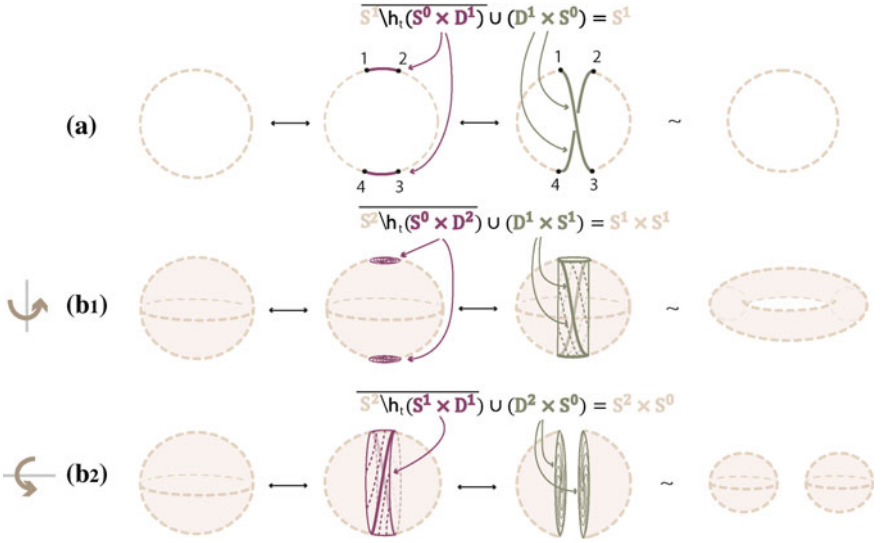


Fig. 15.2 Formal **(a)** 1-dimensional 0-surgery **(b_1)** 2-dimensional 0-surgery and **(b_2)** 2-dimensional 1-surgery using a twisting embedding h_t

15.1.2 2-Dimensional 0-Surgery

Starting with a 2-manifold M, there are two types of surgery. One type is the *2-dimensional 0-surgery*, whereby two discs $S^0 \times D^2$ are removed from M and are replaced in the closure of the remaining manifold by a cylinder $D^1 \times S^1$, which gets attached via a homeomorphism along the common boundary $S^0 \times S^1$ comprising two copies of S^1. The gluing homeomorphism of the common boundary may twist one or both copies of S^1. For $M = S^2$ the above operation changes its homeomorphism type from the 2-sphere to that of the torus. View Fig. 15.1b_1 for the standard embedding h_s and Fig. 15.2b_1 for a twisting embedding h_t. For example, the homeomorphism $\mu : D^2 \to D^2$; $(t_1, t_2) \to (-t_1, -t_2)$ induces the 2-dimensional analogue h_t of the embedding defined in the previous example, namely: $h_t : S^0 \times D^2 = D^2 \amalg D^2 \xrightarrow{1 \amalg \mu} S^0 \times D^2 \hookrightarrow M$ which rotates one D^2 by 180°. When, now, the cylinder $D^1 \times S^1$ is glued along the common boundary $S^0 \times S^1$, the twisting of this boundary induces the twisting of the cylinder, see Fig. 15.2b_1.

15.1.3 2-Dimensional 1-Surgery

The other possibility of 2-dimensional surgery on M is the *2-dimensional 1-surgery*: here a cylinder (or annulus) $S^1 \times D^1$ is removed from M and is replaced in the closure of the remaining manifold by two discs $D^2 \times S^0$ attached along the common boundary $S^1 \times S^0$. For $M = S^2$ the result is two copies of S^2, see Fig. 15.1b_2 for the standard embedding h_s. Unlike Fig. 15.1b_1 where the cylinder is illustrated vertically, in Fig. 15.1b_2, the cylinder is illustrated horizontally. This choice was made so that the instances of 1-dimensional surgery can be obtained by crossections of the instances of both types of 2-dimensional surgeries, see further Remark 15.1. Figure 15.2b_2 illustrates a twisting embedding h_t, where a twisted cylinder is being removed. In that case, taking $D^1 = \{h : h \in [-1, 1]\}$ and homeomorphism ζ:

$$\zeta : S^1 \times D^1 \to S^1 \times D^1;$$
$$\zeta : (t_1, t_2, h) \to \left(t_1 \cos \frac{(1-h)\pi}{2} - t_2 \sin \frac{(1-h)\pi}{2}, t_1 \sin \frac{(1-h)\pi}{2} + t_2 \cos \frac{(1-h)\pi}{2}, h \right)$$

the embedding h_t is defined as: $h_t : S^1 \times D^1 \xrightarrow{\zeta} S^1 \times D^1 \hookrightarrow M$. This operation corresponds to fixing the circle S^1 bounding the right side of the cylinder $S^1 \times D^1$, rotating the circle S^1 bounding the left side of the cylinder by 180° and letting the rotation propagate from left to right. This twisting of the cylinder can be seen by comparing the second instance of Fig. 15.1b_2 with the second instance of Fig. 15.2b_2, but also by comparing the third instance of Fig. 15.1b_1 with the third instance of Fig. 15.2b_1.

It follows from Definition 15.1 that a dual 2-dimensional 0-surgery is a 2-dimensional 1-surgery and vice versa. Hence, Fig. 15.1b$_1$ shows that a 2-dimensional 0-surgery on a sphere is the reverse process of a 2-dimensional 1-surgery on a torus. Similarly, as illustrated in Fig. 15.1b$_2$, a 2-dimensional 1-surgery on a sphere is the reverse process of a 2-dimensional 0-surgery on two spheres. In the figure the symbol $\longleftrightarrow$ depicts surgeries from left to right and their corresponding dual surgeries from right to left.

Remark 15.1 The stages of the process of 2-dimensional 0-surgery on S^2 can be obtained by rotating the stages of 1-dimensional 0-surgeries on S^1 by 180° around a vertical axis, see Fig. 15.1b$_1$. Similarly, the stages of 2-dimensional 1-surgery on S^2 can be obtained by rotating the stages of 1-dimensional 0-surgeries on S^1 by 180° around a horizontal axis, see Fig. 15.1b$_2$. It follows from the above that 1-dimensional 0-surgery can be obtained as a cross-section of either type of 2-dimensional surgery.

15.2 Dynamic Topological Surgery in Natural Processes

In this section we present natural processes exhibiting topological surgery in dimensions 1 and 2 and we incorporate the observed dynamics to a schematic model showing the intermediate steps that are missing from the formal definition. This model extends surgery to a continuous process caused by local forces. Note that these intermediate steps can also be explained through Morse theory, see Remark 15.2 for details.

15.2.1 Dynamic 1-Dimensional Topological Surgery

We find that 1-dimensional 0-surgery is present in phenomena where 1-dimensional splicing and reconnection occurs. It can be seen for example during meiosis when new combinations of genes are produced, see Fig. 15.3c, and during magnetic reconnection, the phenomena whereby cosmic magnetic field lines from different magnetic domains are spliced to one another, changing their pattern of conductivity with respect to the sources, see Fig. 15.3b from [3]. It is worth mentioning that 1-dimensional 0-surgery is also present during the reconnection of vortex tubes in a viscous fluid and quantized vortex tubes in superfluid helium. As mentioned in [5], these cases have some common qualitative features with the magnetic reconnection shown in Fig. 15.3b.

In fact, all the above phenomena have similar dynamics. Namely, 1-dimensional 0-surgery is a continuous process for all of them. Furthermore, as detailed in [2], in most cases, surgery is the result of local forces. These common features are captured by our model in Fig. 15.3a which describes the process of dynamic 1-dimensional 0-surgery locally. The process starts with the two points (in red) specified on any 1-dimensional

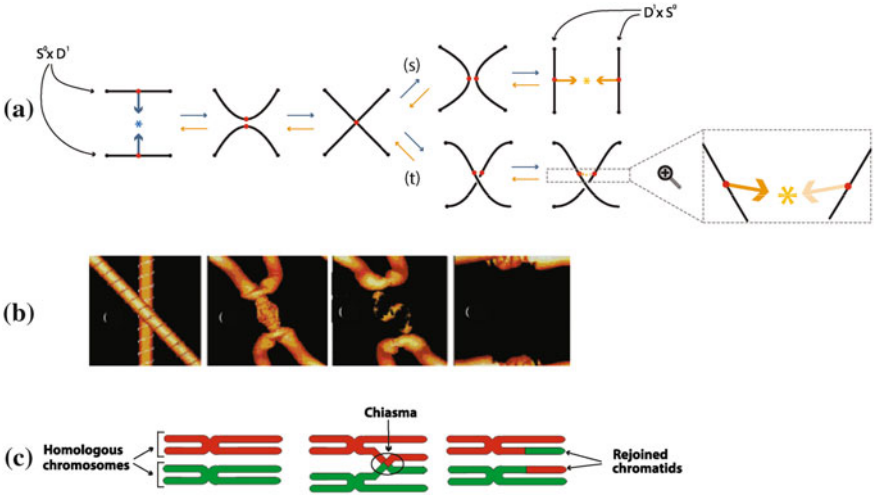


Fig. 15.3 **a** Dynamic 1-dimensional surgery locally **b** The reconnection of cosmic magnetic lines **c** Crossing over of chromosomes during meiosis

manifold, on which attracting forces are applied (in blue). We assume that these forces are caused by an attracting center (also in blue). Then, the two segments $S^0 \times D^1$, which are neighborhoods of the two points, get close to one another. When the specified points (or centers) of the two segments reach the attracting center they touch and recoupling takes place, giving rise to the two final segments $D^1 \times S^0$, which split apart. In Fig. 15.3a, case (s) corresponds to the identity embedding h_s described in Sect. 15.1.1, while (t) corresponds to the twisting embedding h_t described in Sect. 15.1.1.

As also mentioned in Sect. 15.1.1, the dual case is also a 1-dimensional 0-surgery, as it removes segments $D^1 \times S^0$ and replaces them by segments $S^0 \times D^1$. This is the reverse process which starts from the end and is illustrated in Fig. 15.3a as a result of the orange forces and attracting center which are applied on the 'complementary' points.

Finally, for details about the described dynamics and attracting forces in the aforementioned phenomena, the reader is referred to the analysis done in [2].

Remark 15.2 It is worth mentioning that the intermediate steps of surgery presented in Fig. 15.3a can also be viewed in the context of Morse theory [7]. By using the local form of a Morse function, we can visualize the process of surgery by varying parameter t of equation $x^2 - y^2 = t$. For $t = -1$ it is the hyperbola shown in the second stage of Fig. 15.3a where the two segments get close to one another. For $t = 0$ it is the two straight lines where the reconnection takes place as shown in the third stage of Fig. 15.3a while for $t = 1$ it represents the hyperbola of the two final segments shown in case (s) of the fourth stage of Fig. 15.3a. This sequence can be generalized for higher dimensional surgeries as well, however, in this paper we will not use this approach as we are focusing on the introduction of forces and of the attracting center.

15.2.2 Dynamic 2-Dimensional Topological Surgery

We find that 2-dimensional surgery is present in phenomena where 2-dimensional merging and recoupling occurs. For example, *2-dimensional 0-surgery* can be seen during the formation of tornadoes, see Fig. 15.4b (this phenomenon will be detailed in Sect. 15.4.2.2). Further, it can be seen in the formation of Falaco solitons, see Fig. 15.4c (note that each Falaco soliton consists of a pair of locally unstable but globally stabilized contra-rotating identations in the water-air discontinuity surface of a swimming pool, see [4] for details). It can also be seen in gene transfer in bacteria where the donor cell produces a connecting tube called a 'pilus' which attaches to the recipient cell, see Fig. 15.4d; in drop coalescence, the phenomenon where two dispersed drops merge into one, and in the formation of black holes (this phenomena will be discussed in Sect. 15.4.2.1), see Fig. 15.8(ii) (Source:www.black-holes.org).

On the other hand, *2-dimensional 1-surgery* can be seen during soap bubble splitting, where a soap bubble splits into two smaller bubbles, see Fig. 15.4e (Source: soapbubble.dk); when the tension applied on metal specimens by tensile forces results in the phenomena of necking and then fracture, see Fig. 15.4f; also in the biological process of mitosis, where a cell splits into two new cells, see Fig. 15.4g.

Phenomena exhibiting 2-dimensional 0-surgery are the results of *two colinear attracting forces which 'create' a cylinder*. These phenomena have similar dynamics and are characterized by their continuity and the attracting forces causing them, see [2] for details. These common features are captured by our model in Fig. 15.4a

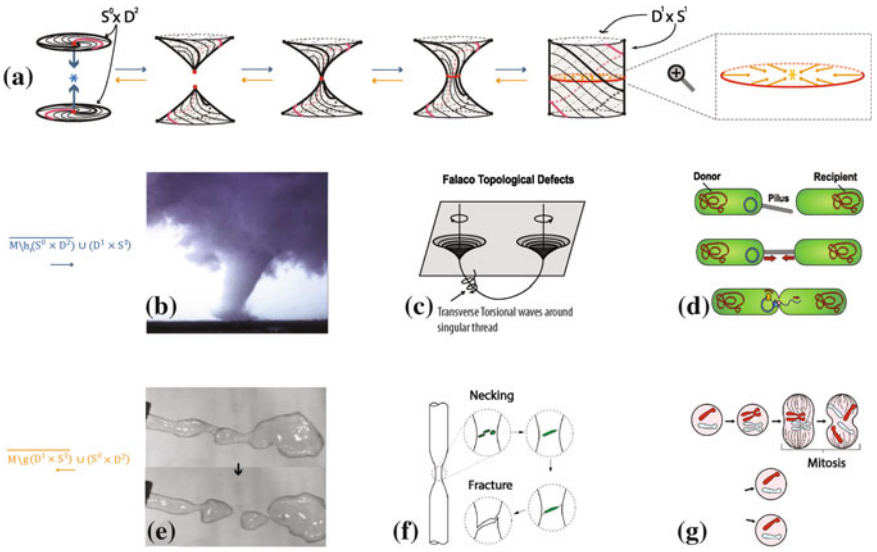


Fig. 15.4 **a** Dynamic 2-dimensional surgery locally **b** Tornadoes **c** Falaco solitons **d** Gene transfer in bacteria **e** Soap bubble splitting **f** Fracture **g** Mitosis

which describes both cases of dynamic 2-dimensional surgeries locally. Note that we have used non-trivial embeddings (recall Sect. 15.1.2) which are more appropriate for natural processes involving twisting, such as tornadoes and Falaco solitons. In Fig. 15.4a, both initial discs have been rotated in opposite directions by an angle of $3\pi/4$ and we can see how this rotation induces the twisting of angle $3\pi/2$ of the final cylinder.

The process of *dynamic 2-dimensional 0-surgery* starts with two points, or poles, specified on the manifold (in red) on which attracting forces caused by an attracting center are applied (in blue). Then, the two discs $S^0 \times D^2$, neighbourhoods of the two poles, approach each other. When the centers of the two discs touch, recoupling takes place and the discs get transformed into the final cylinder $D^1 \times S^1$, see Fig. 15.4a. The cylinder created during 2-dimensional 0-surgery can take various forms. For example, it is a tubular vortex of air in the case of tornadoes, a transverse torsional wave in the case of Falaco solitons and a pilus joining the genes in gene transfer in bacteria.

On the other hand, phenomena exhibiting 2-dimensional 1-surgery are the result of *an infinitum of coplanar attracting forces which 'collapse' a cylinder*, see Fig. 15.4a from the end. As mentioned in Sect. 15.1.3, the dual case of 2-dimensional 0-surgery is the 2-dimensional 1-surgery and vice versa. This is illustrated in Fig. 15.4a where the reverse process is the *2-dimensional 1-surgery* which starts with the cylinder and a specified circular region (in red) on which attracting forces caused by an attracting center are applied (in orange). A 'necking' occurs in the middle which degenerates into a point and finally tears apart creating two discs $S^0 \times D^2$. This cylinder can be embedded, for example, in the region of the bubble's surface where splitting occurs, on the region of metal specimens where necking and fracture occurs or on the equator of the cell which is about to undergo a mitotic process.

Remark 15.3 From Definition 15.1 we know that surgery always starts by removing a thickened sphere $S^n \times D^{m-n}$ from the initial manifold. As seen in Fig. 15.4a, in the case of 2-dimensional 0-surgery, forces (in blue) are applied on two points, or S^0, whose thickening comprises the two discs, while in the case of the 2-dimensional 1-surgery, forces (in orange) are applied on a circle S^1, whose thickening is the cylinder. In other words the forces that model a 2-dimensional n-surgery are always applied to the core n-embedding $e = h_| : S^n = S^n \times \{0\} \hookrightarrow M$ of the framed n-embedding $h : S^n \times D^{2-n} \hookrightarrow M$. Also, note that Remark 15.1 is also true here. One can obtain Fig. 15.4a by rotating Fig. 15.3a and this extends also to the dynamics and forces. For instance, by rotating the two points, or S^0, on which the pair of forces of 1-dimensional 0-surgery acts (shown in red in the last instance of Fig. 15.3a) by 180° around a vertical axis we get the circle, or S^1, on which the infinitum of coplanar attracting forces of 2-dimensional 1-surgery acts (shown in red in the last instance of Fig. 15.4a).

Finally, it is worth pointing out that these local dynamics produce different manifolds depending on the initial manifold where they act. For example, 2-dimensional 0-surgery transforms an $S^0 \times S^2$ to an S^2 by adding a cylinder during gene transfer

in bacteria (see Fig. 15.4d) but can also transform an S^2 to a torus by 'drilling out' a cylinder during the formation of Falaco solitons (see Fig. 15.4c) in which case S^2 is the pool of water and the cylinder is the boundary of the tubular neighborhood around the thread joining the two poles.

15.3 Defining Solid Topological Surgery

Looking closer at the phenomena exhibiting 2-dimensional surgery shown in Fig. 15.4, one can see that, with the exception of soap bubble splitting that involves surfaces, all others involve 3-dimensional manifolds. For instance, what really happens during a mitotic process is that a solid cylindrical region located in the center of the cell collapses and a D^3 is transformed into an $S^0 \times D^3$. Similarly, during tornado formation, the created cylinder is not just a cylindrical surface $D^1 \times S^1$ but a solid cylinder $D^2 \times S^1$ containing many layers of air (this phenomena will be detailed in Sect. 15.4.2.2). Of course we can say that, for phenomena involving 3-dimensional manifolds, the outer layer of the initial manifold is undergoing 2-dimensional surgery. In this section we will define topologically what happens to the whole manifold.

The need of such a definition is also present in dimension 1 for modeling phenomena such as the merging of oil slicks and tension on membranes (or soap films). These phenomena undergo the process of 1-dimensional 0-surgery but happen on surfaces instead of 1-manifolds.

We will now introduce the notion of solid surgery (in both dimensions 1 and 2) where the interior of the initial manifold is filled in. There is one key difference compared to the dynamic surgeries discussed in the previous section. While the local dynamics described in Figs. 15.3 and 15.4 can be embedded in any manifold, here we also have to fix the initial manifold in order to define solid surgery. For example, as we will see next, we define separately the processes of *solid 1-dimensional 0-surgery on* D^2 and *solid 1-dimensional 0-surgery on* $D^2 \times S^0$. However, the underlying features are common in both.

15.3.1 Solid 1-Dimensional Topological Surgery

The process of solid 1-dimensional 0-surgery on D^2 is equivalent to performing 1-dimensional 0-surgeries on the whole continuum of concentric circles included in D^2, see Fig. 15.5a. More precisely, and introducing at the same time dynamics, we define:

Definition 15.2 *Solid 1-dimensional 0-surgery on* D^2 is the following process. We start with the 2-disc of radius 1 with polar layering:

$$D^2 = \cup_{0<r\leq 1} S^1_r \cup \{P\},$$

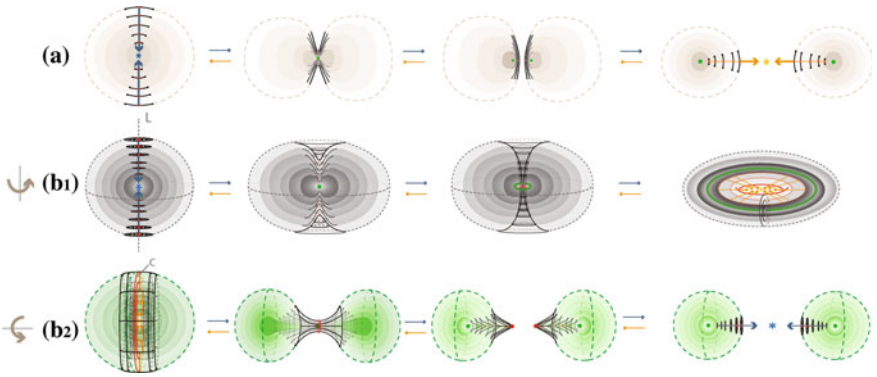


Fig. 15.5 **(a)** Solid 1-dimensional 0-surgery on D^2 **($\mathbf{b_1}$)** Solid 2-dimensional 0-surgery on on D^3 **($\mathbf{b_2}$)** Solid 2-dimensional 1-surgery on D^3

where r the radius of a circle and P the limit point of the circles, which is the center of the disc and also the circle of radius zero. We specify colinear pairs of antipodal points, all on the same diameter, with neighbourhoods of analogous lengths, on which the same colinear attracting forces act. See Fig. 15.5a where these forces and the attracting center are shown in blue. Then the antipodal segments get closer to one another or, equivalently, closer to the attracting center. Note that here, the attracting center coincides with the limit point of all concentric circles, which is shown in green from the second instance and on. Then, we perform 1-dimensional 0-surgery on the whole continuum of concentric circles. We define 1-dimensional 0-surgery on the limit point P to be the two limit points of the resulting surgeries. That is, the effect of *solid 1-dimensional 0-surgery on a point is the creation of two new points*. Next, the segments reconnect until we have two copies of D^2.

The above process is the same as first removing the center P from D^2, doing the 1-dimensional 0-surgeries and then taking the closure of the resulting space. The resulting manifold is

$$\chi(D^2) := \cup_{0<r\leq 1}\chi(S^1_r) \cup \chi(P),$$

which comprises two copies of D^2.

We also have the reverse process of the above, namely, *solid 1-dimensional 0-surgery on two discs* $D^2 \times S^0$. This process is the result of the orange forces and attracting center which are applied on the 'complementary' points, see Fig. 15.5a in reverse order. This operation is equivalent to performing 1-dimensional 0-surgery on the whole continuum of pairs of concentric circles in $D^2 \sqcup D^2$. We only need to define solid 1-dimensional 0-surgery on two limit points to be the limit point P of the resulting surgeries. That is, the effect of *solid 1-dimensional 0-surgery on two points is their merging into one point*. The above process is the same as first removing the centers from the $D^2 \times S^0$, doing the 1-dimensional 0-surgeries and then taking the closure of the resulting space. The resulting manifold is

$$\chi^{-1}(D^2 \times S^0) := \cup_{0<r\leq 1} \chi^{-1}(S^1_r \times S^0) \cup \chi^{-1}(P \times S^0),$$

which comprises one copy of D^2.

15.3.2 Solid 2-Dimensional Topological Surgery

Moving up one dimension, there are two types of solid 2-dimensional surgery on the 3-ball, D^3, analogous to the two types of 2-dimensional surgery. More precisely we have:

Definition 15.3 We start with the 3-ball of radius 1 with polar layering:

$$D^3 = \cup_{0<r\leq 1} S^2_r \cup \{P\},$$

where r the radius of the 2-sphere S^2_r and P the limit point of the spheres, that is, their common center and the center of the ball. *Solid 2-dimensional 0-surgery on* D^3 is the topological procedure whereby 2-dimensional 0-surgery takes place on each spherical layer that D^3 is made of. More precisely, as illustrated in Fig. 15.5b_1, on all spheres S^2_r colinear pairs of antipodal points are specified, all on the same diameter, on which the same colinear attracting forces act. The poles have disc neighborhoods of analogous areas. Then, 2-dimensional 0-surgeries are performed on the whole continuum of the concentric spheres using the same embedding h (recall Sect. 15.1.2). Moreover, 2-dimensional 0-surgery on the limit point P is defined to be the limit circle of the nested tori resulting from the continuum of 2-dimensional 0-surgeries. That is, the effect of *2-dimensional 0-surgery on a point is defined to be the creation of a circle*.

The process is characterized on one hand by the 1-dimensional core L of the solid cylinder which joins the two selected antipodal points of the outer shell and intersects each spherical layer at its two corresponding antipodal points, and on the other hand by the embedding h. The process results in a continuum of layered tori and can be viewed as drilling out a tunnel along L according to h. Note that in Fig. 15.5, the identity embedding has been used. However, a twisting embedding, which is the case shown in Fig. 15.4a, agrees with our intuition that, for opening a hole, *drilling with twisting* seems to be the easiest way. Examples of these two embeddings can be found in Sect. 15.1.2.

Furthermore, *solid 2-dimensional 1-surgery on* D^3 is the topological procedure where on all spheres S^2_r nested cylindrical peels of the solid cylinder of analogous areas are specified and the same coplanar attracting forces act on all spheres, see Fig. 15.5b_2. Then, 2-dimensional 1-surgeries are performed on the whole continuum of the concentric spheres using the same embedding h. Moreover, 2-dimensional 1-surgery on the limit point P is defined to be the two limit points of the nested pairs of 2-spheres resulting from the continuum of 2-dimensional surgeries. That is, the effect of *2-dimensional 1-surgery on a point is the creation of two new points*. The

process is characterized by the 2-dimensional central disc of the solid cylinder and the embedding h, and it can be viewed as squeezing the central disc C or, equivalently, as pulling apart the left and right hemispheres with possible twists, if h is a twisting embedding. This agrees with our intuition that for cutting a solid object apart, *pulling with twisting* seems to be the easiest way. Examples of the identity and the twisting embedding can be found in Sect. 15.1.3.

For both types of solid 2-dimensional surgery, the above process is the same as: first removing the center P from D^3, performing the 2-dimensional surgeries and then taking the closure of the resulting space. Namely we obtain:

$$\chi(D^3) := \cup_{0<r\leq 1}\chi(S^2_r) \cup \chi(P),$$

which is a solid torus in the case of solid 2-dimensional 0-surgery and two copies of D^3 in the case of solid 2-dimensional 1-surgery.

As seen in Fig. 15.5, we also have the two dual solid 2-dimensional surgeries, which represent the reverse processes. As already mentioned in Sect. 15.1.3, the dual case of 2-dimensional 0-surgery is the 2-dimensional 1-surgery and vice versa. More precisely:

Definition 15.4 The *dual case of solid 2-dimensional 0-surgery on* D^3 is the solid 2-dimensional 1-surgery on a solid torus $D^2 \times S^1$. This is the reverse process shown in Fig. 15.5b_1 which results from the orange forces and attracting center. Given that the solid torus can be written as a union of nested tori together with the core circle: $D^2 \times S^1 = (\cup_{0<r\leq 1} S^1_r \cup \{0\}) \times S^1$, solid 2-dimensional 1-surgeries are performed on each toroidal layer starting from specified annular peels of analogous sizes where the same coplanar forces act on the central rings of the annuli. These forces are caused by the same attracting center lying outside the torus. It only remain to define the solid 2-dimensional 1-surgery on the limit circle to be the limit point P of the resulting surgeries. That is, the effect of *solid 2-dimensional 1-surgery on the core circle is that it collapses into one point, the attracting center*. The above process is the same as first removing the core circle from $D^2 \times S^1$, doing the 2-dimensional 1-surgeries on the layered tori, with the same coplanar acting forces, and then taking the closure of the resulting space. Hence, the resulting manifold is

$$\chi^{-1}(D^2 \times S^1) := \cup_{0<r\leq 1}\chi^{-1}(S^1_r \times S^1) \cup \chi^{-1}(\{0\} \times S^1),$$

which comprises one copy of D^3.

Further, *the dual case of solid 2-dimensional 1-surgery on* D^3 is the solid 2-dimensional 0-surgery on two 3-balls D^3. This is the reverse process shown in Fig. 15.5b_2 which results from the blue forces and attracting center. We only need to define the solid 2-dimensional 0-surgery on two limit points to be the limit point P of the resulting surgeries. That is, as in solid 1-dimensional surgery (see Fig. 15.5a), the effect of *solid 2-dimensional 0-surgery on two points is their merging into one point*. The above process is the same as first removing the centers from the $D^3 \times S^0$,

doing the 2-dimensional 0-surgeries on the nested spheres with the same colinear forces and then taking the closure of the resulting space. The resulting manifold is

$$\chi^{-1}(D^3 \times S^0) := \cup_{0<r\leq 1}\chi^{-1}(S^2_r \times S^0) \cup \chi^{-1}(P \times S^0),$$

which comprises one copy of D^3.

Note that Remarks 15.1 and 15.3 are also true here. One can obtain Fig. 15.5b$_1$ and Fig. 15.5b$_2$ by rotating Fig. 15.5a respectively by 180° around a vertical axis and by 180° around a horizontal axis.

Remark 15.4 The notions of 2-dimensional (resp. solid 2-dimensional) surgery, can be generalized from S^2 (resp. D^3) to a surface (resp. a handlebody) of genus g creating a surface (resp. a handlebody) of genus $g \pm 1$ or a disconnected surface (resp. handlebody).

15.3.3 A Dynamical System Modeling Solid 2-Dimensional 0-Surgery

In [2, 6] we present and analyze a dynamical system for modeling 2-dimensional 0-surgery. This system was introduced in [12] and is a generalization of the classical Lotka–Volterra system in three dimensions where the classic predator-prey population model is generalized to a two-predator and one-prey system. The parameters of the system affect the dynamics of the populations and they are analyzed in order to determine the bifurcation properties of the system.

In Fig. 15.6 we reproduce the numerical simulations done in [12, 13], illustrating how by changing the parameters space, solid 2-dimensional 0-surgery is performed on the trajectories of the system. We have used the same colors as in Fig. 15.5b$_1$

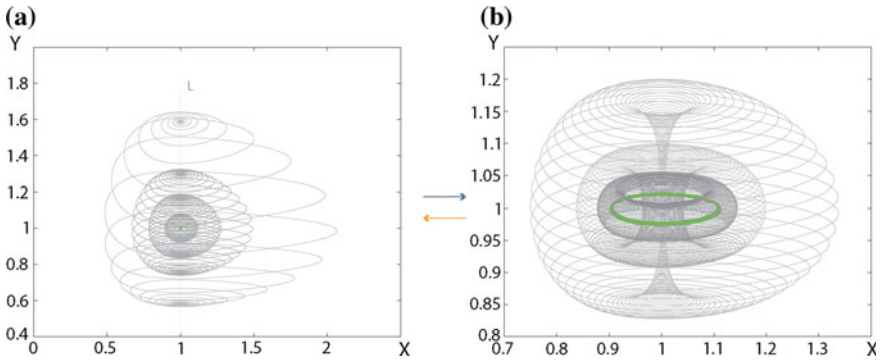


Fig. 15.6 Solid 2-dimensional 0-surgery by changing parameter space from **a**–**b**

to make comparison easier. What is even more striking is that the new topological definition stating that the effect of 2-dimensional 0-surgery on a point is defined to be the creation of a circle (recall Sect. 15.3.2) also has a meaning in the language of dynamical systems. More precisely, the limit point in the spherical nesting of trajectories shown in green in Fig. 15.6a is a steady state point and the core of the toroidal nesting of trajectories shown in green in Fig. 15.6b is a limit cycle. This connection is detailed in [2, 6].

15.4 Defining Embedded Topological Surgery

As mentioned in the Introduction, we noticed that the ambient space is also involved in some natural processes exhibiting surgery. As we will see in this section, depending on the dimension of the manifold, the ambient space either leaves 'room' for the initial manifold to assume a more complicated configuration or it participates more actively in the process. Independently of dimensions, embedding surgery has the advantage that it allows us to view surgery as a process happening inside a space instead of abstractly. We define it as follows:

Definition 15.5 *An embedded m-dimensional n-surgery* is a m-dimensional n-surgery where the initial manifold is an m-embedding $e : M \hookrightarrow S^d$, $d \geq m$ of some m-manifold M, and the result is also viewed as embedded in S^d. Namely, according to Definition 15.1:

$$M' = \chi(e(M)) = \overline{e(M) \setminus h(S^n \times D^{m-n})} \cup_{h|_{S^n \times S^{m-n-1}}} D^{n+1} \times S^{m-n-1} \hookrightarrow S^d.$$

Since in this analysis we focus on phenomena exhibiting embedded 1- and 2-dimensional surgery in 3-space, from now on we fix $d = 3$ and, for our purposes, we consider S^3 or $\mathbb{R}^3$ as our standard 3-space.

15.4.1 Embedded 1-Dimensional Surgery

In dimension 1, the notion of embedded surgery allows the topological modeling of phenomena with more complicated initial 1-manifolds. Let us demonstrate this with the example of site-specific **DNA recombination**. In this process, the initial manifold is a (circular or linear) DNA molecule. With the help of certain enzymes, site-specific recombination performs a 1-dimensional 0-surgery on the molecule, causing possible knotting or linking of the molecule.

The first electron microscope picture of knotted DNA was presented in [14]. In this experimental study, we see how genetically engineered circular DNA molecules can form DNA knots and links through the action of a certain recombination enzyme. A similar picture is presented in Fig. 15.7, where site-specific recombination of a DNA

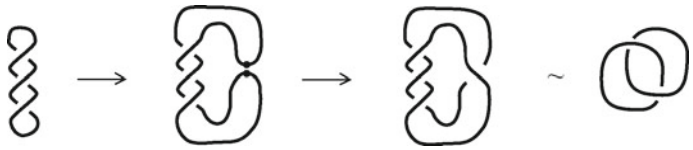


Fig. 15.7 DNA Recombination as an example of embedded 1-dimensional 0-surgery

molecule produces the Hopf link. It is worth mentioning that there are infinitely many knot types and that 1-dimensional 0-surgery on a knot may change the knot type or even result in a two-component link (as shown in Fig. 15.7). Since a knot is by definition an embedding of $M = S^1$ in S^3 or $\mathbb{R}^3$, in this case embedded 1-dimensional surgery is the so-called *knot surgery*. A good introductory book on knot theory is [1] among many others.

We can summarize the above by stating that for $M = S^1$, embedding in S^3 allows the initial manifold to become any type of knot. More generally, in dimension 1 the ambient space which is of codimension 2 gives enough 'room' for the initial 1-manifold to assume a more complicated homeomorphic configuration.

Remark 15.5 Of course we also have, in theory, the notion of embedded solid 1-dimensional 0-surgery whereby the initial manifold is an embedding of a disc in 3-space.

15.4.2 Embedded 2-Dimensional Surgery

Passing now to 2-dimensional surgeries, let us first note that an embedding of a sphere $M = S^2$ in S^3 presents no knotting because knotting requires embeddings of codimension 2. However, in this case the ambient space plays a different role. Namely, embedding 2-dimensional surgeries *allows the complementary space of the initial manifold to participate actively in the process*. Indeed, while some natural phenomena undergoing surgery can be viewed as 'local', in the sense that they can be considered independently from the surrounding space, some others are intrinsically related to the surrounding space. This relation can be both *causal*, in the sense that the ambient space is involved in the triggering of the forces causing surgery, and *consequential*, in the sense that the forces causing surgery, can have an impact on the ambient space in which they take place.

Let us recall here that the ambient space S^3 can be viewed as $\mathbb{R}^3$ with all points at infinity compactified to one single point: $S^3 = \mathbb{R}^3 \cup \{\infty\}$. Further, it can be viewed as the union of two 3-balls along the common boundary: $S^3 = B^3 \cup D^3$ where a neighbourhood of the point at infinity can stand for one of the two 3-balls. Finally, S^3 can be viewed as the union of two solid tori along their common boundary: $S^3 = V_1 \cup V_2$.

As mentioned in the introduction of Sect. 15.3, in most natural phenomena that exhibit 2-dimensional surgery, the initial manifold is a *solid* 3-dimensional object. Hence, in the next subsections, we describe natural phenomena undergoing solid 2-dimensional surgeries which exhibit the causal or consequential relation to the ambient space mentioned above and are therefore better described by considering them as embedded in S^3 or in $\mathbb{R}^3$. In parallel, we describe how these processes are altering the whole space S^3 or $\mathbb{R}^3$.

15.4.2.1 A Topological Model for Black Hole Formation

Let us start by considering the formation of **black holes**. Most black holes are formed from the remnants of a large star that dies in a supernova explosion. Their gravitational field is so strong that not even light can escape. In the simulation of a black hole formation in [8], the density distribution at the core of a collapsing massive star is shown. Figure 15.8(ii) shows three instants of this simulation, which indicate that matter performs solid 2-dimensional 0-surgery as it collapses into a black hole. In fact, matter collapses at the center of attraction of the initial manifold $M = D^3$ creating the singularity, that is, the center of the black hole (shown as a black dot in instance (c) of Fig. 15.8(ii)), which is surrounded by the toroidal accretion disc (shown in white in instance (c) of Fig. 15.8(ii)). Let us be reminded here that an accretion disc is a rotating disc of matter formed by accretion.

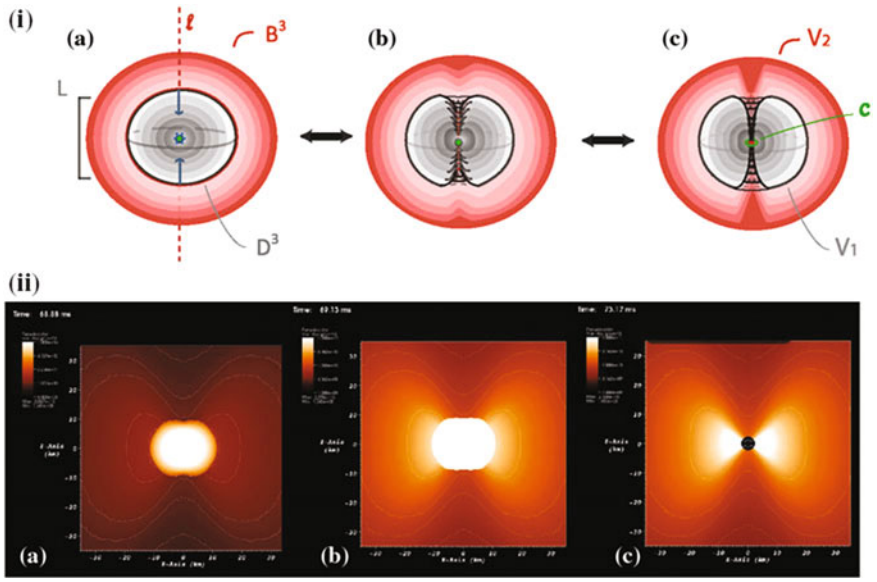


Fig. 15.8 **(i)** Embedded solid 2-dimensional 0-surgery on $M = D^3$ (in $\mathbb{R}^3$) **(ii)** Black hole formation

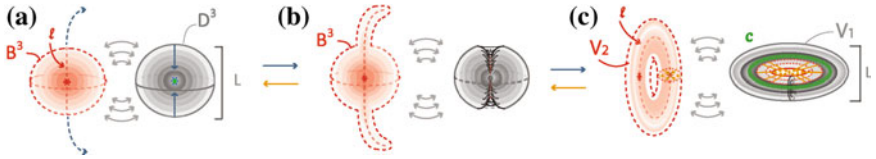


Fig. 15.9 Embedded solid 2-dimensional 0-surgery on $M = D^3$ (in S^3)

Note now that the strong gravitational forces have altered the space surrounding the initial star and that the singularity is created outside the final solid torus. This means that the process of surgery in this phenomenon has moreover altered matter outside the manifold in which it occurs. In other words, *the effect of the forces causing surgery propagates to the complement space, thus causing a more global change in 3-space*. This fact makes black hole formation a phenomenon that topologically undergoes embedded solid 2-dimensional 0-surgery.

In Fig. 15.8(i), we present a schematic model of embedded solid 2-dimensional 0-surgery on $M = D^3$. From the descriptions of S^3 in Sect. 15.4.2, it becomes apparent that embedded solid 2-dimensional 0-surgery on one 3-ball describes the passage from the two-ball description to the two-solid tori description of S^3. This can be seen in $\mathbb{R}^3$ in instances (a) – (c) of Fig. 15.8(i) but is more obvious by looking at instances (a) – (c) of Fig. 15.9 which show the corresponding view in S^3.

We will now detail the instances of the process of embedded solid 2-dimensional 0-surgery on $M = D^3$ by referring to both the view in S^3 and the corresponding decompacified view in $\mathbb{R}^3$. Let $M = D^3$ be the solid ball having arc L as a diameter and the complement space be the other solid ball B^3 containing the point at infinity; see instances (a) of Fig. 15.9 and (a) of Fig. 15.8. Note that, in both cases B^3 represents the hole space outside D^3 which means that the spherical nesting of B^3 in instance (a) of Fig. 15.8 extends to infinity, even though only a subset of B^3 is shown. This joining arc L is seen as part of a simple closed curve l passing by the point at infinity. In instances (b) of Fig. 15.9 and (b) of Fig. 15.8, we see the 'drilling' along L as a result of the attracting forces. This is exactly the same process as in Fig. 15.5b_1 if we restrict it to D^3. But since we have embedded the process in S^3 or $\mathbb{R}^3$, the complement space B^3 participates in the process and, in fact, it is also undergoing solid 2-dimensional 0-surgery. Indeed, the 'matter' that is being drilled out from the interior of D^3 can be viewed as 'matter' of the outer sphere B^3 invading D^3. In instances (c) of Fig. 15.9 and (c) of Fig. 15.8, we can see that, as surgery transforms the solid ball D^3 into the solid torus V_1, B^3 is transformed into V_2. That is, the nesting of concentric spheres of D^3 (respectively of B^3) is transformed into the nesting of concentric tori in the interior of V_1 (respectively of V_2). The point at the origin (in green), which is also the attracting center, turns into the core curve c of V_1 (in green) which, by Definition 15.3 is 2-dimensional 0-surgery on a point. As seen in instance (c) of Fig. 15.9 and (c) of Fig. 15.8(i), the result of surgery is the two solid tori V_1 and V_2 forming S^3.

The described process can be viewed as a double surgery resulting from a single attracting center which is inside the first 3-ball D^3 and outside the second 3-ball B^3. This attracting center is illustrated (in blue) in instance (a) of Fig. 15.8 but also in (a) of Fig. 15.9, where it is shown that the colinear attracting forces causing the double surgery can be viewed as acting on D^3 (the two blue arrows) and also as acting on the complement space B^3 (the two dotted blue arrows), since they are applied on the common boundary of the two 3-balls. Note that in both cases, the attracting center coincides with the limit point of the spherical layers that D^3 is made of, that is, their common center and the center of D^3 (shown in green in (a) of Fig. 15.8 and (a) of Fig. 15.9). For more details on the descriptions of S^3 and their relation to surgery, the reader is referred to [2].

The reverse process of embedded solid 2-dimensional 0-surgery on D^3 is an embedded solid 2-dimensional 1-surgery on the solid torus V_2, see instances of Fig. 15.9 in reverse order. This process is the embedded analog of the solid 2-dimensional 1-surgery on a solid torus $D^2 \times S^1$ defined in Definition 15.4 and shown in Fig. 15.5b_1 in reverse order. Here too, the process can be viewed as a double surgery resulting from one attracting center which is outside the first solid torus V_1 and inside the second solid torus V_2. This attracting center is illustrated (in orange) in instance (c) of Fig. 15.9 where it is shown that the coplanar forces causing surgery are applied on the common boundary of V_1 and V_2 and can be viewed as attracting forces along a longitude when acting on V_1 and as attracting forces along a meridian when acting on the complement space V_2.

One can now directly appreciate the correspondence of the physical phenomena (instances (a), (b), (c) of Fig. 15.8(ii)) with our schematic model (instances (a), (b), (c) of Fig. 15.8(i)). Indeed, if one looks at the formation of black holes and examines it as an isolated event in space, this process shows a decompactified view of the passage from a two 3-ball description of S^3, that is, the core of the star and the surrounding space, to a two solid tori description, namely the toroidal accretion disc surrounding the black hole (shown in white in instance (c) of Fig. 15.8(ii)) and the surrounding space.

Remark 15.6 It is worth pinning down the following spatial duality of embedded solid 2-dimensional 0-surgery for $M = D^3$: the attraction of two points lying on the boundary of segment L by the center of D^3 can be equivalently viewed in the complement space as the repulsion of these points by the center of B^3 (that is, the point at infinity) on the boundary of the segment $l - L$ (or the segments, if viewed in $\mathbb{R}^3$). Hence, the aforementioned duality tells us that the attracting forces from the attracting center that are collapsing the core of the star can be equivalently viewed as repelling forces from the point at infinity lying in the surrounding space.

15.4.2.2 A Topological Model for the Formation of Tornadoes

Another example of global phenomenon is the formation of **tornadoes**, recall Fig. 15.4b. As mentioned in Sect. 15.3 this phenomenon can be modelled by solid

Fig. 15.10 **a** Attracting force between the cloud and the earth **b** Funnel-shaped clouds **c** Tornado

2-dimensional 0-surgery, recall Fig. 15.5b$_2$ (from right to left). However, here, the initial manifold is different than D^3. Indeed, if we consider a 3-ball around a point of the cloud and another 3-ball around a point on the ground, then the initial manifold is $M = D^3 \times S^0$. If certain meteorological conditions are met, an attracting force between the cloud and the earth beneath is created. This force is shown in blue in see Fig. 15.10a. Then, funnel-shaped clouds start descending toward the ground, see Fig. 15.10b. Once they reach it, they become tornadoes, see Fig. 15.10c. The only difference compared to our model is that here the attracting center is on the ground, see Fig. 15.10a, and only one of the two 3-balls (the 3-ball of cloud) is deformed by the attraction. This lack of symmetry in the process can be obviously explained by the big difference in the density of the materials.

During this process, a solid cylinder $D^2 \times S^1$ containing many layers of air is created. Each layer of air revolves in a helicoidal motion which is modeled using a twisting embedding as shown in Fig. 15.4a (for an example of a twisting embedding, the reader is referred Sect. 15.1.2). Although all these layers undergo local dynamic 2-dimensional 0-surgeries which are triggered by local forces (shown in blue in Fig. 15.10a), these local forces are not enough to explain the dynamics of the phenomenon. Indeed, the process is triggered by the difference in the conditions of the lower and upper atmosphere which create an air cycle. *This air cycle lies in the complement space of the initial manifold* $M = D^3 \times S^0$ *and of the solid cylinder* $D^2 \times S^1$, *but is also involved in the creation of the funnel-shaped clouds that will join the two initial 3-balls*. Therefore in this phenomenon, surgery is the outcome of global changes and this fact makes tornado formation an example of embedded solid 2-dimensional 0-surgery on $M = D^3 \times S^0$.

It is worth mentioning that the complement space containing the aforementioned air cycle is also undergoing solid 2-dimensional 0-surgery. The process can be seen in $\mathbb{R}^3$ in instances (a) to (d) of Fig. 15.11 while the corresponding view in S^3 is shown in instances (a′) to (d′) of Fig. 15.11.

More precisely, let us name the two initial 3-balls D_1^3 and D_2^3, hence $M = D^3 \times S^0 = D_1^3 \sqcup D_2^3$. Further, let B^3 be the complement of D_1^3 in S^3. This setup is shown in (a′) of Fig. 15.11 where S^3 is viewed as the union of the two 3-balls $D_1^3 \cup B^3$, and here too, B^3 represents everything outside D_1^3. The complement space of the initial manifold, $S^3 \setminus M = B^3 \setminus D_2^3$, is the 3-ball B^3 where D_2^3 has been removed from its interior and its boundary consists in two spheres $S^2 \times S^0$, one bounding B^3 or, equivalently, D_1^3 (the outside sphere) and one bounding D_2^3 (the inside sphere).

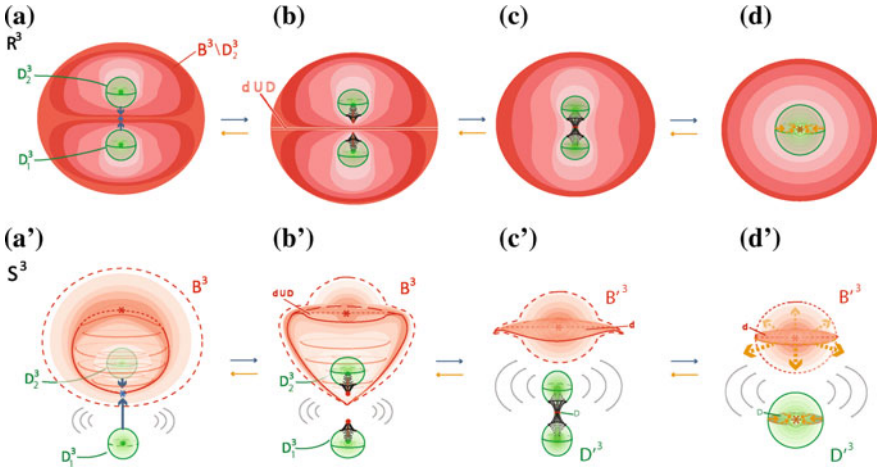


Fig. 15.11 Embedded solid 2-dimensional 0-surgery on $D_1^3 \sqcup D_2^3$ (from left to right) and embedded solid 2-dimensional 1-surgery on D'^3 (from right to left)

Next, D_1^3 and D_2^3 approach each other, see Fig. 15.11b′. In (c′) of Fig. 15.11, D_1^3 and D_2^3 merge and become the new 3-ball D'^3, see Fig. 15.11c′ or Fig. 15.11d′ for a homeomorphic representation.

At the moment of merging, the spherical boundary of D_2^3 punctures the boundary of B^3; see the passage from (b′) – (c′) of Fig. 15.11. As a result, the complement space is transformed from $B^3 \setminus D_2^3$ to the new deformed 3-ball B'^3, see Fig. 15.11c′ or Fig. 15.11d′ for a homeomorphic representation. Note that, although the complement space undergoes a type of surgery that is different from the ones defined in Sect. 15.3 and shown in Fig. 15.5, it can still be defined analogously. In short, we have a double solid 2-dimensional 0-surgery which turns $M = D_1^3 \sqcup D_2^3$ into D'^3 and the complement space $S^3 \setminus (D_1^3 \sqcup D_2^3)$ into B'^3. This process is initiated by the attracting center shown (in blue) in Fig. 15.11a′. The created colinear forces can be viewed as acting on $D_1^3 \sqcup D_2^3$ or, equivalently, as acting on the complement space $S^3 \setminus (D_1^3 \sqcup D_2^3)$ (see the two blue arrows for both cases).

Going back to the formation of tornadoes, the above process describes what happens to the complement space and provides a topological description of the behavior of the air cycle during the formation of tornadoes. The complement space $B^3 \setminus D_2^3$ in $\mathbb{R}^3$ is shown in red in Fig. 15.11a and its behavior during the process can be seen in instances (b) – (d) of Fig. 15.11. Note that in Fig. 15.11a, $B^3 \setminus D_2^3$ represents the hole space outside D_1^3, which means that the red layers of Fig. 15.11a extend to infinity and only a subset is shown.

15.4.2.3 Embedded Solid 2-Dimensional 1-Surgery on $M = D^3$

We will now discuss the process of *embedded solid 2-dimensional 1-surgery* in S^3. Taking $M = D'^3$ as the initial manifold, embedded solid 2-dimensional 1-surgery is the reverse process of embedded solid 2-dimensional 0-surgery on $D^3 \times S^0$ and is illustrated in Fig. 15.11 from right to left. The process is initiated by the attracting center shown (in orange) in (d′) of Fig. 15.11. The created coplanar attracting forces are applied on the circle which is the common boundary of the meridian of D'^3 and the meridian of B'^3 and they can be viewed as acting on the meridional disc D of the 3-ball D'^3 (see orange arrows) or, equivalently, in the complement space, on the meridional disc d of B'^3 (see dotted orange arrows). As a result of these forces, in Fig. 15.11c′, we see that while disc D of D'^3 is getting squeezed, disc d of B'^3 is enlarged. In Fig. 15.11b′, the central disc d of B'^3 engulfs disc D and becomes $d \cup D$, which is a separating plane in $\mathbb{R}^3$, see Fig. 15.11b. At this point the initial 3-ball D'^3 is split in two new 3-balls D_1^3 and D_2^3; see Fig. 15.11b′ or 15.11a′ for a homeomorphic representation. The center point of D'^3 (which coincides with the orange attracting center) evolves into the two centers of D_1^3 and D_2^3 (in green) which by Definition 15.3, is 2-dimensional 1-surgery on a point. This is exactly the same process as in Fig. 15.5b_2 if we restrict it to D'^3, but since we are in S^3, the complement space B'^3 is also undergoing, by symmetry, solid 2-dimensional 1-surgery.

All natural phenomena undergoing embedded solid 2-dimensional 1-surgery take place in the ambient 3-space. The converse, however, is not true. For example, the phenomena exhibiting 2-dimensional 1-surgery discussed in Sect. 15.2.2 are all embedded in 3-space, but they do not exhibit the intrinsic properties of embedded 2-dimensional surgery, since they do not demonstrate the causal or consequential effects discussed in Sect. 15.4.2 involving the ambient space. Yet one could, for example, imagine taking a solid material specimen, stress it until necking occurs and then immerse it in some liquid until its pressure causes fracture to the specimen. In this case the complement space is the liquid and it triggers the process of surgery. Therefore, this is an example of embedded solid 2-dimensional 1-surgery where surgery is the outcome of global changes.

Remark 15.7 Note that the spatial duality described in embedded solid 2-dimensional 0-surgery, in Remark 15.6, is also present in embedded solid 2-dimensional 1-surgery. Namely, the attracting forces from the circular boundary of the central disc D to the center of D'^3 shown in (d′) of Fig. 15.11, can be equivalently viewed in the complement space as repelling forces from the center of B'^3 (that is, the point at infinity) to the boundary of the central disc d, which coincides with the boundary of D.

