HANDBOOK OF ALGEBRA

Volume 6

edited by
M. HAZEWINKEL
CWI, Amsterdam

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

North-Holland is an imprint of Elsevier

North-Holland is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
Linacre House, Jordan Hill, Oxford OX2 8DP, UK
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands

First edition 2009

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

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978-0-444-53257-2
ISSN: 1570-7954

Printed and bound in Hungary
09 10 10 9 8 7 6 5 4 3 2 1

Typeset by: diacriTech, India

Preface

Basic philosophy

Algebra, as we know it today (2008), consists of a great many ideas, concepts, and results, and this was also the case in 1995 when this handbook started (which does not mean that nothing has happened in those 12 years; on the contrary, the field of algebra and its applications has developed at a furious pace).

A reasonable estimate of the number of all the various different concepts, ideas, definitions, constructions, results, . . . in algebra would be somewhere between 50 000 and 200 000. Many of these have been named and many more could (and perhaps should) have a "name" or other convenient designation. Even a nonspecialist is quite likely to encounter most of these, either somewhere in the published literature in the form of an idea, definition, theorem, algorithm. . . or to hear about them, often in somewhat vague terms, and to feel the need for more information. In such a case, if the concept relates to algebra, then one should be able to find something in this handbook, at least enough to judge whether it is worth the trouble to try to find out more. In addition to the primary information, the numerous references to important articles, books, or lecture notes should help the reader find out as much as desired.

As a further tool, the index is perhaps more extensive than usual, and is definitely not limited to definitions, (famous) named theorems, and the like.

For the purposes of this handbook, "algebra" is more or less defined as the union of the following areas of the Mathematics Subject Classification Scheme:

- 20 (Group theory)
- 19 (K-theory this will be treated at an intermediate level; a separate Handbook of K-Theory, which goes into far more detail than the section planned for this Handbook of Algebra is under consideration)
- 18 (Category theory and homological algebra, including some of the uses of categories in computer science, often classified somewhere in Section 68)
- 17 (Nonassociative rings and algebras, especially Lie algebras)
- 16 (Associative rings and algebras)
- 15 (Linear and multilinear algebra. Matrix theory)
- 13 (Commutative rings and algebras; here, there is a fine line to tread between commutative algebras and algebraic geometry; algebraic geometry is definitely not a topic that will be dealt within any detail in this handbook; one day, there will, hopefully, be a separate handbook on that topic)
- 12 (Field theory and polynomials)

– 11 [Number theory, the part that also used to be classified under 12 (Algebraic number theory)]
– 08 (General algebraic systems)
– 06 (Order, lattices, ordered algebraic structures; certain parts; but not topics specific to Boolean algebras as there is a separate three-volume Handbook of Boolean Algebras)

Planning

Originally (1992), we expected to cover the whole field in a systematic way. Volume 1 would be devoted to what is now called Section 1 (see below), Volume 2 to Section 2, and so on. A quite detailed and comprehensive plan was made in terms of topics that needed to be covered and authors to be invited. That turned out to be an inefficient approach. Different authors have different priorities, and to wait for the last contribution to a volume, as planned originally, would have resulted in long delays. Instead, there is now a dynamic evolving plan. This also makes it possible to take new developments into account.

Chapters are still by invitation only according to the then current version of the plan, but the various chapters are published as they arrive, allowing for faster publication. Thus, in this volume 6 of the Handbook of Algebra, the reader will find contributions from five sections.

As the plan is dynamic, suggestions from users, both as to topics that could or should be covered, and authors are most welcome and will be given serious consideration by the board and editor.

The list of sections looks as follows:

Section 1: Linear algebra. Fields. Algebraic number theory
Section 2: Category theory. Homological and homotopical algebra. Methods from logic (algebraic model theory)
Section 3: Commutative and associative rings and algebras
Section 4: Other algebraic structures. Nonassociative rings and algebras. Commutative and associative rings and algebras with extra structure
Section 5: Groups and semigroups
Section 6: Representations and invariant theory
Section 7: Machine computation. Algorithms. Tables
Section 8: Applied algebra
Section 9: History of algebra

For the detailed plan (2008 version), the reader is referred to the Outline of the Series following this preface.

The individual chapters

It is not the intention that the handbook as a whole can also be a substitute undergraduate or even graduate, textbook. Indeed, the treatments of the various topics will

be too much dense and professional for that. Basically, the level should be graduate and up, and such material as can be found in P.M. Cohn's three-volume textbook "Algebra" (Wiley) should, as a rule, be assumed known. The most important function of the chapters in this handbook is to provide professional mathematicians working in a different area with a sufficiency of information on the topic in question if and when it is needed.

Each of the chapters combines some of the features of both a graduate-level textbook and a research-level survey. Not all the ingredients mentioned below will be appropriate in each case, but authors have been asked to include the following:

- Introduction (including motivation and historical remarks).
- Outline of the chapter.
- Basic concepts, definitions, and results (these may be accompanied by proofs or (usually better) ideas/sketches of the proofs when space permits).
- Comments on the relevance of the results, relations to other results, and applications.
- Review of the relevant literature; possibly complete with the opinions of the author on recent developments and future directions.
- Extensive bibliography (several hundred items will not be exceptional).

The present

Volume 1 appeared in 1995 (copyright 1996), Volume 2 in 2000, Volume 3 in 2003, Volume 4 in 2005 (copyright 2006), Volume 5 in 2007. Volume 7 is planned for 2009. Thereafter, we aim at one volume every 2 years (or better).

The future

Of course, ideally, a comprehensive series of books like this should be interactive and have a hypertext structure to make finding material and navigation through it immediate and intuitive. It should also incorporate the various algorithms in implemented form as well as permit a certain amount of dialog with the reader. Plans for such an interactive, hypertext, CDROM (DVD)-based (or web-based) version certainly exist, but the realization is still a nontrivial number of years in the future.

Kvoseliai, July 2008 Michiel Hazewinkel

Kaum nennt man die Dinge beim richtigen Namen
so verlieren sie ihren gefährlichen Zauber

(You have but to know an object by its proper name
for it to lose its dangerous magic)

Elias Canetti

Outline of the series

(as of July 2008)

Plilosophy and principles of the Handbook of Algebra

Compared to the outline in Volume 1, this version differs in several aspects.

First, there is a major shift in emphasis away from completeness as far as the more elementary material is concerned and toward more emphasis on recent developments and active areas. Second, the plan is now more dynamic in that there is no longer a fixed list of topics to be covered, determined long in advance. Instead, there is a more flexible nonrigid list that can and does change in response to new developments and availability of authors.

The new policy, which started with Volume 2, is to work with a dynamic list of topics that should be covered, to arrange these in sections and larger groups according to the major divisions into which algebra falls, and to publish collections of contributions (i.e. chapters) as they become available from the invited authors.

The coding below is by style and is as follows:

– **Author(s) in bold,** followed by chapter title: articles (chapters) that have been received and are published or are being published in this volume.
– *Chapter title in italic:* chapters that are being written.
– Chapter title in plain text: topics that should be covered but for which no author has yet been definitely contracted.

Chapters that are included in Volumes 1–6 have a (x; yy pp) after them, where "x" is the volume number and "yy" is the number of pages.

Compared to the plan that appeared in Volume 1, the section on "Representation and invariant theory" has been thoroughly revised.

Compared with the outline that appeared in Volume 4, section 4H (Rings and algebras with additional structure has been split into two parts: 4H: Hopf algebras and related structures; 4I: Other rings and algebras with additional structure. The old section 4I (Witt vectors) has been absorbed into the section on Hopf algebras. The other changes of this current version compared to the one in Volume 4 (2006) are relatively minor: mostly the addition of a number of topics.

Compared with the outline in Volume 5, there are only minor changes.

Planned publishing schedule (as of September 2007)

1996: Volume 1 (published)
2001: Volume 2 (published)
2003: Volume 3 (published)
2006: Volume 4 (published)
2007 (last quarter): Volume 5 (published)
2008: (last quarter) Volume 6
Further volumes, roughly, at the rate of one every year.

Contents of the Handbook of Algebra

Section 1: Linear algebra. Fields. Algebraic number theory
Section 2: Category theory. Homological and homotopical algebra
Section 3: Commutative and associative rings and algebras
Section 4: Other algebraic structures. Nonassociative rings and algebras. Commutative and associative rings and algebras with additional structure.
Section 5: Groups and semigroups
Section 6: Representations and invariant theory
Section 7: Machine computation. Algorithms. Tables
Section 8: Applied algebra
Section 9: History of algebra

More detailed plans for Sections 1–9 follow below. Very little in the way of detailed plans has been made so far for Sections 7–9.

Experience has shown that often 40–50 pages is a suitable length for a chapter, but both longer and shorter chapters have already occurred and authors are basically free in this matter.

It is not the intention that the handbook as a whole will also be a substitute undergraduate textbook. Basically, the level should be graduate and up, and such material as can be found in P.M. Cohn's three-volume textbook "Algebra" (Wiley) can as a rule be assumed known. A main function of the articles in the handbook is to provide professional mathematicians working in a different area with a sufficiency of information on the topic in question if and when he needs it.

In the following, there is a more detailed description of a number of sections in various volumes, which have more elementary sounding titles. These are partly intended as quick condensed surveys, establishing basic definitions and notations among other things but mainly as opportunities to discuss modern developments and new kinds of applications.

Mostly, each chapter should combine some of the features of a graduate–level textbook and a research-level survey. Not all the ingredients mentioned below will be appropriate in each case, but basically the authors should strive to include the following:

- Introduction (including motivation and historical remarks).
- Outline of the chapter.
- Basic concepts, definitions, and results [these can be accompanied by proofs or (usually better) ideas/sketches of the proofs if space permits].
- Comments on the relevance of the results, relations to other results, and applications.
- Review of the relevant literature; possibly complete with the opinions of the author on recent developments and future directions.
- Extensive bibliography (several hundred items will not be exceptional).

Often, it will be better to present a result in terms of its central core instead of the widest possible but technically involved generalization; the latter can be mentioned later, for instance in terms of an appropriate reference.

Section 1. Linear algebra. Fields. Algebraic number theory

A. Linear algebra

G.P. Egorychev, Van der Waerden conjecture and applications (1; 22 pp)
V.L. Girko, Random matrices (1; 52 pp)
Alexander E. Guterman, Matrix invariants over semirings (6; 35 pp)
J.A. Hermida-Alonso, Linear algebra over commutative rings (3; 49 pp)
A.N. Malyshev, Matrix equations. Factorization of matrices (1; 38 pp)
L. Rodman, Matrix functions (1; 38 pp)
Correction to the chapter by **L. Rodman,** Matrix functions (3; 1 p)
Kazimierz Szymiczek, Quadratic forms (6; 43 pp)
Linear inequalities
Orderings (*partial and total*) *on vectors and matrices*
Structured matrices
Matrices over the integers and other (*special*) *rings and fields*
Quasideterminants and determinants over noncommutative fields
Nonnegative matrices, positive definite matrices, and doubly nonnegative matrices
Generalized inverses of matrices
Matrix invariants over noncommutative rings

B. Linear (in)dependence

J.P.S. Kung, Matroids (1; 28 pp)

C. Algebras arising from vector spaces

Clifford algebras
Other algebras arising from vector spaces

D. Fields, Galois theory, and algebraic number theory

(There is also a chapter planned on ordered fields in Section 4I; see Section 4H for a chapter on Picard–Vessiot theory)

Toma Albu, From field theoretic to abstract co-Galois theory (5; 81 pp)
J.K. Deveney, J.N. Mordeson, Higher derivation Galois theory of inseparable field extensions (1; 34 pp)
I. Fesenko, Complete discrete valuation fields. Abelian local class field theories (1; 48 pp)
M. Jarden, Infinite Galois theory (1; 52 pp)
R. Lidl, H. Niederreiter, Finite fields and their applications (1; 44 pp)
W. Narkiewicz, Global class field theory (1; 30 pp)
H. van Tilborg, Finite fields and error correcting codes (1; 28 pp)
Skew fields and division rings. Brauer group
Topological and valued fields. Valuation theory
Zeta and L-functions of fields and related topics
Structure of Galois modules
Constructive Galois theory
Dessins d'enfants
Hopf–Galois theory

E. Non-Abelian class field theory and the Langlands program

(to be arranged in several chapters by Y. Ihara)

F. Generalizations of fields and related objects

U. Hebisch, H.J. Weinert, Semirings and semifields (1; 38 pp)
G. Pilz, Near rings and near fields (1; 36 pp)

Section 2. Category theory. Homological and homotopical algebra. Methods from logic

A. Category theory

S. Mac Lane, I. Moerdijk, Topos theory (1; 28 pp)
Ernie Manes, Monads of sets (3; 48 pp)
Martin Markl, Operad and PROPs (5; 54 pp)
Boris I. Plotkin, Algebra, categories, and databases (2; 71 pp)
Peter S. Scott, Some aspects of categories in computer science (2; 71 pp)
Ross Street, Categorical structures (1; 50 pp)
Algebraic structures in braided categories

B. Homological algebra. Cohomology. Cohomological methods in algebra. Homotopical algebra

Ronald Brown, Crossed complexes and higher homotopy groupoids as noncommutative tools for higher dimensional local-to-global problems (6; 32 pp)

Jon F. Carlson, The cohomology of groups (1; 30 pp)
A. Generalov, Relative homological algebra. Cohomology of categories, posets, and coalgebras (1; 28 pp)
A. Ya Helemskii, Homology for the algebras of analysis (2; 143 pp)
J.F. Jardine, Homotopy and homotopical algebra (1; 32 pp)
B. Keller, Derived categories and their uses (1; 32 pp)
Volodymyr Lyubashenko, Oleksandr Manchuk, *A(\infty)-algebras, -categories, and -functors* (5; 46 pp)
Boris V. Novikov, 0-cohomology of semigroups (5; 23 pp)
Galois cohomology (*H. Koch*)
Cohomology of commutative and associative algebras
Cohomology of Lie algebras
Cohomology of group schemes

C. Algebraic K-theory

Aderemi Kuku, Classical Algebraic K-theory: the functors K_0, K_1, K_2 (3; 55 pp)
Aderemi Kuku, *Algebraic K-theory: the higher K-functors* (4; 106 pp)
Grothendieck groups
K_2 and symbols
KK-theory and EXT
Hilbert C-modules*
Index theory for elliptic operators over C algebras*
Simplicial algebraic K-theory
Chern character in algebraic K-theory
Noncommutative differential geometry
K-theory of noncommutative rings
Algebraic L-theory
Cyclic cohomology
Asymptotic morphisms and E-theory
Hirzebruch formulae

D. Model theoretic algebra

(see also Paul C. Eklof, Whitehead modules, in Section 3B)
Mike Prest, Model theory for algebra (3; 31 pp)
Mike Prest, Model theory and modules (3; 34 pp)
Frank O. Wagner, Stable groups (2; 36 pp)
Logical properties of fields and applications
Recursive algebras
Logical properties of Boolean algebras
The A-E-K theorem and its relatives

E. Rings up to homotopy. Homotopic algebra

Rings up to homotopy
Simplicial algebras

Section 3. Commutative and associative rings and algebras

A. Commutative rings and algebras

(See also Carl Faith, Coherent rings and annihilator conditions in matrix and polynomial rings in Section 3B)
J.P. Lafon, Ideals and modules (1; 24 pp)
Dorin Popescu, Artin approximation (2; 45 pp)
Rafael H. Villareal, Monomial algebras and polyhedral geometry (3; 62 pp)
General theory. Radicals, prime ideals, etc. Local rings (general). Finiteness and chain conditions
Extensions. Galois theory of rings
Modules with quadratic form
Homological algebra and commutative rings. Ext, Tor, etc. Special properties (p.i.d., factorial, Gorenstein, Cohen–Macauley, Bezout, Fatou, Japanese, excellent, Ore, Prüfer, Dedekind, . . . and their interrelations)
Finite commutative rings and algebras (see also the chapter by A.A. Nechaev in Section 3B)
Localization. Local-global theory
Rings associated to combinatorial and partial order structures (straightening laws, Hodge algebras, shellability, . . .)
Witt rings of quadratic forms and real spectra

B. Associative rings and algebras

(see also "Freeness theorem for groups, rings, and Lie algebras" in Section 5A)
Victor Bavula, Filter dimension (4; 28 pp)
Paul M. Cohn, Polynomial and power series rings. Free algebras, firs, and semifirs (1; 30 pp)
Paul C. Eklof, Whitehead modules (3; 23 pp)
Edgar E. Enochs, Flat covers (3; 21 pp)
Alberto Facchini, The Krull–Schmidt theorem (3; 42 pp)
Carl Faith, Coherent rings and annihilation conditions in matrix and polynomial rings (3; 31 pp)
V.K. Kharchenko, Simple, prime, and semiprime rings (1; 52 pp)
V.K. Kharchenko, Fixed rings and noncommutative invariant theory (2; 27 pp)
T.Y. Lam, Hamilton's quaternions (3; 26 pp)
Aleksandr A. Necchaev, Finite rings with applications (5; 108 pp)
Sudarshan K. Sehgal, Group rings and algebras (3; 96 pp).
Askar A. Tuganbaev, Modules with distributive submodule lattice (2; 25 pp)
Askar A. Tuganbaev, Serial and distributive modules and rings (2; 25 pp)
Askar A. Tuganbaev, Modules with the exchange property and exchange rings (2; 25 pp)
Askar A. Tuganbaev, Max rings and V-rings (3; 23 pp)

Askar A. Tuganbaev, Semiregular, weakly regular, and πregular rings (3; 25 pp)
A. van den Essen, Algebraic microlocalization and modules with regular singularities over filtered rings (1; 28 pp)
F. Van Oystaeyen, Separable algebras (2; 66 pp)
K. Yamagata, Frobenius rings (1; 48 pp)
Classification of Artinian algebras and rings
General theory of associative rings and algebras
Rings of quotients. Noncommutative localization. Torsion theories
von Neumann regular rings
Lattices of submodules
PI rings
Generalized identities
Endomorphism rings, rings of linear transformations, matrix rings
Homological classification of (*noncommutative*) *rings*
Dimension theory
Duality. Morita-duality
Commutants of differential operators
Rings of differential operators
Graded and filtered rings and modules (also commutative)
Goldie's theorem, Noetherian rings, and related rings
Sheaves in ring theory
Koszul algebras
Hamiltonian algebras
Algebraic asymptotics
Antiautomorphisms, especially for simple central algebras

C. Coalgebras

(See also the chapter by Tomasz Brzezinski in Section 4H)
W. Michaelis, Coassociative coalgebras (3; 120 pp)
Co-Lie-algebras

D. Deformation theory of rings and algebras (including Lie algebras)

Yu Khakimdzanov, Varieties of Lie algebras (2; 35 pp)
Deformation theory of rings and algebras (*general*)
Deformation theoretic quantization

Section 4. Other algebraic structures. Nonassociative rings and algebras. Commutative and associative algebras with extra structure

A. Lattices and partially ordered sets

David Kruml, Jan Paseka, Algebraic and categorical aspects of quantales (5; 43 pp)
Ales Pultr, Frames (3; 63 pp)

B. Boolean algebras

C. Universal algebra and general algebraic concepts and constructions

Universal algebra (*general*)
Radicals

D. Varieties of algebras, groups . . .

(See also Yu Khakimdzanov, Varieties of Lie algebras in Section 3D)
V.A. Artamonov, Varieties of algebras (2; 30 pp)
V.A. Artamonov, Quasi-varieties (3; 19 pp)
Varieties of groups
Varieties of semigroups

E. Lie algebras

(See also "Freeness theorems for groups, rings, and Lie algebras" in Section 5A)
Yu A. Bahturin, M.V. Zaitsev, A.A. Mikhailov, Infinite-dimensional Lie superalgebras (2; 30 pp)
Michel Goze, Yusupdjan Khakimdjanov, Nilpotent and solvable Lie algebras (2; 51 pp)
A.T. Molev, Gel'fand–Tsetlin bases (4; 69 pp). Could also have been placed in Section 6D.
Christophe Reutenauer, Free Lie algebras (3; 21 pp)
General structure theory
Classification theory of semisimple Lie algebras over **R** and **C**
The exceptional Lie algebras
Universal enveloping algebras
Modular (ss) Lie algebras (including classification)
Infinite dimensional Lie algebras (general)
Kac–Moody Lie algebras
Finitary Lie algebras
Standard bases
Affine Lie algebras and Lie super algebras and their representations
Kostka polynomials

F. Jordan algebras (finite and infinite dimensional and including their cohomology theory)

G. Other nonassociative algebras (Malcev, alternative, Lie admissible . . .)

Mal'tsev algebras
Alternative algebras

H. Hopf algebras and related structure

(see also "Hopf–Galois theory" in Section 1D)
(see also "Co-Galois theory" in Section 1D)
(see also "Algebraic structures on braided categories" in Section 2A)

(see also "Representation theory of semisimple Hopf algebras" in Section 6D)
Katsutoshi Amano, Akaira Masuoka, Mitsuhiro Takeuchi, Hopf algebraic approach to Picard–Vessiot theory (6; 50 pp)
Gabriella Böhm, Hopf algebroids (6; 44 pp)
Tomasz Brzeziński, Comodules and corings (6; 72 pp)
Mia Cohen a.o., Hopf algebras (general) (4; 87 pp)
Michiel Hazewinkel, Witt vectors. Part 1 (6; 148 pp)
Jae-Hoon Kwon, Crystal graphs and the combinatorics of Young tableaux (6; 32 pp)
Dominiue Manchon, Hopf algebras in renormalization (5; 67 pp)
Akira Masuoka, Classification of semisimple Hopf algebras (5; 29 pp)
A.I. Molev, Yangians and their applications (3; 54 pp)
Frédéric Patras, Lambda-rings (3; 34 pp)
Classification of pointed Hopf algebras
Quantum groups (*general*)
Formal groups
p-divisible groups
Combinatorial Hopf algebras
Symmetric functions
Special functions and q-special functions, 1–2 variable case
Quantum groups and multiparameter q-special functions
Noncommutative geometry à la Connes
Noncommutative geometry from the algebraic point of view
Noncommutative geometry, categorical point of view
Solomon descent algebras
Hopf algebras and operads
Noncommutative symmetric functions and quasisymmetric functions
Witt vectors. Part 2
Trees, dendriform algebras, and dialgebras
Quantum differential geometry, quantum calculus, quantum group approach to noncommutative geometry
Connes–Baum theory

I. Other rings and algebras with additional structure

Alexander Levin, Difference algebra (4; 100 pp)
Graded and super algebras (commutative, associative; for Lie superalgebras, see Section 4E)
Topological rings
Ordered and lattice-ordered rings and algebras.
Rings and algebras with involution. C*-algebras
Differential algebra
Baxter algebras
Ordered fields

Hypergroups
Stratified algebras

Section 5. Groups and semigroups

A. Groups

(see also "Groups and semigroups of automata transformations" in Section 5B)
Laurent Bartholdi, Rostislav I. Grigorchuk, Zoran Sunik, Branch groups (3; 129 pp)
Meinolf Geck, Gunter Malle, Reflection groups. Coxeter groups (4; 38 pp)
A.V. Mikhalev, A.P. Mishina, Infinite Abelian groups: methods and results (2; 48 pp)
V.I. Senashov, Groups with finiteness conditions (4; 27 pp)
M.C. Tamburini. E.V. semirnov, Hurwitz groups and Hurwitz generation (4; 28 pp)
V.V. Vershinin, Braid groups (4; 22 pp)
Simple groups, sporadic groups
Representations of the finite simple groups
Diagram methods in group theory (*Presentations of groups*)
Abstract (finite) groups. Structure theory. Special subgroups. Extensions and decompositions
Solvable (soluble) groups, nilpotent groups, p-groups. Fitting classes, Schunck classes. Thompson's list of minimal nonsolvable groups to be included
Infinite soluble groups
Word and conjugacy problems
Burnside problem
Combinatorial group theory
Distance transitive graphs and their groups
Free groups (including actions on trees)
Formations
Infinite groups. Local properties
The infinite dimensional classical groups
Infinite simple groups
Algebraic groups. The classical groups. Chevalley groups
Chevalley groups over rings
Other groups of matrices. Discrete subgroups
Groups with BN-pair, Tits buildings ...
Groups and (finite combinatorial) geometry
"Additive" group theory
Probabilistic techniques and results in group theory
Self-similar groups
Frobenius group
Just-infinite groups
Automorphism groups of groups
Automorphism groups of algebras and rings

Freeness theorems (*in groups and rings and Lie algebras*)
Groups with perscribed systems of subgroups
Automatic groups
Groups with minimality and maximality conditions (School of N.S. Chernikov)
Lattice-ordered groups
Linearly and totally ordered groups
Finitary groups
Random groups
Hyperbolic groups
Infinite dimensional groups

B. Semigroups

(See also the chapter by Boris Novikov in Section 2B)
Semigroup theory. Ideals, radicals, structure theory
Semigroups and automata theory and linguistics
Groups and semigroups of automata transformations

C. Algebraic formal language theory. Combinatorics of words
D. Loops, quasigroups, heaps. . .

Quasigroups in combinatorics

E. Combinatorial group theory and topology

(See also: "Diagram methods in group theory" in Section 5A)

Section 6. Representation and invariant theory

A. Representation theory. General

Representation theory of rings, groups, algebras (general)
Modular representation theory (general)
Representations of Lie groups and Lie algebras. General
Multiplicity free representations

B. Representation theory of finite and discrete groups and algebras

Representation theory of finite groups in characteristic zero
Modular representation theory of finite groups. Blocks
Representation theory of the symmetric groups (both in characteristic zero and modular)
Representation theory of the finite Chevalley groups (both in characteristic zero and modular)
Modular representation theory of Lie algebras

C. Representation theory of "continuous groups" (linear algebraic groups, Lie groups, loop groups. . .) and the corresponding algebras

(see also Gel'fand–Tsetlin bases for classical Lie algebras in Section 4E by A.I. Molev)

A.U. Klimyk, Infinite dimensional representations of quantum algebras (2; 26 pp)

Representation theory of compact topolgical groups

Representation theory of locally compact topological groups

Representation theory of SL2(R). . .

Representation theory of the classical groups. Classical invariant theory

Classical and transcendental invariant theory

Reductive groups and their representation theory

Unitary representation theory of Lie groups

Finite dimensional representation theory of the ss Lie algebras (in characteristic zero); structure theory of semisimple Lie algebras

Infinite dimensional representation theory of ss Lie algebras. Verma modules

Representation of Lie algebras. Analytic methods

Representations of solvable and nilpotent Lie algebras. The Kirillov orbit method

Orbit method, Dixmier map. . . for ss Lie algebras

Representation theory of the exceptional Lie groups and Lie algebras

Representation theory of "classical" quantum groups

Duality in representation theory

Representation theory of loop groups and higher dimensional analogues, gauge groups, and current algebras

Representation theory of Kac–Moody algebras

Invariants of nonlinear representations of Lie groups

Representation theory of infinite dimensional groups like GL_∞

Metaplectic representation theory

D. Representation theory of algebras

Xueqing Chen, Tomás Pospíchal, Ki-Bong Nam, Quivers and representations (6; 53 pp)

Representations of rings and algebras by sections of sheafs

Representation theory of algebras (Quivers, Auslander–Reiten sequences, almost split sequences. . .)

Tame algebras

Ringel–Hall algebras

Composition algebras

Quasi-hereditary algebras

Cellular algebras

Representation theory of (semisimple) Hopf algebras

E. Abstract and functorial representation theory

Serge Bouc, Burnside rings (2; 60 pp)

P. Webb, A guide to Mackey functors (2; 32 pp)
Abstract representation theory

F. Representation theory and combinatorics
G. Representations of semigroups

Representation of discrete semigroups
Representations of Lie semigroups

H. Hecke algebras

Hecke algebras

I. Invariant theory

Section 7. Machine computation. Algorithms. Tables. Counting algebraic structures

Some notes on this volume: Besides some general article(s) on machine computation in algebra, this volume should contain specific articles on the computational aspects of the various larger topics occurring in the main volumes, as well as the basic corresponding tables. There should also be a general survey on the various available symbolic algebra computation packages

Georgy P. Egorychev, Eugene V. Zima, Integral representation and algorithms for closed form summation (5; 75 pp)
The CoCoA computer algebra system
Groebner bases and their applications
Software for automatic groups

Section 8. Applied Algebra

Marlos Viana, Canonicak decompositions and invariants for data analysis (6; 16 pp)

Section 9. History of algebra

(see also K.T. Lam, Hamilton's quaternions, in Section 3A)
History of coalgebras and Hopf algebras
Development of algebra in the 19th century

List of Contributors

Alexander E. Guterman, Faculty of Algebra, Department of Mathematics and Mechanics, Moscow State University, Moscow, GSP-2, Russia, *e-mail: guterman@list.ru*

Kazimierz Szymiczek, Department of Mathematics, University of Silesia, Katowice, Poland, *e-mail: szymiczek@math.us.edu.pl*

Ronald Brown, School of Computer Science, University of Wales, Dean St. Bangor, Gwynedd LL57 1UT, UK, *e-mail: ronnie.profbrown@btinternet.com*.

Katsutoshi Amano, 8-29 Ida-Sammai-cho, Kawasaki, Kanagawa 211-0037, Japan, *e-mail: ma-pfybdb-61201@agate.ne.jp*

Akaira Masuoka, Department of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan, *e-mail: akira@math.tsukuba.ac.jp*

Mitsuhiro Takeuchi, Department of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan, *e-mail: takeu@math.tsukuba.ac.jp*

Gabriella Böhm, Research Institute for Particle and Nuclear Physics Budapest, POB 49, H-1525 Budapest, Hungary, *e-mail: G.Bohm@rmki.kfki.hu*

Tomasz Brzezinski, Department of Mathematics, Swansea University, Singleton Park, Swansea SA2 8PP, UK, *e-mail: T.Brzezinski@swansea.ac.uk*

Michiel Hazewinkel, Burg. s'Jacob iaan 18, 1401BR Bussum, The Netherlands, *e-mail: michiel.hazewinkel@xs4all.nl*

Jae-Hoon Kwon, Department of Mathematics, University of Seoul, 90 Cheonnong Dong, Dongdaemungu, Seoul 130-743, Korea, *e-mail: jhkwon@uos.ac.kr*

Xueqing Chen, Department of Mathematics and Computer Sciences, University of Wisconsin-Whitewater, 800 West Main Street, Whitewater, WI 53190, USA, *e-mail: chenx@uww.edu*

Tomáš Pospíchal, Department of Mathematics and Statistics, University of Windsor, Windsor, Ontario N9B 3P4, Canada, *e-mail: tpospich@uwindsor.ca*

Ki-Bong Nam, Department of Mathematics and Computer Sciences, University of Wisconsin-Whitewater, 800 West Main Street, Whitewater, WI 53190, USA, *e-mail: namk@uww.edu*

Marlos Viana, University of Illinois at Chicago Eye Center, UIC Eye Center, 1855 West Taylor Street (m/c648), Chicago, IL 60612, USA, *e-mail: viana@gmail.com, viana@uic.edu*

Matrix Invariants over Semirings*

Alexander E. Guterman

Faculty of Algebra, Department of Mathematics and Mechanics,
Moscow State University, Moscow, 119992, GSP-2, Russia
E-mail: guterman@list.ru

Contents

*2000 Mathematics Subject Classification: 15A45, 15A33, 15A03.
The work is partially supported by RFBR grant N. 08-01-00693

HANDBOOK OF ALGEBRA, VOL. 6
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(08)00201-5

1. Introduction

In the last decades, matrices with entries from various semirings have attracted attention of many researchers working in both theoretical and applied mathematics. The concept of a semiring is defined as follows:

DEFINITION 1.1. A *semiring* is a set $\mathcal{S}$ with two binary operations, addition and multiplication, such that

- $\mathcal{S}$ is an Abelian monoid under addition (identity denoted by 0);
- $\mathcal{S}$ is a semigroup under multiplication (identity, if any, denoted by 1);
- multiplication is distributive over addition on both sides;
- $s0 = 0s = 0$ for all $s \in \mathcal{S}$.

In this chapter, we always assume that there is a multiplicative identity 1 in $\mathcal{S}$, which is different from 0.

Briefly, a semiring differs from a ring by the fact that not every element is required to have the additive inverse. The most common examples of semirings that are not rings are the non-negative integers $\mathbb{Z}_+$, the non-negative rationals $\mathbb{Q}_+$, and the non-negative reals $\mathbb{R}_+$, with the usual addition and multiplication. There are classical examples of non-numerical semirings as well. Probably, the first such example appeared in the work of Dedekind [1] in connection with the algebra of ideals of a commutative ring (one can add and multiply ideals, but it is not possible to subtract them). Later, Vandiver [2] put forward the concept of semirings as the best algebraic structure, which includes both rings and bounded distributive lattices. Other important examples of semirings are Boolean algebras, max-algebras, tropical semirings, and fuzzy scalars. See the monographs [3–5] for more details.

It should be noted that the majority of examples of semirings arise in various applications of algebra. Among them, there are automata theory [6, 7], optimization theory [8, 9], combinatorial optimization [10–12], optimal control [13, 14], discrete event systems [15], operations research [13], ergodic control [16, 17], mathematical economics [16], assignment problem [11, 18], graph theory [11, 19], and algebraic geometry [20–22]. For a detailed index of applications, one may consult [4, pp. 355–356].

DEFINITION 1.2. A semiring $\mathcal{S}$ is called *commutative* if the multiplication in $\mathcal{S}$ is commutative.

DEFINITION 1.3. A semiring $\mathcal{S}$ is called *antinegative* (or *zero-sum-free*) if $a + b = 0$ implies that $a = b = 0$.

This means that the zero element is the only element with an additive inverse.

DEFINITION 1.4. We say that a semiring $\mathcal{S}$ has no *zero divisors* if from $ab = 0$ in $\mathcal{S}$ it follows that $a = 0$ or $b = 0$ (or both).

DEFINITION 1.5. A semiring $\mathcal{S}$ is called *Boolean* if $\mathcal{S}$ is isomorphic to a suitable set of subsets of a given set M, the sum of two subsets being their union, and the product being their intersection. The zero element is the empty set and the identity element is the whole set M.

It is straightforward to see that Boolean semirings are commutative and antinegative. If a Boolean semiring $\mathcal{S}$ consists of only two subsets of M: the empty subset and M, then it is called a binary Boolean semiring (or $\{0, 1\}$-semiring) and is denoted by $\mathbf{B}$. Note that the binary Boolean semiring does not have zero divisors.

DEFINITION 1.6. A semiring is called a *chain* semiring if the set $\mathcal{S}$ is totally ordered with universal lower and upper bounds, and the operations are defined by $a + b = \max\{a, b\}$ and $a \cdot b = \min\{a, b\}$.

It is straightforward to see that any chain semiring $\mathcal{S}$ is a Boolean semiring on a set of appropriate subsets of $\mathcal{S}$. Namely, let us consider the set M of all elements in $\mathcal{S}$. We choose all those subsets of M that consist of all elements of $\mathcal{S}$, which are strictly lower than a given element. Chain semirings are examples of Boolean semirings without zero divisors.

DEFINITION 1.7. A semiring is called a *max-algebra* if the set $\mathcal{S}$ is an ordered group with a multiplication $*$ and an order relation $\leq$, and operations in $\mathcal{S}$ are defined as follows: $a + b = \max\{a, b\}$, $a \cdot b = a * b$, for any $a, b \in \mathcal{S}$.

The most common example of max-algebras is $\mathbb{R}_{\max} := (\mathbb{R} \cup \{-\infty\}, \leq, +)$.

In a similar way, *min-algebras* can be defined, with the only difference being $a + b = \min\{a, b\}$. These algebras are sometimes called *tropical algebras*.

2. Matrices and determinants

Let $\mathcal{M}_{m,n}(\mathcal{S})$ denote the set of $m \times n$ matrices with entries from the semiring $\mathcal{S}$, $\mathcal{M}_n(\mathcal{S}) = \mathcal{M}_{n,n}(\mathcal{S})$. Under the natural (and usual) definitions of matrix addition and multiplication, $\mathcal{M}_n(\mathcal{S})$ is obviously a semiring. Matrix theory over semirings has been an object of intensive study during the last decades; see for example the monographs [3, 23] and references therein. For more recent results on this subject, see [20, 24–30].

The matrix I_n is the $n \times n$ identity matrix, $J_{m,n}$ is the $m \times n$ matrix with all elements equal to 1, and $O_{m,n}$ is the $m \times n$ zero matrix. We omit the subscripts when the order is obvious from the context and we write I, J, and O. The matrix $E_{i,j}$ denotes the matrix with 1 at the (i, j)th position and zeros elsewhere, $C_i = \sum_{j=1}^{m} E_{i,j}$ is the ith column matrix, and $R_j = \sum_{i=1}^{n} E_{i,j}$ is the jth row matrix. For $A \in \mathcal{M}_{m,n}(\mathcal{S})$

and $B \in \mathcal{M}_{m,k}(\mathcal{S})$, we denote by $(A|B) \in \mathcal{M}_{m,n+k}(\mathcal{S})$ the matrix whose rows are obtained by the union of columns of A and B.

The development of linear algebra over semirings certainly requires an analog of the determinant function (see [4]). However, it turns out that even over commutative semirings without zero divisors, the classical determinant cannot be defined as over fields and commutative rings. The main problem lies in the fact that, in semirings that are not rings not all elements possess an additive inverse. A natural replacement of the determinant function for matrices over commutative semirings is the bideterminant known since 1972 (see [31] and also [4, 32–34]). The bideterminant is useful for the solution of various pure algebraic problems (see [34]). The bideterminant is important for applications as well, for example, to solve systems of linear equations (see [32,35]), and in connections with graph theory (see [31]). For example, in [36], on the basis of the properties of max-algebraic determinants, it is shown that, the problem of the verification if a matrix is sign-nonsingular is polynomially equivalent to the problem of deciding whether the digraph of the corresponding matrix contains a cycle of even length. See also monographs [4, 5, 35] for more detailed and self-contained information.

Definition 2.1. [4, Chapter 19] The *bideterminant* of a matrix $A = [a_{i,j}] \in \mathcal{M}_n(\mathcal{S})$ is the pair $(\|A\|^+, \|A\|^-)$, where

$$\|A\|^+ = \sum_{\sigma \in A_n} a_{1,\sigma(1)} \cdots a_{n,\sigma(n)}, \ \|A\|^- = \sum_{\sigma \in S_n \setminus A_n} a_{1,\sigma(1)} \cdots a_{n,\sigma(n)}.$$

Here, S_n denotes the symmetric group on the set $\{1, \ldots, n\}$ and A_n denotes its subgroup of even permutations.

It is known that the bideterminant function possesses some natural properties. Namely, it is invariant under transposition, and for any scalar $\alpha \in \mathcal{S}$, $(\|\alpha A\|^+, \|\alpha A\|^-) = (\alpha^n \|A\|^+, \alpha^n \|A\|^-)$. However, some basic properties of the determinant are no longer true for the bideterminant. For example, if A is invertible, then $\|A\|^+ \neq \|A\|^-$, but the converse is not always true.

Example 2.2. [32] Let us consider $A = E_{1,1} + 2E_{1,2} + 3E_{2,1} + 4E_{2,2} \in M_2(\mathcal{S})$, where $\mathcal{S} = (\mathbb{Q}_+, \max, \cdot)$, i.e. the set of non-negative rationals with the standard multiplication and the addition defined by $a + b = \max\{a, b\}$. Then $(\|A\|^+, \|A\|^-) = (4, 6)$ but A is not invertible.

The bideterminant is not multiplicative in general. However, the following weaker version of this property is true.

Proposition 2.3. [30] *The following equality holds for all* $A, B \in \mathcal{M}_n(\mathcal{S})$

$$\|AB\|^+ + \|A\|^+\|B\|^- + \|A\|^-\|B\|^+ = \|AB\|^- + \|A\|^+\|B\|^+ + \|A\|^-\|B\|^-,$$

and, more generally, for $A_1, \ldots, A_s \in \mathcal{M}_n(\mathcal{S})$, it holds that

$$\|A_1,\ A_2 \cdots A_s\|^+ + \sum_{\substack{t_1,\ldots,t_s=\pm \\ t_1 \cdots t_s=-}} \|A_1\|^{t_1}\|A_2\|^{t_2}\cdots\|A_s\|^{t_s}$$
$$= \|A_1 A_2 \cdots A_s\|^- + \sum_{\substack{t_1,\ldots,t_s=\pm \\ t_1 \cdots t_s=+}} \|A_1\|^{t_1}\|A_2\|^{t_2}\cdots\|A_s\|^{t_s}$$

It also appears that the following function can be used instead of the determinant for matrices over semirings.

DEFINITION 2.4. The *permanent* of a matrix $A = [a_{i,j}] \in \mathcal{M}_n(\mathcal{S})$ is

$$\mathrm{per}\,(\mathrm{A}) = \sum_{\sigma \in \mathrm{S_n}} \mathrm{a}_{1,\sigma(1)} \cdots \mathrm{a}_{\mathrm{n},\sigma(\mathrm{n})}.$$

Also, if $\mathcal{S}$ is a subsemiring of a certain commutative ring $\mathcal{R}$, then for $A \in \mathcal{M}_n(\mathcal{S})$, the determinant $\det(A) = \sum_{\sigma \in S_n} (-1)^\sigma a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} = \|A\|^+ - \|A\|^-$ is well defined in $\mathcal{R}$.

In the following, we will concentrate on the recognition of different matrix properties via the determinant function.

2.1. *Singularity and determinant*

Let $\mathcal{R}$ be an associative ring. A matrix, $A \in \mathcal{M}_{m,n}(\mathcal{R})$, with entries from $\mathcal{R}$ is called *right singular* if there exists a nonzero element $\mathbf{x} \in \mathcal{R}^n$ such that $A\mathbf{x} = \mathbf{0}$, and A is called *left singular* if there exists a nonzero element $\mathbf{y} \in \mathcal{R}^m$ such that $\mathbf{y}^t A = \mathbf{0}$.

It is proved in [37, Theorem 9.1] that both these definitions coincide for matrices over commutative rings and for square matrices are equivalent to the statement that $\det A$ is a zero divisor in $\mathcal{R}$, where $\det A$ is the usual determinant of a square matrix over a commutative ring.

The notion of singularity for matrices over semirings can be defined in a similar way:

DEFINITION 2.5. A matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is *$\mathcal{S}$-right singular* if $A\mathbf{x} = \mathbf{0}$ for some nonzero $\mathbf{x} \in \mathcal{S}^n$.

DEFINITION 2.6. A matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is *$\mathcal{S}$-left singular* if $\mathbf{x}^t A = \mathbf{0}^t$ for some nonzero $\mathbf{x} \in \mathcal{S}^m$.

DEFINITION 2.7. A matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is *$\mathcal{S}$-singular* if A is either $\mathcal{S}$-left singular or $\mathcal{S}$-right singular.

The next example shows that even over antinegative commutative semirings without zero divisors, there exist matrices that are $\mathcal{S}$-left singular and are not $\mathcal{S}$-right singular or vice versa.

EXAMPLE 2.8. [24] Let $A = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \in \mathcal{M}_2(\mathbb{Z}_+)$. We have that $A\mathbf{x} = \mathbf{0}$ forces $\mathbf{x} = \mathbf{0}$ since $\mathbb{Z}_+$ is an antinegative semiring, but $[1, 0]A = [0, 0]$. Similarly $\mathbf{x}^t B = \mathbf{0}$ forces $\mathbf{x} = \mathbf{0}$ while $B[0, 1]^t = [0, 0]^t$.

DEFINITION 2.9. A matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is *$\mathcal{S}$-nonsingular* if A is neither $\mathcal{S}$-left singular nor $\mathcal{S}$-right singular.

LEMMA 2.10. [24, Lemma 3.8] *Let $\mathcal{S}$ be an antinegative semiring. Then the following conditions are equivalent for any matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$:*

1. *A is $\mathcal{S}$-singular.*
2. *A has a zero row or a zero column.*

Note that if $\mathcal{S}$ is commutative and A is an $\mathcal{S}$-singular square matrix, then $(\|A\|^+, \|A\|^-) = (0, 0)$. However, the following example shows that there are $\mathcal{S}$-nonsingular matrices with bideterminant equal to $(0, 0)$.

EXAMPLE 2.11. [24] Over any commutative antinegative semiring,

$$\left\| \begin{matrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \right\|^+ = 0 = \left\| \begin{matrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \right\|^- .$$

However, if the semiring $\mathcal{S}$ is also a subsemiring of an associative ring $\mathcal{R}$ without zero divisors, we can consider the following notion of singularity as well.

DEFINITION 2.12. We say that a matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is *$\mathcal{R}$-right singular* if $A\mathbf{x} = \mathbf{0}$ for some nonzero $\mathbf{x} \in \mathcal{R}^n$.

DEFINITION 2.13. We say that a matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is *$\mathcal{R}$-left singular* if $\mathbf{x}^t A = \mathbf{0}$ for some nonzero $\mathbf{x} \in \mathcal{R}^m$.

It is straightforward to see that if a semiring $\mathcal{S}$ is a subsemiring of a certain ring $\mathcal{R}$, then $\mathcal{R}$-right (left) singularity follows from $\mathcal{S}$-right (left) singularity. However, the following example shows that there are $\mathcal{S}$-nonsingular matrices, which are $\mathcal{R}$-singular.

EXAMPLE 2.14. For any n, the matrix $J_n = \sum_{i,j=1}^{n} E_{i,j} \in \mathcal{M}_n(\mathbb{Z}_+)$ is $\mathbb{Z}$-left singular and $\mathbb{Z}$-right singular but $\mathbb{Z}_+$-nonsingular.

DEFINITION 2.15. We say that a matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is *$\mathcal{R}$-singular* if A is *$\mathcal{R}$-left singular* or *$\mathcal{R}$-right singular*.

DEFINITION 2.16. A matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is *$\mathcal{R}$-nonsingular* if A is not $\mathcal{R}$-singular.

Note that similarly to the situation over fields, all nonsquare matrices are $\mathcal{R}$-singular.

Assume in addition that $\mathcal{R}$ is commutative. Then $\det(A) \in \mathcal{R}$ is well defined and the following analog of [37, Theorem 9.1] can be obtained.

LEMMA 2.17. [24, Lemma 5.3] *If $\mathcal{R}$ is an associative commutative ring then for a matrix $A \in \mathcal{M}_n(\mathcal{S})$ the following conditions are equivalent:*

1. *A is $\mathcal{R}$-right singular;*
2. *A is $\mathcal{R}$-left singular;*
3. $\det A = 0$.

DEFINITION 2.18. An element g from a certain multiplicative system $\mathcal{G}$ with identity element $1_{\mathcal{G}}$ is *invertible* (resp., *left invertible, right invertible*) if there is an element $f \in \mathcal{G}$ such that $fg = gf = 1_{\mathcal{G}}$ (resp., $fg = 1_{\mathcal{G}}$, $gf = 1_{\mathcal{G}}$).

DEFINITION 2.19. A matrix $A \in \mathcal{M}_n(\mathcal{S})$ is called *monomial* if it has exactly one nonzero element in each row and column.

It follows from [38] that a matrix over an antinegative semiring is invertible (resp., left invertible, right invertible) if and only if it is a monomial matrix such that all its nonzero elements are invertible (resp., left invertible, right invertible).

For matrices over max-algebras, the following notion of singularity is often in use (see [28] for the details).

DEFINITION 2.20. Let $\mathcal{S}$ be a max-algebra (operations are denoted by max and $+$). A matrix $A = [a_{ij}] \in M_n(\mathcal{S})$ is said to be *tropically singular* if the maximum in the expression for the permanent

$$\operatorname{per}(A) = \max_{\sigma \in S_n} \{a_{1\sigma(1)} + \cdots + a_{n\sigma(n)}\}$$

is achieved at least twice.

This can be generalized to the case of an arbitrary antinegative semiring $\mathcal{S}$ in the following way.