Remark 15.8 One can sum up the processes described in this section as follows. The process of embedded solid 2-dimensional 0-surgery on D^3 consists in taking a solid cylinder such that the part $S^0 \times D^2$ of its boundary lies in the boundary of D^3, removing it from D^3 and adding it to B^3. Similarly, the reverse process of embedded solid 2-dimensional 1-surgery on V_2 consists of taking a solid cylinder such that the part $S^1 \times D^1$ of its boundary lies in the boundary of V_2, removing it from V_2

and adding it to V_1. Following the same pattern, embedded solid 2-dimensional 1-surgery on $M = D^3$ consists of taking a solid cylinder in D'^3 such that the part $S^1 \times D^1$ of its boundary lies in the boundary of D'^3, removing it from D'^3 and adding it to B'^3. Similarly, the reverse process of embedded solid 2-dimensional 0-surgery on $S^3 \setminus (D_1^3 \sqcup D_2^3)$ consists of taking a solid cylinder such that the two parts $S^0 \times D^2 = D_1^2 \sqcup D_2^2$ of its boundary lie in the corresponding two parts of the boundary of $S^3 \setminus (D_1^3 \sqcup D_2^3)$, removing it from $S^3 \setminus (D_1^3 \sqcup D_2^3)$ and adding it to $D_1^3 \sqcup D_2^3$. Note that, for clarity, in the above descriptions the attracting centers causing surgery are always inside the initial manifold. Of course a similar description starting with the complement space as an initial manifold and the attracting center outside of it would also have been correct.

15.5 Conclusions

In this paper, inspired by natural phenomena exhibiting topological surgery, we introduced the notions of dynamic, solid and embedded surgery in dimensions 1, 2 and 3. There are many more natural phenomena exhibiting surgery and we believe that the understanding of their underpinning topology will lead to the better understanding of the phenomena themselves, as well as to new mathematical notions, which will in turn lead to new physical implications.

Acknowledgements We wish to express our gratitude to Louis H. Kauffman and Cameron McA.Gordon for many fruitful conversations on topological surgery. We would also like to thank the Referee for his/her positive comments and for helping us clarify some key notions. We further wish to acknowledge that this research has been co-financed by the European Union (European Social Fund - ESF) and the Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES: Reinforcement of the interdisciplinary and/or inter-institutional research and innovation.

References

1. Adams, C.: The Knot Book, An Elementary Introduction to the Mathematical Theory of Knots. American Mathematical Society (2004)
2. Antoniou, S., Lambropoulou, S.: Extending Topological Surgery to Natural Processes and Dynamical Systems. PLOS ONE (2017). https://doi.org/10.1371/journal.pone.0183993
3. Dahlburg, R.B., Antiochos, S.K.: Reconnection of antiparallel magnetic flux tubes. J. Geophys. Res. **100**(A9), 16991–16998 (1995). https://doi.org/10.1029/95JA01613
4. Kiehn, R.M.: Non Equilibrium Thermodynamics. Non-equilibrium systems and irreversible processes - Adventures in applied topology, pp. 147–150. University of Houston Copyright CSDC Inc (2004)
5. Laing, C.E., Ricca, R.L., Sumners, D.: Conservation of writhe helicity under anti-parallel reconnection. Scientific Reports 5, No 9224 (2014). https://doi.org/10.1038/srep09224

6. Lambropoulou, S., Antoniou, S.: Topological surgery, dynamics and applications to natural processes. J. Knot Theory Ramif (2016). https://doi.org/10.1142/S0218216517430027
7. Milnor, J.: Morse Theory. Princeton University Press (1963)
8. Ott, C.D., et al.: Dynamics and gravitational wave signature of collapsar formation. Phys. Rev. Lett. **106**, 161103 (2011). https://doi.org/10.1103/PhysRevLett.106.161103
9. Prasolov, V.V., Sossinsky, A.B.: Knots, Links, Braids and 3-Manifolds. AMS translations of mathematical monographs (1997)
10. Ranicki, A.: Algebraic and Geometric Surgery. Clarendon Press, Oxford Mathematical Monographs (2002)
11. Rolfsen, D.: Knots and Links. Publish or Perish Inc, AMS Chelsea Publishing (2003)
12. Samardzija, N., Greller, L.: Explosive route to chaos through a fractal torus in a generalized lotka-volterra model. Bull. Math. Biol. **50**(5), 465–491 (1988). https://doi.org/10.1007/BF02458847
13. Samardzija, N., Greller, L.: Nested tori in a 3-variable mass action model. Proc. R. Soc. Lond. Ser. A Math. Phys. Sci. **439**(1907), 637–647 (1992). https://doi.org/10.1098/rspa.1992.0173
14. Wasserman, S.A., Dungan, J.M., Cozzarelli, N.R.: Discovery of a predicted DNA knot substantiates a model for site-specific recombination. Science **229**, 171–174 (1985)

Part III
Algebraic Modeling of Computational Structures

Chapter 16
Automorphisms of Curves

Jannis A. Antoniadis and Aristides Kontogeorgis

Abstract This is a survey article concerning the groups of automorphisms of curves defined over algebraically closed fields of positive characteristic, their representations and applications to their deformation theory.

16.1 Introduction

By an (algebraic) curve we will mean a projective non-singular one-dimensional variety, defined over an algebraically closed field k of characteristic $p \geq 0$. Over the field $\mathbb{C}$ of complex numbers the notion of a projective algebraic curve coincides with the notion of compact Riemann surface. Every compact Riemann surface X is known to be an orientable two-dimensional real manifold and to any such surface we can attach a natural number $g_X \in \mathbb{N} \cup \{0\}$, called the genus, which topologically counts the number of holes of the surface X. Over an arbitrary field of positive characteristic we can still define the genus, by setting g_X to be the dimension of the space of global holomorphic differentials $H^0(X, \Omega_X)$, although a topological interpretation is less clear, see [72]. In Sect. 16.2.2 a topological interpretation can be given as the number of cycles of the graph of analytic reduction.

An automorphism of a curve X is an isomorphism $\sigma : X \to X$, and the set $\mathrm{Aut}_k(X)$ of all automorphisms form a group under composition. Since we assumed that the constant field is algebraically closed we will omit the index k from the notation, and we will denote the automorphism group by $\mathrm{Aut}(X)$.

If the genus is zero, then X is isomorphic to the projective line $\mathbb{P}^1$ and the automorphism group is the group of Möbius transformations $\mathrm{PGL}(2, k)$, which is infinite. If

J.A. Antoniadis
Department of Mathematics & Applied Mathematics, University of Crete,
Voutes Campus, 70013 Heraklion, Greece
e-mail: antoniad@math.uoc.gr

A. Kontogeorgis (✉)
Department of Mathematics, National and Kapodistrian University of Athens,
Panepistimioupolis, 15784 Athens, Greece
e-mail: kontogar@math.uoa.gr

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_16

$g_X = 1$, then the curve X admits a group structure and X acts on X by translations, hence $X \subset \mathrm{Aut}(X)$, so X is infinite as well.

When $g_X \geq 2$ the automorphism group is finite. In characteristic zero this can be proved by applying the Riemann–Roch theorem, in order to see that every automorphism fixing more than $2g_X + 2$ points is the identity, and then obtaining a faithful representation on the set of Weierstrass points. For the case of positive characteristic, the same argument does not work. In order to prove that the automorphism group is finite, we first prove that the decomposition group $G(P)$ is finite, for every $P \in X$. Then we need the existence of a finite non-empty $\mathrm{Aut}(X)$-invariant set Σ. We can use as Σ the set of Weierstrass points, since they are invariant under the action of the automorphism group. A notion of Weierstrass points in positive characteristic was given first by F.K. Schmidt [75] using the theory of Hasse derivatives. For a modern account of this topic we refer to [26, Sect. 6], [29, Chap. 11].

Moreover, it is known that for any finite group G there exists a curve X such that $\mathrm{Aut}(X) \cong G$, see [52]. Notice that most of the curves have trivial automorphism group [57, 67], since curves with non-trivial automorphisms correspond to singularities of the moduli space of curves of fixed genus. However, finding specific examples of curves without automorphisms is not easy, see [65].

Understanding the automorphism group is an interesting problem on its own and has many applications to counting points, moduli problems, etc. In Sect. 16.2 we will present results concerning the order of the group and we will give upper bounds in terms of constants of topological nature, like the genus and the p-rank of the Jacobian.

In Sect. 16.5 we will study automorphisms of relative curves $\pi : \mathscr{X} \to \mathrm{Spec} R$, where R is a discrete valued ring with algebraically closed residue field. We will restrict ourselves to maps π which have fibers X of genus $g_X \geq 2$ and which vary "nicely". Both these properties are formulated by the notion of "*stable curve*". The precise definition is given in [14, Definition I.1]. If the relative curve is stable, then the automorphism group of the special fibre contain the automorphism group of the generic fibre [14, Lemma I.12]. The study of automorphisms of relative curves is a difficult problem even at the infinitesimal level. In this section we also discuss reduction, lifting and the deformation problem. Automorphisms of relative curves are related to the representation theory of the automorphism group on several natural objects of the curve like global sections of global polydifferentials and this will be explained in Sect. 16.3. In Sect. 16.6 we study integral representations of a fibrewise action in relative holomorphic polydifferentials.

The theory of automorphisms of curves is a vast object of study and this article does not have the ambition to describe it completely. It is rather focused on subjects closer to the research interests of the authors. For more general information about automorphisms of curves in characteristic zero we refer to [15, Chap. V], while for curves defined over fields of positive characteristic we refer to [29, Chap. 11], [13, par. 14.3].

16.2 Size of the Group

The automorphism group of a curve X of genus $g_X \geq 2$ is finite. However if the genus is fixed, then the size of the groups that can appear is bounded. In characteristic zero, for $g_X \geq 2$, using the theory of Riemann–Hurwitz formula and by a case by case examination, Hurwitz [31] proved the bound:

$$|\mathrm{Aut}(X)| \leq 84(g_X - 1). \tag{16.1}$$

16.2.1 Ramification Filtration

In order to explain the situation in characteristic $p > 0$ we have to introduce the ramification filtration at a closed point $x \in X$. Let G be a subgroup of $\mathrm{Aut}(X)$ and let $m_{X,x}$ be the maximal ideal of the local ring $\mathscr{O}_{X,x}$. We will denote by $k(x)$ the residue field of x. For $i \geq 0$, the ith lower ramification subgroup $G_{x,i}$ of G at x is the subgroup of all elements $\sigma \in G$ which fix x and which act trivially on $\mathscr{O}_{X,x}/m_{X,x}^{i+1}$. These groups form a decreasing finite sequence

$$G_{x,0} \rhd G_{x,1} \rhd \cdots \rhd G_{x,n} \rhd G_{x,n+1} = \{1\}, \qquad n \in \mathbb{N}. \tag{16.2}$$

When the characteristic $p = 0$ it is known that $G_{x,1} = \{1\}$. In general $G_{x,0}/G_{x,1}$ is a cyclic group of order prime to the characteristic, while for $i \geq 1$ $G_{x,i}/G_{x,i+1}$ is an elementary abelian group, i.e. isomorphic to the direct sum of finitely many cyclic groups of order p. If $G_{x,1} = \{1\}$ for every $x \in X$, then the cover $X \to X/G$ is said to be *tame*, otherwise it is called *wild*. If $G_{x,2} = \{1\}$, then the ramification is called *weak*.

The Riemann–Hurwitz formula [82, Sect. 3.4 p. 90] relates the genera of the curves X and $X/G = Y$ as follows:

$$2(g_X - 1) = 2(g_Y - 1)|G| + \sum_{x \in X} \sum_{i=1}^{\infty} \left(|G_{x,i}| - 1\right). \tag{16.3}$$

Notice that this equation can be obtained by taking degrees on Eq. (16.11).

For tame covers the Hurwitz bound remains the same. For the general wildly ramified curve H. Stichtenoth [80, 81] proved that the following bound holds:

$$|\mathrm{Aut}(X)| \leq 16 g_X^4, \tag{16.4}$$

with the Hermitian Fermat curve as only exception, given by the equation:

$$0 = x^{p^h+1} + y^{p^h+1} + z^{p^h+1} = \underline{x}\left(\underline{x}^t\right)^{p^h}, \text{ where we have set } \underline{x} = (x, y, z). \tag{16.5}$$

Notice that for $h = 0$ the above Fermat curve is a quadratic form, while for $h > 0$ it behaves as Frobenious shifted quadratic form and has $\mathrm{PGU}(3, p^{2h})$ as automorphism group, [43, 51]. The result of H. Stichtenoth was improved by H. Henn [27] who proved that

$$|\mathrm{Aut}(X)| \leq 8g_X^3, \tag{16.6}$$

with a finite list of exceptions. The result of Henn contained a gap which was filled by M. Giulietti and Gábor Korchmáros, see [20].

All exceptions in the list of Henn, have a large p-subgroup compared to its genus. C. Lehr and M. Matignon [50] defined the notion of "big action", when the $\mathrm{Aut}(X)$ contains a p-subgroup P, such that

$$|P| > \frac{2p}{p-1} g_X. \tag{16.7}$$

M. Matignon and M. Rocher [55, 70, 71], and M. Giulietti and G. Korchmáros [19] studied and classified "big actions" defined by an equation similar to Eq. (16.7).

In characteristic $p > 0$ the p-rank of the Jacobian γ_X plays a role analogous to the rank of the homology group and as a matter of fact $0 \leq \gamma_X \leq g_X$. Curves with $g_X = \gamma_X$ are called ordinary and they form a Zariski-dense set in the moduli space of curves of fixed genus. For such curves S. Nakajima [62] proved the bound:

$$|\mathrm{Aut}(X)| \leq 84(g_X - 1)g_X. \tag{16.8}$$

He further notices that his bound could not be best possible and by studying the Artin–Schreier–Mumford curve

$$(x^{p^h} - x)(y^{p^h} - y) = c, \qquad c \in k \tag{16.9}$$

he conjectured that the best possible bound is given by a cubic polynomial in $\sqrt{g_X}$.

16.2.2 Mumford Curves

It is well known that an algebraic curve X, defined over $\mathbb{C}$ can be uniformized by a discrete subgroup Γ of $\mathrm{PSL}(2, \mathbb{R})$, i.e. $X \cong \Gamma \backslash \mathbb{H}$, and the Hurwitz upper bound given in Eq. (16.1) is equivalent to Siegel's lower bound $\pi/21$ on the volume of the fundamental domain of a Fuchsian group ([15, Exercise 6 p. 245], [49]).

Let K be a non-archimedean valued field. D. Mumford [59] showed that curves defined over K, whose stable reduction is split multiplicative, i.e. a union of rational curves intersecting at $\bar{K}$-rational points, are isomorphic to an analytic space of the form $\Gamma \backslash (\mathbb{P}^1_K - \mathscr{L}_\Gamma)$, where Γ is a discontinuous group in $\mathrm{PGL}(2, K)$ and $\mathscr{L}_\Gamma$ is the set of limit points. The automorphism group of the curve X is then isomorphic to the group N/Γ, where N equals the normalizer of $\Gamma \in \mathrm{PGL}(2, K)$, [12], [18, p. 216].

We will call such curves *Mumford curves*. Notice that not all curves defined over K admit such a uniformisation. For example the Artin–Schreier Mumford curve has split multiplicative reduction and is a Mumford curve only if $|c| < 1$. The uniformization theory can give stronger results when applied to Mumford curves.

Herrlich [28] has shown that for p-adic Mumford curves of genus $g_X \geq 2$ and $p \geq 7$ the Hurwitz bound can be strengthened to $12(g_X - 1)$.

Notice that by the work of Manin-Drinfeld [53] and Gerritzen [17], Mumford curves are known to be ordinary, therefore the Nakajima bound given in Eq. (16.8) holds. For Mumford curves defined over non-archimedean fields of positive characteristic G. Cornelissen, F. Kato and the second author [12], proved that Nakajima's conjecture was correct for Mumford curves and the following bound holds:

$$|\mathrm{Aut}(X)| \leq \max\left\{12(g_X - 1), 2\sqrt{g_X}(\sqrt{g_X} + 1)^2\right\}. \tag{16.10}$$

They also classified those curves for which $|\mathrm{Aut}(X)| \geq 12(g_X - 1)$. Moreover, the above bound is best possible since it is attained for the Artin–Schreier–Mumford curves given by Eq. (16.9).

This theorem can also be reformulated in the style of Siegel lower bound as follows: the $\mu(N)$ invariant [37, Eq. 2] of its normalizer N is bounded from below by

$$\mu(N) \geq \min\left\{1/12, \frac{\sqrt{g_X} - 1}{2\sqrt{g_X}(\sqrt{g_X} + 1)}\right\}.$$

Notice that $\mu(N)$ plays the role of a Gauss–Bonnet "volume" and the index $[N : \Gamma]$ which equals the order of automorphism group can be evaluated in terms of theorems of HNN groups as in Theorem 2 in [37].

Concerning the Nakajima conjecture for ordinary curves X over a field of characteristic $p > 0$, R. Guralnik and M. Zieve in a Workshop in Leiden on Automorphisms of curves in 2004, announced that there exists a sharp bound of the order of $g_X^{8/5}$ for $|\mathrm{Aut}(X)|$.

For automorphisms groups of Mumford curves with a specific structure we can have better bounds. For example S. Nakajima in [61] used the Hasse-Arf theorem in order to prove that

$$|\mathrm{Aut}(X)| \leq 4g_X + 4,$$

and this bound has been further improved for abelian automorphisms groups of Mumford curves by V. Rotger and the second author in [47], to the bound

$$|\mathrm{Aut}(X)| \leq 4(g_X - 1).$$

16.3 Representation Theory

The next step is to understand representations of $\mathrm{Aut}(X)$ in some naturally defined vector spaces. Let Ω_X denote the sheaf of relative differentials of X over k and by $H^0(X, \Omega_X^{\otimes m})$ the space of global holomorphic polydifferentials of X. The automorphism group acts on both Ω_X and $H^0(X, \Omega_X^{\otimes m})$, therefore $H^0(X, \Omega_X^{\otimes m})$ becomes a $k[G]$-module of k-dimension equal to $(2m-1)(g-1)$ if $m \neq 1$ or g if $m = 1$. By the work of B. Köck and J. Tait [41] we know that this action is faithful, unless $\mathrm{Aut}(X)$ contains a hyperelliptic involution and either $m = 1$ and $p = 2$ or $m = 2$ and $g_X = 2$.

It is a classical problem proposed first by Hecke [25], to analyse the $k[G]$-module structure of $H^0(X, \Omega_X^{\otimes m})$, i.e. analyse the indecomposable components together with their multiplicities. If the characteristic does not divide $|G|$, this problem was solved by Chevalley and Weil [9].

If the ramification of $X \to X/G$ is tame, then Nakajima [60, Theorem 2] and, independently, Kani [32, Theorem 3] determined the $k[G]$-module structure of $H^0(X, \Omega_X)$. B. Köck in [40] studied weakly ramified covers, he generalized Kani's and Nakajima's work and corrected a criterion for the projectivity of the space of holomorphic differentials given by Kani, see remark 2.4b.

K. Ward in [85] studied the Galois module structure of holomorphic differentials for the cyclotomic function fields obtained by the torsion points of Carlitz modules C_M for a totally split polynomial $M \in \mathbb{F}_q(T)$.

The case when G is a cyclic group was first studied by Valentini and Madan [84, Theorem 1] who considered cyclic p-groups (and also revisited cyclic groups of order prime to the characteristic, [84, Theorem 2]). The case of a general cyclic G was treated by S. Karanikolopoulos and the second author [35, Theorem 7]. A different, general approach to determining the decomposition of global sections of coherent $\mathcal{O}_X - G$-modules into decomposable direct summands was developed by Borne in [7], using the notion of rings with several objects. Some formulas concerning the case of cyclic groups and curves are given in [7, Sect. 7.2].

The situation in positive characteristic is more difficult, because phenomena of modular representation theory appear; for example, the notion of irreducible representation is different than the notion of indecomposable representation. Moreover wild ramification appears: the decomposition groups are not cyclic groups and higher ramification groups appear, see Eq. (16.2). Also the classification of non-cyclic p-groups even for the simplest group $G = \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, for a prime $p > 2$ is considered to be impossible [3, p.13 Sect. 1.2].

For each closed point $x \in X$, let $m_{X,x}$ be the maximal ideal of the local ring $\mathcal{O}_{X,x}$ and let $k(x)$ be the residue field of x. The fundamental character of the inertia group $G_{x,0}$ of x is the character $\theta_x : G_{x,0} \to k(x)^* = \mathrm{Aut}(m_{X,x}/m_{X,x}^2)$ giving the action of $G_{x,0}$ on the cotangent space of x. Here θ_x factors through the maximal p'-quotient $G_{x,0}/G_{x,1}$ of $G_{x,1}$.

In [5] F. Bleher, T. Chinburg and the second author studied the structure of $H^0(X, \Omega_X)$, when G is any group such that the p-Sylow subgroup of G is cyclic. It turns out that the $k[G]$-module structure depends only on the ramification data and the fundamental characters of closed points of X, ramified under the action of G.

16.4 Bases of Holomorphic Differentials

16.4.1 Boseck Theory

A strategy for studying the $k[G]$-module structure is to first write explicit bases of the spaces $H^0(X, \Omega_X^{\otimes m})$. Usually, a curve with a non-trivial automorphism group comes with a natural Galois cover $\pi : X \to X/H = Y$, where Y is a known curve (usually $\mathbb{P}^1$ or an elliptic curve) and H is a subgroup of the full automorphism group G. In this way the divisor of the H-invariant differential $\mathrm{div}\pi^* dx$ can be computed in terms of the pullback formula [24, prop. 2.3 p. 301]

$$\mathrm{div}\pi^* dx \cong \pi^* \mathrm{div}(dx) + R_{X/Y}, \tag{16.11}$$

where $R_{X/Y}$ denotes the ramification divisor of the cover.

Once the divisor $\mathrm{div}(dx)$ is computed, finding the space of holomorphic (poly) differentials is the same as computing the Riemann–Roch space $L(\mathrm{div}dx)$. This method was used by H. Boseck in [8], who gave precise formulas for both Kummer and Artin–Schreier extensions of the projective line. Once a basis is constructed, one has to identify the indecomposable summands. For the case of cyclic group action the last computation essentially is equivalent to the computation of the Jordan normal form. Notice that Boseck's article has an error concerning the computation of Weierstrass points, see the article of A. Garcia [16] for more details.

This method was used in [35, 84] and also by articles of Rzedowski–Calderón Villa-Salvador and Madan [73] and Marques and Ward [54] for some other groups under additional hypotheses on the cover $X \to X/G$.

16.4.2 Mumford Curves

For the case of Mumford curves there is a pure group theoretic approach to the determination of global sections of holomorphic differentials initiated by the work of V. Drinfeld and Y. Manin [53]. For holomorphic polydifferentials there is also a group theoretic approach, the theory of harmonic measures studied by J. Teitelbaum and P. Schneider see [76, 83].

Let K be a complete non-archimedean valued field, $\Gamma \subseteq \mathrm{PGL}(2, K)$ be a Schottky subgroup, and X_Γ the Mumford curve obtained from Γ. We will denote by N the normalizer of Γ in $\mathrm{PGL}(2, K)$. The quotient group $G = N/\Gamma$, which acts on X_Γ from the left, is the automorphism group $\mathrm{Aut}(X_\Gamma)$ of X_Γ over K. Recall that Γ is a free group of finite rank, whose rank, say g, is equal to the genus of X_Γ. Let us fix a free generating set $\{\gamma_1, \ldots, \gamma_g\}$ of Γ. For any right $K[\Gamma]$-module P, each derivation $d\colon \Gamma \to P$ is uniquely determined by its values $h_i = d(\gamma_i)$ for $1 \le i \le g$, and conversely, since Γ is free, such values $h_i \in P$ can be freely chosen to obtain a derivation d; indeed, once h_i's are chosen, then $d(w)$ for any $w \in \Gamma$ is uniquely determined by the derivation rules.

For a positive integer n, we consider the $2n-1$ dimensional vector space of polynomials $P_n \subseteq K[T]$ of degree $\le 2(n-1)$. The group $\mathrm{PGL}(2, K)$ acts on P_n from the right as follows: for $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PGL}(2, K)$ and $F \in P_n$, we define

$$F^\gamma(T) := \frac{(cT+d)^{2(n-1)}}{(ad-bc)^{n-1}} F\left(\frac{aT+b}{cT+d}\right) \in K[T]. \tag{16.12}$$

Now, consider the $(2n-1)g_{X_\Gamma}$-dimensional space $\mathrm{Der}(\Gamma, P_n)$ of derivations, which can be seen as an N-module as follows: for $\delta \in N$ and $d \in \mathrm{Der}(\Gamma, P_n)$, define

$$(d^\delta)(\gamma) = [d(\delta\gamma\delta^{-1})]^\delta \tag{16.13}$$

for $\gamma \in \Gamma$. There is then a well defined $G = N/\Gamma$ action on the group cohomology $H^1(\Gamma, P_n)$, since Γ acts trivially modulo principal derivations.

Theorem 16.1 ([83, Theorem 1]) *For any $n \ge 1$, the space $H^0(X_\Gamma, \Omega_{X_\Gamma}^{\otimes n})$ of n-differentials on the curve X_Γ is naturally isomorphic to the space group cohomology $H^1(\Gamma, P_n)$. Moreover, this identification is G-equivariant with respect to the natural right G-action on $H^0(X_\Gamma, \Omega_{X_\Gamma}^{\otimes n})$.* □

F. Kato and the second author [38] used this approach to study the $K[G]$-module structure of polydifferentials for the case of Artin–Schreier–Mumford curves, where $N = A * B$, $\Gamma = [A, B]$ and $A, B \subset \mathrm{PGL}(2, K)$ are cyclic groups of order p generated by

$$\varepsilon_A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \varepsilon_B = \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}, \tag{16.14}$$

respectively, where $s \in K^\times$ and $|s| > 1$.

16.5 Curves in Families

16.5.1 Stable Curves

Let $\mathscr{X} \to R$ be a family of curves of genus $g \geq 2$ over a base scheme $S := \mathrm{Spec} R$, where R is a discrete valuation ring with algebraically closed residue field. For every point $P : \mathrm{Spec} k \to S$, we will consider the *absolute* automorphism group of the fibre P to be the automorphism group $\mathrm{Aut}_{\bar{k}}(\mathscr{X} \times_S \mathrm{Spec}\bar{k})$ where $\bar{k}$ is the algebraic closure of k. Any automorphism σ acts like the identity on $\bar{k}$ so in our setting there is no $\mathrm{Gal}(\bar{k}/k)$ contribution to the automorphism group of any special fibre. The following theorem due to P. Deligne and D. Mumford [14, Lemma I.12] compares the automorphism groups of the generic and special fibres:

Theorem 16.2 *Consider a stable curve $\mathscr{X} \to S$ and let $\mathscr{X}_\eta$ denote its generic fibre. Every automorphism $\phi : \mathscr{X}_\eta \to \mathscr{X}_\eta$ can be extended to an automorphism $\phi : \mathscr{X} \to \mathscr{X}$.*

The example of the Fermat curve given in Eq. (16.5), shows that the automorphism group of the special fibre can be strictly bigger. A special fibre $\mathscr{X}_p := \mathscr{X} \times_S S/p$ with $\mathrm{Aut}(\mathscr{X}_p) > \mathrm{Aut}(\mathscr{X}_\eta)$ will be called *exceptional*. In general we know that there are finite many exceptional fibres and it is an interesting problem to determine exactly all of them.

There are some results towards this problem for some curves of arithmetic interest. A. Adler [1] and C.S. Rajan [68] proved for the modular curves $X(N)$, that $X(11)_3 := X(11) \times_{\mathrm{Spec}\mathbb{Z}} \mathrm{Spec}\mathbb{F}_3$ has the Mathieu group M_{11} as the full automorphism group. C. Ritzenthaler in [69] and P. Bending, A. Carmina, R. Guralnick in [2] studied the automorphism groups of the reductions $X(q)_p$ of modular curves $X(q)$ for various primes p. It turns out that the reduction $X(7)_3$ of $X(7)$ at the prime 3 has an automorphism group $\mathrm{PGU}(3, 3)$, and $X(7)_3$ and $X(11)_3$ are the only cases where $\mathrm{Aut}X(q)_p > \mathrm{Aut}X(q) \cong \mathrm{PSL}(2, p)$. Also Y. Yang together with the second author in [48] studied special fibers of hyperelliptic modular curves.

In this spirit, a particular interesting problem is the lifting of automorphisms in characteristic zero: Let X be a curve defined over a field of characteristic p and a group $G \subset \mathrm{Aut}(X)$. Is there a smooth family $\mathscr{X} \to \mathrm{Spec}S$, where S is a local ring with closed point k and generic point a field of characteristic zero, such that G acts fibrewise on the family and the special fibre is the initial curve X?

These types of lifting problems where initiated by J.P. Serre in [79] in his attempt (before étale cohomology was invented) to define an appropriate cohomology theory, which could solve the Weil conjectures.

The answer is no for general G. For example in zero characteristic the Hurwitz bound holds, while in positive characteristic there are known examples of automorphism groups that exceed this bound. Frans Oort in 1987 conjectured that such a lift always exists if the group G is cyclic. This was known in the literature as the Oort Conjecture until recently. Florian Pop proved in [66] that this conjecture is true in a stronger sense: in the case where G has only cyclic groups as inertia groups. We

must mention that Pop's proof is based on recent results by Obus and Wewers [64]. For a survey article and for a complete list of the protagonists for this effort see [63] and the historical note in [66].

16.5.2 Deformations of Curves

We will now explain infinitesimal deformation problems from the viewpoint of M. Schlessinger [74]. A deformation of the curve X is a relative curve $\mathscr{X} \to \mathrm{Spec}(R)$ (proper, smooth) over a local ring R with maximal ideal m and $R/m \cong k$, such that $X \cong \mathscr{X} \times_{\mathrm{Spec}R} \mathrm{Spec}R/m$, i.e. we have the following commutative diagram:

$$\begin{array}{ccc} X \cong \mathscr{X} \times_{\mathrm{Spec}R} \mathrm{Spec}R/m & \longrightarrow & \mathscr{X} \\ \downarrow & & \downarrow \\ \mathrm{Spec}(k) \cong R/m & \longrightarrow & \mathrm{Spec}(R) \end{array}$$

Two deformations $\mathscr{X}_1$ $\mathscr{X}_2$ are considered to be equivalent if there is an isomorphism $\psi : \mathscr{X}_1 \to \mathscr{X}_2$ making the diagram

$$\begin{array}{ccc} \mathscr{X}_1 & \xrightarrow{\psi} & \mathscr{X}_2 \\ & \searrow \quad \swarrow & \\ & \mathrm{Spec}R & \end{array}$$

commutative, such that ψ gives the identity on the special fibres.

Definition 16.1 We consider a deformation functor from the category $\mathscr{C}$ of local Artin algebras R with $R/m_R \cong k$, to the category of sets:

$$D : \mathscr{C} \to \mathrm{Sets},$$

$$R \mapsto \left\{ \begin{array}{l} \text{Equivalence classes of} \\ \text{deformations of } X \text{ over } R \end{array} \right\}.$$

We define the tangent space to the deformation functor to be $D(k[\varepsilon]/\langle\varepsilon^2\rangle)$.

The space $D(k[\varepsilon]/\langle\varepsilon^2\rangle)$ is known to be a vector space [74] and by Chech theory and affine triviality we can show [23, p. 89] that:

$$D(k[\varepsilon]/\langle\varepsilon^2\rangle) = H^1(X, T_X), \tag{16.15}$$

where $T_X \cong \Omega_X^*$ is the tangent sheaf of the curve X.

We now fix a pair (X, G) of curves together with a subgroup G of the automorphism group. A deformation of (X, G) over the local ring R is a deformation of the curve X over R together with a group isomorphism $G \to \mathrm{Aut}_R(\mathscr{X})$, such that there is a G-equivariant isomorphism ϕ from the fibre over the closed point of A to the original curve X:

$$\phi : \mathscr{X} \otimes_{\mathrm{Spec}(A)} \mathrm{Spec}(k) \to X.$$

The notion of equivalence of (X, G) deformations is similar to the non equivariant case, but we now assume that ψ is also G-equivariant. A deformation functor is then defined:

$$D_{(X,G)} : \mathscr{C} \to \mathrm{Sets},$$

$$R \mapsto \left\{ \begin{array}{l} \text{Equivalence classes} \\ \text{of deformations of} \\ \text{couples } (X, G) \text{ over } R \end{array} \right\}$$

J. Bertin and A. Mézard [4] proved that there is an equivariant analogon of Eq. (16.15)

$$T := D_{(X,G)}(k[\varepsilon]/\langle\varepsilon^2\rangle) \cong H^1(G, X, T_X),$$

where $H^1(G, X, T_X)$ is Grothendiecks's equivariant cohomology as defined in [21]. Cohomology theories appear as derived functors of appropriate left exact functors. For example for group cohomology we apply the functor of invariant elements of G-modules, and for Zarisky cohomology the functor of global sections. Grothendieck's equivariant cohomology [21] appears naturally when we consider the composition of two left exact functors. In this setting we consider both the functor of global sections and the functor of group invariants.

Geometrically the space $H^1(X, T_X)$ can be interpreted as the tangent space to the moduli space of curves of genus g_X, computed at the point-curve X. It consists of equivalence classes of infinitesimal deformations of the curve X. Similarly the space $H^1(G, X, T_X)$ can be interpreted as the subspace of $H^1(X, T_X)$ consisted of G-invariant elements, which give rise to infinitesimal deformations acted on by G (Fig. 16.1).

16.5.3 Dimension of the Tangent Space to the Deformation Functor

The study of the space $D_{(X,G)}(k[\varepsilon])$ can be reduced to the short exact sequence [4]:

$$0 \to H^1\left(X/G, \pi_*^G(T_X)\right) \to H^1(G, X, T_X) \to H^0\left(X/G, R^1\pi_*^G(T_X)\right) \to 0.$$

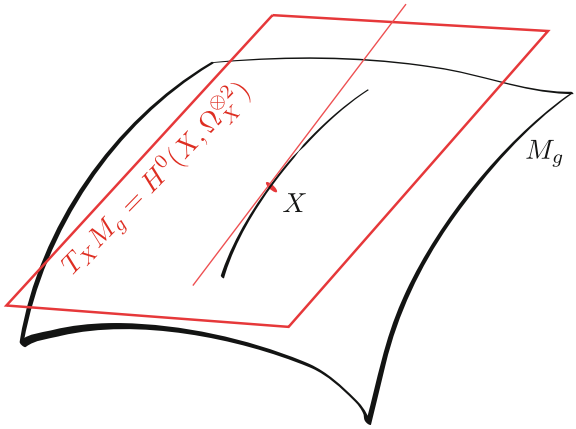


Fig. 16.1 Tangent space to the deformation functor

Suppose that in the cover $X \to X/G$ there are r ramified points $x_1, \ldots, x_r$ and set $e_i^{(\mu)} = |G_{x_i,\mu}|$, for $i = 1, \ldots, r$, $\mu \in \mathbb{N}$. The first factor can be computed using Riemann–Roch theorem [4]

$$\dim H^1(X/G, \pi_*^G(T_X)) = 3g_{X/G} - 3 + \sum_{\mu=1}^{r} \left\lceil \sum_{i=0}^{n_\mu} \frac{e_i^{(\mu)} - 1}{e_0^{(\mu)}} \right\rceil.$$

The second functor can be expressed in terms of group cohomology:

$$H^0\left(X/G, R^1\pi_*^G(T_X)\right) \cong \bigoplus_{i=1}^{r} H^1(G_{0,x_i}, \hat{T}_{X,x_i}), \tag{16.16}$$

where the later sum runs over all wildly ramified points and by $H^1(G_{0,x_i}, \hat{T}_{X,x_i})$ we mean the first cohomology groups, and $\hat{T}_{X,x_i} = k[[t]]d/dt$ is the local tangent space at x_i, while the action of G is the adjoint action:

$$\left(f(t)\frac{d}{dt}\right)^{\sigma} = f(t)^{\sigma}\sigma\frac{d}{dt}\sigma^{-1} = f(t)^{\sigma}\sigma\left(\frac{d\sigma^{-1}(t)}{dt}\right)\frac{d}{dt}.$$

The computation of group cohomology in Eq. (16.16) is manageable only for explicit covers, in particular for Artin–Schreier extensions [11]. One idea exploited by the second author in [44] is to use that the decomposition group G_{x_i} admits a ramification filtration given in Eq. (16.2), where the successive quotients are elementary abelian groups given by Artin–Schreier extensions.

Therefore one can use the Lyndon-Hochshild-Serre spectral sequence [30] which connects the cohomology of the extensions of groups

$$1 \to H \to G \to G/H \to 1,$$

giving rise to a 5-term exact sequence:

$$0 \to H^1(G/H, A^H) \xrightarrow{\inf} H^1(G, A) \xrightarrow{\text{res}} H^1(H, A)^{G/H} \xrightarrow{\text{tg}} H^2(G/H, A^H) \xrightarrow{\inf} H^2(G, A)$$

Unfortunately, the transgression map can be effectively computed only in special cases like the next theorem [44], which limits the usage of this method.

Theorem 16.3 *If G is an abelian group and $G/H \cong \mathbb{Z}/p$, $G \cong G/H \times H$ then the transgression map is zero.*

16.5.4 Representation Theory and the Tangent Space

Serre duality allows us to compute

$$H^1(X, T_X) \cong H^0(X, \Omega^{\otimes 2})^*, \tag{16.17}$$

and the dimension of the later space can be effectively computed using Riemann–Roch theorem to be $3g_X - 3$. In [45] the second author proposes an equivariant form of Eq. (16.17)

$$D_{\text{gl}}(k[\varepsilon]/\langle\varepsilon^2\rangle) = H^1(X, T_X)^G \cong H^0(X, \Omega_X^{\otimes 2})_G.$$

Notice that the space of invariants becomes the space of the co-invariants on the dual space, where for a G-module A, the spaces of invariants and coinvariants respectively, are given by

$$A^G := \{a \in A : a^g = a\} \qquad A_G := A/\langle ga - a : a \in A, g \in G\rangle.$$

For $G = \mathbb{Z}/p$ we have $A^G \cong A_G$, but if G is a more complicated group like $G = \mathbb{Z}/p \times \cdots \times \mathbb{Z}/p$ we can have $A^G \ncong A_G$.

The idea of this construction is that the knowledge of $k[G]$-module structure can lead to the computation of dim $D_{\text{gl}}(k[\varepsilon])$.

S. Karanikolopoulos [33] pursued this idea by studying elementary abelian extensions given as Artin-Schreier extensions: $F/K(x)$ with

$$y^{p^n} - y = \frac{g(x)}{(x - a_1)^{\Phi(1)} \cdots (x - a_s)^{\Phi(s)}}$$

using a modified Boseck construction in order to compute the Galois module structure of global polydifferentials. It turns out that

$$H^0(X, \Omega_X^{\otimes m}) \cong \bigoplus_{j=1}^{p^n} W_j^{d_j},$$

where

$$\Gamma_k(m) = \sum_{i=1}^{s} \left\lfloor \frac{m(p^n-1)(\Phi(i)+1) - k\Phi(i)}{p^n} \right\rfloor,$$

$$d_{p^n} = \Gamma_{p^n-1}(m) - 2m + 1, \qquad d_j = \Gamma_{j-1}(m) - \Gamma_j(m), \qquad j = 1, \ldots, p^n - 1,$$

$$W_j = \langle \theta_0, \ldots, \theta_{j-1} \rangle_K, \qquad \sigma_\alpha(\theta_i) = \sum_{\ell=0}^{i} \binom{i}{\ell} \alpha^{i-\ell} \theta_\ell.$$

Moreover if j has the p-adic expansion $j = \sum_{i=1}^{n} a_i p^i$ and χ be the map

$$\chi : \{0, \ldots, p-1\} \to \{0, 1\}$$

defined by:

$$\chi(a) := \begin{cases} 1 & \text{if } a \neq 0, \\ 0 & \text{if } a = 0 \end{cases}$$

then

$$\dim((W_j)_G) = \sum_{i=1}^{n} \chi(a_i).$$

Finally

$$\dim(H^1(X, G, T_X)) = \begin{cases} s(n+2) - 3 & \text{if } p > 3 \\ s(n+1) - 3 & \text{if } p = 3 \\ sn - 3 & \text{if } p = 2 \end{cases}$$

16.5.5 Weakly Ramified Covers

For the case of weakly ramified covers B. Köck [40] proved that one can extend the global section of holomorphic polydifferentials $H^0(X, \Omega_X^{\otimes m}(D))$ by considering a suitable G-invariant divisor D such that the Euler characteristic $\chi(G, X, \Omega_X(D))$ lifts to a class in the Grothendieck group of projective $k[G]$-modules. This implies that if $H^1(X, \Omega_X(D)) = 0$ vanishes then $H^0(G, \Omega_X(D))$ is projective.

B. Köck together with the second author in [42] used this idea in order to write the short exact sequence

$$0 \to H^0(X, \Omega_X^{\otimes 2}) \to H^0(X, \Omega_X^{\otimes 2}(D)) \to H^0(X, \Sigma) \to 0,$$

where D is selected so that the G-module $H^0(X, \Omega^{\otimes 2}(D))$ is projective, and Σ is a skyscraper sheaf supported at ramified points. Then, the coinvariant functor is applied and the following long exact homology sequence is obtained:

$$0 \to H_1(G, H^0(X, \Sigma)) \to H^0(X, \Omega_X^{\otimes 2})_G \to H^0(X, \Omega_X^{\otimes 2})(D))_G \to H^0(X, \Sigma)_G \to 0.$$

Using the above exact sequence, they arrived at the dimension formula, where g_Y is the genus of the quotient curve $Y = X/G$:

$$\dim H^0(X, \Omega_X^{\otimes 2})_G = 3g_Y - 2 + \sum_{j=1}^{r} \log_p |G(x_j)| + \begin{cases} 2r & \text{if } p > 3 \\ r & \text{if } p = 2 \text{ or } 3. \end{cases}$$

16.5.6 *Galois Weierstrass Points and Harbater–Katz–Gabber Covers*

In 1986 Ian Morrison and Henry Pinkham [58], connected the $k[G]$-structure of the space $H^0(X, \Omega_X)$ to the theory of Weierstrass semigroups of Galois Weierstrass points for the case of Riemann surfaces. A point P on a compact Riemann surface is called *Galois Weierstrass*, if for a meromorphic function f on X such that $(f)_\infty = dP$, where d is the least pole number in the Weierstrass semigroup at P the induced cover $f : X \to \mathbb{P}^1$ is Galois. Morrison and Pinkham's study was based on the monodromy representation of the Galois group at a ramified point, on the fact that the stabilizer of a point in characteristic zero is cyclic and on character theory of cyclic groups. The character of the associated representation of the group $G := \mathrm{Gal}(f)$ is called a Hurwitz character of G, and the authors were able to classify all such characters.

As a wild replacement of the Galois Weierstrass points we can consider the Harbater–Katz–Gabbercovers. A p-order Harbater-Katz-Gabber cover, which from now on will be called HKG-cover, see [22], is a Galois cover $X \to \mathbb{P}^1$ with Galois group a p-group G which has a unique totally ramified point.

Let G act on the complete local ring $k[[t]]$. The Harbater–Katz–Gabber compactification theorem [22, 39], asserts that there is a HKG-cover $X_{\mathrm{HKG}} \to \mathbb{P}^1$ ramified only at one point P of X with Galois group $G = \mathrm{Gal}(X_{\mathrm{HKG}}/\mathbb{P}^1) = G_0$ such that $G_0(P) = G_0$ and the action of G_0 on the completed local ring $\hat{\mathcal{O}}_{X_{\mathrm{HKG}},P}$ coincides with the original action of G_0 on $\mathcal{O}$. There is a lot of recent interest on HKG-covers see [6, 10].

By considering the Harbater–Katz–Gabber compactification to an action on the local ring $k[[t]]$, we have the advantage to attach global invariants, like genus, p-rank, differentials etc., in the local case. Also finite subgroups of the automorphism group $\mathrm{Aut}k[[t]]$, which is a difficult object to understand (and is a crucial object in understanding the deformation theory of curves with automorphisms, see [4]) become subgroups of $\mathrm{GL}(V)$ for a finite dimensional vector space V.

More precisely, let P be a ramified point and let $G(P)$ be the decomposition group at P. There is a representation:

$$\rho : G(P) \to \mathrm{Aut}(k[[t]]),$$

expressing the action of the decomposition group to the completed local ring at a point. The local deformation functor is defined:

$$D_P : \mathscr{C} \to \text{Sets}, R \mapsto \left\{ \begin{array}{l} \text{lifts } G(P) \to \text{Aut}(R[[t]]) \, of \, \rho \bmod - \\ \text{ulo conjugation with an element} \\ \text{of } \ker(\text{Aut}R[[t]] \to k[[t]]) \end{array} \right\}$$

The representation ρ maps $G(P)$ inside the group of automorphisms of formal powerseries, which is a group hard to understand. The following theorem introduced by the second author in [46] gives us a linear representation instead.

Theorem 16.4 *Let P be a fully ramified point of $X \to X/G_1(P)$. Assume that $g_X \geq 2$, $p \geq 2, 3$. Consider the Weierstrass semigroup at P up to the first pole number m_r not divisible by p:*

$$0 = m_0 < \cdots < m_{r-1} < m_r,$$

and select functions in $k(X)$ $f_0, \ldots, f_r$ with $(f_i)_\infty = m_i P$. Then the natural representation

$$\rho : G_1(P) \to \text{GL}(L(m_r P))$$

is faithful.

This theorem allows us to write in explicit form the action on the formal powerseries ring. Indeed, by Hensel's lemma we can select the uniformizer t such that $f_r = t^{-m}$, $m = m_r$. Then the action is given in closed form:

$$\sigma(t) = t\left(1 + t^m \sum_{v=1}^{r} a_{v,r} f_v\right)^{-1/m}.$$

This allows us to work with a general linear group instead of $\text{Aut}(k[[t]])$ and define a representation functor of linear Galois representations as used in the proof of Fermat's last theorem [56].

16.5.7 Representation Filtration

S. Karanikolopoulos and the second author in [34] defined a filtration similar to the ramification filtration, the representation filtration. More precisely for each $0 \leq i \leq r$, consider the representations:

$$\rho_i : G_1(P) \to \text{GL}(L(m_i P)),$$

which give rise to the decreasing sequence of groups:

$$G_1(P) = \ker \rho_0 \supseteq \ker \rho_1 \supseteq \ker \rho_2 \supseteq \cdots \supseteq \ker \rho_r = \{1\},$$

corresponding to the tower of function fields:

$$F^{G_1(P)} = F^{\ker \rho_0} \subseteq F^{\ker \rho_1} \subseteq \cdots \subseteq F^{\ker \rho_r} = F.$$

Theorem 16.5 *If $X \to X/G$ is a HKG-cover, then the representation and the ramification filtrations coincide.*

Select a function $f_{i_0} \in k(X)$ such that $k(X)^G = k(f_{i_0})$.

$$\operatorname{div}(df_{i_0}^{\otimes m}) = \left(-2mp^{h_0} + m \sum_{i=1}^{n} (b_i - b_{i-1})(p^{h-1} - 1) \right) P,$$

where

$$b_0 = -1,\ p^{h_0} = |G_1(P)|,\ p^{h_i} = |\ker \rho_{c_{i+1}}| = |G_{b_{i+1}}|,\ \text{for } i \geq 1.$$

The following theorems give some information for $k[G]$-module structure of holomorphic polydifferentials for the case of HKG-covers.

Theorem 16.6 *For every pole number μ select a function f_μ such that $(f_\mu)_\infty = \mu P$. The set*

$$\{f_\mu df_{i_0}^{\otimes m} : \deg(f_i) \leq m(2g_X - 2)\}$$

forms a basis for the space of m-holomorphic (poly)differentials of X.

Theorem 16.7 *The module $H^0(X, \Omega_X^{\otimes m})$ is a direct sum of $N = \left\lfloor \frac{m(2g-2)}{p^{h_0}} \right\rfloor$ direct indecomposable summands.*

Corollary: If $|G_1(P)| \geq m(2g-2)$, then $N = 1$. In particular curves with *big-action* (in the sense of M. Matignon-M. Rocher) have one indecomposable summand.

16.6 Integral Representation Theory

Suppose that a relative curve $\mathscr{X} \to \operatorname{Spec} R$ with a fibrewise action of G is given. When R is a principal ideal domain then one can show that the spaces $M_n = H^0(\mathscr{X}, \Omega_{\mathscr{X}}^{\otimes n})$ are free R-modules.

Problem: Describe the module structure of M_n within the theory of *integral representations*. Notice that usually the term *integral representation* is reserved for $\mathbb{Z}[G]$-modules. Our situation is a little bit easier since we work over complete local rings, and we also add the eigenvalues $R = W(k)(\zeta_n)$.

S. Karanikolopoulos and the second author in [36] used the model of Bertin-Mézard [4] based on the work of Sekiguchi, Oort and Suwa theory [77, 78] in order to

study this problem for cyclic groups. More precisely, the generic fibre for the Bertin-Mézard model is a Kummer extenrion defined over the Witt ring $S := W(k)(\zeta_p)$ of k with a p-th root of unity ζ_p adjoined, given by:

$$(X + \lambda^{-1})^p = x^{-m} + \lambda^{-p},$$

where $\lambda = \zeta_p - 1$ such that $\lambda \equiv 0 \mod m_S$. We set $m = pq - l$, $0 < l \leq p - 1$ and $\lambda X + 1 = y/x^q$. The model then becomes

$$y^p = (\lambda^p + x^m)x^l = \lambda^p x^l + x^{qp}.$$

More generally x^q can be replaced by $a(x) = x^q + x_1 x^{q-1} + \cdots + x_q$, where $x_q = 0$ if $l \neq 1$, and consider the Kummer extension

$$(\lambda\xi + a(x))^p = \lambda^p x^l + a(x)^p,$$

where $\xi = Xa(x)$, $y = \lambda\xi + a(x) = a(x)(\lambda X + 1)$. This more general model is given by

$$y^p = \lambda^p x^l + a(x)^p = x^l(\lambda^p + a(x)^p x^{-l}).$$

Let R denote the Oort-Sekiguchi-Suwa factor of the versal deformation ring [4]:

$$R = \begin{cases} W(k)[\zeta_p][x_1, \ldots, x_q] & \text{if } l = 1 \\ W(k)[\zeta_p][x_1, \ldots, x_{q-1}] & \text{if } l > 1 \end{cases}$$

The Bertin-Mézard model is a relative curve $\mathscr{X} \to \mathrm{Spec}R$, where the horizontal branch locus is given in Fig. 16.2.

Using the theory of Boseck for the generic fibre of R we see that the set of differentials of the form

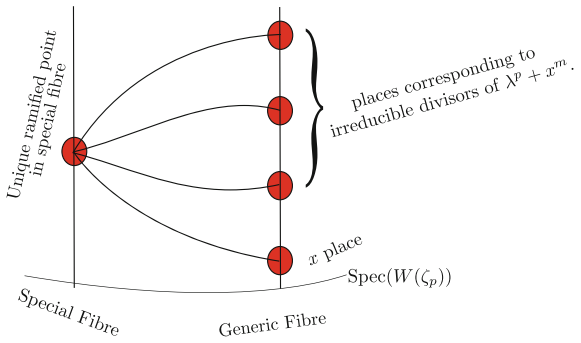


Fig. 16.2 Splitting the branch locus

$$x^N a(x)^a \frac{(\lambda X+1)^a}{a(x)^{p-1}(\lambda X+1)^{p-1}} dx, \tag{16.18}$$

where

$$0 \le a < p-1 \text{ and } l - \left\lceil \frac{(1+a)l}{p} \right\rceil \le N \le (p-1-a)q - 2, \tag{16.19}$$

forms a basis of holomorphic differentials. This base is not suitable for taking the reduction modulo the maximal ideal of the ring $S = W(k)[\zeta]$ since in the reduction $\lambda = 0$. The idea is to change the basis of the generic fibre so that no λ appears in the numerator of the differentials. Then we use the special fibre Boseck basis to show that the reductions are holomorphic, therefore the relative differentials are indeed holomorphic over Spec$R$.

In this way, we arrive at:

Theorem 16.8 *Let σ be an automorphism of $\mathscr{X}$ of order $p \neq 2$ and conductor m with $m = pq - l$, $1 \le q$, $1 \le l \le p-1$. Consider the modules*

$$V_{a_0,a_1} :=_S \langle (\lambda X+1)^{a_0} X^i : 0 \le i < a_1 \rangle.$$

which are indecomposable $S[G]$-modules and define $V_a := V_{1-p,a}$.

The free R-module $H^0(\mathscr{X}, \Omega_{\mathscr{X}})$ has the following $R[G]$ structure:

$$H^0(\mathscr{X}, \Omega_{\mathscr{X}}) = \bigoplus_{\nu=0}^{p-2} V_\nu^{\delta_\nu},$$

where

$$\delta_\nu = \begin{cases} q + \left\lceil \frac{(\nu+1)l}{p} \right\rceil - \left\lceil \frac{(2+\nu)l}{p} \right\rceil & \text{if } \nu \le p-3, \\ q-1 & \text{if } \nu = p-2. \end{cases}$$

Acknowledgements The authors were supported by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) Research Funding Program: THALES: Reinforcement of the interdisciplinary and/or interinstitutional research and innovation.
The authors would like to thank the referees for their suggestions on the improvement of the text. We would like to also thank Professor P. Roquette for introducing us to the fascinating subject of automorphisms of curves 25 years ago.

References

1. Adler, A.: The Mathieu group M_{11} and the modular curve $X(11)$. Proc. Lond. Math. Soc. (3), **74**(1), 1–28 (1997)
2. Bending, P., Camina, A., Guralnick, R.: Automorphisms of the modular curve. In: Progress in Galois Theory, Dev. Math. vol. 12, pp. 25–37. Springer, New York (2005)

3. Benson, D.: Representations of Elementary Abelian p-Groups and Vector Bundles (2015)
4. Bertin, J., Mézard, A.: Déformations formelles des revêtements sauvagement ramifiés de courbes algébriques. Invent. Math. **141**(1), 195–238 (2000)
5. Bleher, F., Chinburg, T., Kontogeorgis, A.: On holomorphic differentials of curves (In preparation)
6. Bleher, F.M., Chinburg, T., Poonen, B., Symonds, P.: Automorphisms of Harbater-Katz-Gabber curves. Mathematische Annalen (to appear)
7. Borne, N.: Cohomology of G-sheaves in positive characteristic. Adv. Math. **201**(2), 454–515 (2006)
8. Boseck, H.: Zur Theorie der Weierstrasspunkte. Math. Nachr. **19**, 29–63 (1958)
9. Chevalley, C., Weil, A., Hecke, E.: Über das verhalten der integrale 1. gattung bei automorphismen des funktionenkörpers. Abh. Math. Semin. Univ. Hambg. (1934)
10. Chinburg, T., Guralnick, R., Harbater, D.: Oort groups and lifting problems. Compos. Math. **144**(04), 849–866 (2008)
11. Cornelissen, G., Kato, F.: Equivariant deformation of Mumford curves and of ordinary curves in positive characteristic. Duke Math. J. **116**(3), 431–470 (2003)
12. Cornelissen, G., Kato, F., Kontogeorgis, A.: Discontinuous groups in positive characteristic and automorphisms of Mumford curves. Math. Ann. **320**(1), 55–85 (2001)
13. Daniel, G., Salvador, V.: Mathematics: Theory & Applications. Topics in the Theory of Algebraic Function Fields. Birkhäuser, Boston (2006)
14. Deligne, P., Mumford, D.: The irreducibility of the space of curves of given genus. Inst. Hautes Études Sci. Publ. Math. **36**, 75–109 (1969)
15. Farkas, H.M., Kra, I.: Graduate Texts in Mathematics. Riemann surfaces, vol. 71. Springer, New York (1980)
16. García, A.: On Weierstrass points on Artin-Schreier extensions of $k(x)$. Math. Nachr. **144**, 233–239 (1989)
17. Gerritzen, L.: On non-archimedean representations of abelian varieties. Math. Ann. **196**(4), 323–346 (1972)
18. Gerritzen, L., van der Put, M.: Lecture Notes in Mathematics. Schottky groups and Mumford curves, vol. 817. Springer, Berlin (1980)
19. Giulietti, M., Korchmáros, G.: Algebraic curves with a large non-tame automorphism group fixing no point. Trans. Amer. Math. Soc. **362**(11), 5983–5983 (2010)
20. Giulietti, M., Korchmáros, G.: Nakajima's remark on Henn's proof. Electronic Notes Discret. Math. **40**, 135–138 (2013)
21. Grothendieck, A.: Sur quelques points d'algèbre homologique. Tôhoku Math. J. **2**(9), 119–221 (1957)
22. Harbater, D.: Moduli of p-covers of curves. Comm. Algebra **8**(12), 1095–1122 (1980)
23. Harris, J., Morrison, I.: Graduate Texts in Mathematics. Moduli of curves, vol. 187. Springer, New York (1998)
24. Hartshorne, R.: Graduate Texts in Mathematics. Algebraic geometry, vol. 52. Springer, New York (1977)
25. Hecke, E.: Über ein Fundamentalproblem aus der Theorie der elliptischen Modulfunktionen. Abh. Math. Sem. Univ. Hamburg **6**(1), 235–257 (1928)
26. Heinrich Matzat, B.: Ein Vortrag Ueber Weierstasspunkte (1975)
27. Henn, H.-W.: Funktionenkörper mit grosser Automorphismengruppe. J. Reine Angew. Math. **302**, 96–115 (1978)
28. Herrlich, F.: Die Ordnung der Automorphismengruppe einer p-adischen Schottkykurve. Math. Ann. **246**(2), 125–130 (1979/1980)
29. Hirschfeld, J.W.P., Korchmáros, G., Torres, F.: Algebraic Curves over a Finite Field. Princeton Series in Applied Mathematics. Princeton University Press, Princeton (2008)
30. Hochschild, G., Serre, J.-P.: Cohomology of group extensions. Trans. Amer. Math. Soc. **74**, 110–134 (1953)
31. Hurwitz, A.: Ueber algebraische gebilde mit eindeutigen transformationen in sich. Mathematische Annalen **41**(3), 403–442 (1892)

32. Kani, E.: The Galois-module structure of the space of holomorphic differentials of a curve. J. Reine Angew. Math. **367**, 187–206 (1986)
33. Karanikolopoulos, S.: On holomorphic polydifferentials in positive characteristic. Math. Nachr. **285**(7), 852–877 (2012)
34. Karanikolopoulos, S., Kontogeorgis, A.: Automorphisms of Curves and Weierstrass semigroups for Harbater-Katz-Gabber covers. Submitted
35. Karanikolopoulos, S., Kontogeorgis, A.: Representation of cyclic groups in positive characteristic and Weierstrass semigroups. J. Number Theory **133**(1), 158–175 (2013)
36. Karanikolopoulos, S., Kontogeorgis, A.: Integral representations of cyclic groups acting on relative holomorphic differentials of deformations of curves with automorphisms. Proc. Amer. Math. Soc. **142**(7), 2369–2383 (2014)
37. Karrass, A., Pietrowski, A., Solitar, D.: Finite and infinite cyclic extensions of free groups. J. Austral. Math. Soc. **16**, 458–466 (1973). Collection of articles dedicated to the memory of Hanna Neumann, IV
38. Kato, F., Kontogeorgis, A.: On the Galois-module structure of polydifferentials of Artin-Schreier-Mumford curves, modular and integral representation theory (In preparation)
39. Katz, N.M.: Local-to-global extensions of representations of fundamental groups. Ann. Inst. Fourier (Grenoble) **36**(4), 69–106 (1986)
40. Köck, B.: Galois structure of Zariski cohomology for weakly ramified covers of curves. Amer. J. Math. **126**(5), 1085–1107 (2004)
41. Köck, B., Tait, J.: Faithfulness of actions on Riemann-Roch spaces. Canad. J. Math. **67**(4), 848–869 (2015)
42. Koeck, B., Kontogeorgis, A.: Quadratic differentials and equivariant deformation theory of curves. Annales de L'institut Fourier **62**(3), 1015–1043 (2012)
43. Kontogeorgis, A.: Automorphisms of Fermat-like varieties. Manuscripta Math. **107**(2), 187–205 (2002)
44. Kontogeorgis, A.: On the tangent space of the deformation functor of curves with automorphisms. Algebra Number Theory **1**(2), 119–161 (2007)
45. Kontogeorgis, A.: Polydifferentials and the deformation functor of curves with automorphisms. J. Pure Appl. Algebra **210**(2), 551–558 (2007)
46. Kontogeorgis, A.: The ramification sequence for a fixed point of an automorphism of a curve and the Weierstrass gap sequence. Math. Z. **259**(3), 471–479 (2008)
47. Kontogeorgis, A., Rotger, V.: On abelian automorphism groups of Mumford curves. Bull. Lond. Math. Soc. **40**(3), 353–362 (2008)
48. Kontogeorgis, A., Yang, Y.: Automorphisms of hyperelliptic modular curves $X_0(N)$ in positive characteristic. LMS J. Comput. Math. **13**, 144–163 (2010)
49. Lehner, J.: Discontinuous Groups and Automorphic Functions. Mathematical Surveys and Monographs. American Mathematical Society, Providence (1964)
50. Lehr, C., Matignon, M.: Automorphism groups for p-cyclic covers of the affine line. Compos. Math. **141**(5), 1213–1237 (2005)
51. Leopoldt, H.-W.: Über die Automorphismengruppe des Fermatkörpers. J. Number Theory **56**(2), 256–282 (1996)
52. Madden, D.J., Valentini, R.C.: The group of automorphisms of algebraic function fields. J. Reine Angew. Math. **343**, 162–168 (1983)
53. Manin, Yu., Drinfeld, V.: Periods of p-adic Schottky groups. J. Reine Angew. Math. **262/263**, 239–247 (1973). Collection of articles dedicated to Helmut Hasse on his seventy-fifth birthday
54. Marques, S., Ward, K.: Holomorphic differentials of solvable galois towers of curves over a perfect field. Preprint (2015). arXiv:1507.07023v2
55. Matignon, M., Rocher, M.: On smooth curves endowed with a large automorphism p-group in characteristic $p > 0$. Algebra Number Theory **2**, 887–926 (2008)
56. Mazur, B.: Deformation theory of Galois representations (in Modular forms and Fermat's last theorem), pp. xx+582 (1997). Papers from the Instructional Conference on Number Theory and Arithmetic Geometry held at Boston University, Boston, MA, 9–18 Aug 1995

57. Monsky, P.: The automorphism groups of algebraic curves. Ph.D. Thesis University of Chicago (1962)
58. Morrison, I., Pinkham, H.: Galois Weierstrass points and Hurwitz characters. Ann. Math. (2). **124**(3), 591–625 (1986)
59. Mumford, D.: An analytic construction of degenerating curves over complete local rings. Compositio Math. **24**, 129–174 (1972)
60. Nakajima, S.: Galois module structure of cohomology groups for tamely ramified coverings of algebraic varieties. J. Number Theory **22**(1), 115–123 (1986)
61. Nakajima, S.: On abelian automorphism groups of algebraic curves. J. Lond. Math. Soc. (2). **36**(1), 23–32 (1987)
62. Nakajima, S.: p-ranks and automorphism groups of algebraic curves. Trans. Amer. Math. Soc. **303**(2), 595–607 (1987)
63. Obus, A.: The (local) lifting problem for curves. 8 May 2011. arXiv:1105.1530
64. Obus, A., Wewers, S.: Cyclic extensions and the local lifting problem. Ann. Math. (2). **180**(1), 233–284 (2014)
65. Poonen, B.: Varieties without extra automorphisms I: Curves. Math. Res. Lett. **7**, 67–76 (2000)
66. Pop, F.: The Oort conjecture on lifting covers of curves. Ann. Math. (2). **180**(1), 285–322 (2014)
67. Popp, H.: The singularities of the moduli schemes of curves. J. Number Theory. **1**(1), 90–107 (1969)
68. Rajan, C.S.: Automorphisms of $X(11)$ over characteristic 3 and the Mathieu group M_{11}. J. Ramanujan Math. Soc. **13**(1), 63–72 (1998)
69. Ritzenthaler, C.: Automorphismes des courbes modulaires $X(n)$ en caractéristique p. Manuscripta Math. **109**(1), 49–62 (2002)
70. Rocher, M.: Large p-groups actions with $|g|/g^2 > 4/(p^2-1)^2$. 22 Apr 2008. arXiv:0804.3494
71. Rocher, M.: Large p-group actions with a p-elementary abelian derived group. J. Algebra **321**(2), 704–740 (2009)
72. Roquette, P.: The Riemann hypothesis in characteristic p, its origin and development I. The formation of the zeta-functions of Artin and of F.K. Schmidt. Hamburger Beitraege zur Gesichte der Mathematik, Mitt. Math. Ges. Hamburg. **21**, 79–157 (2002)
73. Rzedowski-Calderón, M., Villa-Salvador, G., Madan, M.L.: Galois module structure of holomorphic differentials in characteristic p. Arch. Math. (Basel) **66**(2), 150–156 (1996)
74. Schlessinger, M.: Functors of Artin rings. Trans. Amer. Math. Soc. **130**, 208–222 (1968)
75. Schmid, H.L.: Über die automorphismen eines algebraischen funktionenkörpers von primzahlcharakteristik. J. Reine Angew. Math. **179**, 5–15 (1938)
76. Schneider, P.: Rigid-analytic L-transforms. Number theory, Noordwijkerhout 1983 (Noordwijkerhout, 1983). Lecture Notes in Mathematics, vol. 1068, pp. 216–230. Springer, Berlin (1984)
77. Sekiguchi, T., Suwa, N.: Théories de Kummer-Artin-Schreier-Witt. C. R. Acad. Sci. Paris Sér. I Math. **319**(2), 105–110 (1994)
78. Sekiguchi, T., Suwa, N.: Théorie de Kummer-Artin-Schreier et applications. J. Théor. Nombres Bordeaux. **7**(1), 177–189 (1995). Les Dix-huitièmes Journées Arithmétiques (Bordeaux, 1993)
79. Serre, J.-P.: Sur la topologie des variétés algébriques en caractéristique p. In: Symposium Internacional de Topología algebraica International Symposium on Algebraic Topology, pp. 24–53. Universidad Nacional Autónoma de México and UNESCO, Mexico City (1958)
80. Stichtenoth, H.: Über die Automorphismengruppe eines algebraischen Funktionenkörpers von Primzahlcharakteristik. I. Eine Abschätzung der Ordnung der Automorphismengruppe. Arch. Math. (Basel) **24**, 527–544 (1973)
81. Stichtenoth, H.: Über die Automorphismengruppe eines algebraischen Funktionenkörpers von Primzahlcharakteristik. II. Ein spezieller Typ von Funktionenkörpern. Arch. Math. (Basel) **24**, 615–631 (1973)
82. Stichtenoth, H.: Graduate Texts in Mathematics. Algebraic Function Fields and Codes, vol. 254, 2nd edn. Springer, Berlin (2009)

83. Teitelbaum, J.T.: Values of p-adic L-functions and a p-adic Poisson kernel. Invent. Math. **101**(2), 395–410 (1990)
84. Valentini, R.C., Madan, M.L.: Automorphisms and holomorphic differentials in characteristic p. J. Number Theory **13**(1), 106–115 (1981)
85. Ward, K.: Explicit galois representations of automorphisms on holomorphic differentials in characteristic p. Finite Fields Appl. **44**, 34–55 (2017)

Chapter 17
Building and Integrating Semantic Theories over Institutions

Nicola Angius, Maria Dimarogkona and Petros Stefaneas

Abstract This paper constitutes a first attempt at constructing semantic theories over institutions and examining the logical relations holding between different such theories. Our results show that this approach can be very useful for theoretical computer science (and may also contribute to the current philosophical debate regarding the semantic and the syntactic presentation of scientific theories). First we provide a definition of semantic theories in the institution theory framework - in terms of a set of models satisfying a given set of sentences - using the language-independent satisfaction relation characterizing institutions (Definition 17.3). Then we give a proof of the logical equivalence holding between the syntactic and the semantic presentation of a theory, based on the Galois connection holding between sentences and models (Theorem 17.1). We also show how to integrate and combine semantic theories using colimits (Theorem 17.2). Finally we establish when the output of a model-based software verification method applied to a semantic theory over an institution also holds for a semantic theory defined over a different institution (Theorem 17.3).

Keywords Abstract model theory · Semantic view of theories
Institution theory · Algebraic specifications

17.1 Introduction

Abstract model theory studies the general properties of extensions of first-order logic and their models, focusing on the satisfaction relation holding between sentences

N. Angius (✉)
Department of History, Human Sciences, and Education, University of Sassari, Sassari, Italy
e-mail: nangius@uniss.it

M. Dimarogkona · P. Stefaneas
Department of Applied Mathematics, National Technical University of Athens, Zografou Campus, 15780 Athens, Greece
e-mail: mariadim@centrall.ntua.gr

P. Stefaneas
e-mail: petros@math.ntua.gr

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_17

and models. The various properties of different logics are usually explored by the construction of a *theory* over a logical system. In standard abstract model theory [2], theories are conceived as syntactic entities introduced by some set of statements. In Institutions Theory [13], a Σ-theory over an institution is defined by a signature Σ and a set of Σ-sentences that are closed under semantic entailment. Morphisms of the form $\Sigma \rightarrow \Sigma'$ provide language translations for theories over institutions. Σ-models are those structures that satisfy a given set of Σ-sentences, and every Σ-theory is associated with a collection of Σ-models satisfying Σ-statements in the Σ-theory. The duality between sets of Σ-sentences and sets of Σ-models can be defined in terms of *Galois connections*.

In this paper we use the duality characterizing sentences and models (Galois connection) to define a theory over an institution as a semantic entity. Drawing inspiration from the work of [20, 21, 23], semantic theories are here defined by a signature Σ and a set of Σ-models. Our first main result is a proof of the logical equivalence holding between the syntactic and the semantic presentation of a theory over the same institution (Theorem 17.1 in Sect. 17.2). Theorem 17.1 provides a categorical framework to the discussion about the logical relation holding between syntactic and semantic theories [10, 14, 15].

The theory of institutions has been successfully applied, among other areas of theoretical computer science, to the field of algebraic specification [8]. In particular, theories over institutions are used to provide program specifications based on sets of sentences describing allowed program behaviors [19]. One of the main results in [13] was that of constructing, using institutions, "larger"specifications from "smaller" specifications. More specifically, given two institutions I and I' whose signatures can be integrated or combined, syntactic theories over I can be integrated and combined via colimits with syntactic theories over I', thereby denoting "larger"specifications. Building software specifications from small specifications is very helpful when developing modular systems, such as those involving object oriented programs. In such cases, system specifications can be obtained by combining and integrating class specifications [9].

However, many program specifications used today in software verification are defined semantically in terms of state transition systems, including Kripke Structures or Büchi Automata. This is often the case when model checking is performed to evaluate program correctness [3]. By defining a theory over an institution semantically we acquire a means by which to build up extended and combined state transition systems using elementary semantic specifications (Theorem 17.2 in Sect. 17.3).

Finally we examine the logical relations holding between semantic theories introduced using different institutions. By establishing proper morphisms between the two different institutions, Goguen and Burstall have managed to determine when a theorem prover successfully applied on a theory over the first institution can be successfully applied on theories over the second institution [13]. Correspondingly, Theorem 17.3 (Sect. 17.4) establishes when the output of a model-checker applied to a semantic theory over an institution also holds for a semantic theory defined over a different institution.

17.2 Syntactic and Semantic Definition of Theories and Theory Morphisms

Institutions define logical systems abstracting away from any actual linguistic formulation, focusing on the abstract satisfaction relation holding between sentences and models. This is achieved using category theory to introduce a category **Sign** of signatures, a set Set of sentences, a category $\mathbf{Cat}^{op}$ of models, and a satisfaction relation $\models_\Sigma$. The main intuition underpinning institution theory, and well expressed by Tarski's [22] motto 'truth is invariant under the change of notation', is that given any two signatures Σ and Σ' in **Sign**, a Σ-sentence f is satisfied by a Σ-model m iff the Σ' translation f' of f is satisfied by the Σ' translation m' of m. Institutions are formally defined as follows:

Definition 17.1 (*Institution*) An institution $I = (\mathbf{Sign}, Sen, \mathbf{Mod}, \models_\Sigma)$ is defined by:

1. a category **Sign** having signatures as objects, and signature morphisms as arrows;
2. a functor $Sen : \mathbf{Sign} \to Set$ mapping each signature Σ in **Sign** to a set of Σ-sentences;
3. a functor $Mod : \mathbf{Sign} \to \mathbf{Cat}^{op}$ assigning to each Σ in **Sign** a category whose objects are Σ-models and whose morphisms are Σ-model morphisms;
4. a relation $\models_\Sigma \subseteq |Mod(\Sigma)| \times Sen(\Sigma)$ for each $\Sigma \in |\mathbf{Sign}|$, called Σ-satisfaction, such that for each signature morphism $\phi : \Sigma \to \Sigma'$, the **satisfaction condition** $m' \models_{\Sigma'} Sen(\phi)(e)$ iff $Mod(\phi)(m') \models_\Sigma e$ holds for each $m' \in \left|Mod(\Sigma')\right|$ and each $e \in Sen(\Sigma)$. □

According to Goguen and Burstall [13, p. 11], a specification provides a *mathematical theory* of a program behavior in terms of the set of sentences describing this behavior. Given an institution I, mathematical theories of this sort are introduced by a signature Σ and a closed set of Σ-sentences.

Definition 17.2 (*Syntactic (standard) theory over an institution*)

1. A **syntactic** Σ-theory **presentation** is a pair $\langle \Sigma, E \rangle$ where Σ is a signature and E is a set of Σ-sentences.
2. A Σ-model A satisfies a syntactic theory presentation $\langle \Sigma, E \rangle$ if A satisfies each sentence in E, in which case we write $A \models E$.
3. If E is a set of Σ-sentences, let E^* be the set[1] of all Σ-models that satisfy each sentence in E.
4. If M is a set of Σ-models, let M^* be the set of all Σ-sentences that are satisfied by each model in M; M^* also denotes $\langle \Sigma, M^* \rangle$.
5. By the closure of a set E of Σ-sentences we mean the set E^{**}, written $E^\bullet$.[2]

[1] In the terminology of axiomatic set theory, E^* is a class of Σ-models, rather than a set [12]. For the sake of simplicity, this paper will nonetheless consider E^* as a set.

[2] $E^\bullet$ is a model-theoretic closure of the set of Σ-sentences. For some institutions, including equational logic, a corresponding proof-theoretic notion can be given, insofar as there is a complete set of inference rules [13].

6. A set E of Σ-sentences is closed iff $E = E^{\bullet}$. Then a Σ-theory is a syntactic theory presentation $\langle \Sigma, E \rangle$ such that E is closed.
7. The **syntactic theory** presented by the syntactic presentation $\langle \Sigma, E \rangle$ is $\langle \Sigma, E^{\bullet} \rangle$. □

Now let us define a theory semantically, that is, as the set of models representing a program's potential behaviors. It should be noted how semantic Σ-theories here defined are still mathematical theories of programs, insofar as the Σ-models introducing them prescribe program's computations, and they do not represent observed executions. However, the following definition of a semantic theory using institution can be used to build empirical theories of computational systems as well [1].

Definition 17.3 (*Semantic theory over an institution*)

1. A **semantic** Σ-theory **presentation** is a pair $\langle \Sigma, M \rangle$ where Σ is a signature and M is a set of Σ-models.
2. A Σ-model A satisfies a set of sentences E if A satisfies each sentence in E, in which case we write $A \models E$.
3. If M is a set of Σ-models, let M^{*} be the set of all Σ-sentences that are satisfied by each model in M.
4. If E is a set of Σ-sentences, let E^{*} be the set of all Σ-models that satisfy each sentence in E.
5. By the closure of a set M of Σ-models we mean the set M^{**}, written M°.[3]
6. A set M of Σ-sentences is closed iff $M = M^{\circ}$. Then a Σ-theory is a semantic theory presentation $\langle \Sigma, M \rangle$ such that M is closed.
7. The **semantic theory** presented by the semantic presentation $\langle \Sigma, M \rangle$ is $\langle \Sigma, M^{\circ} \rangle$. □

Definitions 17.2 and 17.3 above show the particular duality existing between a set of Σ-sentences and the closed set of Σ-models satisfying those sentences, and between a set of Σ-models and the closed set of Σ-sentences satisfied by those models. Such duality constitutes a *Galois connection*, reflecting 'the view that syntax and semantics are adjoint' [16]. In institution theory the Galois connection is formulated as follows:

Proposition 17.1 (Galois connection) *The two functions* **:sets of* Σ*-sentences* $\rightarrow$ *sets of* Σ*-models,* **:sets of* Σ*-models* $\rightarrow$ *sets of* Σ*-sentences, given in Definitions 17.2 and 17.3, form what is known as a Galois connection, in that they satisfy the following properties, for any sets* E, E' *of* Σ*-sentences and sets* M, M' *of* Σ*-models:*

1. $E \subseteq E'$ *implies* $E'^{*} \subseteq E^{*}$
2. $M \subseteq M'$ *implies* $M'^{*} \subseteq M^{*}$
3. $E \subseteq E^{**}$
4. $M \subseteq M^{**}$

[3] In the equational institution, closed sets of models are usually called *varieties*. A set of models is called closed iff its objects are all the models of some set of sentences.

These imply the following properties:

5. $E^* = E^{***}$
6. $M^* = M^{***}$
7. *There is a dual (i.e. inclusion reversing) isomorphism between the closed set of sentences and the closed set of models.*
8. $(\cup_n, E_n)^* = \cap E_n^*$
9. $\phi(E^*) = (\phi E)^*$, *for* $\phi : \Sigma \to \Sigma'$ *a signature morphism.* □

Proof The proofs of 1 and 2 are straightforward. We now prove 3.
Assume that $E \subseteq Sen(\Sigma)$. Then $E^* = \{m \mid (\forall e \in E)\, m \models e\}$.
It follows that $E^{**} = \left\{e' \mid (\forall m \in E^*)\, m \models e'\right\} = \left\{e' \mid (\forall m)\, m \in E^* \Rightarrow m \models e'\right\} = \left\{e' \mid (\forall m)\, [(\forall e \in E)\, m \models e] \Rightarrow m \models e'\right\}$.
If $e' \in E$, and if $(\forall e \in E)\, m \models e$, then $m \models e'$. It follows that $\left\{e' \mid e' \in E\right\} = E$.

In a similar way we prove 4, and 5–9 are familiar from lattice theory [17]. In fact 8 and 9 follow easily from 7. □

Corollary 17.1 below shows that any syntactically presented theory T determines an equivalent semantically presented theory T^* and every semantically presented theory V determines an equivalent syntactically presented theory V^*; in other words, the syntactic and semantic presentation of a theory, as defined above, are essentially equivalent.

Corollary 17.1 (from Galois connection) *For every syntactic* Σ*-theory* $T = \langle \Sigma, E^\bullet \rangle$ *and every semantic* Σ*-theory* $V = \langle \Sigma, M^\circ \rangle$*:*

1. T *determines the semantic theory* $T^* = \langle \Sigma, E^* \rangle$*, consisting of all the* Σ*-models that satisfy all the sentences in* T*.* ($E^* = E^{***}$*: Galois connection 5);*
2. V *determines the syntactic theory* $V^* = \langle \Sigma, M^* \rangle$*, consisting of those* Σ*-sentences satisfied by all models in* V*.* ($M^* = M^{***}$*: Galois connection 6).* □

In the theory of institutions, (syntactic) theories give rise to the category **Th** of (syntactic) theories whose morphisms from a syntactic theory $T = \langle \Sigma, E^\bullet \rangle$ to another syntactic theory $T' = \langle \Sigma', E'^\bullet \rangle$ are signature morphisms of the form $\phi : \Sigma \to \Sigma'$ such that, for any sentence $e \in E^\bullet$, $\phi(e) \in E'^\bullet$. In a similar way, semantic theories give rise to the category **Vth** of semantic theories.

Definition 17.4 (*category of semantic theories*) If V and V' are semantic theories, say $\langle \Sigma, M^\circ \rangle$ and $\langle \Sigma', M'^\circ \rangle$, then a semantic theory morphism from V to V' is a signature morphism $G : \Sigma \to \Sigma'$ such that $G(m)$ is in M'° for each m in M°; we will write $G : V \to V'$. The **category of semantic theories** has semantic theories as objects and semantic theory morphisms as morphisms, with their composition and identities defined as for signature morphisms; let us denote it **Vth** (it is easy to see that this is a category). □

In the context of a current philosophical debate regarding the syntactic and semantic presentation of scientific theories, Halvorson [14] advanced the thesis that semantic presentations differ from the corresponding syntactic ones in that two equivalent

syntactic theories may result to be non-equivalent when model-theoretically presented and, *vice versa*, two equivalent semantic theories may turn to be non-equivalent when syntactically defined. The notion of model-theoretical equivalence that allowed supporting such claims involves isomorphisms between non-interpreted structures [10]. Only when models are interpreted in a fixed language, isomorphisms enable one to establish equivalence relations between syntactic and corresponding semantic theories [15]. However, according to some construal of the notion of semantic theories (such as the one in [24]), the latter are to be identified with a set of language independent structures and the standard notion of isomorphism in model theory cannot be applied to prove the logic equivalence of syntactic and semantic presentations of a given theory [15]. An institutions-based formalisation of theories allows for a straightforward definition of both syntactic and semantic-theoretical equivalence in a completely language independent context.

Definition 17.5 (*equivalence of syntactic theories*) Given an institution I, two syntactic theories $T = \langle \Sigma, E^\bullet \rangle$ and $T' = \langle \Sigma', E'^\bullet \rangle$ over I are said to be equivalent iff there exists $F : T \to T'$ such that F is an isomorphism.

A morphism $F : T \to T'$ in category **Th** is an isomorphism if there exists another morphism $H : T' \to T$ such that $H \circ F = 1_T$ and $F \circ H = 1_{T'}$, where 1_T and $1_{T'}$ are the identity morphisms of T and T' respectively. □

Definition 17.6 (*equivalence of semantic theories*) Given an institution I, two semantic theories $V = \langle \Sigma, M^\circ \rangle$ and $V' = \langle \Sigma', M'^\circ \rangle$ over I are said to be equivalent iff there exists $G : V \to V'$ such that G is an isomorphism.

A morphism $G : V \to V'$ in category **Vth** is an isomorphism if there exists another morphism $Z : V' \to V$ such that $Z \circ G = 1_V$ and $G \circ Z = 1_{V'}$, where 1_V and $1_{V'}$ are the identity morphisms of T and T' respectively. □

Theorem 17.1 *Given an institution I, two syntactic theories $T = \langle \Sigma, E^\bullet \rangle$ and $T' = \langle \Sigma', E'^\bullet \rangle$ over I are equivalent iff the semantic theories $T^* = \langle \Sigma, E^* \rangle$, and $T'^* = \langle \Sigma', E'^* \rangle$ that they define are equivalent.*

Proof "⇒ Suppose that $T \cong T'$. From corollary of Galois connection we know that T and T' determine the equivalent semantic theories T^*,T'^*, that is, $T \cong T^*$ and $T' \cong T'^*$. Because equivalence relations are symmetric we have that $T^* \cong T$. Thus $T^* \cong T \cong T' \cong T'^*$ from which we get $T^* \cong T'^*$ (equivalence relations are transitive)."

"⇐" Suppose that the semantic theories T^*,T'^* determined by T, T' are equivalent, that is $T^* \cong T'^*$. We know (from corollary of Galois connection) that $T \cong T^*$ and $T' \cong T'^*$. Because equivalence relations are symmetric we have that $T'^* \cong T'$. Thus $T \cong T^* \cong T'^* \cong T'$, from which we get $T \cong T'$ (equivalence relations are transitive). □

17.3 Putting Semantic Theories Together

Theories over an institution can be used to describe a program's behavior and thereby to provide system specifications representing the allowed computations. One of the main benefits coming from the application of category theory to software specification languages is that of exploiting common categorical constructs, chiefly colimits, to construct overall program specifications from modular, smaller, specifications [11]. Users requirements for complex systems created the need for specifications to become as modular as possible, built from small, understandable and re-usable pieces. This not only facilitates understanding and writing specifications, but also proving theorems about them, as well as proving that a given program actually satisfies its specification.

Institutions allowed to considerably decrease the complexity of building software specifications by obtaining the required modularity using parameterized abstractions. The original approach developed in [4, 5] was based on a syntactic specification language called *Clear*. Today, many specifications are model-theoretically conceived, especially when dealing with model-based verification methods such as model-checking [6]. In model-checking, system specifications are mostly provided in terms of Kripke Structures, i.e. labelled state transition systems depicting the allowed computations. Model-checking and many other formal methods are largely accomplished modularly by formally verifying each program's module in isolation [18]. By formalising the concept of a semantic theory using institutions we get a means by which to compose several model specifications into larger model specifications. This may be managed by considering proper theory-morphisms among different semantic theories, each providing a module specification [1].

Definition 17.7 (*diagrams and colimits in Vth*)

- A *diagram* $D : G \to \mathbf{Vth}$ in the category **Vth** of semantic theories is a *graph* $G = (N, E, l, m)$ defined by a set of nodes N, a set of edges $E \subseteq N \times N$ between nodes, a function $l : N \to obj(\mathbf{Vth})$ labelling each node $n \in N$ by an object D_n of **Vth**, that is, a semantic theory $D_n = \langle \Sigma_n, M_n^\circ \rangle$, and a function $m : E \to morph(\mathbf{Vth})$ labelling each edge $e \in E$ from node n to node n' by a theory morphism $D(e)$ in **Vth** from D_n to $D_{n'}$.
- A *cone* α in **Vth** over the diagram D consists of an object A of **Vth**, that is, a semantic theory $A = \langle \Sigma_A, M_A^\circ \rangle$, and a family of morphisms $a_n : D_n \leftarrow A$, one for each $n \in N$ such that for each edge $e : n \to n'$, the diagram in Fig. 17.1 commutes in **Vth**. We call diagram D base of the cone α, A its apex, graph G its shape, and we write $\alpha : D \Rightarrow A$.
- One can consider now the category of cones **Cone**$(D, \mathbf{Vth})$ over D in **Vth** whose objects are cones with base D and whose morphisms are *cone morphisms*. Given two cones α and β with base D and apexes $A = \langle \Sigma_A, M_A^\circ \rangle$ and $B = \langle \Sigma_B, M_B^\circ \rangle$, a morpshism of cones $\alpha \to \beta$ is a theory morphism $f : A \to B$ in **Vth** such that for each node in G the diagram of Fig. 17.2 commutes in **Vth**.
- Finally, a *colimit* of D in **Vth** is an initial object in **Cone**$(D, \mathbf{Vth})$. □

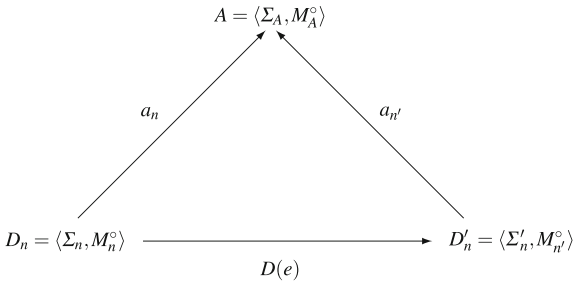


Fig. 17.1 A cone α over a diagram D in the category of semantic theories **Vth**

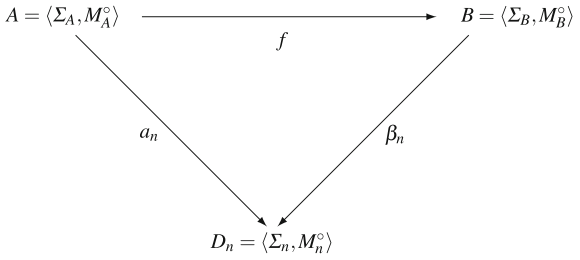


Fig. 17.2 A diagram showing a morphism of cones $\alpha \to \beta$

Colimits in the category of semantic theories **Vth** provide larger specifications resulting from the combination of starting module-semantic theories. Consistently with the analysis of [13], the existence of such colimits may be ascertained by verifying that colimits are given in the corresponding category of signatures **Sign**. Informally, if there is a theory morphism in **Vth** for every signature morphism in **Sign**, we assume that colimits in **Sign** reflect colimits in **Vth**. To prove this formally, let us first introduce the forgetful functor $Sign : \mathbf{Vth} \to \mathbf{Sign}$ defined by mappings $V = \langle \Sigma, M^\circ \rangle \to \Sigma$ from semantic theories to signatures, and $\psi \to \psi$ from theory morphisms to signature morphisms. Then:

Theorem 17.2 *The forgetful functor* $Sign : \mathbf{Vth} \to \mathbf{Sign}$ *reflects colimits.*

Proof Suppose that $D : G \to \mathbf{Vth}$ is a diagram in **Vth**, say with $D_n = \langle \Sigma_n, M_n \rangle$ for $n \in |G|$. Let $D' = Sign \circ D : G \to \mathbf{Sign}$ be the corresponding diagram in **Sign**, in which $D'_n = \Sigma_n$.

Now let $\alpha' : D' \Rightarrow \Sigma$ be a colimit cone for D'. We have to find a colimit cone α for D such that $Sign(\alpha) = \alpha'$. We define $D = \langle \Sigma, M \rangle$ where $M = (\cup_{n \in |G|} \alpha_n(M_n))^\circ$ and we define $\alpha_n = \alpha'_n$ for all $n \in |G|$. Then each α_n is a theory morphism and we claim that $\alpha = \langle \alpha_n : \langle \Sigma_n, M_n \rangle \to \langle \Sigma, M \rangle | n \in |G| \rangle$ is a colimit cone over D in **Vth**. For suppose that $\beta'' = \langle \beta_n : \langle \Sigma_n, M_n \rangle \to \langle \Pi, O \rangle | n \in |G| \rangle$ is another cone over D in **Vth**. Then applying $Sign$ to everything, we get a unique $\phi : \Sigma \to \Pi$ such that $\alpha_n \circ \phi = \beta_n$ for each $n \in |G|$. Thus, there is at most one $\psi : \langle \Sigma, M \rangle \to \langle \Pi, O \rangle$ such that $\alpha_n \circ \psi = \beta_n$ for all $n \in |G|$, namely ϕ. Therefore all we need to show is that ϕ is a theory morphism. Because $\beta_n : \langle \Sigma_n, M_n \rangle \to \langle \Pi, O \rangle$ is a theory morphism, we have $\beta_n(M_n) \subseteq O$. Therefore $\cup_{n \in |G|} \beta_n(M_n) \subseteq 0$ and so

$$\phi(M) = \phi\left(\left(\bigcup_n \alpha_n M(n)\right)^\circ\right) \subseteq \left(\bigcup_n \alpha_n \circ \phi M(n)\right)^\circ = \left(\bigcup_n \beta_n M(n)\right)^\circ \subseteq 0$$

where the first inclusion follows from the Closure Lemma.[4] □

17.4 Using More Than One Institution

Institutions theory allows not only to examine relations between theories over the same institution in terms of theory morphisms, but also to consider relations that can be established between theories over different institutions. Indeed, the same program specification may be formalised into different languages requiring the usage of different institutions, especially when different formal verification techniques are to be applied, and different institutions might be essential when providing specifications for different modules of a program. *Institution morphisms* are mappings holding between two heterogeneous institutions which are commonly defined in the following way:

Definition 17.8 (*Institution morphism*) Given two institutions $I = (\mathbf{Sign}, Sen, \mathbf{Mod}, \models_\Sigma)$ and $I' = (\mathbf{Sign}', Sen', \mathbf{Mod}', \models'_\Sigma)$, an institution morphism $\phi : I \to I'$ is defined by:

1. a functor $\phi : \mathbf{Sign} \to \mathbf{Sign}'$ expressing a translation of signatures;
2. a natural transformation $\alpha : \phi \circ Sen' \Rightarrow Sen$, that is, a natural family of functions $\alpha_\Sigma : Sen'(\phi(\Sigma)) \to Sen(\Sigma)$;
3. a natural transformation $\beta : \mathbf{Mod} \Rightarrow \phi \circ \mathbf{Mod}'$, that is, a natural family of functors $\beta_\Sigma : \mathbf{Mod}(\Sigma) \to \mathbf{Mod}'(\phi(\Sigma))$, such that the following satisfaction condition holds

$$A \models_\Sigma \alpha_\Sigma(e') \text{ iff } \beta_\Sigma(A) \models'_{\phi(\Sigma)} e'$$

for A a Σ-model from I and e' a $\phi(\Sigma)$-sentence from I'. □

The natural transformations of 2 and 3 above translate sentences and models satisfying those sentences according to some signature translation provided by the proper functor; note that the satisfaction condition is still required to be independent from those translations.

Establishing morphisms between different institutions providing specifications for the same program (module) allowed institutions theory to be very useful in defining when a *theorem prover* appropriate for a particular institution could be used on specifications within a different institution, this being a further fundamental result advanced in [13] in the application of the Theory of Institution to the program verification problem. Similarly, when program specifications are expressed semantically

[4] Closure Lemma: For any morphism $\phi : \Sigma \to \Sigma'$ and sets F, F' of $\Sigma-$ sentences: $\phi(F^\circ) \subseteq \phi(F)^\circ$.

using institutions, morphisms holding between different institutions for the same program (module) may establish when the output of a model checker applied to a particular institution also holds for another.

Definition 17.9 (*sound institution morphism*) An institution morphism $\phi : I \rightarrow I'$ is sound iff for every signature Σ' and every Σ'-model m' from I', there are a signature Σ and a Σ-model m from I such that $m^{'} = \beta_{\Sigma}(m)$. □

For example, the institution morphism $FOEQ \rightarrow EQ$, from first order equational logic to equational logic, is sound.

Theorem 17.3 *Given a sound institution morphism $\phi : I \rightarrow I'$, for any model $m' \in M'^{\circ}$ in a theory $\langle \Sigma^{'}, M'^{\circ} \rangle$ and any set of sentences E', if $m' \models E'$, then there is a model $m \in M^{\circ}$ in $\langle \Sigma, M^{\circ} \rangle$ such that $m \models \alpha_{\Sigma}(E')$.*

Proof Assume that $m^{'} \models E^{'}$. Because ϕ is a sound institution morphism we have that there is a signature Σ and a Σ-model m such that $m^{'} = \beta_{\Sigma}(m)$. Thus we have $\beta_{\Sigma}(m) \models E^{'} \iff m \models \alpha_{\Sigma}(E^{'})$ because of the definition of an institution morphism (point 3). □

17.5 Conclusions and Future Developments

This paper showed how institutions theory can be used to construct semantic theories formalising programs' module specifications. Proving that syntactic and semantic theories over an institution are logically equivalent, paves the way for extending extant work on algebraic specifications to the case of semantic specifications. We provided some initial results in this direction, showing how to integrate semantic theories over the same institution using colimits, and to establish when two semantic theories built over different institutions share the same properties (i.e. satisfy the same set of sentences, independently of the chosen signature).

Future developments will deal with the problem of representing, and specifying, new program behaviors resulting from the interactions of different program modules. This may be achieved by considering categorical constructs between semantic theories over different institutions, in particular, colimit theories defining models representing, besides the executions of the source modules, module-interaction computations. Constructs of this sort are very useful in modular specification activities, in case different logics are required to provide module specifications. To do so, it will be required to build *duplex institutions* which be model-theoretically defined, that is, the new institutions are to be obtained by linking models of theories over the source and target institutions. Depending on the kind of morphism one establishes between the two starting institutions, duplex institutions may specify different typologies of module interactions in programming and verification. Indeed, institution morphisms in the category of institutions **Ins** can be used to model refinements, integrations, and compositions of modules in a program [9]. Semantic theories over integrating

or composing duplex institutions will thereby be able to provide specifications over those module interactions and compositions.

As stated in the introduction, theories representing specified program behaviors are defined here on the basis of the semantic approach to scientific theories developed in philosophy of science [20, 21, 23]. The examination of the logical relations holding among different semantic theories advanced in this paper sheds new light on the nature of semantic theories and their logical properties, which are usually examined using model theory.

One straightforward application concerns the debate, mentioned in Sect. 17.2, about the theoretical equivalence between syntactic and semantic presentations of theories. Institutions allow one to define a notion of logical equivalence which is language-independent, thereby satisfying the language-independence condition demanded for semantic theories. Moreover, morphisms in the categories **Th** and **Vth**, and specified functors mapping those morphisms, can be used to clarify the relations holding between sentences and their model classes. Indeed, a related problem in formal philosophy of science is that of establishing whether that is a one-to-many or a many-to-many relation [14, 15]. Institutions theory can be applied to specify the categorical structure of the many-to-many relation between sentences and model classes. Another area of application for institutions theory finally concerns the modular nature of many scientific theories (see for instance [7]). Morphisms in the category **Vth** of semantic theories may be used to define logic relations holding among the different models of the modular theory. Not only: refinements, integrations, and compositions of models in a modular theory can be expressed by considering institution morphisms in the category **Ins** of institutions. Refinements in particular can be very useful to provide a logical analysis of the theory-revision processes for semantic theories.

References

1. Angius, N., Stefaneas, P.: The Logical Structure of Modular Semantic Theories of Software Systems. J. Log. Comput (Forthcoming)
2. Barwise, J.K.: Axioms for abstract model theory. Ann. Math. Log. **7**(2), 221–265 (1974)
3. Baier, C., Katoen, J.P.: Principles of model checking. MIT press, Cambridge (2008)
4. Burstall, R.M., Goguen, J.A.: Putting theories together to make specifications. In: Proceedings of the 5th International Joint Conference on Artificial Intelligence, Volume 2, pp. 1045–1058. Morgan Kaufmann, San Francisco (1977)
5. Burstall, R.M., Goguen, J.A.: The semantics of Clear, a specification language. In: Abstract Software Specifications, pp. 292–332, Springer, Heidelberg (1980)
6. Clarke, E.M., Grumberg, O., Peled, D.A.: Model checking. MIT Press, Cambridge (1999)
7. Darrigol, O.: The modular structure of physical theories. Synth. **162**(2), 195–223 (2008)
8. Diaconescu, R.: Three decades of institution theory. Universal logic: an anthology, pp. 309–322 (2012)
9. Diaconescu, R., Goguen, J., Stefaneas, P.: Logical support for modularization. Second Annual Workshop on Logical Enviroments, pp. 83–100 (1993)
10. Glymour, C.: Theoretical equivalence and the semantic view of theories. Philos. Sci. **80**(2), 286–297 (2013)

11. Goguen, J.A.: A categorical manifesto. Math. Struct. Comput. Sci. **1**(1), 49–67 (1991)
12. Goguen, J.A., Burstall, R.M.: Introducing institutions. In: Clarke, E., Kozen, D. (eds). Proc. Logics of Programming Workshop, vol. 164 of LNCS, pp. 221–256. Springer, Heidelberg (1984)
13. Goguen, J.A., Burstall, R.M.: Institutions: Abstract model theory for specification and programming. J. ACM (JACM) **39**(1), 95–146 (1992)
14. Halvorson, H.: What scientific theories could not be. Philos. Sci. **79**(2), 183–206 (2012)
15. Halvorson, H.: The semantic view, if plausible, is syntactic. Philos. Sci **80**(3), 475–478 (2013)
16. Lawvere, F.W.: Adjointness in foundations. Dialect. **23**(3–4), 281–296 (1969)
17. Birkhoff, G.: On the structure of abstract algebras. Proc. Camb. Philos. Soc. **31**, 433–454 (1935)
18. Mller, P.: Modular Specification and Verification of Object Oriented Programs. Springer, Berlin (2002)
19. Sanella, D., Tarlecki, A.: Foundations of Algebraic Specifications and Formal Software Development. Springer, Berlin (2012)
20. Suppe, F.: The semantic conception of theories and scientific realism. University of Illinois Press, Urbana (1989)
21. Suppes, P.: A comparison of the meaning and uses of models in mathematics and the empirical sciences. Synth. **12**, 287301 (1960)
22. Tarski, A.: The semantic conception of truth: and the foundations of semantics. Philos. Phenomenol. Res. **4**(3), 341–376 (1944)
23. van Fraassen, B.C.: The Scientific Image. Oxford University Press, Oxford (1980)
24. van Fraassen, B.C.: Laws and Symmetry. Oxford University Press, Oxford (1989)

Chapter 18
Generic Constructions and Generic Limits

Sergey V. Sudoplatov, Yiannis Kiouvrekis and Petros Stefaneas

Abstract We collect results related to generic constructions and generic limits for semantic and syntactic cases. It is considered both by pure model theory approach and by the institutional approach.

Appeared first in Fraïssé's papers [7, 8] generic constructions allowed, via generic limits, to solve a series of known model-theoretic problems and to apply for the solutions of related questions.

At present survey we consider both semantic and syntactic generic constructions and their applications. These constructions are considered both from pure model theory and institutional viewpoints.

The survey is organized as follows. In Sect. 18.1 we recall the semantic generic constructions and Theorem on existence of generic structure being a generic limit. Section 18.2 contains references for some known applications of semantic generic

S.V. Sudoplatov (✉)
Sobolev Institute of Mathematics, 4, Acad. Koptyug Avenue,
Novosibirsk 630090, Russia
e-mail: sudoplat@math.nsc.ru

S.V. Sudoplatov
Novosibirsk State Technical University, 20, K.Marx Avenue,
Novosibirsk 630073, Russia

S.V. Sudoplatov
Novosibirsk State University, 2, Pirogova Street,
Novosibirsk 630090, Russia

S.V. Sudoplatov
Institute of Mathematics and Mathematical Modeling,
125, Pushkina Street, Almaty 050010, Kazakhstan

Y. Kiouvrekis · P. Stefaneas
Department of Applied Mathematics, National Technical University of Athens,
Zografou Campus, 15780 Athens, Greece
e-mail: yiannisq@central.ntua.gr

P. Stefaneas
e-mail: petros@math.ntua.gr

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_18

constructions. In Sect. 18.3 we describe, in details, the syntactic approach to generic constructions, some valuable variations of this approach and results on properties of generic structures. Section 18.4 deals with so-called adapted generic classes allowing to construct structures following the structures of formulas in given diagrams. Properties of structural diagrams and corresponding generic structures are described in Sect. 18.5. In Sect. 18.6, we propose an application of the syntactic approach for the class of cubic theories producing the description of spectrum functions for the class of cubic theories. In Sect. 18.7, we consider related topological objects for generic classes producing a classification of models with respect these objects. An institutional generalization of generic constructions and generic limits is considered in Sect. 18.8.

18.1 Semantic Generic Constructions and Generic Limits

Definition 18.1 [1, 13, 24, 42] Let $\mathbf{K}_0$ be a class of finite structures of at most countable predicate language, endowed with a partial order relation $\preceq$ which is invariant under the transition to isomorphic structures, connoting the property of being a *self-sufficient structure*, or *strong substructure*. The class $(\mathbf{K}_0; \preceq)$ is called *generic*, *generative*, or *amalgamation* if it satisfies the following axioms:

(1) if $\mathcal{A} \preceq \mathcal{B}$, then $\mathcal{A} \subseteq \mathcal{B}$;

(2) if $\mathcal{A} \preceq \mathcal{C}$, $\mathcal{B} \in \mathbf{K}_0$, and $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{C}$, then $\mathcal{A} \preceq \mathcal{B}$;

(3) $\emptyset$ is the least element of the system $(\mathbf{K}_0; \preceq)$;

(4) (the *amalgamation property*) for any structures $\mathcal{A}$, $\mathcal{B}, \mathcal{C} \in \mathbf{K}_0$, having embeddings $f_0 : \mathcal{A} \to \mathcal{B}$ and $g_0 : \mathcal{A} \to \mathcal{C}$ such that $f_0(\mathcal{A}) \preceq \mathcal{B}$ and $g_0(\mathcal{A}) \preceq \mathcal{C}$, there are a structure $\mathcal{D} \in \mathbf{K}_0$ and embeddings $f_1 : \mathcal{B} \to \mathcal{D}$ and $g_1 : \mathcal{C} \to \mathcal{D}$ for which $f_1(\mathcal{B}) \preceq \mathcal{D}$, $g_1(\mathcal{C}) \preceq \mathcal{D}$ and $f_0 \circ f_1 = g_0 \circ g_1$; the structure $\mathcal{D}$ is called the *amalgam* of $\mathcal{B}$ and $\mathcal{C}$ over the structure $\mathcal{A}$ and the four (f_0, g_0, f_1, g_1).

With the partially ordered class $(\mathbf{K}_0; \preceq)$, being defined above and with at most countably many isomorphism types, taking finite structures in $\mathbf{K}_0$ and using the *amalgamation* (i. e., embedding the structures $\mathcal{B}$ and $\mathcal{C}$ over $\mathcal{A}$ in structures $\mathcal{D}$ so as to comply with the amalgamation property), we construct step-by-step a countable $(\mathbf{K}_0; \preceq)$-*generic structure* $\mathcal{M}$, that is also called a *Hrushovski limit* for $(\mathbf{K}_0; \preceq)$), i. e., a structure satisfying the following conditions:

(a) for any finite substructure $\mathcal{A} \subseteq \mathcal{M}$, there is a structure $\mathcal{B} \in \mathbf{K}_0$, $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{M}$, for which $\mathcal{B} \preceq \mathcal{M}$, i. e., $\mathcal{B} \preceq \mathcal{B}'$ for any structure $\mathcal{B}' \in \mathbf{K}_0$ with $\mathcal{B} \subseteq \mathcal{B}' \subseteq \mathcal{M}$;

(b) for any finite substructure $\mathcal{A} \subseteq \mathcal{M}$ and any structure $\mathcal{B} \in \mathbf{K}_0$ such that $\mathcal{A} \preceq \mathcal{B}$, there is a structure $\mathcal{B}' \preceq \mathcal{M}$ for which $\mathcal{B} \simeq_{\mathcal{A}} \mathcal{B}'$.

Note that for particular cases, when a generic class $(\mathbf{K}_0; \leqslant)$ have the relation $\leqslant$ which is equal to $\subseteq$ (i. e., $\mathbf{K}_0$ is closed under substructures), the class $(\mathbf{K}_0; \leqslant)$ is called the *Fraïssé class* and the Hrushovski limit is called the *Fraïssé limit* [9, 15]. In such a case finite structures can be replaced by finitely generated structures.

By the definition of Hrushovski limit, the following theorem holds.

Theorem 18.1 ([1, Theorem 2.12]) *For any partially ordered class* $(\mathbf{K}_0; \preceq)$*, satisfying conditions* 1–4 *and having at most countably many isomorphism types, there exists a* $(\mathbf{K}_0; \preceq)$*-generic structure.*

The scheme above represents a *semantic approach* to constructing a generic structure $\mathscr{M}$ and the corresponding *generic theory* $\mathrm{Th}(\mathscr{M})$.

The utility of the semantic approach for realizations of desired model-theoretic properties has confirmed by numerous examples (see the bibliography, reflected in the historical survey) for the cases when predicates are independent, i. e., are not definable in terms of each other.

18.2 Applications for Model-Theoretic Purposes

Fraïssé limits used in constructions of ω-categorical structures with quantifier elimination, random and induced structures [15].

E. Hrushovski [16] has refuted Zilber conjecture constructing examples of strongly minimal theories which are not locally modular and in which infinite groups are not interpreted. His original construction, being a basis for constructions of related examples and for the further solutions of others known model theoretical problems, stimulated intensive investigations of Hrushovski construction itself and its various modifications (in wide sense), capable to create "pathological" theories with given properties as well as axiomatic bases and their applications, allowing to determine applicability bounds for that construction (see [38] for references).

Topological applications of semantic generic constructions are represented in [17–19].

18.3 Syntactic Generic Constructions and Generic Limits

18.3.1 Notions and Properties

We consider collections of sentence and formulas in first order logic over a language Σ. Thus, as usual, $\vdash$ means proof from no hypotheses deducing $\vdash \varphi$ for a formula φ of language Σ, which may contain function symbols and constants. If deducing φ, hypotheses in a set Φ of formulas can be used, we write $\Phi \vdash \varphi$. Usually Σ will be fixed in context and not mentioned explicitly.