DEFINITION 2.21. A matrix $A = [a_{ij}] \in M_n(\mathcal{S})$ is said to be *tropically singular* if there exists a subset $\mathcal{T} \in S_n$ such that

$$\sum_{\sigma\in\mathcal{T}} a_{1\sigma(1)}\cdots a_{n\sigma(n)} = \sum_{\sigma\in S_n\setminus\mathcal{T}} a_{1\sigma(1)}\cdots a_{n\sigma(n)}.$$

We use these concepts in Section 4.1 to deal with the notion of rank.

3. Semimodules: bases and dimension

Semimodules over semiring are analogs of vector spaces over fields and modules over rings. These are discussed more precisely in the following.

DEFINITION 3.1. Given a semiring $\mathcal{S}$, we define a *left semimodule*, U, over $\mathcal{S}$ to be an abelian monoid with identity element, $\mathbf{0} \in U$. In addition, U is equipped with a function,

$$\begin{array}{ccc} \mathcal{S}\times U & \to & U \\ (s,\mathbf{u}) & \to & s\mathbf{u}, \end{array}$$

called *scalar multiplication* such that for all $\mathbf{u}$ and $\mathbf{v}$ in U and $r, s \in \mathcal{S}$ we have

1. $(sr)\mathbf{u} = s(r\mathbf{u})$,
2. $(s+r)\mathbf{u} = s\mathbf{u} + r\mathbf{u}$,
3. $s(\mathbf{u}+\mathbf{v}) = s\mathbf{u} + s\mathbf{v}$,
4. $1\mathbf{u} = \mathbf{u}$,
5. $s\mathbf{0} = \mathbf{0} = 0\mathbf{u}$.

In a similar way, a *right semimodule* can be defined.

In particular, under the natural definitions of matrix addition and multiplication with scalars, $\mathcal{M}_{m,n}(\mathcal{S})$ is obviously a semimodule over $\mathcal{S}$.

DEFINITION 3.2. An element $\mathbf{u}$ in a semimodule U is called a *left* (resp., *right*) *linear combination* of elements from a certain subset $P \subseteq U$ if there exists $k \in \mathbb{N}$, $s_1, \ldots, s_k \in \mathcal{S}$, $\mathbf{u}_1, \ldots, \mathbf{u}_k \in P$ such that $\mathbf{u} = \sum_{i=1}^{k} s_i\mathbf{u}_i \left(\text{resp., } \mathbf{u} = \sum_{i=1}^{k} \mathbf{u}_i s_i\right)$. In this case, $\sum_{i=1}^{k} s_i\mathbf{u}_i \left(\sum_{i=1}^{k} \mathbf{u}_i s_i\right)$ is called a *left (right) linear combination* of the elements $\mathbf{u}_1, \ldots, \mathbf{u}_k$ from P with coefficients $s_1, \ldots, s_k$ from $\mathcal{S}$.

DEFINITION 3.3. A *left* (*right*) *linear span*, $\langle P\rangle_{\mathcal{S}}$, of the set P is the set of all left (right) linear combinations of elements from P with coefficients from $\mathcal{S}$. We say that the set P *generates* a subset $V \subseteq U$ if $V \subseteq \langle P\rangle_{\mathcal{S}}$. Note that by definition, all linear combinations, we consider, are finite and nonempty. We will not specially point out the semiring of the coefficients if it will be clear from the context.

As over fields and rings, we denote the linear span of the set

$$\{[1,0,\dots,0]^t,[0,1,0,\dots,0]^t,\dots,[0,\dots,0,1]^t\}$$

by $\mathcal{S}^n$; here, A^t denotes the transposed matrix to the matrix A.

Note that, in contrast with vector spaces over fields, there are several ways to define a notion of independence for semimodules. For example, in [27,37,39], the following definition is used.

DEFINITION 3.4. A set of elements, P, from a semimodule U over a semiring $\mathcal{S}$ is called *left* (*right*) *linearly independent* if there is no element in P that can be expressed as a left (right) linear combination of the other elements of P with the coefficients from $\mathcal{S}$. The set P is called *linearly dependent* if it is not linearly independent.

However, in [16, 19, 32], the following definition is used.

DEFINITION 3.5. A system of elements, $\mathbf{u}_1,\dots,\mathbf{u}_k$, in a semimodule U is *left* (resp., *right*) *linearly dependent in the Gondran–Minoux sense* if there exist two subsets $I, J \subseteq K := \{1,\dots,k\}$, $I \cap J = \emptyset$, $I \cup J = K$, and scalars $\alpha_1,\dots,\alpha_k \in \mathcal{S}$, not all equal to 0, such that $\sum_{i\in I}\alpha_i\mathbf{u}_i = \sum_{j\in J}\alpha_j\mathbf{u}_j$ $\left(\text{resp.,}\ \sum_{i\in I}\mathbf{u}_i\alpha_i = \sum_{j\in J}\mathbf{u}_j\alpha_j\right)$.

REMARK 3.6. For any semiring, this definition is stronger than the first one. Namely, any system of elements, which is independent in the sense of Definition 3.5, is evidently independent in the sense of Definition 3.4, but the converse is not true as the following example shows.

EXAMPLE 3.7. It is shown in [40, Lemma 3.11] that the system

$$\left\{\mathbf{u}_1=\begin{pmatrix}1\\0\\0\end{pmatrix},\mathbf{u}_2=\begin{pmatrix}0\\1\\0\end{pmatrix},\mathbf{u}_3=\begin{pmatrix}1\\0\\1\end{pmatrix},\mathbf{u}_4=\begin{pmatrix}0\\1\\1\end{pmatrix}\right\}$$

is linearly independent in the sense of Definition 3.4 over any antinegative semiring. However, $\mathbf{u}_1+\mathbf{u}_4=\mathbf{u}_2+\mathbf{u}_3$, that is, this system is linearly dependent in the sense of Definition 3.5.

In [41], the following definition of linear dependence is introduced:

DEFINITION 3.8. A system of elements, $\mathbf{u}_1,\dots,\mathbf{u}_k$, $\mathbf{u}_i=[u_i^1,\dots,u_i^n]^t$, $i=1,\dots,k$, in a semimodule U is *left* (resp., *right*) *linearly dependent in the sense of Izhakian* if there exist two series of subsets $I_l, J_l \subseteq K := \{1,\dots,k\}$, $I_l \cap J_l = \emptyset$, $I_l \cup J_l = K$, $l=1,\dots,n$, and scalars $\alpha_1,\dots,\alpha_k \in \mathbb{R}$, not all equal to 0, such that $\sum_{i\in I_l}\alpha_i u_i^l = \sum_{j\in J_l}\alpha_j u_j^l$ $\left(\text{resp.,}\ \sum_{i\in I_l}u_i^l\alpha_i = \sum_{j\in J_l}u_j^l\alpha_j\right)$ for all l, $1\le l\le n$. A system is called *linearly independent* if it is not linearly dependent.

REMARK 3.9. Definition 3.8 is stronger than Definition 3.5. Namely, it is straightforward to see that a system of elements, which is independent in the sense of Definition 3.8, is also independent in the sense of Definition 3.5; however, the converse is not always true as it is shown by the following example.

EXAMPLE 3.10. Let $\mathcal{S}$ be an arbitrary antinegative semiring. Let us consider the system of vectors $[0, 1, 1]^t$, $[1, 0, 1]^t$, $[1, 1, 0]^t$ with the entries in $\mathcal{S}$. By the antinegativity, it is easy to see that these vectors are linearly independent in the sense of Definitions 3.4 and 3.5. However, taking $I_1 = \{1, 2\}$, $J_1 = \{3\}$, $I_2 = \{1, 2\}$, $J_2 = \{3\}$, $I_3 = \{1\}$, $J_3 = \{2, 3\}$, and the coefficients $(1, 1, 1)$, we see that these vectors are linearly dependent in the sense of Definition 3.8.

As in the case of vector spaces over fields, we can introduce the following definition.

DEFINITION 3.11. A collection, B, of elements is called a *left* (*right*) *basis* of the semimodule M if these elements are left (right) linearly independent in the sense of Definition 3.4 and their left (right) linear span is M.

REMARK 3.12. It may seem that we can have three different definitions of basis if we use here Definition 3.5 or Definition 3.8 instead of Definition 3.4. However, it appears that this is not the case. Namely, in the following, we show that any finitely generated module has a basis in the sense of Definition 3.11, but the same fact does not hold if we consider Definition 3.5 or Definition 3.8 instead of Definition 3.4 (see Remark 3.14). Moreover, it is easy to see that if we consider similar definitions of basis using the notion of independence from Definition 3.5 or 3.8, then any B, which is a basis in this sense, is a basis in the sense of Definition 3.11.

In the following we will consider *left* linear independence and *left* basis without pointing this out anymore. However, all those considerations are valid for right linear independence and right basis as well.

LEMMA 3.13. *Let $\mathcal{S}$ be a semiring and U be a finitely generated semimodule over $\mathcal{S}$. Then for the notion of linear dependence introduced in Definition 3.4, there exists a basis of U.*

PROOF. Let M be a certain finite generating set for U. If M is linearly independent in the sense of Definition 3.4, then we are done. Assume that M is linearly dependent. Then there exist $\mathbf{m} \in M$, $\mathbf{m}_1, \ldots, \mathbf{m}_k \in M$, $\mathbf{m}_i \neq \mathbf{m}$ for $i = 1, \ldots, k$, and $s_1, \ldots, s_k \in \mathcal{S}$ such that $\mathbf{m} = \sum_{i=1}^{k} s_i \mathbf{m}_i$. It follows that $M_1 = M \setminus \{\mathbf{m}\}$ is a generating set for U also. Continuing this process, we complete the proof. □

REMARK 3.14. Let us show that there are spaces without a basis if we consider linear independence in the sense of Definition 3.5 or Definition 3.8. We consider the linear span U_4 of the system of vectors

$$\left\{ \mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{u}_4 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$$

(the same vectors as in Example 3.7). Assume that there is a basis in this space in the sense of Definition 3.5. Thus, it is a basis in the sense of Definition 3.4 by Remark 3.6. However, it is proved in [40, Lemma 3.11] that any basis of U_4 contains $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3,$ and $\mathbf{u}_4$, which are linearly dependent in the sense of Definition 3.5. The same considerations hold for linear dependence in the sense of Definition 3.8.

In [27], it is proved that if $\mathcal{S}$ is a max-algebra, U is a finitely generated semimodule over $\mathcal{S}$, then any basis of U in the sense of Definition 3.4 has the same number of elements. The following example shows that this does not hold for more general classes of semirings, even under the condition of antinegativity.

EXAMPLE 3.15. Let $\mathcal{S} = \mathbb{Z}[\sqrt{7}]_+$ be the set of non-negative reals of the form $a+b\sqrt{7}$, where $a, b \in \mathbb{Z}$. We consider the semimodule U over $\mathcal{S}$, which is generated by the elements $\{3-\sqrt{7},\ \sqrt{7}-2\}$. The set $\{3-\sqrt{7},\ \sqrt{7}-2\}$ is a basis for U in the sense of Definition 3.4, which consists of two elements, but $1 = \left(3-\sqrt{7}\right) + \left(\sqrt{7}-2\right)$ also generates U over $\mathcal{S}$. Thus, $\{1\}$ is a basis consisting of one element.

In view of these examples, we give the following definition.

DEFINITION 3.16. Let U be a semimodule over a certain semiring $\mathcal{S}$. The (*left*) *dimension*, $\dim(U)$, is the minimal number of elements in any (left) basis of U.

By Lemma 3.13, any finitely generated semimodule has finite dimension. However, the following example shows that even a three-dimensional semimodule may have an infinite-dimensional subsemimodule.

EXAMPLE 3.17. It is shown in [27] that in the sense of Definition 3.4, the vectors $\begin{bmatrix} x_i \\ 0 \\ -x_i \end{bmatrix} \in \mathbb{R}^3_{\max}$, $i = 1, 2, \ldots, m$, are linearly independent for any m for different x_i.

Thus, in some cases, it is more informative to consider the maximal number of linear independent vectors in the system than to consider its dimension. From this point of view, all three definitions, Definitions 3.4, 3.5, and 3.8, work. It is easy to see that if $\mathcal{S}$ is a subsemiring of a certain field, then any $n+1$ vectors in $\mathcal{S}^n$ are linearly

dependent in the sense of Definition 3.5. Also, it is proved in [42] that any $n+1$ vectors in $\mathbb{R}^n_{\max}$ are linearly dependent in the sense of Definition 3.5. Hence, in both cases, they are dependent in the sense of Definition 3.8. In particular, a situation as in Example 3.17 cannot appear in these cases.

4. Rank functions

DEFINITION 4.1. We say that a matrix, A, *dominates* a matrix, B, $A \geq B$, if and only if $b_{i,j} \neq 0$ implies that $a_{i,j} \neq 0$.

DEFINITION 4.2. If $A = [a_{i,j}]$, $B = [b_{i,j}] \in \mathcal{M}_{m,n}(\mathcal{S})$, $A \geq B$, we denote by $A \backslash B$ the matrix $C = [c_{i,j}] \in \mathcal{M}_{m,n}(\mathcal{S})$ such that

$$c_{i,j} = \begin{cases} 0 & \text{if } b_{i,j} \neq 0 \\ a_{i,j} & \text{otherwise} \end{cases}.$$

Many authors have investigated various rank functions for matrices over semirings and their properties [40, 44–53].

It is known that over a field, the concept of a matrix rank has a geometrical interpretation as the dimension of the image space of the corresponding linear transformation. Over semirings, the situation is more involved, namely, this geometric approach leads to some surprising properties of the corresponding rank function. For example, it can appear that the rank of a submatrix is greater than the rank of a matrix. The reason is that the dimension of a certain subsemimodule may exceed in general the dimension of the whole semimodule, cf. Example 3.17 for matrices with entries from a max-algebra and Example 3.7. For completeness, we give it here again in more details:

EXAMPLE 4.3. [40] Let $\mathcal{S}$ be an arbitrary antinegative semiring without zero divisors. Consider the semimodule U_4 over $\mathcal{S}$ generated by the vectors $\mathbf{a}_1 = [0, 1, 0]^t$, $\mathbf{a}_2 = [0, 0, 1]^t$, $\mathbf{a}_3 = [1, 0, 1]^t$, and $\mathbf{a}_4 = [1, 1, 0]^t$ from Example 3.7. Then $\dim U_4 = 4$, however, U_4 is a proper subsemimodule of the three-dimensional $\mathcal{S}$-semimodule $\mathcal{S}^3$.

Indeed, it is easy to see that none of these vectors is a linear combination of the others. The reason why U_4 has no basis of less than four elements is that $[1, 0, 0]^t$ cannot be expressed as a linear combination of vectors from U_4 due to the antinegativity (see also [40, Lemma 3.11] for the detailed proof).

Based on Example 4.3, we may easily construct the following matrix example:

EXAMPLE 4.4. Let $\mathcal{S}$ be any antinegative semiring without zero divisors; we consider the matrix

$$Y = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

and its proper submatrix

$$X = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}.$$

The image of the linear operator corresponding to Y is the whole space $\mathcal{S}^3$; so its dimension is three, however, the dimension of the image of the linear operator corresponding to X is equal to 4 by Example 4.3.

In the following, we collect different definitions of a matrix rank over semirings. Each of them has some advantages and disadvantages in comparison with the usual rank function over a field. However, all these functions are vital to solve some problems arising in the theory of matrices over semirings and their applications.

4.1. *Definitions*

We start with one of the most important semiring rank functions.

Definition 4.5. A matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is of *factor rank* k ($\mathrm{f}(A) = k$) if there exist matrices $B \in \mathcal{M}_{m,k}(\mathcal{S})$ and $C \in \mathcal{M}_{k,n}(\mathcal{S})$ with $A = BC$, and k is the smallest positive integer for which such a factorization exists. By definition, the only matrix with the factor rank zero is the zero matrix, O.

Let us note (see [45, 54] for the proof) that the factor rank of A is equal to the minimum number of factor rank-1 matrices whose sum is A. For any submatrix B of A, it holds that $\mathrm{f}(B) \leq \mathrm{f}(A)$ [49].

The notion of factor rank is very important in different applications. It was used in [55] for demographical investigations, in [56] for combinatorial optimization problems, and in [54] for statistics.

If $\mathcal{S}$ is a subsemiring of a certain field, then there is the usual rank function $\rho(A)$ for any matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$.

The geometric approach leads to the following definitions based on the notion of linear independence.

Definition 4.6. The matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is of *row rank* k ($\mathrm{r}(A) = k$) if the dimension of the linear span of the rows of A is equal to k.

Definition 4.7. The matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is of *column rank* k ($\mathrm{c}(A) = k$) if the dimension of the linear span of the columns of A is equal to k.

It is proved in [49] that $\mathrm{r}(B) \leq \mathrm{r}(A)$ if B is obtained by deleting some columns of A and $\mathrm{c}(B) \leq \mathrm{c}(A)$ if B is obtained by deleting some rows of A. However, as

Example 4.4 shows, in general, the rank of a submatrix can be greater than the rank of a matrix.

The following example shows that all aforesaid functions may differ for matrices over semirings.

EXAMPLE 4.8. [39] Let

$$A = \begin{bmatrix} 2 & 3 & 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 & 0 & 2 \\ 0 & 0 & 2 & 0 & 4 & 2 \\ 2 & 3 & 1 & 0 & 2 & 2 \end{bmatrix} \in \mathcal{M}_{5,6}(\mathbb{Z}_+).$$

Then $\rho(A) = 3, \mathrm{f}(A) = 4, \mathrm{r}(A) = 5, \mathrm{c}(A) = 6$.

Since a basis of the linear span of a system of elements may not lie in this system (Examples 5.3 and 5.4 below), we consider also the following definitions of row and column ranks.

DEFINITION 4.9. The matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is said to be of *spanning row rank* k ($\mathrm{sr}(A) = k$) if the minimal number of rows that span all rows of A is k.

DEFINITION 4.10. The matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is said to be of *spanning column rank* k ($\mathrm{sc}(A) = k$) if the minimal number of columns that span all columns of A is k.

The following definitions depend on Definitions 3.4, 3.5, or 3.8 of linear independence, correspondingly.

DEFINITION 4.11. The matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is said to be of *maximal row rank* k ($\mathrm{mr}_1(A) = k$, $\mathrm{mr}_2(A) = k$, or $\mathrm{mr}_3(A) = k$) if it has k linearly independent rows, in the sense of Definitions 3.4, 3.5, or 3.8, respectively, and any $(k+1)$ rows are linearly dependent.

DEFINITION 4.12. The matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is said to be of *maximal column rank* k ($\mathrm{mc}_i(A) = k$, $i = 1, 2, 3$) if it has k linearly independent columns and any $(k+1)$ columns are linearly dependent with respect to Definitions 3.4, 3.5, or 3.8.

Also to avoid the obstruction of Example 4.4, we may use the following definitions:

DEFINITION 4.13. The matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is said to be of *enveloping row rank* k ($\mathrm{er}(A) = k$) if the minimal dimension of any semimodule containing all rows of A is equal to k.

DEFINITION 4.14. The matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is said to be of *enveloping column rank* k ($\mathrm{ec}(A) = k$) if the minimal dimension of any semimodule containing all columns of A is equal to k.

It will be shown later (see Proposition 5.1) that actually $\mathrm{er}(A) = \mathrm{ec}(A) = \mathrm{f}(A)$ for matrices over arbitrary semirings; $\mathrm{r}(A) = \mathrm{mr}_1(A) = \mathrm{sr}(A)$ [resp., $\mathrm{c}(A) = \mathrm{mc}_1(A) = \mathrm{sc}(A)$] if $\mathcal{S}$ is a max-algebra [27].

Let us recall that for matrices over fields, all notions introduced earlier give the same concept; however, they can be different for matrices over semirings. In the following sections, we provide some examples illustrating these difference and investigate how these functions are related to each other.

The definition of rank over fields based on the notion of maximal nonsingular minors has the following analogs in the semiring case:

DEFINITION 4.15. Let $A \in \mathcal{M}_{m,n}(\mathcal{S})$. It is said that A is of *tropical rank* k ($\mathrm{trop}(A) = k$) if k is the largest number such that A has a $(k \times k)$-submatrix, which is tropically nonsingular (see Definition 2.21).

DEFINITION 4.16. Let $A \in \mathcal{M}_{m,n}(\mathcal{S})$. It is said that A is of *symmetric rank* or *symmetric tropical rank* k ($\mathrm{rk}_{\mathrm{sym}}(A) = k$) if k is the largest number such that A has a $(k \times k)$-submatrix A', satisfying $|A'|^+ \neq |A'|^-$.

REMARK 4.17. In other words, symmetric tropical rank differs from the tropical rank by fixing $\mathcal{T}$ from Definition 2.21 to be the alternating group $A_i \subset S_i$ for all $(i \times i)$-submatrices of A, $i = 1, \ldots, n$.

PROPOSITION 4.18. *Let* $A \in \mathcal{M}_{m,n}(\mathcal{S})$. *Then* $\mathrm{sym}(A) \leq \mathrm{trop}(A)$.

PROOF. The proof follows easily from Remark 4.17. □

Now we consider the so-called *combinatorial ranks*, which are very useful for graph theory, transversal theory, and communication networks (see [4, 57, 58] and references therein). Note that these ranks do not coincide with the usual rank function even in the case where $\mathcal{S}$ is a field.

DEFINITION 4.19. A *line* of a matrix is its row or column.

DEFINITION 4.20. A matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is of *term rank* k ($\mathrm{t}(A) = k$) if the minimum number of lines needed to include all nonzero elements of A is equal to k.

Let us denote by $\mathrm{t_c}(A)$ the least number of columns needed to include all nonzero elements of A and by $\mathrm{t_r}(A)$ the least number of rows needed to include all nonzero elements of A.

DEFINITION 4.21. A *generalized diagonal* of a matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is a set of $\min\{m, n\}$ positions in A such that no row or column contains two of these positions.

PROPOSITION 4.22 (König Theorem). [57, Theorem 1.2.1] *Let $\mathcal{S}$ be a semiring, $A \in \mathcal{M}_{m,n}(\mathcal{S})$. Then the term rank of A is the maximum number of nonzero entries in some generalized diagonal of A.*

Also the term rank can be characterized in terms of a zero submatrix of maximal size.

PROPOSITION 4.23. [59] *Let $\mathcal{S}$ be a semiring, $A \in \mathcal{M}_{m,n}(\mathcal{S})$. Then $t(A) = k$ if and only if there exist positive integers s, r, $s + r = n + m - k$ such that for a certain permutation of rows and columns of A contains a submatrix $O_{r,s}$ and no permutation of rows and columns of A contains an $O_{p,q}$ if $p + q > n + m - k$.*

The following "dual" matrix invariant is also known.

DEFINITION 4.24. The matrix $A \in \mathcal{M}_{m,n}(\mathcal{S})$ is said to be of *zero-term rank k* ($\mathrm{z}(A) = k$) if the minimum number of lines needed to include all zero elements of A is equal to k.

In the case when $\mathcal{S}$ is the max-algebra, $\mathbb{R}_{\max}$, also the notion of *Kapranov rank*, $K(A)$, is useful (see [28, 60] for the details).

We should note that the list of rank functions introduced here is not exhausting. We presented here only the most common definitions. However, there are still other natural ways to define rank functions over semirings.

5. Relations between different rank functions

If the semiring $\mathcal{S}$ coincides with a field, then

$$\rho(A) = \mathrm{f}(A) = \mathrm{r}(A) = \mathrm{c}(A) = \mathrm{sr}(A) = \mathrm{sc}(A) = \mathrm{mr}_i(A) = \mathrm{mc}_i(A),$$

$i = 1, 2$. Over more general semirings, the situation is more complicated.

PROPOSITION 5.1. *Let $A \in \mathcal{M}_{m,n}(\mathcal{S})$, then* $\mathrm{er}(A) = \mathrm{f}(A) = \mathrm{ec}(A)$.

PROOF. Let us check that $\mathrm{er}(A) \leq \mathrm{f}(A)$. Write $\mathrm{f}(A) = \mathrm{f}$. Then $A = BC$ for certain $B \in \mathcal{M}_{m,f}(\mathcal{S})$ and $C \in \mathcal{M}_{f,n}(\mathcal{S})$. Then the row space of A is contained in the row space of C; hence $\mathrm{er}(A) \leq \mathrm{er}(C) \leq \mathrm{f}$.

Let us show that the converse inequality also holds. We write $\mathrm{er}(A) = r$. By definition, there exist vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathcal{S}^n$ and elements $\alpha_{i,j} \in \mathcal{S}$, $i = 1, \ldots, m$, $j = 1, \ldots, r$ such that the ith row of A is equal to $\sum_{j=1}^{r} \alpha_{i,j}\mathbf{v}_j^t$. Thus $A = BC$, where

$$B = [\alpha_{i,j}],\ C = \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_r \end{bmatrix}.$$

Similar considerations with the columns of A show that $\mathrm{ec}(A) = \mathrm{f}(A)$. □

PROPOSITION 5.2. *Let* $A \in \mathcal{M}_{m,n}(\mathcal{S})$. *Then* $\mathrm{r}(A) \leq \mathrm{sr}(A) \leq \mathrm{mr}_1(A) \leq m$ *and* $\mathrm{ec}(A) \leq \mathrm{c}(A) \leq \mathrm{sc}(A) \leq \mathrm{mc}_1(A) \leq n$.

PROOF. These statements follow directly from the definitions of these rank functions. □

The maximal column rank associated with Definition 3.4 may actually exceed the spanning column rank, and the spanning column rank may exceed the column rank over some semirings shown in the following examples.

EXAMPLE 5.3. [51] Let us consider $A = [3 - \sqrt{7}, \sqrt{7} - 2] \in \mathcal{M}_{1,2}(\mathbb{Z}[\sqrt{7}]_+)$. Thus $\mathrm{sc}(A) = 2$, since $3 - \sqrt{7} \neq \alpha(\sqrt{7} - 2)$ and $\alpha(3 - \sqrt{7}) \neq \sqrt{7} - 2$ in $\mathbb{Z}[\sqrt{7}]_+$. However, $\mathrm{c}(A) = 1$ since $1 = (3 - \sqrt{7}) + (\sqrt{7} - 2)$ generates the column space of A.

EXAMPLE 5.4. [45] We consider $A = [4 - \sqrt{7}, \sqrt{7} - 2, 1] \in \mathcal{M}_{1,3}(\mathbb{Z}[\sqrt{7}]_+)$. Thus $\mathrm{sc}(A) = 1$, since 1 spans all columns of A. However, just like in the previous example, one can see that $\mathrm{mc}(A) = 2$.

The same statements are also true for the row ranks.

PROPOSITION 5.5. *Let* $A \in \mathcal{M}_{m,n}(\mathcal{S})$ *then* $\mathrm{f}(A) \leq \min\{\mathrm{r}(A), \mathrm{c}(A)\}$.

PROOF. See [45, Proposition 3.1.1]. □

Let us note that this inequality can be strict. Moreover, unlike the case with rank over fields, row and column ranks may not coincide even over commutative semirings; moreover, column ranks may be greater than the number of rows, just like row ranks may be greater than the number of columns.

EXAMPLE 5.6. Let $\mathcal{S}$ be any antinegative semiring without zero divisors; we consider the matrix X defined in Example 4.4. Then it follows from Example 4.3 and Proposition 5.2 that $\mathrm{r}(X) = \mathrm{sr}(X) = \mathrm{mr}_1(X) = 3 = m$ while $\mathrm{c}(X) = \mathrm{sc}(X) = \mathrm{mc}_1(X) = 4 = n > m$.

PROPOSITION 5.7. *Let* $A \in \mathcal{M}_{m,n}(\mathcal{S})$ *then* $\mathrm{f}(A) \leq t(A)$.

PROOF. See [45, Proposition 3.1.3]. □

Note that the inequality $\max\{\mathrm{r}(A), \mathrm{c}(A)\} \leq \mathrm{t}(A)$ does not hold in general.

EXAMPLE 5.8. For the matrix X given in Example 4.4, one has that $\mathrm{r}(A) = \mathrm{t}(A) = 3$, $\mathrm{c}(A) = 4$.

The inequality $\min\{\mathrm{r}(A), \mathrm{c}(A)\} \leq \mathrm{t}(A)$ does not hold over $\mathbb{Z}_+$.

EXAMPLE 5.9. [45] For

$$A = \begin{bmatrix} 3 & 5 & 7 \\ 5 & 0 & 0 \\ 7 & 0 & 0 \end{bmatrix} \in \mathcal{M}_{3,3}(\mathbb{Z}_+),$$

one has that $\mathrm{r}(A) = \mathrm{c}(A) = 3$, $\mathrm{t}(A) = 2$.

PROPOSITION 5.10. *Let $\mathcal{S}$ be either a max-algebra or an antinegative subsemiring of a field, or some other semiring, for which any $n+1$ vectors from $\mathcal{S}^n$ are linearly dependent, $A \in \mathcal{M}_{m,n}(\mathcal{S})$. Then* $\mathrm{mr}_2(A) \leq f(A)$.

PROOF. Repeat the arguments from [42, Lemma 5.9]. □

The factor rank of some matrices can be greater than the Gondran–Minoux rank.

EXAMPLE 5.11. Let us consider the matrix

$$K_4 := \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \in \mathcal{M}_4(\mathbb{R}_{\max}) \text{ or } K_4 \in \mathcal{M}_4(\mathbf{B}).$$

Then it is easy to see that the sum of first two rows (columns) is the vector of all ones, and it is equal to the sum of last two rows (columns). Thus $\mathrm{mr}_1(K_4) \leq 3$. However, $\mathrm{f}(K_4) = 4$ (see [42] or [28]).

PROPOSITION 5.12. *Let $\mathcal{S}$ be a max-algebra, $A \in \mathcal{M}_{m,n}(\mathcal{S})$. Then*

$$\mathrm{rk}_{\mathrm{sym}}(A) \leq \mathrm{mr}_2(A).$$

PROOF. The result follows from [19]; see also [45, Theorem 1']. □

PROPOSITION 5.13. *Let $\mathcal{S}$ be an antinegative subsemiring of a field, $A \in \mathcal{M}_{m,n}(\mathcal{S})$. Then* $\mathrm{mr}_2(A) = \rho(A) = \mathrm{rk}_{\mathrm{sym}}(A)$.

PROOF. Let us prove the first equality. Let some vectors over a semiring $\mathcal{S} \subset \mathbb{F}$ be linearly dependent in the sense of Definition 3.5. Here, $\mathbb{F}$ is a field. Then it is straightforward to see that they are linearly dependent over $\mathbb{F}$, i.e., $\mathrm{mr}_2(A) \leq \rho(A)$ for any A. Conversely, if $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are $\mathbb{F}$-linearly dependent, then they are linearly dependent in the field of fractions of the ring $\mathcal{S} \cup (-\mathcal{S})$, since over fields dimension does not depend on the field extension. Thus, multiplying by the common denominator of all

coefficients, we get a linear relation between $\mathbf{v}_1, \ldots, \mathbf{v}_k$ over $\mathcal{S} \cup (-\mathcal{S})$. It is nontrivial, since $\mathrm{char}(\mathbb{F}) = 0$ by the antinegativity of $\mathcal{S}$. Now, taking the summands with negative coefficients to the other side of the equality sign, we obtain that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly dependent in the sense of Definition 3.5, i.e., $\rho(A) \leq \mathrm{mr}_2(A)$, which completes the proof of the first inequality.

Now, $\rho(A)$ is the size of a maximal nonzero minor $|\hat{A}|$ of A, i.e., this is a maximal submatrix with $|\hat{A}|^+ \neq |\hat{A}|^-$, which proves the second inequality. □

The following is the main result of [41], see also [42] and [43].

PROPOSITION 5.14. *Let $\mathcal{S}$ be a max-algebra, $A \in \mathcal{M}_{m,n}(\mathcal{S})$. Then* $\mathrm{mr}_3(A) = \mathrm{trop}(A) = \mathrm{mc}_3(A)$.

PROPOSITION 5.15. *Let $A \in \mathcal{M}_{m,n}(\mathcal{S})$, then* $\mathrm{mr}_3(A) \leq \mathrm{mr}_2(A) \leq \mathrm{mr}_1(A)$.

PROOF. This follows from Remarks 3.6 and 3.9. □

Now, let us show that $\mathrm{mr}_3(A)$ can be less than $\mathrm{mr}_2(A)$ both over max-algebras and antinegative subsemirings of fields.

EXAMPLE 5.17. In the case where $\mathcal{S}$ is a max-algebra, let us consider the matrix

$$K_3 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Then the rows (columns) of this matrix are linearly dependent in the sense of Definition 3.8 with the coefficients 0, 0, 0. Let us recall that 0 is the identity element of a max-algebra; so this is nontrivial linear dependence. However, $|K_3|^+ = 3 \neq 2 = |K_3|^-$ and by Proposition 5.12, we have that $\mathrm{mr}_2(K_3) \geq \mathrm{rk}_{\mathrm{sym}}(K_3) = 3$.

In the case where $\mathcal{S}$ is the semiring of nonnegative integers with natural addition and multiplication, we consider the same matrix K_3. Then its rows (columns) are linearly dependent in the sense of Definition 3.8 with the coefficients 1, 1, 1; however, $\det(K_3) = 2$ evaluated over $\mathbb{Z}$; thus, by Proposition 5.13, it holds that $\mathrm{mr}_2(A) = 3$.

The statement below follows easily from Remark 3.12 and the above inequalities.

COROLLARY 5.17.

- *If $\mathcal{S}$ is a max-algebra and the row space of A has a basis for linear independence as defined by Definition 3.8, then*

$$\begin{aligned} \mathrm{mr}_3(A) = \mathrm{trop}(A) = \mathrm{rk}_{\mathrm{sym}}(A) = K(A) = \mathrm{mr}_2(A) = \mathrm{mc}_2(A) \\ = \mathrm{f}(A) = \mathrm{r}(A) = \mathrm{c}(A) = \mathrm{mr}_1(A) = \mathrm{mc}_1(A) = \mathrm{sr}(A) = \mathrm{sc}(A). \end{aligned}$$

- *If $\mathcal{S}$ possesses the property that any $n+1$ vectors from $\mathcal{S}^n$ are linearly dependent and that the row space of A has a basis for linear independence as defined by Definition 3.5, then*

$$\mathrm{mr}_2(A) = \mathrm{f}(A) = \mathrm{r}(A).$$

In [47], the authors investigate how the column rank and the factor rank of matrices over certain algebraic systems change as the algebraic system changes. For different classes of semirings in [49], upper bounds are found for the sets of numbers r such that for all $A \in \mathcal{M}_{m,n}(\mathcal{S})$ with $\mathrm{c}(A) = r$, it follows that $\mathrm{f}(A) = r$. It appears that these upper bounds depend considerably on the semiring $\mathcal{S}$.

REMARK 5.18. It is straightforward to see that if $\mathcal{S} = \mathbf{B}$ is a binary Boolean semiring, then $\mathrm{z}(A) = \mathrm{t}(J \setminus A)$ for any $A \in \mathcal{M}_{m,n}(\mathcal{S})$.

6. Arithmetic behavior of rank

The behavior of the usual rank function ρ over fields with respect to matrix multiplication, addition, and union is given by the following classical inequalities: The *rank-sum inequalities*:

$$|\ \rho(A) - \rho(B)\ | \leq \rho(A+B) \leq \rho(A) + \rho(B);$$

Sylvester's laws:

$$\rho(A) + \rho(B) - n \leq \rho(AB) \leq \min\{\rho(A), \rho(B)\};$$

the *rank inequalities for matrix union*:

$$\max\{\rho(A), \rho(B)\} \leq \rho(A|B) \leq \rho(A) + \rho(B),$$

where A and B are conformal matrices with entries from a field.

Following [45], we investigate the behavior of different rank functions over semirings with respect to matrix addition and multiplication. It turns out that the arithmetic properties of semiring rank functions depend deeply on the structure of the semiring of entries.

DEFINITION 6.1. We say that a bound is *exact* if there are matrices such that equality holds and *best possible* if for any given r and s, there are matrices of ranks r and s, respectively, such that the equality holds.

6.1. *Rank-sum inequalities*

Let us show that the standard lower bound for the rank of a sum of two matrices is not valid in general.

EXAMPLE 6.2. [45] Let $\mathcal{S}$ be a Boolean semiring. Then it is possible that

$$\mathrm{f}(A+B) \not\geq |\mathrm{f}(A) - \mathrm{f}(B)|.$$

Let us consider $A = K_7 = \begin{bmatrix} 0&1&1&1&1&1&1\\ 1&0&1&1&1&1&1\\ 1&1&0&1&1&1&1\\ 1&1&1&0&1&1&1\\ 1&1&1&1&0&1&1\\ 1&1&1&1&1&0&1\\ 1&1&1&1&1&1&0 \end{bmatrix}$ and $B = I_7$, then $1 = \mathrm{f}(J_7) = \mathrm{f}(A+B)$, but over any Boolean semiring, $\mathrm{f}(K_7) \leq 5$ due to the factorization

$$K_7 = \begin{bmatrix} 1&0&1&0&1\\ 1&1&0&0&0\\ 0&1&0&1&1\\ 0&1&1&1&0\\ 0&1&1&0&1\\ 0&0&1&1&1\\ 1&0&1&1&0 \end{bmatrix} \begin{bmatrix} 0&0&1&1&1&1&0\\ 1&0&0&0&0&1&1\\ 0&1&1&0&0&0&0\\ 1&1&0&0&1&0&0\\ 0&0&0&1&0&0&1 \end{bmatrix}$$

(see [61] for more details). Thus,

$$|\mathrm{f}(A) - \mathrm{f}(B)| = \mathrm{f}(B) - \mathrm{f}(A) \geq 7 - 5 = 2 > 1 = \mathrm{f}(A+B).$$

However, the following bounds are true.

PROPOSITION 6.3. [45] *Let $\mathcal{S}$ be an antinegative semiring, $A, B \in \mathcal{M}_{m,n}(\mathcal{S})$. Then,*

1. $\mathrm{f}(A+B) \leq \min\{\mathrm{f}(A) + \mathrm{f}(B), m, n\}$;
2. $\mathrm{f}(A+B) \geq \begin{cases} \mathrm{f}(A) & \text{if} \quad B = O \\ \mathrm{f}(B) & \text{if} \quad A = O \\ 1 & \text{if} \quad A \neq O \text{ and } B \neq O \end{cases}$.

These bounds are exact, the upper bound is best possible and the lower bound is best possible over Boolean semirings.

In the case when $\mathcal{S}$ is a subsemiring of $\mathbb{R}_+$, the positive reals with the usual operations, the standard lower bound for the rank of sum of two matrices is again not valid because of the following example.

EXAMPLE 6.4. [45] Let $r, s \geq 4$ and $s < n - 4$. Let us consider

$$A' = \begin{bmatrix} 1&2&3&4\\ 1&1&1&1\\ 1&0&1&0\\ 0&0&2&2 \end{bmatrix}$$

and

$$B' = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

Note that $\mathrm{f}(A') = 4$, $\rho(A') = 3$, $\mathrm{f}(B') = \rho(B') = 1$, and $\mathrm{f}(A' + B') = \rho(A' + B') = 2$.
Let

$$A = \begin{bmatrix} A' & O_{4,r-4} & O_{4,n-r} \\ O_{r-4,4} & L_{r-4} & O_{r-4,n-r} \\ O_{m-r,4} & O_{m-r,r-4} & O_{m-r,n-r} \end{bmatrix}$$

and

$$B = \begin{bmatrix} B' & O_{4,1} & O_{4,s-1} & O_{4,n-s-4} \\ O_{s-1,4} & O_{s-1,1} & U_{s-1} & O_{s-1,n-s-4} \\ O_{m-s-3,4} & O_{m-s-3,1} & O_{m-s-3,s-1} & O_{m-s-3,n-s-4} \end{bmatrix},$$

where $U_k = \sum_{1 \le i \le j \le k} E_{i,j}$, $L_k = \sum_{1 \le j < i \le k} E_{i,j} \in \mathcal{M}_k(\mathcal{S})$. Here, we have that $\mathrm{f}(A) = r$, $\rho(A) = r - 1$, $\mathrm{f}(B) = \rho(B) = s$, and $\mathrm{f}(A + B) = \mid r - s \mid -1 < \mid \mathrm{f}(A) - \mathrm{f}(B) \mid$ if $\mathcal{S}$ is a subsemiring of $\mathbb{R}_+$. Note that if $r = s + 3$, reversing the roles of A' and B' in A and B also gives a corresponding example.

However, the following is true.

PROPOSITION 6.5. [45] *Let $\mathcal{S} \subseteq \mathbb{R}_+$, $A, B \in \mathcal{M}_{m,n}(\mathcal{S})$. Then*

$$\mathrm{f}(A + B) \ge |\rho(A) - \rho(B)|.$$

This bound is exact and best possible.

The following inequalities are true for the term rank.

PROPOSITION 6.6. [45] *Let $\mathcal{S}$ be an arbitrary semiring. For any matrices $A, B \in \mathcal{M}_{m,n}(\mathcal{S})$, we have*

$$\mathrm{t}(A + B) \le \min\{\mathrm{t}(A) + \mathrm{t}(B), m, n\}.$$

This bound is exact and best possible.

The following example shows that a nontrivial, additive lower bound does not hold over an arbitrary semiring.

EXAMPLE 6.7. [45] Let $A = B = J_{m,n}$ over a field whose characteristic is equal to 2. Then $\mathrm{t}(A + B) = \mathrm{t}(O) = 0$.

However for antinegative semirings there is a lower bound, which is even better than the standard one. Namely, the following is true.

PROPOSITION 6.8. [45] *Let $\mathcal{S}$ be an antinegative semiring. For any matrices $A, B \in \mathcal{M}_{m,n}(\mathcal{S})$, the following inequality holds:*

$$\mathrm{t}(A+B) \geq \max\{\mathrm{t}(A), t(B)\}.$$

This bound is exact and best possible.

PROPOSITION 6.9. [45] *Let $\mathcal{S}$ be an antinegative semiring. For $A, B \in \mathcal{M}_{m,n}(\mathcal{S})$, one has that*

$$0 \leq \mathrm{z}(A+B) \leq \min\{\mathrm{z}(A), \mathrm{z}(B)\}.$$

These bounds are exact and best possible.

PROOF. Both bounds follow directly from the definition of the zero-term rank function. To check that the lower bound is exact and best possible for each pair (r,s), $0 \leq r, s \leq \min\{m,n\}$, let us consider the family of matrices $A_r = J \setminus \left(\sum_{i=1}^{r} E_{i,i}\right)$, $B_s = J \setminus \left(\sum_{i=1}^{s} E_{i,i+1}\right)$ if $s < \min\{m,n\}$, and $B_s = J \setminus \left(\sum_{i=1}^{s-1} E_{i,i+1} + E_{s,1}\right)$ if $s = \min\{m,n\}$. Then $\mathrm{z}(A_r) = r, \mathrm{z}(B_s) = s$ by the definition and $\mathrm{z}(A_r + B_s) = 0$ by antinegativity. Similarly, it can be checked that the upper bound is exact and best possible. □

As the following example shows, the standard upper bound for the additive inequalities is not valid for row and column ranks.

EXAMPLE 6.10. Consider

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 2 & 2 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \\ 0 & 6 & 2 \\ 0 & 4 & 6 \end{bmatrix}.$$

Then it is easy to see that $\mathrm{r}(A) = \mathrm{r}(B) = \mathrm{sr}(A) = \mathrm{sr}(B) = \mathrm{mr}_1(A) = 2$. However,

$$\mathrm{r}(A+B) = \mathrm{r}\begin{bmatrix} 1 & 2 & 2 \\ 0 & 3 & 0 \\ 1 & 1 & 2 \\ 1 & 6 & 2 \\ 0 & 5 & 6 \end{bmatrix} = 5 = \mathrm{sr}(A+B) = \mathrm{mr}_1(A+B)$$

over $\mathbb{Z}_+$.

PROPOSITION 6.11. [45] *Let $\mathcal{S}$ be an antinegative semiring. Then for $O \neq A, B \in \mathcal{M}_{m,n}(\mathcal{S})$, one has that*

$$1 \leq \mathrm{c}(A+B), \mathrm{r}(A+B), \mathrm{sr}(A+B), \mathrm{sc}(A+B), \mathrm{mr_i}(A+B), \mathrm{mc_i}(A+B), i = 1, 2, 3.$$

These bounds are exact over any antinegative semiring and best possible over Boolean semirings if $i = 1$.

PROOF. Again we are going to show that the lower bound is best possible. Let $\mathcal{S}$ be a Boolean semiring. For each pair (r, s), $0 \leq r, s \leq m$, we consider the matrices $A_r = J \setminus \left(\sum_{i=1}^{r} E_{i,i}\right)$, $B_s = J \setminus \left(\sum_{i=1}^{s} E_{i,i+1}\right)$ if $s < m$ and $B_s = J \setminus \left(\sum_{i=1}^{s-1} E_{i,i+1}\right) + E_{s,1}$ if $s = m$. Then

$$\mathrm{c}(A_r) = \mathrm{r}(A_r) = \mathrm{sr}(A_r) = \mathrm{sc}(A_r) = \mathrm{mr}_1(A_r) = \mathrm{mc}_1(A_r) = r,$$
$$\mathrm{c}(B_s) = \mathrm{r}(B_s) = \mathrm{sr}(B_s) = \mathrm{sc}(B_s) = \mathrm{mr}_1(B_s) = \mathrm{mc}_1(B_s) = s,$$

and $A_r + B_s = J$ has row and column ranks equal to 1. Thus, these bounds are best possible over Boolean semirings.

The rest of the proof follows directly from the definitions. □

PROPOSITION 6.12. [45] *Let $\mathcal{S}$ be a subsemiring in $\mathbb{R}_+$. Then for $A, B \in \mathcal{M}_{m,n}(\mathcal{S})$, one has that*

$$\mathrm{c}(A+B), \mathrm{r}(A+B), \mathrm{sr}(A+B), \mathrm{sc}(A+B), \mathrm{mr}_1(A+B),$$
$$\mathrm{mc}_1(A+B) \geq |\rho(A) - \rho(B)|.$$

These bounds are exact and best possible.

6.2. *Sylvester inequalities*

First, we show that the analog of the Sylvester lower bound does not hold for the factor rank.

EXAMPLE 6.13. Let $\mathcal{S}$ be a Boolean semiring,

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

and $B = A^t$, then $\mathrm{f}(A) = \mathrm{f}(B) = n$ and $\mathrm{f}(AB) = 1 \not\geq \mathrm{f}(A) + \mathrm{f}(B) - n = n$ since $AB = J$.

However, we can prove the following.

PROPOSITION 6.14. [45] *Let $\mathcal{S}$ be an antinegative semiring, $A \in \mathcal{M}_{m,n}(\mathcal{S})$, $B \in \mathcal{M}_{n,k}(\mathcal{S})$. Then*

1. $\mathrm{f}(AB) \leq \min\{\mathrm{f}(A), \mathrm{f}(B)\}$,
2. $\mathrm{f}(AB) \geq \begin{cases} 0 & \text{if} \quad \mathrm{f}(A) + \mathrm{f}(B) \leq n \\ 1 & \text{if} \quad \mathrm{f}(A) + \mathrm{f}(B) > n, \end{cases}$

provided that $\mathcal{S}$ has no zero divisors.
These bounds are exact, the upper bound is best possible, and the lower bound is best possible over Boolean semirings.

The next example demonstrates that the standard lower bounds of the Sylvester inequality are not valid in the case $\mathcal{S} \subseteq \mathbb{R}_+$.

EXAMPLE 6.15. [45] Let us consider

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 4 \\ 1 & 1 & 4 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

then $\rho(A) = 3$, $\mathrm{f}(A) = 4$, $\rho(B) = \mathrm{f}(B) = 2$, and

$$AB = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \\ 2 & 2 & 4 & 4 \\ 2 & 2 & 4 & 4 \end{bmatrix},$$

so $\mathrm{f}(AB) = 1$. Thus, $1 = \mathrm{f}(AB) \not\geq \mathrm{f}(A) + \mathrm{f}(B) - n = 4 + 2 - 4 = 2$ and $6 = 4 + 2 = \mathrm{f}(AI) + \mathrm{f}(IB) \not\leq \mathrm{f}(AIB) + \mathrm{f}(I) = 1 + 4 = 5$.