Below we write $X, Y, Z, \ldots$ for finite sets of variables, and denote by $A, B, C, \ldots$ finite sets of elements, as well as finite sets in structures, or else the structures with finite universes themselves.

In diagrams, $A, B, C, \ldots$ denote finite sets of constant symbols disjoint from the constant symbols in Σ and $\Sigma(A)$ is the vocabulary with the constants from

A adjoined. $\Phi(A), \Psi(B), \mathrm{X}(C)$ stand for Σ-*diagrams* (of sets A, B, C), that is, *consistent sets of* $\Sigma(A)$-, $\Sigma(B)$-, $\Sigma(C)$-*sentences*, respectively.

Below we assume that for any considered diagram $\Phi(A)$, if a_1, a_2 are distinct elements in A then $\neg(a_1 \approx a_2) \in \Phi(A)$. This means that if c is a constant symbol in Σ, then there is at most one element $a \in A$ such that $(a \approx c) \in \Phi(A)$.

If $\Phi(A)$ is a diagram and B is a set, we denote by $\Phi(A)|_B$ the set $\{\varphi(\bar{a}) \in \Phi(A) \mid \bar{a} \in B\}$. Similarly, for a language Σ, we denote by $\Phi(A)|_\Sigma$ the restriction of $\Phi(A)$ to the set of formulas in the language Σ.

Definition 18.2 [30, 31, 34, 36, 38] We denote by $[\Phi(A)]^A_B$ the diagram $\Phi(B)$ obtained by replacing a subset $A' \subseteq A$ by a set $B' \subseteq B$ of constants disjoint from Σ and with $|A'| = |B'|$, where $A \setminus A' = B \setminus B'$. Similarly we call the consistent set of formulas denoted by $[\Phi(A)]^A_X$ the type $\Phi(X)$ if it is the result of a bijective substitution into $\Phi(A)$ of variables of X for the constants in A. In this case, we say that $\Phi(B)$ is a *copy* of $\Phi(A)$ and a *representative* of $\Phi(X)$. We also denote the diagram $\Phi(A)$ by $[\Phi(X)]^X_A$.

Remark 18.1 If the vocabulary contains functional symbols then diagrams $\Phi(A)$ containing equalities and inequalities of terms can generate both finite and infinite structures. The same effect is observed for purely predicate vocabularies if it is written in $\Phi(A)$ that the model for $\Phi(A)$ should be infinite. For instance, diagrams containing axioms for finitely axiomatizable theories have this property.

By the definition, for any diagram $\Phi(A)$, each constant symbol in Σ appears in some formula of $\Phi(A)$. Thus, $\Phi(A)$ can be considered as $\Phi(A \cup K)$, where K is the set of constant symbols in Σ.

We now give conditions on a partial ordering of a collection of diagrams which suffice for it to determine a structure. We modify some of the conditions for structures by d to signify they are conditions on diagrams not structures.

Definition 18.3 [30, 31, 34, 36, 38] Let Σ be a vocabulary. We say that $(\mathbf{D}_0; \preceq)$ (or $\mathbf{D}_0$) is *generic*, or *generative*, if $\mathbf{D}_0$ is a class of Σ-diagrams of finite sets so that $\mathbf{D}_0$ is partially ordered by a binary relation $\preceq$ such that $\preceq$ is preserved by bijective substitutions, i. e., if $\Phi(A) \preceq \Psi(B)$, and $A' \subseteq B'$ such that $[\Phi(A)]^A_{A'} = \Phi(A')$ and $[\Psi(B)]^B_{B'} = \Psi(B')$ are defined, then $[\Phi(A)]^A_{A'}$, $[\Psi(B)]^B_{B'}$ are in $\mathbf{D}_0$ and $[\Phi(A)]^A_{A'} \preceq [\Psi(B)]^B_{B'}$.[1] Furthermore:

(i) if $\Phi(A) \in \mathbf{D}_0$ then for any quantifier free formula $\varphi(\bar{x})$ and any tuple $\bar{a} \in A$ either $\varphi(\bar{a}) \in \Phi(A)$ or $\neg\varphi(\bar{a}) \in \Phi(A)$;

(ii) if $\Phi \preceq \Psi$ then $\Phi \subseteq \Psi$[2];

(iii) if $\Phi \preceq \mathrm{X}$, $\Psi \in \mathbf{D}_0$, and $\Phi \subseteq \Psi \subseteq \mathrm{X}$, then $\Phi \preceq \Psi$;

(iv) some diagram $\Phi_0(\varnothing)$ is the least element of $(\mathbf{D}_0; \leqslant)$;

[1] Note that $\mathbf{D}_0$ is closed under bijective substitutions since $\preceq$ is preserved by bijective substitutions and $\preceq$ is reflexive.

[2] Note that $\Phi(A) \preceq \Psi(B)$ implies $A \subseteq B$, since if $a \in A$ then $(a \approx a) \in \Phi(A)$, so $\Phi(A) \preceq \Psi(B)$ implies $\Phi(A) \subseteq \Psi(B)$ and we have $(a \approx a) \in \Psi(B)$, whence $a \in B$.

(v) (the *d-amalgamation property*) for any diagrams $\Phi(A)$, $\Psi(B)$, $\mathrm{X}(C) \in \mathbf{D}_0$, if there exist injections $f_0 : A \to B$ and $g_0 : A \to C$ with $[\Phi(A)]^A_{f_0(A)} \preceq \Psi(B)$ and $[\Phi(A)]^A_{g_0(A)} \preceq \mathrm{X}(C)$, then there are a diagram $\Theta(D) \in \mathbf{D}_0$ and injections $f_1 : B \to D$ and $g_1 : C \to D$ for which $[\Psi(B)]^B_{f_1(B)} \preceq \Theta(D)$, $[\mathrm{X}(C)]^C_{g_1(C)} \preceq \Theta(D)$ and $f_0 \circ f_1 = g_0 \circ g_1$; the diagram $\Theta(D)$ is called the *amalgam* of $\Psi(B)$ and $\mathrm{X}(C)$ over the diagram $\Phi(A)$ and witnessed by the four maps (f_0, g_0, f_1, g_1);

(vi) (the *local realizability property*) if $\Phi(A) \in \mathbf{D}_0$ and $\Phi(A) \vdash \exists x\, \varphi(x)$, then there are a diagram $\Psi(B) \in \mathbf{D}_0$, $\Phi(A) \preceq \Psi(B)$, and an element $b \in B$ for which $\Psi(B) \vdash \varphi(b)$;

(vii) (the *d-uniqueness property*) for any diagrams $\Phi(A)$, $\Psi(B) \in \mathbf{D}_0$ if $A \subseteq B$ and the set $\Phi(A) \cup \Psi(B)$ is consistent then $\Phi(A) = \{\varphi(\bar{b}) \in \Psi(B) \mid \bar{b} \in A\}$.

A diagram Φ is called a *strong subdiagram* of a diagram Ψ if $\Phi \preceq \Psi$.

A diagram $\Phi(A)$ is said to be (*strongly*) *embeddable* in a diagram $\Psi(B)$ if there is an injection $f : A \to B$ such that $[\Phi(A)]^A_{f(A)} \subseteq \Psi(B)$ ($[\Phi(A)]^A_{f(A)} \preceq \Psi(B)$). The injection f, in this instance, is called a (*strong*) *embedding* of diagram $\Phi(A)$ in diagram $\Psi(B)$ and is denoted by $f : \Phi(A) \to \Psi(B)$. A diagram $\Phi(A)$ is said to be (*strongly*) *embeddable* in a structure $\mathcal{M}$ if $\Phi(A)$ is (strongly) embeddable in some diagram $\Psi(B)$, where $\mathcal{M} \models \Psi(B)$. The corresponding embedding $f : \Phi(A) \to \Psi(B)$, in this case, is called a (*strong*) *embedding* of diagram $\Phi(A)$ in structure $\mathcal{M}$ and is denoted by $f : \Phi(A) \to \mathcal{M}$.

Let $\mathbf{D}_0$ be a class of diagrams, $\mathbf{P}_0$ be a class of structures of some language, and $\mathcal{M}$ be a structure in $\mathbf{P}_0$. The class $\mathbf{D}_0$ is *cofinal* in the structure $\mathcal{M}$ if for each finite set $A \subseteq M$, there are a finite set B, $A \subseteq B \subseteq M$, and a diagram $\Phi(B) \in \mathbf{D}_0$ such that $\mathcal{M} \models \Phi(B)$. The class $\mathbf{D}_0$ is *cofinal* in $\mathbf{P}_0$ if $\mathbf{D}_0$ is cofinal in every structure of $\mathbf{P}_0$. We denote by $\mathbf{K}(\mathbf{D}_0)$ the class of all structures $\mathcal{M}$ with the condition that $\mathbf{D}_0$ is cofinal in $\mathcal{M}$, and by $\mathbf{P}$ a subclass of $\mathbf{K}(\mathbf{D}_0)$ such that each diagram $\Phi \in \mathbf{D}_0$ is true in some structure in $\mathbf{P}$.

Now we extend the relation $\preceq$ from the generative class $(\mathbf{D}_0; \preceq)$ to a class of subsets of structures in the class $\mathbf{K}(\mathbf{D}_0)$.

Let $\mathcal{M}$ be a structure in $\mathbf{K}(\mathbf{D}_0)$, A and B be finite sets in $\mathcal{M}$ with $A \subseteq B$. We call A a *strong subset* of the set B (in the structure $\mathcal{M}$), and write $A \preceq B$, if there exist diagrams $\Phi(A)$, $\Psi(B) \in \mathbf{D}_0$, for which $\Phi(A) \preceq \Psi(B)$ and $\mathcal{M} \models \Psi(B)$.

A finite set A is called a *strong subset* of a set $M_0 \subseteq M$ (in the structure $\mathcal{M}$), where $A \subseteq M_0$, if $A \preceq B$ for any finite set B such that $A \subseteq B \subseteq M_0$ and $\Phi(A) \subseteq \Psi(B)$ for some diagrams $\Phi(A)$, $\Psi(B) \in \mathbf{D}_0$ with $\mathcal{M} \models \Psi(B)$. If A is a strong subset of M_0 then, as above, we write $A \preceq M_0$. If $A \preceq M$ in $\mathcal{M}$ then we refer to A as a *self-sufficient set* (in $\mathcal{M}$).

Notice that, by the d-uniqueness property, the diagrams $\Phi(A)$ and $\Psi(B)$ specified in the definition of strong subsets are defined uniquely. A diagram $\Phi(A) \in \mathbf{D}_0$, corresponding to a self-sufficient set A in $\mathcal{M}$, is said to be a *self-sufficient diagram* (in $\mathcal{M}$).

Definition 18.4 [30, 31, 34, 36, 38] A class $(\mathbf{D}_0; \preceq)$ possesses the *joint embedding property* (JEP) if for any diagrams $\Phi(A), \Psi(B) \in \mathbf{D}_0$, there is a diagram $\mathrm{X}(C) \in \mathbf{D}_0$ such that $\Phi(A)$ and $\Psi(B)$ are strongly embeddable in $\mathrm{X}(C)$.

Clearly, every generative class has JEP since JEP means the d-amalgamation property over the empty set.

Definition 18.5 [30, 31, 34, 36, 38] A structure $\mathcal{M} \in \mathbf{P}$ has *finite closures* with respect to the class $(\mathbf{D}_0; \preceq)$, or is *finitely generated over* Σ, if any finite set $A \subseteq M$ is contained in some finite self-sufficient set in $\mathcal{M}$, i. e., there is a finite set B with $A \subseteq B \subseteq M$ and $\Psi(B) \in \mathbf{D}_0$ such that $\mathcal{M} \models \Psi(B)$ and $\Psi(B) \preceq \mathrm{X}(C)$ for any $\mathrm{X}(C) \in \mathbf{D}_0$ with $\mathcal{M} \models \mathrm{X}(C)$ and $\Psi(B) \subseteq \mathrm{X}(C)$. A class $\mathbf{P}$ has *finite closures* with respect to the class $(\mathbf{D}_0; \preceq)$, or is *finitely generated over* Σ, if each structure in $\mathbf{P}$ has finite closures (with respect to $(\mathbf{D}_0; \preceq)$).

Clearly, an at most countable structure $\mathcal{M}$ has finite closures with respect to $(\mathbf{D}_0; \preceq)$ if and only if $M = \bigcup\limits_{i \in \omega} A_i$ for some self-sufficient sets A_i with $A_i \preceq A_{i+1}$, $i \in \omega$.

Note that the finite closure property is defined modulo Σ and does not correlate with the cardinalities of algebraic closures. For instance, if Σ contains infinitely many constant symbols then $\mathrm{acl}(A)$ is always infinite whereas a finite set A can or can not be extended to a self-sufficient set.

Besides, for the finite closures of sets A we consider finite self-sufficient extensions B in a given structure $\mathcal{M}$ with respect to $(\mathbf{D}_0; \preceq)$ only and B can be both a universe of a substructure of $\mathcal{M}$ or not. Moreover, it is permitted that corresponding diagrams $\Psi(B)$ can have only finite, finite and infinite, or only infinite models.

Thus, for instance, a finitely axiomatizable theory without finite models and with a generative class $(\mathbf{D}_0; \subseteq)$, containing diagrams for all finite sets and with axioms in diagrams, has identical finite closures whereas each diagram in $\mathbf{D}_0$ has only infinite models.

Definition 18.6 [30, 31, 34, 36, 38] A structure $\mathcal{M} \in \mathbf{K}(\mathbf{D}_0)$ is $(\mathbf{D}_0; \preceq)$-*generic*, or a *generic limit for the class* $(\mathbf{D}_0; \preceq)$ and denoted by $\mathrm{glim}(\mathbf{D}_0; \preceq)$, if it satisfies the following conditions:

(a) $\mathcal{M}$ has finite closures with respect to $\mathbf{D}_0$;

(b) if $A \subseteq M$ is a finite set, $\Phi(A), \Psi(B) \in \mathbf{D}_0$, $\mathcal{M} \models \Phi(A)$ and $\Phi(A) \preceq \Psi(B)$, then there exists a set $B' \preceq M$ such that $A \subseteq B'$ and $\mathcal{M} \models \Psi(B')$.

Theorem 18.2 ([34, 36, 38]) *For any generative class* $(\mathbf{D}_0; \preceq)$ *with at most countably many diagrams whose copies form* $\mathbf{D}_0$*, there exists a* $(\mathbf{D}_0; \preceq)$*-generic structure.*

Note that [23, Theorem 2.1] implies that Theorem 18.2 does not hold for some generative classes $(\mathbf{D}_0; \preceq)$ with uncountable languages and corresponding to semantic amalgamation classes. It means that for some uncountable languages there are amalgamation classes $(\mathbf{K}_0; \subseteq)$ without Fraïssé limits, i. e., without $(\mathbf{K}_0; \subseteq)$-generic structures. In such a case the class $(\mathbf{K}_0; \subseteq)$ can be extended to an amalgamation

class $(\mathbf{K}_0'; \subseteq)$ with $(\mathbf{K}_0'; \subseteq)$-generic structure. The class $\mathbf{K}_0'$ can be defined taking all finitely generated substructures of a structure $\mathscr{M}$ which collects copies of all structures in $\mathbf{K}_0$ and should have new isomorphism types with respect to $\mathbf{K}_0$. The class $\mathbf{K}_0'$ can be considered as a "closure" of the class $\mathbf{K}_0$ (with respect to $\mathscr{M}$). Clearly, $\mathbf{K}_0'$ can vary depending on $\mathscr{M}$ and the question is: is it possible to find the least $\mathbf{K}_0' \supset \mathbf{K}_0$, or at least a minimal one, with $(\mathbf{K}_0'; \subseteq)$-generic structure? The same question arises for generative classes $(\mathbf{D}_0; \preceq)$ with uncountable languages, modulo classes generating isomorphic structures.

Considering these questions we note that the only obstacle for the existence of a $(\mathbf{D}_0; \preceq)$-generic structure, actually noticed by Kudaibergenov [23], is an imbalance between cardinalities of definable sets forced by diagrams in $\mathbf{D}_0$ with their amalgams and cardinalities of definable links forced by formulas in these diagrams. By Theorem 18.2 this imbalance can start with generative classes $(\mathbf{D}_0; \preceq)$ having at least uncountably many copies which are not linked by substitutions. In particular, it can be forced by generative classes with uncountable vocabularies. Thus, if the least (minimal) "closure" exists, then it can be achieved adding the least (minimal) class of diagrams which removes that imbalance.

Theorem 18.3 ([30]) *Every ω-homogeneous structure $\mathscr{M}$ is $(\mathbf{D}_0; \preceq)$-generic for some generative class $(\mathbf{D}_0; \preceq)$.*

Thus any first-order theory has a generic model and therefore can be represented by it.

Definition 18.7 [30, 31, 34, 36, 38] A generic class $(\mathbf{D}_0; \preceq)$ is *self-sufficient* if the following *axiom of self-sufficiency* holds:

(viii) if $\Phi, \Psi, \mathrm{X} \in \mathbf{D}_0$, $\Phi \preceq \Psi$, and $\mathrm{X} \subseteq \Psi$, then $\Phi \cap \mathrm{X} \preceq \mathrm{X}$.

Theorem 18.4 ([34, 36, 38]) *Let $(\mathbf{D}_0; \preceq)$ be a self-sufficient class, $\overline{\mathscr{M}}$ be at most countable $(\mathbf{D}_0; \preceq)$-generic structure, and $\mathbf{K}$ be the class of all models of $T = \mathrm{Th}(\overline{\mathscr{M}})$ which has finite closures. Then the generic structure $\overline{\mathscr{M}}$ is homogeneous.*

Thus, since any ω-homogeneous structure can be considered as generic with respect to a generic class with complete diagrams, a countable structure $\mathscr{M}$ is homogeneous if and only if it is generic for an appropriate self-sufficient generic class $(\mathbf{D}_0; \preceq)$.

18.3.2 Pre-generative Classes

Definition 18.8 [30] Consider the following modification of the d-amalgamation property for a class $(\mathbf{D}_0; \preceq)$:

(v′) for any diagrams $\Phi(A), \Psi(B), \mathrm{X}(C) \in \mathbf{D}_0$, if there exist injections $f_0 : A \to B$ and $g_0 : A \to C$ with $[\Phi(A)]^A_{f_0(A)} \preceq \Psi(B)$ and $[\Phi(A)]^A_{g_0(A)} \preceq \mathrm{X}(C)$ such that $B \setminus A = B \setminus f_0(A), C \setminus A = C \setminus g_0(A), (B \setminus A) \cap (C \setminus A) = \emptyset$ and $[\Psi(B)]^{f_0(A)}_A \cup$

$[\mathrm{X}(C)]_A^{g_0(A)}$ is consistent, then there are a diagram $\Theta(D) \in \mathbf{D}_0$ and injections $f_1 : B \to D$ and $g_1 : C \to D$ for which $[\Psi(B)]^B_{f_1(B)} \preceq \Theta(D)$, $[\mathrm{X}(C)]^C_{g_1(C)} \preceq \Theta(D)$ and $f_0 \circ f_1 = g_0 \circ g_1$.

Note that replacing the d-amalgamation property in the definition of generative class by (v′) it suffices to use only identical embeddings for the construction of $(\mathbf{D}_0; \preceq)$-generic structure.

By the definition every generative class $(\mathbf{D}_0; \preceq)$ is generated by a set $\mathbf{D}'_0$ of diagrams in $\mathbf{D}_0$ such that each $\Phi(A) \in \mathbf{D}_0$ has a copy $\Phi(A') \in \mathbf{D}'_0$.

Let $\mathbf{D}'_0$ be a class (in particular, a set) of diagrams $\Phi(A)$ over finite sets A, in a language Σ, such that a set of some copies for all elements in $\mathbf{D}'_0$ is consistent and for each $\Phi(A) \in \mathbf{D}'_0$, $\varphi(\bar{a}) \in \Phi(A)$ or $\neg\varphi(\bar{a}) \in \Phi(A)$ for any quantifier-free formula $\varphi(\bar{x})$ and any tuple $\bar{a} \in A$.

Definition 18.9 [30] We say that $\mathbf{D}'_0$ is *pre-generic*, or *pre-generative*, if it is equipped with a partial order $\preceq'_0$ satisfying conditions (ii), (iii), (vii) for $(\mathbf{D}'_0; \preceq_0)$ as well as the *property of invariance* for $\preceq_0$ under bijective substitutions as follows:

if $\Phi(A) \preceq_0 \Psi(B)$, $A' \subseteq B'$ and $[\Phi(A)]^A_{A'}, [\Psi(B)]^B_{B'} \in \mathbf{D}'_0$ then $[\Phi(A)]^A_{A'} \preceq_0 [\Psi(B)]^B_{B'}$.

In this case, we also say that $(\mathbf{D}'_0; \preceq_0)$ is *pre-generic*, or *pre-generative*.

By the definition every generative class is pre-generative but not vice versa.

Having the axiom (viii) for a pre-generative class $(\mathbf{D}'_0; \preceq_0)$ we also say that $(\mathbf{D}'_0; \preceq_0)$ is *self-sufficient*.

Definition 18.10 [30] A (pre-)generative class $(\mathbf{D}'_0; \preceq_0)$ is *regular* if for any copies $\Phi_1(A_1), \ldots, \Phi_n(A_n)$ of elements in $\mathbf{D}'_0$ with consistent $\Phi_1(A_1) \cup \ldots \cup \Phi_n(A_n)$, we have $\Phi_1(A_1) \cap \ldots \cap \Phi_n(A_n) = \Phi_i(A_i)|_{(A_1 \cap \ldots \cap A_n)}$, $i = 1, \ldots, n$.

A (pre-)generative class $(\mathbf{D}'_0; \preceq_0)$ is *non-refinable* if for any $\Phi(A) \in \mathbf{D}'_0$ and B with $B \subset A$, there is $\Psi(B) \in \mathbf{D}'_0$ such that $\Psi(B) \subset \Phi(A)$, and $\preceq_0 = \subseteq$.

By the axiom of d-uniqueness, every non-refinable (pre-)generative class is regular and self-sufficient. Note that each consistent set of complete diagrams, in a given language, is regular and its closure under complete subdiagrams is non-refinable.

Clearly, if $(\mathbf{D}_0; \preceq)$ is a generative class then for any $\mathbf{D}'_0 \subseteq \mathbf{D}_0$ the restriction $(\mathbf{D}_0; \preceq)|_{\mathbf{D}'_0}$ is pre-generative. At the same time the following theorem holds:

Theorem 18.5 ([30]) *Any regular pre-generative class* $(\mathbf{D}'_0; \preceq_0)$ *can be extended to a non-refinable generative class* $(\mathbf{D}_0; \preceq)$.

18.4 Adapted Generative Classes

Definition 18.11 [34, 36, 38] Let $(\mathbf{D}_0; \preceq)$ and $(\mathbf{D}'_0; \preceq')$ be generative classes of languages Σ and Σ', respectively, with $\Sigma \subseteq \Sigma'$. We say that the class $(\mathbf{D}'_0; \preceq')$

dominates the class $(\mathbf{D}_0; \preceq)$, and write $\mathbf{D}_0 \leq \mathbf{D}'_0$, if for any diagram $\Phi(A) \in \mathbf{D}_0$ there is a diagram $\Phi'(A') \in \mathbf{D}'_0$ such that $\Phi(A) \subseteq \Phi'(A')$, and the condition of there being some systems, which are extensions over A, together with available information on interrelations of elements in these extensions written in the diagram $\Phi(A)$, implies that the same extensions exist over A, and that similar information is available on interrelations of elements in those extensions written in the diagram $\Phi'(A')$.

If $\mathbf{D}_0 \leq \mathbf{D}'_0$ and $\mathbf{D}'_0 \leq \mathbf{D}_0$ we say that generative classes $(\mathbf{D}_0; \preceq)$ and $(\mathbf{D}'_0; \preceq')$ are *domination-equivalent* and write $\mathbf{D}_0 \sim \mathbf{D}'_0$.

It is easy to see that $\sim$ is an equivalence relation, and by uniqueness of homogeneous structures realizing same set of types if generative classes are domination-equivalent then these classes produce isomorphic generic structures. We have the converse implication for quantifier-free generative classes.

At the same time, there are generative classes, being not $\sim$-equivalent but forming isomorphic generic structures. For instance, the structure $\langle \mathbf{Q}; \leq \rangle$, having a finitely axiomatizable theory with an axiom φ_0, is generated both by the quantifier-free generative class $(\mathbf{D}_0; \preceq)$ (whose diagrams describe $\leq$-links of elements for finite subsets of $\mathbf{Q}$) and by the generative class $(\mathbf{D}'_0; \preceq')$, where each diagram contains the axiom φ_0. Clearly, $\mathbf{D}_0 \leq \mathbf{D}'_0$ and $\mathbf{D}'_0 \not\preceq \mathbf{D}_0$.

Similar the skolemization of theories we define *special* generative classes adapted to the semantic of given formulae.

Definition 18.12 [31] Let $(\mathbf{D}_0; \preceq)$ be a generative class, $\mathcal{M}$ be a $(\mathbf{D}_0; \preceq)$-generic structure,

$$\varphi(x_1, \ldots, x_m) \rightleftharpoons \mathrm{Q}_1 y_1 \ldots \mathrm{Q}_n y_n \psi(x_1, \ldots, x_m, y_1, \ldots, y_n)$$

be a formula in prenex normal form, $\mathrm{Q}_i \in \{\forall, \exists\}$, ψ is a quantifier-free formula, $\mathcal{M} \models \varphi(a_1, \ldots, a_m)$ for some $a_1, \ldots, a_m \in M$. $(\mathbf{D}_0; \preceq)$ is called $\varphi(a_1, \ldots, a_m)$-*adapted*, or briefly φ-*adapted*, if it satisfies the following conditions:

(1) if φ is an existential formula then for any diagram $\Phi(A) \in \mathbf{D}_0$, where $A \neq \emptyset$ and $\Phi(A) \cup \{\varphi(a_1, \ldots, a_m)\}$ is consistent, there are $a'_1, \ldots, a'_n \in A$ such that $\Phi(A) \vdash \psi(a_1, \ldots, a_m, a'_1, \ldots, a'_n)$;

(2) if φ has the form $\forall y'_1, \ldots, y'_k \exists y''_1, \ldots, y''_l \chi$ and for $\Phi(A) \in \mathbf{D}_0$, $\Phi(A) \vdash \varphi(a_1, \ldots, a_m)$ then for any $\Psi(B) \in \mathbf{D}_0$ with $\Psi(B) \supset \Phi(A)$, and for any $a'_1, \ldots, a'_k \in A$, there are $b_1, \ldots, b_l \in B$ such that

$$\Psi(B) \vdash \chi(a_1, \ldots, a_m, a'_1, \ldots, a'_k, b_1, \ldots, b_l).$$

Theorem 18.6 ([31]) *For any generative class* $(\mathbf{D}_0; \preceq)$ *of language* Σ *and for any formula* φ *satisfied in a* $(\mathbf{D}_0; \preceq)$*-generic structure, there is a* φ*-adapted generative class* $(\mathbf{D}'_0; \preceq')$ *such that* $\mathbf{D}'_0 \sim \mathbf{D}_0$.

18.5 Structural Diagrams and Canonical Structures

Definition 18.13 [31] Let $(\mathbf{D}_0; \preceq)$ be a generative class in a language Σ, $\Phi(A) \in \mathbf{D}_0$.

The diagram $\Phi(A)$ is called *structural* if it satisfies the following modification of the local realizability property:

(vi$'$) if $\Phi(A) \vdash \exists x\, \varphi(x)$ then there is a constant term $t(\bar{a})$, $\bar{a} \in A$, such that $\Phi(A) \vdash \varphi(t(\bar{a}))$.

Theorem 18.7 ([31]) *For any diagram $\Phi(A) \in \mathbf{D}_0$, $A \neq \emptyset$, the following conditions are equivalent:*

(1) *$\Phi(A)$ is structural;*

(2) *there exists a structure $\mathcal{M}$ consisting of some constant terms in the language $\Sigma \cup A$ and such that $\mathcal{M} \models \Phi(A)$.*

By Theorem 18.7, any structural diagram $\Phi(A)$, for $A \neq \emptyset$, defines the algebra with the universe $N/\sim$ being the restriction of $\mathcal{N}/\sim$ to the functional sublanguage and *finitely* generated by A (relative constant symbols in Σ). At the same time, for quantifier-free diagrams, this condition is sufficient:

Corollary 18.1 [31] *Any quantifier-free diagram $\Phi(A) \in \mathbf{D}_0$ is structural.*

Theorem 18.7 also implies

Corollary 18.2 [31] *If a generative class $(\mathbf{D}_0; \preceq)$ consists of structural diagrams then the $(\mathbf{D}_0; \preceq)$-generic structure is isomorphic to a union of representations of canonical structures for diagrams in $\mathbf{D}_0$.*

Corollary 18.3 [31] *If $(\mathbf{D}_0; \preceq)$ is a quantifier-free generative class then the $(\mathbf{D}_0; \preceq)$-generic structure is isomorphic to a union of representations of canonical structures for diagrams in $\mathbf{D}_0$.*

Definition 18.14 [31] A diagram $\Phi(A) \in \mathbf{D}_0$ is called *self-structural* if $A \neq \emptyset$ and $\Phi(A)$ satisfies the following:

(vi$''$) if $\Phi(A) \vdash \exists x\, \varphi(x)$ then there is an element $a \in A$ such that $\Phi(A) \vdash \varphi(a)$.

Proposition 18.1 ([31]) *For any diagram $\Phi(A) \in \mathbf{D}_0$, $A \neq \emptyset$, the following conditions are equivalent:*

(1) *$\Phi(A)$ is self-structural;*

(2) *if $\Phi(A) \vdash \exists x\, \varphi(x)$ then $\Phi(A) \vdash \exists x \left(\varphi(x) \wedge \bigvee\limits_{a \in A} (x \approx a)\right)$;*

(3) *there exists a structure $\mathcal{A}$ having the universe A and such that $\mathcal{A} \models \Phi(A)$.*

Proposition 18.1 implies

Corollary 18.4 [31] *A quantifier-free diagram $\Phi(A) \in \mathbf{D}_0$, where $A \neq \emptyset$, is self-structural if and only if for the sublanguage Σ_F consisting of all functional symbols in Σ, there exists an algebra $\mathcal{A}$ having the universe A and such that $\mathcal{A} \models \Phi(A)|_{\Sigma_F}$.*

Corollary 18.5 [31] *Any quantifier-free diagram $\Phi(A) \in \mathbf{D}_0$ in a relational language, where $A \neq \emptyset$, is self-structural.*

Corollary 18.6 [31] *If a generative class $(\mathbf{D}_0; \preceq)$, in a language Σ, consists of a diagram $\Phi_0(\emptyset)$ and of self-structural diagrams (over non-empty sets) then $(\mathbf{D}_0; \preceq)$ is φ-adapted for any formula φ in the language Σ.*

Proposition 18.2 ([31]) *Let $\mathcal{M}$ be a $(\mathbf{D}_0; \preceq)$-generic structure, for a quantifier-free generative class $(\mathbf{D}_0; \preceq)$, such that its restriction to the sublanguage Σ_F consisting of all functional symbols in Σ is a locally finite algebra. Then $(\mathbf{D}_0; \preceq)$ is domination-equivalent to a generative class $(\mathbf{D}'_0; \preceq')$ consisting of a diagram $\Phi_0(\emptyset)$ and only self-structural diagrams $\Phi(A)$ for $A \neq \emptyset$.*

18.5.1 Generative Classes for Finite Structures

Let $(\mathbf{D}_0; \preceq)$ is a generative class. If $\Phi(A)$ is a diagram in $\mathbf{D}_0$ having a proper extension in $\mathbf{D}_0$ and such that $A \neq \emptyset$ then removing $\Phi(A)$ and all its copies from $(\mathbf{D}_0; \preceq)$ we get the generative class $(\mathbf{D}_0; \preceq) \setminus \{\Phi(A)\}$ domination-equivalent to $(\mathbf{D}_0; \preceq)$. At the same time, staying in the family of generative classes, we can not remove from $(\mathbf{D}_0; \preceq)$ the diagram $\Phi_0(\emptyset)$ as well as, preserving the domination-equivalence, a *maximal* diagram $\Phi(A)$ such that it does not have proper extensions in $(\mathbf{D}_0; \preceq)$. Clearly, maximal diagrams $\Phi(A)$ are self-structural and, moreover, describe $(\mathbf{D}_0; \preceq)$-generic structures with the universe $A \cup C$, where C is the set of constant symbols. It means that $\Phi_0(\emptyset)$ and copies of $\Phi(A)$ form a *minimal* generative class being domination equivalent to $(\mathbf{D}_0; \preceq)$. Moreover, the maximal diagram $\Phi(A)$ producing a representation, with the universe $A \cup C$, of $\Phi(A)$-canonical structure can be chosen quantifier-free.

Thus we get

Theorem 18.8 ([31]) *For a generative class $(\mathbf{D}_0; \preceq)$ with a language having a finite set C of pairwise distinct constants, the following conditions are equivalent:*

(1) *the $(\mathbf{D}_0; \preceq)$-generic structure is finite;*

(2) *$(\mathbf{D}_0; \preceq)$ has maximal diagrams;*

(3) *$(\mathbf{D}_0; \preceq)$ is domination-equivalent to a minimal generative class consisting of a diagram $\Phi_0(\emptyset)$ and of copies of a self-structural diagram $\Phi(A)$;*

(4) *the $(\mathbf{D}_0; \preceq)$-generic structure is isomorphic, for a quantifier-free diagram $\Phi(A)$, to a representation, with the universe $A \cup C$, of $\Phi(A)$-canonical structure.*

18.6 Applications of Syntactic Generic Constructions

Syntactic generic constructions were used to get a classification of countable models of complete theories. This classification is divided in classifications of Ehrenfeucht

theories (i.e., theories with finitely many but more than one countable models) [33, 38], of small theories (i.e., countable theories with countably many types) [35, 37, 38], and of theories with continuum many types [26, 38].

Using syntactic generic construction for Ehrenfeucht theories and modifying Hrushovski–Herwig construction [13], the known Lachlan's problem on the existence of stable (i.e., without infinite definable orders) Ehrenfeucht theory was solved:

Theorem 18.9 *For any natural $n \geq 3$ there is a stable Ehrenfeucht theory with exactly n countable models.*

Illustrating the syntactic approach for generic constructions, spectra and structures of models of cubic theories are described [32, 39].

18.7 Classes of Structures and Their Generic Limits

18.7.1 Neighbourhoods

Let $(\mathbf{D}_0; \preceq)$ be a self-sufficient generic class of a language Σ and $\Phi(A)$ be a diagram in $\mathbf{D}_0$. Denote by $[\![\Phi(A)]\!]$ the class $\{\mathcal{M} \mid \mathcal{M} \models \Phi(A)\}$ of all *potential* structures $\mathcal{M}$, of the language Σ, containing A and satisfying $\Phi(A)$.

By the definition, the classes $[\![\Phi(A)]\!]$ are neighborhoods for structures satisfying $\Phi(A)$. Some variations of these neighborhoods and related topologies introduced in [14].

Proposition 18.3 ([40]) *For any consistent diagrams $\Phi(A), \Psi(B) \in \mathbf{D}_0$, $\Phi(A) \subseteq \Psi(B)$ if and only if $[\![\Phi(A)]\!] \supseteq [\![\Psi(B)]\!]$, moreover, $\Phi(A) \subset \Psi(B)$ if and only if $[\![\Phi(A)]\!] \supset [\![\Psi(B)]\!]$.*

By the definition, classes $[\![\Phi(A)]\!]$ are closed under isomorphisms and it is natural to consider isomorphism types instead of structures in $[\![\Phi(A)]\!]$.

For a diagram $\Phi(A) \in \mathbf{D}_0$ we denote

- by $[\![\Phi(A)]\!]_{\mathrm{Iso}}$ the quotient of $[\![\Phi(A)]\!]$ by the isomorphism relation;
- by $[\![\Phi(A)]\!]_{\mathrm{Iso},\lambda}$ the restriction of $[\![\Phi(A)]\!]_{\mathrm{Iso}}$ to the isomorphism types of structures of cardinality $\leq \lambda$;
- by $[\![\Phi(A)]\!]_{\mathrm{Iso},h}$ the restriction of $[\![\Phi(A)]\!]_{\mathrm{Iso}}$ to the isomorphism types of ω-homogeneous structures;
- by $[\![\Phi(A)]\!]_{\mathrm{Iso},h,\lambda}$ the restriction of $[\![\Phi(A)]\!]_{\mathrm{Iso},h}$ to the isomorphism types of structures of cardinality $\leq \lambda$.

By the definition, we have the following relations:

$$[\![\Phi(A)]\!]_{\mathrm{Iso}} \supseteq [\![\Phi(A)]\!]_{\mathrm{Iso},\lambda}, \tag{18.1}$$

$$[\![\Phi(A)]\!]_{\mathrm{Iso}} \supseteq [\![\Phi(A)]\!]_{\mathrm{Iso},h}, \tag{18.2}$$

$$[\![\Phi(A)]\!]_{\mathrm{Iso},\lambda} \cap [\![\Phi(A)]\!]_{\mathrm{Iso},h} = [\![\Phi(A)]\!]_{\mathrm{Iso},h,\lambda}). \tag{18.3}$$

The inclusion (18.1) is strict if and only if $[\![\Phi(A)]\!]$ contains infinite structures or $\Phi(A)$ is satisfied by a finite structure having a cardinality $> \lambda$. The strict inclusion (18.2) means that $\Phi(A)$ is satisfied by a structure $\mathscr{M}$ having a theory $\mathrm{Th}(\mathscr{M})$ with models which are not ω-homogeneous. In particular, that theory is not totally categorical.

Since there are generic structures of a cardinality $\leq \max\{|\Sigma|, \omega\}$ and each structure of cardinality $\leq \max\{|\Sigma|, \omega\}$ has a ω-homogeneous elementary extension preserving the cardinality $\leq \max\{|\Sigma|, \omega\}$, by Proposition 18.3 and relations (18.1)–(18.3) we have

Corollary 18.7 [40] *For any consistent diagrams $\Phi(A), \Psi(B) \in \mathbf{D}_0$, the following conditions are equivalent:*

(1) $[\Phi(A)]^A_{A'} \subseteq \Psi(B)$ *for* $A' \subseteq B$*;*
(2) $[\![\Phi(A)]\!]_{\mathrm{Iso}} \supseteq [\![\Psi(B)]\!]_{\mathrm{Iso}}$*;*
(3) $[\![\Phi(A)]\!]_{\mathrm{Iso},\max\{|\Sigma|,\omega\}} \supseteq [\![\Psi(B)]\!]_{\mathrm{Iso},\max\{|\Sigma|,\omega\}}$*;*
(4) $[\![\Phi(A)]\!]_{\mathrm{Iso},h} \supseteq [\![\Psi(B)]\!]_{\mathrm{Iso},h}$*;*
(5) $[\![\Phi(A)]\!]_{\mathrm{Iso},h,\max\{|\Sigma|,\omega\}} \supseteq [\![\Psi(B)]\!]_{\mathrm{Iso},h,\max\{|\Sigma|,\omega\}}$.

At the same time, strict inclusions $[\Phi(A)]^A_{A'} \subset \Psi(B)$ may not imply $[\![\Phi(A)]\!]_{\mathrm{Iso}} \supset [\![\Psi(B)]\!]_{\mathrm{Iso}}$. For instance, if $\Phi(A)$ contains a totally categorical countable theory T, of language Σ, then $[\![\Phi(A)]\!]_{\mathrm{Iso}}$ and $[\![\Psi(B)]\!]_{\mathrm{Iso}}$ contain unique element for every cardinality $\lambda \geq \omega$, and all models are infinite. Thus, $[\![\Phi(A)]\!]_{\mathrm{Iso}} = [\![\Psi(B)]\!]_{\mathrm{Iso}}$ as well as $[\![\Phi(A)]\!]_{\mathrm{Iso},\omega} = [\![\Psi(B)]\!]_{\mathrm{Iso},\omega}$, $[\![\Phi(A)]\!]_{\mathrm{Iso},h} = [\![\Psi(B)]\!]_{\mathrm{Iso},h}$ and $[\![\Phi(A)]\!]_{\mathrm{Iso},h,\omega} = [\![\Psi(B)]\!]_{\mathrm{Iso},h,\omega}$.

As classes $[\![\Phi(A)]\!]$ are neighborhoods for structures satisfying $\Phi(A)$, classes $[\![\Phi(A)]\!]_{\mathrm{Iso}}$, $[\![\Phi(A)]\!]_{\mathrm{Iso},\lambda}$, $[\![\Phi(A)]\!]_{\mathrm{Iso},h}$, $[\![\Phi(A)]\!]_{\mathrm{Iso},h,\lambda}$ form neighborhoods for corresponding isomorphism types.

Let $\Phi(A), \Psi(B), \mathrm{X}(C), \Theta(D) \in \mathbf{D}_0$, and $\Theta(D)$ be an amalgam of $\Psi(B)$ and $\mathrm{X}(C)$ over $\Phi(A)$. The neighborhood $[\![\Theta(D)]\!]$, $[\![\Theta(D)]\!]_{\mathrm{Iso}}$ and its restrictions are called *amalgams* of corresponding neighborhoods of $\Psi(B)$ and $\mathrm{X}(C)$ over neighborhoods of $\Phi(A)$. By Corollary 18.7, we have the following inclusions for Iso-neighborhoods: $[\![\Theta(D)]\!]_{\mathrm{Iso}} \subseteq [\![\Psi(B)]\!]_{\mathrm{Iso}} \cap [\![\mathrm{X}(C)]\!]_{\mathrm{Iso}}$, $[\![\Psi(B)]\!]_{\mathrm{Iso}} \cup [\![\mathrm{X}(C)]\!]_{\mathrm{Iso}} \subseteq [\![\Phi(A)]\!]_{\mathrm{Iso}}$. Similar inclusions hold for λ- h- and (h,λ)-neighborhoods. Thus, for the generic class $(\mathbf{D}_0; \preceq)$, there are partially ordered, by inclusion, downward directed sets A_{Iso}, $A_{\mathrm{Iso},\lambda}$, $A_{\mathrm{Iso},h}$, $A_{\mathrm{Iso},h,\lambda}$ (with the greatest elements corresponding to $\Phi_0(\emptyset)$) of corresponding neighborhoods, each of which consists of isomorphism types of potential structures and contains the isomorphism type of $(\mathbf{D}_0; \preceq)$-generic structure $\mathscr{M}$ (for $\lambda \geq |M|$).

Below we consider limits with respect to these posets, i. e., intersections

$$\begin{gathered}\bigcap_{\Phi(A)\in\mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso}}, \quad \bigcap_{\Phi(A)\in\mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},\lambda}, \\ \bigcap_{\Phi(A)\in\mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},h}, \quad \bigcap_{\Phi(A)\in\mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},h,\lambda}.\end{gathered} \tag{18.4}$$

Since these intersections can be represented as unions for *completions* $\mathbf{D}'_0$ of $\mathbf{D}_0$, obtained by adding of complete theories T to diagrams in $\mathbf{D}_0$, we can assume that $\mathbf{D}_0$ is already *complete*, i. e., all diagrams in $\mathbf{D}_0$ contain a complete theory T. Moreover, since all neighborhoods are singletons for complete classes having finite generic structures, we shall additionally assume that $(\mathbf{D}_0; \preceq)$-generic structures are infinite. In such a case, intersections (18.4) contain isomorphism types of models of T realizing sets of types $[\Phi(A)]^A_X$ for $\Phi(A) \in \mathbf{D}_0$. Thus, the problem of description for intersections (18.4) is reduced to the problem of description for models of T realizing a given set of types.

By Löwenheim–Skolem theorem, intersections $\bigcap\limits_{\Phi(A)\in\mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso}}$ have models of unbounded cardinalities, are not sets, and, in particular, can not be singletons. By same reason, intersections $\bigcap\limits_{\Phi(A)\in\mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},h,\lambda}$ are not singletons for $\lambda > \max\{|\Sigma|, \omega\}$, as well as we have at least two structures in $\bigcap\limits_{\Phi(A)\in\mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},h,\omega}$ if a $(\mathbf{D}_0; \preceq)$-generic structure $\mathcal{M}$ is at most countable and there are types in $S(T)$ that are not realized by diagrams in $\mathbf{D}_0$. At the same time, if a countable $(\mathbf{D}_0; \preceq)$-generic structure $\mathcal{M}$ is saturated then $\bigcap\limits_{\Phi(A)\in\mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},h,\omega}$ is a singleton containing the isomorphism type of $\mathcal{M}$, since there exists unique homogeneous structure realizing exactly given set of types. Thus we have

Theorem 18.10 ([40]) *Let $(\mathbf{D}_0; \preceq)$ be a complete self-sufficient generic class and $\mathcal{M}$ be a countable $(\mathbf{D}_0; \preceq)$-generic structure. The following conditions are equivalent:*

(1) $\bigcap\limits_{\Phi(A)\in\mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},h,\omega}$ *is a singleton (consisting of the isomorphism type of $\mathcal{M}$);*

(2) *the structure $\mathcal{M}$ is saturated.*

Theorem 18.10 means that intersection of neighborhoods for isomorphism types of countable homogeneous models contains unique limit point if and only if this point, being a generic limit, corresponds to the saturated model.

Definition 18.15 ([2]) A type $p(\bar{x}) \in S(T)$ is said to be *powerful* in a theory T if every model $\mathcal{M}$ of T realizing p also realizes every type $q \in S(T)$, that is, $\mathcal{M} \models S(T)$.

By Theorem 18.10, the set $Z = \bigcap\limits_{\Phi(A)\in\mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},\omega}$ is a singleton if and only if it consists only of isomorphism type of a saturated structure $\mathcal{M}$. Since there exists a countable saturated structure $\mathcal{M}$ if and only if the theory $T = \mathrm{Th}(\mathcal{M})$ is small (i. e., $|S(T)| = \omega$), and all countable models of small T are isomorphic to unions of elementary chains of prime models over tuples [38] (these unions are either again prime models over tuples or *limit* models), we have $|Z| = 1$ if and only if T is either countably categorical or T does not have powerful types and there is unique, up to isomorphism, union of elementary chains of prime models over tuples which realizes all types in $S(T)$.

Indeed, if T is countably categorical then $|Z| = 1$. If there are no powerful types, the uniqueness of considered elementary chains implies that Z is a singleton consisting of isomorphism type of saturated model. Conversely, if T is not countably categorical and there is a powerful type p then the isomorphism type of model $\mathscr{M}_p$, prime over a realization of p, belongs to Z. We have $\mathscr{M}_p \not\succ \mathscr{M}$, isomorphism types for $\mathscr{M}_p$ and $\mathscr{M}$ both belong to Z, whence $|Z| \geq 2$. We get again the inequality $|Z| \geq 2$ if there are at least two non-isomorphic unions of elementary chains of prime models over tuples. Thus we have

Theorem 18.11 ([40]) *Let* $(\mathbf{D}_0; \preceq)$ *be a complete self-sufficient generic class and* $\mathscr{M}$ *be a countable* $(\mathbf{D}_0; \preceq)$*-generic structure. The following conditions are equivalent:*

(1) $\bigcap\limits_{\Phi(A)\in\mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},\omega}$ *is a singleton (consisting of the isomorphism type of* $\mathscr{M}$*);*

(2) *the structure* $\mathscr{M}$ *is saturated and the theory* $T = \mathrm{Th}(\mathscr{M})$ *is either countably categorical or there are no powerful types of* T *and there is unique, up to isomorphism, union of elementary chains of prime models over tuples which realizes all types in* $S(T)$*.*

18.7.2 Classification of Countable Models of Complete Theories for Neighborhood Structures

Classification of countable models of complete theories [36, 38] allows to describe intersections (1) for countable models as well as links between them. We consider the description for theories T with finite Rudin–Keisler preorders. It is known [36, 38] that every countable model of T is either prime over a finite set or limit over a type. Recall that the preordered set $\mathrm{CM}(T)$ consists of all isomorphism types of countable models of T and the Rudin–Keisler preorder $\leq_{\mathrm{RK}}$ defined by the following way: for isomorphism types $\mathbf{M}_1$ and $\mathbf{M}_2$, $\mathbf{M}_1 \leq_{\mathrm{RK}} \mathbf{M}_2$ if and only if for models $\mathscr{M}_1 \in \mathbf{M}_1$ and $\mathscr{M}_2 \in \mathbf{M}_2$, $\mathscr{M}_2$ realizes all types in $S(\emptyset)$ being realized in $\mathscr{M}_1$. Thus the Rudin–Keisler preorders $\leq_{\mathrm{RK}}$ define links between models reflecting corresponding links between sets of realized types. Moreover, the following theorems hold:

Theorem 18.12 ([38]) *The set of isomorphism types for countable model of an arbitrary small theory* T *with a finite Rudin–Keisler preorder has a partition into classes* $\widetilde{\mathbf{PM}} \cup \mathbf{L}$ *forming a partially ordered set* $\mathrm{CM}(T)$ *with a least and a greatest elements, where* $\widetilde{\mathbf{PM}}$ *are sets of isomorphism types for prime models over tuples,* $\mathbf{L}$ *are isomorphism types of limit models related to* $\widetilde{\mathbf{PM}}$*. Each class* $\widetilde{\mathbf{PM}} \cup \mathbf{L}$ *contains at most one isomorphism type of homogeneous model. Moreover, the following conditions are satisfied:*

(a) *the least element of* $\mathrm{CM}(T)$ *is a singleton and contains the isomorphism type of prime model;*

(b) *the greatest element of* $\mathrm{CM}(T)$ *contains the isomorphism type of the homogeneous saturated model, which is limit if* $|\mathrm{CM}(T)| > 1$*;*

(c) *each element* $\widetilde{\mathbf{PM}} \cup \mathbf{L}$ *is either a singleton and does not contain isomorphism types of limit models, or the sets* $\widetilde{\mathbf{PM}}$ *and* $\mathbf{L}$ *are non-empty and, having a homogeneous model, the isomorphism type of that homogeneous model belongs to* $\mathbf{L}$.

Theorem 18.13 ([38]) *For any finite partially ordered set* $\langle X; \leq \rangle$ *with a least and a greatest elements, there is a small theory* T *such that the restriction* $\mathrm{CM}(T)_h$ *of* $\mathrm{CM}(T)$ *to the set of isomorphism types of homogeneous models of* T *is isomorphic to* $\langle X; \leq \rangle$.

Let T be a generic theory, $\mathbf{D}$ be the class of all complete self-sufficient generic classes $\mathbf{D}_0$ such that T is $(\mathbf{D}_0; \preceq)$-generic for appropriate $\preceq$. Denote by $B(T)_{\mathrm{Iso}}$ the structure $\langle B(T)_{\mathrm{Iso}}; \leq \rangle$, where $B(T)_{\mathrm{Iso}}$ consists of all intersections $\bigcap\limits_{\Phi(A) \in \mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso}}$ for $\mathbf{D}_0 \in \mathbf{D}$, and $\leq = \subseteq$. Similarly we denote by $B(T)_{\mathrm{Iso},\lambda}$ the structure $\langle B(T)_{\mathrm{Iso},\lambda}; \leq \rangle$, where $B(T)_{\mathrm{Iso},\lambda}$ consists of all intersections $\bigcap\limits_{\Phi(A) \in \mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},\lambda}$; by $B(T)_{\mathrm{Iso},h}$ the structure $\langle B(T)_{\mathrm{Iso},h}; \leq \rangle$, where $B(T)_{\mathrm{Iso},h}$ consists of all intersections $\bigcap\limits_{\Phi(A) \in \mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},h}$, and $B(T)_{\mathrm{Iso},h,\lambda}$ is the structure $\langle B(T)_{\mathrm{Iso},h,\lambda}; \leq \rangle$, where $B(T)_{\mathrm{Iso},h,\lambda}$ consists of all intersections $\bigcap\limits_{\Phi(A) \in \mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},h,\lambda}$, $\mathbf{D}_0 \in \mathbf{D}$, $\leq = \subseteq$.

Note that for any $(\mathbf{D}_0; \preceq)$-generic structure $\mathscr{M}$ and for any $(\mathbf{D}'_0; \preceq')$-generic structure $\mathscr{M}'$, where $\mathbf{D}_0, \mathbf{D}'_0 \in \mathbf{D}$ and $|M| \leq |M'| \leq \lambda$, $\mathscr{M}$ is elementary embeddable in $\mathscr{M}'$ if and only if all types realized in $\mathscr{M}$ are also realized in $\mathscr{M}'$. It means that $\bigcap\limits_{\Phi(A) \in \mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},h,\lambda} \supseteq \bigcap\limits_{\Phi'(A') \in \mathbf{D}'_0} [\![\Phi'(A')]\!]_{\mathrm{Iso},h,\lambda}$. Thus, for any countable complete theory T without finite models, we have a canonical isomorphism ϕ between the poset $\mathrm{CM}(T)_h$ and the poset $B(T)_{\mathrm{Iso},h,\omega}$ mapping the isomorphism type of countable $(\mathbf{D}_0; \preceq)$-generic homogeneous model $\mathscr{M}$ to $\bigcap\limits_{\Phi(A) \in \mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},h,\omega}$. Having the isomorphism $\phi\colon \mathrm{CM}(T)_h \xrightarrow{\sim} B(T)_{\mathrm{Iso},h,\omega}$, by Theorem 18.13, we get

Theorem 18.14 ([40]) *For any finite partially ordered set* $\langle X; \leq \rangle$ *with a least and a greatest elements, there is a small theory* T *such that the structure* $B(T)_{\mathrm{Iso},h,\omega}$ *is isomorphic to* $\langle X; \leq \rangle$.