However, the following generalization for the lower bound is true.

PROPOSITION 6.16. [45] *Let $\mathcal{S} \subseteq \mathbb{R}_+$, $A \in \mathcal{M}_{m,n}(\mathcal{S})$, $B \in \mathcal{M}_{n,k}(\mathcal{S})$. Then*

$$\mathrm{f}(AB) \geq \begin{cases} 0 & \text{if} \quad \rho(A) + \rho(B) \leq n, \\ \rho(A) + \rho(B) - n & \text{if} \quad \rho(A) + \rho(B) > n. \end{cases}$$

This bound is exact and best possible.

Let us turn to the term rank now.

EXAMPLE 6.17. The inequality $\mathrm{t}(AB) \leq min(\mathrm{t}(A), \mathrm{t}(B))$ does not hold. For a counterexample, take $A = C_1$, $B = R_1$. Then $\mathrm{t}(AB) = \mathrm{t}(J_n) = n > 1$.

Also, there is no nontrivial multiplicative lower bound over an arbitrary semiring.

EXAMPLE 6.18. Set $A = B = J_n$ over a field whose characteristic is a divisor of n. Then $\mathrm{t}(AB) = \mathrm{t}(nJ_n) = 0$.

However, the following inequalities are true.

PROPOSITION 6.19. [45] *Let $\mathcal{S}$ be an antinegative semiring. Then for any $A \in \mathcal{M}_{m,n}(\mathcal{S})$, $B \in \mathcal{M}_{n,k}(\mathcal{S})$, the inequalities*

$$\mathrm{t}(AB) \leq \min(\mathrm{t}_r(A), \mathrm{t}_c(B))$$

and

$$\mathrm{t}(AB) \geq \begin{cases} 0 & if \quad \mathrm{t}(A) + \mathrm{t}(B) \leq n, \\ \mathrm{t}(A) + \mathrm{t}(B) - n & if \quad \mathrm{t}(A) + \mathrm{t}(B) > n \end{cases}$$

hold. These are exact and best possible bounds.

PROPOSITION 6.20. [45] *Let $\mathcal{S}$ be an antinegative semiring. For $A \in \mathcal{M}_{m,n}(\mathcal{S})$, $B \in \mathcal{M}_{n,k}(\mathcal{S})$, we have*

$$0 \leq \mathrm{z}(AB) \leq \min\{\mathrm{z}(A) + \mathrm{z}(B), k, m\}.$$

These bounds are exact and best possible for $n > 2$.

PROPOSITION 6.21. [45] *Let $\mathcal{S}$ be an antinegative semiring without zero divisors. For $O \neq A \in \mathcal{M}_{m,n}(\mathcal{S})$, $O \neq B \in \mathcal{M}_{n,k}(\mathcal{S})$, and $\mathrm{c}(A) + \mathrm{r}(B) > n$, one has that*

$$1 \leq \mathrm{c}(AB), \mathrm{r}(AB), \mathrm{sr}(AB), \mathrm{sc}(AB), \mathrm{mr}_1(AB), \mathrm{mc}_1(AB).$$

These bounds are exact over any antinegative semiring without zero divisors and best possible over Boolean semirings.

PROOF. For $O \neq A \in \mathcal{M}_{m,n}(\mathcal{S})$ and $O \neq B \in \mathcal{M}_{n,k}(\mathcal{S})$, the matrix A has at least $\mathrm{c}(A)$ ($\mathrm{sc}(A)$, $\mathrm{mc}_1(A)$) nonzero columns while B has at least $\mathrm{r}(B)$ ($\mathrm{sr}(B)$, $\mathrm{mr}_1(B)$) nonzero rows. Thus, if $\mathrm{c}(A) + \mathrm{r}(B) > n$, $AB \neq O$ and hence these bounds are established. For the proof of exactness, let us take $A = B = E_{1,1}$.

Let $\mathcal{S}$ be a Boolean semiring. For each pair (r, s), $1 \leq r \leq \min\{m, n\}$, $1 \leq s \leq \min\{k, n\}$, let us consider the matrices

$$A_r = \sum_{i=1}^{r} E_{i,i} + \sum_{i=1}^{m} E_{i,1}, \quad B_s = \sum_{i=1}^{s} E_{i,i} + \sum_{i=1}^{n} E_{1,i}.$$

Then

$$\mathrm{c}(A_r) = \mathrm{r}(A_r) = \mathrm{sr}(A_r) = \mathrm{sc}(A_r) = \mathrm{mr}_1(A_r) = \mathrm{mc}_1(A_r) = r,$$

$$\mathrm{c}(B_s) = \mathrm{r}(B_s) = \mathrm{sr}(B_s) = \mathrm{sc}(B_s) = \mathrm{mr}_1(B_s) = \mathrm{mc}_1(B_s) = s$$

by definition, and $A_r B_s = J$. Thus $\mathrm{c}(A_r B_s) = \mathrm{r}(A_r B_s) = \mathrm{sr}(A_r B_s) = \mathrm{sc}(A_r B_s) = \mathrm{mr}_1(A_r B_s) = \mathrm{mc}_1(A_r B_s) = 1$. □

Note that the condition that $\mathrm{c}(A) + \mathrm{r}(B) > n$ is necessary because for $A = I_k \oplus O$ and $B = O \oplus I_j$, we have $AB = O$ whenever $k + j \leq n$.

PROPOSITION 6.22. [45] *Let $\mathcal{S}$ be a subsemiring in $\mathbb{R}_+$. Then for $A \in \mathcal{M}_{m,n}(\mathcal{S})$ and $B \in \mathcal{M}_{n,k}(\mathcal{S})$, one has that*

$$\begin{matrix} \mathrm{c}(AB), \mathrm{sc}(AB), \mathrm{mc}_1(AB) \\ \mathrm{r}(AB), \mathrm{sr}(AB), \mathrm{mr}_1(AB) \end{matrix} \geq \begin{cases} 0 & if \quad \rho(A)+\rho(B) \leq n, \\ \rho(A)+\rho(B)-n & if \quad \rho(A)+\rho(B) > n. \end{cases}$$

These bounds are exact and best possible.

The following example, given in [51] for the spanning column rank, shows that standard analog for the upper bound of the rank of product of two matrices do not work for row and column ranks.

EXAMPLE 6.23. [51] Let $A = [3, 7, 7] \in \mathcal{M}_{1,3}(\mathbb{Z}_+)$, $B = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$. Then $\mathrm{c}(A) = \mathrm{sc}(A) = \mathrm{mc}(A) = 2$, $\mathrm{c}(B) = \mathrm{sc}(B) = \mathrm{mc}(B) = 3$, and $\mathrm{c}(AB) = \mathrm{c}(3, 10, 17) = \mathrm{sc}(AB) = \mathrm{mc}(AB) = 3$.

However, the following upper bounds are proved in [51].

PROPOSITION 6.24. [51] *Let $\mathcal{S}$ be an antinegative semiring. Then for $A \in \mathcal{M}_{m,n}(\mathcal{S})$ and $B \in \mathcal{M}_{n,k}(\mathcal{S})$, one has that*

$$\mathrm{c}(AB) \leq \mathrm{c}(B), \mathrm{sc}(AB) \leq \mathrm{sc}(B), \mathrm{mc}_1(AB) \leq \mathrm{mc}_1(B),$$
$$\mathrm{r}(AB) \leq \mathrm{r}(A), \mathrm{sr}(AB) \leq \mathrm{sr}(A), \mathrm{mr}_1(AB) \leq \mathrm{mr}_1(A).$$

These bounds are exact and best possible.

Similarly, the following can be proved.

PROPOSITION 6.25. *Let $\mathcal{S}$ be an antinegative semiring. Then for $A \in \mathcal{M}_{m,n}(\mathcal{S})$, $B \in \mathcal{M}_{n,k}(\mathcal{S})$, one has that*

$$\mathrm{mc}_i(AB) \leq \mathrm{mc}_i(B), \ \mathrm{mr}_i(AB) \leq \mathrm{mr}_i(A), \ i = 2, 3.$$

These bounds are exact and best possible.

PROPOSITION 6.26. [42] *Let $\mathcal{S}$ be a max-algebra then the following holds*

$$\mathrm{rk}_{\mathrm{sym}}(AB) \leq \min\{\mathrm{rk}_{\mathrm{sym}}(A), \mathrm{rk}_{\mathrm{sym}}(B)\}.$$

6.3. *Ranks of matrix union*

In [53, Lemma 3.17, Lemma 3.20], the following is proved.

PROPOSITION 6.27. *Let $\mathcal{S}$ be an antinegative semiring. For any $A \in \mathcal{M}_{m,n}(\mathcal{S})$, and $B \in \mathcal{M}_{m,k}(\mathcal{S})$, the following inequalities for the rank of a matrix union hold. These inequalities are sharp and the upper bounds are best possible:*

$$\max\{\mathrm{f}(A), \mathrm{f}(B)\} \leq \mathrm{f}(A|B) \leq \mathrm{f}(A) + \mathrm{f}(B),$$

$$\max\{\mathrm{t}(A), \mathrm{t}(B)\} \leq \mathrm{t}(A|B) \leq \mathrm{t}(A) + \mathrm{t}(B).$$

PROPOSITION 6.28. *Let $\mathcal{S}$ be an antinegative semiring. Then for any $A \in \mathcal{M}_{m,n}(\mathcal{S})$, $B \in \mathcal{M}_{m,k}(\mathcal{S})$, it holds that* $\max\{\mathrm{r}(A), \mathrm{r}(B)\} \leq \mathrm{r}(A|B)$.

PROOF. This is straightforward by the definition of row rank. □

Let us show that the row rank of the matrix union $(A|B)$ can be bigger than the sum of the row ranks of A and B.

EXAMPLE 6.29. Let us consider the following two matrices

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Then it is straightforward to see that $\mathrm{r}(A) = 2$, $\mathrm{r}(B) = 1$. Note that by Example 4.4, it holds that

$$\mathrm{r}(A|B) = 4 > 3 = \mathrm{r}(A) + \mathrm{r}(B).$$

PROPOSITION 6.30. *Let $\mathcal{S}$ be an antinegative semiring. Then for any $A \in \mathcal{M}_{m,n}(\mathcal{S})$, $B \in \mathcal{M}_{m,k}(\mathcal{S})$, it holds that* $\mathrm{c}(A|B) \leq \mathrm{c}(A) + \mathrm{c}(B)$.

PROOF. This follows directly from the definition of column rank. □

The following example shows that the column rank of the matrix union $(A|B)$ can be less than the minimum of column ranks of A and B.

EXAMPLE 6.31. Let us consider the matrices

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}.$$

Then one has by Example 4.4 that $\mathrm{c}(A) = 4$, $\mathrm{c}(B) = 4$, but $\mathrm{c}(A|B) = 3 < 4$.

For the other row and column ranks, the situation is analogous.

Propositions 6.32 and 6.33 follow directly from the definitions of bi-determinant and tropical singularity, respectively.

PROPOSITION 6.32. *Let $\mathcal{S}$ be an antinegative semiring. For any $A \in \mathcal{M}_{m,n}(\mathcal{S})$, $B \in \mathcal{M}_{m,k}(\mathcal{S})$, the following inequalities for the rank of matrix union hold. These inequalities are sharp and the upper bound is best possible:*

$$\max\{\mathrm{rk}_{\mathrm{sym}}(A), \mathrm{rk}_{\mathrm{sym}}(B)\} \leq \mathrm{rk}_{\mathrm{sym}}(A|B) \leq \mathrm{rk}_{\mathrm{sym}}(A) + \mathrm{rk}_{\mathrm{sym}}(B).$$

PROPOSITION 6.33. *Let $\mathcal{S}$ be an antinegative semiring. For any $A \in \mathcal{M}_{m,n}(\mathcal{S})$, $B \in \mathcal{M}_{m,k}(\mathcal{S})$, the following inequalities for the rank of matrix union hold. These inequalities are sharp and the upper bound is best possible:*

$$\max\{\mathrm{trop}(A), \mathrm{trop}(B) \leq \mathrm{trop}(A|B) \leq \mathrm{trop}(A) + \mathrm{trop}(B).$$

Acknowledgments

The author is grateful to Marianne Akian, LeRoy Beasley, Peter Butkovič, Stephane Gaubert, and Jonathan Golan for interesting discussions during the work on this project.

References

[1] R. Dedekind, *Über die Theorie der ganzen algebraiscen Zahlen*, Supplement XI to P. G. Lejeune Dirichlet: Vorlesung über Zahlentheorie, vol. 4 Aufl., Druck und Verlag, Braunschweig, 1894.

[2] H.S. Vandiver, Note on a simple type of algebra in which the cancellation law of addition does not hold, Bull. Am. Math. Soc. **40** (1934) 914–920.

[3] K. Glazek, *A Guide to the Literature on Semirings and their Applications in Mathematics and Information Sciences*, Kluwer Academic Publishers, Dordrecht 2002.

[4] J.S. Golan, *Semirings and their Applications, Updated and Expanded Version of the Theory of Semirings, with Applications to Mathematics and Theoretical Computer Science*, Kluwer Academic Publishers, Dordrecht, 1999.

[5] U. Hebisch, H.J. Weinert, *Semirings: Algebraic Theory and Applications in Computer Science, Series in Algebra*, vol. **V**, World Scientific Publishing Co. Inc. River Edge. NJ, 1998, 361 pp.

[6] J.H. Conway, *Regular Algebra and Finite Machines*, Chapman and Hall, London, 1971.

[7] S. Eilenberg, *Automata, Languages, and Machines* A, Academic Press, New York, 1974.

[8] R.A. Cuninghame-Green, *Minimax Algebra*, Lecture Notes in Econ. Math. Syst., vol. **166**, Springer-Verlag, Berlin, 1979.

[9] V.N. Kolokol'tsov, V. Maslov, *Idempotent Analysis and Applications*, Kluwer, Dordrecht, 1997.

[10] A.I. Barvinok, *Combinatorial Optimization and Computations in the ring of Polynomials,* DIMACS Technical Report 93–13, 1993.

[11] P. Butkovič, Max-algebra: the linear algebra of combinatorics? Linear Algebra Appl. **367** (2003) 315–335.

[12] B. Sturmfels, *Solving Systems of Polynomial Equations, CBMS Regional Conference Series in Mathematics*, vol. **97**, Amer. Math. Soc., Providence R.I., 2002.

[13] N. Bacaer, *Modéles mathématiques pour l'optimisation des rotations,* Comptes Rendus de l'Académie d'Agriculture de France **89** (3) (2003) 52.

[14] G. Cohen, D. Dubois, J.-P. Quadrat, M. Viot, A linear system theoretic view of discrete event processes and its use for performance evaluation in manufacturing, IEEE Trans. Automat. Contr. AC-**30** (1985) 210–220.

[15] F. Baccelli, G. Cohen, G. Olsder, J. Quadrat, *Synchronization and Linearity — an Algebra for Discrete Event Systems*, John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, 1992.
[16] M. Akian, R. Bapat, S. Gaubert, Max-plus algebras, *Handbook of Linear Algebra*, Discrete Mathematics and its Applications Chapman & Hall/CRC, Boca Raton, FL, 2006.
[17] M. Akian, S. Gaubert, *Spectral theorems for convex monotone homogeneous maps, and ergodic control*, Nonlinear Anal. 52 (2) (2003) 637–679.
[18] R.E. Burkard, P. Butkovič, Max algebra and the linear assignment problem, Math. Program. Ser. B **98** (2003) 415–429.
[19] M. Gondran, M. Minoux, *Graphs and Algorithms*, Wiley-Interscience, New York, 1984.
[20] M. Develin, B. Sturmfels, Tropical convexity, *Doc. Math.* **9** (2004) 1–27.
[21] G. Mikhalkin, Real algebraic curves, the moment map and amoebas, Ann. Math. **151** (2) (2000) 309–326.
[22] O. Viro, Dequantization of real algebraic geometry on logarithmic paper, European Congress of Mathematics, vol. I (Barcelona, 2000), Progr. Math., vol. **201**, Birkhäuser, Basel, 2001, pp. 135–146.
[23] K.H. Kim, *Boolean Matrix Theory and Applications*, Pure Appl. Math., vol. **70**, Marcel Dekker, New York, 1982.
[24] L.B. Beasley, A.E. Guterman, S.-G. Lee, S.-Z. Song, Frobenius and Dieudonné theorems over semirings, Linear Multi. Algebra **55** (1) (2007) 19–34.
[25] L.B. Beasley, A.E. Guterman, S.-G. Lee, S.-Z. Song, Linear preservers of zeros of matrix polynomials, Linear Algebra Appl. **401** (2005) 325–340.
[26] L.B. Beasley, A.E. Guterman, S.-G. Lee, S.-Z. Song, Linear transformations preserving the Grassmannian over $M_n(\mathbf{Z}+)$, Linear Algebra Appl. **393** (2004) 39–46.
[27] R.A. Cuninghame-Green, P. Butkovič, Bases in max-algebra, Linear Algebra Appl. **389** (2004) 107–120.
[28] M. Develin, F. Santos, B. Sturmfels, On the rank of a tropical matrix, in: E. Goodman, J. Pach, E. Welzl (Eds.), *Discrete and Computational Geometry*, MSRI Publications, Cambridge University Press, 2005.
[29] A.E. Guterman, Transformations of non-negative integer matrices preserving the determinant, Uspehi Mat. Nauk **58** (6) (2003) 147–148 (in Russian).
[30] P.L. Poplin, R.E. Hartwig, Determinantal identities over commutative semirings, Linear Algebra Appl. **387** (2004) 99–132.
[31] J. Kuntzmann, *Théorie des réseaux (graphes)*, Dunod, Paris, 1972.
[32] S. Ghosh, Matrices over semirings, *Inf. Sci.* **90** (1996) 221–230.
[33] M. Gondran, M. Minoux, Linear algebra in diods: a survey of recent results, Ann. Discrete Math. **19** (1984) 147–164.
[34] C. Reutenauer, H. Straubing, Inversion of matrices over a commutative semiring, J. Algebra **88** (1984) 350–360.
[35] J.S. Golan, *Semirings and Affine Equations over them: Theory and Applications*, Math. Appl., vol. **556**, Kluwer Academic Publishers, Dordrecht, 2003.
[36] S. Gaubert, P. Butkovic, Sign-nonsingular matrices and matrices with unbalanced determinant in symmetrised semirings, Linear Algebra Appl. **301** (1999) 195–201.
[37] W.C. Brown, *Matrices Over Commutative Rings*, Marcel Dekker inc., New York, 1993.
[38] S.N. Il'in, *Invertible matrices over non-associative rings*, Universal Algebra Appl., Peremena, Volgograd, 2000, pp. 81–89 (in Russian).
[39] S. Pierce, et al., A survey of linear preserver problems, Linear and Multilinear Algebra **33** (1992) 1–119.
[40] L.B. Beasley, A.E. Guterman, Y.-B. Jun, S.-Z. Song, Linear preservers of extremes of rank inequalities over semirings: row and column ranks, Linear Algebra Appl. **413** (2006) 495–509.
[41] Z. Izhakian, *The tropical rank of a tropical matrix*, arXiv:math.AC/0604208.
[42] M. Akian, S. Gaubert, A.E. Guterman, *Rank functions for matrices over max-algebras*, Contemp. Math. Amer., Math. Soc. Providence, RI, to appear, arXiv:0812.3496.
[43] M. Akian, S. Gaubert, A.E. Guterman, *Tropical linear independence and mean payoff games*, Preprint, 2008.

[44] L.B. Beasley, A.E. Guterman, Linear preservers of extremes of rank inequalities over semirings: The factor rank, J. Math. Sci. **131** (5) (2005) 5919–5938.
[45] L.B. Beasley, A.E. Guterman, Rank inequalities over semirings, J. Korean Math. Soc. **42** (2) (2005) 223–241.
[46] L.B. Beasley, A.E. Guterman, S.-C. Yi, Linear preservers of extremes of rank inequalities over semirings: term rank and zero-term rank, J. Math. Sci. **137** (6) (2006) 5179–5191.
[47] L.B. Beasley, S.J. Kirkland, B.L. Shader, Rank Comparisons, Linear Algebra Appl. **221** (1995) 171–188.
[48] L.B. Beasley, C.-K. Li, S. Pierce, Miscellaneous preserver problems, Linear Mult. Algebra **33** (1992) 109–119.
[49] L.B. Beasley, N.J. Pullman, Semiring rank versus column rank, Linear Algebra Appl. **101** (1988) 33–48.
[50] D.A. Gregory, N.J. Pullman, Semiring rank: Boolean rank and nonnegative rank factorization, J. Combin. Inf. Syst. Sci. **8** (1983) 223–233.
[51] S.-G Huang, S.-Z. Song, Spanning column ranks and theiz preservers of nonnegative matrices, Linear Algebra Appl. **254** (1997) 485–495.
[52] O.A. Pshenitsyna, Factor and term ranks of matrix union over semirings, Fundam. Appl. Math. **9** (3) (2003) 175–197.
[53] V.L. Watts, Boolean rank of Kronecker products, Linear Algebra Appl. **336** (2001) 261–264.
[54] J.E. Cohen, U.G. Rothblum, Nonnegative ranks, decompositions, and factorizations of nonnegative matrices, Linear Algebra Appl. **190** (1993) 149–168.
[55] L. Henry, Nuptiality, Theor. Popul. Biol. **3** (1972) 135–152.
[56] M. Yannakakis, Expressing combinatorial optimization problems by linear programs, Proc. 20th Ann. ACM Symp. the Theory Comput. (1998) 223–228.
[57] R. Brualdi, H. Ryser, *Combinatorial Matrix Theory*, Cambridge University Press, Cambridge, 1991.
[58] V.N. Sachkov, V.E. Tarakanov, *Combinatorics of Nonnegative Matrices*, Prog. Theor. Appl. Math., vol. **2**, Moscow, TVP, 2000.
[59] H. Minc, *Permanents*, Addison-Wesley, Reading, Mass, 1978.
[60] K.H. Kim, F.W. Roush, *Kapranov rank vs. tropical rank*, arXiv: math.CO/0503044 v2.
[61] D. deCaen, D. Gregory, N. Pullman, The Boolean rank of zero one matrices, Proc. Third Caribb. Conf. Combin. Comput., Barbados (1981) 169–173.

Quadratic Forms

Kazimierz Szymiczek

University of Silesia, Katowice, Poland
E-mail: szymiczek@math.us.edu.pl

Contents

HANDBOOK OF ALGEBRA, VOL. 6
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(08)00202-7

1. Introduction

Quadratic forms constitute a large domain of research, with roots in classical mathematics and remarkable developments in the last decades. Its origins go back to Fermat and Euler, and at the times of Gauss there existed already a deep theory of quadratic forms with integer coefficients. A new stimulus was provided by two of the Hilbert problems (11th and 17th) on quadratic forms resolved in the 1920s by Hasse and Artin. Together with the important 1937 paper by Witt [240], this laid the foundation for the development of the algebraic theory of quadratic forms with the leading role played by the theory of quadratic forms over fields of characteristic different from two. This survey concentrates on that motor force behind the spectacular developments we have witnessed lately in the subject of quadratic forms.

Sections 2–4 contain introductory material. Here, we set up the notation, terminology, recall the fundamentals of the theory, and provide motivation for the material discussed in the next sections. In Section 5, we discuss briefly the Milnor Conjecture and its solution by Voevodsky. In Section 6, the classification results for quadratic forms and Witt rings are presented. While classification of quadratic forms up to equivalence is an obvious central problem of the theory, its solution does not suffice to determine the structure of the Witt ring. We consider the classification of Witt rings over a field F to be the ultimate goal of the theory and report on the results over those fields where the theory is complete.

Section 7 discusses selected applications of function fields of quadratic forms. This technique brings to quadratic form theory the methods of algebraic geometry, which proved to be extremely important and successful. The subsection on the $u-$invariant shows the strength of the function fields compared with the earlier more elementary and direct approach.

Section 8 discusses the properties of particular quadratic forms, the sums of squares. Here, three most important themes are brought up which are as follows: the level of a field or ring, the composition of sums of squares, and the Pythagoras number of fields and rings.

In Section 9, we briefly discuss the most natural generalization of the original Witt ring, the Witt ring of a commutative ring. As a vehicle, we have chosen two problems. First, for an integral domain R and its field of fractions F, we report on the results on the natural Witt ring homomorphism $W(R) \to W(F)$. Second, for a Dedekind domain R, there is an exact Knebusch–Milnor sequence

$$0 \to W(R) \xrightarrow{i} W(F) \xrightarrow{\partial} \coprod_{\mathfrak{p}} W(R/\mathfrak{p}),$$

and the problem we discuss is the computation of the cokernel of ∂.

In Section 10, we describe the abstract approaches to quadratic form theory that emerged from the work on the classification of Witt rings and from the attempts to generalize Witt rings to the context of semilocal rings. We explain the most important unsolved problem of the abstract theory, the elementary type conjecture. Its solution would give a full classification of Witt rings of fields with finite number of square classes. We also present the basic result on the structure of reduced Witt rings.

In Section 11, we sketch the developments leading to the Witt's 1937 paper, beginning with Minkowski and Hilbert.

While the literature on algebraic theory of quadratic forms is huge and grows steadily, there are several books synthesizing the work in the area. Of these, we mention only three: Scharlau [204] covering Hermitian forms with an extensive literature up to 1982; Lam [143], covering almost all areas of quadratic form theory over fields of characteristic $\neq 2$; and Elman, Karpenko, and Merkurjev [65], a new important book written from the viewpoint of algebraic geometry (not available at the time of writing this chapter).

In this survey, we do not discuss the areas of research that properly belong to quadratic form theory over domains other than fields, such as

- classical theory of quadratic forms over rings of integers in an algebraic number field, especially over $\mathbb{Z}$, for this see [11, 37, 83, 114, 125, 174, 181], and the literature cited there in,
- quadratic forms over algebraic geometric varieties, see [3, 7, 33, 120, 180, 186, 187, 223]
- quadratic and Hermitian forms over general rings, polynomial rings, and affine algebras [126, 204].

We also do not discuss the areas of research close to quadratic form theory and naturally related and motivated by the developments in quadratic form theory, such as

- ordered fields, preorderings, and valuation theory [141, 142]
- central simple algebras, linear algebraic groups [107]
- Galois theory [1, 82, 104, 105, 140, 175, 176]
- real commutative algebra, semialgebraic geometry, and algebraic topology [63, 106, 163, 198, 207]
- higher level orderings of fields, Witt rings of higher level [17, 20, 21, 185]
- forms of higher degree, higher levels, and higher Pythagoras numbers [18, 29, 85, 166, 203].

2. Quadratic forms over fields

In this section, we consider quadratic forms and bilinear forms over fields of characteristic different from 2. Given a finite dimensional vector space V over the field F, a quadratic form q on V is a function $q : V \to F$ such that the associated function $b_q : V \times V \to F$ defined by $b_q(u, v) = \frac{1}{2}(q(u + v) - q(u) - q(v))$ is bilinear and $q(av) = a^2 q(v)$ for all $a \in F$ and $v \in V$. The pair (V, q) is then said to be a *quadratic space* and (V, b_q) a *bilinear space* over the field F. Two vectors $u, v \in V$ are said to be orthogonal if $b_q(u, v) = 0$.

Two quadratic spaces (V_1, q_1) and (V_2, q_2) over F are said to be *isometric* when there is a vector space isomorphism $\varphi : V_1 \to V_2$ satisfying $q_2(\varphi(v)) = q_1(v)$ for all $v \in V_1$. The two quadratic forms q_1 and q_2 are then said to be *equivalent*, written $q_1 \cong q_2$. For a quadratic form q on V, the set of nonzero values of the form q is

denoted as $D_F(q)$. The elements in $D_F(q)$ are said to be *represented* by q over F. Since $q(av) = a^2 q(v)$ for $a \in F$, it follows that the set $D_F(q)$ consists of whole cosets of the multiplicative group $\dot{F}$ modulo the subgroup $\dot{F}^2$ of squares. Hence, $D_F(q)$ can be viewed as a subset of the group $\dot{F}/\dot{F}^2$ of square classes of F.

For a quadratic space (V, q), the dimension of V is said to be the dimension of the quadratic form q, written as $\dim q$. If $\mathcal{B} = \{v_1, \ldots, v_n\}$ is a basis for V, the matrix $B = [b_q(v_i, v_j)]$ is said to be the matrix of q with respect to $\mathcal{B}$. Distinct bases $\mathcal{B}_1$ and $\mathcal{B}_2$ produce *congruent* matrices B_1 and $B_2 = PB_1P^T$, where P is a nonsingular matrix. Hence, $\det B_1$ and $\det B_2$, if nonzero, lie in the same coset $(\det B)\dot{F}^2$, which is said to be the *determinant* of the form q (or of the space (V, q)), written as $\det q$. If $\det B = 0$ for some basis $\mathcal{B}$, we set $\det q = 0$. Forms with nonzero determinant are called *nonsingular*. This is equivalent to the fact that the map

$$V \to V^*, \quad v \mapsto b_q(\ , v)$$

is an isomorphism of V onto the dual space V^*. Each quadratic form q over a field of characteristic not 2 can be *diagonalized*. That is, there is a basis $\mathcal{B}$ for V such that the matrix of q with respect to $\mathcal{B}$ is a diagonal matrix. In this case, $\mathcal{B}$ consists of pairwise orthogonal vectors. We often identify quadratic forms with their diagonal matrices.

If (V_1, q_1) and (V_2, q_2) are quadratic spaces, then so is $(V_1 \oplus V_2, q_1 \perp q_2)$, where $(q_1 \perp q_2)(v_1, v_2) = q_1(v_1) + q_2(v_2)$. This is called the *orthogonal direct sum* of forms q_1 and q_2. Also, the tensor product $V_1 \otimes V_2$ can be equipped with the structure of a quadratic space $(V_1 \otimes V_2, q)$, where the associated bilinear form b_q equals the tensor product $b_{q_1} \otimes b_{q_2}$. Hence,

$$b_q(v_1 \otimes v_2, v_1' \otimes v_2') = b_{q_1}(v_1, v_1') \cdot b_{q_2}(v_2, v_2')$$

for all simple tensors $v_1 \otimes v_2, v_1' \otimes v_2'$ in $V_1 \otimes V_2$. The form q is then called the *tensor product* of the quadratic forms q_1 and q_2, written as $q_1 \otimes q_2$. Direct orthogonal sum and tensor product of nonsingular quadratic forms are likewise nonsingular. If $q_1 = (a_1, \ldots, a_n)$ and $q_2 = (b_1, \ldots, b_m)$ are diagonalized forms, then

$$q_1 \perp q_2 = (a_1, \ldots, a_n, b_1, \ldots, b_m), \ \ q_1 \otimes q_2 = (a_1b_1, \ldots, a_nb_1, a_2b_1, \ldots, a_nb_m).$$

The very useful Witt cancellation theorem states that if q is nonsingular and $q \perp q_1 \cong q \perp q_2$, then $q_1 \cong q_2$.

A quadratic form $(1, a)$, $a \in \dot{F}$, is called one-fold Pfister form, and an n-fold tensor product $(1, a_1) \otimes \cdots \otimes (1, a_n)$ of one-fold Pfister form is called an n-fold Pfister form. A quadratic form q is said to be *isotropic* if there exists a nonzero vector $v \in V$ so that $q(v) = 0$. A simple but fundamental example of a nonsingular isotropic form is the *hyperbolic plane*. This is the two-dimensional form h with diagonalization $(1, -1)$ in some basis of the plane. If a quadratic form q is isotropic, then it splits off a hyperbolic plane h_1, i.e., $q \cong h_1 \perp q_1$ for some quadratic form q_1. Continuing with q_1, we ultimately obtain a decomposition

$$q \cong h_1 \perp \cdots \perp h_i \perp q_a,$$

where $h_1, \ldots, h_i$ are hyperbolic planes and q_a is *anisotropic*. By Witt cancellation, the form q_a, called the anisotropic part of q, is unique up to isometry, and also, the

number i (called the Witt index of q) is unique. If $q_a = 0$, the form q is said to be *hyperbolic*. It is an important property of Pfister forms that if a Pfister form is isotropic, then it is necessarily hyperbolic. Another fundamental property is that for a Pfister form q, the value set $D_F(q)$ is a group under multiplication. This implies immediately that the level of a nonformally real field is always a power of two (see Section 8.1). Here, the level $s(A)$ of a ring A is the minimal number of summands in a representation of $-1 \in A$ as a sum of squares of elements in A (or ∞ if such a representation does not exist).

Two quadratic forms q and g over F are said to be *similar* (or Witt equivalent), written as $q \sim g$, if their anisotropic parts are isometric, $q_a \cong g_a$. An easy observation is that for quadratic forms q and g over F,

$$\dim q = \dim g \quad \text{and} \quad q \sim g \Rightarrow q \cong g. \tag{2.1}$$

Two fundamental problems of quadratic form theory are the representation problem and the classification problem. The representation problem over a field F asks for an explicit determination of the elements represented by any quadratic form over F. The classification problem over a field F asks for criteria for deciding whether any given quadratic forms over F are equivalent or not.

If the representation problem is solved over a field F, the equivalence problem can be solved by induction on dimension. Indeed, for two forms q and g of the same dimension n, take any element $a \in D_F(q)$ and check whether $a \in D_F(g)$ or not. If a is not represented by g, the forms q and g are not equivalent. If a is represented by g, the forms q and g can be diagonalized with the element a occurring on the diagonal. Hence, $q = (a) \perp q_1$ and $g = (a) \perp g_1$ for some $(n-1)$-dimensional forms q_1, g_1, and by the Witt cancellation theorem, $q \cong g \Longleftrightarrow q_1 \cong g_1$. Thus, the question is reduced to a lower dimension.

Another elementary observation is that the representation problem is equivalent to the *isotropy* problem: to decide whether or not a given quadratic form is isotropic. For $a \in D_F(q) \Longleftrightarrow q \perp (-a)$ is isotropic.

3. Witt rings of fields

We continue to assume that F is a field of characteristic different from 2. For a nonsingular quadratic form q over F, we write $\langle q \rangle$ for the class of quadratic forms similar to q. This is the Witt class of q. Witt classes of nonsingular quadratic forms over F with addition and multiplication induced by direct orthogonal sum and tensor product form a commutative ring called the Witt ring of the field F and denoted as $W(F)$.[1] When the characteristic of F equals 2, a similar construction produces the Witt ring of nonsingular *bilinear forms* over F, also denoted as $W(F)$. Below we refer to this ring when we allow characteristic 2.

[1] The same notation, $W(A)$, is used for the ring of big Witt vectors over a ring or field A, a totally different concept. See Michiel Hazewinkel, *Witt vectors*. Part 1, this volume.

The Witt ring has the following description in terms of generators and relations. Consider the integral group ring $\mathbb{Z}[\dot{F}/\dot{F}^2]$ of the group of square classes of the field F. Then,

$$W(F) = \mathbb{Z}[\dot{F}/\dot{F}^2]/J,$$

where J is the ideal in the group ring $\mathbb{Z}[\dot{F}/\dot{F}^2]$ generated by the set,

$$\{[a] + [b] - [c] - [d] : (a, b) \cong (c, d)\} \cup \{[1] + [-1]\}.$$

This is a straightforward consequence of the Witt theorem on chain equivalence of quadratic forms (see [240, Satz 7]).

For a Witt class $\langle q\rangle$, we define the *dimension-index*

$$e\langle q\rangle = \dim q \pmod 2 \in \mathbb{Z}/2\mathbb{Z}$$

and the *discriminant*

$$d\langle q\rangle = (-1)^{\frac{1}{2}n(n-1)} \det q \in \dot{F}/\dot{F}^2,$$

where $n = \dim q$. These are well-defined invariants of the similarity class. Moreover,

$$e : W(F) \to \mathbb{Z}/2\mathbb{Z}$$

is a ring epimorphism. Its kernel consists of the Witt classes $\langle q\rangle$ with even-dimensional q and is said to be the *fundamental ideal* of the Witt ring $W(F)$, written as $I(F)$. Further, $d : W(F) \to \dot{F}/\dot{F}^2$ is a well-defined map, but it fails to be a homomorphism of the additive group $W(F)$ (for instance, take $F = \mathbb{R}$). However, if we restrict the discriminant map to the fundamental ideal $I(F)$, we obtain a surjective group homomorphism

$$d : I(F) \to \dot{F}/\dot{F}^2$$

of the additive group $I(F)$ onto the group of square classes. Interestingly enough, the kernel of this homomorphism equals $I^2(F)$, the additive group of the square of the fundamental ideal. This prompts looking at higher powers $I^n(F)$ of the fundamental ideal. An elementary observation is that $I^n(F)$ is generated as an Abelian group by the Witt classes of all n-fold Pfister forms over F. Much deeper and difficult to prove is the following theorem known as the Hauptsatz. It was proved by Arason and Pfister [5] in 1971.

THEOREM 3.1 (Arason–Pfister Hauptsatz). *For a positive-dimensional anisotropic quadratic form q over a field F, if $\langle q\rangle \in I^n(F)$, then* $\dim q \geq 2^n$.

As a consequence, we get the intersection property in the Witt ring $W(F)$ of the field F,

$$\bigcap_{n=0}^{\infty} I^n(F) = 0.$$

3.1. *Prime ideals*

Prime ideals of the Witt ring are described as follows (Harrison [84], Lorenz and Leicht [159]).

THEOREM 3.2. *Let F be a field of characteristic not* 2.

(a) *If the Witt ring $W(F)$ has a prime ideal $\mathfrak{p} \neq I(F)$, then the field F is formally real and the set*

$$P := \{\, a \in \dot{F} : \langle 1, -a\rangle \in \mathfrak{p} \,\}$$

defines an ordering of the field F.

(b) *Let F be a formally real field and let P be an ordering of F. Let $\mathfrak{p}_0$ be the ideal of the Witt ring $W(F)$ generated by the set*

$$\{\, \langle 1, -a\rangle \in W(F) : a \in P \,\}.$$

Then, $\mathfrak{p}_0$ is a minimal prime ideal of the Witt ring $W(F)$.
Moreover, $\mathfrak{p}_0 \subset I(F)$ and $\mathfrak{p}_0 \neq I(F)$.

Hence, MinSpec $W(F) = \{I(F)\}$ is a singleton set when the field F is nonreal, and MinSpec $W(F)$ contains at least one nonmaximal minimal prime ideal when the field F is formally real.

Each ordering P of a formally real field F gives rise to a signature homomorphism

$$\operatorname{sgn}_P : W(F) \to \mathbb{Z}$$

sending the class $\langle q\rangle$ to the signature of the form q for the ordering P. The map

$$\sigma : X(F) \to \text{MinSpec } W(F), \quad P \mapsto \ker \operatorname{sgn}_P$$

is a bijective correspondence between the set $X(F)$ of all orderings of the field F and the minimal prime ideals of the Witt ring $W(F)$.
The inverse map for σ is the map π defined as follows:

$$\pi : \text{MinSpec } W(F) \to X(F), \quad \mathfrak{p} \mapsto P = \{a \in \dot{F} : \langle 1, -a\rangle \in \mathfrak{p}\}.$$

The set $X(F)$ can be given the induced Zariski topology from the prime spectrum and this turns $X(F)$ into a Boolean space (compact, Hausdorff, and totally disconnected). Craven [42] proved that every Boolean space can be obtained as the space of orderings of some field. For a realization of Boolean spaces as spaces of orderings of higher level, see Osiak [183, 184].

3.2. *Nilradical, torsion, divisors of zero, and units*

For a nilpotent element $x \in W(F)$, we obviously have $e(x) = 0$ and hence Nil $W(F) \subseteq I(F)$. When F is non-formally real (nonreal, for short) field, we actually have Nil $W(F) = I(F)$. When F is formally real, we observe that nilpotent elements

in $W(F)$ have zero signatures at every ordering of F, and so the nilradical Nil $W(F)$ is contained in the intersection of the kernels of all signature homomorphisms. Since the latter are all minimal prime ideals in $W(F)$, their intersection actually equals the nilradical Nil $W(F)$.

For a nonreal field F of level s, the unit element $\langle 1\rangle \in W(F)$ has finite order $2s$ in the additive group $W(F)$ and so $2sW(F) = 0$. This is an immediate consequence of the theory of Pfister forms. Thus, $W(F)$ is torsion group and every element is two-primary torsion. When F is formally real, torsion elements have zero signatures at all orderings of the field, so it follows that Tors $W(F) \subseteq$ Nil $W(F)$. A deeper result says that, in fact, for every formally real field F,

$$\text{Tors } W(F) = \text{Nil } W(F).$$

This is a consequence of the local–global principle for torsion elements of Witt rings first proved by Pfister [191].

THEOREM 3.3 (Pfister local–global principle). *Let F be a formally real field and let q be an anisotropic quadratic form over F. For every ordering P of the field F, choose a real closure F_P inducing the ordering P on F. The following statements are equivalent.*

(a) $\langle q\rangle$ *is a torsion element of the Witt ring* $W(F)$.
(b) $\langle q\rangle_{F_P} = 0$ *in the Witt ring* $W(F_P)$, *for every ordering* P *of the field* F.
(c) $\operatorname{sgn}_P\langle q\rangle = 0$ *for all orderings* P *of the field* F.

Moreover, every torsion element in $W(F)$ is two-primary torsion.

A field F is said to be Pythagorean when every sum of squares of nonzero elements of F is a square of a nonzero element in F. Thus, a Pythagorean field is automatically formally real. For all fields F except for Pythagorean fields, the set $\mathrm{ZD}(W(F))$ of zero divisors of the Witt ring $W(F)$ coincides with the ideal $I(F)$. And for a Pythagorean field F, an element of the Witt ring of F is a zero divisor in $W(F)$ if and only if it lies in a minimal prime ideal of the ring $W(F)$. Hence for a Pythagorean field F,

$$\mathrm{ZD}(W(F)) = \bigcup\{\ker \operatorname{sgn}_P : P \in X(F)\}.$$

The following theorem gives an important characterization of Pythagorean fields in terms of their Witt rings.

THEOREM 3.4. *For a formally real field F,*

$$F \text{ is Pythagorean} \iff \text{Tors } W(F) = 0 \iff \text{Nil } W(F) = 0.$$

This theorem implies a solution of the classification problem for quadratic forms over Pythagorean fields. If q and g are nonsingular quadratic forms over a Pythagorean field F, then $q \cong g$ if and only if $\dim q = \dim g$ and $\operatorname{sgn}_P q = \operatorname{sgn}_P g$ for every ordering P of the field F.

The units $\mathrm{U}(W(F))$ of the Witt ring $W(F)$ can be determined as follows.

THEOREM 3.5. *If the field F is nonreal, then* $\mathrm{U}(W(F)) = 1 + I(F)$.
If F is formally real, then $\langle q\rangle \in \mathrm{U}(W(F))$ *if and only if* $\operatorname{sgn}_P\langle q\rangle = \pm 1$ *for all orderings P of F.*

Finally, for any field F, 0 and 1 are the only idempotents of the Witt ring $W(F)$.

The proofs of the results in this and preceding sections can be found in any standard text on algebraic theory of quadratic forms (for instance, [143, 204, 228]).

4. Hasse and Witt invariants

For a nonsingular diagonal quadratic form $q = (a_1, \ldots, a_n)$ with the entries a_i in a field F of characteristic $\neq 2$, we define the *Hasse algebra* $\mathrm{H}(q)$ of the form q as the following tensor product of quaternion algebras:

$$\mathrm{H}(q) := \bigotimes_{1\le i\le j\le n} \left(\frac{a_i, a_j}{F}\right).$$

Here, for $a, b \in \dot{F}$, the symbol $(\frac{a,b}{F})$ denotes the quaternion algebra over F with structure constants a, b. This is a four-dimensional $F-$algebra with a basis $1, i, j, k$ and the multiplication rules $i^2 = a$, $j^2 = b$, $ij = k = -ji$.
If $q = (a_1, \ldots, a_n)$ and $g = (b_1, \ldots, b_n)$ are equivalent quadratic forms, then $\mathrm{H}(q) \cong \mathrm{H}(g)$. In other words, the Hasse algebra of a quadratic form q is uniquely determined up to algebra isomorphism and is an equivalence invariant. To calculate the Hasse algebra one uses the following observation

$$\mathrm{H}(q \perp g) \cong \mathrm{H}(q) \otimes \mathrm{H}(g) \otimes \left(\frac{\det q, \det g}{F}\right) \otimes M_\ell(F), \tag{4.1}$$

where ℓ is an appropriate positive integer. The formula (4.1) suggests that it is suitable to pass from the Hasse algebra to the similarity class of the algebra in the Brauer group $\mathrm{Br}(F)$ of the field F.

For a nonsingular quadratic form $q = (a_1, \ldots, a_n)$ over a field F of characteristic $\neq 2$, the *Hasse invariant* $\mathrm{h}(q)$ of the form q is defined to be the similarity class of the Hasse algebra $\mathrm{H}(q)$ of the form q in the Brauer group of the field F :

$$\mathrm{h}(q) := [\mathrm{H}(q)] = \prod_{1\le i\le j\le n} [a_i, a_j] \in \mathrm{Br}(F).$$

Here $[a_i, a_j]$ stands for the similarity class of the quaternion algebra $\left(\frac{a_i,a_j}{F}\right)$ in the Brauer group of the field F.

THEOREM 4.1 (Classification theorem for quadratic forms of dimension ≤ 3). *Let q and g be nonsingular quadratic forms over a field F of characteristic $\neq 2$ and let* $\dim q \le 3$ *and* $\dim g \le 3$. *Then, we have*

$$q \cong g \quad \Leftrightarrow \quad \dim q = \dim g, \ \ \det q = \det g, \ \ \mathrm{h}(q) = \mathrm{h}(g).$$

The Hasse invariant, like the determinant, is an invariant for equivalence but not for similarity of quadratic forms. However, as in the case of the determinant, one can correct this by multiplying the Hasse invariant with a suitable factor. This works in all dimensions, but we shall restrict attention to even-dimensional forms. We define

$$\mathrm{w}(q) := \mathrm{h}(q) \cdot [-1,-1]^{n(n+1)/2}, \quad \text{where } \dim q = 2n. \tag{4.2}$$

It turns out that the new function is a similarity invariant on the set of even-dimensional forms with discriminant $1 \in \dot{F}/\dot{F}^2$ and this set is precisely $I^2(F)$. Hence for $\langle q\rangle \in I^2(F)$, we set $\mathrm{w}(\langle q\rangle) := \mathrm{w}(q) \in \mathrm{Br}(F)$, and call it the *Witt invariant* of the form q, and of the class $\langle q\rangle \in I^2(F)$.
Thus,

$$\mathrm{w}(q) = \begin{cases} \mathrm{h}(q)\cdot[-1,-1] & \text{when } 2n \equiv 2,4 \bmod 8, \\ \mathrm{h}(q) & \text{when } 2n \equiv 0,6 \bmod 8. \end{cases}$$

Notice that $\mathrm{w}(q) \in \mathrm{Quat}\, F$, the subgroup of $\mathrm{Br}(F)$ generated by classes of quaternion algebras. Since

$$\mathrm{w}(x,-ax,-bx,abx) = \mathrm{w}(1,-a,-b,ab) = [a,b],$$
$$\mathrm{w}((1,-a)\otimes(1,-b)\otimes(1,-c)) = 1,$$

for all $x, a, b, c \in \dot{F}$, it easily follows that the map $\mathrm{w} : I^2(F) \to \mathrm{Quat}\, F$ is a group epimorphism and

$$I^3(F) \subseteq \ker \mathrm{w}.$$

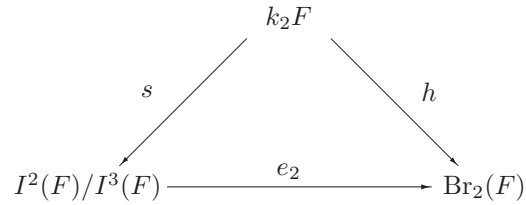
It has been of central importance for quadratic form theory whether actually $I^3(F) = \ker \mathrm{w}$. This was proved for various classes of fields F including local and global fields and fields of transcendence degree ≤ 1 over a real closed field.

A more general approach presented in [69] showed that $I^3(F) = \ker \mathrm{w}$ for all 1-amenable fields. Here F is said to be 1-amenable if for any multiquadratic extension K of F, i.e. $K = F(\sqrt{a_1}, \cdots, \sqrt{a_r})$, $a_i \in \dot{F}$, the kernel $\ker(W(F) \to W(K))$ of the natural ring homomorphism $W(F) \to W(K)$ is generated as an ideal by the one-fold Pfister forms $\langle 1, -a_i\rangle$. A field extension K/F is said to be excellent if, for any quadratic form q over F, the anisotropic part of $q \otimes K$ is defined over F. F is 1-amenable if every multiquadratic extension is excellent. Excellence and amenability were investigated in [70, 71], but then the first examples of fields, which are not 1-amenable were found in [72], thus showing that $I^3(F) = \ker \mathrm{w}$ cannot be deduced in general from 1-amenability. Further, counterexamples to amenability and excellence were given in [149, 210–212].

The question whether $I^3(F) = \ker \mathrm{w}$ was resolved by Merkurjev [168] in 1981 as a part of an even more spectacular result discussed below.

We shall use the Milnor K-theory group K_2F and its factor group, the elementary Abelian two-group $k_2F := K_2F/2K_2F$ generated by the cosets of *symbols* $\{a, b\}$ with $a, b \in \dot{F}$. The correspondence $\{a, b\} \mapsto (\frac{a,b}{F})$ induces a group homomorphism $h : k_2F \to \mathrm{Br}_2(F)$, where $\mathrm{Br}_2(F)$ is the subgroup of the Brauer group $\mathrm{Br}(F)$ consisting of elements of order ≤ 2. However, the correspondence $\{a, b\} \mapsto \langle 1, -a, -b, ab\rangle$

induces a group homomorphism $s : k_2F \to I^2(F)/I^3(F)$, and the Witt invariant induces a homomorphism $e_2 : I^2(F)/I^3(F) \to \text{Quat } F \subseteq \text{Br}_2(F)$. These homomorphisms yield the commutative diagram



It was proved by Milnor [173] that s is an isomorphism. Merkurjev's contribution is the following important result.

THEOREM 4.2 (Merkurjev [168]). *The homomorphism h is an isomorphism for all fields F of characteristic $\neq 2$.*

It follows that the homomorphism e_2 is also an isomorphism and consequently $I^3(F) = \ker \text{w}$. Thus, we get a complete characterization of quadratic forms in $I^3(F)$: these are even dimensional forms with trivial discriminant and trivial Witt invariant. Another consequence is an improvement on the classification theorem 4.1 contained in Theorem 6.2 below.

Merkurjev's theorem shows that $\text{Quat } F = h(k_2F) = \text{Br}_2(F)$, which solves an old problem in the theory of algebras (any element of order 2 in the Brauer group $B(F)$ is expressible as a product of classes of quaternion algebras, or, equivalently, any central division algebra with an involution over F is similar to a tensor product of quaternion algebras). Merkurjev's proof that $h : k_2F \to \text{Br}_2(F)$ is an isomorphism depended on a result of Suslin on the quadratic norm residue symbol available at that time only through deep results in Quillen's K-theory. New proofs were given subsequently by Arason [4], Wadsworth [238], and Merkurjev [171] and they do not depend on Quillen's K-theory.

While the setup of Merkurjev's theorem can be generalized to commutative rings, there is, however, no analog to Merkurjev's theorem. Parimala and Sridharan have given in [188] an example of a commutative ring R for which the Clifford algebra map (corresponding to the Witt invariant w) from strictly even rank quadratic forms over R to $\text{Br}_2(R)$ is not surjective.

5. Milnor's Conjecture

In a paper published in 1970, Milnor [173] observed interrelations among Witt rings, K-theory, and Galois cohomology. For a field F of characteristic not 2, the K-theory groups are Milnor's K_nF and the factor groups $k_nF := K_nF/2K_nF$, and the cohomology groups are the Galois cohomology $H^n(F) = H^n(\text{Gal}\,(\overline{F}/F), \mathbb{Z}/2\mathbb{Z})$, where $\overline{F}$ denotes a separable closure of F. Milnor defined homomorphisms

$$s_n : k_nF \to I^n(F)/I^{n+1}(F)$$

induced by the map sending the pure symbol $\{a_1, \ldots, a_n\} \in K_n F$ onto the Witt class of the Pfister form $(1, -a_1) \otimes \cdots \otimes (1, -a_n) \in I^n(F)$. Since $I^n(F)$ is additively generated by all n-fold Pfister forms, every s_n is surjective, and Milnor proved that s_1 and s_2 are bijective. He also defined homomorphisms

$$h_n : k_n F \to H^n(F)$$

induced by the map sending the pure symbol $\{a_1, \ldots, a_n\} \in K_n F$ onto the cup product $(a_1) \cup \cdots \cup (a_n) \in H^n(F)$.

For $n = 0, 1, 2$, the groups $H^n(F)$ can be identified with $\mathbb{Z}/2\mathbb{Z}$, $\dot{F}/\dot{F}^2$, $\mathrm{Br}_2 F$, and dimension index, discriminant, and Witt invariant can be viewed as surjective group homomorphisms

$$e^n : I^n(F) \to H^n(F)$$

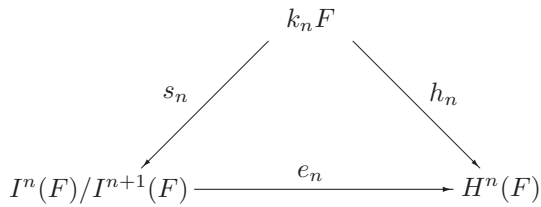
satisfying $\ker e^n = I^{n+1}(F)$ (for $n = 2$, this assumes Merkurjev's theorem). Hence in these cases, we have isomorphisms

$$e_n : I^n(F)/I^{n+1}(F) \to H^n(F).$$

It was an open problem whether such isomorphisms exist for $n \geq 3$.

The existence of these isomorphisms is of importance for the classification of quadratic forms, and so it is highly desirable to extend these existence results for other values of n. In 1975, Arason [2] proved the existence of the group homomorphism e^3, and in 1989 Jacob and Rost [103] and independently Szyjewski [222] proved the existence of e^4.

The homomorphisms s_n, h_n, and the supposed maps e_n combine into the diagram



The Milnor conjecture asserts that s_n and h_n are isomorphisms for all n and all fields F of characteristic not 2. Thus, the difficult and, in fact, intractable question in quadratic form theory about the existence of the invariants e_n has been transferred to K-theory and cohomology theory.

The bijectivity of h_n for all n was proved by Voevodsky in 1996 and the bijectivity of s_n shortly thereafter by Orlov, Vishik and Voevodsky. Hence, for any n and any field F of characteristic not 2, the map

$$e_n = h_n \circ s_n^{-1} : I^n(F)/I^{n+1}(F) \to H^n(F)$$

is an isomorphism sending the cosets $(1, -a_1) \otimes \cdots \otimes (1, -a_n) + I^{n+1}(F)$ of Pfister forms onto the cup products $(a_1) \cup \cdots \cup (a_n) \in H^n(F)$.

The final versions of their papers are now available [182, 236]. There are also several accounts of Voevodsky's work explaining with various degree of detail the proofs of Milnor's conjecture (see, Morel [179]). Therefore, we shall not discuss here the papers [182, 236]. We refer to the survey by Pfister [199] and to the paper by Arason and Elman [6] for consequences of Voevodsky's results in quadratic form theory. We shall mention a few newer developments in the following sections.

6. Classification

The classification of quadratic forms up to equivalence may be considered the central problem of quadratic form theory. By (2.1), this is essentially the same as classification of quadratic forms up to similarity. Hence, solving the classification problem for forms over a field F, we determine the elements of the Witt ring $W(F)$. However, this says nothing about the ring structure of $W(F)$. Hence, it is natural to split the classification problem into two parts. First, classify quadratic forms up to equivalence, and second, classify the Witt rings of fields up to isomorphism.

6.1. *Equivalence of quadratic forms*

The classical invariants of quadratic forms are dimension, determinant, Hasse invariant, and the total signature. The latter is the totality of all signatures sgn_P defined by all orderings P of the field in question. Total signature occurs only for formally real fields. The class of fields for which classical invariants classify quadratic forms was characterized by Elman and Lam [68] as follows.

THEOREM 6.1. *Quadratic forms over a field F are classified by their dimension, determinant, Hasse invariant, and total signature if and only if the ideal $I^3(F)$ of the Witt ring $W(F)$ is torsion-free.*

When the field F is nonreal, the Witt ring $W(F)$ is the torsion so that torsion-freeness of $I^3(F)$ is to be interpreted as $I^3(F) = 0$. Observe that, since $I^3(F)$ is torsion-free for all local and global fields, this theorem recaptures the classical results of Hasse on the classification of quadratic forms over local and global fields (see Section 11 for Hasse's results). Theorem 6.1 was proved before Merkurjev's Theorem 4.2. Using Merkurjev's theorem, the proof would become much easier. A refined version of the proof not using Merkurjev's theorem is given in Lam's book [143, pp. 440–443]. For more general classification results involving the classical invariants, see [16] (Hermitian forms over central simple algebras with involution) and [158] (classification of central simple algebras with involution).

Voevodsky's theorem gives new possibilities for the classification of quadratic forms. When we want to check whether two quadratic forms q, g over a field F of characteristic different from 2 are equivalent, we may assume that $\dim q = \dim g = n$ and consider the class $\varphi = \langle q \perp -g \rangle$. From (2.1), we have $q \cong g \iff \varphi = 0 \in W(F)$.

We shall write e^n for the composition $I^n(F) \to I^n(F)/I^{n+1}(F) \to H^n(F)$ and view the values of the homomorphisms e^n as invariants of quadratic forms in $I^n(F)$. In order to compute $e^n(\varphi)$, we need to know that $\varphi \in I^n(F)$. The latter is equivalent to requiring that $e^i(\varphi) = 0$ for $i = 0, 1, \ldots, n-1$.

THEOREM 6.2 (General classification theorem). *Let q, g be quadratic forms over a field F of characteristic different from 2 and assume that* $\dim q = \dim g = n$. *Set $\varphi = \langle q \perp -g \rangle$, and let k satisfy $2^k \le 2n < 2^{k+1}$.*

If $e^i(\varphi) \neq 0$ for some $i \le k$, then q and g are not equivalent forms.
If $e^i(\varphi) = 0$ for $i = 1, \ldots, k$, then q and g are equivalent forms.

PROOF. The class φ is even-dimensional and hence lies in $I(F)$ and so $e^0(\varphi) = 0$. If $e^1(\varphi) = 0$, we have $\varphi \in I^2(F)$, and if ℓ is the smallest index for which $e^\ell(\varphi) \neq 0$, then necessarily $\varphi \neq 0 \in W(F)$ and so q and g are inequivalent.
If $e^i(\varphi) = 0$ for $i = 1, \ldots, k$, then by Voevodsky's theorem $\varphi \in I^{k+1}(F)$. However, $\dim \varphi = 2n < 2^{k+1}$ implies $\varphi = 0 \in W(F)$ by the Arason–Pfister Hauptsatz, but $\varphi = 0 \in W(F)$ and $\dim q = \dim g$ imply $q \cong g$. □

6.1.1. Trace forms For a finite separable field extension K/F, the trace functional $\mathrm{tr}_{K/F} : K \to F$ defines a symmetric bilinear form

$$K \times K \to F, \quad (x, y) \mapsto \mathrm{tr}_{K/F}(xy).$$

The associated quadratic form is called the *trace form* and denoted $t_{K/F}$. When F is a number field, trace forms $t_{K/F}$ are quadratic forms over F and so are classified up to isometry by the four classical invariants. Here, $\dim t_{K/F} = [K : F]$, $\det t_{K/F} = N_{K/F}(f'(\theta))$, where $K = F(\theta)$ and f is the minimal polynomial of θ over F. The signature $\mathrm{sgn}_P t_{K/F}$ with respect to an ordering P of F equals the number of the extensions of P to orderings of K. In particular, we always have $\mathrm{sgn}_P t_{K/F} \ge 0$. The details may be found in [34]. The Hasse invariant of $t_{K/F}$ was computed by Serre [208] (see also [34, Chapter II]).

A quadratic form q (the Witt class $\langle q \rangle \in W(F)$) over a number field F is said to be positive if the form (the Witt class) has nonnegative signature with respect to every ordering of F. If F has no orderings, then every quadratic form (and every Witt class) over F is positive. As mentioned above, trace forms (and their Witt classes) are positive. Conner and Perlis [34] proved that in the Witt ring $W(\mathbb{Q})$ the converse holds: if a Witt class $X \in W(\mathbb{Q})$ is positive, then there is a finite extension K of $\mathbb{Q}$ such that $\langle t_{K/\mathbb{Q}} \rangle = X$. For an arbitrary number field F as a base field the result, was proved by Scharlau [205].

A characterization of positive quadratic forms up to isometry was found by Krüskemper [131]. He proved that every positive quadratic form of even degree $n \ge 4$ over a number field F is isometric to a trace form. Then, J.-P. Serre (in a letter to Krüskemper) pointed out that, over a number field, every positive form of odd rank $n > 4$ is isometric to a trace form. Serre's proof is based on a result of Mestre [172] concerning deformations of polynomials leaving trace forms intact. Thus, over

a number field F, all positive forms of dimension $n \geq 4$ are trace forms, and for forms of lower dimension, this was known from [34, III 3.6].

For related work on trace forms, see [36, 73–76, 202, 239].

6.2. *Witt equivalence of fields*

Two fields with isomorphic Witt rings are called *Witt equivalent*. The first question is what kind of affinity of behavior of quadratic forms can we expect over two Witt equivalent fields. The following definition describes the ideal situation where quadratic forms over two fields behave in the same way.

DEFINITION 6.3. Let K and L be fields of characteristic $\neq 2$. We say that K and L are *equivalent with respect to quadratic forms*, when there exists a pair of bijective maps

$$t : \dot{K}/\dot{K}^2 \to \dot{L}/\dot{L}^2 \quad \text{and} \quad T : \mathrm{Cl}(K) \to \mathrm{Cl}(L),$$

where $\mathrm{Cl}(K)$ denotes the set of equivalence classes of nonsingular quadratic forms over K, satisfying the following four conditions:

(A) $T(a_1, \ldots, a_n) = (t(a_1), \ldots, t(a_n))$ for all $a_1, \ldots, a_n \in \dot{K}/\dot{K}^2$.
(B) $\det T(q) = t(\det q)$ for every nonsingular quadratic form q over K.
(C) $D_L(T(q)) = t(D_K(q))$ for every nonsingular quadratic form q over K.
(D) $t(1) = 1$ and $t(-1) = -1$.

In 1970, Harrison [84] proved a criterion for Witt equivalence of fields, which combined with a result of Cordes [38] yielding the following theorem.

THEOREM 6.4 (Harrison criterion). *Let K and L be fields of characteristic $\neq 2$. The following statements are equivalent:*

(a) *K and L are equivalent with respect to quadratic forms.*
(b) *There exists a group isomorphism $t : \dot{K}/\dot{K}^2 \to \dot{L}/\dot{L}^2$ such that $t(-1) = -1$, and for all $a, b \in \dot{K}/\dot{K}^2$,*

$$1 \in D_K(a, b) \;\Leftrightarrow\; 1 \in D_L(t(a), t(b)).$$

(c) $W(K) \cong W(L)$.
(d) $W(K)/I^3(K) \cong W(L)/I^3(L)$.

For proof, see [189] or [228, p. 417]. For a characteristic 2 counterpart of Harrison's criterion, see the paper [12].

EXAMPLE 6.1. Quadratically closed fields are all Witt equivalent and their Witt ring is $\mathbb{Z}/2\mathbb{Z}$. Quadratic forms are classified by their dimension.

Real closed fields are all Witt equivalent and their Witt ring is $\mathbb{Z}$. Quadratic forms are classified by dimension and signature.

Finite fields of characteristic different from 2 are Witt equivalent if and only if they have the same level. Here quadratic forms are classified by dimension and determinant, and there are only two (kinds of) nonisomorphic Witt rings for finite fields of characteristic $\neq 2$. These are $\mathbb{Z}/4\mathbb{Z}$ for fields with the number of elements congruent to 3 (mod 4), and the group ring $\mathbb{Z}/2\mathbb{Z}[\dot{F}/\dot{F}^2]$ for the remaining finite fields of odd characteristic.

Two completions $F_{\mathfrak{p}_1}$ and $F_{\mathfrak{p}_2}$ of a global field F at primes $\mathfrak{p}_1$, $\mathfrak{p}_2$ are Witt equivalent if and only if $|\dot{F}_{\mathfrak{p}_1}/\dot{F}^2_{\mathfrak{p}_1}| = |\dot{F}_{\mathfrak{p}_2}/\dot{F}^2_{\mathfrak{p}_2}|$ and $s(F_{\mathfrak{p}_1}) = s(F_{\mathfrak{p}_2})$ (see [143, Chapter VI]). Here again quadratic forms are classified by the same set of invariants (dimension, determinant, and the Hasse invariant), but there are infinitely many distinct Witt rings for this class of fields (see [143, ChapterVI]).

6.3. *Fields with finite square class number*

From Harrison's criterion it follows that Witt equivalence preserves the order of the square class group of the field, and so classification of fields up to Witt equivalence reduces to classification of fields with a given cardinality of the group of square classes. So it is natural to begin with fields having only finitely many square classes. We have already indicated in Example 6.1 that all fields with one square class are equivalent, and all fields with two square classes split into exactly three classes with respect to Witt equivalence represented by real closed fields and finite fields. These are the only simplest results in a series of classification theorems for fields with a finite number of square classes.

Write $q = q(F) = |\dot{F}/\dot{F}^2|$ for the number of square classes of a field F, and let $w(q)$ be the number of classes of Witt equivalent fields (of characteristic $\neq 2$) with q square classes. Then, as said earlier, $w(1) = 1$ and $w(2) = 3$. The exact values of $w(q)$ are known only for $q \leq 32$. They are shown in the following table:

q	1	2	4	8	16	32
$w(q)$	1	3	6	17	51	155

Witt groups of nonreal fields with $q \leq 8$ were classified by Cordes [38]. For $q = 8$, he found 10 distinct possible Witt groups and for 7 of them he found fields having the prescribed Witt groups. The three remaining fields were constructed by Kula [133]. For formally real fields with $q = 8$, the classification was carried out in [224] and [139]. Seven possible Witt groups were exhibited and for each of them a field with that Witt group was constructed. These results classify Witt *groups*, but the method of classification is based on determining value sets of binary forms and, by Harrison's criterion, these determine the Witt ring structure. However, Cordes [38, Example 7.2] gave an example of two fields K and F with isomorphic Witt groups and nonisomorphic Witt rings. In his example, $q(K) = 8$ and $q(F) = 16$. In fact, there exists a simpler example of this type with $q(K) = 4$ and $q(F) = 8$ (see the cases (3.3) and (4.1) of Theorem 3.2 in [224]).

The fields with $q = 16$ were classified in 1979 in the dissertation of Szczepanik [221] (published in 1985). An interesting fact revealed by that classification is that, for fields with $q = 16$, there are 51 distinct Witt rings, but there are only 45 distinct

Witt groups. Hence, there exist fields having the same number of square classes with isomorphic Witt groups and nonisomorphic Witt rings. These observations show that the Witt ring, and not the Witt group, is the object reflecting properly the behavior of quadratic forms over fields. Hence, one can maintain that generalizations of Witt groups, which do not have natural ring structure, cannot be viewed as part of quadratic form theory.

The classification of Witt rings of fields with $q = 32$ is due to Carson and Marshall [25].

6.4. *Witt equivalence of global fields*

In this section, we consider *global fields* of characteristic not 2. These are either number fields (finite extensions of the rationals) or function fields (finite extensions of rational function fields with a finite field of constants of characteristic not 2). The following property of global fields turns out to be important for investigating Witt equivalence of global fields.

A *Hilbert-symbol equivalence* (HSE, for short) between two global fields K and L is a pair of maps

$$t : \dot{K}/\dot{K}^2 \to \dot{L}/\dot{L}^2, \quad T : \Omega_K \to \Omega_L,$$

where t is an isomorphism of the square class groups and T is a bijective map between the sets of all primes of K and L, preserving quadratic Hilbert symbols[2] in the sense that

$$(a, b)_{\mathfrak{p}} = (ta, tb)_{T\mathfrak{p}} \quad \forall\, a, b \in \dot{K}/\dot{K}^2, \quad \forall\, \mathfrak{p} \in \Omega_K.$$

Using standard results such as the Hasse principle for quadratic forms over global fields and the Harrison criterion for Witt equivalence of fields, one shows that HSE implies Witt equivalence of global fields. Thus, HSE is a set of local conditions for Witt equivalence of global fields. Actually, the two equivalences coincide, but the proof of the converse is not so straightforward (see [189, 226, 229] for three different proofs).

The main result on Witt equivalence of global fields has the form of a local–global principle.

THEOREM 6.5 ([189]). *Two global fields are Witt equivalent if and only if their primes can be paired so that corresponding completions are Witt equivalent.*

Since Witt equivalence of the completions is well understood (see Example 6.1), the local–global principle is an efficient tool for checking Witt equivalence of global fields. The local behavior implies that global *function fields* split into two Witt equivalence classes (depending on the level) and these are disjoint from Witt equivalence classes of number fields. Two other simple consequences of the

[2] For a definition of the quadratic Hilbert symbol see [143, p. 159].

local–global principle, not obvious otherwise, are the following. First, the rational field $\mathbb{Q}$ forms a singleton class, being not Witt equivalent to any other number field, and second, the Witt equivalence of number fields preserves the field degrees over $\mathbb{Q}$.

More generally, as a consequence of the local–global principle, the following is the complete set of invariants of Witt equivalence for a number field F:

$$\big(n, r, s, g\,;\quad (n_i, s_i),\quad i = 1, \dots, g\big),$$

where n is the field degree over $\mathbb{Q}$, r is the number of infinite real primes, s is the level of the field (assuming the value zero when F is formally real), g is the number of dyadic primes, and n_i and s_i are local dyadic degrees and the corresponding local dyadic levels, respectively (see [24]).

Since $n = n_1 + \cdots + n_g$, and with fixed values of local degrees the other invariants can assume only finitely many values, it follows that given the degree n, there are only finitely many Witt equivalent classes of number fields of degree n. The number $w(n)$ of Witt equivalence classes of number fields of degree n is known for $n \le 11$. For example, $w(2) = 7$, $w(3) = 8$, $w(4) = 29$, and $w(9) = 365$ (see [227]).

Remark 6.6. The Witt ring of a global field is a much more sensitive measure than the additive Witt group. One can prove that two formally real algebraic number fields have isomorphic Witt groups if and only if they have the same number of orderings. Two nonreal algebraic number fields have the same Witt group if and only if they have the same level. For quadratic fields, this gives only four Witt groups (while there are seven distinct Witt rings), and for cubic fields, we get only two Witt groups (and eight Witt rings). Another striking difference is that number fields with isomorphic Witt groups can have distinct degrees over $\mathbb{Q}$.

As a result of good understanding of the Witt ring of global field, Czogała and Sładek [55] gave a presentation for the Witt ring $W(F)$ of a global field of characteristic not 2 in terms of square classes, Hilbert symbols and the reduced Witt ring. This had been done earlier for $\mathbb{Q}$ by DeMeyer and Harrison [58, p. 9].

Hilbert-symbol equivalences come in two types: tame and wild. The equivalence (t, T) is said to be *tame* at the finite prime $\mathfrak{p}$ if

$$\operatorname{ord}_{\mathfrak{p}}(a) \equiv \operatorname{ord}_{T\mathfrak{p}}(ta) \pmod 2,$$

for all square classes $a \in \dot{K}/\dot{K}^2$; otherwise (t, T) is *wild* at $\mathfrak{p}$. The equivalence (t, T) is said to be tame if it is tame at every finite prime of K. A main interest in tame equivalence comes from the fact that it induces the following commutative diagram with exact rows

$$\begin{array}{ccccccccc}
0 & \to & W(\mathcal{O}_K) & \to & W(K) & \xrightarrow{\partial_K} & \coprod_{\mathfrak{p}} W(\mathcal{O}_K/\mathfrak{p}) & \to & C(K)/C(K)^2 \to 1 \\
 & & \downarrow & & \downarrow \varphi & & \downarrow \overline{\varphi} & & \downarrow \\
0 & \to & W(\mathcal{O}_L) & \to & W(L) & \xrightarrow{\partial_L} & \coprod_{\mathfrak{p}} W(\mathcal{O}_L/T\mathfrak{p}) & \to & C(L)/C(L)^2 \to 1
\end{array}$$

Here, $\mathcal{O}_K$ is the ring of integers in K, $W(\mathcal{O}_K)$ is the Witt ring of $\mathcal{O}_K$ (defined in Section 9 below), and $C(K)$ is the ideal class group of K. In this diagram, the first three vertical arrows are ring isomorphisms and the fourth is a group isomorphism. It follows that tame equivalence preserves the integral Witt rings and also preserves the 2-ranks of ideal class groups. Tame equivalence was introduced in [189] and studied in [35, 47, 48, 130, 230]. In [231], tameness is studied in the context of semigroups with divisor theory.

Hilbert-symbol equivalence has been generalized to the context of higher degree Hilbert symbols. For this, see the papers [49, 51–53, 218].

Another generalization of HSE was studied by Koprowski in [129]. In algebraic function fields of one variable over a fixed real closed field k, Hilbert symbols are replaced with quaternion algebras. It is proved that there are only two Witt rings for this class of fields so that all formally real function fields over k are Witt equivalent. It is established that two formally real function fields are tamely equivalent if and only if the associated algebraic curves have the same number of semialgebraically connected components. The genus of a function field turns out not to be an invariant of tame equivalence. And finally, there is one special case: for $\mathbb{R}$, the field of real numbers, the rational function field $k(X)$ is tamely equivalent to $\mathbb{R}(X)$ if and only if k is isomorphic to $\mathbb{R}$.

In 1985, Baeza and Moresi showed that any two global fields of characteristic 2 are Witt equivalent (see [12]). It is not difficult to see that a global field of characteristic 2 is never Witt equivalent to a global field of characteristic different from 2.

7. Function fields

7.1. *Selected applications of function fields techniques*

A nonsingular quadratic form q of dimension at least 2 over a field F, viewed as a polynomial over F, is irreducible over F unless q is a hyperbolic plane. Hence, assuming that $\dim q \geq 2$ and q is not a hyperbolic plane, we can consider the integral domain $F[X]/q(X)$ called the *affine algebra* of the quadric $q = 0$ and its field of fractions $F[q]$ called the *function field* of q.

It is obvious that q becomes isotropic over $F[q]$, but it is a nontrivial question as to what other quadratic forms g over F become isotropic over $F[q]$. If g becomes isotropic over $F[q]$, one writes $g > q$, and if g becomes hyperbolic over $F[q]$, one writes $g \gg q$. The starting point of the theory is the following result.

THEOREM 7.1. *If $g \gg q$ and $1 \in D_F(q)$, then $q(X) \cdot g \cong g$ over the rational function field $F(X)$. Moreover, if g is anisotropic, then* $\dim g \geq \dim q$.

From this result, the Arason–Pfister Hauptsatz follows without much difficulty by an inductive argument (see [143, p. 354]). Since an isotropic Pfister form q is hyperbolic, it becomes hyperbolic over $F[q]$. Conversely, from Theorem 7.1 and

the multiplicative properties of Pfister forms it follows that, if $q \gg q$, then q is either a scalar multiple of a Pfister form or is itself a hyperbolic form. According to Lam [143, p. 348], the following three results are amongst the strongest and most interesting results obtained on function fields in the decade.

THEOREM 7.2 (Hoffmann [92]). *Let q and g be quadratic forms over F, and let g be anisotropic. If the dimensions of q and g are separated by a power of two, that is* $\dim g \leq 2^n < \dim q$ *for some $n \geq 1$, then g remains anisotropic over $F[q]$.*

The first Witt index $i_1(q)$ of an anisotropic quadratic form q over F is by definition the (usual) Witt index of the form q over its function field $F[q]$. Clearly, $i_1(q) \geq 1$, and the precise determination of all possible values of the first Witt index is given in the following theorem.

THEOREM 7.3 (Karpenko [110]). *Let q be an anisotropic quadratic form, and let*

$$\dim q - 1 = 2^{n_1} + 2^{n_2} + \cdots + 2^{n_r}$$

be the dyadic expansion of $\dim q - 1$, $n_1 < n_2 < \cdots < n_r$. *Then, for the first Witt index $i_1(q)$, we necessarily have*

$$i_1(q) - 1 = 2^{n_1} + 2^{n_2} + \cdots + 2^{n_s}$$

for some nonnegative $s < r$.

For an anisotropic form q the integer $\dim_{es} q = \dim q - i_1(q) + 1$ is said to be the *essential dimension* of q.

THEOREM 7.4 (Karpenko and Merkurjev [113]). *For anisotropic forms g and q over F, if $g > q$, then* $\dim_{\mathrm{es}} g \geq \dim_{\mathrm{es}} q$ *and equality holds iff $q > g$.*

As an application of their result, Karpenko and Merkurjev proved the following extension of the Arason–Pfister Hauptsatz.

THEOREM 7.5. *For a positive-dimensional anisotropic quadratic form q over a field F, if $\langle q \rangle \in I^n(F)$ and* $\dim q > 2^n$, *then* $\dim q \geq 2^n + 2^{n-1}$.

While Karpenko and Merkurjev's arguments work in any characteristic $\neq 2$, Theorem 7.5 was first proved for fields F of characteristic 0 by Vishik [234].

There is an even more penetrating result solving completely the problem of determining dimensions of anisotropic forms q in $I^n(F)$.

THEOREM 7.6 (Vishik gap theorem). *Any anisotropic form q in $I^n(F)$ has dimension equal to one of the following numbers:*

$$2^{n+1} - 2^{r+1}, \quad 0 \leq r \leq n \quad \text{or} \quad 2^{n+1} + 2k, \quad k \geq 0.$$

It is also possible to realize all these dimensions by anisotropic forms in $I^n(F)$ over a suitable field F. For a proof of Vishik's gap theorem see Karpenko [111].

Knebusch [119] introduced *generic splitting towers* of quadratic forms, which is essentially the following construction leading to higher Witt indices $i_j(q)$ by iterating the process of forming the first Witt index $i_1(q)$ (for fields of arbitrary characteristic see [124]). So, we put $F_0 = F, q_0 = q$, and let $F_1 = F[q]$ be the function field of q. We have defined $i_1(q)$ to be the Witt index of q over F_1, and if, for $j > 0$, q_{j-1} is the anisotropic part of q over F_{j-1}, we set $i_j(q)$ to be the Witt index of q_{j-1} over $F_j = F_{j-1}[q_{j-1}]$ if $\dim q_{j-1} \geq 2$. The smallest h such that $\dim q_h \leq 1$ is called the height of q and $i_1(q), \ldots, i_h(q)$ is the sequence of higher Witt indices of q. The question of possible values for these higher Witt indices remains open. In [112], Karpenko found relations among the 2-adic orders $v_2(i_j(q))$ of the higher Witt indices. He proved that

$$\max_{m<j\leq h} v_2(i_j(q)) \geq v_2(i_m(q)) \geq \min_{m<j\leq h} v_2(i_j(q)) - 1,$$

where the first inequality requires that $i_m(q) + 2\sum_{j=m+1}^{h} i_j(q)$ be not a power of two. See also [98] for the lower bound of the heights of all forms of a given dimension.
As a final example, we discuss characterization of $I^n(F)$ via Knebusch's degree. For an even-dimensional form q of height h, the form q_{h-1} above has itself height 1 and so it is similar to a Pfister form by a theorem of Wadsworth [237]. If q_{h-1} is similar to an n-fold Pfister form, $n = \deg q$ is said to be the Knebusch's *degree* of the form q. Observe that, when q is an n-fold Pfister form, then its height is 0 and hence its degree equals n. Now, following Knebusch [119], we set

$$J_n(F) = \{\langle q\rangle \in I(F) : \deg q \geq n\}.$$

Knebusch [119, Theorem 6.4] proved that $J_n(F)$ is an ideal of the Witt ring $W(F)$. A first consequence of this is that

$$I^n(F) \subseteq J_n(F).$$

To see this, observe that for an anisotropic form q we have $\dim q \geq 2^n$, where $n = \deg q$. Since all n-fold Pfister forms have degree n, they certainly belong to $J_n(F)$, and so, since $J_n(F)$ is an ideal of the Witt ring $W(F)$, we have $I^n(F) \subseteq J_n(F)$.

A second consequence is the Arason–Pfister Hauptsatz. For, if $I^n(F) \subseteq J_n(F)$, then each anisotropic form in $I^n(F)$ has degree at least n, and hence the dimension at least 2^n, which is the Arason–Pfister Hauptsatz.

A natural and important question to ask was whether $I^n(F) = J_n(F)$ holds for all $n \geq 1$. This outstanding problem was solved in the affirmative in Orlov–Vishik–Voevodsky's paper [182, Theorem 4.3].

There are several surveys and texts on quadratic forms emphasizing function fields and their applications in quadratic form theory. First, Lam [143] gives an introduction to the subject and an overview of new results in the area. The Lens Conference volume [100] includes an introduction to motives of quadrics by Vishik, with applications to the splitting patterns of quadratic forms; papers by Izhboldin and Karpenko on Chow

groups of quadrics with application to the construction of fields with u-invariant 9; and a paper by Kahn presenting a general framework for the computation of the unramified cohomology groups of quadrics and other varieties. A forthcoming book [65] by Elman, Karpenko and Merkurjev is a research monograph written from the viewpoint of algebraic geometry and presents quadratic form theory in full generality. Part I of the book was available before publication in preprint form. The publisher's announcement includes the following information. "Part I includes the classical algebraic theory of quadratic and bilinear forms and answers many questions that have been raised in the early stages of the development of the theory. Assuming only a basic course in algebraic geometry, Part II presents the necessary additional topics from algebraic geometry including the theory of Chow groups, Chow motives, and Steenrod operations. These topics are used in Part III to develop a modern geometric theory of quadratic forms."

7.2. *u-invariant*

For nonreal field F, Kaplansky defined $u(F)$ to be the maximal dimension of *universal* quadratic form (which is the same as the maximal dimension of an anisotropic quadratic form) over F. Here, universal means representing all nonzero field elements. Hence, $u(F) = 1$ iff F is quadratically closed, $u(F) = 2$ for a finite field F, $u(F) = 4$ for any local field F, and hence (by Hasse Principle) $u(F) = 4$ for any nonreal global field F. If F is an algebraic function field of transcendence degree n over the complex field $\mathbb{C}$, then $u(F) \leq 2^n$ and $u(F) = 2^n$ when F is finitely generated over $\mathbb{C}$. This is a special case of a more general result proved in 1936 by Tsen [233]. However, using the theory of Pfister forms one easily proves that always $u(F) \neq 3, 5, 7$.

In 1953, Kaplansky [108] conjectured that $u(F)$ is always a power of two (or ∞). He observed that, for a field F complete with respect to a discrete rank one valuation with residue class field K of characteristic different from two, $u(F) = 2u(K)$. This produces examples of fields with u-invariant equal to any given power of 2 (for instance, the iterated power series field $\mathbb{C}((t_1)) \cdots ((t_n))$ has $u = 2^n$). See also the Elman and Lam papers [66, 67].

Kaplansky's conjecture was proved for various classes of fields, but in 1988, much to anybody's surprise, the first examples of nonreal fields with u-invariant equal to 6 were constructed by Merkurjev [169]. Actually, any field K is contained in a field F with $u(F) = 6$. Merkurjev's construction uses function field techniques for six-dimensional quadratic forms of determinant -1. These arise naturally as "Albert forms" of biquaternion algebras. If $B = \left(\frac{a,b}{F}\right)$ and $C = \left(\frac{c,d}{F}\right)$ are quaternion algebras over F, then the Albert form of their tensor product $A = B \otimes_F C$ is the form $q_A = (-a, -b, ab, c, d, -cd)$. The crucial property of q_A is that, A is a division algebra iff q_A is anisotropic over F (see [127, Chapter IV, Section 16] for a comprehensive discussion of Albert forms). The main ingredient of Merkurjev's construction is the following theorem.

THEOREM 7.7 (Merkurjev). *If the Albert form q_A is anisotropic over F, then for any nonsingular form g over F of dimension ≥ 7, the form q_A remains anisotropic over the function field $F[g]$.*

For a comprehensive survey and details of proofs, see Lam [144] and also [143, pp. 484–495].

Further, in 1991, Merkurjev [170] proved that for any even number $2n$ there exists a field (of cohomological dimension two) with u-invariant $2n$. This construction led to the study of index-reduction formulas for division algebras and algebraic groups. In case of fields of characteristic 2, fields with any even u-invariant were constructed by Mammone, Tignol, and Wadsworth [160] using the ideas of Merkurjev's construction of fields with $u = 6$.

The next challenge was to decide whether there exist fields with odd u-invariant. Izhboldin [99] constructed the first examples of fields with $u = 9$. It is now known that $u(F)$ can assume any value of the form $2^r + 1$ for $r \geq 3$. This was proved by Vishik [235] using a new method which also allows constructing fields with even u-invariant without using the Merkurjev index reduction formula.

A natural problem in the study of the u-invariant is to find estimates for u under finite field extensions. The best general result in this direction is the following theorem.

THEOREM 7.8 (Leep [145]). *For a nonreal field F and any field extension K/F of degree $n = [K : F] < \infty$, we have $u(K) \leq \frac{1}{2}(n + 1)u(F)$.*

The proof depends on a generalization of the u-invariant to the context of systems of quadratic forms. See [147, 177].

The u-invariant considered so far is an invariant of nonreal fields. Already Kaplansky in 1953 stated that "the formally real case probably has a parallel theory, which may be worth separate study." For a formally real field F, there are at least three distinct generalizations of the u-invariant. We shall confine attention to the one introduced by Elman and Lam in [66]. The u-invariant of F is the supremum of the dimensions of anisotropic forms over F having signature zero with respect to every ordering of F. By the Pfister local–global principle (Theorem 3.3), $u(F)$ is the supremum of dimensions of anisotropic torsion forms over F. Torsion forms are necessarily even-dimensional, so $u(F)$, if finite, is an even integer. For power series fields, the equality $u(F((t))) = 2u(F)$ still holds; hence, there are formally real fields with the u-invariant 2^n for any integer n. But also here Kaplansky's conjecture fails to hold. Modifying Merkurjev's construction Lam [144, Theorem 5.2] shows that there exist formally real fields with u-invariant 6. Then, in a (handwritten) manuscript, "Some consequences of Merkurjev's work on function fields" (1989), Lam adopted the arguments of the preprint version of Merkurjev's paper [170] to the formally real field context and proved that there exist formally real fields with $u = 2n$, any given even integer. Similar results were published a little later by Hornix [96].

A survey of results on isotropy questions over function fields of quadrics is available (see Hoffmann [94]). It presents a uniform approach to constructing fields with prescribed invariants (level, Pythagoras number, u-invariant, including $u = 9$).

8. Sums of squares

8.1. *Level*

Recall that for a ring A, the level $s(A)$ is the smallest natural number n such that $-1 \in A$ is a sum of n squares in A or ∞ if -1 is not a sum of squares in A. We have mentioned already in Section 2 that the level of a nonformally real field is always a power of two. This was proved by Pfister as a simple consequence of the fact that for a Pfister form q, the value set $D_F(q)$ is a group under multiplication. In particular, nonzero sums of 2^n squares of elements of F form a group under multiplication. Now, assume that F is a nonreal field with $s = s(F)$ and let $2^n \leq s < 2^{n+1}$. Then $-1 = A + B$, where A is a sum of 2^n and B a sum of less than 2^n squares in F. By the group property, it follows that $-1 = (1 + B)/A$ is a sum of 2^n squares in F, and so $s = 2^n$. This proves the first part of the following theorem.

THEOREM 8.1 (Pfister level theorem [190]). *For any nonreal field F, its level $s(F)$ is a power of* 2. *For any $s = 2^n$, there exists a nonreal field F with $s(F) = s$.*

The function field $\mathbb{R}[q]$ of the quadratic form $q = X_1^2 + \cdots + X_m^2$ over $\mathbb{R}$, where $2^n \leq m - 1 < 2^{n+1}$, can be shown to have level 2^n.

Pfister's level theorem, together with the calculations of the u-invariant and Pythagoras number prior to Merkurjev's breakthrough in 1989, gave the false impression that all sensible quadratic form invariants of fields should be powers of two.[3] In particular, it would be natural to expect that the level of integral domains also assumes only some restricted values. However, there are no restrictions at all, as follows from the following elegant theorem.

THEOREM 8.2 (Dai–Lam–Peng [56]). $s(\mathbb{R}[X_1, \ldots, X_n]/(1 + X_1^2 + \cdots + X_n^2)) = n$.

The method of proof is somewhat astonishing: it is based on the topological Borsuk–Ulam antipodal point theorem. Dai and Lam [57] studied in depth the interrelations between levels in algebra and topology. In topology, the level $s(X)$ of a topological space X endowed with a fix-point-free involution i is the smallest number n with the property that there exists an equivariant map from (X, i) to the sphere $(S^{n-1}, -)$, where $-$ denotes the antipodal involution on the sphere (or $s(X) = \infty$ if such a map does not exist). Then, for instance, the Borsuk–Ulam theorem states that $s(S^{n-1}, -) = n$. Dai and Lam prove a general "Level Theorem" associating with each space (X, i) a commutative $\mathbb{R}-$algebra A_X such that the levels coincide: $s(X) = s(A_X)$. Theorem 8.2 is then obtained by taking $X = S^{n-1}$. For proofs, see also [143, pp. 505–514] and [198, pp. 46–50].

[3] At an Oberwolfach conference on quadratic forms sitting at breakfast, I mentioned in the presence of J.-P. Serre that it is a bit strange that the organizers invited only 29 persons for the conference since in quadratic form theory all good invariants are powers of two. Serre replied: 16 would be enough.

Lewis [150] constructed examples of quaternion division algebras whose levels take any prescribed value 2^k or $2^k + 1$. Whether the level of a division quaternion algebra can take other values was an open question. Recently, Hoffmann [95] proved that there are infinitely many values not of the form 2^k or $2^k + 1$ that occur as levels of quaternion algebras. This uses function field techniques and a result of Karpenko and Merkurjev [113].

For a survey on levels of fields and quaternion algebras, see Lewis [156] and the list of references there. For the relation between the commutativity of a division algebra and representability of 0 as a sum of squares, see [148].

8.2. *Composition*

A composition formula of size $[r, s, n]$ is a sum of squares formula of the type

$$(x_1^2 + x_2^2 + \cdots + x_r^2) \cdot (y_1^2 + y_2^2 + \cdots + y_s^2) = z_1^2 + z_2^2 + \cdots + z_n^2,$$

where $X = (x_1, x_2, \ldots, x_r)$ and $Y = (y_1, y_2, \ldots, y_s)$ are systems of indeterminates and each $z_k = z_k(X, Y)$ is a bilinear form in X and Y. A triple $[r, s, n]$ is *admissible* over a field F if a composition formula of size $[r, s, n]$ exists with the bilinear forms z_k in X and Y having coefficients in F.

Formulas of size $[n, n, n]$ are known to exist for $n = 1, 2, 4, 8$, and they are related to multiplication in the field of complex numbers, and in quaternion and octonion algebras. Hurwitz proved in 1898 that $[n, n, n]$ formulas exist over $\mathbb{C}$ only for $n = 1, 2, 4, 8$. Actually, Hurwitz's arguments work over any field of characteristic $\neq 2$.

Another classical result is the Hurwitz–Radon theorem for the size $[r, n, n]$. For this to be admissible over $\mathbb{R}$ or $\mathbb{C}$, it is necessary and sufficient that $r \leq \rho(n)$, where $\rho(n)$ is the "Hurwitz–Radon function" defined as follows: if $n = 2^{4a+b} n_0$, where n_0 is odd and $0 \leq b \leq 3$, then $\rho(n) = 8a + 2^b$. Again, Hurwitz's ideas can be generalized to any field of characteristic $\neq 2$.

In 1940, using topological methods, Hopf proved the following result.

THEOREM 8.3. *If $[r, s, n]$ is admissible over $\mathbb{R}$, then the binomial coefficient $\binom{n}{k}$ is even whenever $n - s < k < r$.*

The parity condition on binomial coefficients in the above theorem is said to be the hopf condition. It has been known that for any formally real field F, and even for all fields of characteristic 0, if $[r, s, n]$ is admissible over F, then the Hopf condition holds. However, generalizing this to fields of prime characteristic was an unsolved problem. As it turned out recently, extending Hopf's theorem to fields of prime characteristic required the use of motivic cohomology. Dugger and Isaksen [64] proved in 2007 that the Hopf conditions are necessary for admissibility of $[r, s, n]$ over an arbitrary field F of characteristic $\neq 2$.

A comprehensive presentation of all aspects of the composition of quadratic forms is contained in the book Shapiro [209]. See also [127, Chapter VIII].

8.3. *Pythagoras number*

The Pythagoras number of a ring R, denoted $P(R)$, is the smallest positive integer n such that every sum of squares in R can be written as a sum of $\leq n$ squares. The Pythagoras number of a nonreal field F with level s assumes one of the following two values: s or $s+1$. Indeed, any element of the field can be written as a difference of two squares $x^2 + (-1)y^2$ and -1 is a sum of s squares. Moreover, using standard arguments, one can show that for any $s = 2^n$ there are fields of level s and Pythagoras number s, $s+1$, respectively (see [143, p. 396]). The formally real fields have a much more difficult and interesting theory of Pythagoras numbers. Prestel [201] showed that for any number m, there exist uniquely ordered formally real subfields F and K of $\mathbb{R}$ with $P(F) = 2^m$ and $P(K) = 2^m + 1$. For a long time, it remained an open problem whether the Pythagoras number of a formally real field, if finite, must be of the form 2^m or $2^m + 1$. A complete solution to this problem came as a great surprise not so because of the author but rather because of the answer.

THEOREM 8.4 (Hoffmann [93]). *For any $n \geq 1$, there exists a formally real field F with $P(F) = n$.*

In fact, a more precise result was obtained, namely a field exists which is uniquely ordered, has $P(F) = n$ and has any prescribed *even* value of the $u-$invariant. The method of proof relies on function field technique used by Merkurjev to prove the existence of fields with $u = 6$. Lam [143, pp. 495–499] explains the approach in the example of construction of fields with Pythagoras numbers 6 and 7.

Artin [8] proved in 1927 that every positive semidefinite rational function in $K = \mathbb{R}(X_1, \ldots, X_n)$ is a sum of squares of rational functions. His approach does not say, however, whether the Pythagoras number $P(K)$ is finite. This was clarified 40 years later by Pfister.

THEOREM 8.5 (Pfister [192]). $P(\mathbb{R}(X_1, \ldots, X_n)) \leq 2^n$.

The three main ingredients of Pfister's proof are the Artin result, the theorem of Tsen, and Pfister's theory of multiplicative (Pfister) forms. While this is an important quantitative result, yet all is not well. We know the *exact* value of $P(\mathbb{R}(X_1, \ldots, X_n))$ only for $n = 1$ and $n = 2$. Here, $P(\mathbb{R}(X_1)) = 2$ is elementary algebra, and $P(\mathbb{R}(X_1, X_2)) = 4$ is a very difficult result proved by Cassels, Ellison and Pfister [27] in 1971 by using the theory of elliptic curves over $\mathbb{R}(X)$. On the other hand, for the polynomial rings over $\mathbb{R}$ and $\mathbb{Z}$, the situation is clarified by the following theorem.