Using Theorem 18.12 the isomorphism ϕ produces the isomorphism

$$\psi : \mathrm{CM}(T)_h \xrightarrow{\sim} B(T)_{\mathrm{Iso},\omega}$$

such that $\psi(x) = \bigcap\limits_{\Phi(A) \in \mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},\omega}$ for $x \in \mathrm{CM}(T)_h$ with

$$\phi(x) = \bigcap_{\Phi(A) \in \mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},h,\omega}.$$

Thus we obtain the following

Theorem 18.15 ([40]) *For any finite partially ordered set $\langle X; \leq \rangle$ with a least and a greatest elements, there is a small theory T such that the structure $B(T)_{\mathrm{Iso},\omega}$ is isomorphic to $\langle X; \leq \rangle$.*

18.7.3 Spectra of Theories and Cardinalities Of Intersections of Neighborhoods

In this section we consider complete countable theories T without finite models.

As usual we denote by $I(T, \lambda)$ the *spectrum function* giving the number of non-isomorphic models $\mathscr{M}$ of T with $|\mathscr{M}| = \lambda$. It is known that $I(T, \omega) \in (\omega \setminus \{0, 2\}) \cup \{\omega, \omega_1, 2^\omega\}$.

Clearly, if T has a prime model $\mathscr{M}_0$ being $(\mathbf{D}_0; \preceq)$-generic for a complete self-sufficient class $(\mathbf{D}_0; \preceq)$, then $\left| \bigcap\limits_{\Phi(A) \in \mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},\omega} \right| = I(T, \omega)$. Thus this cardinality belongs to $(\omega \setminus \{0, 2\}) \cup \{\omega, \omega_1, 2^\omega\}$.

More generally, for any $\lambda \geq \omega$ we have

$$\left| \bigcap_{\Phi(A) \in \mathbf{D}_0} [\![\Phi(A)]\!]_{\mathrm{Iso},\lambda} \right| = \sum_{\omega \leq \mu \leq \lambda} I(T, \mu). \tag{18.5}$$

Using correspondences, as in Sect. 18.3, for homogeneous models and neighborhoods for generic classes, the equality (18.5) can be adapted for an arbitrary generic class.

Let β be an ordinal number, λ a limit ordinal number, and κ a cardinal number. We define $\Im_\beta(\kappa)$ recursively: $\Im_0(\kappa) = \kappa$, $\Im_{\beta+1}(\kappa) = 2^{\Im_\beta(\kappa)}$, and $\Im_\lambda(\kappa) = \bigcup\limits_{\beta < \lambda} \Im_\beta(\kappa)$.

Theorem 18.16 ([12, 28]) *For any complete theory T in a countable language and without finite models the uncountable spectrum $\aleph_\alpha \mapsto I(T, \aleph_\alpha)$ $(\alpha > 0)$ is the minimum of the map $\aleph_\alpha \mapsto 2^{\aleph_\alpha}$ and one of the following maps:*

(1) $2^{\aleph_\alpha}$;
(2) $\Im_{d+1}(|\alpha + \omega|)$ *for some* $d \in [\omega, \omega_1)$;
(3) $\Im_{d-1}(|\alpha + \omega|^{2^{\aleph_0}})$ *for some finite* $d > 0$;
(4) $\Im_{d-1}(|\alpha + \omega|^{\aleph_0} + \Im_2(\aleph_0))$ *for some finite* $d > 0$;
(5) $\Im_{d-1}(|\alpha + \omega| + \Im_2(\aleph_0))$ *for some finite* $d > 0$;
(6) $\Im_{d-1}(|\alpha + \omega|^{\aleph_0})$ *for some finite* $d > 0$;
(7) $\Im_{d-1}(|\alpha + \omega| + 2^{\aleph_0})$ *for some finite* $d > 1$;
(8) $\Im_{d-1}(|\alpha + \omega|)$ *for some finite* $d > 0$;
(9) $\Im_{d-2}(|\alpha + \omega|^{|\alpha+1|})$ *for some finite* $d > 1$;
(10) *identically* $\Im_2(\aleph_0)$;
(11) $|\alpha|$ *for infinite* α *and, for finite* α, $|(\alpha + 1)^n / G| - |\alpha^n / G|$ *for some finite* $n > 1$ *and some normal subgroup* G *of* $\mathrm{Sym}(n)$;
(12) *identically* 1.

Each of these 12 *possibilities occurs as the spectrum of some complete theory in a countable language.*

By Theorem 18.16 we have an explicit description for the sum $\sum_{\omega \le \mu \le \lambda} I(T, \mu)$ in (18.5).

18.8 A Path to Institutional Generalization of Generic Limits

The Theory of Institutions was introduced by Goguen and Burstall [11] as a systematic way of studying logical systems using category theory. Institution theory is an abstract mathematical study of formal logical systems that is not committed to any particular concrete logical system. The definition of Institutions takes into account both syntax and semantics as well as the relationship between them. Furthermore the high level of abstraction allows the accommodation of not only well-established logical systems but also unconventional systems.

The aim of this section is to presenting the basic definitions and results of Institution Theory.

18.8.1 The Concept of Institution Theory

We shall start with the definitions, which are the basic tools employed in our demonstration. As noted in [11], a *specification* provides a mathematical theory of the behaviour of a program and if a theory consists of all the sentences that are true under that behaviour then it is important to define the fundamental properties of theories over an arbitrary institution.

Definition 18.16 ([11]) An *Institution* $\mathscr{I} = \left(\mathbf{Sig}^{\mathscr{I}}, \mathbf{Sen}^{\mathscr{I}}, \mathbf{Mod}^{\mathscr{I}}, \models^{\mathscr{I}}\right)$ consists of:

1. A category $\mathbf{Sig}^{\mathscr{I}}$, whose objects are called *signatures*,
2. A functor $\mathbf{Sen}^{\mathscr{I}} : \mathbf{Sig}^{\mathscr{I}} \to \mathbf{Set}$ giving for each signature a set whose elements are called *sentences* over that signature,
3. A functor $\mathbf{Mod}^{\mathscr{I}} : \left(\mathbf{Sig}^{\mathscr{I}}\right)^{op} \to \mathbf{CAT}$ giving for each signature Σ a category whose objects are called Σ-*models* and whose arrows are called Σ-*morphisms*, and
4. A relation $\models^{\mathscr{I}}_{\Sigma} \subseteq \left|\mathbf{Mod}^{\mathscr{I}}(\Sigma)\right| \times \mathbf{Sen}^{\mathscr{I}}(\Sigma)$ for each $\Sigma \in \left|\mathbf{Sig}^{\mathscr{I}}\right|$, called Σ-*satisfaction*

such that for each morphism $\phi : \Sigma \to \Sigma'$ in $\mathbf{Sig}^{\mathscr{I}}$, the *satisfaction condition*

$$M' \models^{\mathscr{I}}_{\Sigma'} \mathbf{Sen}^{\mathscr{I}}(\phi)(\rho) \quad \text{if and only if} \quad \mathbf{Mod}^{\mathscr{I}}(\phi)(M') \models^{\mathscr{I}}_{\Sigma} \rho$$

holds for each $M' \in \left|\mathbf{Mod}^{\mathscr{I}}\right|$ and $\rho \in \mathbf{Sen}^{\mathscr{I}}(\Sigma)$.

Example 18.1 (*Propositional Logic* [11]) As example we will give **PL** the institution of Propositional Logic.

- The category $\mathbf{Sig}^{\mathbf{PL}}$ has as objects sets of propositional variables and the arrows are the functions between them.
- A signature morphism σ is a mapping between the propositional variables.
- The functor $\mathbf{Sen}^{\mathbf{PL}}$ acts and mapping for each signature Σ the $\mathbf{Sen}^{\mathbf{PL}}(\Sigma)$ the set of propositional variables from Σ and connectives for conjunction, disjunction, implication and negation.
- The $\mathbf{Sen}^{\mathbf{PL}}(\sigma)$ is the extension of σ to all formulas.
- Models of Σ are truth valuations, i.e. mappings from Σ into the standard Boolean algebra $Bool = \{0, 1\}$.
- A model morphism between Σ-models M and m' exists iff $M(p) \leq M'(p)$.
- Given $\sigma : \Sigma_1 \to \Sigma_2$ and a Σ_2-model $M_2 : \Sigma_2 \to Bool$, then the reduct $M_2 \restriction_\sigma$ is the composition $M_2 \circ \sigma$.
- And $M \models^{PL}_{\Sigma} \phi$ if and only if ϕ evaluates 1 under the standard extension of M to all formulas.

18.8.1.1 The Galois Connection

Regarding Goguen's fundamental paper, we observed that one of the most useful results is duality between theories and models. In institutional framework a Σ-theory is a set of Σ-sentences, and respectively a Σ-model class is a class of Σ-models. A more formal definition follows.

Definition 18.17 ([11]) Let $\mathscr{I}$ be a fixed but arbitrary institution. Then

1. A Σ-*presentation* is a pair $\langle \Sigma, E \rangle$, where Σ is a signature and E is collection of Σ-sentences.
2. A Σ-model M *satisfies* a presentation $\langle \Sigma, E \rangle$ if it satisfies each sentence in E; we write $M \models E$ in this case.
3. Given a collection E of Σ-sentences, let E^* be the collection of all Σ-models that satisfy each sentence in E.
4. Given a collection M of Σ-models, let M^* be the collection of all Σ-sentences that are satisfied by each model in M; also let M^* denote $\langle \Sigma, M^* \rangle$ called the *theory* of M.
5. The *closure* of a collection E of Σ-sentences is E^{**}, denoted $E^\bullet$.
6. A collection E of Σ-sentences is *closed* if and only if $E = E^\bullet$.
7. A Σ-*theory* is a presentation $\langle \Sigma, E \rangle$ such that E is closed.
8. The Σ-theory *presented* by a presentation $\langle \Sigma, E \rangle$ is $\langle \Sigma, E^\bullet \rangle$.
9. A Σ-sentence e is *semantically entailed* by a collection E of Σ-sentences, written $E \models e$, if and only if $e \in E^\bullet$.

Proposition 18.4 *For any institution $\mathscr{I}$ and signature Σ, the closed Σ-theories, and the closed Σ-model classes, are complete lattices under inclusion. Moreover there is a dual isomorphism between these complete lattices.*

18.8.1.2 Internal Logic

The semantics of Boolean connectives as well quantifiers have been introduced by Tarrlecki [41].

Definition 18.18 (*Internal Boolean Connectives* [4]) Let Σ a signature in an institution then:

- the Σ-sentence ϕ is a (semantic) negation of ψ when $\phi^* = |\mathbf{Mod}(\Sigma)| \setminus \psi^*$;
- the Σ-sentence ϕ is the (semantic) conjunction of the Σ-sentences ψ_1 and ψ_2 when $\phi^* = \psi_1^* \cap \psi_2^*$.

Remark 18.2 The least Boolean connectives, such as disjunction $\vee$, implication $\Rightarrow$, equivalence $\Leftrightarrow$, etc. can be derived as usually from negations and conjunctions.

Definition 18.19 (*Internal Quantifiers* [4]) For any signature morphism $\chi : \Sigma \to \Sigma'$ in an arbitrary institution:

- a Σ-sentence ϕ is a (semantic) *existential χ-quantification* of a χ-sentence ψ when $\phi^* = (\psi^*)\restriction_\chi$; in this case we write ϕ as $\exists \chi \psi$;
- a Σ-sentence ϕ is a (semantic) *universal χ-quantification* of a χ-sentence ϕ when $\phi^* = |\mathbf{Mod}(\Sigma)| \setminus \left(\left|\mathbf{Mod}(\Sigma')\right| \setminus \psi^*\right)\restriction_\chi$; in this case we write ϕ as $\forall \chi \psi$.

In order to complete the generalizations at the level of Institution theory, it is necessary to define the *basic sentences* and the *representable signature morphisms*.

Definition 18.20 (*Basic sentences* [4]) Given a signature Σ in any institution, a Σ-sentence e is basic if there exists a Σ-model M_e, called the basic modal of e such that for each Σ-model M , $M \models_\Sigma e$ if and only if there exists a modal homomorphism $M_e \to M$. The sentence e is epic basic when the homomorphism $M_e \to M$ is an epi.

Definition 18.21 (*Representable signature morphims* [4]) In any institution, a signature morphism $\chi : \Sigma \to \Sigma'$ is representable if and only of there exists a Σ-model M_χ and a isomorphism i_χ of categories such that the following diagram commutes.

18.8.2 Institution-Independent Model Theory

Several important model theory methods and results have been developed in the abstract institutional framework. These results form the field of *Institution-independent model theory or Institutional model theory*.

The institutional approach to completeness of generalized first order logic at the level of a framework which is institutional-independent is a breakthrough for abstract model theory. This methodology allows us to obtain in a universal manner completeness results for several logical systems.

In [10] the writers have developed an institutional expression and proofing for completeness of infinitary first order logics. It is shown in [25] that there exists an institutional version of Gödel's Completeness Theorem and directions for generalized completeness in a branch of non-standard logical systems. Furthermore in [3], Downward Löwenheim–Skolem and Omitting Types Theorems are studied at the institutional framework.

In standard model theory, the method of diagrams is one of the most popular methods. Here we present it in the institutional framework.

Definition 18.22 (*The methods of diagrams* [29]) An institution $\mathscr{I}$ has diagrams when for each signature Σ and each Σ-model M there exists a signature Σ_M and a signature morphism $i_\Sigma(M) : \Sigma \to \Sigma_M$, functorial in Σ and M, and a set E_M of Σ_M-sentences such that $Mod(\Sigma_M, E_M)$ and the comma category $M/Mod(\Sigma)$ are naturally isomorphic, i.e., the diagram bellow commutes,

The signature morphism $i_\Sigma(M) : \Sigma \to \Sigma_M$ is called the elementary extension of Σ via M and the set E_M of Σ_M-sentences is called the diagram of the model M.

The institution-independent method of diagrams has been used widely in institutional model theory. For example, the co-limits of models, interpolation, definability, Tarski's elementary chain theorem, saturated models, axiomatizability [4] etc.

In the paper [29] Diaconescu, Stefaneas developed possible semantics at the categorical abstract model theoretic level provided by the institution theory framework. They also extended the Institution-independent method of Ultraproducts to possible worlds and proved a fundamental preservation theorem for abstract modal satisfaction.

The fundamental ultraproducts theorem gives a preservation property for the satisfaction condition by ultraproducts of models for a variety of logical systems, including unconventional ones for which such development was difficult to establish.

Formal proof systems also constitute a basic part of institution theory. There are several results as generation of freely proof systems, the freely addition of quantification to any proof system such that its sentence part has a syntax for quantifiers, compactness of the proof systems, automatic transfer of compactness [29] and recently the study of entailment topology of a proof system as a complete semantic [20].

18.8.3 Institutional Path

Ravi Rajani is developing [27] the theory of generic structures at a general, unified level but it is not strictly categorical. Furthermore, Kubis in [22] is developing a category-theoretic approach to homogeneous structures and in [21] introducing the Katetov functors which provides a universal way to construct Fraïssé limits.

Our goal is to unify all these different approaches under the Institution-Independent model theory. The institution theory gives us a tool to expand the main results of the generic constructions and generic limits in abstract logical systems.

Regarding the theory of generic limits and generic constructions for the standard first order logic we will work in the institutional first order logic framework, based on $\mathscr{D}$-first order fragment [10]. The first goal is to merge the institutional definition of diagrams with the standard theory of diagrams under an universal way. Furthermore, we will attempt to give abstract categorical conditions on partial ordering of a collection of Σ-diagrams with respect to Definition 18.3. Therefore we introduce the Σ_d-amalgamation property, the semantic existential *local realizability property* and the Σ_d-uniqueness property expanding Definition 18.3.

As far as cardinalities of generic structures as well as their languages can unbounded it is natural to generalize generic constructions for the restrictions $\mathscr{I}_h = \left(\mathbf{Sig}^{\mathscr{I}_h}, \mathbf{Sen}^{\mathscr{I}_h}, \mathbf{Mod}^{\mathscr{I}_h}, \models^{\mathscr{I}_h}\right)$ of institutions $\mathscr{I}$ to the class of homogeneous structures. These restrictions $\mathscr{I}_h$ are called *h-institutions*.

Taking a *h*-institution $\mathscr{I}_h$ and using Theorem 18.3 we can replace theories and their homogeneous models by appropriate generative classes obtaining a *g-institution* $\mathscr{I}_g$. Since by Theorems 18.2 and 18.3 generative classes allow to reconstruct countable generic structures, up to isomorphism, and their theories, we have

Theorem 18.17 *Any h-institution $\mathscr{I}_h$ can be transformed to a g-institution $\mathscr{I}_g$ such that countable models in $\mathbf{Mod}^{\mathscr{I}_h}$ can be reconstructed by generative classes in $\mathscr{I}_g$, up to isomorphism, and sentences in $\mathbf{Sen}^{\mathscr{I}}$ satisfied by these models as well.*

An advantage of *g*-institutions is that these categories are syntactic, do not contain semantic objects allowing to describe semantic links in syntactic way.

Again applying Theorems 18.2 and 18.3 we immediately get the following theorem clarifying links between institutions and generic structures via first order fragments of these institutions.

Theorem 18.18 *Let $\mathscr{I} = \left(\mathbf{Sig}^{\mathscr{I}}, \mathbf{Sen}^{\mathscr{I}}, \mathbf{Mod}^{\mathscr{I}}, \models^{\mathscr{I}}\right)$ be an Institution with $\mathscr{D} \subseteq \mathbf{Sig}$ of signature morphisms then every countable $\mathscr{D}$-**First Order Fragment** has generic models and therefore can be represented by them.*

Acknowledgements The first author was partially supported by the Grants Council (under RF President) for State Aid of Leading Scientific Schools (grant NSh-6848.2016.1) and by Committee of Science in Education and Science Ministry of the Republic of Kazakhstan (Grant No. 0830/GF4). The second author was partially supported by Special Account for Research Grants of National Technical University of Athens.

References

1. Baldwin, J.T., Shi, N.: Stable generic structures. Ann. Pure Appl. Logic. **79**, 1–35 (1996)
2. Benda, M.: Remarks on countable models. Fund. Math. **81**, 107–119 (1974)

3. Daniel Gain. Forcing, downward Lwenheim-Skolem and Omitting Types theorems, institutionally Logica Universalis 2013/12/12
4. Diaconescu, R.: Institution-independent Model Theory. In: Studies in Universal Logic. Springer, Birkhauser (2008)
5. Diaconescu, R.: Institution-independent ultraproducts. Fundam. Inform. **55**(3-4):321–348
6. Diaconescu, R. Goguen, J., Stefaneas, P.: Logical support for modularisation. Logical Environments pages , Cambridge 1993. In Proceedings of a Workshop held in Edinburg, Scotland, 83–130 May 1991
7. Fraïssé, R.: Sur certaines relations qui généralisent l'ordre des nombres rationnels. C.R. Acad. Sci. Paris. **237**, 540–542 (1953)
8. Fraïssé, R.: Sur lextension aux relations de quelques propriétés des ordres. Annales Scientifiques de lÉcole Normale Supérieure. Troisième Série. **71**, 363–388 (1954)
9. Fraïssé, R.: Theory of Relations. North-Holland, Amsterdam (1986)
10. Gaina, Daniel, Petria, Marius: J. Logic Comput. **20**(6), 1165–1186 (2010)
11. Goguen, J., Burstall, R.: Institutions: Abstract model theory for specification and programming. J. Assoc. Comput. Mach. **39**, 95–146 (1992)
12. Hart, B., Hrushovski, E., Laskowski, M.S.: The uncountable spectra of countable theories. Ann. Math. **152**, 207–257 (2000)
13. Herwig, B.: Weight ω in stable theories with few types. J. Symbolic Logic. **60**, 353–373 (1995)
14. Hjorth, G., Kechris, A.S.: Borel equivalence relations and classifications of countable models. Ann. Pure Appl. Logic. **82**, 221–272 (1996)
15. Hodges W.: Model Theory. Cambridge University Press (1993)
16. Hrushovski, E.: A new strongly minimal set. Ann. Pure Appl. Logic. **62**, 147–166 (1993)
17. Kechris, A.S., Rosendal, C.: Turbulence, amalgamation and generic automorphisms of homogeneous structures. Proc. London Math. Soc. **94**, 302–350 (2007)
18. Kechris, A.S., Pestov, V.G., Todorcevic, S.: Fraïssé limits, Ramsey theory, and topological dynamics of automorphism groups. Geom. Funct. Anal. **15**, 106–189 (2005)
19. Kechris, A.S., Sokić, M., Todorcevic, S.: Ramsey Properties Of Finite Measure Algebras And Topological Dynamics Of The Group Of Measure Preserving Automorphisms: Some Results And An Open Problem. California Institute of Technology, Pasadena (2015)
20. Kiouvrekis, Y., Stefaneas, P. Topological semantics in institutions with proofs // Algebra and Model Theory 10. In: Pinus, A.G., Ponomaryov, K.N., Sudoplatov, S.V., Timoshenko, E.I. (eds.) Collection of papers. pp. 92–100. Novosibirsk, NSTU (2015)
21. Kubi, W., Maulovi, D.: Appl. Categor. Struct. (2016). 10.1007/s10485-016-9461-z
22. Kubis W.: Fraïssé sequences: category-theoretic approach to universal homogeneous structures. Ann. Pure Appl. Logic **165**(11), 2007
23. Kudaibergenov, K.Zh: On Fraïssé Theorem For An Uncountable Class Of Finitely Generated Structures. Almaty (2016) (Preprint)
24. Kueker, D.W., Laskowski, C.: On generic structures. Notre Dame J. of Formal Logic. **60**, 175–183 (1992)
25. Petria, M.: An institutional version of Gdels completeness theorem. In: Mossakowski, T., Montanari, U., Haveraaen, M., (eds). CALCO, pp. 409–424 (2007)
26. Popkov, R.A., Sudoplatov, S.V.: Distributions of countable models of complete theories with continuum many types. Sib. Electron. Math. Rep. **12**, 267–291 (2015)
27. Rajani, R. Generic Structures, ILLC Scientific Publications, pp-2003–08
28. Shelah, S.: Classification Theory And The Number Of Non-isomorphic Models. North-Holland, Amsterdam (1990)
29. Stefaneas, Petros, Diaconescu, Razvan: Ultraproducts and possible worlds semantics in institutions. Theor. Comput. Sci. **379**(1), 210–230 (2007)
30. Sudoplatov, S.V.: Generative and pre-generative classes. In: Proceedings of the 10 th Panhellenic Logic Symposium, June 11–15, 2015, Samos, Greece. University of Aegean, University of Crete, and University of Athens. 30–34 (2015)
31. Sudoplatov, S.V.: Generative classes generated by sets of diagrams. In: Algebra and Model Theory 10. Collection of papers. NSTU, Novosibirsk. 163–174 (2015)

32. Sudoplatov, S.V.: Group polygonometries. NSTU, Novosibirsk (2011, 2013)
33. Sudoplatov, S.V.: Complete theories with finitely many countable models II. Algebr. Logic. **45**, 180–200 (2006)
34. Sudoplatov, S.V.: Syntactic approach to constructions of generic models. Algebr. Logic. **46**, 134–146 (2007)
35. Sudoplatov, S.V.: On the number of countable models of complete theories with finite Rudin-Keisler preorders. Sib. Math. J. **48**, 334–338 (2007)
36. Sudoplatov, S.V.: The Lachlan Problem. NSTU, Novosibirsk (2009)
37. Sudoplatov, S.V.: Hypergraphs of prime models and distributions of countable models of small theories. J. Math. Sci. **169**, 680–695 (2010)
38. Sudoplatov, S.V.: Classification of Countable Models of Complete Theories. NSTU, Novosibirsk (2014)
39. Sudoplatov, S.V.: Models of cubic theories. Bull. Sect. Logic. **43**, 19–34 (2014)
40. Sudoplatov, S.V.: Classes of structures and their generic limits. Lobachevskii J. Math. **36**, 426–433 (2015)
41. Tarlecki, Andrzej: Quasi-varieties in abstract algebraic institutions. J. Comput. Syst. Sci. **33**(3), 333–360 (1986)
42. Wagner F.: Relational strucrures and dimensions. In: Automorphisms Of First Order Structures. Clarendon Press, Oxford (1994)

Chapter 19
On Combining Algebraic Specifications with First-Order Logic via Athena

Katerina Ksystra, Nikos Triantafyllou and Petros Stefaneas

Abstract We present a verification framework developed by researchers of the National Technical University of Athens as part of the Research Project Thalis "Algebraic Modeling of Topological and Computational Structures and Applications". The proposed framework combines two different specification and theorem-proving systems, in order to facilitate the modeling and analysis of critical software systems. On the one hand, the CafeOBJ algebraic specification language offers executable, composable specifications, and insightful information about the proofs of desired invariant properties. On the other hand, Athena, an interactive theorem-proving system, provides automation and soundness guarantees for its results, as well as detailed structured proofs. Although having conducted complicated case studies (references to which are provided in the paper), here we focus on explaining the steps of the proposed hybrid methodology as clearly as possible, through an illustrative example of a simple mutual exclusion protocol.

Keywords Algebraic specifications · CafeOBJ
Observational transition systems · Proof scores · First-order logic · Athena

K. Ksystra (✉) · N. Triantafyllou · P. Stefaneas
Department of Applied Mathematics, National Technical University of Athens,
Zografou Campus, 15780 Athens, Greece
e-mail: katksy@central.ntua.gr

N. Triantafyllou
e-mail: nitriant@central.ntua.gr

P. Stefaneas
e-mail: petros@math.ntua.gr

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_19

19.1 Introduction

Algebraic specification techniques have been developed in the area of formal methods and several algebraic specification languages and processors have been proposed. In algebraic specification methods, systems are specified based on algebraic modeling, and then the specifications are verified against requirements using algebraic techniques. Algebraic specification languages such as CafeOBJ [1], Maude [2] and CASL [3] have well-known advantages for modeling and reasoning about digital systems. The specifications are relatively simple, readable and writable, and can be executed and analyzed in various ways to provide valuable information to the modelers.

Sometimes, however, specification languages can be more effective when integrated with more conventional theorem-proving systems. CASL, for instance, has been interfaced with HOL/Isabelle [4], through HOL-CASL [5]. Also the Heterogeneous Tool Set (Hets) [6] has connections with the algebraic specification languages Maude and CASL and external theorem provers (SPASS, Vampire, Darwin, KRHyper and MathServe).

In this paper we present a methodology that combines the CafeOBJ algebraic specification language with the Athena theorem proving system [7], and a common interface that trasforms equational specifications written in CafeOBJ into Athena specifications. The aim of the proposed framework is to exploit both the nice properties of CafeOBJ specifications and the soundness and automation offered by Athena. This framework was developed by researchers of the National Technical University of Athens as part of the Research Project Thalis "Algebraic Modeling of Topological and Computational Structures and Applications".

In more details, CafeOBJ provides mechanized implementations of Observational Transition Systems (OTSs), a species of behavioral specifications, that allow users to specify distributed systems using multi-sorted conditional equational logic with subsorting [8]. The specifications are executable via rewriting, which is useful for building up computational intuitions about the underlying system. In addition, CafeOBJ allows users to compose proof scores that establish certain invariant properties, typically by induction. Athena on the other hand, is a system based on general polymorphic multi-sorted first-order logic. It integrates computation and deduction, allows for readable and highly structured proofs, guarantees the soundness of results that have been proved, and also has built-in mechanisms for general model-checking and theorem-proving, as well as seamless connections to state-of-the-art external systems for both.

By combining these two methodologies we wish to combine the strengths of CafeOBJ, most notably succinct, composable, executable specifications based on conditional equational logic with those of Athena, namely, structured and readable proofs, soundness of the results, and greater automation both for proof and for counterexample discovery. The goal of this paper is to give an overview of the theoretical foundations of our framework (Sect. 19.2) and to provide an easy to follow introduction to the proposed methodology (together with a tool that automates the

transformation process from OTS/CafeOBJ into Athena specifications) using a simple and illustrative example (Sect. 19.3). Finally, we discuss related work and advantages of the proposed framework in Sect. 19.4 and future directions in Sect. 19.5.

We should mention here that the proposed methodology has been successfully applied to much larger and complicated use cases as well. To this end we encourage the interested reader to visit [9] for the application of the proposed methodology to the verification of the Alternating Bit Protocol, which is considered the traditional benchmark used for testing mechanical verification of protocols [10].

19.2 Theoretical Background

19.2.1 CafeOBJ Algebraic Specification Language

CafeOBJ [1] is an algebraic specification language and processor that can be used to specify abstract data types and abstract state machines. The basic units of a CafeOBJ specification are its modules. There are two kinds of modules in CafeOBJ, tight and loose modules. A tight module only accepts the smallest implementation that satisfies what are specified in the module, while a loose module can accept any implementations (that satisfy them). A tight module is declared with the keyword $mod!$, and a loose module with the keyword $mod*$.

In CafeOBJ modules, we can declare module imports, sorts, operators, variables and equations. Operators without arguments are called constants. Built-in operators denoting logical connectives can be used to declare negation, conjunction, disjunction, implication, and exclusive disjunction. Operators can have attributes such as *comm* that specifies that the binary operator is commutative. Operators are declared with the keyword *op* (or *ops* if there are many). The constructor operators of the sorts are declared with the attribute *constr*. The non-constructor operators, or some properties of the operators are defined in equations. Conditional equations can also be declared inside a module, using the keyword *ceq*.

19.2.2 Observational Transition Systems (OTSs)

An Observational Transition System (OTS) is a transition system written in terms of equations and is a proper subclass of behavioral specifications [11]. Assuming that there exists a universal state space Y and that each datatype we need to use has been declared in advance, an OTS S is defined as the triplet $S = \langle O, I, T \rangle$ where [12]:

- O is a set of observers. Each $o \in O$ is a function $o : Y D_{o1} \dots D_{om} \rightarrow Do$, that takes as input a state of the system and maybe other datatypes (not necessarily in that order) and return a datatype value. Given an OTS S and two states u_1, u_2 the equivalence between them is defined with respect to the values returned by

the observers, i.e. $u_1 =_S u_2$ if and only if for each $o \in O$, $o(u_1, x_1, \ldots, x_m) = o(u_2, x_1, \ldots, x_m)$ for all $x_1 \in Do_1, \ldots, x_m \in Do_m$.
- I is the set of initial states, such that $I \subseteq Y$.
- T is a set of conditional transitions. Each $t \in T$ is a function $t : Y\ D_{t1} \ldots D_{tn} \rightarrow Y$. Each transition t, together with any other parameters, preserves the equivalence between two states, i.e. if $u_1 =_S u_2$, then for each $t \in T$, $t(u_1, y_1, \ldots, y_n) =_S t(u_2, y_1, \ldots, y_n)$ for all $y_1 \in D_{t1}, \ldots, y_n \in D_{tn}$. Each t has an effective condition $c\text{-}t : Y\ D_{t1} \ldots D_{tn} \rightarrow Bool$. If $\sim c\text{-}t(u, y_1, \ldots, y_n)$, then $t(u, y_1, \ldots, y_n) =_S u$. $t(u, y_1, \ldots, y_n)$ is called a successor state of a state u. We write $u \rightsquigarrow_S u'$ iff a state $u' \in$Y is a successor state of a state $u \in Y$.

An *execution* of an OTS S is an infinite sequence $u_0, u_1, \ldots$ of states satisfying:

– *Initiation*: $u_0 \in I$ and
– *Consecution*: For each $i \in N$, there exists $t \in T$ such that $u_{i+1} =_S t(u_i)$.

Let ϵ_s be the set of all executions obtained from S. A state $u \in Y$ appears in an execution $u_0, u_1, \ldots$ of an OTS S, denoted by $u \in u_0, u_1, \ldots$ if there exists $i \in N$ such that $u =_S u_i$. A state $u \in Y$ is called reachable with respect to an OTS S, if and only if there exists an execution $e \in \epsilon_s$ such that $u \in e$. Let R_S be the set of all reachable states with respect to S.[1]

Roughly speaking, in the OTS/CafeOBJ method the *transitions* between the states of the system are modelled with constructor operators. The structure of a state is abstracted by the *observers*, each one returning some observable information about the state. The meaning of an observer is formally described by means of (conditional) equations [14].

19.2.3 Proof Scores in CafeOBJ

CafeOBJ is equipped with its processor called the CafeOBJ system, which is used as an interactive theorem prover. The CafeOBJ system verifies the desired properties by using the equations of the theory that defines an OTS as left to right re-write rules. This method is called proof score approach and is computer-human interactive [15]. A proof score is a plan to verify that a property holds for a specification. This is implemented as a set of instructions written by a human to a proof engine, such that when executed, and if everything evaluates as expected, a desired theorem is proved.

In order to prove an invariant property using CafeOBJ the following steps need to be taken. First, we formally express the property we want to prove as a predicate in CafeOBJ terms in a module. Next, we write the inductive step as a predicate that defines that if the invariant holds in an arbitrary state s then that implies that it holds in its successor state s'. Then we ask CafeOBJ to prove via term rewriting (using the

[1] R_S is the type denoting the set of all reachable states wrt S. Also Sys denotes R_S but not Y if the constructor-based logic is adopted, which is the current logic underlying the OTS/CafeOBJ method [13].

reduce command), if the property holds for an arbitrary initial state. Finally, using all the transition rules in turn, we must instantiate s' and ask CafeOBJ to prove the inductive step for each case.

After asking CafeOBJ to prove such an expression three results might be returned by the system. If true is returned this means that the proof is successful. If a CafeOBJ term is returned, this means that there exist some terms that the system cannot fully reduce. The user must then split the case, by stating that the returned term equals to true and false in turn (computer-human interactive method). This creates two new proof obligations and is known as case splitting. Finally, if false is returned, either the property does not hold, or the case that returned false is unreachable for our system.

19.2.4 Athena

Athena [7] is an interactive theorem proving environment for polymorphic multi-sorted first-order logic, with separate (but intertwined) languages for computation and deduction. The main tool for constructing proofs is the method call. A method call represents an inference step and can be primitive or complex. Methods can accept as arguments other methods and/or procedures, and thus are higher-order.

Athena proofs are machine-checkable and expressed in a true natural-deduction style, a style of proof that was explicitly designed to capture the essential aspects of mathematical reasoning as it has been practiced for thousands of years [16], whose semantics are formalized in terms of *assumption bases*. An assumption base is a set of sentences serving as working hypotheses—typically axioms and previously derived theorems. Initially the system starts with a small assumption base, containing defining axioms for some built-in function symbols. When an axiom is postulated or a theorem is proved, the corresponding sentence is inserted into the assumption base. During the evaluation of a proof, methods interact with the assumption base, checking to see if some arguments are present in the set and/or making new entries [17].

Another important point here is that Athena guarantees the soundness of every defined method, meaning that if and when a method call manages to derive a conclusion p, we can be assured that p is a logical consequence of the assumption base in which the call occurred. This guarantee stems from the formal semantics of Athena. The larger point to keep in mind is that defined methods can never produce a result that is not logically entailed by the assumption base [16].

Additionally, Athena is integrated with model builders. Model generation is useful for consistency checking, and in particular for falsifying conjectures: If we are not sure whether a formula follows from a given set of premises, we can try to find a counter-model, i.e., a model in which all the premises are true but the formula in question is false. Model generators can be used to find such models automatically. If we manage to find a counterexample in Athena, then any attempt to prove the property would be pointless and therefore model checking can be quite a time saver [16].

Athena provides also a significant degree of proof automation, through seamless interfaces to powerful automated theorem provers like Vampire [18] and SPASS [19], as well as SAT and SMT solvers.

19.3 Proposed Framework: Combining CafeOBJ and Athena Environments

The OTS/CafeOBJ approach presents many advantages, the most important of which, in our opinion and experience, is that the proof scores can easily and effectively guide the user to discover the required case splittings and occasional lemmas for the proof. However, we believe that it could benefit by being combined with proof systems like Athena, for the following reasons.

At first, each proof conducted in Athena is checked for soundness by the system. Thus, Athena could act as a validator for the proofs conducted in CafeOBJ. Also, Athena as we have already mentioned is integrated with some powerful automated theorem provers (ATPs). Thus, via Athena, access to these highly efficient ATPs can be enabled for OTS/CafeOBJ specifications as well, which could help to streamline the verification process. Finally, Athena uses a Fitch-style natural-deduction proof system. Expressing OTS/CafeOBJ in a high-level and structured natural-deduction style such as Athena's could help to make them comprehensible and accessible to a wider audience.

Accordingly, we propose a methodology which combines both environments (CafeOBJ and Athena) and consists of the following steps:

- Step 1: Create the OTS specification in CafeOBJ.
- Step 2: Generate an equivalent Athena specification, automatically.
- Step 3: Attempt to falsify the property of interest in Athena. If unsuccessful proceed to the next step, else either the property or the specification should be revised.
- Step 4: Apply the proof score methodology in CafeOBJ until a lemma is required. Discover a candidate lemma in CafeOBJ that can discard the problematic case.
- Step 5: Attempt to falsify the candidate lemma using Athena. If unsuccessful proceed to the next step, else either the property or the specification should be revised.
- Step 6: Iterate steps 4 to 5 until the proof scores methodology is completed for the property in question and also for all the lemmas used.
- Step 7: Using the insights gained by the proof scores (case splittings and lemmas) generate an Athena proof and check its soundness.

With the proposed methodology (steps 1 to 3), the user could save considerable time by first attempting to falsify the property in question. If indeed a counter example is returned by Athena, then either the property in question is not invariant for the specification, or there might be a bug in the specification itself. In the first case, the proof is completed and the property falsified. In the second case the output of Athena usually provides sufficient information for the discovery and correction of the bug.

During the verification of complex systems with the proof score methodology, it is possible that the user will consider as lemmas, properties which while successfully discard the problematic cases, are not invariant for the specification. This is usually discovered at a late stage of the verification of the lemmas, in which case the proof needs to be recreated using a different candidate lemma. This, can at times become an important time sink for the verification effort. Athena, could potentially inform the user of the error in the candidate lemmas at a much earlier stage (steps 4 to 5) and thus save the user valuable time.

Also once the proof score methodology is completed, by transferring the insights gained by it to Athena we can easily create a formal proof for the property in question. Athena can thus inform us about the soundness of the proof or point to reasoning errors.

Finally, as we mentioned earlier the final proof constructed in Athena will be in a style easy to understand and thus could help provide explanation as to why the property is invariant or not for the system, in other words act as a sort of software documentation.

19.3.1 Rules of Translation

In order to obtain an Athena specification from an OTS/CafeOBJ specification we have defined an appropriate translation schema. The basic units of OTS/CafeOBJ specifications and their transformation into Athena notation are shown in Table 19.1.

In more details, in Athena initial algebras (with tight semantics) are specified using the keyword *datatype*, whereas an arbitrary carrier (with loose semantics) is introduced with the *domain* keyword. An induction principle is automatically generated for every new datatype. States are formalized as a *structure* and state transitions as its constructors. Structures are very much like datatypes, except that there may be confusion, i.e. different constructor terms might denote one and the same object. An induction principle is also automatically generated (and, of course, valid). We continue with the declaration of the observers. They are defined as functions with the constraint that they take as input a state (and maybe additional input) and return

Table 19.1 The basic units of OTS/CafeOBJ specifications and their transformation into Athena notation

OTS/CafeOBJ notation	Athena notation
Tight modules	Datatypes
Loose modules	Domains
State	Structure
Initial state	State constructor
State transitions	State constructors
Observers	Functions
Equations	Axioms

some datatype. Finally, the equations that define the initial state, as well as the pre- and post-conditions of the state transitions are defined as axioms.

Informally, an arbitrary OTS specification written in CafeOBJ terms is translated into an Athena specification through the operator `cafe2athena` as follows:

Datatype modules:

```
cafe2athena(mod! M1 {[m1] ...})
=
datatype m1 cafe2athena(...)

cafe2athena(mod* M2 {[m2] ...})
=
domain m2 cafe2athena(...)
```

OTS modules (sort representing state space, initial state, transitions):

```
cafe2athena(mod S {*[Sys]* op init : → Sys .
bop a : Sys V_j1 ... V_jn → Sys ...})
=
structure Sys := init | (a Sys V_j1 ... V_jn)
```

Observers:

```
cafe2athena(bop o : Sys V_i1 ... V_in → V)
=
declare o := [Sys V_i1 ... V_in] → V
```

Variables:

```
cafe2athena(Var x_i1 : V_i1 ... Var x_in : V_in)
=
define [x_i1 ... x_in] := [?V_i1 ... ?V_in]
```

Init axiom:

```
cafe2athena(eq o(init, x_i1, ..., x_in) = f(x_i1, ..., x_in))
=
assert* init-axiom := ((o init x_i1 ... x_in) = f x_i1 ... x_in)
```

where x_{i1},...,x_{in} are V_{i1} ...V_{in} sorted CafeOBJ variables and $f(x_{i1}, ..., x_{in})$ is the value of the observer at the initial state.

Effective condition:

```
cafe2athena(op c-a : Sys V_j1 ... V_jn → Bool .
eq c-a(w, x_j1, ..., x_jn) = g(w, x_j1, ..., x_jn) .)
=
define c-a := lambda (w x_j1 ... x_jn) (g w x_j1 ... x_jn)
```

where w is a hidden sorted variable and x_{j1}, ..., x_{jn} are V_{j1}...V_{jn} sorted CafeOBJ variables.

Transition axioms:

```
cafe2athena(eq o(a(w, x_j1, ..., x_jn), x_i1, ..., x_in) = e-a(w, x_j1, ..., x_jn, x_i1, ..., x_in)
     if c-a(w, x_j1, ..., x_jn) .)
=
assert a-axiom :=
((o(a w x_j1 ...x_jn) x_i1 ... x_in) = (e-a w x_j1 ... x_jn x_i1 ... x_in)
     if (c-a w x_j1 ... x_jn))
```

where $e\text{-}a(w, x_{j1}, ..., x_{jn}, x_{i1}, ..., x_{in})$ is the changed value of the observer after the application of the transition. The definition of a proof score in Athena terms can be found in Appendix 5

19.3.1.1 Semantic Correctness of the Translation.

Some readers may have noticed that we restricted the translation to non-parametric modules. The parametric module system does not add to the expressiveness of the language, however not supporting it may result in an overhead in specification code. By enforcing this restriction on the style of the OTS/CafeOBJ specifications it is safe to claim the semantic correctness of the translation, since in both languages the underlying semantics is basically Order-sorted Conditional Equational Logic (in which constructors are explicitly used). For details about the semantic correctness of the translation we refer readers to [9].

19.3.1.2 Cafe2Athena Tool.

In order to make the proposed methodology more agile, we have developed a tool that takes as input an OTS/CafeOBJ specification and automatically produces an Athena specification, implementing the rules of translation we previously presented. The tool is written in Java and hides the details of the translation, since the user loads a CafeOBJ specification file, presses the "Translate to Athena" button and gets the corresponding Athena specification. Cafe2Athena tool was a deliverable of the Research Program Thalis "Algebraic modeling of topological and computational structures" and can be downloaded from [20].

19.3.2 Illustrating Example: A Mutual Exclusion Protocol Using an Atomic Instruction (Mutex)

We present in this section the application of the proposed methodology to a mutex algorithm so as to explain it better and demonstrate its effectiveness. In this system, we have a set of processes, each of which is executing code. A process, at any system state, is either in some critical section of the code or in some remainder (waiting)

section. Also, we have two transitions; the first corresponds to a process entering the critical section and the second to a process exiting the critical section (and entering the remainder section).

When a process p enters its critical section, the resulting state becomes *locked*. When p exits the critical section, the resulting state is unlocked. For p to enter its critical section in some state s, p must be *enabled* in s. A process p is *enabled* in s iff p is in its remainder section in s and s is not locked. This is, therefore, the effective condition of the *enter* state transition for a given process. The effective condition of the *exit* transition is for the process to be in its critical section. We also have two observer functions, one that takes a state s and a process id p and tells us what section of the code p is executing in s (critical or remainder), and a function that takes a state s and tells us whether s is locked.

19.3.2.1 Step 1. Specification in CafeOBJ.

The specification of the mutex OTS in CafeOBJ can be seen in Appendix 5

19.3.2.2 Step 2. Specification in Athena.

Using the 'Cafe2Athena' tool we obtain the specification of the mutex OTS in Athena. Some parts of the specification are explained below.

A domain of process identifiers (`Pid`) and a datatype for code labels (`cs` and `rs`, for critical and remainder section, respectively) have been introduced:

```
domain Pid
datatype Label := rs | cs
assert label-axioms := (datatype-axioms"Label")
```

Here, `rs` and `cs` are the (nullary) *constructors* of the `Label` algebra. The `datatype-axioms` of `Label` are quantified sentences that assert no-confusion and no-junk conditions for the constructors. The effect of the `assert` command is to insert those conditions into the global *assumption base*. States are formalized as a structure and the state transitions as constructors of this structure, as follows:

```
structure State := init | (enter Pid State)
| (exit Pid State)
```

The declaration of the observer functions can be seen below:

```
declare at: [Pid State] -> Label
declare locked: [State] -> Boolean
```

The "*at*" function tells us the label of a given process in a given state. Binary function symbols can be used in infix in Athena (the default notation is prefix), so if

p and *s* are terms of sort *Pid* and *State*, respectively, then (*p ats*) gives us the label of *p* in state *s*. The term *(lockeds)* tells us whether or not *s* is locked.

We now present the axioms that define the initial state, as well as the pre- and post-conditions of the two state transitions (entering and exiting). First, appropriate variables for the given sorts are defined.

```
define [i j s s'] := [?i:Pid ?j:Pid ?s s':State]
assert* init-axioms := [(_ at init = rs)
                        (~ locked init)]
```

The initial-state axioms are simple enough: every process in the initial state is in the remainder section, and the initial state is not locked. For the transition axioms of *enter* we see it is helpful to define the effective condition as a separate procedure, called `enabled-at`:

```
define (enabled-at i s) := (s at i = rs & ~ locked s)
```

The axioms for the *enter* and *exit* transitions, respectively, are presented below:

```
assert* enter-axioms :=
 [(i enabled-at s ==> locked i enter s)
  (i enabled-at s ==> i at i enter s = cs)
  (i enabled-at s & j =/= i ==> j at i enter s = j at s)
  (~ i enabled-at s ==> i enter s = s)]

assert* exit-axioms :=
 [(i at s = cs ==> i at i exit s = cs)
  (i at s = cs & j =/= i ==> j at i enter s = j at s)
  (i at s = cs ==> ~ locked i exit s)
  (i at s =/= cs ==> i exit s = s)]
```

19.3.2.3 Step 3. Define the Desired Goal and Falsify It with Athena.

The desired mutual exclusion property satisfied by the algorithm, is that there is at most one process in the critical section at any given moment. This property can be rephrased as *"if there are two processes in the critical section, then those processes are identical"*. In Athena, the desired goal is defined as follows:

```
define (goal-property s) :=
    (forall i j . i at s = cs & j at s = cs ==> i = j)
define goal := (forall s . goal-property s)
```

Before we try to prove a conjecture *p*, it is often useful to first try to falsify it in Athena by finding a counterexample to it. This falls under a class of techniques collectively known as model checking, or model building. If we manage to find a counterexample, then any attempt to prove *p* would be useless (because Athena's proofs are guaranteed to be sound, so we can never prove something that doesnt

follow logically from the assumption base, and if we find a counterexample to p then p does not follow from the assumption base), and therefore model checking can be quite a time saver. Moreover, the feedback provided by the model checker, given in the form of a specific counterexample to the conjecture, is often very valuable in helping us to debug and, in general, to better understand the theory or system we are developing [16].

Athena is integrated with a number of external systems that can be used for model building/checking, most notably SMT and SAT solvers, but there is also a built-in model checker which is perhaps the simplest to use and can be surprisingly effective [16].

To falsify a conjecture p, we simply call (`falsify` p N). Here N is the desired *quantifier bound*, namely, the number of values of the corresponding sort that we wish to examine (in connection with the truth value of p) at each quantifier of p. When falsification fails within the given bound, the term *'failure* is returned. When falsification succeeds, it returns specific values for the quantified variables that make the conjecture false. Let us see in our example, if we can falsify the goal by examining 100 states:

```
> (falsify  goal 100)
List: ['success
|{
?i:Pid := Pid_1
?j:Pid := Pid_2
?s:State := (enter Pid_2
        (exit Pid_1
              (enter Pid_1 init)))
}|]
```

As we can see, Athena falsified the property in question and returned a state in which the goal is violated. This state can be reached after the application of the transitions *enter*, *exit* and *enter* in the initial state of our system, as Athena informed us `(enter p2 exit p1 enter p1 init)`.

In order to understand better why the property is violated in a particular state we have written a procedure, `simulate`, that takes as input a sequence of states $s_1, \ldots, s_n$, and prints out each state in the sequence by applying all observer functions to the given state, and specifically by evaluating the applications of those observer functions (by using the relevant axioms as rewrite rules).

Since we know which transitions cause the falsification of the goal, we can use the simulate method to see the value of the observers in this problematic state. In our case, calling `simulate (p2 enter p1 exit p1 enter init)`, results in the following output:

```
State after (p1 enter init):
locked:                                    p1 at:
------------                               -------------------
true                                         cs

State after (p1 exit (p1 enter init)):
locked:                                    p1 at:
------------                               -------------------
false                                        cs

State after (p2 enter (p1 exit (p1 enter init))):
locked:                                    p2 at:
------------                               -------------------
true                                         cs
```

If we observe the returned values in this state we see that while $p2$ exits the critical section it basically remains in the critical section. Thus we understand that there must exist an error in the definition of the exit axioms (in particular, the first axiom was mis-written as `(i at s = cs ==> i at i exit s = cs)` instead of `(i at s = cs ==> i at i exit s = rs)`). After redefining the axiom, we falsify again the desired goal, and Athena returns the term *'failure*. Thus we proceed to the next step of our methodology.

19.3.2.4 Step 4. Start the Proof of the Desired Goal Using Proof Scores Approach (until a Lemma Is Needed).

We start to work with proof scores in CafeOBJ, until we reach the following case where CafeOBJ returns false:

```
open ISTEP .
-- arbitrary values
op k : -> Pid .
-- assumptions
-- eq c-enter(s,k) = true .
eq at(s,k) = rs .
eq locked(s) = false .
eq i = k .
eq (j = k) = false .
eq at(s,j) = cs .
-- successor state
eq s' = enter(s,k) .
red istep1 .
close
```

This means that in order to discard this case we must use a lemma. By taking a close look at the equations that define this state we can observe that `eq at(s,j) = cs` and `eq locked(s) = false` cannot hold together. This means that a possible lemma could be the following: `inv2(S,J) = (at(S,J) = cs implies locked(S) = true)`. To test if this lemma actually discards the problematic

case we use the above proof score with the reduction `red inv2(s,j) implies istep1` and CafeOBJ returns true.

19.3.2.5 Step 5. Falsify the Discovered Lemma with Athena.

After defining the lemma in Athena, we try to falsify it and Athena returns the term *'failure*, so we continue the proof score using this lemma.

19.3.2.6 Step 6. Continue the Proof Using Proof Scores in CafeOBJ.

Checking the rest of the cases in CafeOBJ will result in the completion of the proof score. It is interesting to point out that in our example, during the proof score of the lemma the original property under verification was required as part of the inductive hypothesis. This case is shown below.

```
open ISTEP .
-- arbitrary values
op k : -> Pid .
-- assumptions
eq at(s,k) = cs .
eq (j = k) = false .
eq at(s,j) = cs .
-- successor state
eq s' = exit(s,k) .
red inv1(s,j,k) implies istep2 .
close
```

19.3.2.7 Step 7. Create an Athena Proof Based on the Gained Insights.

The above pattern in a proof score denotes a situation where simultaneous induction must performed. Together with the case splits and lemmas used, such information is essential to the construction of the Athena proof. Also, by taking a closer look at the invariant and the lemma used ((`i at s = cs and j at s = cs ==> i = j` and `j at s = cs ==> locked s`, respectively) it is not difficult to understand that a strengthened goal can be formulated out of them: `i at s = cs & j at s = cs ==> i = j & locked s`, which we will attempt to verify in Athena. The new goal is defined as follows:

```
define (new-goal-property s) :=
(forall i j . i at s = cs & j at s = cs ==> i = j & locked s)
define new-goal := (forall s . new-goal-property s)
```

A method for completely automated inductive reasoning is automatically defined whenever a new structure or datatype is introduced, as we have already mentioned. The name of the method is the name of the corresponding structure joined with the string `-induction`, in lower case. If we try to prove the new goal using the `state-induction` we get the following output, which means that the goal is automatically proved.

```
> (!state-induction new-goal)
Theorem: (forall ?s:State
            (forall ?i:Pid
              (forall ?j:Pid
                (if (and (= (at ?i:Pid ?s:State)
                            cs)
                         (= (at ?j:Pid ?s:State)
                            cs))
                    (and (= ?i:Pid ?j:Pid)
                         (locked ?s:State))))))
```

However, a completely automatic proof does not shed much light on a system's workings. A structured proof that derives the desired result in a piecemeal fashion can be much more valuable in *explaining* the underlying system, i.e., in explaining why a given property holds. In addition, the effort invested in constructing the proof often pays off in increased understanding, and also in the discovery of errors, unintended consequences of design constraints, and so on.

Athena uses a natural-deduction style of proof, as we have already mentioned, which allows for structured and readable proofs that resemble in certain key respects the informal proofs one encounters in practice [16]. Part of a detailed structured Athena proof of the goal property for our example is shown below.[2]

```
by-induction (forall s . new-goal s) {
  init => (!spf (new-goal init) (ab))
| (s' as (k enter s)) =>
   pick-any i:Pid j:Pid
    let {IH := (forall i j . i at s = cs & j at s = cs
                          ==> i = j & locked s)}
    assume hyp := (i at s' = cs & j at s' = cs)
     conclude goal := (i = j & locked s')
      (!two-cases
       assume case1 := (k enabled-at s)
        let {s'-locked := (!chain-> [case1
                              ==> (locked s') [enter-axioms]]);
             s-not-locked := (!chain-> [case1
                              ==> (~ locked s) [right-and]]);
             i=j := (!by-contradiction (i = j)
                      assume h := (i =/= j)
                   let {D := {(i =/= k | j =/= k) from h}}
                        (!cases D
                          assume i=/=k := (i =/= k)
                           (!M i=/=k case1 IH)
                          assume j=/=k := (j =/= k)
                           (!M j=/=k case1 IH)))}
```

[2]For the full proof we refer readers to Appendix 5.

```
        (!both i=j s'-locked)
      assume case2 := (~ k enabled-at s)
        (!direct-ih hyp s case2 IH enter-axioms))
...
}
```

In this way, Athena takes us one step closer to the goal of enabling proofs that can serve to explain and communicate our reasoning, but which are nevertheless entirely formal and checkable by machine [16].