THEOREM 8.6 (Choi, Dai, Lam, and Reznick [28]). $P(\mathbb{R}[X_1, \ldots, X_n]) = \infty$ *for* $n \geq 2$ *and* $P(\mathbb{Z}[X_1, \ldots, X_n]) = \infty$ *for* $n \geq 1$.

Another natural question is to find the Pythagoras numbers $P(\mathbb{Q}(X_1, \ldots, X_n))$ and $P(\mathbb{Q}[X_1, \ldots, X_n])$. Landau proved in 1906 that $P(\mathbb{Q}[X_1]) \leq 8$ and Pourchet [200] in 1971 established $P(\mathbb{Q}[X_1]) = 5$. According to a result of Cassels [26], for an

arbitrary field F, one has $P(F[X_1]) = P(F(X_1))$. Hence, $P(\mathbb{Q}(X_1)) = 5$. For $n = 2$ and $n = 3$, it was proved by Colliot-Thélène and Jannsen [32] that

$$P(\mathbb{Q}(X_1, X_2)) \leq 8 \quad \text{and} \quad P(\mathbb{Q}(X_1, X_2, X_3)) \leq 16.$$

A conditional result in [32], modulo Milnor's conjecture and a conjecture of Kato, asserts that $P(\mathbb{Q}(X_1, \ldots, X_n)) \leq 2^{n+1}$ for $n \geq 2$. For further details, see [199] and [77].

9. Witt rings of rings

9.1. *Witt rings of commutative rings*

Knebusch [115] generalized in 1970 the construction of the Witt ring $W(R)$ of a field R to cover the case when R is a commutative ring. Instead of nonsingular symmetric bilinear forms on finite dimensional vector spaces, one considers finitely generated projective modules over R equipped with nonsingular symmetric bilinear forms, called *inner product spaces*. For two inner product spaces (S, α) and (T, β) their *orthogonal direct sum*, written $(S, \alpha) \perp (T, \beta)$ or $S \perp T$, is defined to be the direct sum of the modules S and T with bilinear form γ, defined by the equation

$$\gamma(s_1 \oplus t_1, s_2 \oplus t_2) = \alpha(s_1, s_2) + \beta(t_1, t_2).$$

For inner product spaces S and T, $S \perp T$ is also an inner product space. For an inner product space (S, α) and for a submodule N of S, consider the *orthogonal complement* $N^{\perp}$ of N in S,

$$N^{\perp} := \{s \in S : \alpha(s, N) = 0\}.$$

The inner product space (S, α) is said to be *metabolic* if there exist submodules M and N of S such that,

$$S = M \oplus N \quad \text{and} \quad N = N^{\perp}.$$

Two inner product spaces M and N are said to be *similar*, or *Witt equivalent*, written as $M \sim N$, if there exist metabolic spaces S and T so that $M \perp S$ is isometric to $N \perp T$. Similarity is an equivalence relation and we write $W(R)$ for the collection of all similarity classes (Witt classes) of inner product spaces over the commutative ring R. The class containing the space (M, β) is written as $\langle M, \beta \rangle$ or simply $\langle M \rangle$. We define two operations in $W(R)$, the addition and the multiplication of classes by setting

$$\langle M \rangle + \langle N \rangle = \langle M \perp N \rangle \quad \text{and} \quad \langle M \rangle \cdot \langle N \rangle = \langle M \otimes N \rangle.$$

These operations do not depend on the choice of representatives of the similarity classes, and the sum and the product of similarity classes of two inner product spaces are the similarity classes of inner product spaces. Thus, we have well-defined addition and multiplication operations in $W(R)$. The collection $W(R)$ of all Witt classes of inner

product spaces over R forms a commutative ring, using the orthogonal sum as addition operation and the tensor product as multiplication operation.

When R is a field, this construction is identical with the original construction of Witt presented in Section 3. Details of the construction and further material may be found in [10, 163, 174, 204].

Any ring homomorphism $R \to R'$ induces a homomorphism $W(R) \to W(R')$ of Witt rings. For an integral domain R and its field of fractions F, consider the ring homomorphism

$$\varphi : W(R) \to W(F)$$

induced by the inclusion $R \hookrightarrow F$. If R is a Dedekind domain it is known that $\ker \varphi$ is zero. This was first proved by M. Knebusch ([115, Satz 11.1.1]).

Also in some other cases, this homomorphism is injective. The most general situation known where this happens is the case when R is a regular domain of dimension ≤ 3 (theorem of Pardon (1984) and Ojanguren (1982), see [126, p. 472]). However, φ is no longer injective when the dimension of R is four. A counterexample can be found in [128].

A more general case was considered by Craven, Rosenberg, and Ware [41]. They proved that when R is a regular Noetherian domain of an arbitrary Krull dimension, then $\ker \varphi$ is a nilideal. Here, the regularity condition is not necessary since, as shown in [30], the kernel of φ is a nilideal also for nonmaximal orders in global fields, and these are not regular rings.

On the other hand, for each order $\mathbb{Z}[fi]$, $f > 1$, of the Gaussian field $\mathbb{Q}(i)$, the natural ring homomorphism $W(\mathbb{Z}[fi]) \to W(\mathbb{Q}(i))$ is not injective, and there are other results of that type [31]. They confirm in part the conjecture that for an algebraic number field F and its order $\mathcal{O}$ the natural ring homomorphism $W(\mathcal{O}) \to W(F)$ is injective if and only if $\mathcal{O}$ is the maximal order of F.

All the results on nilpotency of the kernel of natural homomorphisms of Witt rings depend on the following local–global principle proved by Dress [62]. For a ring R, and a prime ideal $\mathfrak{p}$ of R, we write $R_{\mathfrak{p}}$ for the localization of R at $\mathfrak{p}$.

For a commutative ring R and any inner product R-space E, the class $\langle E\rangle \in W(R)$ is nilpotent if and only if $\langle E_{R_{\mathfrak{p}}}\rangle \in W(R_{\mathfrak{p}})$ is nilpotent for all maximal ideals $\mathfrak{p}$ in R.

From this result, facts about prime ideal structure in Witt rings over fields can be extended to commutative rings (see, [163]).

Knebusch, Rosenberg, and Ware in [122, 123] generalized most of the results in Section 3 to the case of Witt rings of semilocal rings (for this, see also [116]), and of Dedekind rings. Actually, they consider semilocal and Dedekind rings with involutions and work with more general Witt rings described in [121]. Their work culminated in two papers by Knebusch [117, 118], where finally the notion of a real closure of a commutative ring was discovered and a ring-theoretical counterpart of the Artin–Schreier theory of formally real fields with its relation to Witt rings of fields was found. For a self-contained treatment of quadratic and bilinear forms over semilocal rings and of Witt rings of semilocal rings, see Baeza [10] and Marshall [163].

9.2. *Witt rings of Dedekind domains*

Let F be the field of fractions of a Dedekind domain $\mathcal{O}$. As mentioned in the preceding section, Knebusch [115] proved that the natural ring homomorphism

$$W(\mathcal{O}) \xrightarrow{i} W(F)$$

is injective. Milnor extended the result in the following way. With $\mathfrak{p}$ running over the set of all nonzero prime ideals of $\mathcal{O}$, we have the following exact sequence for the Witt groups $W(\mathcal{O})$, $W(F)$:

$$0 \to W(\mathcal{O}) \xrightarrow{i} W(F) \xrightarrow{\partial} \coprod_{\mathfrak{p}} W(\mathcal{O}/\mathfrak{p}), \tag{9.1}$$

where $\partial = \partial_{\mathcal{O}}$ is the *total residue homomorphism*. This is said to be the *Knebusch–Milnor exact sequence*. The background, definitions of the maps involved, and proofs can be found in [174, Ch. IV].

The sequence (9.1) can be extended to the right in some important special cases. First of all, when $\mathcal{O}$ is the ring of integers of a number field F and $C = C(F)$ is the ideal class group, Milnor proved that with a suitable choice of the homomorphism λ the sequence

$$0 \to W(\mathcal{O}) \xrightarrow{i} W(F) \xrightarrow{\partial} \coprod_{\mathfrak{p}} W(\mathcal{O}/\mathfrak{p}) \xrightarrow{\lambda} C/C^2 \to 0 \tag{9.2}$$

is exact. Hence, the cokernel of the total residue homomorphism ∂ is the finite elementary Abelian 2-group C/C^2. For a proof, see [174, pp. 93–94] and [204, p. 227].

When $\mathcal{O} = k[X]$ is the ring of polynomials in one indeterminate over a field k and $F = k(X)$ is the field of rational functions over k, the cokernel is known to be the zero group (Milnor's theorem, see [204, p. 211]). When $\mathcal{O}$ is the ring of polynomial functions on the circle $\mathcal{O} = \mathbb{R}[X, Y]/(X^2 + Y^2 - 1)$, then $\operatorname{coker} \partial = \mathbb{Z}$ is the infinite cyclic group ([174, p. 94]).

Pfister [197] studied the Knebusch–Milnor sequence in the case of the quadratic extension $\mathcal{O} = k[X, \sqrt{aX^2 + b}\,]$ of the polynomial ring $k[X]$ such that for $F = k(X, \sqrt{aX^2 + b}\,)$, the field of fractions, F/k is a function field of genus 0 but is not a rational function field. Pfister [197, Theorem 5] proved that in that case

$$\operatorname{coker} \partial \cong \operatorname{ann}\langle 1, -a, -b, ab\rangle / \operatorname{ann}\langle 1, -a\rangle,$$

where the annihilators are taken in the Witt ring $W(k)$. Using this result, one can construct a Dedekind domain $\mathcal{O}$ with the cokernel of $\partial_{\mathcal{O}}$ containing torsion elements of any given two-power order (see [232] for details).

The description of the cokernels of the total residue homomorphisms $\partial_{\mathcal{O}}$ for all Dedekind rings $\mathcal{O}$ seems to be out of reach. A more realistic open problem is to characterize the class of Abelian groups, which occur as cokernels of total residue homomorphisms $\partial_{\mathcal{O}}$ for all Dedekind rings $\mathcal{O}$.

Very little is known about Witt equivalence for general Dedekind domains. The structure of Witt rings of Hasse domains in global fields is determined in [50], and

also, sufficient conditions for Witt equivalence of Hasse domains are given. In [54], an analog of the Theorem 6.5 was proved for Witt equivalence of semilocal subrings of a Dedekind domain.

9.3. *Witt rings of skew fields and algebras*

For a skew field A, Craven [46] defined the Witt ring $W(A)$ as the abstract Witt ring (see Section 10.1 below)

$$W(A) = \mathbb{Z}[S(A)]/J,$$

where $S(A)$ is the subgroup of the multiplicative group $A^\times$ generated by the set of all squares $A^{\times 2}$. Here, J is the ideal generated by the set

$$\{[1] + [a] - [1 + a] - [a(1 + a)] : a \in S(A), a \neq 0, -1\} \cup \{[1] + [-1]\}.$$

As in the commutative case, for an ordered skew field A there is a bijective correspondence between the set of minimal prime ideals of the ring $W(A)$ and the set $X(A)$ of all orderings of A. The set $X(A)$ can be given the induced Zariski topology from the prime spectrum and this makes $X(A)$ into a Boolean space. Craven [45] proved that every Boolean space can be obtained as the space of orderings of a noncommutative skew field. He also proved that every field F can be embedded in a noncommutative skew field A such that the induced ring homomorphism $W(F) \to W(A)$ is an isomorphism. However in [46], an example is produced of a skew field D, for which, given any commutative subfield $F \hookrightarrow D$, the induced ring homomorphism $W(F) \to W(D)$ has nontrivial kernel.

Another interesting question is whether for a given noncommutative skew field A, there is a field F such that $W(A) \cong W(F)$. Sładek proved that this is so when (a) A is a quaternion algebra over a Pythagorean field or over a local field ([213]) or (b) A is a division algebra over a global field and has odd dimension ([216]).

When the dimension of A is even, $W(A)$ is a homomorphic image of the Witt ring of a field ([217]).

Sładek [215] proved a noncommutative counterpart of Springer's theorem on the structure of Witt groups (and rings) of fields complete with respect to a discrete valuation. As a consequence, he showed that there is a noncommutative field A complete with respect to a discrete valuation and such that the Witt ring $W(A)$ is not isomorphic to the Witt ring of any field ([214]).

For Witt rings of central simple algebras, see Lewis and Tignol [157]. For orderings of higher level and reduced Witt rings of skew fields, see Craven [46].

9.4. *More general Witt rings and groups*

Knebusch defined in [120] the Witt rings of arbitrary schemes. Balmer [13, 14] introduced Witt groups of triangulated categories with duality. These generalize Knebusch's Witt groups of schemes. A discussion of the interrelations amongst various generalizations of the concept of Witt group can be found in Balmer [15].

10. Abstract Witt rings

10.1. *Elementary type conjecture*

We have mentioned in Section 3 that the Witt ring of a field F has a presentation by generators and relations $W(F) = \mathbb{Z}[G]/J$, where $G = \dot{F}/\dot{F}^2$ is the square class group of F and J is an ideal of the group ring $\mathbb{Z}[G]$. It is an important property of the Witt ring that the torsion subgroup of the additive group of the ring is 2-primary. These observations led Knebusch, Rosenberg, and Ware [121] to the following definition.

An *abstract Witt ring* is a ring R with a presentation

$$R = \mathbb{Z}[G]/J,$$

where G is an elementary 2-group and the subgroup of torsion elements in R is 2-primary. It turns out that for such abstract Witt rings one can prove most of the structure results, which Pfister proved for the Witt rings of fields.

Abstract Witt rings form a category where a morphism $h : \mathbb{Z}[G]/J \to \mathbb{Z}[G_1]/J_1$ between two objects is a ring homomorphism sending each coset $g + J$ for $g \in G$ onto a coset $g_1 + J_1$ for some $g_1 \in G_1$. Denote this category as $\mathcal{AWR}$. Clearly, Witt rings of fields are objects of this category. If R is an abstract Witt ring and there exists a field F such that R and $W(F)$ are isomorphic, then we say that the Witt ring R is realized by the field F.

Another approach to abstract Witt rings uses *linked quaternionic mappings* as the basic concept of the theory (see [162, 167]). The categories of abstract Witt rings and linked quaternionic mappings are equivalent.

The first fact showing advantages of the abstract approach to Witt rings is the following observation: *finite products exist in the category* $\mathcal{AWR}$. Indeed, given abstract Witt rings $R_1, \ldots, R_n$, the subring R of the Cartesian product $R_1 \times \cdots \times R_n$ consisting of elements $(f_1, \ldots, f_n)$ satisfying $\dim f_i \equiv \dim f_j \pmod 2$ for all $i, j \leq n$, is the product of $R_1, \ldots, R_n$ in the category $\mathcal{AWR}$ (fiber product over $\mathbb{Z}/2\mathbb{Z}$).

There is a second operation in $\mathcal{APW}$, the group extension of a Witt ring. If R is a Witt ring, then the group ring $R[\Delta]$, where Δ is an elementary 2-group, is also an abstract Witt ring. These two operations are said to be the *elementary operations* on abstract Witt rings. For a given list of *basic* Witt rings, say $R_1, \ldots, R_m$, the elementary operations generate a class of Witt rings constructible from the basic rings by means of the elementary operations. For example, all Witt rings of fields with square class number $q \leq 8$ can be obtained by elementary operations from the following list of basic Witt rings:

$$\mathbb{Z}, \quad \mathbb{Z}/2\mathbb{Z}, \quad \mathbb{Z}/4\mathbb{Z}, \quad \mathbb{L},$$

where $\mathbb{L}$ is the Witt ring of the field $\mathbb{Q}_2$ of 2-adic numbers. The four basic rings cannot be constructed from any simpler Witt rings by elementary operations and they are realized by the fields $\mathbb{R}, \mathbb{C}, \mathbb{F}_3, \mathbb{Q}_2$, respectively. A further analysis of the situation leads to the conclusion that the list of basic Witt rings should include also the Witt rings $\mathbb{L}_n$, $n \geq 1$, of finite extensions of degree n of the field $\mathbb{Q}_2$ of 2-adic numbers.

If $n = 2k$ is even, there are two nonisomorphic Witt rings denoted $\mathbb{L}_{2k,0}$ and $\mathbb{L}_{2k,1}$ corresponding to fields with level $s = 1$ or 2, respectively.

All these Witt rings are indecomposable in terms of elementary operations. The Witt rings obtained from

$$\mathbb{Z}, \quad \mathbb{Z}/2\mathbb{Z}, \quad \mathbb{Z}/4\mathbb{Z}, \quad \mathbb{L}_{2k,0}, \quad \mathbb{L}_{2k,1}, \quad \mathbb{L}_{2k-1} \quad k \geq 2$$

by elementary operations are said to be of *elementary type*. With this list of basic Witt rings, we have the following *elementary type conjecture*:

Each finitely generated abstract Witt ring is of elementary type.

A restricted form of the conjecture states that *each finitely generated abstract Witt ring, which is realized by a field, is of elementary type.* The classification results for fields with respect to Witt equivalence described in Section 6.3 confirm the ETC for Witt rings of fields with square class number $q(F) \leq 32$.

The ETC has also been proved for torsion-free abstract Witt rings. As we know from Theorem 3.4, the Witt ring of a field is torsion-free if and only if the field is Pythagorean. Hence, the Witt ring of any Pythagorean field with finite square class number can be obtained from the Witt ring of the reals $W(\mathbb{R}) = \mathbb{Z}$ by means of a sequence of elementary operations. This was first proved by Bröcker [23] and Craven [43, 44] in the field case (for reduced Witt rings, see Section 10.2 below) and then by Marshall ([161]), and independently in the dissertation of Sładek in 1979, in the general abstract case.

Elementary operations and ETC are particularly (and, perhaps, exclusively) important for Witt rings of fields (abstract Witt rings realizable by fields). A fundamental question arises whether rings obtained by elementary operations from Witt rings realizable by fields are likewise realizable by fields. In case of group extensions, a positive answer follows from results of Springer [220] on the structure of the Witt group of a nondyadic complete discretely valuated field F. Springer's theorem asserts that $W(F) \cong W(\overline{F}) \oplus W(\overline{F})$, where $\overline{F}$ is the residue class field of F (for an analog of Springer's theorem for fields of characteristic 0, carrying a 2-Henselian valuation with residue field of characteristic 2, see Jacob [101, 102]).

Roughly speaking, group extensions of the Witt ring $W(F)$ are realized by (iterated) formal power series fields over F. Realizability of products is more difficult and was proved by Kula using valuation theoretic means.

THEOREM 10.1 (Kula [132, 134]). *The product of Witt rings of fields is the Witt ring of a field.*

This theorem and the validity of ETC would give a complete solution to the classification problem of Witt rings of fields with finite groups of square classes.

The complexity of ETC for the Witt ring of a field F is measured by the square class number $q(F) = |\dot{F}/\dot{F}^2|$. As we have mentioned earlier, for $q \leq 32$, ETC holds. Another measure of complexity of ETC for a field F is the number $Q = Q(F)$ of quaternion algebras over F. The most difficult are the cases where Q is a power of two. The case $Q = 2$ is already nontrivial and was investigated by

Fröhlich, Kaplansky [109], and Cordes [39] (a definitive exposition in [162, pp. 95–99]). Kaplansky introduced the concept of a *radical* of any field F. This is the set of all $a \in \dot{F}$ for which the quaternion[4] algebra $(a, b/F)$ splits for all $b \in \dot{F}$. Equivalently, the radical is the intersection of the value sets of all one-fold Pfister forms over F. Kaplansky studied quadratic form theory over fields F with $Q = 2$, and for formally real fields, these are exactly the fields with the index of the radical in $\dot{F}$ equal to two. The higher radicals were introduced in [241], and Kaplansky's results were generalized to the context of n-fold Pfister forms in [225].

Cordes [40] proved that, if $Q(F) = 4$, the ETC holds for F, and Kula [136] proved that, if $Q(F) = 8$, the ETC holds for F. Another important piece of evidence for the ETC comes from the work of Jacob and Ware [105]. They proved that, for a field F with finite square class group, if $W(F)$ is a finite direct product of Witt rings, then the Witt ring $W(K)$ of any quadratic extension $K = F(\sqrt{a})$ is a finite direct product of abstract Witt rings describable in terms of factors of $W(F)$. In particular, when the Witt ring $W(F)$ is of elementary type, then for any $a \in \dot{F}$ the Witt ring $W(F(\sqrt{a}))$ is of elementary type.

Kula [137] studied the number of Witt rings for fields with a given number of square classes and its asymptotics. Scharlau [206] defined the generating function for a field F with finite group of square classes as the power series $1 + \sum_{i>0} c_i x^i$, where c_i is the number of equivalence classes of n-dimensional quadratic forms over F. If the Witt ring $W(F)$ is of elementary type, he proved that the series is a rational function with the only pole at $x = 1$. Kula [135] proved that this is so for an arbitrary field F with finite square class group. Kula, Marshall, and Sładek [138] investigated the structure of direct limits of torsion-free abstract Witt rings.

Various ring properties of Witt rings of fields and abstract Witt rings have been studied, such as units [154,164], level and sums of squares [153,155], automorphisms [146], annihilating polynomials [97,151,152], primary ideals, ideal class groups and Picard groups [78–80], and projectivity of $W(K)$ over $W(F)$ for a field extension K/F [81].

The best source for abstract Witt rings is Marshall's book [162]. For a comprehensive survey of work related to elementary type conjecture, citing 117 papers, see Marshall [165]. For an approach to an axiomatization of quadratic form theory in terms of the isometry relation on binary forms, see Dickmann and Miraglia [59,60].

10.2. *Reduced theory of quadratic forms*

For a formally real field F, there exists a complete theory of quadratic forms relative to a fixed *preordering* of the field. Here, a preordering in F is a proper subset T of F closed with respect to addition and multiplication and satisfying $\dot{F}^2 \subseteq T$. Any

[4] Actually, Kaplansky avoids quaternion algebras and uses Hilbert symbols instead. As to the term *radical of field*, Kaplansky says "Since any conventional concept of radical is ridiculous for a field, I trust there will be no confusion."

preordering T in F is contained in at least one ordering P of F (i.e., a preordering P satisfying additionally $P \cup -P = F$). The set of all orderings of F extending a given preordering T is denoted as X/T. Using the Kuratowski–Zorn Lemma, one shows that $T = \bigcap P$, where as P ranges over X/T. In case of $T = \sum \dot{F}^2$, this recaptures the Artin–Schreier characterization of totally positive elements in an ordered field.

A T-form is just a finite sequence $f = (a_1, \ldots, a_n)$ of elements in $\dot{F}$ and at each ordering $P \in X/T$, we have the signature $\operatorname{sgn}_P f$. Two forms f and g are said to be T-isometric if they have the same dimension n and have the same signatures with respect to any ordering P extending T. With these definitions, one can now follow the classical quadratic form theory pattern over fields and define the counterparts of all the classical concepts including the Witt ring, denoted as $W_T(F)$. In the special case, when $T = \sum \dot{F}^2$, it may be shown that $W_T(F) \cong W(F)/W_t(F)$, and this ring is said to be the *reduced* Witt ring of the field F.

The set $X = X/\sum \dot{F}^2$ of all orderings of the field F can be viewed as a topological space with the induced Zariski topology from the prime spectrum of the Witt ring $W(F)$. Alternatively, the *Harrison sets*

$$H(a) = \{P \in X : a \in P\}, \quad a \in \dot{F}$$

form a subbasis. X is a Boolean space and for each preordering T, the set X/T is closed, hence also a Boolean space. We now consider the ring $C(X/T, \mathbb{Z})$ of continuous functions from the space X/T to $\mathbb{Z}$ (with discrete topology) and the map

$$c_T : W_T(F) \to C(X/T, \mathbb{Z}), \quad c_T(f) = \hat{f},$$

where $\hat{f} : X/T \to \mathbb{Z}$, $\hat{f}(P) = \operatorname{sgn}_P f$. One checks that c_T is an injective ring homomorphism and it is of great interest to describe precisely the image of c_T. The functions in the image are said to be *represented* over T and the representation problem asks for a description of represented functions.

A crucial role in solving the representation problem is played by a special type of preorderings called *fans*. A preordering T is a *fan* when, for any subset S of F, if $-1 \notin S$ and $\dot{S}$ is a subgroup of $\dot{F}$, then S is a preordering in F. The ultimate result reads as follows.

THEOREM 10.2 (Representation Theorem, Becker and Bröcker [19]). *Let T_0 be a preordering in F. Then a function $f \in C(X/T_0), \mathbb{Z})$ is represented over T_0 iff for every fan $T \supseteq T_0$ of finite index in F,*

$$\sum_{P \in X/T} f(P) \equiv 0 \pmod{|X/T|}.$$

A comprehensive survey of the reduced theory, the relations to valuation theory, and detailed proofs are provided in Lam's notes [142]. There are 164 references there, but Lam says the list is incomplete and one should consult the bibliography in his survey on ordered fields [141] where there are 223 references.

11. Historical

Since the time of Fermat, eminent mathematicians were always interested in phenomena connected with quadratic forms. However, quadratic forms were considered usually over the integers or over the fields $\mathbb{R}$ and $\mathbb{C}$. The equivalence problem over integers is extraordinarily difficult and over $\mathbb{R}$ and $\mathbb{C}$ it is trivial. Hence, finding criteria for equivalence of quadratic forms was not the central problem. This changed at the end of 19th century when Minkowski [178] solved the equivalence problem for quadratic forms over the rationals, and Hilbert included in his list of 23 problems the following eleventh problem.[5]

Our present knowledge of the theory of quadratic number fields puts us in a position to attack successfully the theory of quadratic forms with any number of variables and with any algebraic numerical coefficients. This leads in particular to the interesting problem: to solve a given quadratic equation with algebraic numerical coefficients in any number of variables by integral or fractional numbers belonging to the algebraic realm of rationality determined by the coefficients.

Thus, Hilbert asks for solution of the representation problem over algebraic number fields and over rings of algebraic integers. As we know, in the case of fields, a solution of the representation problem implies a solution of the equivalence problem; hence, Hilbert's question addresses the more general problem.

Hilbert's high opinion on the *present knowledge* is certainly, at least in part, based on the results of his 1899 paper [91]. He presented there a complete theory of the Hilbert symbol in relative quadratic extensions of nonreal algebraic number fields with odd class number. For such a field K, he proved ([91, Section 44, Satz 66]) that the equation $ax^2 + by^2 = 1$ has a solution in K if and only if for every prime ideal $\mathfrak{p}$ of K we have $(a, b/\mathfrak{p}) = 1$. This condition is equivalent to the solvability of the congruence $ax^2 + by^2 \equiv 1 \pmod{\mathfrak{p}^n}$ in integers x, y in K for all exponents n. Hilbert does not say so, but the latter is equivalent to the solvability of the equation $ax^2 + by^2 = 1$ in the $\mathfrak{p}$-adic completions of the field K. Thus, Hilbert's result is a prototype of a local–global principle: the equation $ax^2 + by^2 = 1$ is solvable in the field K if and only if it is solvable in all completions of K.

Five papers of Hasse [86–90] give a complete solution to the algebraic part of Hilbert's problem XI. Hasse does not mention the Hilbert problem at all, but there is no doubt that his results form a *theory of quadratic forms* over number fields, as demanded by Hilbert. Hasse shows that two quadratic forms are equivalent over a number field F if and only if they are equivalent over all completions $F_\mathfrak{p}$ of F (where $\mathfrak{p}$ runs through all primes of F, including infinite primes). A similar principle holds for the representation problem: a quadratic form over number field F represents an element of the field F if and only if it represents that element over all completions $F_\mathfrak{p}$ of F. Thus, the solutions of the equivalence and representation problems over the number field F depend exclusively on solutions of these problems in the local fields, $F_\mathfrak{p}$. For local fields, Hasse found a complete solution of both problems in

[5] Translation taken from BAMS **8** (1902), 437–479.

principle equivalent to what follows from Theorem 6.1 before. In this way, the Hasse local–global principle becomes an effective tool for solving the equivalence and representation problems for quadratic forms over algebraic number fields.

Another important event in the 1920s was Artin's solution of Hilbert's 17th problem titled by Hilbert "Expression of definite forms by squares."

For a field F, let $S(F)$ denote the set of all sums of squares of elements of F, and let $T(F)$ be the set of all elements of F, that is, nonnegative at every ordering of F. Further, for the rational function field $K = F(X_1, \ldots, X_n)$ over an ordered field F, let $D(K)$ be the set of positive semidefinite elements of K. Thus, $D(K)$ consists of rational functions $f(X_1, \ldots, X_n) \in K$ with $f(a_1, \ldots, a_n)$ non-negative in F for all n-tuples $a_1, \ldots, a_n \in F$ in the domain of f. Clearly,

$$S(K) \subseteq T(K) \quad \text{and} \quad S(K) \subseteq D(K).$$

Hilbert's 17th problem asked whether $S(K) = D(K)$ when $F = \mathbb{R}$. Thus, the question was whether every positive semidefinite rational function over $\mathbb{R}$ can be written as the sum of squares of rational functions over $\mathbb{R}$. Obviously, this question reduces to representing positive semidefinite *polynomials* over $\mathbb{R}$ as sums of squares of *rational functions* over $\mathbb{R}$. Already Hilbert knew that there are semidefinite polynomials, which cannot be written as sums of squares of polynomials, but the first explicit example of such a polynomial was found only in 1967 by T. Motzkin,

$$f = 1 + X^2Y^4 + X^4Y^2 - 3X^2Y^2.$$

Hilbert also knew that the answer to his question is in the positive when $n = 2$. Hilbert's problem was solved by E. Artin in 1927. Two main results in his paper [8] assert that, first, $S(K) = T(K)$ for all fields K [with the convention that $T(K) = K$ for a nonreal field K], and second, $T(K) = D(K)$ for any rational function field $K = F(X_1, \ldots, X_n)$, where F is a number field with exactly one ordering, for example, $F = \mathbb{R}$. Thus, $S(K) = T(K) = D(K)$ for any $K = F(X_1, \ldots, X_n)$, where F is a subfield of $\mathbb{R}$ with a unique ordering. Artin's solution was based on the theory of formally real fields founded by Artin and Schreier in [9]. Brauer [22] comments on these results as follows: "His method was as remarkable as the result. It was perhaps the first triumph of what is sometimes called *abstract* algebra."

After Hasse and Artin, the third important paper of that time, and perhaps the most influential paper on quadratic forms, is the 1937 Ernst Witt paper [240]. He introduced the geometric language of "metric spaces," that is, vector spaces with quadratic (and bilinear) forms, and proved a series of theorems which became a standard foundation of quadratic form theory. Thus, he proves the orthogonal complement theorem, the existence of orthogonal bases, the cancellation theorem (Satz 4), the theorem on decomposition of isotropic forms into a direct sum of hyperbolic and anisotropic forms. Then he proves that classes of similar forms form a commutative ring (Satz 6) and proves the chain equivalence theorem (Satz 7). In the second part of the paper, he introduces the Clifford algebra of an arbitrary quadratic form (not just sum of squares) and uses it to prove that the Hasse algebra of a quadratic form is a well-defined equivalence invariant. It is amusing that Witt used for this Clifford algebras

and did not notice that his chain equivalence theorem (Satz 7) does the job effortlessly. Then he proceeds to a series of classification results of which the first (Satz 11) is Theorem 4.1 above. In the third part of the paper Witt gives new, concise and lucid proof of the Hasse principle. In particular, he deduces the local–global principle for equivalence from the principle for the representation problem using his cancellation theorem. The final section of the paper solves the classification problem over real algebraic function fields in one variable.

While Witt's paper is of utmost importance, it is interesting and surprising that at some points it was anticipated by a 1907 Dickson paper [61]. As reported by Scharlau [207], Dickson not only found the diagonalizability of quadratic forms over fields of characteristic $\neq 2$ but also proved the cancellation theorem! It is a melancholy thought to realize that it took almost 100 years to recognize Dickson's contribution.

After Witt's paper, there has been very little activity in quadratic form theory with sporadic, albeit important, publications such as Kaplansky's paper [108] mentioned in Section 7.2 and Springer [219, 220]. Then in 1965, the first paper of Pfister [190] appeared with the solution of the level problem for fields. In his next papers [191, 192], the local–global principle (Theorem 3.3) for torsion forms was proved and the quantitative solution of Hilbert's 17th problem was obtained (Theorem 8.5). These results attracted the attention of many researchers and were the starting point of continual activity in quadratic form theory over fields. The progress was surveyed by Pfister himself in at least five surveys [193–196, 199]. Also Scharlau [207] gives an overview, which concentrates on "some aspects neglected in the existing literature."

References

[1] A. Adem, W. Gao, D.B. Karagueuzian, J.Ináč, Field theory and the cohomology of some Galois groups, J. Algebra **235** (2001), 608–635.

[2] J.K. Arason, Cohomologische Invarianten quadratischer Formen, J. Algebra **36** (1975), 448–491.

[3] J.K. Arason, Der Wittring projektiver Räume, Math. Ann. **253** (1980), 205–212.

[4] J.K. Arason, *A proof of Merkurjev's theorem.* CMS Conference Proceedings, vol. **4**, American Mathematical Society Providence, RI, 1984, pp. 121–130; *Quadratic and Hermitian forms*, Hamilton, Ontario, 1983.

[5] J.K. Arason, A. Pfister, Beweis des Krullschen Durchschnittsatzes für den Wittring, Invent. Math., **12** (1971), 173–176.

[6] J.K. Arason, R. Elman, Powers of the fundamental ideal in the Witt ring, J. Algebra **239** (2001), no. 1, 150–160.

[7] J.K. Arason, R. Elman, B. Jacob, On the Witt ring of elliptic curves, Proc. Symp. Pure Math. **58**, Part II (1995), 1–25.

[8] E. Artin, Über die Zerlegung definiter Funktionen in Quadrate, Abh. Math. Sem. Univ. Hamburg **5** (1927), 100–115.

[9] E. Artin, O. Schreier, Algebraische Konstruktion reeller Körper, Abh. Math. Sem. Univ. Hamburg **5** (1927), 85–99.

[10] R. Baeza, Lecture Notes in Mathematics, vol. **655**: *Quadratic forms over semilocal rings*, Springer Verlag, Berlin, 1978.

[11] R. Baeza, J.S. Hsia, B. Jacob, A. Prestel, eds., Contemporary Mathematics, vol. **344**: *Algebraic and Arithmetic Theory of Quadratic forms.* American Mathematical Society Providence, RI, 2004.

[12] R. Baeza, R. Moresi, On the Witt-equivalence of fields of characteristic 2, J. Algebra **92** (1985), 446–453.
[13] P. Balmer, Triangular Witt groups, I: The 12−term localization exact sequence, K-theory **19** (2000), 311–363.
[14] P. Balmer, Triangular Witt groups, II: From usual to derived, Math. Zeit. **236** (2001), 351–382.
[15] P. Balmer, Handbook of K-theory, vol. **2**: Witt Groups, Springer Verlag, 2005, pp. 539–576.
[16] E. Bayer-Fluckiger, R. Parimala, Galois cohomology of the classical groups over fields of cohomological dimension ≤ 2, Invent. Math. **122** (1995), 195–229.
[17] E. Becker, Hereditarily Pythagorean Fields and Orderings of Higher Level, vol. **29**: *Monografias de Matemática*, Instituto de Matemática Aplicada, Rio de Janeiro, 1978.
[18] E. Becker, Lecture Notes in Mathematics, vol. **959**: *The real holomorphy ring and sums of* 2*nth powers*, 1982, pp. 139–181.
[19] E. Becker and L. Bröcker, On the description of the reduced Witt ring, J. Algebra **52** (1978), 328–346.
[20] E. Becker, J. Harman, and A. Rosenberg, Signatures of fields and extwnsion theory, J. Reine Angew. Math. **330** (1982), 53–75.
[21] E. Becker, A. Rosenberg, Reduced forms and reduced Witt rings of higher level, J. Algebra **92** (1985), 477–503.
[22] R. Brauer, Emil Artin, Bull. Am. Math. Soc. **73** (1967), 27–43.
[23] L. Bröcker, Über die Anzahl der Anordnungen eines kommutativen Körpers, Arch. Math. **29** (1977), 458–464.
[24] J.P. Carpenter, Finiteness theorems for forms over global fields, Math. Zeit. **209** (1992), 153–166.
[25] A.B. Carson, M.A. Marshall, Decomposition of Witt rings, Can. J. Math. **34** (1982), 1276–1302.
[26] J.W.S. Cassels, On the representation of rational functions as sums of squares, Acta Arith. **9** (1964), 79–82.
[27] J.W.S. Cassels, W.J. Ellison, A. Pfister, On sums of squares and on elliptic curves over function fields, J. Number Theory **3** (1971), 125–149.
[28] M.D. Choi, Z.D. Dai, T.Y. Lam, B. Reznick, The Pythagoras number of some affine algebras and local algebras, J. Reine Angew. Math. **336** (1982), 45–82.
[29] M.D. Choi, T.Y. Lam, A. Prestel, B. Reznick, Sums of 2*m*th powers of rational functions in one variable over real closed fields, Math. Zeit. **221** (1996), 93–112.
[30] M. Ciemała, K. Szymiczek, On natural homomorphisms of Witt rings, Proc. Am. Math. Soc. **133** (2005), 2519–2523.
[31] M. Ciemała, K. Szymiczek, On injectivity of natural homomorphisms of Witt rings, Ann. Math. Siles. **21** (2007), 15–30.
[32] J.-L. Colliot-Thélène, U. Jannsen, Sommes de carrés dans les corps de fonctions, C.R. Acad. Sci. Paris **312** (1991), 759–762.
[33] J.-L. Colliot-Thélène, R. Sujatha, The unramified Witt group of real anisotropic quadrics, Proc. Symp. Pure Math. **58**, Part II (1995), 127–147.
[34] P.E. Conner, R. Perlis, *A Survey of Trace Forms of Algebraic Number Fields*, World Scientific, Singapore, 1984.
[35] P.E. Conner, R. Perlis, K. Szymiczek, Wild sets and 2-ranks of class groups, Acta Arith. **79** (1997), 83–91.
[36] P.E. Conner, N. Yui, The additive characters of the Witt ring of an algebraic number field, Can. J. Math. **40** (1988), 546–588.
[37] J.H. Conway, N.J.A. Sloane, Grundlehren der mathematischen Wissenschaften, vol. **209**: *Sphere Packings, Lattices, and Groups*, 3rd ed., Springer-Verlag, New York, 1988.
[38] C.M. Cordes, The Witt group and equivalence of fields with respect to quadratic forms, J. Algebra **26** (1973), 400–421.
[39] C.M. Cordes, Quadratic forms over nonformally real fields with a finite number of quaternion algebras, Pacific J. Math. **63** (1976), 357–365.
[40] C.M. Cordes, Quadratic forms over fields with four quaternion algebras, Acta Arith. **41** (1982), 55–70.

[41] T.C. Craven, A. Rosenberg, R. Ware, The map of the Witt ring of a domain into the Witt ring of its field of fractions, Proc. Am. Math. Soc. **51** (1975), 25–30.
[42] T.C. Craven, The Boolean space of orderings of a field, Trans. Am. Math. Soc. **209** (1975), 225–235.
[43] T.C. Craven, Characterizing reduced Witt rings of fields, J. Algebra **53** (1978), 68–77.
[44] T.C. Craven, Characterizing reduced Witt rings, II. Pacific J. Math. **80** (1979), 341–349.
[45] T.C. Craven, On the topological space of all orderings of a skew field, Arch. Math. **36** (1981), 234–238.
[46] T.C. Craven, Witt rings and orderings of skew fields, J. Algebra **77** (1982), 74–96.
[47] A. Czogała, On reciprocity equivalence of quadratic number fields, Acta Arith. **58** (1991), 27–46.
[48] A. Czogała, On integral Witt equivalence of algberaic number fields, Acta Math. Inf. Univ. Ostrav. **4** (1996), 7–21.
[49] A. Czogała, Higher degree tame Hilbert-symbol equivalence of number fields, Abh. Math. Sem. Univ. Hamburg **69** (1999), 175–185.
[50] A. Czogała, Witt rings of Hasse domains of global fields, J. Algebra **244** (2001), 604–630.
[51] A. Czogała, Hilbert-symbol equivalence of global function fields, Math. Slovaca **51** (2001), 393–401.
[52] A. Czogała, A. Sładek, Higher degree Hilbert-symbol equivalence of number fields, I. Tatra Mount. Math. Publ. **11** (1997), 77–88.
[53] A. Czogała, A. Sładek, Higher degree Hilbert-symbol equivalence of number fields, II. J. Number Theory **72** (1998), 363–376.
[54] A. Czogała, B. Rothkegel, Witt equivalence of semilocal dedekind domains in global fields, Abh. Math. Sem. Univ. Hamburg **77** (2007), 1–24.
[55] A. Czogała, A. Sładek, Witt rings of global fields, Commun. Algebra **31** (2003), 3195–3205.
[56] Z.D. Dai, T.Y. Lam, C.K. Peng, Levels in algebra and topology, Bull. Am. Math. Soc. **3** (1980), 845–848.
[57] Z.D. Dai, T.Y. Lam, Levels in algebra and topology, Comment. Math. Helv. **59** (1984), 376–424.
[58] F.R. DeMeyer, D.K. Harrison, R. Miranda, Quadratic forms over $\mathbb{Q}$ and Galois extensions of commutative rings, Mem. Am. Math. Soc. **77**, no. 394, 1989, 1–63.
[59] M.A. Dickmann, F. Miraglia, Special groups: Boolean-theoretic methods in the theory of quadratic forms, Mem. Am. Math. Soc. **145**, no. 689, 2000.
[60] M.A. Dickmann, F. Miraglia, Algebraic K-theory of special groups, J. Pure Appl. Algebra **204** (2006), 195-234.
[61] L.E. Dickson, On quadratic forms in a general field, Bull. Am. Math. Soc. **14** (1907), 108–115.
[62] A.W.M. Dress, The weak local global principle in algebraic K-theory, Commun. Algebra **3** (1975), no. 7, 615–661.
[63] D.W. Dubois, T. Recio, eds., Contemporary Mathematics, vol. **8**: *Ordered Fields and Real Algebraic Geometry*, American Mathematical Society, Providence, RI, 1982, pp. 1–360.
[64] D. Dugger, D. Isaksen, The Hopf condition for bilinear forms over arbitrary fields, Ann. Math. **165** (2007), 943–964.
[65] R. Elman, N. Karpenko, A. Merkurjev, AMS Colloquium Publications, vol. **56**: *The Algebraic and Geometric Theory of Quadratic Forms*, AMS, Providence, RI, 2008.
[66] R. Elman, T.Y. Lam, Quadratic forms and the u-invariant, I. Math. Zeit. **131** (1973), 283–304.
[67] R. Elman, T.Y. Lam, Quadratic forms and the u-invariant, II. Invent. Math. **21** (1973), 125–137.
[68] R. Elman, T.Y. Lam, Classification theorems for quadratic forms over fields, Comment. Math. Helv. **49** (1974), 373–381.
[69] R. Elman, T.Y. Lam, A.R. Wadsworth, Amenable fields and Pfister extensions, Queen's Papers in Pure and Appl. Math. **46** (1976), 445–492.
[70] R. Elman, T.Y. Lam, A.R. Wadsworth, Function fields of Pfister forms, Invent. Math. **51** (1979), 61–75.
[71] R. Elman, T.Y. Lam, A.R. Wadsworth, Quadratic forms under multiquadratic extensions, Indag. Math. **42** (1980), 131–145.
[72] R. Elman, T.Y. Lam, J.-P. Tignol, A.R. Wadsworth, Witt rings and Brauer groups under multiquadratic extensions, I. Am. J. Math. **105** (1983), 1119–1170.

[73] M. Epkenhans, Trace forms of normal extensions of local fields, Linear and Multilinear Algebra **24** (1989), 103–116.
[74] M. Epkenhans, Trace forms of normal extensions of algebraic number fields, Linear and Multilinear Algebra **25** (1989), 309–320.
[75] M. Epkenhans, On the ramification set of a positive quadratic form over an algebraic number field, Acta Arith. **66** (1994), 133–145.
[76] D.R. Estes, J. Hurrelbrink, R. Perlis, Total positivity and algebraic Witt classes, Comment. Math. Helvetici **60** (1985), 284–290.
[77] J. Fernando, J. Ruiz, C. Scheiderer, Sums of squares in real rings, Trans. Am. Math. Soc. **356** (2004), 2663–2684.
[78] R.W. Fitzgerald, Primary ideals in Witt rings, J. Algebra **96** (1985), 368–385.
[79] R.W. Fitzgerald, Ideal class groups of Witt rings, J. Algebra **124** (1989), 506–520.
[80] R.W. Fitzgerald, Picard groups of Witt rings, Math. Zeit. **296** (1991), 303–319.
[81] R.W. Fitzgerald, Witt rings under odd degree extensions, Pacific J. Math. **158** (1993), 121–143.
[82] W. Gao, J. Mináč, Milnor's conjecture and Galois theory, I. Fields Inst. Commun. **16** (1997), 95–110.
[83] L.J. Gerstein, Basic Quadratic Forms, Graduate Studies in Math., vol. **90**. Am. Math. Soc., Providence, RI., 2008.
[84] D.K. Harrison, *Witt Rings*, University of Kentucky, Lexington, Kentucky, 1970. Notes by Joel Cunningham.
[85] D.K. Harrison, B. Pareigis, Witt rings of higher degree forms, Commun. Algebra **16** (1988), 1275–1313.
[86] H. Hasse, Über die Darstellbarkeit von Zahlen durch quadratische Formen im Körper der rationalen Zahlen, J. Reine Angew. Math. **152** (1923), 129–148.
[87] H. Hasse, Über die Äquivalenz quadratischer Formen im Körper der rationalen Zahlen, J. Reine Angew. Math. **152** (1923), 205–224.
[88] H. Hasse, Symmetrische Matrizen im Körper der rationalen Zahlen, J. Reine Angew. Math. **153** (1924), 12–43.
[89] H. Hasse, Darstellbarkeit von Zahlen durch quadratische Formen in einem beliebigen algebraischen Zahlkörper, J. Reine Angew. Math. **153** (1924), 113–130.
[90] H. Hasse, Äquivalenz quadratischer Formen in einem beliebigen algebraischen Zahlkörper, J. Reine Angew. Math. **153** (1924), 158–162.
[91] D. Hilbert, Über die Theorie des relativ quadratischen Zahlkörpers, Math. Ann. **51** (1899), 1–127.
[92] D.W. Hoffmann, Isotropy of quadratic forms over the function field of a quadric, Math. Zeit. **220** (1995), 461–476.
[93] D.W. Hoffmann, Pythagoras numbers of fields, J. Am. Math. Soc. **12** (1999), 839–848.
[94] D.W. Hoffmann, Isotropy of quadratic forms and field invariants, Contemp. Math. **272** (2000), 73–102.
[95] D.W. Hoffmann, *Levels of quaternion algebras*, Arch. Math. (Basel) **90** (2008), 401–411.
[96] E.A.M. Hornix, Formally real fields with prescribed invariants in the theory of quadratic forms, Indag. Math. (N.S.) **2** (1991), no. 1, 65–78.
[97] J. Hurrelbrink, Annihilating polynomials for group rings and Witt rings, Can. Math. Bull. **32** (1989), 412–416.
[98] J. Hurrelbrink, N.A. Karpenko, U. Rehman, The minimal height of quadratic forms of given dimension, Arch. Math. **87** (2006), 522–529.
[99] O.T. Izhboldin, Fields of u-invariant 9, Ann. Math. **154** (2001), 529–587.
[100] O.T. Izhboldin, B. Kahn, N.A. Karpenko, A. Vishik, Lecture Notes in Mathematics, vol. **1835**: *Geometric Methods in the Algebraic Theory of Quadratic Forms*, Springer, Berlin, Heidelberg, 2004.
[101] B. Jacob, Quadratic forms over dyadic valued fields, I. The graded Witt ring, Pacific J. Math. **126** (1987), 21–79.
[102] B. Jacob, Quadratic forms over dyadic valued fields, II. Relative rigidity and Galois cohomology, J. Algebra **148** (1992), 162–202.

[103] W.B. Jacob, M. Rost, Degree four cohomological invariants for quadratic forms, Invent. Math. **96** (1989), 551–570.
[104] W.B. Jacob, R. Ware, A recursive description of the maximal pro-2 Galois group via Witt rings, Math. Zeit **200** (1989), 379–396.
[105] W.B. Jacob, R. Ware, Realizing dyadic factors of elementary type Witt rings and pro−2 Galois groups, Math. Zeit. **208** (1991), 193–208.
[106] W.B. Jacob, T.Y. Lam, R.O. Robson, eds., Contemporary Mathematics, vol. **155**: *Recent Advances in Real Algebraic Geometry and Quadratic Forms*, American Mathematical Society, Providence, RI, 1994.
[107] W.B. Jacob, A. Rosenberg, ed., *K-theory and algebraic geometry: Connections with quadratic forms and division algebras*, in Proceedings of Symposia in Pure Mathematics, American Mathematical Society, Providence, RI, 1995.
[108] I. Kaplansky, Quadratic forms, J. Math. Soc. Japan **5** (1953), 200–207.
[109] I. Kaplansky, Fröhlich's local quadratic forms, J. Reine Angew. Math. **239/240** (1969), 74–77.
[110] N. Karpenko, On the first Witt index of quadratic forms, Invent. Math. **153** (2003), 455–462.
[111] N. Karpenko, Holes in I^n, Ann. Sci. École Norm. Sup. (4) **37** (2004), no. 6, 973–1002.
[112] N. Karpenko, *A relation between higher Witt indices*, Proceedings of the St.Petersburg Mathematical Society. vol. XI, pp. 77–86, Amer. Math. Soc. Transl. Ser. 2, **218**, American Mathematical Society, Providence, RI, 2006.
[113] N. Karpenko, A. Merkurjev, Essential dimensions of quadrics, Invent. Math. **153** (2003), 361–372.
[114] M.-H. Kim, J.S. Hsia, Y. Kitaoka, R. Schulze-Pillot, ed., Contemporary Mathematics, Integral quadratic forms and lattices, vol. **249**, American Mathematical Society, Providence, RI, 1999.
[115] M. Knebusch, *Grothendieck- und Wittringe von nichtausgearteten symmetrischen Bilinearformen.* Sitzungsber. Heidelb. Akad. Wiss. Mat.-naturw. Kl. vol. **3**. Springer Verlag, Berlin, 1969/1970.
[116] M. Knebusch, Runde Formen über semilokalen Ringen, Math. Ann. **193** (1971), 21–34.
[117] M. Knebusch, Real closures of commutative rings, I. J. Reine Angew. Math. **274/275** (1975), 61–89.
[118] M. Knebusch, Real closures of commutative rings, II. J. Reine Angew. Math. **286/287** (1976), 278–313.
[119] M. Knebusch, Generic splitting of quadratic forms, I. Proc. London Math. Soc. (3) **33** (1) (1976), 65–93.
[120] M. Knebusch, Symmetric bilinear forms over algebraic varieties, Queen's Papers in Pure and Appl. Math. **46** (1977), 103–283.
[121] M. Knebusch, A. Rosenberg, R. Ware, Structure of Witt rings and quotients of Abelian group rings, Am. J. Math. **94** (1972), 119–155.
[122] M. Knebusch, A. Rosenberg, R. Ware, Grothendieck and Witt rings of hermitian forms over Dedekind rings, Pacific J. Math. **43** (1972), 657–673.
[123] M. Knebusch, A. Rosenberg, R. Ware, Signatures on semi-local rings, J. Algebra **26** (1973), 208–250.
[124] M. Knebusch, U. Rehmann, Generic splitting towers and generic splitting preparation of quadratic forms, Contemp. Math. **272** (2000), 173–199.
[125] M. Kneser, *Quadratische Formen*, Springer-Verlag, Berlin, 2002.
[126] M.-A. Knus, *Quadratic and Hermitian forms over rings*, Springer-Verlag, Berlin-Heidelberg-New York, 1991.
[127] M.-A. Knus, A. Merkurjev, M. Rost, J.-P. Tignol, American Mathematical Society Colloqium Publication, vol. **44**: *The book of involutions*. American Mathematical Society, Providence, RI, 1998.
[128] M.-A. Knus, M. Ojanguren, R. Sridharan, Quadratic forms and Azumaya algebras, J. Reine Angew. Math. **303/304** (1978), 231–248.
[129] P. Koprowski, Witt equivalence of algebraic function fields over real closed fields, Math. Zeit. **242** (2002), 323–345.
[130] P. Koprowski, On existence of tame Harrison map, Math. Slovaca **57** (2007), 407–414.
[131] M. Krüskemper, Algebraic number field extensions with prescribed trace forms, J. Number Theory **40** (1992), 120–124.
[132] M. Kula, Fields with prescribed quadratic form schemes, Math. Zeit. **167** (1979), 201–212.

[133] M. Kula, Fields with non-trivial Kaplansky's radical and finite square class number, Acta Arith. **37** (1981), 411–418.
[134] M. Kula, Fields and quadratic form schemes, Ann. Math. Siles. **1** (1985), 7–22.
[135] M. Kula, Crystal growth and Witt rings, J. Algebra **136** (1991), 190–196.
[136] M. Kula, *Finitely generated Witt rings*, Uniw. Śląski w Katowicach **1207** (1991), 1–52.
[137] M. Kula, Counting Witt rings, J. Algebra **206** (1998), 568–587.
[138] M. Kula, M. Marshall, A. Sładek, Direct limits of finite spaces of orderings, Pacific J. Math. **112** (1984), 391–406.
[139] M. Kula, L. Szczepanik, K. Szymiczek, Quadratic forms over formally real fields with eight square classes, Manuscripta Math. **29** (1979), 295–303.
[140] J. Labute, N. Lemire, J. Mináč, J. Swallow, Demuškin groups, Galois modules, and the elementary type conjecture, J. Algebra **304** (2006), 1130–1146.
[141] T.Y. Lam, *The theory of ordered fields*, in: B. McDonald, ed., *Ring Theory and Algebra III*, Lecture Notes in Pure and Applied Mathematics, vol. **55**, Dekker, New York, 1980, pp. 1–152,
[142] T.Y. Lam, Conference Board of Mathematical Sciences Lecture Notes Series, vol. **52**: *Orderings, Valuations and Quadratic Forms*, American Mathematical Society, Providence, RI, 1983.
[143] T.Y. Lam, Graduate Studies in Mathematics, vol. **67**: *Introduction to Quadratic Forms over Fields*, American Mathematical Society, Providence, RI, 2005.
[144] T.Y. Lam, *Fields of u-invariant 6 after A. Merkurjev*, in: L. Rowen, *Ring Theory 1989*. Proc. Symp. and Workshop, Jerusalem/Isr. 1988/89, *Israel Mathematical Conference Proceedings* **1**, pages 12–30. Weizmann Science Press of Israel, 1989.
[145] D.B. Leep, Systems of quadratic forms, J. Reine Angew. Math. **350** (1984), 109–116.
[146] D.B. Leep, M.A. Marshall, Isomorphisms and automorphisms of Witt rings, Can. Math. Bull. **31** (1988), 250–256.
[147] D.B. Leep, A.S. Merkurjev, Growth of the u-invariant under algebraic extensions, Contemp. Math. **155** (1994), 327–332.
[148] D.B. Leep, D.B. Shapiro, A.R. Wadsworth, Sums of squares in division algebras, Math. Zeit. **190** (1985), 151–162.
[149] D.B. Leep, T.L. Smith, Witt kernels of triquadratic extensions, Contemp. Math. **344** (2004), 249–256.
[150] D.W. Lewis, Levels and sublevels of division algebras, Proc. Roy. Irish Acad. Sect. A **87** (1987), 103–106.
[151] D.W. Lewis, Witt rings as integral rings, Invent. Math. **90** (1987), 631–633.
[152] D.W. Lewis, New proofs of the structure theorems for Witt rings, Expos. Math. **7** (1989), 83–88.
[153] D.W. Lewis, Sums of squares in Witt rings, J. Pure Appl. Algebra **69** (1990), 67–72.
[154] D.W. Lewis, Units in Witt rings, Commun. Algebra **18** (1990), 3295–3306.
[155] D.W. Lewis, The level of a Witt ring, Linear and Multilinear Algebra **27** (1990), 163–165.
[156] D.W. Lewis, *Levels of fields and quaternion algebras—a survey.* Théorie des nombres, Années 1996/97–1997/98, 9 pp., Publ. Math. UFR Sci. Tech. Besançon, Univ. Franche-Comté, Besançon, 1999.
[157] D.W. Lewis, J.-P. Tignol, Square class groups and Witt rings of central simple algebras, J. Algebra **154** (1993), 360–376.
[158] D.W. Lewis, J.-P. Tignol, Classification theorems for central simple algebras with involution, Manuscripta Math. **100** (1999), 259–276.
[159] F. Lorenz, J. Leicht, Die Primideale des Wittschen ringes, Invent. Math. **10** (1970), 82–88.
[160] P. Mammone, J.-P. Tignol, A.R. Wadsworth, Fields of characteristic 2 with prescribed u-invariants, Math. Ann. **290** (1991), 109–128.
[161] M.A. Marshall, Classification of finite spaces of orderings, Can. J Math. **31** (1979), 320–330.
[162] M.A. Marshall, Queen's Papers in Pure and Applied Mathematics, vol. **57**: *Abstract Witt Rings*, Queen's University, Kingston, Ontario, 1980.
[163] M.A. Marshall, Queen's Papers in Pure and Applied Mathematics, vol. **71**: *Bilinear Forms and Orderings on Commutative Rings*. Queen's University, Kingston, Ontario, 1985.
[164] M.A. Marshall, Exponentials and logarithms on Witt rings, Pacific J. Math. **127** (1987), 127–140.

[165] M.A. Marshall, The elementary type conjecture in quadratic form theory, Contemp. Math. **344** (2004), 275–293.
[166] M.A. Marshall, V. Powers, Higher level form schemes, Commun. Algebra **21** (1993), 4083–4102.
[167] M. Marshall, J. Yucas, Linked quaternionic mappings and their associated Witt rings, Pacific J. Math. **95** (1981), 411–425.
[168] A.S. Merkurjev, *On the norm residue symbol of degree* 2. Dokl. Akad. Nauk SSSR **261** (1981), 542–547. English Translation: Soviet Math. Dokl. **24** (1981), 546–551.
[169] A.S. Merkurjev, *Kaplansky's conjecture in the theory of quadratic forms.* (Russian). Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **175** (1989), Koltsa i Moduli **3**, 75–89, 163–164; translation in J. Soviet Math. **57** (1991), no. 6, 3489–3497.
[170] A.S. Merkurjev, *Simple algebras and quadratic forms.* (Russian). Izv. Akad. Nauk SSSR, Ser. Mat. **55** (1991), 218–224.
[171] A.S. Merkurjev, *On the norm residue homomorphism of degree two*, Proceedings of the St. Petersburg Mathematical Society. vol. XII, pp. 103–124, Am. Math. Soc. Transl. Ser. 2, **219**, American Mathematica Society, Providence, RI, 2006.
[172] J.-F. Mestre, Extensions régulières de $\mathbb{Q}(t)$ de groupe de Galois $\tilde{A}_n$, J. Algebra **131** (1990), 483–495.
[173] J. Milnor, Algebraic K-Theory and Quadratic Forms, Invent. Math. **9** (1970), 318–344.
[174] J. Milnor, D. Husemoller, *Symmetric Bilinear Forms*. Ergebnisse der Mathematik und ihrer Grenzgebiete vol. **73**. Springer-Verlag, Berlin Heidelberg New York, 1973.
[175] J.Ináč, M. Spira, Witt rings and Galois groups, Ann. of Math. **144** (1996), 35–60.
[176] J.Ináč, T. Smith, Decomposition of Witt rings and Galois groups, Can. J Math. **47** (1995), 1274–1289.
[177] J.Ináč, A.R. Wadsworth, The u-invariant for algebraic extensions, Proc. Symp. Pure Math. **58**, Part II (1995), 333–358.
[178] H. Minkowski, *Ueber die Bedingungen, unter welchen zwei quadratische Formen mit rationalen Coeffizienten rational in einander transformirt werden knnen.* (Auszug aus einem von Herrn H. Minkowski in Bonn an Herrn Adolf Hurwitz gerichteten Briefe), J. Reine Angew. Math. **106** (1890), 5–26.
[179] F. Morel, Voevodsky's proof of Milnor's conjecture, Bull. Am. Math. Soc. **35** (1998), 123–143.
[180] M. Ojanguren, R. Parimala, R. Sridharan, Symplectic bundles over affine surfaces, Comment. Math. Helv. **61** (1986), 491–500.
[181] O.T. O'Meara, Grundlehren der mathematischen Wissenschaften, vol. **117**: *Introduction to Quadratic Forms*, 3rd ed., Springer-Verlag, Berlin Göttingen Heidelberg, 2000.
[182] D. Orlov, A. Vishik,V. Voevodsky, An exact sequence for $K_*^M/2$ with applications to quadratic forms, Ann. of Math. **165** (2007), 1–13.
[183] K. Osiak, A Cantor cube as a space of higher level orderings, Tatra Mount. Math. Publ. **32** (2005), 71–84.
[184] K. Osiak, The Boolean space of higher level orderings, Fund. Math. **196** (2007), 101–117.
[185] K. Osiak, A. Sładek, A note on number of orderings of higher level, Arch. Math. **86** (2006), 101–110.
[186] R. Parimala, Witt groups of conics, elliptic, and hyperelliptic curves, J. Number Theory **28** (1988), 69–93.
[187] R. Parimala, R. Sujatha, Witt groups of hyperelliptic curves, Comment. Math. Helv. **65** (1990), 559–580.
[188] R. Parimala, R. Sridharan, Non-surjectivity of the Clifford invariant map, Proc. Indian Acad. Sci. (Math. Sci.) **104** (1994), 49–56.
[189] R. Perlis, K. Szymiczek, P.E. Conner, R. Litherland, Matching Witts with global fields, Contemp. Math. **155** (1994), 365–387.
[190] A. Pfister, Zur Darstellung von -1 als Summe von Quadraten in einem Körper, J. London Math. Soc. **40** (1965), 159–165.
[191] A. Pfister, Quadratische formen in beliebigen Körpern, Invent. Math. **1** (1966), 116–132.
[192] A. Pfister, Zur Darstellung definiter Funktionen als summe von quadraten, Invent. Math. **4** (1967), 229–237.

[193] A. Pfister, Neuere entwicklungen in der theorie der quadratischen formen, Jber. Deutsch. Math.-Verein. **74** (1972), 131–142.
[194] A. Pfister, Über einige neuere ergebnisse aus der algebraischen theorie der quadratischen formen, Abh. Braunschw. Wiss. Gesellschaft **32** (1982), 61–68.
[195] A. Pfister, *Some remarks on the historical development of the algebraic theory of quadratic forms.* CMS Conference Proceedings, vol. **4**, American Mathematica Society, Providence, RI, 1984. pp. 1–16; Quadratic and Hermitian forms (Hamilton, Ontario, 1983).
[196] A. Pfister, Quadratische formen, Dok. Gesch. Math. **6** (1990), 657–671.
[197] A. Pfister, Quadratic lattices in function fields of genus 0, Proc. London Math. Soc. (3)**66** (1993), 257–278.
[198] A. Pfister, *Lecture Notes Series*, vol. **217**: *Quadratic Forms with Applications to Algebraic Geometry and Topology.* London Math. Soc. Cambridge Univ. Press, Cambridge, 1995.
[199] A. Pfister, on the milnor conjectures: history, influence, applications, Jber. Dt. Math.-Verein. **102** (2000), 15–41.
[200] Y. Pourchet, Sur la représentation en somme de carrés des polynômes à une indéterminée sur un corps de nombres algébriques, Acta Arith. **19** (1971), 89–104.
[201] A. Prestel, Remarks on the Pythagoras and Hasse number of real fields, J. Reine Angew. Math. **303/304** (1978), 284–294.
[202] A. Prestel, On trace forms of algebraic function fields, Rocky Mountain J. Math. **19** (1989), 897–911.
[203] P. Revoy, Niveaux d'ordre supérieur des corps et des anneaux, C.R. Acad. Sci. Paris Sér. I Math. **307** (1988), 203–206.
[204] W. Scharlau, Grundlehren der mathematischen Wissenschaften, vol. **270**: *Quadratic and Hermitian Forms*, Springer-Verlag, Berlin Heidelberg New York Tokyo, 1985.
[205] W. Scharlau, On trace forms of algebraic number fields, Math. Zeit. **196** (1987), 125–127.
[206] W. Scharlau, Generating functions of finitely generated Witt rings, Acta Arith. **54** (1989), 51–59.
[207] W. Scharlau, On the history of the algebraic theory of quadratic forms, Contemp. Math. **272** (2000), 229–259.
[208] J.-P. Serre, L'invariant de Witt de la forme $\mathrm{tr}(x^2)$, Comment. Math. Helvetici **59** (1984), 651–676.
[209] D.B. Shapiro, *Compositions of Quadratic Forms*, De Gruyter Expositions in Mathematics, vol. **32**, W. de Gruyter & Co., Berlin, 2000.
[210] D.B. Shapiro, J.-P. Tignol, A.R. Wadsworth, Witt rings and Brauer groups under multiquadratic extensions, II. J. Algebra **78** (1982), 58–90.
[211] A.S. Sivatski, Nonexcellence of multiquadratic field extensions, J. Algebra **275** (2004), 859–866.
[212] A.S. Sivatski, Nonexcellence of the function field of the product of two conics, K-Theory **34** (2005), 209–218.
[213] A. Sładek, Witt rings of quaternion algebras, J. Algebra **103** (1986), 267–272.
[214] A. Sładek, *Witt rings of complete skew fields*. II, Wiss. Beitr. Martin-Luther-Univ. Halle-Wittenberg, Halle (Saale) **33** (1986), 239–243.
[215] A. Sładek, Witt rings of complete skew fields, Pacific J. Math. **132** (1988), 391–399.
[216] A. Sładek, Witt rings of division algebras over global fields, Commun. Algebra **18** (1990), 2159–2175.
[217] A. Sładek, Witt rings of even degree division algebras over number fields, Commun. Algebra **22** (1994), 3531–3543.
[218] A. Sładek, Higher degree Harrison equivalence and Milnor K-functor, Acta Math. Inf. Univ. Ostrav. **6** (1998), 183–189.
[219] T.A. Springer, Sur les formes quadratiques d'indice zéro, C.R. Acad. Sci. Paris **234** (1952), 1517–1519.
[220] T.A. Springer, Quadratic forms over a field with a discrete valuation, Indag. Math. **17** (1955), 352–362.
[221] L. Szczepanik, Fields and quadratic form schemes with the index of radical not exceeding 16, Ann. Math. Siles. **1** (1985), 23–46.

[222] M. Szyjewski, *The fifth invariant of quadratic forms.* (Russian), Dokl. Akad. Nauk SSSR **308** (1989), 542–545. Translation in: Soviet Math. Doklady **40** (1990), 355–358.
[223] M. Szyjewski, On the Witt ring of a relative projective line, Coll. Math. **75** (1998), 53–78.
[224] K. Szymiczek, Quadratic forms over fields with finite square class number, Acta Arith. **28** (1975), 195–221.
[225] K. Szymiczek, Generalized Hilbert fields, J. Reine Angew. Math. **329** (1981), 58–65.
[226] K. Szymiczek, Matching Witts locally and globally, Math. Slovaca **41** (1991), 315–330.
[227] K. Szymiczek, Witt equivalence of global fields, Commun. Algebra **19** (1991), 1125–1149.
[228] K. Szymiczek, Algebra, Logic and Applications, vol. **7**: *Bilinear Algebra. An introduction to the algebraic theory of quadratic forms*, Gordon and Breach Science Publishers, Amsterdam, 1997.
[229] K. Szymiczek, Hilbert-symbol equivalence of number fields, Tatra Mount. Math. Publ. **11** (1997), 7–16.
[230] K. Szymiczek, A characterization of tame Hilbert-symbol equivalence, Acta Math. Inf. Univ. Ostrav. **6** (1998), 191–201.
[231] K. Szymiczek, Tame equivalence and wild sets, Semigroup Forum **60** (2000), 260–270.
[232] K. Szymiczek, Ten problems on quadratic forms, Acta Math. Inf. Univ. Ostrav. **10** (2002), 133–143.
[233] C.C. Tsen, Zur Stufentheorie der quasialgebraisch-Abgeschlossenheit kommutativer Körper, J. Chin. Math. Soc. **1** (1936), 81–92.
[234] A. Vishik, On dimensions of quadratic forms, Dokl. Akad. Nauk **373** (2000), no. 4, 445–447.
[235] A. Vishik, *Fields of u-invariant* $2^r + 1$. Preprint No. 229, October 2006. Linear Algebraic Groups and Related Structures preprint server (http://www.math.uni-bielefeld.de/LAG/).
[236] V. Voevodsky, *On* $2-$*torsion in motovic cohomology.* Preprint 2001. Available at K-theory Preprint Archives http://www.math.uiuc.edu/K-theory/0502/
[237] A.R. Wadsworth, Similarity of quadratic forms and isomorphism of their function fields, Trans. Am. Math. Soc. **208** (1975), 352–358.
[238] A.R. Wadsworth, Merkurjev's elementary proof of Merkurjev's theorem, Contemp. Math. **55** (1986), 741–776.
[239] W.C. Waterhouse, Scaled trace forms over number fields, Arch. Math. **47** (1986), 229–231.
[240] E. Witt, Theorie der quadratischen Formen in beliebigen Körpern, J. Reine Angew. Math. **176** (1937), 31–44.
[241] J.L. Yucas, Quadratic forms and radicals of fields, Acta Arith. **39** (1981), 313–322.

Uncited References

[1] T.C. Craven, Orderings, valuations, and hermitian forms over *-fields, Proc. Symp. Pure Math. **58**, Part II (1995), 149–160.
[2] R. Elman, A. Prestel, Reduced stability of the Witt ring of a field and its Pythagorean closure, Am. J. Math. **106** (1984), 1237–1260.
[3] R.W. Fitzgerald, J. Yucas, Combinatorial techniques and abstract Witt rings, I. J. Algebra **114** (1988), 40–52.
[4] R.W. Fitzgerald, J. Yucas, Combinatorial techniques and abstract Witt rings, II. Rocky Mountain J. Math. **19** (1989), 687–708.
[5] U. Jannsen, R. Sujatha, Levels of function fields of surfaces over number fields, J. Algebra **251** (2002), no. 1, 350–357.
[6] I. Kaplansky, J. Shaker, Abstract quadratic forms, Can. J. Math. **21** (1969), 1218–1233.
[7] M.-A. Knus, T.Y. Lam, D.B. Shapiro, J.-P. Tignol, Discriminants of involutions on biquaternion algebras, Proc. Symp. Pure Math. **58**, Part II (1995), 279–303.
[8] T.Y. Lam, *The Algebraic Theory of Quadratic Forms*, W.A. Benjamin, Addison-Wesley, Reading, Mass., 1973. Second printing with revisions, 1980.

[9] T.Y. Lam, D.B. Leep, J.-P. Tignol, Biquaternion algebras and quartic extensions, Math. Ann. **62** (1991), 272–285.
[10] F. Lorenz, Lecture Notes in Mathematics, vol. 130: Quadratische Formen über Körpern, Springer-Verlag, Berlin-Heidelberg-New York, 1970, 2007.
[11] P. Mammone, D.B. Shapiro, The Albert quadratic form for an algebra of degree four, Proc. Am. Math. Soc. **105** (1989), 525–530.
[12] A. Prestel, Lecture Notes in Mathematics, vol. **1093**: *Lectures on Formally Real Fields*, Springer-Verlag, Berlin Heidelberg New York Tokyo, 1984.
[13] D.B. Shapiro, Products of sums of squares, Expos. Math. **2** (1984), 235–261.

Crossed Complexes and Higher Homotopy Groupoids as Noncommutative Tools for Higher Dimensional Local-to-Global Problems*

Ronald Brown

School of Computer Science, Bangor University, Bangor, Gwynedd LL57 1UT, U.K.
E-mail: r.brown@bangor.ac.uk

Contents

*This is a revised version (2008) of a paper published in Fields Institute Communications 43 (2004) 101–130, which was an extended account of a lecture given at the meeting on "Categorical Structures for Descent, Galois Theory, Hopf algebras and semiabelian categories," Fields Institute, September 23–28, 2002. The author is grateful for support from the Fields Institute and a Leverhulme Emeritus Research Fellowship, 2002–2004 and to M. Hazewinkel for helpful comments on a draft.

HANDBOOK OF ALGEBRA, VOL. 6
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(08)00203-9

Introduction

An aim is to give a survey and explain the origins of results obtained by R. Brown and P.J. Higgins and others over the years 1974–2008, and to point to applications and related areas. These results yield an account of some basic algebraic topology on the border between homology and homotopy; it differs from the standard account through the use of *crossed complexes*, rather than chain complexes, as a fundamental notion. In this way, one obtains comparatively quickly[1] not only classical results such as the Brouwer degree and the relative Hurewicz theorem but also noncommutative results on second relative homotopy groups, as well as higher dimensional results involving the fundamental group through its actions and presentations. A basic tool is the fundamental crossed complex ΠX_* of the filtered space X_*, which in the case X_0 is a singleton is fairly classical; when applied to the skeletal filtration of a CW-complex X, Π gives a more powerful version of the usual cellular chains of the universal cover of X because it contains non-Abelian information in dimensions 1 and 2 and has good realization properties. It also gives a replacement for singular chains by taking X to be the geometric realization of a singular complex of a space.

One of the major results is a homotopy classification theorem (4.1.9), which generalizes a classical theorem of Eilenberg–Mac Lane, though this does require results on geometric realizations of cubical sets.

A replacement for the excision theorem in homology is obtained by using cubical methods to prove a Higher Homotopy van Kampen Theorem (HHvKT)[2] for the fundamental crossed complex functor Π on filtered spaces. This theorem is a higher dimensional version of the van Kampen Theorem (vKT) on the fundamental group of a space with base point, [1][3], which is a classical example of a

noncommutative local-to-global theorem,

and is the initial motivation for the work described here. The vKT determines completely the fundamental group $\pi_1(X, x)$ of a space X with base point, which is the union of open sets U, V whose intersection is path-connected and contains the base point x; the "local information" is on the morphisms of fundamental groups induced by the inclusions $U \cap V \to U, U \cap V \to V$. The importance of this result reflects the importance of the fundamental group in algebraic topology, algebraic geometry, complex analysis, and many other subjects. Indeed, the origin of the fundamental group was in Poincaré's work on monodromy for complex variable theory.

Essential to this use of crossed complexes, particularly for conjecturing and proving local-to-global theorems, is a construction of a *cubical higher homotopy groupoid*, with properties described by *an algebra of cubes*. There are applications to local-to-global problems in homotopy theory, which are more powerful than available by

[1] This comparison is based on the fact that the methods do not require singular homology or simplicial approximation.

[2] We originally called this a generalized van Kampen Theorem, but this new term was suggested in 2007 by Jim Stasheff.

[3] An earlier version for simplicial complexes is due to Seifert.

purely classical tools while shedding light on those tools. It is hoped that this account will increase the interest in the possibility of wider applications of these methods and results since homotopical methods play a key role in many areas.

Background in higher homotopy groups

Topologists in the early part of the 20th century were well aware that

- the noncommutativity of the fundamental group was useful in geometric applications;
- for path-connected X, there was an isomorphism

$$H_1(X) \cong \pi_1(X, x)^{\mathrm{ab}};$$

- the Abelian homology groups $H_n(X)$ existed for all $n \geqslant 0$.

Consequently, there was a desire to generalize the noncommutative fundamental group to all dimensions.

In 1932, Čech submitted a paper on higher homotopy groups $\pi_n(X, x)$ to the ICM at Zurich, but it was quickly proved that these groups were Abelian for $n \geqslant 2$, and on these grounds, Čech was persuaded to withdraw his paper so that only a small paragraph appeared in the Proceedings [2]. We now see the reason for this commutativity as the result (Eckmann–Hilton) that a group internal to the category of groups is just an Abelian group. Thus, since 1932, the vision of a noncommutative higher dimensional version of the fundamental group has been generally considered to be a mirage. Before we go back to the vKT, we explain in the next section how nevertheless work on crossed modules did introduce noncommutative structures relevant to topology in dimension 2.

Work of Hurewicz, [3], led to a strong development of higher homotopy groups. The fundamental group still came into the picture with its action on the higher homotopy groups, which I once heard J.H.C. Whitehead remark (1957) was especially fascinating for the early workers in homotopy theory. Much of Whitehead's work was intended to extend to higher dimensions the methods of combinatorial group theory of the 1930s, hence the title of his papers were as follows: "Combinatorial homotopy, I, II" [4, 5]. The first of these two papers has been very influential and is part of the basic structure of algebraic topology. It is the development of work of the second paper, which we explain here.

The paper by Whitehead on "Simple homotopy types" [6], which deals with higher dimensional analogs of Tietze transformations, has a final section using crossed complexes. We refer to this again later in Section 15.

It is hoped also that this survey will be useful background to work on the van Kampen Theorem for diagrams of spaces in [7], which uses a form of higher homotopy groupoid that is in an important sense much more powerful than that given here since it encompasses n-adic information; however, current expositions are still restricted to the reduced (one base point) case, the proof uses advanced tools of algebraic topology, and the result was suggested by the work exposed here.

1. Crossed modules

In the years 1941–1950, Whitehead developed work on crossed modules to represent the structure of the boundary map of the relative homotopy group

$$\pi_2(X, A, x) \to \pi_1(A, x), \tag{1}$$

in which both groups can be noncommutative. Here is the definition he found.

A *crossed module* is a morphism of groups $\mu : M \to P$ together with an action $(m, p) \mapsto m^p$ of the group P on the group M satisfying the following two axioms:

CM1) $\mu(m^p) = p^{-1}(\mu m)p$
CM2) $n^{-1}mn = m^{\mu n}$

for all $m, n \in M$, $p \in P$.

Standard algebraic examples of crossed modules are as follows:

(i) an inclusion of a normal subgroup, with action given by conjugation;
(ii) the inner automorphism map $\chi : M \to \mathrm{Aut}\, M$, in which χm is the automorphism $n \mapsto m^{-1}nm$;
(iii) the zero map $M \to P$ where M is a P-module;
(iv) an epimorphism $M \to P$ with kernel contained in the center of M.

Simple consequences of the axioms for a crossed module $\mu : M \to P$ are as follows:

1.1. $\mathrm{Im}\,\mu$ *is normal in* P.

1.2. $\mathrm{Ker}\,\mu$ *is central in* M *and is acted on trivially by* $\mathrm{Im}\,\mu$ *so that* $\mathrm{Ker}\,\mu$ *inherits an action of* $M/\mathrm{Im}\,\mu$.

Another important algebraic construction is the *free crossed* P*-module*

$$\partial : C(\omega) \to P$$

determined by a function $\omega : R \to P$, where P is a group and R is a set. The group $C(\omega)$ is generated by elements $(r, p) \in R \times P$ with the relations

$$(r, p)^{-1}(s, q)^{-1}(r, p)(s, qp^{-1}(\omega r)p);$$

the action is given by $(r, p)^q = (r, pq)$; and the boundary morphism is given by $\partial(r, p) = p^{-1}(\omega r)p$, for all $(r, p), (s, q) \in R \times P$.

A major result of Whitehead was

THEOREM W. [5]. *If the space* $X = A \cup \{e^2_r\}_{r\in R}$ *is obtained from* A *by attaching 2-cells by maps* $f_r : (S^1, 1) \to (A, x)$*, then the crossed module of* (1) *is isomorphic to the free crossed* $\pi_1(A, x)$*-module on the classes of the attaching maps of the 2-cells.*

Whitehead's proof, which stretched over three papers, 1941–1949, used transversality and knot theory—an exposition is given in [8]. Mac Lane and Whitehead [9]

used this result as part of their proof that crossed modules capture all homotopy 2-types (they used the term "3-types").

The title of the paper in which the first intimation of Theorem W appeared was "On adding relations to homotopy groups" [10]. This indicates a search for higher dimensional vKTs.

The concept of free crossed module gives a noncommutative context for *chains of syzygies*. The latter idea, in the case of modules over polynomial rings, is one of the origins of homological algebra through the notion of *free resolution*. Here is how similar ideas can be applied to groups. Pioneering work here, independent of Whitehead, was by Peiffer [11] and Reidemeister [12]. See [13] for an exposition of these ideas.

Suppose $\mathcal{P} = \langle X \mid \omega \rangle$ is a presentation of a group G so that X is a set of generators of G and $\omega : R \to F(X)$ is a function, whose image is called the set of relators of the presentation. Then we have an exact sequence

$$1 \xrightarrow{i} N(\omega R) \xrightarrow{\phi} F(X) \longrightarrow G \longrightarrow 1,$$

where $N(\omega R)$ is the normal closure in $F(X)$ of the set ωR of relators. The above work of Reidemeister, Peiffer, and Whitehead showed that to obtain the next level of syzygies, one should consider the free crossed $F(X)$-module $\partial : C(\omega) \to F(X)$ since this takes into account the operations of $F(X)$ on its normal subgroup $N(\omega R)$. Elements of $C(\omega)$ are a kind of "formal consequences of the relators" so that the relation between the elements of $C(\omega)$ and those of $N(\omega R)$ is analogous to the relation between the elements of $F(X)$ and those of G. It follows from the rules for a crossed module that the kernel of ∂ is a G-module, called the module of *identities among relations*, and sometimes written as $\pi(\mathcal{P})$; there is considerable work on computing it [13–17, 126]. By splicing to ∂ a free G-module resolution of $\pi(\mathcal{P})$, one obtains what is called a *free crossed resolution* of the group G. We explain later (Proposition 15.3) why these resolutions have better realization properties than the usual resolutions by chain complexes of G-modules. They are relevant to the Schreier extension theory [18].

This notion of using crossed modules as the first stage of syzygies in fact represents a wider tradition in homological algebra, in the work of Frölich and Lue [19, 20].

Crossed modules also occurred in other contexts, notably in representing elements of the cohomology group $H^3(G, M)$ of a group G with coefficients in a G-module M [21] and as coefficients in Dedecker's theory of non-Abelian cohomology [22]. The notion of free crossed resolution has been exploited by Huebschmann [23–25] to represent cohomology classes in $H^n(G, M)$ of a group G with coefficients in a G-module M and also to calculate with these.

The HHvKT can make it easier to compute a crossed module arising from some topological situation, such as an induced crossed module [26, 27], or a coproduct crossed module [28], than the cohomology class in $H^3(G, M)$ the crossed module represents. To obtain information on such a cohomology element, it is useful to work with a small free crossed resolution of G, and this is one motivation for developing methods for calculating such resolutions. However, it is not so clear what a *calculation* of such a cohomology element would amount to although it is interesting to know

whether the element is nonzero or what its order is. Thus, the use of algebraic models of cohomology classes may yield easier computations than the use of cocycles, and this somewhat inverts traditional approaches.

Since crossed modules are algebraic objects generalizing groups, it is natural to consider the problem of explicit calculations by extending techniques of computational group theory. Substantial work on this has been done by C.D. Wensley using the program GAP [29, 30].

2. The fundamental groupoid on a set of base points

A change in prospects for higher order noncommutative invariants was suggested by Higgins' paper [31] and leading to work of the writer published in 1967 [32]. This showed that the vKT could be formulated for the *fundamental groupoid* $\pi_1(X, X_0)$ *on a set* X_0 *of base points*, thus enabling computations in the nonconnected case, including those in Van Kampen's original paper [1]. This successful use of groupoids in dimension 1 suggested the question of the use of groupoids in higher homotopy theory and, in particular, the question of the existence of *higher homotopy groupoids*.

To see how this research programme could progress, it is useful to consider the statement and special features of this generalized vKT for the fundamental groupoid. If X_0 is a set, and X is a space, then $\pi_1(X, X_0)$ denotes the fundamental groupoid on the set $X \cap X_0$ of base points. This allows the set X_0 to be chosen in a way appropriate to the geometry. For example, if the circle S^1 is written as the union of two semicircles $E_+ \cup E_-$, then the intersection $\{-1, 1\}$ of the semicircles is not connected, so it is not clear where to take the base point. Instead, one takes $X_0 = \{-1, 1\}$, and so has two base points. This flexibility is very important in computations, and this example of S^1 was a motivating example for this development. As another example, you might like to consider the difference between the quotients of the actions of $\mathbb{Z}_2$ on the group $\pi_1(S^1, 1)$ and on the groupoid $\pi_1(S^1, \{-1, 1\})$, where the action is induced by complex conjugation on S^1. Relevant work on *orbit groupoids* has been developed by Higgins and Taylor [33, 34], (under useful conditions, the fundamental groupoid of the orbit space is the orbit groupoid of the fundamental groupoid [35, 11.2.3]).

Consideration of a set of base points leads to the following theorem (Theorem 2.1):

THEOREM 2.1. [32] *Let the space X be the union of open sets U, V with intersection W, and let X_0 be a subset of X meeting each path component of U, V, W. Then,*
(C) (connectivity) X_0 *meets each path component of X and*
(I) (isomorphism) *the diagram of groupoid morphisms induced by inclusions*

$$\begin{array}{ccc} \pi_1(W, X_0) & \xrightarrow{i} & \pi_1(U, X_0) \\ {\scriptstyle j}\downarrow & & \downarrow{\scriptstyle j'} \\ \pi_1(V, X_0) & \xrightarrow[i']{} & \pi_1(X, X_0) \end{array} \tag{2}$$

is a pushout of groupoids.

From this theorem, one can compute a particular fundamental group $\pi_1(X, x_0)$ using combinatorial information on the graph of intersections of path components of U, V, W, but for this, it is useful to develop the algebra of groupoids. Notice two special features of this result, which are as follows.

(i) The computation of the invariant you may want, a fundamental group, is obtained from the computation of a larger structure, and so part of the work is to give methods for computing the smaller structure from the larger one. This usually involves noncanonical choices, for example, that of a maximal tree in a connected graph. The work on applying groupoids to groups gives many examples of this [31, 32, 36, 37].

(ii) The fact that the computation can be done is surprising in two ways: (a) the fundamental group is computed *precisely*, even though the information for it uses input in two dimensions, namely 0 and 1. This is contrary to the experience in homological algebra and algebraic topology, where the interaction of several dimensions involves exact sequences or spectral sequences, which give information only up to extension, and (b) the result is a noncommutative invariant, which is usually even more difficult to compute precisely.

The reason for the success seems to be that the fundamental groupoid $\pi_1(X, X_0)$ contains information in dimensions 0 and 1, and so can adequately reflect the geometry of the intersections of the path components of U, V, W and of the morphisms induced by the inclusions of W in U and V.

This suggested the question of whether these methods could be extended successfully to higher dimensions.

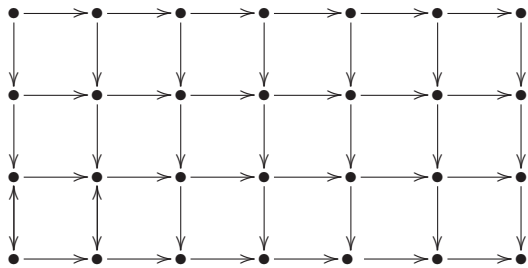
Part of the initial evidence for this quest was the intuitions in the proof of this groupoid vKT, which seemed to use three main ideas to verify the universal property of a pushout for diagram (2). So, suppose given morphisms of groupoids f_U, f_V from $\pi_1(U, X_0)$, $\pi_1(V, X_0)$ to a groupoid G, satisfying $f_U i = f_V j$. We have to construct a morphism $f : \pi_1(X, X_0) \to G$ such that $fi' = f_U$, $fj' = f_V$ and prove f is unique. We concentrate on the construction.

- One needs a "deformation," or "filling," argument: given a path $a : (I, \dot{I}) \to (X, X_0)$, one can write $a = a_1 + \cdots + a_n$ where each a_i maps into U or V, but a_i will not necessarily have end points in X_0. So one has to deform each a_i to a_i' in U, V, or W, using the connectivity condition so that each a_i' has end points in X_0, and $a' = a_1' + \cdots + a_n'$ is well defined. Then, one can construct using f_U or f_V an image of each a_i' in G and hence of the composite, called $F(a) \in G$, of these images. Note that we subdivide in X and then put together again in G (this uses the condition $f_U i = f_V j$ to prove that the elements of G are composable), and this part can be summarized as follows:
- Groupoids provided a convenient *algebraic inverse to subdivision*. Note that the usual exposition in terms only of the fundamental group uses loops, that is, paths which start and finish at the same point. An appropriate analogy is that if one goes on a train journey from Bangor and back to Bangor, one usually wants to

stop off at intermediate stations; this breaking and continuing a journey is better described in terms of groupoids rather than groups.

Next, one has to prove that $F(a)$ depends only on the class of a in the fundamental groupoid. This involves a homotopy relative end points $h : a \simeq b$, considered as a map $I^2 \to X$; subdivide h as $h = [h_{ij}]$ so that each h_{ij} maps into U, V, or W; deform h to $h' = [h'_{ij}]$ (keeping in U, V, W) so that each h'_{ij} maps the vertices to X_0 and so determines a commutative square[4] in one of $\pi_1(Q, X_0)$ for $Q = U, V, W$. Move these commutative squares over to G using f_U, f_V and recompose them (this is possible again because of the condition $f_U i = f_V j$), noting that

- in a groupoid, *any composition of commutative squares is commutative*. Here, a "big" composition of commutative squares is represented by a diagram such as



(3)

and one checks that if each individual square is commutative, so also is the boundary square (later called a 2-shell) of the compositions of the boundary edges.

---

[4] We need the notion of *commutative square* in a category C. This is a quadruple $\begin{pmatrix} & c & \\ a & & d \\ & b & \end{pmatrix}$ of arrows in C, called "edges" of the square such that $ab = cd$, that is, these compositions are defined and agreed. The commutative squares in C form a *double category* $\square\mathsf{C}$ in that they compose "vertically"

$$\begin{pmatrix} & c & \\ a & & d \\ & b & \end{pmatrix} \circ_1 \begin{pmatrix} & b & \\ a' & & d' \\ & e & \end{pmatrix} = \begin{pmatrix} & c & \\ aa' & & dd' \\ & e & \end{pmatrix}$$

and "horizontally"

$$\begin{pmatrix} & c & \\ a & & d \\ & b & \end{pmatrix} \circ_2 \begin{pmatrix} & c' & \\ d & & f \\ & b' & \end{pmatrix} = \begin{pmatrix} & cc' & \\ a & & f \\ & bb' & \end{pmatrix}$$

This notion of $\square\mathsf{C}$ was defined by C. Ehresmann in papers and in [38]. Note the obvious geometric conditions for these compositions to be defined. Similarly, one has geometric conditions for a rectangular array (c_{ij}), $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$, of commutative squares to have a well-defined composition, and then their "multiple composition," written as $[c_{ij}]$, is also a commutative square, whose edges are compositions of the "edges" along the outside boundary of the array. It is easy to give formal definitions of all this.

Two opposite sides of the composite commutative square in G so obtained are identities because h was a homotopy relative to end points, and the other two sides are $F(a),\ F(b)$. This proves that $F(a) = F(b)$ in G.

Thus, the argument can be summarized as follows: a path or homotopy is divided into small pieces, then deformed so that these pieces can be packaged and moved over to G, where they are reassembled. There seems to be an analogy with the processing of an e-mail.

Notable applications of the groupoid theorem were: (i) to give a proof of a formula in van Kampen's paper of the fundamental group of a space, which is the union of two connected spaces with non connected intersection, see [35, 8.4.9]; and (ii) to show the topological utility of the construction by Higgins [36] of the groupoid $f_*(G)$ over Y_0 induced from a groupoid G over X_0 by a function $f : X_0 \to Y_0$. (Accounts of these with the notation $U_f(G)$ rather than $f_*(G)$ are given in [35, 36].) This latter construction is regarded as a "change of base," and analogs in higher dimensions yielded generalizations of the relative Hurewicz theorem and of the Theorem W, using induced modules and crossed modules.

There is another approach to the vKT, which goes via the theory of covering spaces and the equivalence between covering spaces of a reasonable space X and functors $\pi_1(X) \to \mathsf{Set}$ [35]. See for example [39] for an exposition of the relation with traditional Galois theory and [40] for a modern account in which *Galois groupoids* make an essential appearance. The paper [41] gives a general formulation of conditions for the theorem to hold in the case $X_0 = X$ in terms of the map $U \sqcup V \to X$ being an "effective global descent morphism" (the theorem is given in the generality of lextensive categories). This work has been developed for toposes [42]. Analogous interpretations for higher dimensional vKTs are not known.

The justification of the breaking of a paradigm in changing from groups to groupoids is several fold: the elegance and power of the results; the increased linking with other uses of groupoids [43]; and the opening out of new possibilities in higher dimensions, which allowed for new results and calculations in homotopy theory and suggested new algebraic constructions. The important and extensive work of Charles Ehresmann in using groupoids in geometric situations (bundles, foliations, germes, ...) should also be stated (see his collected works of which [44] is volume 1 and a survey [45]).

3. The search for higher homotopy groupoids

Contemplation of the proof of the groupoid vKT in the last section suggested that a higher dimensional version should exist, though this version amounted to an idea of a proof in search of a theorem. Further evidence was the proof by J.F. Adams of the cellular approximation theorem given in [35]. This type of subdivision argument failed to give algebraic information apparently because of a lack of an appropriate higher homotopy groupoid, that is a gadget to capture what might be the underlying "algebra of cubes." In the end, the results exactly encapsulated this intuition.

One intuition was that in groupoids, we are dealing with a partial algebraic structure,[5] in which composition is defined for two arrows if and only if the source of one arrow is the target of the other. This seems to generalize easily to directed squares, in which two such are composable horizontally if and only if the left-hand side of one is the right-hand side of the other (and similarly vertically).

However, the formulation of a theorem in higher dimensions required specification of the *topological data*, the *algebraic data*, and of a functor

$$\Pi : (\text{topological data}) \to (\text{algebraic data}),$$

which would allow the expression of these ideas for the proof.

Experiments were made in the years 1967–1973 to define some functor Π from spaces to some kind of double groupoid, using compositions of squares in two directions, but these proved abortive. However, considerable progress was made in work with Chris Spencer in 1971–1973 on investigating the algebra of double groupoids [47] and on showing a relation to crossed modules. Further evidence was provided when it was found, [48], that group objects in the category of groupoids (or groupoid objects in the category of groups, either of which are often now called "2-groups") are equivalent to crossed modules and, in particular, are not necessarily commutative objects. It turned out that this result was known to the Grothendieck school in the 1960s, but not published.

We review next a notion of double category, which is not the most general but is appropriate in many cases. It was called an *edge symmetric double category* in [49].

In the first place, a double category, K, consists of a triple of category structures

$$(K_2, K_1, \partial_1^-, \partial_1^+, \circ_1, \varepsilon_1), \quad (K_2, K_1, \partial_2^-, \partial_2^+, \circ_2, \varepsilon_2),$$
$$(K_1, K_0, \partial^-, \partial^+, \circ, \varepsilon),$$

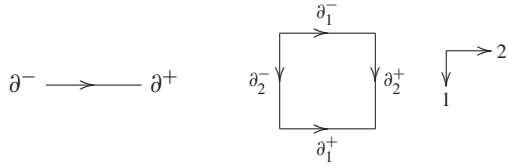
as partly shown in the diagram

$$\begin{array}{ccc} K_2 & \overset{\partial_2^-}{\underset{\partial_2^+}{\rightrightarrows}} & K_1 \\ {\scriptstyle \partial_1^+}\downarrow\downarrow{\scriptstyle \partial_1^-} & & {\scriptstyle \partial^+}\downarrow\downarrow{\scriptstyle \partial^-} \\ K_1 & \underset{\partial^+}{\overset{\partial^-}{\rightrightarrows}} & K_0 \end{array} \tag{4}$$

The elements of K_0, K_1, and K_2 will be called, respectively, *points or objects, edges, and squares*. The maps $\partial^{\pm}$, $\partial_i^{\pm}$, $i = 1, 2$, will be called *face maps*, the maps ε_i :

[5] The study of partial algebraic operations was initiated in [46]. We can now suggest a reasonable definition of "higher dimensional algebra" as dealing with families of algebraic operations, whose domains of definitions are given by geometric conditions.