19.4 Discussion

Some approaches that deal with the integration of algebraic specifications with more conventional proving systems are briefly presented. A methodology that offers parsing, static analysis and proof management is the tool Hets - the Heterogeneous Tool Set [6]. Hets has among others, connections with the algebraic specification languages Maude and CASL and external theorem provers and is based on a graph of logics and logic translations. Another interesting methodology for proving inductive properties of OTSs that aims at automating the proof scores approach to verification, can be found in [14]. In this paper, authors revise the entailment system of proof scores and enrich it with proof rules and tactics. Also, a prototype tool (Constructor-based Inductive Theorem Prover - CITP) implementing the methodology is demonstrated. CITP is implemented in Maude.

This last approach is the closest to ours but with a different point of view. While both methodologies aim at providing soundness to the OTS/CafeOBJ method, CITP focuses on the automation of the proofs whereas we are more concerned with combining the benefits of CafeOBJ and Athena specification and verification methodologies into one. For example, in [14] authors discharge automatically a desired property and the required lemmas but they do not describe how the lemmas are formulated. Also they do not support detailed proofs or simulation of the system's behavior. Thus, we believe that our approach offers better explanation and understanding of the verification of the desired properties and the behavior of the specified system.

In the following we summarize the benefits of the proposed methodology; One advantage is that the proof scores approach can provide feedback to the user when a proof fails, and can be used to discover the required case splittings and occasional lemmas for the proof, in contrast with most automated theorem provers including Athena. Another advantage is that you can use the model-checking tools of Athena to obtain some first insights about the specified system and thus save valuable time. Also, the simulate procedure, can become really helpful in understanding how the specified system behaves. Especially when you deal with a complex system where it is almost impossible to "follow" its execution process, such visualization techniques can provide a clear overview of the system and help in the discovery of possible errors. For example in [21] authors state that among the lessons learned from a non trivial, real life protocol case study of the OTS/CafeOBJ method, is the delay arose

from errors in the specification and the expression of system's property. These errors would be found easily with the proposed integration. In addition, the structured proofs supported by Athena can provide valuable information and explanation as to why a property is invariant or not for the specified system. Another important advantage of our approach is that Athena does a thorough check of the overall proof and provides a guarantee that if and when you get a theorem, that result follows logically from the assumption base and the primitive methods (which in this case include the external ATP). This is really helpful because on the other hand, with CafeOBJ's proof scores approach there is much greater room for human oversight. Finally, the automation offered by Athena through its connection with external systems is another advantage of the proposed framework.

19.5 Conclusion

We proposed a hybrid methodology that combines the proof scores approach of CafeOBJ with Athena's reasoning and verification tools, together with a tool that translates equational CafeOBJ specifications into Athena code, and demonstrated our approach with a mutex algorithm. Also we presented several features of the methodology that can be used to: better understand the specified system, model-check desired properties and verify them via theorem proving, both automatically and in a structured and more detailed way. The proposed method aims at combining the strengths of the languages, CafeOBJ and Athena, by working with proof scores and CafeOBJ but also taking advantage of more conventional formal-methods techniques that have traditionally lied outside of the rewriting community.

This methodology has been applied in larger case studies as well [9] and as future work we plan to push the automation level further for complex systems and to investigate possible connections with tools incorporating various provers and different specification languages. Also, we plan to extend the proposed framework in order to include the behavioral aspects of CafeOBJ as well.

Acknowledgements This research has been co-financed by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALIS. The authors would like to acknowledge the insightful feedback provided by Dr. Konstantine Arkoudas. The authors would also like to warmly thank Prof. Dr. Sofia Lambropoulou, Project Coordinator of the Research Project Thalis "Algebraic Modeling of Topological and Computational Structures and Applications" for their excellent collaboration.

Appendix A

Here we present the CafeOBJ specification of the mutex system.

```
mod* PID {
  [Pid]
  op _=_ : Pid Pid -> Bool {comm}
  var I : Pid
  eq (I = I) = true .}
mod! LABEL {
  [Label]
  ops rs cs : -> Label
  op _=_ : Label Label -> Bool {comm}
  var L : Label
  eq (L = L) = true .
  eq (rs = cs) = false .}
mod* MUTEX {
  pr(LABEL + PID)
  *[Sys]*
  -- an arbitrary initial state
  op init : -> Sys
  -- observation functions
  bop at : Sys Pid -> Label
  bop locked : Sys -> Bool
  -- transition functions
  bop enter : Sys Pid -> Sys
  bop exit : Sys Pid -> Sys
  -- CafeOBJ variables
  var S : Sys
  vars I J : Pid
  -- init
  eq at(init,I) = rs .
  eq locked(init) = false .
  -- enter
  op c-enter : Sys Pid -> Bool
  eq c-enter(S,I) = ((at(S,I) = rs) and not locked(S)) .
  ceq at(enter(S,I),J) = cs if (I = J) and c-enter(S,I) .
  ceq at(enter(S,I),J) = at(S,J) if not((I = J) and c-enter(S,I)) .
  ceq locked(enter(S,I)) = true if c-enter(S,I) .
  ceq enter(S,I) = S if not c-enter(S,I) .
  -- exit
  op c-exit : Sys Pid -> Bool
  eq c-exit(S,I) = (at(S,I) = cs) .
  ceq at(exit(S,I),J) = rs if (I = J) and c-exit(S,I) .
  ceq at(exit(S,I),J) = at(S,J) if not((I = J) and c-exit(S,I)) .
  ceq at(exit(S,I),J) = at(S,J) if not(I = J) and not c-exit(S,I) .
  ceq locked(exit(S,I)) = false if c-exit(S,I) .
  ceq exit(S,I) = S if not c-exit(S,I) .}
```

Appendix B

The declaration of an invariant property in CafeOBJ terms and the definition of the induction schema, are shown below:

```
-- declaration of the invariant property
mod INV {
-- arbitrary values
op s : -> Sys .
ops i j : -> Pid .
-- name of invariant to prove
op inv1 : Sys Pid Pid -> Bool
-- CafeOBJ variables
var S : Sys
vars I J : Pid
-- invariant to prove
eq inv1(S,I,J) = (at(S,I) = cs and at(S,J) = cs implies I = J) .
}
-- declaration of the inductive step
mod ISTEP {
pr(INV)
-- arbitrary values
 op s' : -> Sys
-- name of formula to prove in each inducion case
op istep1 : -> Bool
-- formula to prove in each induction case
eq istep1 = inv1(s,i,j) implies inv1(s',i,j) .
}
```

The definition of the corresponding invariant in Athena is presented here (the induction schema is automatically defined in Athena).

```
define (inv1 s) := (forall ?i ?j . at s ?i = cs & at s ?j = cs
==> ?i = ?j)
```

The proof score of the desired invariant property in CafeOBJ, for the initial state and when a transition, called `enter(s,k)`, is applied can be seen below:

```
-- proof score
-- I. Base case
open INV .
  red inv1(init,i,j) .
close .

-- II. Induction case
-- 1. enter(s,k)
open ISTEP  .
-- arbitrary values
  op k : -> Pid .
-- successor state
  eq s' = enter(s,k) .
-- check
  red istep1 .
close
```

The corresponding proof skeleton in Athena can be defined as follows.

```
by-induction (forall ?s . goal-property ?s) {
    init => (!prove (goal-property init))
    | (state as (enter s k)) =>
    pick-any i:Pid j:Pid
        assume hyp := (state at i = cs & state at j = cs) {
                goal := (i = j);
            goal from (ab)
        }
    }
```

Finally, the following proof scores present a case splitting in CafeOBJ. In the first case we assume that `at(s,k) = cs` while the second proof score assumes its symmetrical case, i.e. `(at(s,k) = cs) = false`.

```
open ISTEP  .
-- arbitrary values
  op k : -> Pid .
-- case 1
eq at(s,k) = cs .
-- successor state
  eq s' = enter(s,k) .
-- check
  red istep1 .
close

open ISTEP  .
-- arbitrary values
  op k : -> Pid .
-- case 2
eq (at(s,k) = cs) = false .
-- successor state
  eq s' = enter(s,k) .
-- check
  red istep1 .
close
```

The same case splitting can be defined in Athena terms as follows:

```
(! two-cases
assume case1 := (k at s = cs)
...
assume case2 := (k at s =/= cs))
```

Appendix C

A detailed structured Athena proof of the (strengthened) goal for our example is shown below.

Theorem 19.1 *For all states s' and processes i and j, if i and j are in their critical sections in s', then $i = j$ and s' is locked.* ■

Proof By structural induction on s'. When s' is the initial state the result is trivial because the antecedent is false, as all processes are in their remainder sections initially. Suppose now that s' is of the form $(k\ enter\ s)$. Pick any processes i and j and assume both are in their critical sections in s'. We then need to show that $i = j$ and that $s' = (k enter s)$ is locked. The inductive hypothesis here is:

$$i\ at\ s = cs\ \&\ j\ at\ s\ = cs\ ==>\ i = j\ \&\ locked\ s \tag{19.1}$$

We distinguish two cases:

1. *Case 1*: k is enabled at s. Then $(k\ ats' = cs)$and $(locked s')$ follow from the *enter* axioms. Thus, we only need to show $i = j$. By contradiction, suppose that $i \neq j$. Then either $i \neq k$ or $j \neq k$. [3] So assume first that $i \neq k$ (the reasoning for the case $j \neq k$ is symmetric). Then, from the *enter* axioms and the assumption that k is enabled at s, we conclude $i\ ats' = i\ ats$, hence $i\ at\ s = cs$. Now applying the inductive hypothesis to the above assumption, we conclude $(locked\ s)$. However, that contradicts the assumption that k is enabled at s, as that assumption means that s is *not* locked.
2. *Case 2*: k is not enabled at s. In that case, by the *enter* axioms, we get

$$(k\ enter\ s = s)$$

   i.e., $s' = s$, and the result now follows directly from the inductive hypothesis.

Finally, suppose that s' is of the form $(k\ exit\ s)$. Again pick any processes i and j and assume both are in their critical sections in s'. We again need to show that $i = j$ and that $s' = (k exit s)$ is locked. The inductive hypothesis here is the same as before, (19.1). We distinguish two cases again, depending on whether or not the effective condition of the *exit* transition holds:

1. *Case 1*: $(k\ at\ s = \text{cs})$. We proceed by contradiction. First, by applying the inductive hypothesis to the conjunction of $(k\ at\ s = cs)$ with itself, we obtain $(locked\ s)$. Also, by the *exit* axioms, we get

$$k\ at\ s' = (k\ exit\ s) = rs$$

   i.e.,

$$k\ at\ s' = rs. \tag{19.2}$$

   The *exit* axioms also imply that s' is *not* locked. We can now conclude that

$$i \neq k \tag{19.3}$$

[3] Clearly, if neither of these hold, i.e., if $i = k$ and $j = k$, then we could also have $i = j$, contradicting our hypothesis.

because otherwise, if $i = k$, the assumption that i is in cs in state s' would contradict (19.2). Hence, by the *exit* axioms, we get

$$i \; at \; s' = i \; at \; s. \tag{19.4}$$

Therefore, from (19.4) and the assumption that i is in cs in s', we get $i \; at \; s = cs$. But now applying the inductive hypothesis to $i \; at \; s = cs$ and to $(k \; at \; s = cs)$ yields $i = k$, contradicting (19.3).

2. *Case 2*: ($k \; at \; s \neq$ cs). In that case the *exit* axioms give ($k \; exit \; s = s$, i.e., $s' = s$, and the result follows directly from the inductive hypothesis.

The above informal proof can be formulated in Athena at the same level of abstraction and with the exact same structure. Moreover, the proof colloquialism "the reasoning for that case is symmetric" that appears in the *enter* transition can be directly accommodated by abstracting the symmetric reasoning into a *method* and then applying that method to multiple instances. Likewise, the treatment of *enter* and *exit* is symmetric when their effective conditions are violated, in which case the result follows directly from the inductive hypothesis, and this commonality too can be easily factored out into a general method. The entire proof, along with these two methods, can be seen below. Note that the proof doesn't use external theorem provers. Instead, it uses Athena's own library *chain* method, which allows for limited proof search. The *chain* method extends the readability benefits of equational chains into arbitrary implication chains.

```
# This is the ''symmetric'' reasoning that appears in two
# places in the treatment of enter. We abstract it here
# into one single generic method M.

define (M inequality enabled-premise IH) :=
  match [inequality enabled-premise] {
    [(~ (i = k)) (((k at s) = rs) & (~ (locked s)))] =>
      (!chain->
       [inequality
    ==> (enabled-premise & inequality) [augment]
    ==> (i at k enter s = i at s)      [enter-axioms]
    ==> (i at s = i at k enter s)      [sym]
    ==> (i at s = cs)                  [(i at k enter s = cs)]
    ==> (i at s = cs & i at s = cs)    [augment]
    ==> (locked s)                     [IH]
    ==> (locked s & ~ locked s)        [augment]
    ==> false                          [prop-taut]])
  }

# This method handles all cases where the effective condition
# is violated.

define (direct-ih hyp s failed-ec IH transition-axioms) :=
 match hyp {
   (((i at s') = cs) & ((j at s') = cs)) =>
     let {s=s' := (!chain->
                   [failed-ec ==> (s' = s) [transition-
                                            axioms]])}
       (!chain-> [hyp ==> (i at s = cs & j at s = cs) [s=s']
```

```
                        ==> (i = j & locked s)          [IH]
                        ==> (i = j & locked s')         [s=s']])
  }

# The main proof:

by-induction (forall s . gp s) {
  init => (!spf (gp init) (ab))
| (s' as (k enter s)) =>
   pick-any i:Pid j:Pid
    let {IH := (forall i j . i at s = cs & j at s = cs
                          ==> i = j & locked s)}
    assume hyp := (i at s' = cs & j at s' = cs)
     conclude goal := (i = j & locked s')
      (!two-cases
       assume case1 := (k enabled-at s)
        let {s'-locked := (!chain-> [case1
                             ==> (locked s') [enter-axioms]]);
             s-not-locked := (!chain-> [case1
                             ==> (~ locked s) [right-and]]);
             i=j := (!by-contradiction (i = j)
                      assume h := (i =/= j)
                   let {D := {(i =/= k | j =/= k) from h}}
                        (!cases D
                          assume i=/=k := (i =/= k)
                           (!M i=/=k case1 IH)
                          assume j=/=k := (j =/= k)
                           (!M j=/=k case1 IH)))}
        (!both i=j s'-locked)
       assume case2 := (~ k enabled-at s)
        (!direct-ih hyp s case2 IH enter-axioms))
| (s' as (k exit s)) =>
  pick-any i:Pid j:Pid
   let {IH := (forall i j . i at s = cs & j at s = cs
                         ==> i = j & locked s)}
    assume hyp := (i at s' = cs & j at s' = cs)
     conclude goal := (i = j & locked s')
      (!two-cases
       assume case1 := (k at s = cs)
        (!by-contradiction goal
         assume -goal := (~ goal)
          let {locked-s :=
                (!chain-> [case1
                        ==> (case1 & case1) [augment]
                        ==> (locked s)      [IH]]);
               p2 := (!chain-> [case1
                             ==> (k at s' = rs) [exit-axioms]]);
               i=/=k :=
                 (!by-contradiction (i =/= k)
                  assume h := (i = k)
                   (!chain-> [p2
                         ==> (i at s' = rs) [h]
                         ==> (cs = rs)      [(i at s' = cs)]
                         ==> (cs = rs & cs =/= rs)  [augment]
                         ==> false             [prop-taut]]))}
          (!chain-> [i=/=k
                ==> (i at s' = i at s)  [exit-axioms]
                ==> (i at s = cs)       [(i at s' = cs)]
                ==> (i at s = cs & k at s = cs) [augment]
```

```
                ==> (i = k)                              [IH]
                ==> (i = k & i =/= k)                    [augment]
                ==> false                                [prop-taut]]))
        assume case2 := (k at s =/= cs)
         (!direct-ih hyp s case2 IH exit-axioms))
}
```

References

1. Diaconescu, R., Futatsugi, K., Ogata, K.: CafeOBJ: Logical foundations and methodologies. Comput. Inform. **22**, 257–283 (2003)
2. Clavel, M., Durn, F., Eker, S., Lincoln, P., Mart-Oliet, N., Meseguer, J., Quesada, J.: Maude: specification and programming in rewriting logic. Maude System documentation (1999)
3. Mossakowski, T., Haxthausen, A.E., Sannella, D., Tarlecki, A.: Casl the Common Algebraic Specification Language. In: Logics of Specification Languages. Part of the series Monographs in Theoretical Computer Science pp. 241–298 (2008)
4. Nipkow, T.: Programming and Proving in Isabelle/HOL. Technical Report (2014)
5. Autexier, S., Mossakowski, T.: Integrating HOL-CASL into the development graph manager MAYA. Frontiers of combining systems. Lect. Notes Comput. Sci. **2309**, 2–17 (2002)
6. Codescu, M., Horozal, F., Kohlhase, M., Mossakowski, T., Rabe, F., Sojakova, K.: Towards Logical Frameworks in the Heterogeneous Tool Set Hets. In Till Mossakowski, Hans-Jrg Kreowski (eds.), Recent Trends in Algebraic Development Techniques, 20th International Workshop, WADT 2010, vol. 7137, 139–159, Lecture Notes in Computer Science. Springer, Berlin (2010)
7. Arkoudas, K.: Athena, proofcentral.org, (2004)
8. CafeOBJ Algrebraic Specification and Verification. https://cafeobj.org/
9. CafeOBJ@NTUA blog. https://cafeobjntua.wordpress.com/2016/05/26/on-combining-algebraic-specications-with-first-order-logic-via-athena/
10. Smith, M., Klarlund, N.: Verification of a Sliding Window Protocol Using IOA and MONA. Research Report RR-3959, INRIA (2000)
11. Goguen, J., Malcolm, G.: A hidden agenda. Technical Report No. CS97-538, Ed.: University of California at San Diego (1997)
12. Ogata, K., Futatsugi, K.: Compositionally writing proof scores of invariants in the OTS/-CafeOBJ method. J. Univers. Comput. Sci. **19**(6), 771–804 (2013)
13. Futatsugi, K., Gaina, D., Ogata, K.: Principles of proof scores in CafeOBJ. Theor. Comput. Sci. **464**, 90112 (2012)
14. Gaina, D., Lucano, D., Ogata, K., Futatsugi, K.: On automation of OTS/CafeOBJ method. In: Specification, Algebra, and Software, LNCS **8373**, 578–602 (2014)
15. Ogata, K., Futatsugi, K.: Proof scores in the OTS/cafeOBJ method. In Proceedings of the Conference on Formal Methods for Open Object-Based Distributed Systems, vol. 2884 170–184 (2003)
16. Arkoudas, K., Musser, D.: Fundamental Proof Methods in Computer Science. MIT Press (2017)
17. Musser, D.: Understanding Athena Proofs
18. Vampire, Web page. www.vprover.org/
19. Spass, Web page. www.spass-prover.org/
20. Algebraic modeling of topological and computational structures (AlModTopCom). http://www.math.ntua.gr/~sofia/ThalisSite/publications.html
21. Ouranos, I., Ogata, K., Stefaneas, P.: TESLA Source Authentication Protocol Verification Experiment in the Timed OTS/CafeOBJ Method: Experiences and Lessons Learned. IEICE Trans. **97**(5), 1160–1170 (2014)

Chapter 20
A Rule-Based Approach for Air Traffic Control in the Vicinity of the Airport

Theodoros Mitsikas, Petros Stefaneas and Iakovos Ouranos

Abstract The constantly augmenting loads of the aviation industry inevitably define the evolutions in the field of Air Traffic Control. Relevant regulations are changing, in order to accommodate the increase in passenger, flight, and cargo numbers. This paper presents the design of an innovative rule base for the Air Traffic Control regulations during the take-off and landing phases, covering both current and future separation standards of ICAO and FAA. The rule base consists of the rules implementing the Air Traffic Control regulations, and a database containing characteristics of airports and aircraft. The proposed rule base constitutes a flexible tool for the computation of the aircraft separation according to current and future regulations, useful in the fields of conflict detection, conflict avoidance, and scheduling aircraft landings. A further application will be as a decision support tool in real-time environments, guaranteeing the enforcement of all the separation standards.

Keywords RuleML · Rule base · Aircraft separation · Air traffic control

AMS Subject Classifications 68Q60 · 68N30 · 68T30 · 03B70

20.1 Introduction

Air Traffic Control (ATC) is an important factor for the airliner operations, guiding the aircraft on the air and on the ground. ATC is responsible for organizing the air traffic flow in an efficient way, while ensuring the safety.

T. Mitsikas (✉) · P. Stefaneas
Department of Applied Mathematics, National Technical University of Athens, Zografou Campus, 15780 Athens, Greece
e-mail: mitsikas@central.ntua.gr

P. Stefaneas
e-mail: petros@math.ntua.gr

I. Ouranos
HCAA, Heraklion International Airport "N.Kazantzakis", 71601 Heraklion, Greece
e-mail: iouranos@central.ntua.gr

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_20

The *separation* of aircraft is the concept of keeping a minimum distance between aircraft to avoid collisions and possible accidents caused by *wake turbulence*. The lift that the aircraft's wing is designed to produce directly affects the intensity and lifespan of the generated vortex. Therefore, the *separation minima* is based upon wake vortex categories of the preceding and the following aircraft which, in turn, are derived from the *maximum takeoff weight* (MTOW) or the *maximum takeoff mass* (MTOM) [8, 10, 13, 14].

Due to the increasing traffic and congested airports, regulation changes are currently being planned, with the aim of increasing airport capacity [6, 12, 14, 17]. The first step is the wider adaptation of RECAT, which recategorizes aircraft and sets new standards for wake turbulence separation minima. Therefore, the wingspan is used as an additional to MTOW/MTOM parameter. As a result, aircraft are placed into six wake vortex categories, common for departure and arrival separation, which enhance both safety and efficiency [8, 11, 14]. The second step is a static separation matrix of distance and time for both arrivals and departures for the common commercial aircraft, called RECAT-2 [7]. The third step, RECAT-3, will provide dynamic pair-wise spacing that will vary with atmospheric conditions and aircraft performance [7]. However, if in the future the concepts of *Free Flight* and *Self-Separation* are employed, the management of air traffic will be revolutionized. They consist of decentralized methods of ATC, using computer communication to reserve parts of the airspace dynamically and automatically in a distributed way [20].

The above mentioned changes mandate the modernization of the current infrastructure, as their application necessitates the operation of relevant computer systems. For example, a specialized tool will be required to apply RECAT-2, because of the size of the separation matrix. Besides, after the realization of Free Flight and Self-Separation concepts, conflict avoidance will be totally automated, relying only on computer systems and computer communication between aircraft.

Our research project aspires to support the evolution of ATC. We propose a rule-based approach to model ATC regulations in the terminal airspace. Our approach allows the validation and verification of its formal properties [15], without compromising the compliance with the regulations. We have used the reference Naf Hornlog implementation of RuleML rule language, OO jDREW, taking advantage of its suitable built-in predicates and functions, and its vast multi-agent and distributed systems compatible API. Furthermore, our approach can serve as a tool for the application of both current and future regulations on the field.

Our main contributions are as follows: (i) we developed rules, derived from the regulations, for the definition of wake vortex categories; (ii) we developed rules concerning the separation minima during the landing phases; (iii) a large database of more than 200 types of aircraft was developed, containing the characteristics required to compute the above. Additionally, we developed a reference airport database, containing characteristics of airports in Greece; (iv) we developed the mathematical background within the rule base, capable of handling the heading of aircraft as well as angles and angle comparison, as a base for future work; (v) we developed rules and a methodology of handling not fully established future regulations, covering different orientations that may follow.

The rest of the paper is organized as follows. Section 20.2 presents related work, while Sect. 20.3 introduces RuleML. Section 20.4, the main part of our paper, presents the design of the rule-base. Finally, Sect. 20.5 concludes the paper and proposes future work.

20.2 Related Work

To the best of our knowledge, no strict rule-based approach for ATC in the airport area exists. However, a convergence point can be found in other lines of research concerning decision support tools, landing scheduling, and Free Flight. Below, the most related previous work is discussed.

Prototype decision support tools for terminal area and controllers, such as FAST [5], have been developed and evaluated in operation with live air traffic. More complete systems have been developed and tested, such as [17]. An overview of a method for formal requirements capture and validation, in the domain of oceanic ATC is presented in [16]. The obtained model focuses on conflict prediction, while being compliant to the regulations governing aircraft separation in oceanic airspace. The design approach, the specification structure, and some examples of the rules and axioms of the formal specification are provided. Those examples, expressed in Many-Sorted First Order Logic or in the Prolog notation, include rules about conflict prediction and aircraft separation. Supplementary, the model was validated by automated processes, formal reasoning and domain experts.

The objective of landing scheduling is to position all arriving aircraft to a runway and to a specific time window, while respecting separation constraints and minimizing delays [2, 5, 6, 18, 19, 21]. Different algorithms and heuristics are used, the simplest one being first-come first-served (FCFS). Aircraft separation is used as a scheduling constraint. However, it has to be underlined that the regulations requirements are not strictly followed, but constant values or simplified forms are used.

Research in the field of Free Flight concept focuses on model principles and different algorithms for conflict detection and avoidance. In [20], a general platform (NAMA), which is oriented towards agent-based Free Flight implementations, was presented. In the same work, a conflict detection and resolution, based on utility functions without any negotiations between agents, was proposed. However, the separation constraints were not explained. In [12], the authors focused on airport airspace. They developed a multi-agent architecture, and a software prototype. The latter implements an ATC system with distributed policy of conflict resolution, predictive analysis and P2P interaction-based autonomous coordination of aircraft's motions. Nevertheless, aircraft classes serve only for the estimation of airspeed range and not for the differences in separation between classes.

20.3 RuleML Basics

RuleML is a markup language, designed to express both top–down and bottom–up rules in XML schema; a shorthand is POSL, that follows a Prolog-like syntax. More specifically, we used OO jDREW, a RuleML implementation that follows Horn Clauses in implication form, supports negation as failure, and is written in the Java programming language.

A Horn Clause in implication form is written as $h \leftarrow p \wedge q \wedge \ldots \wedge r$, where h, p, q, r are atoms. An atom is of the form $r(t_1, \ldots, t_n)$, where r is a predicate of arity n, and t_i are terms.

In POSL, a clause has the general form:

```
h :- p,q,...,r.
```

The head of the clause is h and the body is $p, q, \ldots, r$. A clause is called a fact if the body is empty. Neither disjunction nor negation is supported on the body [3]. In this paper, we take advantage of OO jDREW's built-in equality and inequality predicates, as well as of math functions. The English sentence: *"A customer is premium if their spending has been min 5000 euro in the previous year"*, part of the classic RuleML example which can be found in [4], can be written in POSL form, using inequality predicates, as follows:

```
premium(?customer) :-
 spending(?customer, ?amount,"previous year"),
 greaterThanOrEqual(?amount, 5000^^Real).
```

The relation atoms are `premium` and `spending`, `?customer` and `?amount` are variables representing customer's name and amount spent, respectively, while `"previous year"` and `5000` are constants and `^^Real` is the type. The built-in predicate `greaterThanOrEqual` is satisfied iff the first argument is equal to or greater than the second argument. By asserting e.g., the following fact in the Knowledge Base:

```
spending(Peter, 6000^^Real, "previous year").
```

the query `premium(?x)` binds `Peter` to the variable `?customer`, as shown in Fig. 20.1.

The notation used in this paper is: (i) variables are denoted by x_k, (ii) constants and bindings are lower or upper case alphanumericals, (iii) atoms, predicates and functions can be either in the form $R_k^n(t_1, \ldots t_n)$, or in the simplified form —e.g. $(t_1 < t_2)$— for equality and inequality predicates.

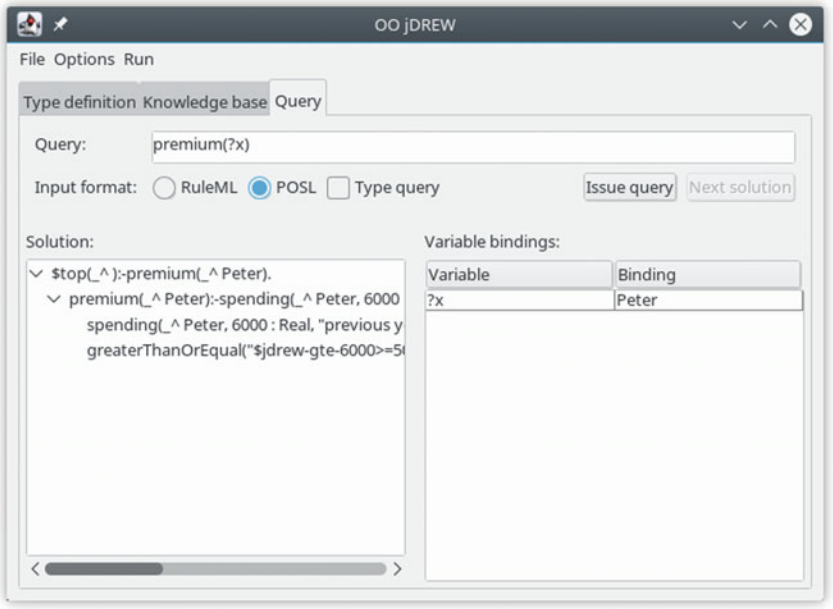


Fig. 20.1 The query execution tab of OO JDREW

20.4 The Rule Base

In this section we present the design of our approach. We express in terms of rules, ATC regulations concerning aircraft separation standards in the airport vicinity under instrument flight rules (IFR). We follow weight classes and the separation of aircraft of different classes according to International Civil Aviation Organization (ICAO) and Federal Aviation Administration (FAA) regulations. Additionally, we decscribe rules regarding airport layout, heading and turns, and weather conditions in order to cover future regulations or complex cases.

The design of the rule base aims to (i) implement ATC regulations adhering to ICAO and FAA standards. (ii) study complex cases (iii) be easily integrated with any existing or future framework (iv) be flexibly adapted to each framework's requirements.

Handling intersection departures, intersecting runways, intersecting flight path operations and parallel approaches is presented as example of how our proposed rule base efficiently solves complex cases.

20.4.1 Database

The database, written in RuleML/POSL, contains more than 200 types of aircraft defined by ICAO aircraft designation (a three- or four-character alphanumeric code designating aircraft), with the characteristics needed to compute the separation minima. The data were obtained from [9]. Each entry has a general form of an atom of arity 4 as follows:

$$aircraftChar(x_{type},\ x_{kg},\ x_{ft},\ x_{kt}) \tag{20.1}$$

where x_{type} denotes the type of the aircraft as defined by ICAO aircraft designation, x_{kg} denotes the MTOM in kilograms, x_{ft} the wingspan of the aircraft measured in feet, and x_{kt} is the approach speed, measured in knots. The corresponding code, for the Airbus A321 (ICAO aircraft designation code: `A321`), in RuleML/POSL, is as follows:

```
aircraftChar(A321,93500^^Real,111.9^^Real,138^^Real).
```

For each characteristic an individual atom of arity 1 is obtained by the following Horn Clauses, written in an implication form:

$$mtom(x_{kg}) \leftarrow aircraftChar(x_{type},\ x_{kg},\ x_{ft},\ x_{kt}) \wedge aircraft(x_{type}) \tag{20.2}$$

$$wingspan(x_{ft}) \leftarrow aircraftChar(x_{type},\ x_{kg},\ x_{ft},\ x_{kt}) \wedge aircraft(x_{type}) \tag{20.3}$$

$$appSpeed(x_{kt}) \leftarrow aircraftChar(x_{type},\ x_{kg},\ x_{ft},\ x_{kt}) \wedge aircraft(x_{type}) \tag{20.4}$$

where $aircraft(x_{type})$ must be defined —e.g. `aircraft(A321)` for the Airbus A321— in the knowledge base as a fact.

Similar atoms are defined for the preceding aircraft:

$$mtomPreceding(x_{kg}) \leftarrow aircraftChar(x_{type},\ x_{kg},\ x_{ft},\ x_{kt}) \wedge precedingAircraft(x_{type}) \tag{20.5}$$

$$wingspanPreceding(x_{ft}) \leftarrow aircraftChar(x_{type},\ x_{kg},\ x_{ft},\ x_{kt}) \wedge precedingAircraft(x_{type}) \tag{20.6}$$

where $precedingAircraft(x_{type})$ must be defined as a fact.

Furthermore, the database contains a reference implementation of airports. Each entry in the database is of the form:

$$airportChar(x_{code}, x_{or_1}, x_{or_2}, x_{or_3}, x_{or_4}, x_{name_1}, x_{name_2}, x_{name_3}, x_{name_4}, x_d, x_{rules}) \quad (20.7)$$

where x_{code} is representing the ICAO airport code (a four-character alphanumeric code designating airports), x_{or_1}, x_{or_2}, x_{or_3} and x_{or_4} are representing the the exact heading of each runway in degrees in ascending order respectively, x_{name_1}, x_{name_2}, x_{name_3}, and x_{name_4} are the variables representing the runway designation, x_d is the distance between the two runways and, the regulations applicable to each airport are denoted by x_{rules}.

Those characteristics are needed to compute the airport layout, the heading during landing, and the initial heading after take-off. Currently, the rule base supports up to two runways in both directions.

As above, atoms of arity 1 are obtained for each characteristic, by the following Horn Clauses, written in an implication form:

$$runwayOneOr(x_{or_1}) \leftarrow airportChar(x_{code}, x_{or_1}, \ldots) \wedge airportName(x_{code}) \quad (20.8)$$

$$\vdots$$

$$runwayOneName\ (x_{name_1}) \leftarrow airportChar(x_{code}, \ldots, x_{name_1}, \ldots) \wedge airportName(x_{code}) \quad (20.9)$$

$$\vdots$$

$$distanceBtwnRunw(x_d) \leftarrow airportChar(x_{code}, .., x_d, ..) \wedge airportName(x_{code}) \quad (20.10)$$

where $airportName(x_{code})$ must be defined as a fact in the knowledge base, e.g. `aiportName(LGTS)` for the Thessaloniki Airport "Macedonia" (ICAO airport code: LGTS), an airport with two intersecting runways (10/28 and 16/34). The corresponding entry in the database, in RuleML/POSL, is

```
airportChar(LGTS, 103.9^^Real, 166.3^^Real,
 283.9^^Real, 346.3^^Real, 10^^String, 16^^String,
 28^^String, 34^^String, 0^^Real, icao).
```

Weather information can be useful for simultaneous operations on different runways, or for future expansion to support RECAT-3, Weather Dependent Separations (WDS), and Time-Based Separation (TBS) in strong headwind conditions [7]. The form of the atoms in the database concerning weather information is as follows:

$$weather(x_{code}, x_{kt}, x_{dir}) \tag{20.11}$$

where x_{code} is representing the ICAO airport code, x_{kt} denotes the wind speed, measured in knots, and x_{dir} denotes the wind direction.

The following is an example according to METAR information of March, 21^{st}, 18:20 UTC, for LGTS:

$$weather(LGTS,\ 3,\ variable). \tag{20.12}$$

As above, atoms of arity 1 can be obtained by the following Horn Clauses, written in an implication form:

$$windSpeed(x_{kt}) \leftarrow weather(x_{code}, x_{kt}, x_{dir}) \wedge airportName(x_{code}) \tag{20.13}$$

$$windDirection(x_{dir}) \leftarrow weather(x_{code}, x_{kt}, x_{dir}) \wedge airportName(x_{code}) \tag{20.14}$$

20.4.2 ICAO Regulations

Current regulations of ICAO categorize aircraft as follows [13, 14]:

Light MTOM of 7000 kg or less.
Medium MTOM of greater than 7000 kg, but less than 136000 kg.
Heavy MTOM of 136000 kg or greater.
Super - A separate designation that currently only refers to the Airbus A380 (MTOM 575000 kg, ICAO designation `A388`).

The categorization can be specified using the following Horn Clauses, (for **Light** and **Medium** categories):

$$icaoCategory(light) \leftarrow mtom(x_{kg}) \wedge (x_{kg} \leq 7000) \tag{20.15}$$

$$icaoCategory(medium) \leftarrow mtom(x_{kg}) \wedge (x_{kg} > 7000) \wedge (x_{kg} < 136000) \tag{20.16}$$

Airbus A380 would normally belong to **Heavy** category. Consequently, the definition of **Heavy** and **Super** classes is additionally using the aircraft type.

$$icaoCategory(heavy) \leftarrow mtom(x_{kg}) \wedge (x_{kg} \geq 136000) \wedge aircraft(x_{type}) \wedge (x_{type} \neq A380) \tag{20.17}$$

$$icaoCategory(super) \leftarrow aircraft(A380) \tag{20.18}$$

Table 20.1 Current ICAO weight categories and associated separation minima [14]

ICAO separation standards (nautical miles)					
		Follower			
		Super	Heavy	Medium	Light
Leader	Super	MRS	6	7	8
	Heavy	MRS	4	5	6
	Medium	MRS	MRS	MRS	5
	Light	MRS	MRS	MRS	MRS

MRS: Minimum Radar Separation

Inequality predicates are built-ins of OO jDREW. The computation of aircraft's category can be made by defining the type of an aircraft, which must be present at the database as a fact —e.g. `aircraft(A321)`— and executing the query `icaoCategory(?x)`. Similar rules exist for the preceding aircraft, defining the atom $icaoCategoryPreceding(x_{class})$.

ICAO separation standards for flights on instrument flight rules (IFR) are presented at Table 20.1.

Those separations standards can be represented by Horn Clauses. For instance, the derived Horn Clauses for the categories **heavy**, **medium** are as follows:

$$icaoSeparation(mrs) \leftarrow icaoCategory(heavy) \wedge icaoCategoryPreceding(medium) \tag{20.19}$$

$$icaoSeparation(5) \leftarrow icaoCategory(medium) \wedge icaoCategoryPreceding(heavy) \tag{20.20}$$

$$icaoSeparation(mrs) \leftarrow icaoCategory(medium) \wedge icaoCategoryPreceding(medium) \tag{20.21}$$

$$icaoSeparation(4) \leftarrow icaoCategory(heavy) \wedge icaoCategoryPreceding(heavy) \tag{20.22}$$

By defining the type of the preceding and following aircraft, the needed separation can be obtained by executing the query `icaoSeparation(?x)`.

20.4.3 FAA Regulations

The methodology for constructing the rules concerning aircraft classes and separation according to FAA regulations is similar. The FAA is using the following classes [10]:

Small - Aircraft of 41000 pounds (19000 kg) or less MTOW.
Large - Aircraft of more than 41000 pounds (19000 kg) MTOW, up to, but not including, 300000 pounds (140000 kg).
Heavy - Aircraft capable of takeoff weights of 300000 pounds (140000 kg) or more.
Super - A separate designation that currently only refers to the Airbus A380 and the Antonov An-225
B757 - Different separation standards are applied for the Boeing 757.

The Horn Clauses defining the above regulations, after conversion of pounds to kilograms has been applied, are:

$$faaCategory(small) \leftarrow mtom(x_{kg}) \wedge (x_{kg} \leq 19000) \tag{20.23}$$

Boeing 757 would normally belong to **Large** class, while Airbus A380 and Antonov A225 would belong to **Heavy** class. Therefore, similar to ICAO categorization, the definition of those classes additionally needs the aircraft type:

$$faaCategory(large) \leftarrow aircraft(x_{type}) \wedge (x_{type} \neq B757) \wedge mtom(x_{kg}) \wedge (x_{kg} > 7000) \wedge (x_{kg} < 136000) \tag{20.24}$$

$$faaCategory(heavy) \leftarrow mtom(x_{kg}) \wedge (x_{kg} \geq 136000) \wedge aircraft(x_{type}) \wedge (x_{type} \neq A380) \wedge (x_{type} \neq A225) \tag{20.25}$$

$$faaCategory(super) \leftarrow aircraft(A380) \vee aircraft(A225) \tag{20.26}$$

$$faaCategory(B757) \leftarrow aircraft(B757) \tag{20.27}$$

The lack of disjunction in the head of Horn Clauses and in RuleML/POSL, mandates two rules for **B757** and **Super** classes, one for each ICAO aircraft code designation (`B752`, `B753` for Boeing 757, `A225` for Antonov An-225). Executing the query `faaCategory(?x)`, answers the class of the aircraft. Similar rules exist for the preceding aircraft, defining the atom $faaCategoryPreceding(x_{class})$.

The separation standards at the runway threshold for flights under IFR are defined by the Table 20.2.

The derived Horn Clauses for e.g. the categories **heavy**, **B757**, are as follows:

$$faaSeparation(5) \leftarrow faaCategory(B757) \wedge faaCategoryPreceeding(heavy) \tag{20.28}$$

$$faaSeparation(4) \leftarrow faaCategory(heavy) \wedge faaCategoryPreceeding(B757) \tag{20.29}$$

$$faaSeparation(4) \leftarrow faaCategory(heavy) \wedge faaCategoryPreceeding(heavy) \tag{20.30}$$

Table 20.2 FAA wake separation standards (nautical miles, at the threshold) [8]

Leader/Follower	Super	Heavy	B757	Large	Small
Super	MRS	6	7	7	8
Heavy	MRS	4	5	5	6
B757	MRS	4	4	4	5
Large	MRS	MRS	MRS	MRS	4
Small	MRS	MRS	MRS	MRS	MRS
MRS: minimum radar separation					

$$faaSeparation(4) \leftarrow faaCategory(B757) \wedge faaCategoryPreceeding(B757) \tag{20.31}$$

20.4.4 RECAT Regulations

For the purposes of wake turbulence separation minima, aircraft are categorized as Category A through Category F. Each aircraft is assigned a category based on wingspan, and maximum takeoff weight (MTOW) [8, 11]:

Category A. Aircraft capable of MTOW of 300,000 pounds or more and wingspan greater than 245 feet.
Category B. Aircraft capable of MTOW of 300,000 pounds or more and wingspan greater than 175 feet and less than or equal to 245 feet.
Category C. Aircraft capable of MTOW of 300,000 pounds or more and wingspan greater than 125 feet and less than or equal to 175 feet.
Category D. Aircraft capable of MTOW of less than 300,000 pounds and wingspan greater than 125 feet and less than or equal to 175 feet; or aircraft with wingspan greater than 90 feet and less than or equal to 125 feet.
Category E. Aircraft capable of MTOW greater than 41,000 pounds with wingspan greater than 65 feet and less than or equal to 90 feet.
Category F. Aircraft capable of MTOW of less than 41,000 pounds and wingspan less than or equal to 125 feet, or aircraft capable of MTOW less than 15,500 pounds regardless of wingspan, or a powered sailplane.

The derived definite clauses in implication form, after conversion of pounds to kilograms are as follows:

$$recat(A) \leftarrow mtom(x_{kg}) \wedge (x_{kg} \geq 136000) \wedge wingspan(x_{ft}) \wedge (x_{ft} > 245) \tag{20.32}$$

$$recat(B) \leftarrow mtom(x_{kg}) \wedge (x_{kg} \geq 136000) \wedge wingspan(x_{ft}) \wedge (x_{ft} \leq 245) \wedge (x_{ft} > 175) \tag{20.33}$$

$$recat(C) \leftarrow mtom(x_{kg}) \wedge (x_{kg} \geq 136000) \wedge wingspan(x_{ft}) \wedge (x_{ft} \leq 175) \wedge (x_{ft} > 125) \quad (20.34)$$

$$recat(D) \leftarrow (wingspan(x_{ft}) \wedge (x_{ft} \leq 125) \wedge (x_{ft} > 90)) \vee (mtom(x_{kg}) \wedge (x_{kg} < 136000) \wedge wingspan(x_{ft}) \wedge (x_{ft} \leq 175) \wedge (x_{ft} > 125)) \quad (20.35)$$

$$recat(E) \leftarrow mtom(x_{kg}) \wedge (x_{kg} > 18000) \wedge wingspan(x_{ft}) \wedge (x_{ft} \leq 90) \wedge (x_{ft} > 65) \quad (20.36)$$

$$recat(F) \leftarrow (mtom(x_{kg}) \wedge (x_{kg} < 7000)) \vee (mtom(x_{kg}) \wedge (x_{kg} \leq 18000) \wedge wingspan(x_{ft}) \wedge (x_{ft} \leq 125)) \quad (20.37)$$

The lack of disjunction in the body of Horn Clauses and in RuleML/POSL, mandates two rules for categories **D** and **F**. Executing the query `recatCategory(?x)` answers about aircraft's category. Similar rules exist for the preceding aircraft, defining the atom $recatPreceding(x_{class})$.

RECAT separation standards for IFR flights are presented in Table 20.3. For instance, the separation values for the pair **B**, **C** are captured from the following Horn Clauses:

$$recatSeparation(mrs) \leftarrow recat(B) \wedge recatPreceding(C) \quad (20.38)$$

$$recatSeparation(4) \leftarrow recat(C) \wedge recatPreceding(B) \quad (20.39)$$

Table 20.3 RECAT wake separation standards (nautical miles) [8, 11]

		Follower					
		A	B	C	D	E	F
Leader	A	MRS	5	6	7	7	8
	B	MRS	3	4	5	5	7
	C	MRS	MRS	MRS	3.5	3.5	6
	D	MRS	MRS	MRS	MRS	MRS	4
	E	MRS	MRS	MRS	MRS	MRS	MRS
	F	MRS	MRS	MRS	MRS	MRS	MRS

MRS: Minimum Radar Separation

20.4.5 Airport Layout

ICAO and FAA rules are covering cases such as intersecting runway/intersecting flight path separation, intersection departures, parallel approaches, etc. Furthermore, different separation standards may exist when aircraft operate in different runways separated by at least 2500 ft or 760 m [10, 13]. Finally, the rule base must cover situations where a runway is closed for maintenance of for emergency reasons. In order to implement the above, it is necessary to define rules concerning the airport layout.

Reasoning on airport layout is realized by using information about airport characteristics described in Sect. 20.4.1, combined with additional rules. Currently, the rule base supports up to two runways, used both ways. In accordance with ICAO and FAA regulations, the basic layouts supported are: (i) single runway, (ii) two intersecting runways, (iii) two close parallel runways, less than 2500 feet (760 m for ICAO regulations) apart, (iv) parallel runways more than or equal to 2500 feet, and (v) non parallel and non intersecting runways, denoted as `npni`.

$$layout(closeparallel) \leftarrow (R_1(x_1) \wedge R_2(x_2) \wedge (x_1 = x_2)) \wedge R_d(x_3 < 2500) \tag{20.40}$$

$$layout(farparallel) \leftarrow (R_1(x_1) \wedge R_2(x_2) \wedge (x_1 = x_2)) \wedge R_d(x_3 \geq 2500) \tag{20.41}$$

$$layout(npni) \leftarrow (R_1(x_1) \wedge R_2(x_2) \wedge (x_1 \neq x_2)) \wedge R_d(x_3 \neq 0) \tag{20.42}$$

where R_{layout} is defined by `layout`, R_1 is the heading of the first runway, defined as `runwayOneOr`, R_2 is the heading of the second runway, defined as `runwayTwoOr` and R_d is the distance between runways, defined by the atom `distanceBtwnRunw`. Equality predicate is built-in of OO jDREW. The query `layout(?x)` is shown in Fig. 20.2.

20.4.6 Angles and Heading

Solving complex cases, as mentioned in Sect. 20.4.5, require reasoning over heading and turns, as well as weather information. Representing spatial information for the purpose of an ATC rule base mandates the use of quantitative terms for angle and heading, since qualitative spatial terms cannot be used for precise arithmetic calculations. Previous work (e.g. [1]), can be useful only in cases of fuzzy terms, as wind heading variations as expressed by METAR information.

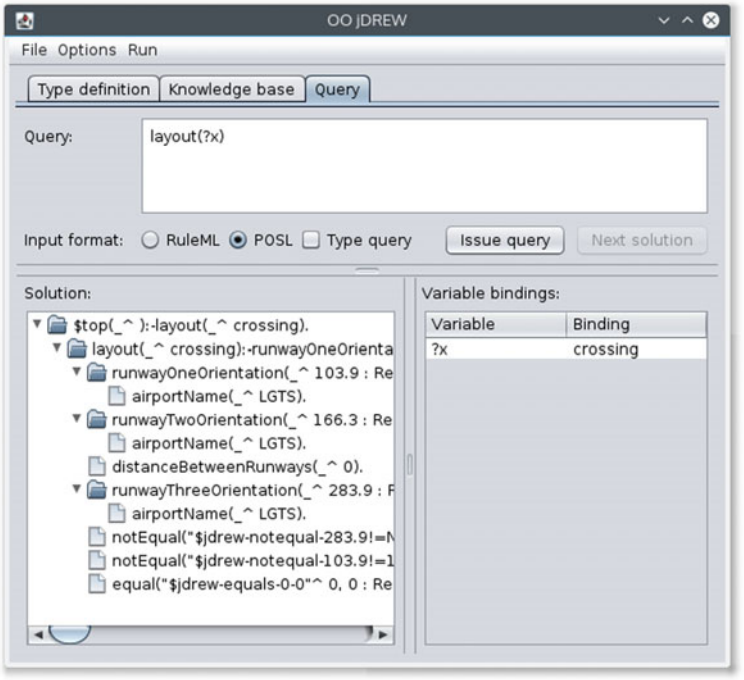


Fig. 20.2 Executing the query about airport layout, for the Thessaloniki Airport "Macedonia"

During landing phase, aircraft's heading is derived from the orientation of the runway on which the aircraft operates (denoted by the atom $onRunway(x_2)$). Four rules are needed to match runway name with the corresponding heading.

$$heading(x_{deg}) \leftarrow runwayOneOr(x_{deg}) \wedge runwayOneName(x_1) \wedge onRunway(x_2) \wedge (x_1 = x_2) \tag{20.43}$$

Heading and angle changes require two Horn Clauses for each computation, due to the possible, greater than 360° result. Angle subtraction also requires additional rules, covering cases from 0° to 180° and from 180° to 360°. Negative values have been taken in to consideration, using the built-in function $abs(x_{abs}, x_1)$, where, given that the first argument is a variable then it will be bound to the absolute value of the second argument.

20.4.7 Future Regulations

The final form of the pair-wise separation matrix of RECAT-2 is not yet known; therefore it was not possible to include those regulations in the rule base. However,

a 120×120 matrix was implemented for testing purposes. The computational time was sufficient for use in real-time environments, or at large simulations ($\sim$5 ms per query on a laptop with AMD A6-6310 APU, 8GB RAM).

Other planned regulation changes as WDS or TBS can be handled by the rule base and the database, according to known information. One advantage of the current rule base is the inclusion of approach speed for each type of aircraft which makes the precise computation of the time-based separation needed possible, in contrast with [12, 17, 21].

20.5 Conclusions and Future Work

A rule base for ATC regulations has been developed for the vicinity of the airport. This rule base, equipped with a large database consisting of characteristics of aircraft and airports, can compute the separation minima during landing phases, as mandated from current and future ATC regulations concerning operations under IFR.

The rule base is derived from interpretation of ATC regulations as Horn Clauses, which allows validation and verification of formal properties. The reasoner used, OO jDREW, provides adequate built-in predicates and functions for the implementation of the subset of ATC regulations studied.

This approach is suitable for building a database of aircraft and airports, containing their characteristics. Furthermore, it is possible to develop rules for categorization of aircraft using only characteristics which are present in the database, according to current ICAO, FAA and RECAT regulations, and subsequently, the implementation of separation tables defined by the above.

Other aspects of regulations and existing models were investigated, leading to the preliminary implementation of a mathematical background capable of handling angles and angle comparison, headings of aircraft, turns, airspeed, airport layout and future regulations.

In the future, we plan to extend the rule base to include a more complete set of regulations, such as cases of separation minima reduction, incident management, and transition zones. Further extensions may include lateral and vertical separation, and 4-D rules, which, given the position, heading, airspeed and time, will generate advisories to avoid possible conflicts. Finally, a collision avoidance method, external or embedded in the rule base, will be necessary to utilize the rule base in Free Flight concepts.