$K_1 \longrightarrow K_2$, $i = 1, 2$, resp. $\varepsilon : K_0 \longrightarrow K_1$ will be called *degeneracies*. The boundaries of an edge and of a square are given by the diagrams



(5)

The partial compositions, $\circ_1$, resp. $\circ_2$, are referred to as *vertical*, resp. *horizontal, composition* of squares, are defined under the obvious geometric conditions, and have the obvious boundaries. The axioms for a double category also include the usual relations of a 2-cubical set (e.g. $\partial^-\partial_2^+ = \partial^+\partial_1^-$) and the *interchange law*. We use matrix notation for compositions as

$$\begin{bmatrix} a \\ c \end{bmatrix} = a \circ_1 c, \quad \begin{bmatrix} a & b \end{bmatrix} = a \circ_2 b,$$

and the crucial interchange law[6] for these two compositions allows one to use matrix notation

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} a & b \end{bmatrix} \\ \begin{bmatrix} c & d \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} a \\ c \end{bmatrix} & \begin{bmatrix} b \\ d \end{bmatrix} \end{bmatrix}$$

for double composites of squares, whenever each row composite and each column composite is defined. We also allow the multiple composition $[a_{ij}]$ of an array (a_{ij}), whenever for all appropriate i, j we have $\partial_1^+ a_{ij} = \partial_1^- a_{i+1,j}$, $\partial_2^+ a_{ij} = \partial_2^- a_{i,j+1}$. A clear advantage of double categories and cubical methods is this easy expression of multiple compositions, which allows for *algebraic inverse to subdivision*, and so applications to local-to-global problems.

The identities with respect to $\circ_1$ (*vertical identities*) are given by ε_1 and will be denoted by $|\ |$. Similarly, we have *horizontal identities* denoted by $\overline{\ \ }$. Elements of the form $\varepsilon_1\varepsilon(a) = \varepsilon_2\varepsilon(a)$ for $a \in K_0$ are called *double degeneracies* and will be denoted by $\square$.

A *morphism of double categories* $f : \mathsf{K} \to \mathsf{L}$ consists of a triple of maps $f_i : K_i \to L_i$, $(i = 0, 1, 2)$, respecting the cubical structure, compositions, and identities.

Whereas it is easy to describe a commutative square of morphisms in a category, it is not possible with this amount of structure to describe a commutative cube of squares in a double category. We first define a cube, or 3-shell, that is, without any condition of commutativity, in a double category.

[6] The interchange implies that a double monoid is simply an Abelian monoid, so partial algebraic operations are essential for the higher dimensional work.

DEFINITION 3.1. Let K be a double category. A *cube* (*3-shell*) in K,

$$\alpha = (\alpha_1^-, \alpha_1^+, \alpha_2^-, \alpha_2^+, \alpha_3^-, \alpha_3^+),$$

consists of squares $\alpha_i^{\pm} \in K_2 \quad (i = 1, 2, 3)$ such that

$$\partial_i^{\sigma}(\alpha_j^{\tau}) = \partial_{j-1}^{\tau}(\alpha_i^{\sigma})$$

for $\sigma, \tau = \pm$ and $1 \leqslant i < j \leqslant 3$. ■

It is also convenient to have the corresponding notion of square, or 2-shell, of arrows in a category. The obvious compositions also makes these into a double category.

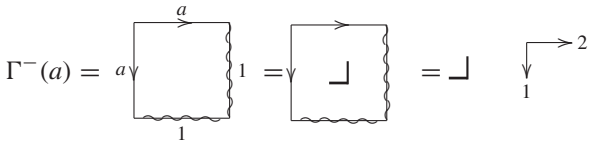
It is not hard to define three compositions of cubes in a double category so that these cubes form a triple category: this is done in [50] or more generally in Section 5 of [51]. A key point is that to define the notion of a *commutative cube*, we need extra structure on a double category. Thus, this step up a dimension is nontrivial, as was first observed in the groupoid case in [52]. The problem is that a cube has six faces, which easily divide into three even and three odd faces. So we cannot say as we might like that "the cube is commutative if the composition of the even faces equals the composition of the odd faces," since there are no such valid compositions.

The intuitive reason for the need of a new basic structure in that in a 2-dimensional situation, we also need to use the possibility of "turning an edge clockwise or anti-clockwise." The structure to do this is as follows.

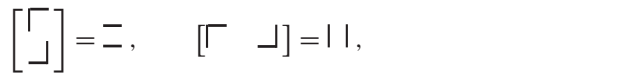
A *connection pair* on a double category K is given by a pair of maps

$$\Gamma^-, \Gamma^+ : K_1 \longrightarrow K_2,$$

whose edges are given by the following diagrams for $a \in K_1$:

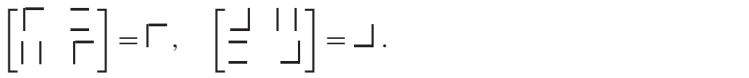


This "hieroglyphic" notation, which was introduced in [53], is useful for expressing the laws these connections satisfy. The first is a pair of cancellation laws which read



which can be understood as "if you turn right and then left, you face the same way," and similarly the other way round. They were introduced in [54]. Note that in this matrix notation, we assume that the edges of the connections are such that the composition is defined.

Two other laws relate the connections to the compositions and read



These can be interpreted as "turning left (or right) with your arm outstretched is the same as turning left (or right)." The term "connections" and the name "transport laws" were because these laws were suggested by the laws for path connections in differential geometry, as explained in [47]. It was proved in [49] that a connection pair on a double category K is equivalent to a "thin structure," namely a morphism of double categories $\Theta : \square K_1 \to K$, which is the identity on the edges. The proof requires some "2-dimensional rewriting" using the connections.

We can now explain what is a "commutative cube" in a double category K with connection pair.

DEFINITION 3.2. Suppose given, in a double category with connections K, a cube (3-shell)

$$\alpha = (\alpha_1^-, \alpha_1^+, \alpha_2^-, \alpha_2^+, \alpha_3^-, \alpha_3^+).$$

We define the *composition of the odd faces* of α to be

$$\partial^{\text{odd}}\alpha = \begin{bmatrix} \ulcorner & \alpha_1^- & \lrcorner \\ \alpha_3^- & \alpha_2^+ & = \\ = & = & = \end{bmatrix} \tag{6}$$

and the *composition of the even faces* of α to be

$$\partial^{\text{even}}\alpha = \begin{bmatrix} = & \alpha_2^- & \alpha_3^+ \\ = & \alpha_1^+ & \lrcorner \\ \ulcorner & = & = \end{bmatrix}. \tag{7}$$

We define α to be *commutative* if it satisfies the homotopy commutativity lemma (HCL), that is,

$$\partial^{\text{odd}}\alpha = \partial^{\text{even}}\alpha. \tag{HCL}$$

This definition can be regarded as a cubical, categorical (rather than groupoid) form of the Homotopy Addition Lemma (HAL) in dimension 3.

You should draw a 3-shell, label all the edges with letters, and see that this equation makes sense in that the 2-shells of each side of equation (HCL) coincide. Notice, however, that these 2-shells do not have coincident partitions along the edges: that is, the edges of this 2-shell in direction 1 are formed from different compositions of the type $1 \circ a$ and $a \circ 1$. This definition is discussed in more detail in [50], is related to other equivalent definitions, and it is proved that compositions of commutative cubes

in the three possible directions are also commutative. These results are extended to all dimensions in [55]; this requires the full structure indicated in Section 9 and also the notion of *thin element* indicated in Section 12.

The initial discovery of connections arose in [47] from relating crossed modules to double groupoids. The first example of a double groupoid was the double groupoid $\square G$ of commutative squares in a group G. The first step in generalizing this construction was to consider quadruples $\left(\begin{smallmatrix} & c & \\ a & & d \\ & b & \end{smallmatrix}\right)$ of elements of G such that $abn = cd$ for some element n of a subgroup N of G. Experiments quickly showed that for the two compositions of such quadruples to be valid, it was necessary and sufficient that N be normal in G. But in this case, the element n is determined by the boundary, or 2-shell, a, b, c, d. In homotopy theory, we require something more general. So we consider a morphism $\mu : N \to G$ of groups and consider quintuples $\left(n : \begin{smallmatrix} & c & \\ a & & d \\ & b & \end{smallmatrix}\right)$ such that $ab\mu(n) = cd$. It then turns out that we get a double groupoid if and only if $\mu : N \to G$ is a crossed module. The next question is which double groupoids arise in this way? It turns out that we need exactly double groupoids with connection pairs, though in this groupoid case we can deduce Γ^- from Γ^+ using inverses in each dimension. This gives the main result of [47], the equivalence between the category of crossed modules and that of double groupoids with connections and one vertex.

These connections were also used in [52] to define a "commutative cube" in a double groupoid with connections using the equation

$$c_1 = \begin{bmatrix} \ulcorner & a_0^{-1} & \urcorner \\ -b_0 & c_0 & b_1 \\ \llcorner & a_1 & \lrcorner \end{bmatrix}$$

representing one face of a cube in terms of the other five and where the other connections $\llcorner, \urcorner$ are obtained from $\lrcorner, \ulcorner$ by using the two inverses in dimension 2. As you might imagine, there are problems in finding a formula in still higher dimensions. In the groupoid case, this is handled by a homotopy addition lemma and thin elements, [51], but in the category case, a formula for just a commutative four-cube is complicated (see [56]).

The blockage of defining a functor Π to double groupoids was resolved after nine years in 1974 in discussions with Higgins, by considering the Whitehead Theorem W. This showed that a 2-dimensional universal property was available in homotopy theory, which was encouraging; it also suggested that a theory to be any good should recover Theorem W. But this theorem was about *relative* homotopy groups. This suggested studying a relative situation $X_* : X_0 \subseteq X_1 \subseteq X$. On looking for the simplest way to get a homotopy functor from this situation using squares, the obvious answer came up: consider maps $(I^2, \partial I^2, \partial\partial I^2) \to (X, X_1, X_0)$, that is, maps of the square, which take the edges into X_1 and the vertices into X_0, and then take homotopy classes of such maps relative to the vertices of I^2 to form a set $\rho_2 X_*$. Of course, this set will not inherit a group structure, but the surprise is that it does inherit the structure of double groupoid with connections—the proof is not entirely trivial, and is given in [52] and the expository article [57]. In the case X_0 is a singleton, the equivalence

of such double groupoids to crossed modules takes ρX_* to the usual second-relative homotopy crossed module.

Thus, a search for a *higher homotopy groupoid* was realized in dimension 2. Connes suggests in [58] that it has been fashionable for mathematicians to disparage groupoids and it might be that a lack of attention to this notion was one reason why such a construction had not been found earlier than 40 years after Hurewicz's papers.

Finding a good homotopy double groupoid led rather quickly, in view of the previous experience, to a substantial account of a 2-dimensional HHvKT [52]. This recovers Theorem W and also leads to new calculations in 2-dimensional homotopy theory, and in fact to some new calculations of 2-types. For a recent summary of some results and some new ones, see the paper published in the J. Symbolic Computation [30]—publication in this journal illustrates that we are interested in using general methods to obtain specific calculations and ones to which there seems no other route.

Once the 2-dimensional case had been completed in 1975, it was easy to conjecture the form of general results for dimensions > 2. These were proved by 1979 and announcements were made in [59] with full details in [51, 60]. However, these results needed a number of new ideas, even just to construct the higher dimensional compositions, and the proof of the HHvKT was quite hard and intricate. Further, for applications such as to explain how the general Π behaved on homotopies, we also needed a theory of tensor products, found in [61], so that the resulting theory is quite complex. It is also remarkable that ideas of Whitehead in [5] played a key role in these results.

4. Main results

Major features of the work over the years with Philip Higgins and others can be summarized in the following diagram of categories and functors:

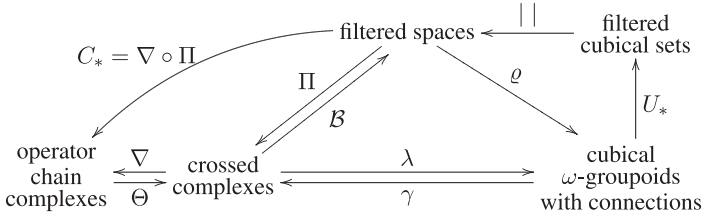


Diagram 4.1.

in which

4.1.1 the categories FTop of filtered spaces, ω-Gpd of cubical ω-groupoids with connections, and Crs of crossed complexes are monoidal closed, and have a notion of homotopy using $\otimes$ and a unit interval object;

4.1.2 ρ, Π are homotopical functors (that is, they are defined in terms of homotopy classes of certain maps) and preserve homotopies;

4.1.3 λ, γ are inverse adjoint equivalences of monoidal closed categories;

4.1.4 there is a natural equivalence $\gamma\rho \simeq \Pi$ so that either ρ or Π can be used as appropriate;

4.1.5 ρ, Π preserve certain colimits and certain tensor products;

4.1.6 the category of chain complexes with a groupoid of operators is monoidal closed, ∇ preserves the monoid structures, and is left adjoint to Θ;

4.1.7 by definition, the *cubical filtered classifying space* is $\mathcal{B}^{\square} = |\,| \circ U_*$, where U_* is the forgetful functor to filtered cubical sets[7] using the filtration of an ω-groupoid by skeleta, and $|\,|$ is geometric realization of a cubical set;

4.1.8 there is a natural equivalence $\Pi \circ \mathcal{B}^{\square} \simeq 1$;

4.1.9 if C is a crossed complex and its cubical classifying space is defined as $B^{\square}C = (\mathcal{B}^{\square}C)_{\infty}$, then for a CW-complex X, and using homotopy as in 4.1.1 for crossed complexes, there is a natural bijection of sets of homotopy classes

$$[X, B^{\square}C] \cong [\Pi X_*, C].$$

Recent applications of the simplicial version of the classifying space are in [62–64].

Here, a *filtered space* consists of a (compactly generated) space X_{∞} and an increasing sequence of subspaces

$$X_* : X_0 \subseteq X_1 \subseteq X_2 \subseteq \cdots \subseteq X_{\infty}.$$

With the obvious morphisms, this gives the category FTop. The tensor product in this category is the usual

$$(X_* \otimes Y_*)_n = \bigcup_{p+q=n} X_p \times Y_q.$$

The closed structure is easy to construct from the law

$$\mathsf{FTop}(X_* \otimes Y_*, Z_*) \cong \mathsf{FTop}(X_*, \mathsf{FTOP}(Y_*, Z_*)).$$

An advantage of this monoidal closed structure is that it allows an enrichment of the category FTop over crossed complexes, ω-Gpd using Π, or ρ applied to $\mathsf{FTOP}(Y_*, Z_*)$.

The structure of *crossed complex* is suggested by the canonical example, the *fundamental crossed complex* ΠX_* of the filtered space X_*. So it is given by a diagram

[7] Cubical sets are defined, analogously to simplicial sets, as functors $K : \square^{op} \to \mathsf{Set}$, where $\square$ is the "box" category with objects I^n and morphisms the compositions of inclusions of faces and of the various projections $I^n \to I^r$ for $n > r$. The *geometric realization* $|K|$ of such a cubical set is obtained by quotienting the disjoint union of the sets $K(I^n) \times I^n$ by the relations defined by the morphisms of $\square$. For more details, see [65], and for variations on the category $\square$ to include for example connections, see [66]. See also Section 9.

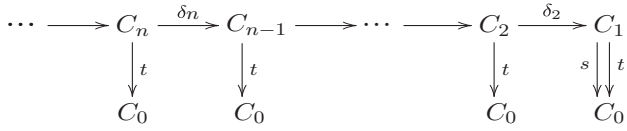


Diagram 4.2.

in which in this example, C_1 is the fundamental groupoid $\pi_1(X_1, X_0)$ of X_1 on the "set of base points" $C_0 = X_0$, while for $n \geqslant 2$, C_n is the family of relative homotopy groups $\{C_n(x)\} = \{\pi_n(X_n, X_{n-1}, x) \mid x \in X_0\}$. The boundary maps are those standard in homotopy theory. There is, for $n \geqslant 2$, an action of the groupoid C_1 on C_n (and of C_1 on the groups $C_1(x)$, $x \in X_0$ by conjugation), the boundary morphisms are operator morphisms, $\delta_{n-1}\delta_n = 0$, $n \geqslant 3$, and the additional axioms are satisfied that

4.3. $b^{-1}cb = c^{\delta_2 b}$, $b, c \in C_2$ so that $\delta_2 : C_2 \to C_1$ is a crossed module (of groupoids);

4.4. if $c \in C_2$ then $\delta_2 c$ acts trivially on C_n for $n \geqslant 3$;

4.5. each group $C_n(x)$ is Abelian for $n \geqslant 3$, and so the family C_n is a C_1-module.

Clearly, we obtain a category Crs of crossed complexes; this category is not so familiar and so we give arguments for using it in the next section.

As algebraic examples of crossed complexes, we have $C = \mathbb{C}(G, n)$ where G is a group, commutative if $n \geqslant 2$, and C is G in dimension n and trivial elsewhere; $C = \mathbb{C}(G, 1 : M, n)$, where G is a group, M is a G-module, $n \geqslant 2$, and C is G in dimension 1, M in dimension n, trivial elsewhere, and zero boundary if $n = 2$; C is a crossed module (of groups) in dimensions 1 and 2 and trivial elsewhere.

A crossed complex C has a fundamental groupoid $\pi_1 C = C_1 / \operatorname{Im} \delta_2$ and also, for $n \geqslant 2$, a family $\{H_n(C, p) | p \in C_0\}$ of homology groups.

5. Why crossed complexes?

- They generalize groupoids and crossed modules to all dimensions. Note that the natural context for second-relative homotopy groups is crossed modules of groupoids rather than groups.
- They are good for modelling CW-complexes.
- Free crossed resolutions enable calculations with small CW-complexes and CW-maps, see Section 15.
- Crossed complexes give a kind of "linear model" of homotopy types, which includes all 2-types. Thus, although they are not the most general model by any means (they do not contain quadratic information such as Whitehead products), this simplicity makes them easier to handle and to relate to classical tools. The new methods and results obtained for crossed complexes can be used as a model

for more complicated situations. This is how a general n-adic Hurewicz theorem was found [67].

- They are convenient for *calculation* and the functor Π is classical, involving relative homotopy groups. We explain some results in this form later.
- They are close to chain complexes with a group(oid) of operators and related to some classical homological algebra (e.g. *chains of syzygies*). In fact if SX is the simplicial singular complex of a space, with its skeletal filtration, then the crossed complex $\Pi(SX)$ can be considered as a slightly noncommutative version of the singular chains of a space.
- The monoidal structure is suggestive of further developments (e.g. *crossed differential algebras*), see [68, 69]. It is used in [70] to give an algebraic model of homotopy 3-types and to discuss automorphisms of crossed modules.
- Crossed complexes have a good homotopy theory, with a *cylinder object, and homotopy colimits*, [71]. The homotopy classification result 4.1.9 generalizes a classical theorem of Eilenberg–Mac Lane. Applications of the simplicial version are given in, for example, [63, 64, 72].
- They have an interesting relation with the Moore complex of simplicial groups and of simplicial groupoids (see Section 18).

6. Why cubical ω-groupoids with connections?

The definition of these objects is more difficult to give but will be indicated in Section 9. Here we explain why these structures are a kind of engine giving the power behind the theory.

- The functor ρ gives a form of *higher homotopy groupoid*, thus confirming the visions of the early topologists.
- They are equivalent to crossed complexes.
- They have a clear *monoidal closed structure*, and a notion of homotopy, from which one can deduce those on crossed complexes, using the equivalence of categories.
- It is easy to relate the functor ρ to tensor products but quite difficult to do this directly for Π.
- Cubical methods, unlike globular or simplicial methods, allow for a simple *algebraic inverse to subdivision*, which is crucial for our local-to-global theorems.
- The additional structure of "connections," and the equivalence with crossed complexes, allows for the sophisticated notion of *commutative cube* and the proof that *multiple compositions of commutative cubes are commutative*. The last fact is a key component of the proof of the HHvKT.
- They yield a construction of a *(cubical) classifying space* $B^{\square}C = (\mathcal{B}^{\square}C)_{\infty}$ of a crossed complex C, which generalizes (cubical) versions of Eilenberg–Mac Lane spaces, including the local coefficient case. This has convenient relation to homotopies.

- There is a current *resurgence of the use of cubes* in, for example, combinatorics, algebraic topology, and concurrency. There is a Dold–Kan type theorem for cubical Abelian groups with connections [73].

7. The equivalence of categories

Let Crs and ω-Gpd denote, respectively, the categories of crossed complexes and ω-groupoids: we use the latter term as an abbreviation of "cubical ω-groupoids with connections." A major part of the work consists in defining these categories and proving their equivalence, which, thus, gives an example of two algebraically defined categories whose equivalence is nontrivial. It is even more subtle than that because the functors $\gamma : \mathsf{Crs} \to \omega\text{−}\mathsf{Gpd}$ and $\lambda : \omega\text{−}\mathsf{Gpd} \to \mathsf{Crs}$ are not hard to define, and it is easy to prove $\gamma\lambda \simeq 1$. The hard part is to prove $\lambda\gamma \simeq 1$, which shows that an ω-groupoid G may be reconstructed from the crossed complex $\gamma(G)$ it contains. The proof involves using the connections to construct a "folding map" $\Phi : G_n \to G_n$, with image $\gamma(G)_n$, and establishing its major properties, including the relations with the compositions. This gives an algebraic form of some old intuitions of several ways of defining relative homotopy groups, for example, using cubes or cells.

On the way, we establish properties of *thin elements*, as those which fold down to 1, and show that G satisfies a strong Kan extension condition, namely that every box has a unique thin filler. This result plays a key role in the proof of the HHvKT for ρ since it is used to show an independence of choice. This part of the proof goes by showing that the two choices can be seen, since we start with a homotopy, as given by the two ends $\partial^{\pm}_{n+1}x$ of an $(n+1)$-cube x. It is then shown by induction, using the method of construction and the above result, that x is degenerate in direction $n+1$. Hence the two ends in that direction coincide.

Properties of the folding map are used also in showing that ΠX_* is actually included in ρX_*, in relating two types of thinness for elements of ρX_*, and in proving a *homotopy addition lemma* in ρX_*.

Any ω-Gpd G has an underlying cubical set UG. If C is a crossed complex, then the cubical set $U(\lambda C)$ is called the *cubical nerve* $N^{\Box}C$ of C. It is a conclusion of the theory that we can also obtain $N^{\Box}C$ as

$$(N^{\Box}C)_n = \mathsf{Crs}(\Pi I^n_*, C),$$

where I^n_* is the usual geometric cube with its standard skeletal filtration. The (cubical) geometric realization $|N^{\Box}C|$ is also called the *cubical classifying space* $B^{\Box}C$ of the crossed complex C. The filtration C^* of C by skeleta gives a filtration $B^{\Box}C^*$ of $B^{\Box}C$ and there is (as in 4.1.6) a natural isomorphism $\Pi(B^{\Box}C^*) \cong C$. Thus, the properties of a crossed complex are those that are universally satisfied by ΠX_*. These proofs use the equivalence of the homotopy categories of Kan[8] cubical sets and of

[8] The notion of Kan cubical set K is also called a cofibrant cubical set. It is an extension condition that any partial r-box in K is the partial boundary of an element of K_r. See, for example, [65], but the idea goes back to the first papers by D. Kan [74, 75].

CW-complexes. We originally took this from the Warwick Masters thesis of S. Hintze, but it is now available with different proofs from Antolini [76] and Jardine [65].

As said above, by taking particular values for C, the classifying space $B^{\square}C$ gives cubical versions of Eilenberg–Mac Lane spaces $K(G, n)$, including the case $n = 1$ and G noncommutative. If C is essentially a crossed module, then $B^{\square}C$ is called the *cubical classifying space* of the crossed module, and in fact realizes the k-invariant of the crossed module.

Another useful result is that if K is a cubical set, then $\rho(|K|_*)$ may be identified with $\rho(K)$, the *free* ω-Gpd *on the cubical set* K, where here $|K|_*$ is the usual filtration by skeleta. On the other hand, our proof that $\Pi(|K|_*)$ is the free crossed complex on the nondegenerate cubes of K uses the generalized HHvKT of the next section.

It is also possible to give simplicial and globular versions of some of the above results because the category of crossed complexes is equivalent also to those of simplicial T-complexes [77] and of globular ∞-groupoids [78]. In fact, the published paper on the classifying space of a crossed complex [79] is given in simplicial terms to link more easily with well-known theories.

8. First main aim of the work: higher Homotopy van Kampen theorems

These theorems give *noncommutative tools for higher dimensional local-to-global problems* yielding a variety of new, often noncommutative, calculations, which *prove* (i.e. test) the theory. We now explain these theorems in a way which strengthens the relation with descent, since that was a theme of the conference at which the talk was given on which this survey is based.

We suppose given an open cover $\mathcal{U} = \{U^\lambda\}_{\lambda\in\Lambda}$ of X. This cover defines a map

$$q : E = \bigsqcup_{\lambda\in\Lambda} U^\lambda \to X$$

and so we can form an augmented simplicial space

$$\check{C}(q) : \cdots E \times_X E \times_X E \;\substack{\longrightarrow\\ \longrightarrow\\ \longrightarrow}\; E \times_X E \;\rightrightarrows\; E \xrightarrow{q} X,$$

where the higher dimensional terms involve disjoint unions of multiple intersections U^ν of the U^λ.

We now suppose given a filtered space X_*, a cover $\mathcal{U}$ as above of $X = X_\infty$, and an augmented simplicial filtered space $\check{C}(q_*)$, involving multiple intersections U^ν_* of the induced filtered spaces.

We still need a connectivity condition.

Definition 8.1. A filtered space X_* is *connected* if and only if the induced maps $\pi_0 X_0 \to \pi_0 X_n$ are surjective and $\pi_n(X_r, X_n, \nu) = 0$ for all $n > 0, r > n$ and $\nu \in X_0$.

THEOREM 8.2 (Main result (HHvKT)). *If U^v_* is connected for all finite intersections U^v of the elements of the open cover, then*
(C) (connectivity) *X_* is connected, and*
(I) (isomorphism) *the following diagram as part of $\rho(\check{C}(q_*))$,*

$$\rho(E_* \times_{X_*} E_*) \rightrightarrows \rho E_* \xrightarrow{\rho(q_*)} \rho X_*, \qquad (c\rho)$$

is a coequalizer diagram. Hence the following diagram of crossed complexes,

$$\Pi(E_* \times_{X_*} E_*) \rightrightarrows \Pi E_* \xrightarrow{\Pi(q_*)} \Pi X_*, \qquad (c\Pi)$$

is also a coequalizer diagram.

So we get calculations of the fundamental crossed complex ΠX_*.

It should be emphasized that to get to and apply, this theorem takes just the two papers [51,60], totalling 58 pages. With this we deduce in the first instance

- the usual vKT for the fundamental groupoid on a set of base points;
- the Brouwer degree theorem ($\pi_n S^n = \mathbb{Z}$);
- the relative Hurewicz theorem;
- Whitehead's theorem that $\pi_n(X \cup \{e^2_\lambda\}, X)$ is a free crossed module;
- an excision result, more general than the previous two, on $\pi_n(A \cup B, A, x)$ as an induced module (crossed module if $n = 2$) when $(A, A \cap B)$ is $(n-1)$-connected.

The assumptions required of the reader are quite small, just some familiarity with CW-complexes. This contrasts with some expositions of basic homotopy theory, where the proof of say the relative Hurewicz theorem requires knowledge of singular homology theory. Of course, it is surprising to get this last theorem without homology, but this is because it is seen as a statement on the morphism of relative homotopy groups

$$\pi_n(X, A, x) \to \pi_n(X \cup CA, CA, x) \cong \pi_n(X \cup CA, x)$$

and is obtained, like our proof of Theorem W, as a special case of an excision result. The reason for this success is that *we use algebraic structures that model the underlying processes of the geometry* more closely than those in common use. These algebraic structures and their relations are quite intricate, as befits the complications of homotopy theory, so the theory is tight knit.

Note also that these results cope well with the action of the fundamental group on higher homotopy groups.

The calculational use of the HHvKT for ΠX_* is enhanced by the relation of Π with tensor products (see Section 15 for more details).

9. The fundamental cubical ω-groupoid ρX_* of a filtered space X_*

Here are the basic elements of the construction.

I^n_*: the n-cube with its skeletal filtration.

Set $R_n X_* = \mathsf{FTop}(I^n_*, X_*)$. This is a *cubical set with compositions, connections, and inversions*.

For $i = 1, \ldots, n$, there are standard:
face maps $\partial_i^{\pm} : R_n X_* \to R_{n-1} X_*$;
degeneracy maps $\varepsilon_i : R_{n-1} X_* \to R_n X_*$
connections $\Gamma_i^{\pm} : R_{n-1} X_* \to R_n X_*$
compositions $a \circ_i b$ defined for $a, b \in R_n X_*$ such that $\partial_i^+ a = \partial_i^- b$
inversions $-_i : R_n \to R_n$.

The connections are induced by $\gamma_i^{\alpha} : I^n \to I^{n-1}$ defined using the monoid structures max, min : $I^2 \to I$. They are essential for many reasons, for example, to discuss the notion of *commutative cube*.

These operations have certain algebraic properties that are easily derived from the geometry and which we do not itemize here (see, e.g. [80]). These were listed first in the Bangor thesis of Al-Agl [81] (In the paper [51], the only basic connections needed are the Γ_i^+, from which the Γ_i^- are derived using the inverses of the groupoid structures.).

Now it is natural and convenient to define $f \equiv g$ for $f, g : I_*^n \to X_*$ to mean f is homotopic to g through filtered maps and relative to the vertices of I^n. This gives a quotient map

$$p : R_n X_* \to \rho_n X_* = (R_n X_* / \equiv).$$

The following results are proved in [60].

9.1. The compositions on RX_* are inherited by ρX_* to give ρX_*, the structure of cubical multiple groupoid with connections.

9.2. The map $p : RX_* \to \rho X_*$ is a Kan fibration of cubical sets.

The proofs of both results use methods of collapsing, which are indicated in the next section. The second result is almost unbelievable. Its proof has to give a systematic method of deforming a cube with the right faces "up to homotopy" into a cube with exactly the right faces, using the given homotopies. In both cases, the assumption that the relation $\equiv$ uses homotopies relative to the vertices is essential to start the induction. (In fact, the paper [60] does not use homotopy relative to the vertices, but imposes an extra condition J_0, that each loop in X_0 is contractible X_1, which again starts the induction. This condition is awkward in applications, for example, to function spaces. A full exposition of the whole story is in preparation [82].)

An essential ingredient in the proof of the HHvKT is the notion of *multiple composition*. We have discussed this already in dimension 2, with a suggestive picture in the diagram (3). In dimension n, the aim is to give algebraic expression to the idea of a cube I^n being subdivided by hyperplanes parallel to the faces into many small cubes, a subdivision with a long history in mathematics.

Let $(m) = (m_1, \ldots, m_n)$ be an n-tuple of positive integers and

$$\phi_{(m)} : I^n \to [0, m_1] \times \cdots \times [0, m_n]$$

be the map $(x_1, \ldots, x_n) \mapsto (m_1 x_1, \ldots, m_n x_n)$. Then a *subdivision of type* (m) of a map $\alpha : I^n \to X$ is a factorization $\alpha = \alpha' \circ \phi_{(m)}$; its *parts* are the cubes $\alpha_{(r)}$, where $(r) = (r_1, \ldots, r_n)$ is an n-tuple of integers with $1 \leqslant r_i \leqslant m_i$, $i = 1, \ldots, n$, and

where $\alpha_{(r)} : I^n \to X$ is given by

$$(x_1, \ldots, x_n) \mapsto \alpha'(x_1 + r_1 - 1, \ldots, x_n + r_n - 1).$$

We then say that α is the *composite* of the cubes $\alpha_{(r)}$ and write $\alpha = [\alpha_{(r)}]$. The *domain* of $\alpha_{(r)}$ is then the set $\{(x_1, \ldots, x_n) \in I^n : r_i - 1 \leqslant x_i \leqslant r_i, 1 \leqslant i \leqslant n\}$. This ability to express "algebraic inverse to subdivision" is one benefit of using cubical methods.

Similarly, in a cubical set with compositions satisfying the interchange law, we can define the multiple composition $[\alpha_{(r)}]$ of a multiple array $(\alpha_{(r)})$ provided the obviously necessary multiple incidence relations of the individual $\alpha_{(r)}$ to their neighbors are satisfied.

Here is an application that is essential in many proofs and seems hard to prove without the techniques involved in 9.2.

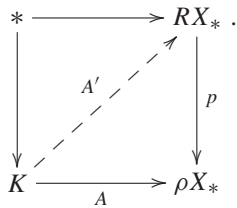
THEOREM 9.3 (Lifting multiple compositions). *Let* $[\alpha_{(r)}]$ *be a multiple composition in* $\rho_n X_*$. *Then representatives* $a_{(r)}$ *of the* $\alpha_{(r)}$ *may be chosen so that the multiple composition* $[a_{(r)}]$ *is well defined in* $R_n X_*$.

Proof: The multiple composition $[\alpha_{(r)}]$ determines a cubical map

$$A : K \to \rho X_*,$$

where the cubical set K corresponds to a representation of the multiple composition by a subdivision of the geometric cube so that top cells $c_{(r)}$ of K are mapped by A to $\alpha_{(r)}$.

Consider the diagram, in which $*$ is a corner vertex of K,



Then K collapses to $*$, written $K \searrow *$. (As an example, see how the subdivision in the diagram (3) may be collapsed row by row to a point.) By the fibration result, A lifts to A', which represents $[a_{(r)}]$, as required.

So we have to explain collapsing.

10. Collapsing

We use a basic notion of collapsing and expanding due to J.H.C. Whitehead [6].

Let $C \subseteq B$ be subcomplexes of I^n. We say C is an *elementary collapse* of B, $B \overset{e}{\searrow} C$, if for some $s \geqslant 1$, there is an s-cell a of B and $(s-1)$-face b of a, the *free face*, such that

$$B = C \cup a, \qquad C \cap a = \dot{a} \setminus b$$

(where $\dot{a}$ denotes the union of the proper faces of a).

We say B_1 *collapses* to B_r, written $B_1 \searrow B_r$, if there is a sequence

$$B_1 \overset{e}{\searrow} B_2 \overset{e}{\searrow} \cdots \overset{e}{\searrow} B_r$$

of elementary collapses.

If C is a subcomplex of B, then

$$B \times I \searrow (B \times \{0\} \cup C \times I)$$

(this is proved by induction on dimension of $B \setminus C$).

Further, I^n collapses to any one of its vertices (this may be proved by induction on n using the first example). These collapsing techniques allow the construction of the extensions of filtered maps and filtered homotopies that are crucial for proving 9.1 that ρX_* does obtain the structure of multiple groupoid.

However, more subtle collapsing techniques using partial boxes are required to prove the fibration theorem 9.2, as partly explained in the next section.

11. Partial boxes

Let C be an r-cell in the n-cube I^n. Two $(r-1)$-faces of C are called *opposite* if they do not meet.

A partial box in C is a subcomplex B of C generated by one $(r-1)$-face b of C (called a *base* of B) and a number, possibly zero, of other $(r-1)$-faces of C, none of which is opposite to b.

The partial box is a box if its $(r-1)$-cells consist of all but one of the $(r-1)$-faces of C.

The proof of the fibration theorem uses a filter homotopy extension property and the following:

PROPOSITION 11.1 (Key Proposition). *Let B, B' be partial boxes in an r-cell C of I^n such that $B' \subseteq B$. Then there is a chain*

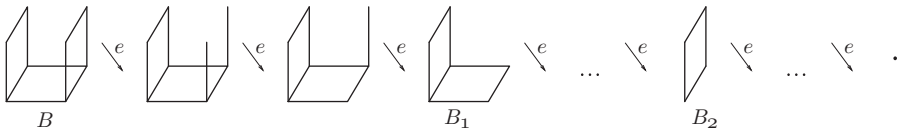
$$B = B_s \searrow B_{s-1} \searrow \cdots \searrow B_1 = B'$$

such that

(i) *each B_i is a partial box in C;*
(ii) *$B_{i+1} = B_i \cup a_i$, where a_i is an $(r-1)$-cell of C not in B_i;*
(iii) *$a_i \cap B_i$ is a partial box in a_i.*

The proof is quite neat and follows the pictures. Induction up such a chain of partial boxes is one of the steps in the proof of the fibration theorem 9.2. The proposition implies that an inclusion of partial boxes is what is known as an anodyne extension [65].

Here is an example of a sequence of collapsings of a partial box B, which illustrate some choices in forming a collapse $B \searrow \mathbf{0}$ through two other partial boxes B_1 and B_2.



The proof of the fibration theorem gives a program for carrying out the deformations needed to do the lifting. In some sense, it implies computing a multiple composition that can be done using collapsing as the guide.

Methods of collapsing generalize methods of trees in dimension 1.

12. Thin elements

Another key concept is that of *thin element* $\alpha \in \rho_n X_*$ for $n \geqslant 2$. The proofs here use strong results of [51].

We say α is *geometrically thin* if it has a *deficient* representative, that is, an $a : I^n_* \to X_*$ such that $a(I^n) \subseteq X_{n-1}$.

We say α is *algebraically thin* if it is a multiple composition of degenerate elements or those coming from repeated (including 0) negatives of connections. Clearly, any multiple composition of algebraically thin elements is thin.

THEOREM 12.1. (i) *Algebraically thin is equivalent to geometrically thin.*
(ii) *In a cubical ω-groupoid with connections, any box has a unique thin filler.*

PROOF. The proof of the forward implication in (i) uses lifting of multiple compositions, in a stronger form than stated above.
The proofs of (ii) and the backward implication in (i) use the full force of the algebraic relation between ω-groupoids and crossed complexes. □

These results allow one to replace arguments with commutative cubes by arguments with thin elements.

13. Sketch proof of the HHvKT

The proof goes by verifying the required universal property. Let $\mathcal{U}$ be an open cover of X as in Theorem 8.2.

We go back to the following diagram, whose top row is part of $\rho(\check{\mathrm{C}}(q_*))$,

$$\begin{array}{ccccc} \rho(E_* \times_{X_*} E_*) & \overset{\partial_0}{\underset{\partial_1}{\rightrightarrows}} & \rho(E_*) & \xrightarrow{\rho(q_*)} & \rho X_* \\ & & & \searrow^{f} & \downarrow{\scriptstyle f'} \\ & & & & G \end{array} \qquad (\mathrm{c}\rho)$$

To prove this top row is a coequalizer diagram, we suppose given a morphism $f : \rho(E_*) \to G$ of cubical ω-groupoids with connection such that $f \circ \partial_0 = f \circ \partial_1$, and we prove that there is a unique morphism $f' : \rho X_* \to G$ such that $f' \circ \rho(q_*) = f$.

To define $f'(\alpha)$ for $\alpha \in \rho X_*$, you subdivide a representative a of α to give $a = [a_{(r)}]$ so that each $a_{(r)}$ lies in an element $U^{(r)}$ of $\mathcal{U}$; use the connectivity conditions and this subdivision to deform a into $b = [b_{(r)}]$ so that

$$b_{(r)} \in R(U^{(r)}_*)$$

and so obtain

$$\beta_{(r)} \in \rho(U^{(r)}_*).$$

The elements

$$f\beta_{(r)} \in G$$

may be composed in G (by the conditions on f) to give an element

$$\theta(\alpha) = [f\beta_{(r)}] \in G.$$

So the proof of the universal property has to use an *algebraic inverse to subdivision*. Again an analogy here is with sending an email: the element you start with is subdivided, deformed so that each part is correctly labelled, the separate parts are sent, and then recombined.

The proof that $\theta(\alpha)$ is independent of the choices made uses crucially properties of thin elements. The key point is *a filter homotopy $h : \alpha \equiv \alpha'$ in $R_n X_*$ gives a deficient element of $R_{n+1} X_*$*.

The method is to do the subdivision and deformation argument on such a homotopy, push the little bits in some

$$\rho_{n+1}(U^{\lambda}_*)$$

(now thin) over to G, combine them, and get a thin element

$$\tau \in G_{n+1},$$

all of whose faces not involving the direction $(n+1)$ are thin *because h was given to be a filter homotopy*. An inductive argument on unique thin fillers of boxes then shows that τ *is degenerate in direction* $(n+1)$ so that the two ends in direction $(n+1)$ are the same.

This ends a rough sketch of the proof of the HHvKT for ρ.

Note that the theory of these forms of multiple groupoids is designed to make this last argument work. We replace a formula for saying a cube h has commutative boundary by a statement that h is thin. It would be very difficult to replace the above argument, on the composition of thin elements, by a higher dimensional manipulation of formulae such as that given in Section 3 for a commutative 3-cube.

Further, the proof does not require knowledge of the existence of all coequalizers, nor does it give a recipe for constructing these in specific examples.

14. Tensor products and homotopies

The construction of the monoidal closed structure on the category ω-Gpd is based on rather formal properties of cubical sets, and the fact that for the cubical set $\mathbb{I}^n$, we have $\mathbb{I}^m \otimes \mathbb{I}^n \cong \mathbb{I}^{m+n}$. The details are given in [61]. The equivalence of categories implies then that the category Crs is also monoidal closed, with a natural isomorphism

$$\mathsf{Crs}(A \otimes B, C) \cong \mathsf{Crs}(A, \mathsf{CRS}(B, C)).$$

Here the 'internal hom' $\mathsf{CRS}(B, C)$ is a crossed complex and its elements in dimension n are: for $n = 0$ the morphisms $B \to C$; for $n = 1$ are homotopies of morphisms; and for $n \geqslant 2$ are forms of higher homotopies. The precise description of these is obtained of course by tracing out in detail the equivalence of categories. It should be emphasized that certain choices are made in constructing this equivalence, and these choices are reflected in the final formulae that are obtained.

An important result is that if X_*, Y_* are filtered spaces, then there is a natural transformation

$$\begin{aligned} \eta : \rho X_* \otimes \rho Y_* &\to \rho(X_* \otimes Y_*) \\ [a] \otimes [b] &\mapsto [a \otimes b], \end{aligned}$$

where if $a : I^m_* \to X_*$, $b : I^n_* \to Y_*$, then $a \otimes b : I^{m+n}_* \to X_* \otimes Y_*$. It is not hard to see, in this cubical setting, that η is well defined. It can also be shown using previous results that η is an isomorphism if X_*, Y_* are the geometric realizations of cubical sets with the usual skeletal filtration.

The equivalence of categories now gives a natural transformation of crossed complexes

$$\eta' : \Pi X_* \otimes \Pi Y_* \to \Pi(X_* \otimes Y_*). \tag{8}$$

It would be hard to construct this directly. It is proved in [79] that η' is an isomorphism if X_*, Y_* are the skeletal filtrations of CW-complexes. The proof uses the HHvKT, and the fact that $A \otimes -$ on crossed complexes has a right adjoint and so preserves colimits. It is proved in [69] that η is an isomorphism if X_*, Y_* are cofibred, connected filtered spaces. This applies, in particular, to the useful case of the filtration $B^{\Box} C^*$ of the classifying space of a crossed complex.

It turns out that the defining rules for the tensor product of crossed complexes, which follows from the above construction, are obtained as follows. We first define a bimorphism of crossed complexes.

Definition 14.1. A *bimorphism* $\theta : (A, B) \to C$ of crossed complexes is a family of maps $\theta : A_m \times B_n \to C_{m+n}$ satisfying the following conditions, where $a \in A_m, b \in B_n, a_1 \in A_1, b_1 \in B_1$ (temporarily using additive notation throughout the definition):

(i)
$$\beta(\theta(a, b)) = \theta(\beta a, \beta b) \text{ for all } a \in A, b \in B.$$

(ii)

$$\theta(a, b^{b_1}) \;=\; \theta(a, b)^{\theta(\beta a, b_1)} \;\text{ if }\; m \geqslant 0, n \geqslant 2,$$

$$\theta(a^{a_1}, b) \;=\; \theta(a, b)^{\theta(a_1, \beta b)} \;\text{ if }\; m \geqslant 2, n \geqslant 0.$$

(iii)

$$\theta(a, b + b') = \begin{cases} \theta(a, b) + \theta(a, b') & \text{if } m = 0, n \geqslant 1 \text{ or } m \geqslant 1, n \geqslant 2, \\ \theta(a, b)^{\theta(\beta a, b')} + \theta(a, b') & \text{if } m \geqslant 1, n = 1, \end{cases}$$

$$\theta(a + a', b) = \begin{cases} \theta(a, b) + \theta(a', b) & \text{if } m \geqslant 1, n = 0 \text{ or } m \geqslant 2, n \geqslant 1, \\ \theta(a', b) + \theta(a, b)^{\theta(a', \beta b)} & \text{if } m = 1, n \geqslant 1. \end{cases}$$

(iv)

$$\delta_{m+n}(\theta(a, b)) = \begin{cases} \theta(\delta_m a, b) + (-)^m \theta(a, \delta_n b) & \text{if } m \geqslant 2, n \geqslant 2, \\ -\theta(a, \delta_n b) - \theta(\beta a, b) + \theta(\alpha a, b)^{\theta(a, \beta b)} & \text{if } m = 1, n \geqslant 2, \\ (-)^{m+1}\theta(a, \beta b) + (-)^m \theta(a, \alpha b)^{\theta(\beta a, b)} + \theta(\delta_m a, b) & \text{if } m \geqslant 2, n = 1, \\ -\theta(\beta a, b) - \theta(a, \alpha b) + \theta(\alpha a, b) + \theta(a, \beta b) & \text{if } m = n = 1. \end{cases}$$

(v)

$$\delta_{m+n}(\theta(a, b)) \;=\; \begin{cases} \theta(a, \delta_n b) & \text{if } m = 0, n \geqslant 2\ , \\ \theta(\delta_m a, b) & \text{if } m \geqslant 2, n = 0\ . \end{cases}$$

(vi)

$$\alpha(\theta(a, b)) \;=\; \theta(a, \alpha b) \quad \text{and} \quad \beta(\theta(a, b)) \;=\; \theta(a, \beta b) \quad \text{if } m = 0, n = 1\ ,$$

$$\alpha(\theta(a, b)) \;=\; \theta(\alpha a, b) \quad \text{and} \quad \beta(\theta(a, b)) \;=\; \theta(\beta a, b) \quad \text{if } m = 1, n = 0\ .$$

The *tensor product* of crossed complexes A, B is given by the universal bimorphism $(A, B) \to A \otimes B$, $(a, b) \mapsto a \otimes b$. The rules for the tensor product are obtained by replacing $\theta(a, b)$ by $a \otimes b$ in the above formulae.

The conventions for these formulae for the tensor product arise from the derivation of the tensor product via the category of cubical ω-groupoids with connections, and the formulae are forced by our conventions for the equivalence of the two categories [51, 61].

The complexity of these formulae is directly related to the complexities of the cell structure of the product $E^m \times E^n$, where the n-cell E^n has cell structure e^0 if $n = 0$, $e^0_+ \cup e^1$ if $n = 1$, and $e^0 \cup e^{n-1} \cup e^n$ if $n \geqslant 2$.

It is proved in [61] that the bifunctor $- \otimes -$ is symmetric and that if a_0 is a vertex of A, then the morphism $B \to A \otimes B,\ b \to a_0 \otimes b$ is injective.

There is a standard groupoid model I of the unit interval, namely the indiscrete groupoid on two objects $0, 1$. This is easily extended trivially to either a crossed complex or an ω-Gpd. So using $\otimes$, we can define a "cylinder object" $\mathsf{I} \otimes -$ in these categories and so a homotopy theory [71].