References

1. Batsakis, S.: Reasoning over 2d and 3d directional relations in OWL: a rule-based approach. In: Theory, Practice, and Applications of Rules on the Web, pp. 37–51. Springer Science + Business Media, Berlin (2013)

2. Beasley, J.E., Krishnamoorthy, M., Sharaiha, Y.M., Abramson, D.: Scheduling aircraft landings - the static case. Transp. sci. **34**(2), 180–197 (2000)
3. Boley, H.: Integrating positional and slotted knowledge on the semantic web. J. Emerg. Technol. Web. Intell. **2**(4), 343–353 (2010)
4. Boley, H., Grosof, B., Tabet, S.: RuleML tutorial. In: The RuleML Initiative (2005). http://www.ruleml.org/papers/tutorial-ruleml.html. (Cited July 2016)
5. Davis, T., Krzeczowski, K., Bergh, C.: The final approach spacing tool. In: Automatic Control in Aerospace 1994, pp. 73–79. Elsevier BV, Amsterdam (1995)
6. Erzberger, H.: Design principles and algorithms for automated air traffic management. Knowl. Based. Funct. Aerosp. Syst. **7**, 2 (1995)
7. EUROCONTROL. Future Airport Operations (2015). https://www.eurocontrol.int/articles/future-airport-operations. (Cited July 2016)
8. FAA. Advisory Circular 90-23G - Aircraft Wake Turbulence (2014)
9. FAA. Aircraft Characteristics Database (2009). http://www.faa.gov/airports/engineering/aircraft_char_database/. (Cited July 2016)
10. FAA. Order JO 7110.65V, Air Traffic Control (2014)
11. FAA. Order JO 7110.659C, Wake Turbulence Recategorization (2016)
12. Gorodetsky, V., Karsaev, O., Samoylov, V., Skormin, V.: Multi-Agent Technology for Air Traffic Control and Incident Management in Airport Airspace (2008)
13. ICAO. Doc 4444-RAC/501, Procedures for Air Navigation Services - Rules of the Air and Air Traffic Services (2007)
14. Lang, S., Tittsworth, J., Bryant, W., Wilson, P., Lepadatu, C., Delisi, D., Lai, D., Greene, G.: Progress on an ICAO wake turbulence re-categorization effort. In: AIAA Atmospheric and Space Environments Conference (2010)
15. Ligeza, A.: Principles of Verification of Rule-Based Systems. Springer, Berlin (2006)
16. McCluskey, T., Porteous, J., Naik, Y., Taylor, C., Jones, S.: A requirements capture method and its use in an air traffic control application. Softw. Pract. Exp. **25**(1), 47–71 (1995)
17. Nikoleris, A., Erzberger, H., Paielli, R.A., Chu, Y.-C.: Autonomous system for air traffic control in terminal airspace. In: 14th AIAA Aviation Technology, Integration, and Operations Conference. American Institute of Aeronautics and Astronautics (AIAA) (2014)
18. Soomer, M., Franx, G.J.: Scheduling aircraft landings using airlines preferences. Eur. J. Operat. Res. **190**(1), 277–291 (2008)
19. Tobias, L., Scoggins, J.: Time-based Air-Traffic Management Using Expert Systems. Technical report, NASA Ames Research Center (1986)
20. Valkanas, G., Natsiavas, P., Bassiliades, N.: A collision detection and resolution multi agent approach using utility functions. In: 2009 Fourth Balkan Conference in Informatics. Institute of Electrical and Electronics Engineers (IEEE) (2009)
21. Veidal, E.: Scheduling Aircraft Landings - The Dynamic Case. Master's thesis, Technical University of Denmark, DTU, DK-2800 Kgs. Lyngby, Denmark (2007)

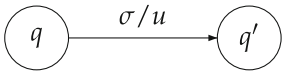
Chapter 21
Sequential Machines and Affine Musical Contours

Marianthi Bozapalidou

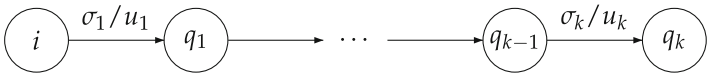
Abstract Affine contours may be viewed as an abstraction of the notion of musical intervals and are closely related to sequential machines. We show that every commutative affine musical contour actually simulates the classical one $c : \mathbb{Z}_{12} \times \mathbb{Z}_{12} \to \mathbb{Z}_{12}$, $c(s,t) = t - s$.

21.1 Sequential Machines and Musical Contours

Sequential machines are mathematical tools suitable to describe and represent various phenomena in music. Classically, a sequential machine $\mathcal{SM}$ can be depicted by a finite directed graph $\mathcal{M}$ whose vertices are its states and whose transitions are of the form



meaning that if $\mathcal{M}$ is in state q and is fed with the input letter σ then it goes to state q' and outputs the string u [3]. In addition the model is deterministic. This implies that for any input string $\sigma_1 \cdots \sigma_k$, there is a unique path



and the emitted string is $u_1 \cdots u_k$ (i denoting the single initial state).

Functions computed by such systems are called sequential and have the fundamental property to preserve initial segments. The preservation of initial segments enables the securing of similarities, necessary for outlining the dynamics of the musical flow, by rendering $\mathcal{SM}$s as a considerable tool to classify musical strings.

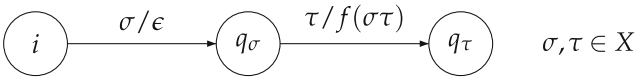
M. Bozapalidou (✉)
Xifilinon 6, 55131 Thessaloniki, Greece
e-mail: bozapalidou@gmail.com

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_21

Let X be an alphabet of musical entities and N a cancellative monoid of minimal features. A contour from X to M is a function $c : X \times X \to M$ assigning a feature $c(x, y)$ to any pair (x, y) of elements of X [1, 10]. Bozapalidou [2] studied contours of the form $c : X \times X \to \mathbb{R}$ (real numbers) counting quantitative features of musical strings.

Consider the contour $c : X \times X \to M$ and its associated function $f : X^* \to M$ defined by the formulas:

- $f(\sigma) = e = f(\varepsilon)$, for all $\sigma \in X$ (e the neutral element of M)
- $f(\sigma_1\sigma_2 \cdots \sigma_k) = c(\sigma_1, \sigma_2) \triangle c(\sigma_2, \sigma_3) \triangle \cdots \triangle c(\sigma_{k-1}, \sigma_k)$, $\sigma_i \in X, k \geq 2$ where the operation $\triangle$ in the right hand side is that of the monoid M. Such a contour function $f : \Sigma^* \to M$ is initial segment preserving and its minimal SM is



21.2 Affine Contours

Following Lewin [8, 9], generalized interval system is a contour $(X, (M, \triangle, e), c)$ which satisfies the additional two axioms:

int_1) for every $x, y, z \in X$

$$c(x, y) \triangle c(y, z) = c(x, z)$$

int_2) for every $x \in X$ and $m \in M$, there exists a unique $y \in X$ such that $c(x, y)=m$.

These two axioms are exactly the definition axioms of an affine space in Geometry, where int_1) is known as the Chasles axiom. Under this light such a system will be more suitably called an *affine contour* (AC).

Proposition 21.1 *If (X, M, c) is an AC, then M is a group.*

Proof Applying int_1) for $x = y = z$ we obtain

$$c(x, x) \triangle c(x, x) = c(x, x) = c(x, x) \triangle e$$

and by left cancellation we get $c(x, x) = e$, where e is the neutral element of M. Furthermore, given $m \in M$, let $x, y \in X$ be such that $c(x, y) = m$ (axiom int_2)). Then

$$m \triangle c(y, x) = c(x, y) \triangle c(y, x) = c(x, x) = e$$

and similarly

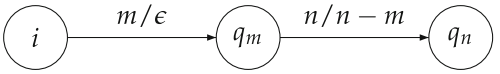
$$c(y,x) \triangle m = c(y,x) \triangle c(x,y) = c(y,y) = e.$$

In other words $c(y,x)$ is the symmetric of the element m and so M is a group.

Example 21.1 The standard AC is $(\mathbb{Z}_{12}, (\mathbb{Z}_{12}, \oplus, 0), dif)$ with

$$dif : \mathbb{Z}_{12} \times \mathbb{Z}_{12} \to \mathbb{Z}_{12}, \quad dif(s,t) = t - s$$

its minimal SM is



$m,n \in \mathbb{Z}_{12}$.

Likewise from any group $(M, \triangle, e)$ we get the AC $(M, (M, \triangle, e), c_M)$ where the interval function $c_M : M \times M \to M$ is given by $c_M(m_1, m_2) = m_2' \triangle m_1'$ (m_1' being the symmetric of m_1 in M). The associated minimal SM is similar to the previous one.

It is interesting to see that any AC over a commutative group is of the form of the previous example (up to isomorphism). To prove this we need some additional notation. For every $x \in X$, axiom int_2) ensures that the function

(1) $\varphi_x : X \to M, \varphi_x(y) = c(x,y)$, for all $y \in X$

is bijective and thus so is its inverse function $\varphi_x^{-1} : M \to X$ given by

(2) $\varphi_x^{-1}(m) = y \iff c(x,y) = m$.

Moreover, it holds

(3) $c(x, \varphi_x^{-1}(m)) = m$ for all $m \in M, x \in X$.

Indeed, we have

$$c(x, \varphi_x^{-1}(m)) = \varphi_x(\varphi_x^{-1}(m)) = (\varphi_x \circ \varphi_x^{-1})(m) = id(m) = m$$

with id denoting identity function. Now a binary operation $\triangle_x : X \times X \to X$ is defined by setting

(4) $y_1 \triangle_x y_2 = \varphi_x^{-1}(c(x,y_1) \triangle c(x,y_2))$, for all $y_1, y_2 \in X$.

Clearly the element x is the neutral element:

$$x \triangle_x y = \varphi_x^{-1}(c(x,x) \triangle c(x,y)) = \varphi_x^{-1}(e \triangle c(x,y)) = \varphi_x^{-1}(c(x,y)) \overset{(2)}{=} y$$

for all $y \in X$, e the neutral element of M. From $\varphi_x(x) = c(x,x) = e$, we get

(5) $\varphi_x^{-1}(e) = x$.

Proposition 21.2 *The triple* $(X, \triangle_x, x)$ *is a group and the function* $\varphi_x^{-1} : M \to X$ *is a group isomorphism.*

Proof We show that the symmetric y' of the element $y \in X$ is $\varphi_x^{-1}(c(y, x))$, i.e. that

$$y \triangle_x \varphi_x^{-1}(c(y, x)) = x \text{ and } \varphi_x^{-1}(c(y, x)) \triangle_x y = x.$$

Indeed, we have

$$\begin{aligned} y \triangle_x \varphi_x^{-1}(c(y, x)) =& \varphi_x^{-1}(c(x, y) \triangle c(x, \varphi_x^{-1}(c(y, x)))) \\ \overset{(3)}{=}& \varphi_x^{-1}(c(x, y) \triangle c(y, x)) = \varphi_x^{-1}(e) \overset{(5)}{=} x. \end{aligned}$$

The associativity law is obtained after long calculations.

The bijective function $\varphi_x : X \to M$ preserves binary operations:

$$\varphi_x(y_1 \triangle_x y_2) = \varphi_x(y_1) \triangle \varphi_x(y_2), \text{ for all } y_1, y_2 \in X.$$

In fact we have

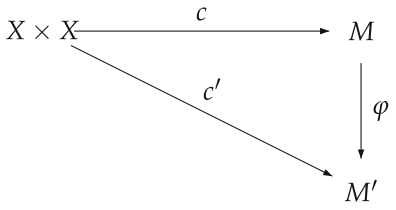
$$\begin{aligned} \varphi_x(y_1 \triangle_x y_2) &= \varphi_x(\varphi_x^{-1}(c(x, y_1) \triangle c(x, y_2))) \\ &= (\varphi_x \circ \varphi_x^{-1})(c(x, y_1) \triangle c(x, y_2)) \\ &= id(c(x, y_1) \triangle c(x, y_2)) = c(x, y_1) \triangle c(x, y_2) = \varphi_x(y_1) \triangle \varphi_x(y_2). \end{aligned}$$

Consequently φ_x is a group isomorphism and thus is its inverse φ_x^{-1}.

Now, a morphism from $AC\ (X, M, c)$ to the $AC\ (X, M', c')$ is a group morphism $\varphi : M \to M'$ commuting with the contour functions, i.e.

$$\varphi_x(c(s, t)) = c'(s, t), \text{ for all } s, t \in X.$$

This means that the next triangle



is commutative, Fiore, Noll and Satyendra [5]. Now we are in a position to establish the main result.

Theorem 21.1 *Any AC* $(X, (M, \triangle, e), c)$*, with* $(M, \triangle, e)$ *commutative group, is the image via the isomorphism* φ_x $(x \in X)$ *of the standard AC* $(X, (X, \triangle_x, x), c_x)$*, where*

$$c_x : X \times X \to X, \;\; c_x(y_1, y_2) = y_2 \vartriangle_x y_1',$$

y_1' being the symmetric element of y_1.

Proof We have to show the equality

$$c(y_1, y_2) = \varphi_x(y_2 \vartriangle_x y_1'), \;\; \text{for all} \;\; y_1, y_2 \in X$$

As we have seen in the proof of the previous proposition, the symmetric of y_1 is the element

(6) $y_1' = \varphi_x^{-1}(c(y_1, x))$.

Then

$$\begin{aligned} y_2 \vartriangle_x y_1' &= \varphi_x^{-1}(c(x, y_2) \vartriangle c(x, y_1')) \;\; \text{by definition of} \;\; \vartriangle_x \\ &= \varphi_x^{-1}(c(x, y_2) \vartriangle c(x, \varphi_x^{-1}(c(y_1, x))) \;\; \text{by (6)} \\ &= \varphi_x^{-1}(c(x, y_2) \vartriangle c(y_1, x)) \;\; \text{by (3)} \\ &= \varphi_x^{-1}(c(y_1, x) \vartriangle c(x, y_2)) \;\; \text{by commutativity} \\ &= \varphi_x^{-1}(c(y_1, y_2)) \;\; \text{by axiom } int_1) \end{aligned}$$

that is

$$y_2 \vartriangle_x y_1' = \varphi_x^{-1}(c(y_1, y_2)).$$

Applying the isomorphism φ_x in both members of the above equality we get

$$\varphi_x(y_2 \vartriangle_x y_1') = c(y_1, y_2)$$

and the proof is complete.

Let us apply the previous theorem to determine all the ACs on $\mathbb{Z}_{12}$. Such an AC is obtained by composing the standard system:

$$diff : \mathbb{Z}_{12} \times \mathbb{Z}_{12} \to \mathbb{Z}_{12} \;\; \text{by an automorphism}$$

of $\mathbb{Z}_{12}$. But all the automorphisms of $\mathbb{Z}_{12}$ are the dilatations

$$h_a : \mathbb{Z}_{12} \to \mathbb{Z}_{12}, \;\; h_a(x) = ax, \;\; \text{for all} \;\; x \in \mathbb{Z}_{12}$$

where $a = 1, 5, 7, 11$. We conclude that four ACs can be defined on $\mathbb{Z}_{12}$, namely,

$$c_a : \mathbb{Z}_{12} \times \mathbb{Z}_{12} \to \mathbb{Z}_{12}, \;\; c_a(s, t) = a(t - s), \;\; s, t \in \mathbb{Z}_{12}, \;\; a = 1, 5, 7, 11.$$

In the sequel we investigate an interesting non-commutative AC taken from Fiore [4]. From now on, upper (lower) case letters will represent major (minor) triad chords.

Consider the group $S(X_{ch})$ of permutations of the set of triad chords $X_{ch} = \{G, g, E\flat, e\flat, B, b\}$ and its subgroup $M(X_{ch})$ generated by the permutations

$$P = \begin{pmatrix} G & E\flat & B & g & e\flat & b \\ g & e\flat & b & G & E\flat & B \end{pmatrix}$$

$$L = \begin{pmatrix} G & E\flat & B & g & e\flat & b \\ b & g & e\flat & E\flat & B & G \end{pmatrix}$$

where $P =$ parallel and $L =$ leading tone exchange. We have

$$PL = \begin{pmatrix} G & E\flat & B & g & e\flat & b \\ B & G & E\flat & e\flat & b & g \end{pmatrix}$$

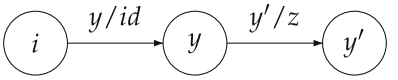
$$LP = \begin{pmatrix} G & E\flat & B & g & e\flat & b \\ E\flat & B & G & b & g & e\flat \end{pmatrix}$$

$$PLP = \begin{pmatrix} G & E\flat & B & g & e\flat & b \\ e\flat & b & g & B & G & E\flat \end{pmatrix}$$

and $M(X_{ch}) = \{id, P, L, PL, LP, PLP\}$ hence the table

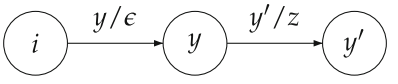
$c(X_{ch})$	G	$E\flat$	B	g	$e\flat$	b
G	id	LP	PL	P	PLP	L
$E\flat$	PL	id	LP	L	P	PLP
B	LP	PL	id	PLP	L	P
g	P	L	PLP	id	PL	LP
$e\flat$	PLP	P	L	LP	id	PL
b	L	PLP	P	PL	LP	id

It is not hard to verify that the triple $(X_{ch}, M(X_{ch}), c(X_{ch}))$ is indeed a AC. It can be viewed as a sequential machine in two different ways, either



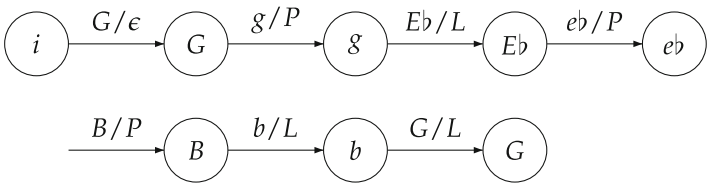
$y, y' \in X_{ch}$, $z \in M(X_{ch})$

with output the group $\mathcal{M}(X_{ch})$ or



$y, y' \in X_{ch}$, $z \in M(X_{ch})$

with output the group $\mathcal{M}(X_{ch})^*$ (id is the identity permutation and ε the empty word). Feeding the latter SM with the string of chords $u = GgE\flat e\flat BbG$, the created path is



and its output $PLPLPL$ classifies u as an hexatonic sequence.

Also, the latter machine simulates the former one, via the pair (id, φ) consisting of the identity state function and the canonical monoid epimorphism $\varphi : \mathcal{M}(X_{ch})^* \to \mathcal{M}(X_{ch})$ interpreting the symbols P and L as the transformations parallel and leading tone exchange.

References

1. Bor, M.: Contour Reduction Algorithms: A Theory of Pitch and Duration Hierarchies for Post-Tonal Music. Ph.D. Thesis, The University of British Columbia Canada (2006)
2. Bozapalidou, M.: Automata and music contour functions. J. Math. Music **7**(3), 195–211 (2013)
3. Eilenberg, S.: Automata, Languages and Machines, vol. A, Academic Press (1974)
4. Fiore, T.M.: Transformational theory: overview presentation. http://www.personal.umd.umich.edu/tmfiore. Accessed 9 Jan 2011
5. Fiore, T., Noll, T., Satyendra, R.: Morphisms of generalized interval systems and PR-groups. J. Math. Music **7**(1), 3–27 (2013)
6. Friedmann, M.: A methology for for the discussion of contur: Its application to Schoenberg's music. J. Music Theory **29**(2), 223–48 (1985)
7. Forte, A.: The Structure of Atonal Music. Yale University Press, New Haven, CT (1973)
8. Lewin, D.: Generalized Musical Intervals and Transformations. New Heaven, CT and London (1987): Yale University Press. Reprinted, Oxford and New York: Oxford University Press (2007)
9. Lewin, D.: Musical Form and Transformation: Four Analytic Essays. New Heaven, CT and London (1993): Yale University Press. Reprinted, Oxford and New York: Oxford University Press (2007)
10. Morris, R.: New directions in the theory and analysis of musical contours. Music Theory Spectr. **15**(2), 205–28 (1993)

Chapter 22
A Formal Representation of Video Content with the Picture Hyperoperation

Antonios Kalampakas, Nikolaos Triantafyllou, Katerina Ksystra and Petros Stefaneas

Abstract A hyperoperation on the pixels of a two-dimensional picture is introduced and studied. In this setup pictures are defined as a specific type of rectangular graphs and the picture hyperoperation is given by virtue of the notion of the path inside such a picture. Using this hyperoperation objects can naturally be defined inside pictures in an algebraic way and this concept can be utilized in order to formally represent and compare video content using algebraic semiotics.

22.1 Introduction

Hyperstructures have been thoroughly examined in the previous decades, for an overview see [6]. In particular, hypergroupoids deriving from binary relations, namely *C-hypergroupoids*, where introduced by Corsini in [3], see also [2, 13, 16] and [17, 19]. Since then, the correlation between hyperstructures and binary relations has been examined by several researchers in [4–9], see also [11, 14, 15, 17–20]. In a broader view, these studies where the first step towards a more general investigation of the correlation between Graph Theory and algebraic hyperstructures. Recently several more definite and interesting results were obtained towards this direction. In

A. Kalampakas
Department of Mathematics, College of Engineering and Technology,
American University of the Middle East, Egaila, Kuwait
e-mail: akalampakas@gmail.com

N. Triantafyllou · K. Ksystra · P. Stefaneas (✉)
Department of Applied Mathematics, National Technical University of Athens,
Zografou Campus, 15780 Athens, Greece
e-mail: petros@math.ntua.gr

N. Triantafyllou
e-mail: nitriant@central.ntua.gr

K. Ksystra
e-mail: katksy@central.ntua.gr

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_22

particular different types of graph hyperoperations, i.e., hyperoperations defined on a specific corresponding graph, where introduced and investigated.

The first graph hyperoperation, namely the *path hyperoperation* was introduced in [12] as a generalization of Corsini's hyperoperation which was associated with a binary relation. More specifically, given a graph $G = (V, R)$, where V is the set of nodes and R the set of edges, we introduce the *path hyperoperation* $\star_G : V \times V \to V$ assigning to every pair $(x, y) \in V \times V$ the set consisting of every element of V that belongs to at least one directed path from x to y. It is clear that for any binary relation $R \subseteq V \times V$ and elements $x, y \in V$, their Corsini product is a subset of $x \star_G y$. In the same paper the connection of the associated path hyperstructure with mixed model assembly line designing was also investigated. The commutativity property for the class of all path hypergroupoids is examined and its connection with Graph Theory is delineated by specifying a sufficient and necessary condition for the corresponding graph in order to obtain a commutative path hyperoperation. As a result a path hypergoupoid is commutative if and only if it can be obtained as a disjoint union of non-partial hypergroupoids.

In this paper we extend this approach in order to introduce and investigate the properties of a hyperoperation defined on the elementary elements of rectangular grids. Many approaches exist in the literature for the representation of two-dimensional pictures in an algebraic formalism. In the present setup a picture is a subset of a rectangular array (frame) which consists of elementary pieces (pixels). This definition is very close to the one introduced by Bozapalidis in [1], see also [10] for a similar representation. We then employ this simple set-theoretic definition of a picture in order to introduce a hyperoperation on pictures similar to the path hyperoperation for graphs [12]. Our aim is to utilize the resulting hyperstructure in order to formally represent and compare video content using algebraic semiotics.

22.2 The Picture Hyperoperation

In this section the notion of a a picture as a subset of a rectangular array consisting of elementary pieces is presented and a natural set-theoretic approach is adopted in order to introduce a hyperoperation on pictures similarly to the path hyperoperation for graphs.

Formally a frame of dimension (m, n), where $m, n \in \mathbb{N}$, is a set

$$F = \{(i, j) \mid i, j \in \mathbb{N},\ 1 \leq i \leq m \text{ and } 1 \leq j \leq n\}$$

Its elements are called pixels. A picture P in the frame F is now defined as a subset $P \subseteq F$. Hence a picture is just a set of pixels. Notice that for the sake of simplicity this definition is appropriate only for black and white pictures but it can be easily generalized for color pictures by taking multiple subsets, i.e., pictures $P \subseteq F$ - one for each color.

Definition 22.1 Given a frame F and pixels $p = (m, n) \in F$ and $p' = (m', n') \in F$, we say that p and p' are *neighbors* in F if it holds

$$m = m' \quad \text{and} \quad |n - n'| = 1$$

or

$$|m - m'| = 1 \quad \text{and} \quad n = n'.$$

Definition 22.2 A *path* inside a picture P is a sequence

$$(p_1, \ldots, p_k)$$

such that $p_1, \ldots, p_k \in P$ and for all $1 \leq i \leq k - 1$, p_i and p_{i+1} are neighbors.

We will employ the above set-theoretic definition of pictures to introduce a picture hyperoperation which operates in a way similar to the path and the simple hyperoperations. More precisely, given an (m, n) frame F and a picture $P \subseteq F$, the *picture hyperoperation* (or picture product)

$$\star_P : F \times F \to \mathscr{P}(F)$$

is defined for every $p, p' \in F$ by:

$$p \star_P p' = \{q \in F \mid q \text{ belongs to a path inside P from } p \text{ to } p'\}.$$

Notice that, from the definition of the path, we deduce that necessary but not sufficient condition for $p \star_P p' \neq \emptyset$ is that $p, p' \in P$.

The so obtained (partial) hyperstructure $(F, \star_P)$ is called the (partial) *picture hypergroupoid* associated with P. In the case that it holds

$$p \star_P p' \neq \emptyset, \quad \text{for all} p, p' \in F,$$

then $\star_P$ is a non-partial hyperoperation and the related picture hypergroupoid is called non-partial. Non-partial picture hypergroupoids are characterized as follows.

Proposition 22.1 *Given a frame F and a picture $P \subseteq F$ the following conditions are equivalent*

(i) $\star_P$ *is non-partial,*
(ii) $\star_P$ *is total,*
(iii) it holds

$$P = F$$

An object inside a picture $P \subseteq F$ is now naturally defined as a subset $O \subseteq P$ such that for every pixel $p \in P$ it holds

$$p \star_P p' \neq \emptyset \text{ if and only if } p' \in O.$$

It is straightforward to check that for every $p, p' \in O$ it holds

$$p \star_P p' = O$$

Moreover,

Proposition 22.2 *If for two objects $O, O' \in P$ it holds $O \cap O' \neq \emptyset$ then $O = O'$.*

It turns out that every picture $P \subseteq F$ can be written as the union of all its objects $O_1, \ldots, O_n$, i.e.,

$$P = O_1 \cup \cdots \cup O_n$$

with $O_i \cap O_j = \emptyset$, for $i \neq j$, $1 \leq i, j \leq n$.

Definition 22.3 Given an (m, n) frame F, a picture $P \subseteq F$ and an object $O \subseteq P$, we say that a pixel $p \in O$ belongs to the *interior* of the object O if for all neighbor pixels $p' \in F$ it holds

$$p \star_P p' \neq \emptyset$$

or equivalently $p' \in O$.

Definition 22.4 Given an (m, n) frame F, a picture $P \subseteq F$ and an object $O \subseteq P$, we say that a pixel $p \in O$ belongs to the *boundary* of the object O, if it does not belong to the interior, i.e., if there exists a neighbor pixel $p' \in F$ such that it holds

$$p \star_P p' = \emptyset$$

or equivalently $p' \notin O$.

As opposed to path and simple path hyperoperations the picture hyperoperation is commutative and associative.

Proposition 22.3 *Every picture hypergroupoid $(F, \star_P)$ is commutative and associative.*

Proof Given $u, v \in P$, we distinguish two cases: if u and v belong to the same object $O \subseteq P$ then $u \star_P v = v \star_P u = O$, else $u \star_P v = v \star_P u = \emptyset$.

Similarly, for associativity, given $u, v, w \in P$, if there exists $O \subseteq P$ with $u, v, w \in O$ then $u \star_P (v \star_P w) = (u \star_P v) \star_P w = O$, else $u \star_P (v \star_P w) = (u \star_P v) \star_P w = \emptyset$.

Acknowledgements This research has been co-financed by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALIS

References

1. Bozapalidis, S.: Picture deformation. Acta Inform. **45**, 1–31 (2008)
2. Corsini, P.: Hypergraphs and hypergroups. Algebra Universalis **35**(4), 548–555 (1996)
3. Corsini, P.: Binary relations and hypergroupoids. Ital. J. Pure Appl. Math. **7**, 11–18 (2000)
4. Corsini, P., Leoreanu, V.: Hypergroups and binary relations. Algebra Universalis **43**, 321–330 (2000)
5. Corsini, P., Leoreanu, V.: Applications of Hyperstructure Theory. Kluwer Academic Publishers, Boston (2002)
6. Corsini, P., Leoreanu, V.: Survey on new topics of hyperstructure theory and its applications. Proc. of 8th Internat. Congress on AHA. 1-37 (2003)
7. Cristea, I., Stefanescu, M.: Binary relations and reduced hypergroups. Discret. Math. **308**, 3537–3544 (2008)
8. Cristea, I., Stefanescu, M.: Hypergroups and n-ary relations. Eur. J. Comb. **31**, 780–789 (2010)
9. Cristea, I., Stefanescu, M., Anghluta, C.: About the fundamental relations defined on the hypergroupoids associated with binary relations. Eur. J.Comb. **32**, 72–81 (2011)
10. Dassow, J., Hinz, F.: Decision problems and regular chain code picture languages. Discret. Appl. Math. **45**, 29–49 (1993)
11. Kalampakas, A., Spartalis, S., Skoulariki, K.: Directed graphs representing isomorphism classes of C-Hypergroupoids. Ratio Math. **23**, 51–64 (2012)
12. Kalampakas, A., Spartalis, S., Tsigkas, A.: The path hyperoperation. An. Stiintifice ale Universitatii Ovidius Constanta, Ser. Math. **22**, 141–153 (2014)
13. Konstantinidou, M., Serafimidis, K.: Sur les P-supertrillis, In: New Frontiers in Hyperstructures and Rel. Algebras, 139-148 Hadronic Press, Palm Harbor U.S.A (1996)
14. Massouros, C.H., Massouros, G.: Hypergroups Associated with graphs and automata. Proceedings of the International Conference on Numerical Analysis and Applied Mathematics, American Institute of Physics (AIP) Conference Proceedings 1168, 164-167 (2009)
15. Massouros C.H., Tsitouras, C.H.: Enumeration of hypercompositional structures defined by binary relations. Italian Journal of Pure and Applied Mathematics, vol. 28, 4354 (2011)
16. Rosenberg, I.: Hypergroups and join spaces determined by relations. Ital. J. Pure Appl. Math. **4**, 93–101 (1998)
17. Spartalis, S.: Hypergroupoids obtained from groupoids with binary relations. Ital. J. Pure Appl. Math. **16**, 201–210 (2004)
18. Spartalis, S.: The hyperoperation relation and the Corsini's partial or not-partial hypergroupoids (A classification). Ital. J. Pure Appl. Math. **24**, 97–112 (2008)
19. Spartalis, S., Mamaloukas, C.: On hyperstructures associated with binary relations. Comput. Math. Appl. **51**, 41–50 (2006)
20. Spartalis, S., Konstantinidou, M., Taouktsoglou, A.: C-hypergroupoids obtained by special binary relations. Comput. Math. Appl. **59**, 2628–2635 (2010)

Chapter 23
Novel Approaches to Medical Information Processing and Analysis

Evi Karali

Abstract The purpose of this article is to present scientific research results of Karali et al (Proceedings of 8th International Conference on Bioinformatics and Bioengineering (BIBE), 2008, [14]), Karali et al (Inf Technol Biomed, 15(13):381–6, 2011, [15]), Karali et al (J Biosci Med (JBM), 1:6–9, 2013, [16]), Karali et al (Int J Comput Vis, 321–331, 1988 [24]) concerning medical imaging, especially in the fields of image reconstruction in Emission Tomography and image segmentation. Image reconstruction in Positron Emission Tomography (PET) uses the collected projection data of the object/patient under examination. Iterative image reconstruction algorithms have been proposed as an alternative to conventional analytical methods. Despite their computational complexity, they become more and more popular, mostly because they can produce images with better contrast-to-noise (CNR) and signal-to-noise (SNR) ratios at a given spatial resolution, compared to analytical techniques. In Sect. 23.1 of this study we present a new iterative algorithm for medical image reconstruction, under the name Image Space Weighted Least Squares (ISWLS) (Karali et al, Proceedings of 8th International Conference on Bioinformatics and Bioengineering (BIBE), 2008, [14]). In (Karali et al, Proceedings of 8th International Conference on Bioinformatics and Bioengineering (BIBE), 2008, [14]) we used phantom data from a prototype small-animal PET system and the methods presented are applied to 2D sinograms. Further, we assessed the performance of the new algorithm by comparing it to the simultaneous versions of known algorithms (EM-ML, ISRA and WLS). All algorithms were compared in terms of cross-correlation coefficient, reconstruction time and CNRs. ISWLS have ISRA's properties in noise manipulation and WLS's acceleration of reconstruction process. As it turned out, ISWLS presents higher CNRs than EM-ML and ISRA for objects of different sizes. Indeed ISWLS shows similar performance to WLS during the first iterations but it has better noise manipulation. Section 23.5 of this study deals with another important field of medical imaging, the image segmentation and in particular the subject of deformable models. Deformable models are widely used segmentation methods with scientifi-

E. Karali (✉)
Department of Automation Engineering, Piraeus University of Applied Science, Athens, Greece
e-mail: ekarali76@hotmail.com

S. Lambropoulou et al. (eds.), *Algebraic Modeling of Topological and Computational Structures and Applications*, Springer Proceedings in Mathematics & Statistics 219, https://doi.org/10.1007/978-3-319-68103-0_23

cally accepted results. In Karali et al (Int J Comput Vis, 321–331, 1988 [24]) various methods of deformable models are compared, namely the classical snake (Kass et al, Int J Comput Vis, 321–331, 1988, [25]), the gradient vector field snake (GVF snake) (Xu, IEEE Proceedings on Computer Society Conference on Computer Vision and Pattern Recognition, 1997, [36]) and the topology-adaptive snake (t-snake) (Mcinerney, Topologically Adaptable Deformable Models For medical Image Analysis, 1997, [29]), as well as the method of self-affine mapping system (Ida and Sambonsugi, IEEE Trans Imag Process, 9(11), 2000 [22]) as an alternative to snake models. In Karali et al (Int J Comput Vis, 321–331, 1988 [24]) we modified the self-affine mapping system algorithm as far as the optimization criterion is concerned. The new version of self-affine mapping system is more suitable for weak edges detection. All methods were applied to glaucomatic retinal images with the purpose of segmenting the optical disk. The methods were compared in terms of segmentation accuracy and speed. Segmentation accuracy is derived from normalized mean square error between real and algorithm extracted contours. Speed is measured by algorithm segmentation time. The classical snake, T-snake and the self-affine mapping system converge quickly on the optic disk boundary comparing to GVF-snake. Moreover the self-affine mapping system presents the smallest normalized mean square error (nmse). As a result, the method of self-affine mapping system presents adequate segmentation time and segmentation accuracy, and significant independence from initialization.

23.1 Part I

23.1.1 Introduction

Medical Imaging is a vital component of a large number of clinical applications. Such applications occur at all stages of medical treatment, from diagnosis to the areas of design, implementation and assessment of the effectiveness of treatment. The development of tomography, the breakthrough of modern computer systems, the evolution of specialized computer signal processing packages and the advancements on medical data analysis, has brought a real revolution in radiology and diagnostic radiology. The result of the medical technology development is the noninvasive obtain of precise functional and/or anatomical information of the interior of the human body.

The modern techniques of computed tomography (x-ray, CT), positron emission tomography (PET), single photon emission tomography (SPECT) use detector arrays mounted or rotatable around the test object in order to collect multiple different angular views (projections) of the object. The collected projection data are used by mathematical algorithms to reconstruct images of the areas of interest in the subject matter. These images are either anatomical or images of biochemical activity of structures of interest.

Positron Emission Tomography (PET) is a medical imaging technique, which utilizes the unique features of $\beta +$ nuclear decay of specific radionuclides for imaging metabolic activity of the anatomy of the tested structures. Radionuclides are produced in a cyclotron and are used for labeling molecules with particular biological interest. The labeled molecules are introduced intravenously in the body under examination and distributed by the bloodstream to tissues in a manner that is determined by their biochemical properties. Specifically when the radioactive atom of a labeled molecule decays a positron e + is emitted, which is annihilated, near the point of generation, by an individual electron e- resulting in the emission of two gamma-ray energy 511 keV each. The two emitted photons travel diametrically and can escape from the human body [17].

A PET system consists of a set of detectors which surround the patient and the aim is to detect and convert the high energy photons emitted to an electrical signal. The electrical signal is then fed to signal processing electronic devices. In a typical PET examination the annihilation events detected, are corrected for various factors and reconstructed into tomographic images using special mathematical algorithms. The result of the reconstruction is a whole tomographic image. The luminance value of each pixel of the image is proportional to the amount of the radioactive molecule in the region showing the pixel. Therefore PET images allow the in vivo quantitative recording of the spatial distribution of the radioactive tracer in the body [5, 18, 19].

Methods of medical data reconstruction that have been developed so far are divided into two major categories [5, 17, 18]:

- Analytical methods, which use the mathematics of the computed tomography, which connect projection data with the spatial distribution of radioactivity within the subject matter.
- Iterative methods, which model the data collection process by the PET scan and attempt, relying on predefined criteria and a number of successful iterations, to approach the real image of the spatial distribution of the radioactive tracer.

The analytical reconstruction methods are based on linear calculation of the inverse Radon transform and offer direct mathematical solution for the image formation. The core of the analytical reconstruction methods is the Filtered Backprojection algorithm (FBP). Variations or extensions of FBP methods are the Fast Volume Reconstruction Algorithm (FAVOR [7]) and the 3D Reprojection algorithm (3DRP) [17].

The analytical methods are standard reconstruction techniques, which are applied in clinical systems. The reason for their prevalence of the preceding decades lies in their low computational cost. The big disadvantage of analytical techniques is their inability to include correction models of all factors involved in the making of PET data (such as the scope of the positron, the production of two gamma rays at angles between them of less than 180°, scattering phenomena, the attenuation, random coincidences, the different performance and sensitivity of the detectors, etc.) during the reconstruction process. Further analytical methods do not maintain the non-negativity of the image values and can export images reconstructed with star-like artifacts [17].

The iterative methods are based on stochastic models for the entire data-making process taking into account all of the physical and technical factors involved during a PET examination. They fall into two major categories, the algebraic techniques (algebraic techniques-ART [12]) and the statistical methods which in turn are divided into maximum likelihood techniques (maximum-likelihood algorithms-ML) and least squares methods (least squares-LS, weighted- least squares-WLS [2], iterated space reconstruction algorithm (ISRA) [3]). The algebraic techniques are the first iterative techniques used. Then the algorithm of maximum likelihood (ML) 1982 by Shepp and Vardi [18] has been applied. Since then, multitude of variants have appeared (SAGE [8], RAMLA [4], OSEM [11], MAP-EM [10]), in order to improve efficiency and reconstruction time and the further improvement of the quality of medical image.

The progress in research and clinical application of iterative techniques is closely associated with the development and optimization of electronic circuits and computer systems. They require large computational reconstruction times, high computing and memory storage. However they have great research interest due to the high image quality they produce, compared to analytical methods. These are techniques that allow modeling of the whole data acquisition process in the reconstruction, and most, especially statistical methods, guarantee positive solutions.

The purpose of this part of the study is to present a new iterative image reconstruction algorithm, under the short name ISWLS [14] (Image Space Weighted Least Squares), produced by the maximization of an objective function. This algorithm was introduced in [14] and studied in [14]. To maximize the objective function, the Kuhn–Tucker condition must be satisfied. ISWLS is expected to have ISRA properties in noise manipulation and WLS acceleration of the reconstruction process. To assess the performance of the new iterative reconstruction method, we have used phantom data produced from simulating a prototype small-animal PET system. We compare ISWLS reconstruction data with those from EM-ML, ISRA, and WLS. We also present the OS (ordered subsets) version of ISWLS (OS-ISWLS) and compare it with the OSEM, OS-ISRA, and OS-WLS [15]. Moreover the MRP (median-root-prior) [1] version of ISWLS is presented and compared with MRP-EMML, MRP-ISRA and MRP-WLS [16]. The methods presented here are applied to 2D sinograms. We have implemented simultaneous versions of the aforementioned algorithms. The simultaneous version of an algorithm is an algorithm where all image pixels are simultaneously updated in every iteration. These methods are of great interest, because of their ability to be implemented in parallel computing architectures, which decreases drastically the total reconstruction time.

23.2 Theory

In general, every iterative method relies on the hypothesis that the projection data **y** are linearly connected to the image **x** of radiopharmaceutical spatial distribution, according to the equation:

$$\mathbf{y} = A^T \mathbf{x} \tag{23.1}$$

where A is a matrix that characterizes the PET system being used for data acquisition. In bibliography this matrix is called *system* or *probability matrix* and it projects image data to sinogram domain (the term sinogram refers to the projection data matrix). Every element α_{ij} of the system matrix A represents the probability an annihilation event emitted in image pixel i to be detected in LOR_j. The significance of the probability matrix lies on the valuable information, related to the data acquisition process, that it can contain (e.g. number of detector rings, number of detector elements in every ring, ring diameter, diameter of transaxial field of view, detector size, image size, spatial and angular sampling).

The most commonly used least squares algorithms, that are based on simultaneous iterative schemes, are ISRA (Image Space Reconstruction Algorithm) and WLS (Weighted Least Squares) with updating step in the kth iteration:

$$\text{ISRA:} \quad x_i^k = x_i^{k-1} \frac{\sum_{j=1}^{M} a_{ij} y_j}{\sum_{j=1}^{M} a_{ij} \sum_{i'=1}^{N} a_{i'j} x_{i'}^{k-1}} \tag{23.2}$$

$$\text{WLS:} \quad x_i^k = x_i^{k-1} \sum_{j=1}^{M} \frac{a_{ij} y_j^2}{\left(\sum_{i'=1}^{N} a_{i'j} x_{i'}^{k-1} \right)^2} \tag{23.3}$$

Expectation Maximization Maximum Likelihood (EM-ML) algorithm has an updating step in the kth iteration:

$$\text{EM-ML:} \quad x_i^k = x_i^{k-1} \sum_{j=1}^{M} \frac{a_{ij} y_j}{\left(\sum_{i'=1}^{N} a_{i'j} x_{i'}^{k-1} \right)} \tag{23.4}$$

23.2.1 ISWLS Algorithm

In the current work, we propose a new algorithm under the short name ISWLS. Consider an image discretized into N pixels and the measured data $\mathbf{y}$ collected by M detector tubes. We can propose the following ISWLS estimator of $\mathbf{x}$ in Eq. 23.1:

$$\hat{\mathrm{x}} = \arg\max_{\mathrm{x}} \phi(\mathrm{x}) \text{ subject to } x_i \geq 0, \quad i = 1, 2, \ldots, N \tag{23.5}$$

where

$$\phi(\mathrm{x}) = \sum_{j=1}^{M} \left[-\frac{\left(y_j - \sum_{i=1}^{N} a_{ij} x_i\right)^3}{3} + \left(y_j - \frac{2}{3}\sum_{i=1}^{N} a_{ij} x_i\right)\left(\sum_{i=1}^{N} a_{ij} x_i\right)^2 \right] \tag{23.6}$$

Under the conditions in problem (23.5), $\hat{\mathbf{x}}$ is a solution if and only if the Kuhn-Tucker condition is satisfied, namely:

$$x_i \frac{\partial \phi(x)}{\partial x_i}\bigg|_{\hat{\mathbf{x}}} = 0 \tag{23.7}$$

where:

$$\begin{aligned} \frac{\partial \phi(x)}{\partial x_i} &= \sum_{j=1}^{M} \left[\left(y_j - \sum_{i'=1}^{N} a_{i'j} x_i\right)^2 a_{ij} + 2a_{ij} y_j \left(\sum_{i'=1}^{N} a_{i'j} x_i\right) - 2a_{ij} \left(\sum_{i'=1}^{N} a_{i'j} x_i\right)^2 \right] \\ &= \sum_{j=1}^{M} \left(a_{ij} y_j^2 - a_{ij} \left(\sum_{i'=1}^{N} a_{i'j} x_i\right)^2 \right) \end{aligned} \tag{23.8}$$

According to (23.7) Eq. 23.8 is written as:

$$x_i \sum_{j=1}^{M} \left(a_{ij} y_j^2 - a_{ij} \left(\sum_{i'=1}^{N} a_{i'j} x_i\right)^2 \right) = 0 \tag{23.9}$$

So, we obtain the fixed point iterative formula for the ith pixel's update as follows:

$$\text{ISWLS:} \quad x_i^k = x_i^{k-1} \frac{\sum_{j=1}^{M} a_{ij} y_j^2}{\sum_{j=1}^{M} a_{ij} \sum_{i'=1}^{N} \left(a_{i'j} x_{i'}^{k-1}\right)^2} \tag{23.10}$$

Moreover, we can derive the ordered subsets version of ISWLS, (OS-ISWLS) with updating scheme in kth iteration for subset S_n:

$$\text{OS-ISWLS:} \quad x_i^k = x_i^{k-1} \frac{\sum_{j \in S_n} a_{ij} y_j^2}{\sum_{j \in S_n} a_{ij} \left(\sum_{i'=1}^{N} a_{i'j} x_{i'}^{k-1}\right)^2} \tag{23.11}$$

The OS version of an iterative algorithm accelerates the reconstruction process without image deterioration. An OS algorithm divides sinogram data into subsets and applies the simple version of the reconstruction algorithm to every subset. The approximation of the image x_s that the simple version provides after processing the s subset consists the initial solution for the next subset s + 1. Every OS iteration is completed after the application of the simple algorithm to all subsets of sinogram data.

In addition we can present the MRP (median-root-prior) [1] version of ISWLS. MRP versions belong to regularization's reconstruction algorithms. These methods take into account a priori information for the radioactivity spatial distribution inside the object under examination. For the reduction of the noise many regularization methods have been proposed, which reduce drastically the noise with a small image resolution reduction. The success of a regularization method depends on the mathematical formula of the prior. Median root prior (MRP) belongs to the most popular priors. It is derived from a Gaussian distribution with mean value the median value of reconstructed image pixels in the vicinity of pixel i. The use of MRP results in noise component reduction while at the same time it preserves the edges.

Suppose that:

$$f(x_i) \approx e^{-b\frac{(x_i-M)^2}{2M}} \tag{23.12}$$

where $M = \text{med}(x_i, i)$, the median value of reconstructed image pixels in the vicinity of pixel i. Then:

$$u(x_i) = \frac{\partial \ln(f(x_i))}{\partial x_i} = -b\frac{x_i - med(x_i, i)}{med(x_i, i)} \tag{23.13}$$

The term $b \in [0,2]$ determines the degree of smoothing in reconstruction images. If $b = 0$ no prior is applied. Big values of b cause over-smoothing, while small values of b result in images with high resolution but with increased noise.

According to the one-step-late philosophy [9], where the prior is applied to the previous radiopharmaceutical distribution estimation, we can extract an empirical iterative formula for the ISWLS algorithm in combination with MRP prior. The new iterative algorithm has updating scheme:

$$x_i^k = \frac{x_i^{k-1}}{1 + b\frac{x_i^{k-1} - med(x_i^{k-1}, i)}{med(x_i^{k-1}, i)}} \frac{\sum_{j=1}^{M} a_{ij} y_j^2}{\sum_{j=1}^{M} a_{ij} \left(\sum_{i'=1}^{N} a_{i'j} x_{i'}^{k-1} \right)^2} \tag{23.14}$$

The MRP-EM-ML updating scheme in kth iteration is:

$$x_i^k = \frac{x_i^{k-1}}{1 + b\frac{x_i^{k-1} - med(x_i^{k-1}, i)}{med(x_i^{k-1}, i)}} \sum_{j=1}^{M} \frac{a_{ij} y_j}{\left(\sum_{i'=1}^{N} a_{i'j} x_{i'}^{k-1} \right)} \tag{23.15}$$

The MRP-ISRA updating scheme in kth iteration is:

$$x_i^k = \frac{x_i^{k-1}}{1 + b\frac{x_i^{k-1} - med(x_i^{k-1}, i)}{med(x_i^{k-1}, i)}} \frac{\sum_{j=1}^{M} a_{ij} y_j}{\sum_{j=1}^{M} a_{ij} \sum_{i'=1}^{N} a_{i'j} x_{i'}^{k-1}} \tag{23.16}$$

The MRP-WLS updating scheme in kth iteration is

$$x_i^k = \frac{x_i^{k-1}}{1 + b\frac{x_i^{k-1} - med(x_i^{k-1}, i)}{med(x_i^{k-1}, i)}} \sum_{j=1}^{M} \frac{a_{ij} y_j^2}{\left(\sum_{i'=1}^{N} a_{i'j} x_{i'}^{k-1}\right)^2} \tag{23.17}$$

23.3 Results

For the evaluation of the iterative reconstruction methods presented in THEORY, projection data of a Derenzo-type phantom have been used. The Derenzo-type phantom consists of sets of rods filled with F^{18}, with diameters 4.8, 4, 3.2, 2.4, 1.6 and 1.2 mm, and the same separation between surfaces in the corresponding sets. The rods were surrounded by plastic (polyethylene). Data were produced using Monte Carlo simulation of a small-animal PET scanner.

Further, 18×10^6 coincidence events were collected. Projection data were binned to a 2D sinogram, 55×170 pixels in size, which means that data from 55 TORs (Tube of Response) per rotation angle were collected and 170 totally angular samples were used. Since the two detector heads rotate from 0° to 180° the angular step size was 1.0647°.

The system matrix was derived from an analytical method and calculated once before reconstruction. Each element a_{ij} was computed as the area of intersection E_{ij}, of TOR_j (Tube-of-Response) with image pixel i. The calculated system matrix is a sparse matrix. It consists of zero-valued elements in majority that have no contribution during iterative reconstruction process. So, only the non-zero elements were stored, resulting in significant reduction in system matrix size and consequently in required storage. The reconstructed 2D images were 128×128 pixels in size, thus the system matrix consisted of $55 \times 170 \times 128 \times 128$ elements with 4.33% sparsity.

The initial image estimate for all algorithms was:

$$x_{o_i} = \frac{\sum_{j=1}^{M} y_j}{N},\ \text{i} = 1, 2, \ldots, N \tag{23.18}$$

where y_j is the value of the jth sinogram element and N represents the total number of image pixels ($N = 128 \times 128$ in this implementation).

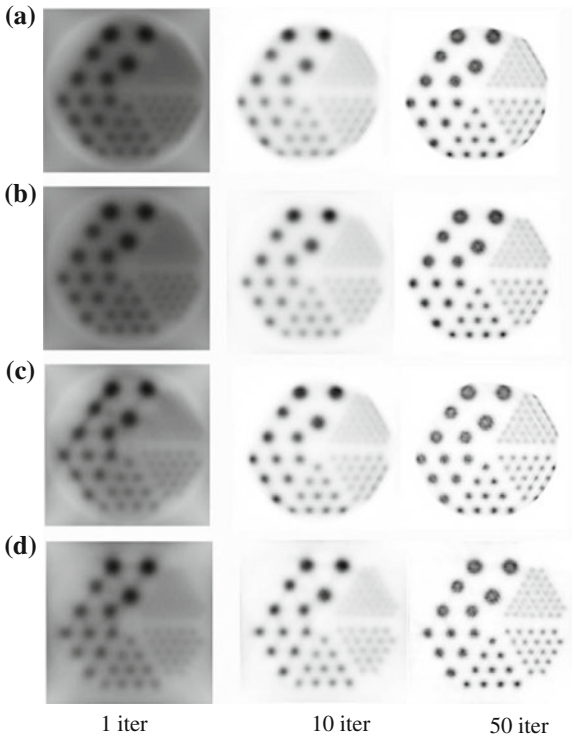


Fig. 23.1 Reconstructed images with: (a) EM-ML, (b) ISRA, (c) WLS, (d) ISWLS, after 1, 10 and 50 iterations respectively

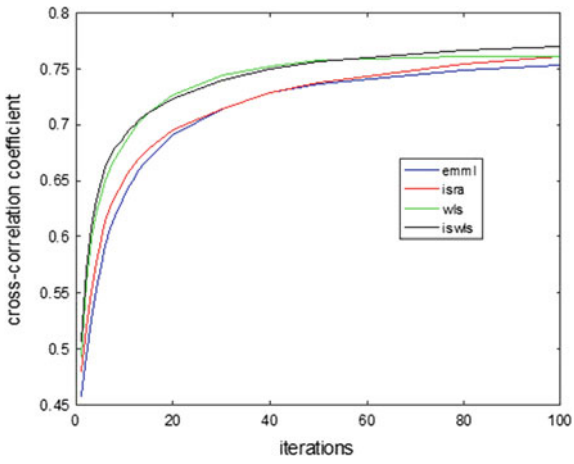


Fig. 23.2 Cross-correlation coefficient versus the number of iterations for EM-ML, ISRA, WLS, and ISWLS

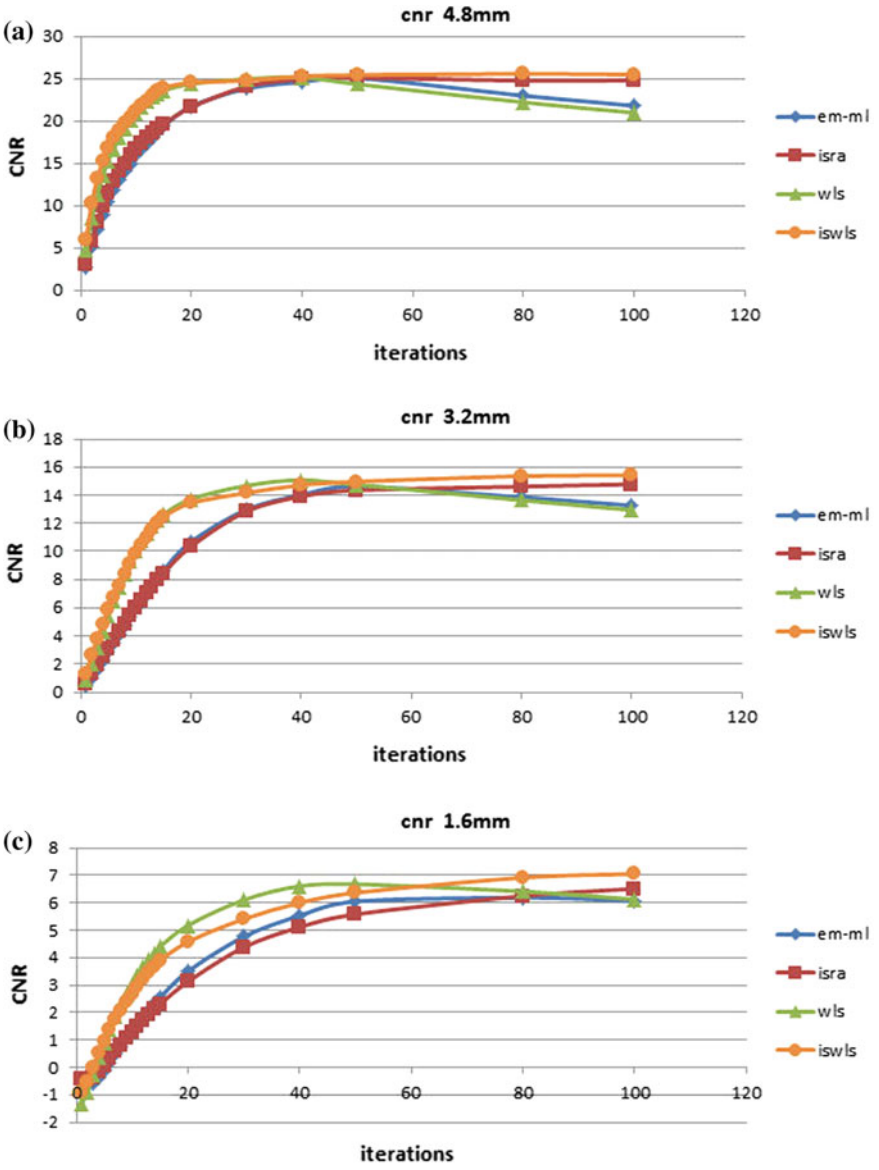


Fig. 23.3 CNRs versus iterations for: (a) 4.8 mm, (b) 3.2 mm and (c) 1.6 mm object diameter

23.3.1 Comparative Evaluation of Normal Versions

Figure 23.1 shows the reconstructed transaxial images with EMML, ISRA, WLS, and ISWLS after 1, 10 and 50 iterations.

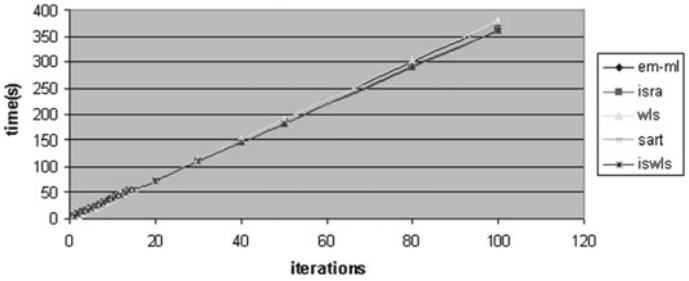


Fig. 23.4 Reconstruction time/slice as a function of the number of iterations

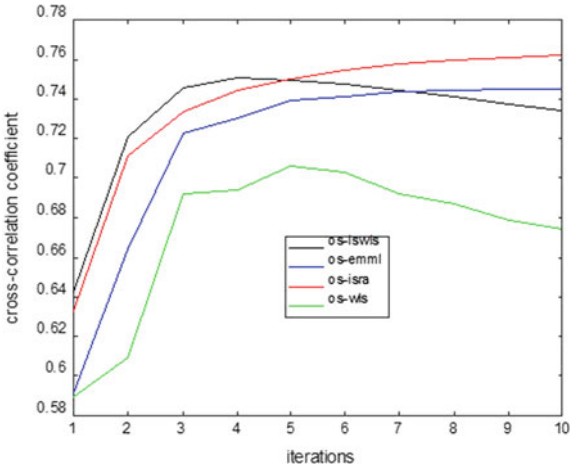


Fig. 23.5 Cross-correlation coefficient of OSEM, OS-ISRA, OS-WLS and OS-ISWLS versus the number of iterations

In Fig. 23.2 cross-correlation coefficient c [13] of every iterative method is plotted, versus the number of iterations. The cross-correlation coefficient c was calculated according to the equation:

$$c = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(Irecon_{ij} - \bar{I}recon\right)\left(Ireal_{ij} - \bar{I}real\right)}{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(Irecon_{ij} - \bar{I}recon\right)^2 \sum_{i=1}^{N}\sum_{j=1}^{N}\left(Ireal_{ij} - \bar{I}real\right)^2}}, \tag{23.19}$$

where $\bar{I}recon$ and $\bar{I}real$ are the reconstructed image and the true phantom activity image mean values, respectively. Cross-correlation coefficient is a similarity measure between reconstructed and real radiodistribution image. Its values are in the range $[-1, 1]$. Value $c = 1$ corresponds to fully correlated images.

Except for the cross-correlation coefficient that shows the average performance of the reconstruction methods, local contrast-to-noise ratios (CNR) [6] for rods with different diameters were calculated. CNRs for 4.8, 3.2, and 1.6 mm rods diameter

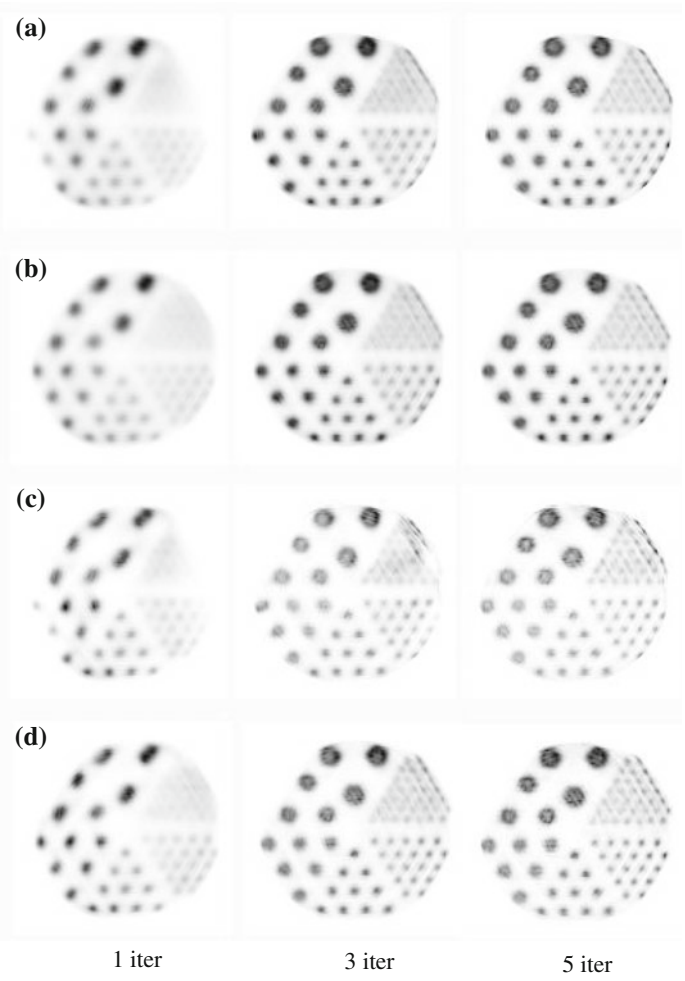


Fig. 23.6 Reconstructed images with: (a) OSEM, (b) OS-ISRA, (c) OS-WLS and (d) OS-ISWLS, after 1, 3 and 5 iterations using 15 subsets

were computed, using squared regions-of-interest (ROIs), 4.55, 3.85 and 1.15 mm in size, respectively. The ROIs were placed inside the corresponding objects. The number of selected ROIs was equal to the number of same sized objects. ROIs of the same sizes were positioned in three different background areas, each.

$$CNR_{ROI} \text{ was defined as: } \quad CNR_{ROI} = \frac{R_{obj_{ROI}} - R_{Backg_{ROI}}}{\sigma_{Backg_{ROI}}} \tag{23.20}$$

where $R_{obj_{ROI}}$ is the mean value of reconstructed objects in the corresponding ROIs, and $R_{Backg_{ROI}}$ is the mean value of the three background ROIs in each case.

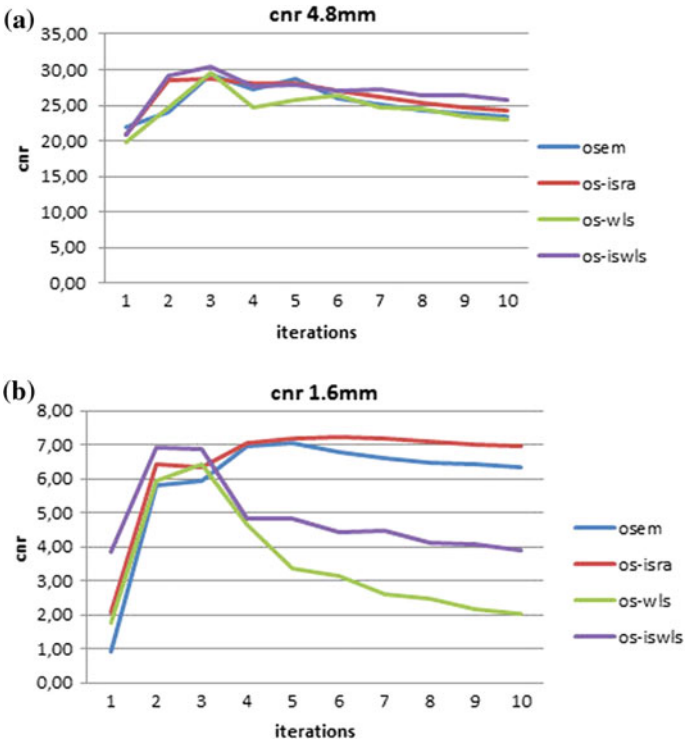


Fig. 23.7 CNRs versus iterations for (a) 4.8 mm, (b) 1.6 mm object diameter for the different OS algorithms

Further, $\sigma_{Backg_{ROI}}$ is the standard deviation of background values in the corresponding ROIs. The graphs in Fig. 23.3 illustrate the variation of CNR_{ROI} with respect to the number of iterations, for the three different objects diameters.

In Fig. 23.4 the reconstruction time for every iterative algorithm is presented as a function of the number of iterations. Reconstruction time calculations were performed on a Pentium M processor 1400 MHz (Intel Corp.) personal computer (RAM 1280 MB) under Windows XP Professional.

23.3.2 Comparative Evaluation of OS Versions

For the evaluation of the ordered subsets, iterative reconstruction methods projection data of the same Derenzo-type phantom has been used as in the comparative study of the simple algorithms.

In Fig. 23.5 cross-correlation coefficient c of every ordered subsets iterative method is plotted versus the number of iterations.

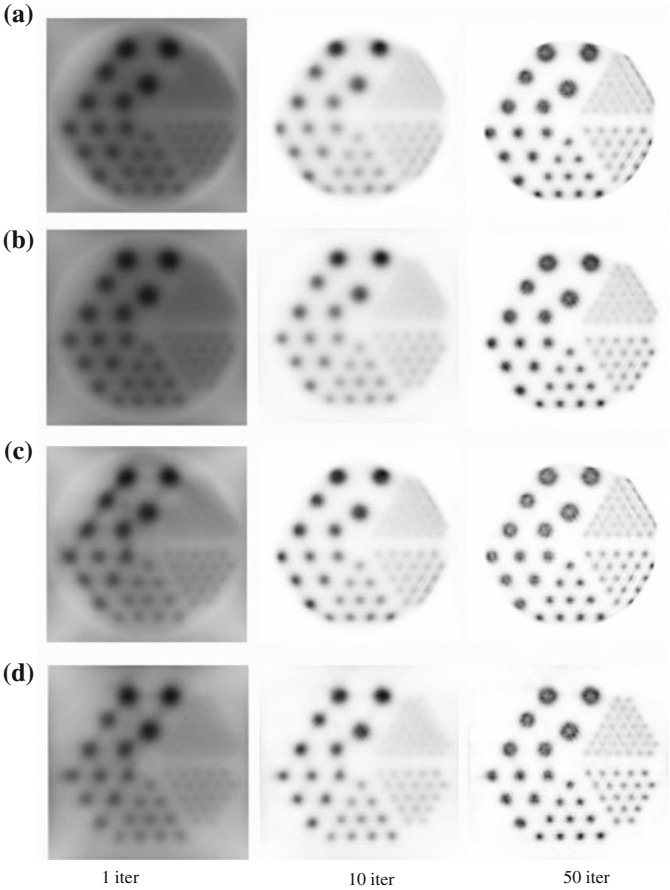


Fig. 23.8 Reconstructed images with: (a) MRP-EMML, (b) MRP-ISRA, (c) MRP-WLS, and (d) MRP-ISWLS, after 1, 10 and 50 iterations respectively

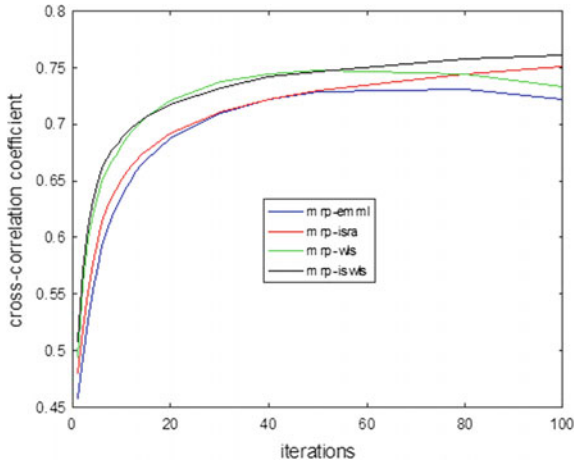


Fig. 23.9 Cross-correlation coefficient versus the number of iterations for MRP-EMML, MRP-ISRA, MRP-WLS, and MRP-ISWLS

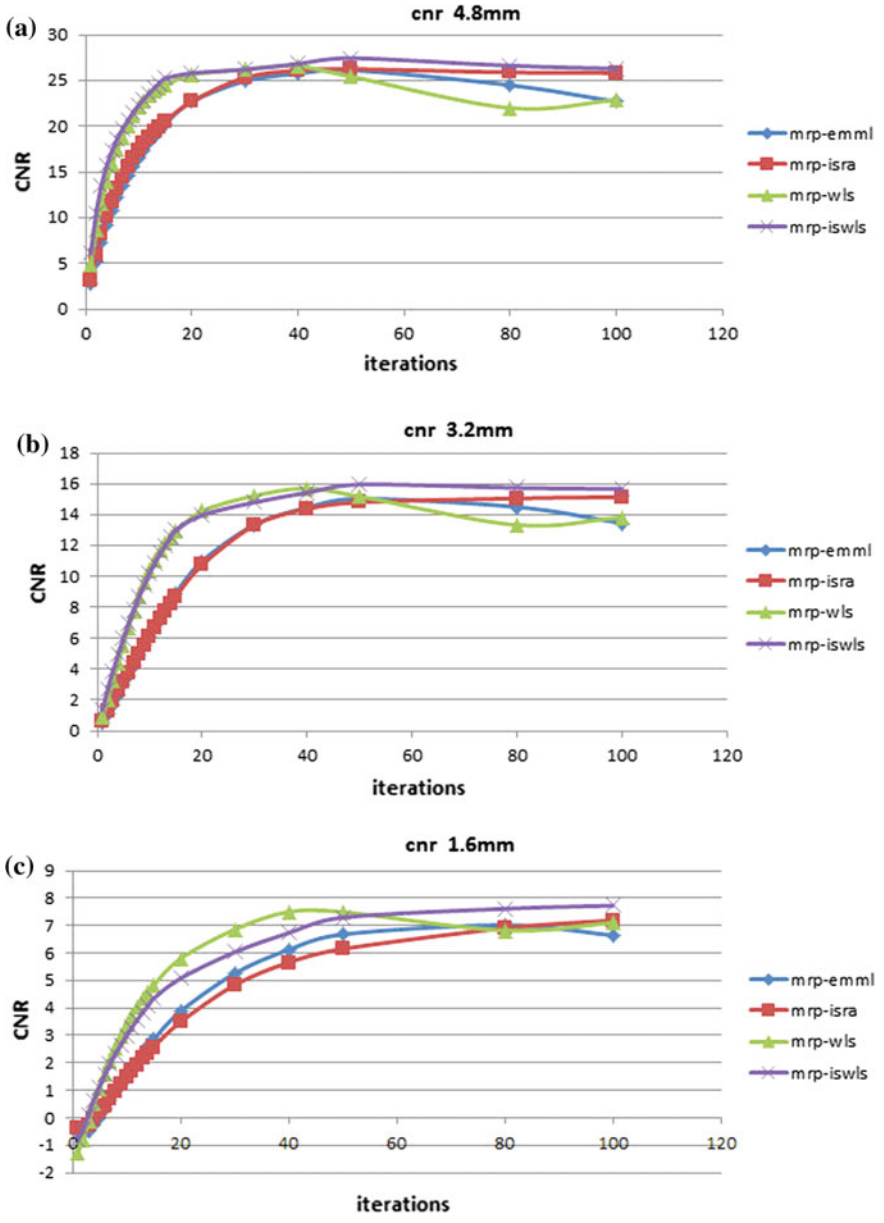


Fig. 23.10 CNRs versus iterations for: (a) 4.8 mm, (b) 3.2 mm and (c) 1.6 mm object diameter

Figure 23.6 presents reconstructed images with OSEM, OS-ISRA, OS-WLS and OS-ISWLS after 1, 3 and 5 iterations for 15 subsets.

In Fig. 23.7 CNRs for two objects of different diameter are plotted versus the number of iterations. CNRs are derived with the same method as explained in Eq. 23.20. Reconstruction time is the same for all ordered subsets algorithms under study. One iteration lasts 29 s for all ordered subsets methods.

23.3.3 Comparative Evaluation of MRP Versions

Figure 23.8 shows the reconstructed transaxial images with MRP-EMML, MRP-ISRA, MRP-WLS and MRP-ISWLS after 1, 10 and 50 iterations.

In Fig. 23.9 cross-correlation coefficient c of every iterative method is plotted versus the number of iterations.

The graphs in Fig. 23.10 illustrate the variation of CNR_{ROI} with respect to the number of iterations for the three different object diameter.

23.4 Discussion

According to results (Figs. 23.2, 23.3) EM-ML and ISRA presents similar behavior at the first 50 iterations. EM-ML converges slowly during the first 50 iterations to the best approximation of the true image. After the 50th iteration the algorithm enhances more small-sized objects but in general image contrast decreases and noise component starts to increase. ISRA on the contrary presents better noise manipulation. ISRA reaches constant CNRs values for big and small-sized objects. WLS shows almost identical noise manipulation as EM-ML algorithm. It reaches faster than EM-ML and ISRA the same CNRs values. ISWLS combines WLS's acceleration and ISRA's noise manipulation.

Reconstruction time of EM-ML and WLS is almost the same as a function of the number of iterations (≈3.8 s/iteration). Although it is not obvious from Fig. 23.4, ISRA and ISWLS are slower than EM-ML and WLS during the first 9 iterations. Their reconstruction speed is gradually improved by increasing the number of iterations. ISWLS and ISRA reconstruction time converges to the others' reconstruction time after 10 iterations. The reason for slow reconstruction process during the first iterations lies in the time needed for backprojection computations ($\sum_{i=1}^{M} a_{ij} y_j$ for ISRA and $\sum_{i=1}^{M} a_{ij} y_j^2$ for ISWLS) in the first iteration.

The OS versions reach faster than the simple versions the same cross-correlation coefficient and CNRs values, as expected. For example they reach the value of 0.75 of cross-correlation coefficient earlier (during the first 10 iterations) than the simple versions, which they present the same results after 50 iterations. In general OS-ISWLS is comparable to OSEM and shows better performance than OS-WLS, according

to CNRs graph, which indicates a better noise manipulation than OS-WLS. ISWLS reaches to the best approximation of the true image during the first 3 iterations for big and small-sized objects. OS-ISRA and OSEM needs few more iterations especially for small-sized objects.

The choice of 15 subsets was made after a comparative study of ISWLS and OS-ISWLS. We compared OS-ISWLS to ISWLS using 3, 9, 15, and 24 subsets. The comparative criterion was CNR, calculated as described in Eq. 23.20. Small number of subsets resulted in CNRs similar to ISWLS. Increasing the number of subsets resulted in images with high CNRs during the first ten iterations. Using 15 or 24 subsets the reconstructed images reach the highest CNR values before ten iterations. ISWLS achieves the same highest CNRs after 50 iterations. We chose 15 subsets to 24 because this number of subsets presented smaller image degradation as the number of iteration increased.

MRP-ISWLS presents higher cross-correlation values than MRP-EM-ML and MRP-ISRA. It shows the same high values of cross-correlation coefficient as MRP-WLS. As illustrated in Fig. 23.10 MRP-ISWLS presents high CNR ratios from the first iterations, higher than MRP-EM-ML and MRP-ISRA. Although it shows similar performance to MRP-WLS, its CNR ratios do not degrade after 50 iterations but tend to be stabilized. So, MRP-ISWLS presents a better noise manipulation than MRP-WLS.

Reconstruction time of MRP-EM-ML and MRP-WLS is almost the same (77.5 s/iteration). MRP-ISRA and MRP-ISWLS are slower than MRP-EM-ML and MRP-WLS during the first 9 iterations (79 s/iteration). Their reconstruction speed is gradually improved by the increasing of iterations number. MRP-ISWLS and MRP-ISRA reconstruction time converges to the others' reconstruction time after 10 iterations. The reason for slow reconstruction process during the first iterations lies in the time needed for backprojection computations in the first iteration.

In order to determine a satisfactory value of b, various values were applied to MRP-ISWLS. These values were 0.001, 0.01, 0.1, 1, 1.5. The value $b = 0.001$ presented higher CNR values in comparison to ISWLS's results.

In this study, data were corrected for scanner sensitivity prior to the reconstruction process. Such a correction could be incorporated during the reconstruction process, but that exceeds our research purpose.

23.5 Part II

23.5.1 Introduction

Extracting useful medical information, for the inside of the human body, from medical images belongs to the field of medical image segmentation. In recent years the computational medical image segmentation plays an increasingly important role in medical imaging. The image segmentation is an important pretreatment step in medical

imaging, especially in automatic detection of anatomical structures and pathological conditions. It is the separation process of structures of interest, the implementation of which resulted in the development of numerous algorithms. Because of the wide variety of objects, shapes and variations in image quality, segmentation remains a difficult process. A global segmentation method, which produces satisfactory results for all imaging applications, does not exist. The medical images more often contain high levels of noise, and defects due to incomplete data collection. This could cause appreciable difficulties when applying classical segmentation methods, such as edge detection and thresholding. As a result, these techniques either fail completely or they need an image processing stage after segmentation to remove false limits of the object of interest. Furthermore, subsequent analysis and interpretation of the segmented structures is prevented by the dimensions of the pixel or the voxel.

Each segmentation algorithm aims to identify the pixels belonging to the object of interest or to determine the pixels of the object contour. Detection methods of the object region are based on the intensity values or the texture of the pixels. Contour determination techniques use the derivative of the image, which has high values at the boundaries of objects.

One of the most popular methods of segmentation is the use of deformable models [20, 40]. According to these techniques, near the region of interest an elastic contour model is placed which, through repetitive processes, is adapted to the actual contour. The widely recognized ability of the deformable models comes from their ability to segment, match and trace images of anatomical structures, utilizing restrictions arising from imaging data combined with a priori information relative to the location, size and shape of these structures. The deformable models are capable of resolving the often considerable variation in biological structures over time and from person to person. For this reason they are the focus of research in the segmentation and have been used extensively in medical applications with promising results. Also, they support intuitive interaction mechanisms that allow experts to interpret the results. The deformable models are especially suitable for segmentation of images characterized by intense artifacts, noise, limited spatial resolution and hardly detectable borders between the structures. However, deformable models may result in false contour if the original outline is not placed close to real. They can also present problems in areas of intense curvature of the contour.

The deformable models in the literature are referred to as snakes, active contours or active surfaces, balloons, deformable contours or deformable surfaces. These models are curves or surfaces defined on image domain that deform under the influence of internal and external forces. Internal forces are related to the curve or the surface itself and are designed to keep the model smooth during the deformation process. External forces adjust the model to the real object boundary and their computation is based on image information. In theory, because of the constrain for smooth contours and their ability to incorporate a priori information of object shape, deformable models can handle noise problems in images and contour discontinuities. So, they permit the description of the contour as a continuous mathematical model and they are able to achieve inter-pixel segmentation accuracy.

In this work various methods of deformable models are presented, namely, the classical snake [21, 25, 28], the gradient vector field snake (GVF snake) [36–39] and the topology-adaptive snake (t-snake) [29–32, 34]. Also, the method of self-affine mapping system [22, 23] is presented as an alternative to the snakes. The self-affine mapping system was implemented using an adapting scheme for determining the size of areas with similarity [22]. The minimum distance was used as optimization criterion which, according to Sect. 23.7, is more suitable for weak edges detection. All methods were applied to glaucomatic retinal images with the purpose of segmenting the optical disk. Moreover, the aforementioned methods were compared in terms of segmentation accuracy.

The optical disk is the optical nerve and the vessels' entrance point in the retina. In fundus gray images it appears as a luminous white area. It has an almost round shape that is interrupted by the outgoing vessels. Sometimes the optical disk has an elliptical shape because of the small but not negligible angle between the planes of the image and the object. Optical disk size varies from patient to patient.

Optical disk segmentation is a very important preprocessing stage of many algorithms, which have been designed for automatic detection of retinal anatomical structures and pathological conditions. For example, the detection methods of some vessels and their junctions start from the optical disk area, where the big vessels lie. These can serve as starting points for the detection of the other vessels [35]. Another important feature in retinal images is macula. Macula's position usually is estimated according to the optical disk's position under the condition that the distance between the macula and the optical disk is constant [26, 27]. Moreover, optical disk camouflage contributes to better and easier lesions detection related to different retinopathies [20]. Furthermore, the optical disk center can be used as a reference point for distance measurements in retinal images. In addition, the segmented optical disk can be a reference area for the registration of retinal images acquired in different time or with a different method. Retinal images registration can reveal changes in vessels' size and disposition inside the optical disk, as well as changes in optical disk size related to serious eye diseases, such as glaucoma and vessels neoplasia [33].

23.6 Theory

23.6.1 Classical Snake

The classical snakes are used widely in image processing, in computer vision and in medical imaging applications for allocating object boundaries. The classical parametric elastic models (classical snakes), due to Kass, Witkin and Terzopoulos, change shape and finally are adapted to the real object boundary according to the minimization process of an energy function. The energy function reaches its global minimum when the active contour is smooth and coincides with the real object boundary.

23.6.2 Gradient Vector Flow Snake (GVF Snake)

The use of classical snake is unfortunately limited because it must be initialized close to the true contour and because of its inefficient convergence in boundary concavities. Xu and Prince proposed an improved snake model in order to overcome classical snakes' limitations. In particular, they introduced a new external force, the gradient vector flow (GVF), that is computed as the diffusion of gradient vectors of a gray or binary edge map of the image. According to Xu and Prince, the usual external forces are conservative forces that make the active contour unable to successfully approximate boundary concavities. GVF is a non- conservative force. Mathematically, it is based on the Helmholtz theorem, according to which a general static vector field can be separated in a conservative and tubular field. GVF snake was designed to have conservative and tubular characteristics in order to present the desirable properties of adequate initialization range and convergence in boundary concavities.

23.6.3 Topologically Adaptable Snake (T-Snake)

T-snake comprises another variant of classical snake. It is based on a space partitioning technique to create a topologically adaptable snake. The difference between the classical snake and T-snake is the use of an affine cell image decomposition, so as to iteratively reparametrize the snake and to make topological transformations. Image is partitioned into a net of discrete triangular cells. As the snake evolves under the influence of internal and external forces, it is reparametrized with a new set of nodes and elements. The reparametrization process consists of an efficient calculation of the intersections of the snake with the image net. These intersection points might be nodes of the updated T-snake. In 2D the T-snake is a 2D curve consisting of N nodes connected in series. The T-snake is a discrete version of the classical snake and retains many of its properties.

23.6.4 Self-Affine Mapping System

The self-affine mapping system is a technique similar to the snake model that adjusts an initial curve to the real object contour, using a self-affine mapping system which is used widely in fractal encoding. This particular method has an advantage over conventional snakes, mostly because of its ability to detect distinct and blurred edges with significant accuracy. It has replaced the process of energy minimization of the classical snake with a contractive self-affine mapping system that is used in the creation of fractal shapes. The parameters of the system are determined after a blockwise self-similarity analysis of the gray image through a simplified algorithm of fractal coding. The use of the self-affine mapping system is due to the fact that the

points of the initial map, when they are positioned near image edges, after iterative contractions of the map, they are finally attached to the edges. This attraction can be exploited for contour extraction that has the shape of a curve of similar points rather than a curve of smooth points which are detected by the snake model.

Suppose an image $g(x)$ is defined in $G \subset R^n$. If there exist affine transformations $a_i : A_i \to R^n$ and $\beta_i : R^1 \to R^1$ so that

$$\forall x \in A_i,\ g(x) = \beta_i(g(a_i(x))),\ i = 1, 2, \ldots, I \tag{23.21}$$

for some image regions $A_i \subset G$ then the texture in A_i is similar to the texture in $a_i(A_i)$ and the image presents self-similarity in these two regions A_i and $a_i(A_i)$. The set $\{A_i, a_i, \beta_i | i = 1, 2, \ldots, I\}$, where I is the total number of regions A_i, is called self-affine model of the image. If Eq. 23.21 holds then it can be written that:

$$\forall x \in a_i(A_i),\ g(x) = \beta_i^{-1}(g(a_i^{-1}(x))),\ i = 1, 2, \ldots, I \tag{23.22}$$

so there arises another self-affine model of the image the set $\{a_i(A_i), a_i^{-1}, \beta_i^{-1} | i = 1, 2, \ldots, I\}$. The transformations a_i dilate maps A_i and β_i are unit operators. If Ω is the set of subsets of G then the self-affine map $S : \Omega \to \Omega$ is defined as:

$$S(X) = \left\{ \bigcup_{i=1}^{I} a_i(A_i \cap X) \right\} \cup C \tag{23.23}$$

where X is a subset of G and C a fixed set. When X is known its intersection with A_i is mapped through the affine transformation a_i and the union of all these mapped regions with C results in the final map $S(X)$.

For a 2D image $n = 2$, A_i are squared image regions, subsets of G, and transformations a_i are defined as:

$$\forall x \in A_i, a_i(x) = r_i(x - \bar{x}_i) + (\tau_i + \bar{x}_i) \tag{23.24}$$

and

$$r_i > 1 \tag{23.25}$$

where $\bar{x}_i$ the central point of A_i. Moreover, the self-affine model assumes that:

$$\beta_i(g(a_i(x))) = p_i g(a_i(x)) + q_i,\ p_i \in [0, 1] \tag{23.26}$$

In order for map S to be determined, the regions A_i are first defined. Then an adequate algorithm searches for the best values of the parameters r_i, p_i, q_i and $\tau_i = (s_i, t_i)$ of the map, so that Eq. 23.25 is satisfied for every A_i, so the self-affine models $\{A_i, a_i, \beta_i\}$ and $\left\{a_i(A_i), a_i^{-1}, \beta_i^{-1}\right\}$ are determined. The searching is performed through a block-matching algorithm. The block-matching algorithm consists of the following steps:

STEP 1: Initialization of r, s and t. Also, the difference $E = g(\mathbf{x}) - \beta_i(g(a_i(\mathbf{x})))$ is assigned a very large value.
STEP 2: For every $\mathbf{x} \in A_i$ the value of $g(a_i(\mathbf{x}))$ is computed. Because the sampling points $\mathbf{x}$ may be between image pixels, the values $g(a_i(\mathbf{x}))$ are computed using bilinear interpolation.
STEP 3: Initialization of p and q.
STEP 4: Computation of $\beta_i(g(a_i(\mathbf{x})))$ for every $\mathbf{x} \in A_i$.
STEP 5: Computation of the difference E. For this computation the Mean Square Error (MSN) or the Absolute Mean Distance (AMD) are usually used. If the new E is smaller than the initial value then the initial value is replaced by the new E and the values of r, s, t, p and q are registered as r_i, s_i, t_i, p_i and q_i respectively.
STEP 6: If all p and q are checked then the algorithm moves to STEP 7, otherwise it goes back to STEP 4.
STEP 7: If all r, s, and t are checked the algorithm is terminated, otherwise it goes back to STEP 2.

As in the conventional snake model this method must be initialized by a rough contour. The pixels inside the initial curve take value 1 and the rest belongs to the background having value 0. This way a binary image is created, called alpha mask. The purpose of the self-affine mapping system is the fitting of the alpha mask contour to the real object boundary. In order for the initial curve b to be attached to the real contour c three conditions must be satisfied:

1. the set c must equal the invariable set S
2. the transformations a_i^{-1} must be systolic
3. the set b should be adequately close to c

Moreover, parameters s, t which are defined during the block-matching process should be determined, so that every set $a_i(A_i)$ contains the corresponding A_i, namely: $-\frac{r-1}{2}e \leq s$ and $t \leq \frac{r-1}{2}e$, where $e$is the size of regions A_i. Finally, the total number of iterations ν should be: $\nu > \frac{\log \frac{e}{2}}{\log r}$.

23.7 Results

23.7.1 New Self-Affine Mapping System Optimization Criterion

In the self-affine mapping system the size of the areas $a_i(A_i)$ was chosen to be twice the size of the areas A_i, namely $r = 2$ so as condition 2 to holds. The searching area for the block-matching process was $[-\frac{n}{4}, \frac{n}{4}]$, where n the size of the area A. When the value of n is small (e.g. $n = 8$) condition 3 is not satisfied, while when it is large (e.g. $n = 32$) condition 3 is satisfied, but the final contour is a rough approximation of the optic disk true boundary and condition 1 is now not satisfied

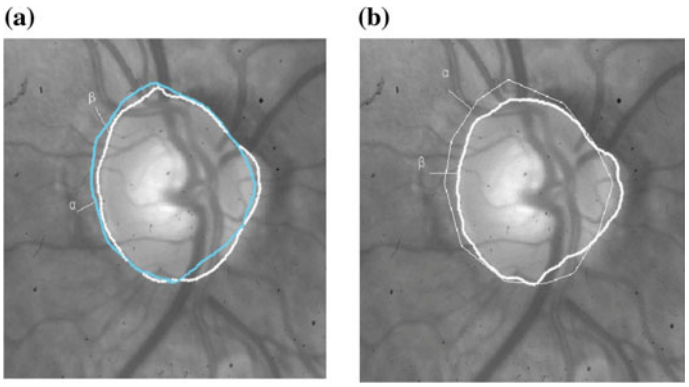


Fig. 23.11 (a) Final contours with $n = 8$ (curve a (blue)) and $n = 32$ (curve β (white)), (b) Final contour using the adapting scheme (curve β), curve a is the initial contour

and fine details of the object's boundary are not detected. So, we chose an adapting scheme where n was assigned an initial big value ($n = 32$), which is gradually decreased to $n = 4$, namely $e_{\max} = 32$ and $e_{\min} = 4$. The number of iterations was set to $\nu = \frac{\log \frac{\varepsilon}{2}}{\log r} + 1$. In Eq. 23.26 p was set to 1 and q to 0. As optimization criteria of measuring the difference between $g(x)$ and $\beta_i(g(a_i(x)))$, two criteria were tested and optically evaluated. The first was the classical AMD and the second was the Minimum Distance (MD).

Figure 23.11 presents the final contours using $n = 8$ and $n = 32$ fused in the same optical image (Fig. 23.11a) and the final contour using the adapting scheme (Fig. 23.11b).

From Fig. 23.11b one can observe the strong attraction of mask boundary points from the vessels in the area of the optic disk. Vessels are strong edges. The AMD function is minimized towards these strong edges and the mask boundary points are caged from the edges of the vessels. Figure 23.12 shows two examples of a self-affine mapping system application, where AMD (left column) and MD (right column) were used as optimization functions. MD application results in the detection of optic disk boundary and not the detection of points that belong to strong edges like the vessels. However, the detection of weak edges reduces the degrees of freedom in mapping out the initial contour.

23.7.2 Comparison of Elastic Models

The different segmentation methods were applied to 26 retinal images 512×512 pixels in size, in order for the optic disk boundary to be extracted. For the implementation of the classical snake the external force, derived from a Gaussian potential field with $\sigma = 3$, was used accompanied with a pressure force with constant weights. In every

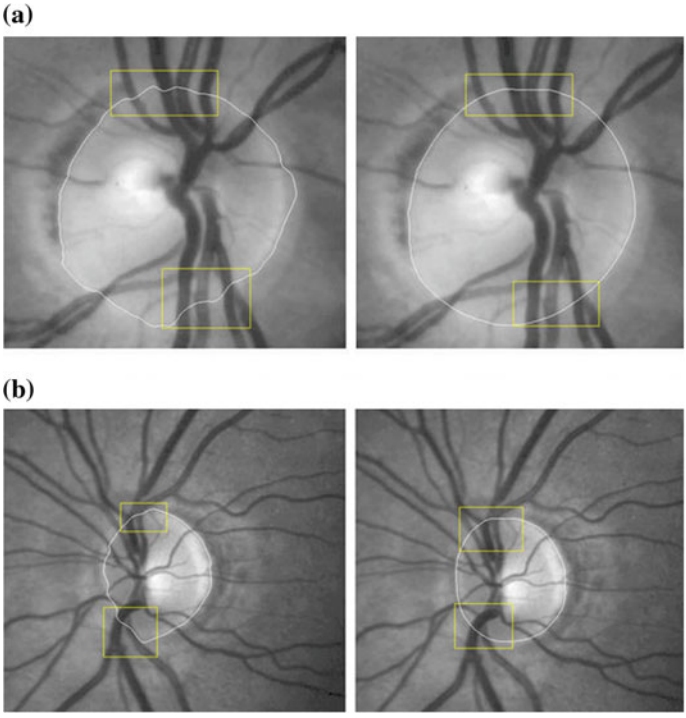


Fig. 23.12 Final contours using AMD (left column) and MD (right column) for (a) big and (b) small sized objects (optic disk)

position of the snake nodes, external force values were calculated using the bilinear interpolation method. Constant weights were also used for the internal forces. Initial contours were placed close to the real object boundary.

The GVF field $\vec{v}(x, y) = [u(x, y), \upsilon(x, y)]$, was calculated according to the equations:

$$u_{i+1} = u_i + \mu 4\nabla^2 u_i - |\nabla f|\,(u_i - f_x) \tag{23.27}$$

$$\upsilon_{i+1} = \upsilon_i + \mu 4\nabla^2 \upsilon_i - |\nabla f|\,\left(\upsilon_i - f_y\right) \tag{23.28}$$

were f_x and $f_y$the first derivatives of the edge map f of the image I. The edge map was derived as $f(x, y) = |\nabla [G_\sigma(x, y) * I(x, y)]|$, with $\sigma = 3$. Initial values for u and υ were $u_{init} = f_x$ and $\upsilon_{init} = f_y$. Moreover, $\mu = 0.29$ and 80 total iterations were used for the calculation of the GVF field for all the 26 applications. Initial contours were placed close to the real ones.

For the implementation of the T-snake the external force derived from a Gaussian potential field with $\sigma = 3$ was used, accompanied with a pressure force with constant weights. The initial seed point was chosen to be inside the object of interest. According to this seed point the algorithm initializes a square snake.

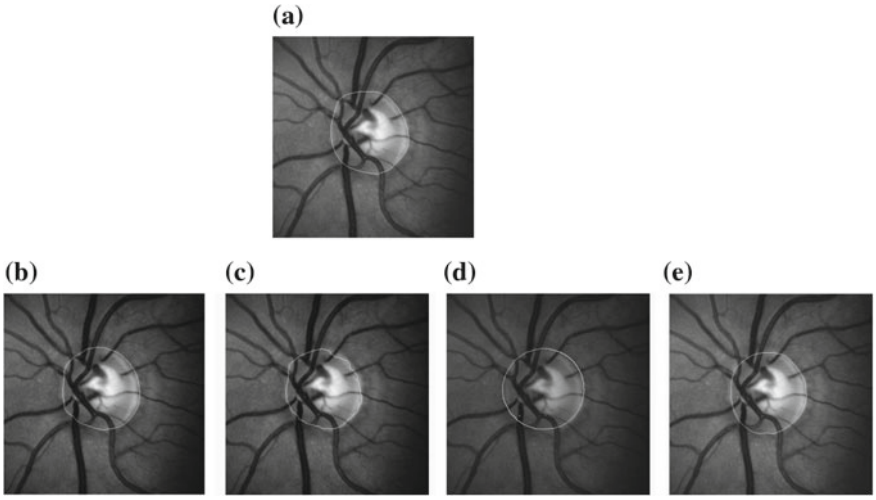


Fig. 23.13 Optic disk extraction with (a) manual by expert, (b) the classical snake ($\alpha = 2$, $\beta = 2$, $\gamma = 1$, w $= 7$, $w_p = 0.05$, 125 total iterations), (c) the GVF snake ($\alpha = 2$, $\beta = 2$, $\gamma = 1$, w $= 7$, $w_p = 0.05$, $\mu = 0.29$, 80 iterations for GVF calculation and 125 total iterations), (d) the T-snake ($\alpha = 20$, $\beta = 20$, w $= 71$, $w_p = 70$, 45 total iterations), (d) the self-affine mapping system ($e_{min} = 4$, $e_{max} = 32$, r $= 2$)

The self-affine mapping system was implemented according to Sect. 23.7.1.

Figure 23.13 presents comparatively optic disk extraction using the classical snake model, the GVF model, the T-snake and the self- affine mapping system.

For the comparative evaluation of the different deformable models normalized mean square error (nmse), between the real optic disk contour (I_{real}) and the final contour ($I_{deformabls}$) extracted by every deformable model method applied to 26 retinal images, was computed. The real contours were drawn by an expert. The nmse was calculated according to the equation:

$$nmse = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(I_{deformable_{ij}} - I_{real})^2}{\sum_{i=1}^{N}\sum_{j=1}^{N} I_{real}^2} \tag{23.29}$$

Fig. 23.14 presents nmse for the four deformable model techniques and for the 26 retinal images.

Figure 23.15 shows the total segmentation time for the four deformable models for the 26 retinal images.

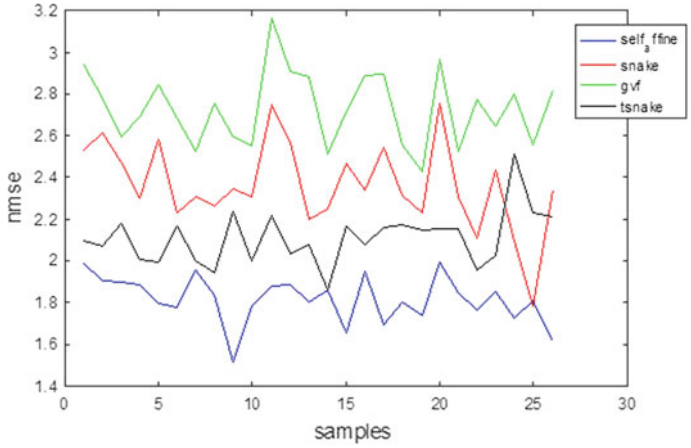
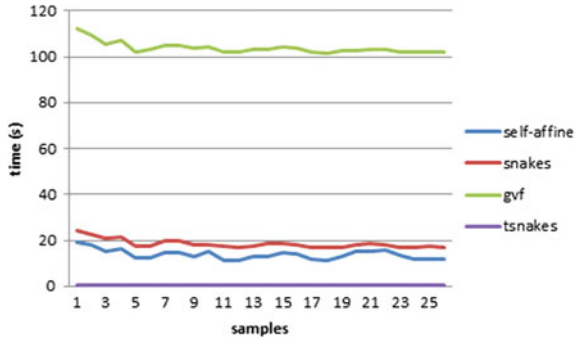


Fig. 23.14 Nmse curves for the four deformable model techniques applied to 26 retinal images

Fig. 23.15 Total segmentation time (s) for the four deformable models applied to 26 retinal images



23.8 Discussion

From Fig. 23.13 the final contour of the classical model is smooth and approximates the real one. GVF snake is an alternative of the classical snake designed to detect complicated boundaries with high curvature and sharp edges. The optic disk boundary does not have such characteristics, it is almost round. The range of the two methods is almost the same. The classical snake and the GVF snake must be initialized close to the real contour. According to Fig. 23.15 the GVF snake algorithm is also slower than the classical snake method, mostly because of the extra time it needs to calculate the external GVF force.

T-snake results in a satisfying optic disk contour (Fig. 23.13d). The biggest advantage of the T-snake algorithm is its range. It is initialized in a point inside the optic disk. Moreover, the total segmentation time of T-snake is small, smaller than any other deformable method. So, T-snake presents a robust and efficient segmentation technique.

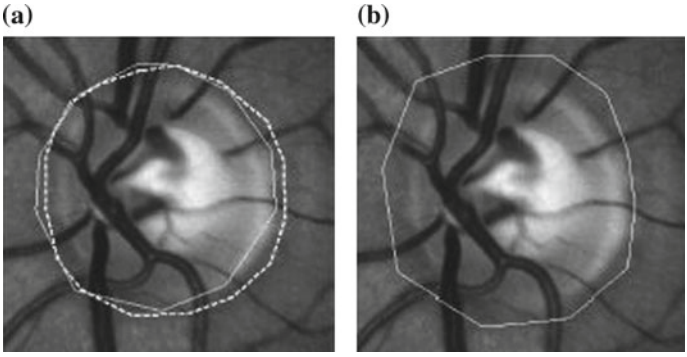


Fig. 23.16 Examples of contour initialization for (a) the classical snake and the GVF snake and (b) the self-affine mapping system

The self-affine mapping system is superior in optic disk boundary extraction than the other techniques, according to Figs. 23.13 and 23.14. Furthermore the algorithm is independent from optic disk size and image intensity. Also, with the choice of minimum distance as a matching criterion the caging of the model from the vessels is avoided. Self-affine mapping system seems to present bigger independence from initialization than the classical snake and the GVF snake (Fig. 23.16). The initial contour is placed away from the real one (Fig. 23.16b). The total segmentation time of the algorithm is also adequate, according to Fig. 23.15, since it is faster than the classical snake and the GVF snake. The self-affine mapping system succeeds also in small nmse values (Fig. 23.14). Another advantage of the self-affine mapping system is that this method is self-terminated, a characteristic that the other deformable methods do not present.

In general self-affine mapping system succeeds in approximating very well the true optic disk boundary. The distance between the real and the initial contour depends on the value of $e_{\max}$. If $e_{\max}$ is increased the degrees of freedom in lining the initial contour are increased.

23.9 Conclusions

Iterative image reconstruction algorithms have been proposed as an alternative to conventional analytical methods. Despite their computational complexity, they become more and more popular, mostly because they can produce images with better contrast-to-noise (CNR) and signal-to-noise ratios at a given spatial resolution, compared to FBP techniques. Iterative methods are able to incorporate a model of all the physical phenomena during the acquisition process, including scanner characteristics. Based on predetermined criteria and after a series of successful iterations, they attempt to find the best approach to the true image of radioactivity spatial distribution.

In this paper, different simultaneous iterative reconstruction schemes were applied to data acquired from a simulation of a small-animal PET scanner. A new iterative scheme was discussed named ISWLS. EM-ML, ISRA, WLS, and ISWLS and their OS and MRP versions were implemented and evaluated in terms of task-dependent measures for quantization and detection. ISWLS combines WLS's reconstruction acceleration with ISRA's good noise manipulation. ISWLS could be preferable for use in 3-D reconstruction applications, where the precalculation of the factor $\sum_{j=1}^{M} a_{ij} y_j$ which is constant, will lessen the computational cost and demands for high computational memory.

Another important field of medical imaging is image segmentation. It is used especially in automatic anatomical structure and pathological areas detection. Because of the variety of object shapes and the variance in image quality, image segmentation remains a difficult task. Every segmentation algorithm aims at the detection of the image pixels that belong to the object of interest. Deformable models are the most popular image segmentation techniques. They are designed to approximate the significant variance of biological structures with time and from person to person.

In this work various methods of deformable models were compared, which are the classical snake, the gradient vector field snake (GVF snake) and the topology-adaptive snake (t-snake). Also, the method of self-affine mapping system was presented as an alternative of the snake models. The self-affine mapping system was implemented using an adapting scheme. Moreover, Minimum Distance was introduced as an optimization criterion more suitable for optic disk boundary detection. All methods were applied to glaucomatic retinal images with the purpose of segmenting the optical disk. The methods were compared in terms of segmentation accuracy. The self-affine mapping system presents efficient segmentation time, segmentation accuracy and significant independence from initialization.

References

Part I

1. Alenius, S., Ruotsalainen, U., Astola, J.: Using local median as the location of the prior distribution in iterative emission tomography image reconstruction. In: Proceedings of the IEEE Medical Image Conference (1997)
2. Anderson, M.M., Mair, B.A., Rao, M., Wu, C.H.: Weighted least-squares reconstruction methods for positron emission tomography. IEEE Trans. Med. Imag. **16**(2), 159–165 (1997)
3. Archer, G.E.B., Titterington, D.M.: The iterative image space reconstruction algorithm (Isra) as an alternative to the EM algorithm for solving positive linear inverse problems. Stat. Sin. **5**, 77–96 (1995)
4. Browne, J., De Pierro, A.R.: A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography. IEEE Trans. Med. Imag. **15**(5), 687–699 (1996)
5. Cherry, S.R., Gambhir, S.S.: Use of positron emission tomography in animal research. ILAR J. **42**(3), 219–231 (2001)
6. Cherry, S.R., Sorenson, J.A., Phelps, M.E.: Physics in Nuclear Medicine. Saunders-Elsevier (2003)

7. Comtat, C.M., Defrise, M., Townsend, D.: The FAVOR algorithm for 3D PET data and its implementation using a network of transputers. Phys. Med. Biol. **38**, 929–944 (1993)
8. Fessler, J.A., Hero, A.O.: Space-alternating generalized expectation-maximization algorithm. IEEE Trans. Signal Process. **42**(10), 2664–2676 (1994). October
9. Green, P.J.: On use of the EM algorithm for penalized likelihood estimation. J. R. Stat. Soc. Ser. B **52**, 443–452 (1990)
10. Hanson, K.M.: Chapter 3: Bayesian and related methods in image reconstruction from incomplete data. In: Stark, H. (ed.) Image Recovery: Theory and Applications. Academic Press, Orlando (1987)
11. Hudson, H.M., Larkin, R.S.: Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans. Med. Imag. **13**(4), 601–609 (1994)
12. Jiang, M., Wang, G.: Convergence studies on iterative algorithms for image reconstruction. IEEE Trans. Med. Imag. **22**(5), 569–579 (2003)
13. Karali, E., Asvestas, P., Nikita, K.S., Matsopoulos, G.K.: Comparison of different global and local automatic registration schemes: an application to retinal images. In: Proceedings of MICCAI 2004. LNCS, vol. 3216, Part 1, pp. 813–820. Springer, Berlin (2004)
14. Karali, E., Pavlopoulos, S., Koutsouris, D.: Assessment of iterative image reconstruction techniques for small-animal PET imaging applications. In: Proceedings of 8th International Conference on Bioinformatics and Bioengineering (BIBE), Athens, Greece, October 2008 (2008)
15. Karali, E., Pavlopoulos, S., Lambropoulou, S., Koutsouris, D.: ISWLS: novel algorithm for image reconstruction in PET. IEEE Trans. Inf. Technol. Biomed. **15**(13), 381–6 (2011)
16. Karali, E., Koutsouris, D.: Towards novel regularization approaches to PET image reconstruction. J. Biosci. Med. (JBM) **1**, 6–9 (2013)
17. Phelps, M.E.: PET Molecular Imaging and its Applications. Springer, New York (2004)
18. Shepp, L.A., Vardi, Y.: Maximum likelihood reconstruction for emission tomography. IEEE Trans. Med. Imag. **MI–1**(2) (1982)
19. Tarantola, G., Zito, F., Gerundini, P.: PET instrumentation and reconstruction algorithms in whole-bode applications. J. Nucl. Med. **44**(5), 756–768 (2003)

Part II

20. Alireza, O., Majid, M.: Barry Thomas and Richard Markham, Colour Morphology and Snakes for Optic Disc Localisation (2002)
21. Cohen, L.D.: On active contour models and balloons. Comput. Vis. Gr. Imag. Process. Imag. Underst. **53**(2), 211–218
22. Ida, T., Sambonsugi, Y.: Self-affine mapping system and its application to object contour extraction. IEEE Trans. Imag. Process. **9**(11) (2000)
23. Ida, T., Sambonsugi, Y., Watanabe, T.: Boundary fitting of extracted objects using LIFS. Syst. Comput. Jpn. **31**(8) (2000)
24. Karali, E., Lambropoulou, S., Koutsouris, D.: Elastic models: a comparative study applied to retinal images. Technology and Health Care, vol. 19, pp. 1–13. IOS Press, Amsterdam (2011)
25. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. Int. J. Comput. Vis. 321–331 (1988)
26. Li, H., Opas, C.: Fundus image features extraction. In: Proceedings of the 22nd Annual EMBS International Conference (2000)
27. Li, H., Opas, C.: Automatic location of optic disk in retinal images. IEEE (2001)
28. McInerney, T., Terzopoulos, D.: Deformable models in medical image analysis: a survey. Med. Imag. Anal. **1**(2), 91–108 (1996)
29. Mcinerney, T.: Topologically adaptable deformable models for medical image analysis. Ph.D. Dissertation, University of Toronto (1997)

30. McInerney, T., Terzopoulos, D.: Medical image segmentation using topologically adaptable surfaces. In: Proceedings of CVRMed'97, France (1997)
31. McInerney, T., Terzopoulos, D.: Topology adaptive deformable surfaces for medical image volume segmentation. IEEE Trans. Med. Imag. **18**(10), 840–850 (1999)
32. McInerney, T., Terzopoulos, D.: T-snakes: topology adaptive snakes. Med. Imag. Anal. **4**, 73–91 (2000)
33. Rosin, P.L., Marshall, D., Morgan, J.E.: Multimodal retinal imaging: new strategies for the detection of glaucoma. In: IEEE ICIP (2002)
34. Terzopoulos, D.: Deformable models: classic, topology-adaptive and generalized formulations
35. Thomas, W., Jean-Claude, K.: Segmentation of color fundus images of the human retina: detection of the optic disk and the vascular tree using morphological techniques (2001)
36. Xu, C., Prince, J.: Gradient vector flow: a new external force for snakes. In: IEEE Proceedings on Computer Society Conference on Computer Vision and Pattern Recognition, pp. 66–71 (1997)
37. Xu, C., Prince, J.: Snakes, shapes and gradient vector flow. IEEE Trans. Imag. Process. **7**(3) (1998)
38. Xu, C., Prince, J.L.: Generalized gradient vector flow external forces for active contours. Signal Process. **7**(1), 131–139 (1998)
39. Xu, C., Prince, J.: Gradient vector flow deformable models. Handbook of Medical Imaging. Academic Press, London (2000)
40. Xu, Yezzi, A., Prince, J.: On the relationship between parametric and geometric active contours. In: Proceedings of the 34th Asilomar Conference on Signals, Systems and Computers, pp. 483-489 (2000)