15. Free crossed complexes and free crossed resolutions

Let C be a crossed complex. A *free basis* B_* for C consists of the following:
B_0 is set which we take to be C_0;
B_1 is a graph with source and target maps $s, t : B_1 \to B_0$ and C_1 is the free groupoid on the graph B_1: that is, B_1 is a subgraph of C_1 and any graph morphism $B_1 \to G$ to a groupoid G extends uniquely to a groupoid morphism $C_1 \to G$;
B_n is, for $n \geqslant 2$, a totally disconnected subgraph of C_n with target map $t : B_n \to B_0$; for $n = 2$, C_2 is the free crossed C_1-module on B_2 while for $n > 2$, C_n is the free $(\pi_1 C)$-module on B_n.

It may be proved using the HHvKT that if X_* is a CW-complex with the skeletal filtration, then ΠX_* is the free crossed complex on the characteristic maps of the cells of X_*. It is proved in [79] that the tensor product of free crossed complexes is free.

A *free crossed resolution* F_* of a groupoid G is a free crossed complex that is aspherical together with an isomorphism $\phi : \pi_1(F_*) \to G$. Analogs of standard methods of homological algebra show that free crossed resolutions of a group are unique up to homotopy equivalence.

To apply this result to free crossed resolutions, we need to replace free crossed resolutions by CW-complexes. A fundamental result for this is the following, which goes back to Whitehead [6] and Wall [83], and which is discussed further by Baues in [84, Chapter VI, Section 7]:

THEOREM 15.1. *Let X_* be a CW-filtered space, and let $\phi : \Pi X_* \to C$ be a homotopy equivalence to a free crossed complex with a preferred free basis. Then, there is a CW-filtered space Y_*, and an isomorphism $\Pi Y_* \cong C$ of crossed complexes with preferred basis such that ϕ is realized by a homotopy equivalence $X_* \to Y_*$.*

In fact, as pointed out by Baues, Wall states his result in terms of chain complexes, but the crossed complex formulation seems more natural, and avoids questions of realizability in dimension 2, which are unsolved for chain complexes.

COROLLARY 15.2. *If A is a free crossed resolution of a group G, then A is realized as free crossed complex with preferred basis by some CW-filtered space Y_*.*

PROOF. We only have to note that the group G has a classifying CW-space BG, whose fundamental crossed complex $\Pi(BG)$ is homotopy equivalent to A. □

Baues also points out in [84, p. 657] an extension of these results, which we can apply to the realization of morphisms of free crossed resolutions. A new proof of this extension is given by Faria Martins in [85], using methods of Ashley [77].

PROPOSITION 15.3. *Let $X = K(G,1)$, $Y = K(H,1)$ be CW-models of Eilenberg–Mac Lane spaces and let $h : \Pi X_* \to \Pi(Y_*)$ be a morphism of their fundamental crossed complexes with the preferred bases given by skeletal filtrations. Then $h = \Pi(g)$ for some cellular $g : X \to Y$.*

PROOF. Certainly, h is homotopic to $\Pi(f)$ for some $f : X \to Y$ since the set of pointed homotopy classes $X \to Y$ is bijective with the morphisms of groups $A \to B$. The result follows from [84, p. 657, (**)] (if f is Π-realizable, then each element in the homotopy class of f is Π-realizable). □

These results are exploited in [86, 87] to calculate free crossed resolutions of the fundamental groupoid of a graph of groups.

An algorithmic approach to the calculation of free crossed resolutions for groups is given in [17] by constructing partial contracting homotopies for the universal cover at the same time as constructing this universal cover inductively. This has been implemented in GAP4 by Heyworth and Wensley [88].

16. Classifying spaces and the homotopy classification of maps

The formal relations of cubical sets and of cubical ω-groupoids with connections and the relation of Kan cubical sets with topological spaces allow the proof of a homotopy classification theorem.

THEOREM 16.1. *If K is a cubical set and G is an ω-groupoid, then there is a natural bijection of sets of homotopy classes*

$$[|K|, |UG|] \cong [\rho(|K|_*), G],$$

where on the left-hand side, we work in the category of spaces, and on the right in ω-groupoids.

Here, $|K|_*$ is the filtration by skeleta of the geometric realization of the cubical set.

We explained earlier how to define a cubical classifying space say $B^{\square}(C)$ of a crossed complex C as $B^{\square}(C) = |UN^{\square}C| = |U\lambda C|$. The properties already stated now give the homotopy classification theorem 4.1.9.

It is shown in [60] that for a CW-complex Y, there is a map $p : Y \to B^{\square}\Pi Y_*$, whose homotopy fibre is n-connected if Y is connected and $\pi_i Y = 0$ for $2 \leqslant i \leqslant n-1$. It follows that if also X is a connected CW-complex with $\dim X \leqslant n$, then p induces a bijection

$$[X, Y] \to [X, B\Pi Y_*].$$

So, under these circumstances, we get a bijection

$$[X, Y] \to [\Pi X_*, \Pi Y_*]. \tag{9}$$

This result, due to Whitehead [5], translates a topological homotopy classification problem to an algebraic one. We explain below how this result can be translated to a result on chain complexes with operators.

It is also possible to define a simplicial nerve $N^{\Delta}(C)$ of a crossed complex C by

$$N^{\Delta}(C)_n = \mathsf{Crs}(\Pi(\Delta^n), C).$$

The *simplicial classifying space* of C is then defined using the simplicial geometric realization

$$B^{\Delta}(C) = |N^{\Delta}(C)|.$$

The properties of this simplicial classifying space are developed in [79], and in particular an analog of 4.1.9 is proved.

The simplicial nerve and an adjointness

$$\mathsf{Crs}(\Pi(L), C) \cong \mathsf{Simp}(L, N^{\Delta}(C))$$

are used in [89, 90] for an equivariant homotopy theory of crossed complexes and their classifying spaces. Important ingredients in this are notions of coherence and an Eilenberg–Zilber type theorem for crossed complexes proved in Tonks' Bangor thesis [91, 92], see also [93].

Labesse in [94] defines a *crossed set*. In fact, a crossed set is exactly a crossed module (of groupoids) $\delta : C \to X \rtimes G$, where G is a group acting on the set X and $X \rtimes G$ is the associated actor groupoid; thus, the simplicial construction from a crossed set described by Larry Breen in [94] is exactly the simplicial nerve of the crossed module, which is regarded as a crossed complex. Hence, the cohomology with coefficients in a crossed set used in [94] is a special case of cohomology with coefficients in a crossed complex, dealt with in [79] (We are grateful to Breen for pointing this out to us in 1999.).

17. Relation with chain complexes with a groupoid of operators

Chain complexes with a group of operators are a well-known tool in algebraic topology, where they arise naturally as the chain complex $C_*\widetilde{X}_*$ of cellular chains of the universal cover $\widetilde{X}_*$ of a reduced CW-complex X_*. The group of operators here is the fundamental group of the space X.

Whitehead in [5] gave an interesting relation between his free crossed complexes (he called them "homotopy systems") and such chain complexes. We refer later to his important homotopy classification results in this area. Here, we explain the relation with the Fox free differential calculus [95].

Let $\mu : M \to P$ be a crossed module of groups and let $G = \mathrm{Coker}\mu$. Then, there is an associated diagram

$$\begin{array}{ccccc} M & \xrightarrow{\mu} & P & \xrightarrow{\phi} & G \\ \downarrow{\scriptstyle h_2} & & \downarrow{\scriptstyle h_1} & & \downarrow{\scriptstyle h_0} \\ M^{\mathrm{ab}} & \xrightarrow[\partial_2]{} & D_\phi & \xrightarrow[\partial_1]{} & \mathbb{Z}[G] \end{array} \tag{10}$$

in which the second row consists of (right) G-modules and module morphisms. Here, h_2 is simply the Abelianization map; $h_1 : P \to D_\phi$ is the universal ϕ-derivation, that is, it satisfies $h_1(pq) = h_1(p)^{\phi q} + h_1(q)$, for all $p, q \in P$, and is universal for this property; and h_0 is the usual derivation $g \mapsto g - 1$. Whitehead in his Lemma 7 of [5] gives this diagram in the case P is a free group, when he takes D_ϕ to be the free G-module on the same generators as the free generators of P. Our formulation, which uses the derived module due to Crowell [96], includes his case. It is remarkable that diagram (10) is a commutative diagram in which the vertical maps are operator morphisms and that the bottom row is defined by this property. The proof in [97] follows essentially Whitehead's proof. The bottom row is exact: this follows from results in [96], and is a reflection of a classical fact on group cohomology, namely the relation between central extensions and the Ext functor, (see [21]). In the case the crossed module is the crossed module $\delta : C(\omega) \to F(X)$ derived from a presentation of a group, then $C(\omega)^{\mathrm{ab}}$ is isomorphic to the free G-module on R, D_ϕ is the free G-module on X, and it is immediate from the above that ∂_2 is the usual derivative $(\partial r/\partial x)$ of Fox's free differential calculus [95]. Thus, Whitehead's results anticipate those of Fox.

It is also proved in [5] that if the restriction $M \to \mu(M)$ of μ has a section, which is a morphism but not necessarily a P-map, then h_2 maps $\mathrm{Ker}\ \mu$ isomorphically to $\mathrm{Ker}\ \partial_2$. This allows calculation of the module of identities among relations by using module methods, and this is commonly exploited, see for example [16] and the references there.

Whitehead introduced the categories CW of reduced CW-complexes, HS of homotopy systems, and FCC of free chain complexes with a group of operators, together with functors

$$\mathsf{CW} \xrightarrow{\Pi} \mathsf{HS} \xrightarrow{C} \mathsf{FCC}.$$

In each of these categories, he introduced notions of homotopy and he proved that C induces an equivalence of the homotopy category of HS with a subcategory of the homotopy category of FCC. Further, $C\Pi X_*$ is isomorphic to the chain complex $C_*\widetilde{X}_*$ of cellular chains of the universal cover of X so that under these circumstances, there is a bijection of sets of homotopy classes

$$[\Pi X_*, \Pi Y_*] \to [C_*\widetilde{X}_*, C_*\widetilde{Y}_*]. \tag{11}$$

This with the bijection (9) can be interpreted as an operator version of the Hopf classification theorem. It is surprisingly little known. It includes results of Olum [98]

published later, and it enables quite useful calculations to be done easily, such as the homotopy classification of maps from a surface to the projective plane [99], and other cases. Thus, we see once again that this general theory leads to specific calculations.

All these results are generalized in [97] to the nonfree case and to the nonreduced case, which requires a groupoid of operators, thus giving functors

$$\mathsf{FTop} \xrightarrow{\Pi} \mathsf{Crs} \xrightarrow{\nabla} \mathsf{Chain}.$$

(The paper [97] uses the notation Δ for this ∇.) One utility of the generalization to groupoids is that the functor ∇ then has a right adjoint and so preserves colimits. An example of this preservation is given in [97, Example 2.10]. The construction of the right adjoint Θ to ∇ builds on a number of constructions used earlier in homological algebra.

The definitions of the categories under consideration to obtain a generalization of the bijection (11) has to be quite careful, since it works in the groupoid case, and not all morphisms of the chain complex are realizable.

This analysis of the relations between these two categories is used in [79] to give an account of cohomology with local coefficients.

It is also proved in [97] that the functor ∇ preserves tensor products, where the tensor in the category Chain is a generalization to modules over groupoids of the usual tensor for chain complexes of modules of groups. Since the tensor product is described explicitly in dimensions $\leqslant 2$ in [61] and $(\nabla C)_n = C_n$ for $n \geqslant 3$, this preservation yields a complete description of the tensor product of crossed complexes.

18. Crossed complexes and simplicial groups and groupoids

The Moore complex NG of a simplicial group G is not in general a (reduced) crossed complex. Let D_nG be the subgroup of G_n generated by degenerate elements. Ashley showed in his thesis [77] that NG is a crossed complex if and only if $(NG)_n \cap (DG)_n = \{1\}$ for all $n \geqslant 1$.

Ehlers and Porter in [100, 101] show that there is a functor C from simplicial groupoids to crossed complexes in which $C(G)_n$ is obtained from $N(G)_n$ by factoring out

$$(NG_n \cap D_n)d_{n+1}(NG_{n+1} \cap D_{n+1}),$$

where the Moore complex is defined so that its differential comes from the last simplicial face operator.

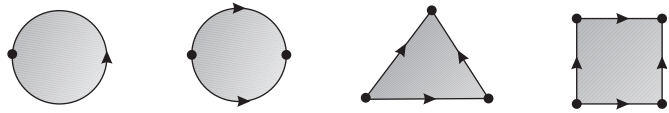
This is one part of an investigation into the Moore complex of a simplicial group, of which the most general investigation is by Carrasco and Cegarra in [102].

An important observation in [103] is that if $N \triangleleft G$ is an inclusion of a normal simplicial subgroup of a simplicial group, then the induced morphism on components $\pi_0(N) \to \pi_0(G)$ obtains the structure of crossed module. This is directly analogous to the fact that if $F \to E \to B$ is a fibration sequence, then the induced morphism of fundamental groups $\pi_1(F, x) \to \pi_1(E, x)$ also obtains the structure of crossed

module. This last fact is relevant to algebraic K-theory, where for a ring R, the homotopy fibration sequence is taken to be $F \to B(GL(R)) \to B(GL(R))^+$.

19. Other homotopy multiple groupoids

A natural question is whether there are other useful forms of higher homotopy groupoids. It is because the geometry of convex sets is so much more complicated in dimensions > 1 than in dimension 1 that new complications emerge for the theories of higher order group theory and of higher homotopy groupoids. We have different geometries, for example, those of disks, globes, simplices, cubes, as shown in dimension 2 in the following diagram.



The cellular decomposition for an n-disk is $D^n = e^0 \cup e^{n-1} \cup e^n$ and that for globes is

$$G^n = e^0_\pm \cup e^1_\pm \cup \cdots \cup e^{n-1}_\pm \cup e^n.$$

The higher dimensional group(oid) theory reflecting the n-disks is that of crossed complexes and that for the n-globes is called *globular ω-groupoids*.

A common notion of higher dimensional category is that of n-category, which generalize the 2-categories studied in the late 1960s. A 2-category C is a category enriched in categories, in the sense that each hom set $\mathsf{C}(x, y)$ is given the structure of category, and there are appropriate axioms. This gives inductively the notion of an n-category as a category enriched in $(n-1)$-categories. This is called a "globular" approach to higher categories. The notion of n-category for all n was axiomatized in [78] and called an ∞-category; its underlying geometry was also axiomatized in this paper as a family of sets S_n, $n \geqslant 0$ with operations

$$D^\alpha_i : S_n \to S_i,\ E_i : S_i \to S_n, \alpha = 0, 1; i = 1, \ldots, n-1$$

satisfying appropriate laws. Such a structure was later called a "globular set" [104], and the term ω-category was used instead of the earlier ∞-category. Difficulties of the globular approach are to define multiple compositions, and also monoidal closed structures, although these are clear in the cubical approach. A globular higher homotopy groupoid of a filtered space has been constructed in [105], deduced from cubical results in [60, 80].

Although the proof of the HHvKT outlined earlier does seem to require cubical methods, there is still a question of the place of globular and simplicial methods in this area. A simplicial analog of the equivalence of categories is given in [77, 106], using Dakin's notion of *simplicial T-complex* [107]. However, it is difficult to describe in detail the notion of tensor product of such structures or to formulate a proof of the

HHvKT theorem in that context. There is a tendency to replace the term T-complex from all this earlier work such as [77, 108] by *complicial set* [109].

It is easy to define a homotopy globular *set* $\rho^{\bigcirc} X_*$ of a filtered space X_*, but it is not quite so clear how to prove directly that the expected compositions are well defined. However, there is a natural graded map

$$i : \rho^{\bigcirc} X_* \to \rho X_* \tag{12}$$

and applying the folding map of [80,81] analogously to methods in [60] allows one to prove that i of (12) is injective. It follows that the compositions on ρX_* are inherited by $\rho^{\bigcirc} X_*$ to make the latter a globular ω-groupoid. The details are in [105].

Loday in 1982 [110] defined the fundamental cat^n-group of an n-cube of spaces (a cat^n-group may be defined as an n-fold category internal to the category of groups) and showed that cat^n-groups model all reduced weak homotopy $(n+1)$-types. Joint work [7] formulated and proved a HHvKT for the cat^n-group functor from n-cubes of spaces. This allows new local to global calculations of certain homotopy n-types [111] and also an n-adic Hurewicz theorem [67]. This work obtains more powerful results than the purely linear theory of crossed complexes. It yields a group-theoretic description of the first nonvanishing homotopy group of a certain $(n+1)$-ad of spaces, so several formulae for the homotopy and homology groups of specific spaces; [112] gives new applications. Porter in [103] gives an interpretation of Loday's results using methods of simplicial groups. There is clearly a lot to do in this area. See [113] for relations of cat^n-groups with homological algebra.

Recently, some absolute homotopy 2-groupoids and double groupoids have been defined, see [114] and the references there, while [115] applies generalized Galois theory to give a new homotopy double groupoid of a map, generalizing previous work of [52]. It is significant that crossed modules have been used in a differential topology situation by Mackaay and Picken [116]. Reinterpretations of these ideas in terms of double groupoids are started in [117].

It seems reasonable to suggest that in the most general case, double groupoids are still somewhat mysterious objects. The paper [118] gives a kind of classification of them.

20. Conclusion and questions

- The emphasis on filtered spaces rather than the absolute case is open to question. But no useful definition of a higher homotopy groupoid of a space with clear uses has been proposed. The work of Loday makes use of n-cubes of filtered spaces (see [110, 119]).
- *Mirroring the geometry by the algebra* is crucial for conjecturing and proving universal properties.
- *Thin elements* are crucial for modelling a concept not so easy to define or handle algebraically, that of commutative cubes. (see also [55, 120]).
- The cubical methods summarized in Section 9 have also been applied in concurrency theory, (see, e.g. [121, 122].)

- *HHvKT theorems* give, when they apply, exact information even in noncommutative situations. The implications of this for homological algebra could be important.
- One construction inspired eventually by this work, the *non-Abelian tensor product of groups*, has a bibliography of 90 papers, since it was defined with Loday in [7].
- Globular methods do fit into this scheme. They have not so far yielded new calculations in homotopy theory (see [105]) but have been applied to directed homotopy theory [121]. Globular methods are the main tool in approaches to weak category theory (see, e.g. [104, 123]) although the potential of cubical methods in that area is hinted at in [120].
- For computations, we really need strict structures (although we do want to compute invariants of homotopy colimits).
- No work seems to have been done on Poincaré duality, that is, on finding special qualities of the fundamental crossed complex of the skeletal filtration of a combinatorial manifold. However, the book by Sharko ([124, Chapter VI]) does use crossed complexes for investigating Morse functions on a manifold.
- In homotopy theory, identifications in low dimensions can affect high-dimensional homotopy. So we need structure in a range of dimensions to model homotopical identifications algebraically. The idea of identifications in low dimensions is reflected in the algebra by "induced constructions."
- In this way, we calculate some crossed modules modeling homotopy 2-types, whereas the corresponding k-invariant is often difficult to calculate.
- The use of crossed complexes in Čech theory is a current project with Jim Glazebrook and Tim Porter.
- **Question:** Are there applications of higher homotopy groupoids in other contexts where the fundamental groupoid is currently used, such as algebraic geometry?
- **Question:** There are uses of double groupoids in differential geometry, for example in Poisson geometry, and in 2-dimensional holonomy [125]. Is there a non-Abelian De Rham theory, using an analog of crossed complexes?
- **Question:** Is there a truly noncommutative integration theory based on limits of multiple compositions of elements of multiple groupoids, in analogy to the way length is defined using limits of sequences of line segments?

References

[1] E.H.v. Kampen, On the connection between the fundamental groups of some related spaces, Amer. J. Math. **55** (1933), 261–267.

[2] E. Čech, 1933, Höherdimensionale Homotopiegruppen, International Congress of Mathematicians (New Series) (4th: 1932: Zurich, Switzerland): Verhandlungen des Internationalen Mathematiker-Kongresses: Zürich 1932: Im Auftrage des Komitees für den Internationalen Mathematiker-Kongress/herausgegeben von Walter Saxer..Zürich: Orell Füssli, 1932–1933..2v.

[3] W. Hurewicz, Beiträge zur Topologie der Deformationen, Nederl. Akad. Wetensch. Proc. Ser. A **38** (1935), 112–119, 521–528, **39** (1936), 117–126, 213–224.

[4] J.H.C. Whitehead, Combinatorial Homotopy I, Bull. Amer. Math. Soc. **55** (1949), 213–245.

[5] J.H.C. Whitehead, Combinatorial Homotopy II, Bull. Amer. Math. Soc. **55** (1949), 453–496.
[6] J.H.C. Whitehead, Simple homotopy types, Amer. J. Math. **72** (1950), 1–57.
[7] R. Brown, J.-L. Loday, Van Kampen theorems for diagrams of spaces, Topology **26** (1987), 311–337.
[8] R. Brown, On the second relative homotopy group of an adjunction space: an exposition of a theorem of J.H.C. Whitehead, J. London Math. Soc. (2) **22** (1980), 146–152.
[9] S. Mac Lane, J.H.C. Whitehead, On the 3-type of a complex, Proc. Nat, Acad. Sci. U.S.A. **36** (1950), 41–48.
[10] J.H.C. Whitehead, On adding relations to homotopy groups, Ann. Math. (2) **42** (1941), 409–428.
[11] R. Peiffer, Uber Identitäten zwischen Relationen, Math. Ann. **121** (1949), 67–99.
[12] K. Reidemeister, Über Identitäten von Relationen, Abh. Math. Sem. Hamburg **16** (1949), 114–118.
[13] R. Brown, J. Huebschmann, Identities among relations, in: R. Brown, T.L. Thickstun, eds., *Low Dimensional Topology*, London Math. Soc Lecture Notes, Cambridge University Press, Cambridge, 1982, pp. 153–202.
[14] S. Pride, Identities among relations, in A.V.E. Ghys, A. Haefliger, eds., *Proc. Workshop on Group Theory from a Geometrical Viewpoint*, International Centre of Theoretical Physics, Trieste, 1990, World Scientific, River Edge, NJ (1991), pp. 687–716.
[15] C. Hog-Angeloni, W. Metzler, eds., Two-dimensional homotopy and combinatorial group theory, London Mathematical Society Lecture Note Series, **197** (1993), Cambridge University Press, Cambridge.
[16] G. Ellis, I. Kholodna, Three-dimensional presentations for the groups of order at most 30, LMS J. Comput. Math. **2** (1999), 93−117+2 appendixes (HTML and source code).
[17] R. Brown, A. Razak Salleh, Free crossed resolutions of groups and presentations of modules of identities among relations, LMS J. Comput. Math. **2** (1999), 28–61 (electronic).
[18] R. Brown, T. Porter, On the Schreier theory of non-abelian extensions: generalisations and computations, Proc. Royal Irish Academy **96A** (1996), 213–227.
[19] A. Fröhlich, Non-Abelian homological algebra. I. Derived functors and satellites, Proc. London Math. Soc. (3) **11** (1961), 239–275.
[20] A.S.T. Lue, Cohomology of groups relative to a variety, J. Algebra **69** (1981), 155–174.
[21] S. Mac Lane, Homology, number **114** in Grundlehren der math. Wiss. **114**, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1963.
[22] P. Dedecker, Sur la cohomologie non abélienne. II, Canad. J. Math. **15** (1963), 84–93.
[23] J. Huebschmann, Crossed n-fold extensions of groups and cohomology, Comment. Math. Helv. **55** (1980), 302–313.
[24] J. Huebschmann, Automorphisms of group extensions and differentials in the Lyndon-Hochschild-Serre spectral sequence, J. Algebra **72** (1981), 296–334.
[25] J. Huebschmann, Group extensions, crossed pairs and an eight term exact sequence, J. fur. reine. und ang. Math. **321** (1981), 150–172.
[26] R. Brown, C.D. Wensley, On finite induced crossed modules, and the homotopy 2-type of mapping cones, Theory Appl. Categories, **1** (1995), 54–70.
[27] R. Brown, C.D. Wensley, Computing crossed modules induced by an inclusion of a normal subgroup, with applications to homotopy 2-types, Theory Appl. Categories, **2** (1996), 3–16.
[28] R. Brown, Coproducts of crossed P-modules: applications to second homotopy groups and to the homology of groups, Topology **23** (1984), 337–345.
[29] The GAP Group, Groups, Algorithms, and Programming, version 4.4, Technical report, http://www.gap-system.org. 2008.
[30] R. Brown, C.D. Wensley, Computations and homotopical applications of induced crossed modules, J. Symb. Comp. **35** (2003), 59–72.
[31] P.J. Higgins, Presentations of Groupoids, with Applications to Groups, Proc. Camb. Phil. Soc. **60** (1964), 7–20.
[32] R. Brown, Groupoids and van Kampen's Theorem, Proc. London Math. Soc. (3) **17** (1967), 385–340.
[33] P.J. Higgins, J. Taylor, The Fundamental Groupoid and Homotopy Crossed Complex of an Orbit Space, in: K.H. Kamps, et al., eds., *Category Theory: Proceedings Gummersbach 1981*, Springer LNM, Berlin-New York, **962** (1982), 115–122.

[34] J. Taylor, Quotients of Groupoids by the Action of a Group, Math. Proc. Camb. Phil. Soc. **103** (1988), 239–249.
[35] R. Brown, Topology and groupoids, Booksurge PLC, S. Carolina, 2006, (previous editions with different titles, 1968, 1988).
[36] P.J. Higgins, Categories and Groupoids, van Nostrand, New York, 1971. Reprints in Theory and Applications of Categories **7** (2005), pp. 1–195.
[37] W. Dicks, E. Ventura, The group fixed by a family of injective endomorphisms of a free group, Contemporary Mathematics, **195** Am. Math. Soc., Providence, RI, (1996) x+81.
[38] C. Ehresmann, Catégories et Structure, Dunod, Paris, 1983.
[39] A. Douady, R. Douady, Algebres et theories Galoisiennes, volume 2, CEDIC, Paris, 1979.
[40] F. Borceux, G. Janelidze, Galois Theories, Cambridge Studies in Advanced Mathematics **72** (2001), Cambridge University Press, Cambridge.
[41] R. Brown, G. Janelidze, Van Kampen theorems for categories of covering morphisms in lextensive categories, J. Pure Appl. Algebra **119** (1997), 255–263.
[42] M. Bunge, S. Lack, Van Kampen theorems for toposes, Adv. Math. **179** (2003), 291–317.
[43] R. Brown, From groups to groupoids: a brief survey, Bull. London Math. Soc. **19** (1987), 113–134.
[44] C. Ehresmann, *Œuvres complètes et commentées. I-1,2. Topologie algébrique et géométrie différentielle, With commentary by W. T. van Est, Michel Zisman, Georges Reeb, Paulette Libermann, René Thom, Jean Pradines, Robert Hermann, Anders Kock, André Haefliger, Jean Bénabou, René Guitart, and Andrée Charles Ehresmann, Edited by Andrée Charles Ehresmann*, Cahiers Topologie Géom. Différentielle **24** (1983) suppll. 1.
[45] R. Brown, *Three themes in the work of Charles Ehresmann: Local-to-global; Groupoids; Higher dimensions*, Proceedings of the 7th Conference on the Geometry and Topology of Manifolds: The Mathematical Legacy of Charles Ehresmann, Bedlewo (Poland) 8.05.2005-15.05.2005, Banach Centre Publications **76** Institute of Mathematics Polish Academy of Sciences, Warsaw, 2007 pp. 51–63. (math.DG/0602499).
[46] P.J. Higgins, Algebras with a scheme of operators, Math. Nachr. **27** (1963), 115–132.
[47] R. Brown, C.B. Spencer, Double groupoids and crossed modules, Cahiers Topologie Géom. Différentielle **17** (1976), 343–362.
[48] R. Brown, C.B. Spencer, $\mathcal{G}$-groupoids, crossed modules and the fundamental groupoid of a topological group, Proc. Kon. Ned. Akad. v. Wet **79** (1976), 296–302.
[49] R. Brown, G.H. Mosa, Double categories, 2-categories, thin structures and connections, Theory Appl. Categories, **5** (1999), No. 7, 163–175 (electronic).
[50] R. Brown, H. Kamps, T. Porter, A homotopy double groupoid of a Hausdorff space II: a van Kampen theorem, Theory Appl. Categories **14** (2005), 200–220.
[51] R. Brown, P.J. Higgins, The algebra of cubes, J. Pure Appl. Alg. **21** (1981), 233–260.
[52] R. Brown, P.J. Higgins, On the connection between the second relative homotopy groups of some related spaces, Proc. London Math. Soc. (3) **36** (1978), 193–212.
[53] R. Brown, Higher dimensional group theory, in: Low dimensional topology, London Math Soc. Lecture Note Series **48** R. Brown, T.L. Thickstun, ed. Cambridge University Press, 1982, 215–238.
[54] C.B. Spencer, An abstract setting for homotopy pushouts and pullbacks, Cahiers Topologie Géom. Différentielle **18** (1977), 409–429.
[55] P.J. Higgins, Thin elements and commutative shells in cubical ω-categories, Theory Appl. Categories, **14** (2005), 60–74.
[56] P. Gaucher, Combinatorics of branchings in higher dimensional automata, Theory Appl. Categories, **8** (2001), 324–376.
[57] R. Brown, Groupoids and crossed objects in algebraic topology, Homology, Homotopy and Applications **1** (1999), 1–78.
[58] A. Connes, Noncommutative geometry, Academic Press Inc., San Diego, CA, 1994.
[59] R. Brown, P.J. Higgins, Sur les complexes croisés d'homotopie associés à quelques espaces filtrés, C. R. Acad. Sci. Paris Sér. A-B **286** (1978), A91–A93.
[60] R. Brown, P.J. Higgins, Colimit theorems for relative homotopy groups, J. Pure Appl. Algebra **22** (1981), 11–41.

[61] R. Brown, P.J. Higgins, Tensor products and homotopies for ω-groupoids and crossed complexes, J. Pure Appl. Algebra **47** (1987), 1–33.
[62] R. Brown, *Exact sequences of fibrations of crossed complexes, homotopy classification of maps, and nonabelian extensions of groups*, J. Homotopy and Related Structures **3** (2008), 331–343.
[63] T. Porter, V. Turaev, Formal Homotopy Quantum Field Theories I: Formal maps and crossed C-algebras, J. Homotopy and Related Structures, **3** (2008), 113–159.
[64] J. Faria Martins, T. Porter, On Yetter's invariant and an extension of the Dijkgraaf-Witten invariant to categorical groups, Theory Appl. Categories **18** (2007), 118–150.
[65] J.F. Jardine, Categorical homotopy theory, Homology, Homotopy Appl. **8** (2006), 71–144.
[66] M. Grandis, L. Mauri, Cubical sets and their site, Theory Applic. Categories **11** (2003), 185–201.
[67] R. Brown, J.-L. Loday, Homotopical excision, and Hurewicz theorems, for n-cubes of spaces, Proc. London Math. Soc. (3) **54** (1987), 176–192.
[68] H.-J. Baues, A. Tonks, On the twisted cobar construction, Math. Proc. Cambridge Philos. Soc. **121** (1997), 229–245.
[69] H.J. Baues, R. Brown, On relative homotopy groups of the product filtration, the James construction, and a formula of Hopf, J. Pure Appl. Algebra **89** (1993), 49–61.
[70] R. Brown, N.D. Gilbert, Algebraic models of 3-types and automorphism structures for crossed modules, Proc. London Math. Soc. (3) **59** (1989), 51–73.
[71] R. Brown, M. Golasiński, A model structure for the homotopy theory of crossed complexes, Cah. Top. Géom. Diff. Cat. **30** (1989), 61–82.
[72] J. Faria Martins, On the homotopy type and fundamental crossed complex of the skeletal filtration of a CW-complex, Homology, Homotopy and Applications **9** (2007), 295–329.
[73] R. Brown, P.J. Higgins, Cubical Abelian groups with connections are equivalent to chain complexes, Homology, Homotopy and Applications **5** (2003), 49–52.
[74] D.M. Kan, Abstract homotopy. I. Proc. Nat. Acad. Sci. U.S.A. **41** (1955), 1092–1096.
[75] D.M. Kan, Abstract homotopy. II. Proc. Nat. Acad. Sci. U.S.A. **42** (1956), 255–258.
[76] R. Antolini, Cubical Structures and Homotopy Theory, Ph.D. thesis, Univ. Warwick, Coventry, 1996.
[77] N. Ashley, Simplicial T-Complexes: a non Abelian version of a theorem of Dold-Kan, Dissertationes Math. **165** (1988), 11–58, (Published version of University of Wales PhD thesis, 1978).
[78] R. Brown, P.J. Higgins, The equivalence of ∞-groupoids and crossed complexes, Cahiers Top. Géom. Diff. **22** (1981), 370–386.
[79] R. Brown, P.J. Higgins, The classifying space of a crossed complex, Math. Proc. Cambridge Philos. Soc. **110** (1991), 95–120.
[80] F. Al-Agl, R. Brown, R. Steiner, Multiple categories: the equivalence between a globular and cubical approach, Adv. Math. **170** (2002), 71–118.
[81] F. Al-Agl, Aspects of Multiple Categories, Ph.D. thesis, University of Wales, Bangor, 1989.
[82] R. Brown, P.J. Higgins, R. Sivera, Nonabelian algebraic topology (2009).
[83] C.T.C. Wall, Finiteness conditions for CW-complexes II, Proc. Roy. Soc. Ser. A, **295** (1966), 149–166.
[84] H.J. Baues, Algebraic Homotopy, volume **15** of Cambridge Studies in Advanced Mathematics, Cambridge Univ. Press, Cambridge, 1989.
[85] J. Faria Martins, A new proof of a theorem of H.J. Baues, Preprint IST, Lisbon, 2007 p. 16.
[86] E.J. Moore, Graphs of Groups: Word Computations and Free Crossed Resolutions, Ph.D. thesis, University of Wales, Bangor, 2001.
[87] R. Brown, E. Moore, T. Porter, C. Wensley, Crossed complexes, and free crossed resolutions for amalgamated sums and HNN-extensions of groups, Georgian Math. J. **9** (2002), 623–644.
[88] A. Heyworth, C.D. Wensley, IdRel - logged rewriting and identities among relators, GAP4, http://www.gap-system.org/Packages/idrel.html, version 2.05, 2008.
[89] R. Brown, M. Golasiński, T. Porter, A. Tonks, Spaces of maps into classifying spaces for equivariant crossed complexes, Indag. Math. (N.S.) **8** (1997), 157–172.
[90] R. Brown, M. Golasiński, T. Porter, A. Tonks, Spaces of maps into classifying spaces for equivariant crossed complexes. II. The general topological group case, K-Theory **23** (2001), 129–155.

[91] A.P. Tonks, Theory and applications of crossed complexes, Ph.D. thesis, University of Wales, Bangor, 1993.
[92] A.P. Tonks, On the Eilenberg–Zilber theorem for crossed complexes, J. Pure Appl. Algebra **179** (2003), 199–220.
[93] R. Brown, R. Sivera, Normalisation for the fundamental crossed complex of a simplicial set, J. Homotopy and Related Structures, Special Issue devoted to the memory of Saunders Mac Lane, **2** (2007), 49–79.
[94] J.-P. Labesse, Cohomologie, stabilisation et changement de base, Astérisque, vi+161, **257** (1999), appendix A by Laurent Clozel and Labesse, and Appendix B by Lawrence Breen.
[95] R.H. Fox, Free differential calculus I: Derivations in the group ring, Ann. Math. **57** (1953), 547–560.
[96] R. Crowell, The derived module of a homomorphism, Adv. Math. **5** (1971), 210–238.
[97] R. Brown, P.J. Higgins, Crossed complexes and chain complexes with operators, Math. Proc. Camb. Phil. Soc. **107** (1990), 33–57.
[98] P. Olum, On mappings into spaces in which certain homotopy groups vanish, Ann. Math. (2) **57** (1953), 561–574.
[99] G.J. Ellis, Homotopy classification the J.H.C. Whitehead way, Exposition. Math. **6** (1988), 97–110.
[100] P.J. Ehlers, T. Porter, Varieties of simplicial groupoids. I. Crossed complexes, J. Pure Appl. Algebra **120** (1997), 221–233.
[101] P.J. Ehlers, T. Porter, Erratum to: Varieties of simplicial groupoids. I. Crossed complexes, J. Pure Appl. Algebra **120** (1997), no. 3, 221–233; J. Pure Appl. Algebra **134** (1999), 207–209.
[102] P. Carrasco, A.M. Cegarra, Group-theoretic Algebraic Models for Homotopy Types, J. Pure Appl. Algebra **75** (1991), 195–235.
[103] T. Porter, n-types of simplicial groups and crossed n-cubes, Topology **32** (1993), 5–24.
[104] R. Street, The petit topos of globular sets, J. Pure Appl. Algebra **154** (2000), 299–315.
[105] R. Brown, A new higher homotopy groupoid: the fundamental globular ω-groupoid of a filtered space, Homotopy, Homology and Applications **10** (2008), 327–343.
[106] G. Nan Tie, A Dold-Kan theorem for crossed complexes, J. Pure Appl. Algebra **56** (1989), 177–194.
[107] M.K. Dakin, Kan complexes and multiple groupoid structures, PhD. Thesis, University of Wales, Bangor, 1976.
[108] R. Brown, P.J. Higgins, Sur les complexes croisés, ω-groupoïdes et T-complexes, *C.R. Acad. Sci. Paris Sér. A.* **285** (1977), 997–999.
[109] D. Verity, Complicial sets characterising the simplicial nerves of strict ω-categories, Mem. Amer. Math. Soc. **193** (2008), no. 905, xvi+184 pp.
[110] J.-L. Loday, Spaces with finitely many homotopy groups, J. Pure Appl. Algebra **24** (1982), 179–202.
[111] R. Brown, Computing homotopy types using crossed n-cubes of groups, in: N. Ray and G. Walker, eds., *Adams Memorial Symposium on Algebraic Topology, 1 (Manchester, 1990)*, volume **175** of *London Math. Soc. Lecture Note Ser.*, 187–210, Cambridge Univ. Press, Cambridge, 1992.
[112] G. Ellis, R. Mikhailov, A colimit of classifying spaces, arXiv:0804.3581, 2008.
[113] J.M. Casas, G. Ellis, M. Ladra, T. Pirashvili, Derived functors and the homology of n-types, J. Algebra **256** (2002), 583–598.
[114] R. Brown, K.A. Hardie, K.H. Kamps, T. Porter, A homotopy double groupoid of a Hausdorff space, Theory Appl. Categ. **10** (2002), 71–93 (electronic).
[115] R. Brown, G. Janelidze, A new homotopy double groupoid of a map of spaces, Appl. Categorical Structures **12** (2004), 63–80.
[116] M. Mackaay, R.F. Picken, Holonomy and parallel transport for Abelian gerbes, Adv. Math. **170** (2002), 287–339.
[117] R. Brown, J.F. Glazebrook, Connections, local subgroupoids, and a holonomy Lie groupoid of a line bundle gerbe, Univ. Iagel. Acta Math. **XLI** (2003), 283–296.
[118] N. Andruskiewitsch, S. Natale, Tensor categories attached to double groupoids, Adv. Math. **200** (2006), 539–583.
[119] R. Steiner, Resolutions of spaces by n-cubes of fibrations, J. London Math. Soc. (2) **34** (1986), 169–176.

[120] R. Steiner, Thin fillers in the cubical nerves of omega-categories, Theory Appl. Categ. **16** (2006), 144–173.
[121] P. Gaucher, E. Goubault, Topological deformation of higher dimensional automata, Homology Homotopy Appl. **5** (2003), 39–82.
[122] L. Fajstrup, M. Rauen, E. Goubault, Algebraic topology and concurrency, Theoret. Comput. Sci. **357** (2006), 241–278.
[123] T. Leinster, Higher operads, higher categories, London Mathematical Society Lecture Note Series, **298** Cambridge University Press, Cambridge, 2004, pp. xiv+433.
[124] V.V. Sharko, Functions on manifolds: Algebraic and topological aspects, Translations of mathematical monographs, **131** Amer. Math. Soc., (1993), x+193.
[125] R. Brown, I. Icen, Towards two dimensional holonomy, Adv. Math. **178** (2003), 141–175.
[126] A. Heyworth, C.D. Wensley, Logged rewriting and identities among relators, in: C.M. Campbell, E.F. Robertso, and G.C. Smith, eds., Groups St. Andrews 2001 in Oxford. vol. I, *London Math. Soc. Lecture Note Ser.*, volume 304. Cambridge Univ. Press, Cambridge 2003, pp. 256–276.

Hopf Algebraic Approach to Picard–Vessiot Theory

Katsutoshi Amano, Akira Masuoka and Mitsuhiro Takeuchi

Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan
E-mails: amano@math.tsukuba.ac.jp; akira@math.tsukuba.ac.jp; and takeu@math.tsukuba.ac.jp

Contents

HANDBOOK OF ALGEBRA, VOL. 6
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(08)00204-0

Introduction

Picard–Vessiot (PV) theory is a Galois theory of linear differential equations; see [1–4]. Let us quickly review its basics along the traditional approach. A field K together with a single derivation is called a differential field. Suppose that the characteristic ch K of K is zero, and the subfield k, say, of constants in K is algebraically closed. Given a linear homogeneous differential equation

$$\partial y = y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y = 0 \quad (a_i \in K)$$

with coefficients in K, there is a unique (up to isomorphism) minimal splitting extension L/K of differential fields. By definition, L has the same constant field k as K, includes an n-dimensional k-subspace of solutions of $\partial y = 0$, and is generated by those solutions over K. Such an extension L/K is called a *PV extension*, just as the minimal splitting field of a (separable) algebraic equation is called as a Galois extension. The group $\mathrm{Aut}_{\mathrm{dif}}(L/K)$ of differential automorphisms of L/K forms a linear algebraic group as a closed subgroup of $GL_n(k)$, the group of linear automorphisms of the solution space. Then, $M \mapsto \mathrm{Aut}_{\mathrm{dif}}(L/M)$ gives a 1–1 correspondence (*Galois correspondence*) between the intermediate differential fields in L/K and the closed subgroups in $\mathrm{Aut}_{\mathrm{dif}}(L/K)$. The extension L/K is said to be *Liouville* if there is a sequence $K = L_0 \subset L_1 \subset \cdots \subset L_r = L$ of differential fields in which each L_i/L_{i-1} is a finite extension or a simple extension $L_i = L_{i-1}(x_i)$, where x_i is an integral, or the exponential of an integral, of some element in L_{i-1}. This condition holds if and only if the connected component in $\mathrm{Aut}_{\mathrm{dif}}(L/K)$ containing 1 is solvable.

Parallel results were obtained by replacing differential fields with partial differential fields [5], Δ-fields [6], and fields with higher derivation [7]. On contrary, the Franke [8] in his attempt of difference PV theory encountered the following difficulty: it often happens that a difference equation with coefficients in a difference field, say K, cannot split in any difference field extension L/K with the same constants. Here, a difference field is a field together with a single automorphism. More than 30 years later, van der Put and Singer [9] overcame the difficulty stated before by extending the framework from difference fields to simple total difference rings and thereby succeeded in establishing the desired difference PV theory.

A Hopf algebraic approach to the PV theory was proposed by Takeuchi [10] in the context of C-ferential fields, where C is a cocommutative coalgebra with specified grouplike 1_C. With C appropriately chosen, differential fields, partial differential fields, Δ-fields, and fields with higher derivation all turn into examples of C-ferential fields. According to this approach, to each PV extension L/K (see below for definition) is associated a commutative Hopf algebra H over the constant field k, which is called the *PV Hopf algebra* for L/K. Quite generally, commutative Hopf algebras are in 1–1 correspondence with affine group schemes. The affine group scheme $\mathbf{G}(L/K) := \mathrm{Sp}\, H$ corresponding to H is called the *PV group scheme* for L/K. This use of commutative Hopf algebras (or affine group schemes) instead of linear algebraic groups makes it possible to avoid certain dispensable assumptions on k and L/K. For example, the Galois correspondence is formulated as a 1–1 correspondence between the intermediate C-ferential fields in L/K and the affine closed subgroup schemes in

$\mathbf{G}(L/K)$, for which k may be arbitrary, and L/K may not be finitely generated. Of course, this result reduces to the Galois correspondence cited in the first paragraph in the special situation there.

Amano and Masuoka [11] pursued Takeuchi's approach, to include difference PV theory as well, working with AS D-module algebras. Here, D denotes a split cocommutative Hopf algebra over a fixed ground field R; it is thus of the form $D^1\#RG$, where G denotes the group of grouplikes in D, and D^1 is the irreducible component containing 1. A D-module algebra is said to be *AS*, if it is Artinian as a ring, and is simple as a D-module algebra; such a D-module algebra K is necessarily the product $K_1 \times \cdots \times K_1$ of finitely many copies of a $D(G_1)$-module field K_1, where $G_1 \subset G$ is the subgroup of the stabilizers of a primitive idempotent in K and $D(G_1) = D^1\#RG_1$. If D^1 is trivial and G is the free group on one generator, then an AS D-module algebra is precisely a simple total ring as in [9]. Roughly speaking, the special situation where G is trivial corresponds to the C-ferential context, but Amano and Masuoka [11] imposes throughout the assumption that D^1 is of Birkhoff–Witt type, which is assumed in [10] only to a small extent. The article [11] does not include [10] as a special case but clarifies how the argument in the C-ferential context can be used to deal with the context of AS D-module algebras. We remark that prior to Amano and Masuoka [11], André [12] gave a unified approach to differential and difference PV theories from the viewpoint of noncommutative differential geometry. Liouville extensions are not discussed in [10–12]. Amano [13] characterized Liouville PV extensions in the context of AS D-module algebras, introducing the notion of Liouville affine group schemes.

The present article brings the three papers [10, 11, 13] together and is divided into the three parts:

Part I: PV theory in the differential context
Part II: PV theory in the C-ferential context
Part III: Unified PV theory

To avoid too much technical argument, we discuss in Part I everything in the differential context. We then explain how the results are generalized in the C-ferential context (Part II) and in the context of AS D-module algebras (Part III). Part I is fairly selfcontained, with almost full proofs given. Only minimal knowledge of Hopf algebras and affine group schemes is assumed there. The language of tensor categories will be used only in Parts II and III. In what follows in the Introduction, we will work in the differential context for simplicity.

One more feature of our Hopf algebraic approach is first to formalize the notion of a PV extension intrinsically (see below), and then to characterize it as a minimal splitting extension of a linear differential equation, just as Emil Artin formalized a Galois extension as a separable normal extension, subsequently characterizing it as a minimal splitting extension of a separable algebraic equation. To explain our formalism of PV extensions, let L/K be an extension of differential fields with the same constant field k. We define L/K to be a *PV extension* (Definition 1.8) if there is a (necessarily unique) differential subring $A \subset L$ including K such that (a) L equals the quotient field $Q(A)$ of A and (b) the constants $H := (A \otimes_K A)_0$ in $A \otimes_K A$ generate the left (or right) A-module $A \otimes_K A$. It is then proved that the natural map

$\mu : A \otimes_k H \to A \otimes_K A$, $\mu(a \otimes h) = (a \otimes 1) \cdot h$ is an isomorphism so that H is a k-form of the left A-module $A \otimes_K A$. The natural A-coring structure on $A \otimes_K A$ induces on H a structure of a commutative k-Hopf algebra; this H is precisely the PV Hopf algebra cited before. Moreover, A/K turns into a right H-Galois extension [14, Definition 8.1.1] for which the canonical isomorphism $A \otimes_K A \xrightarrow{\simeq} A \otimes_k H$ required by the definition coincides with the inverse of μ. Thus, our approach is based on an idea of Hopf–Galois theory.

According to our definition, PV extensions need not be finitely generated. We prove that L/K is finitely generated PV if and only if L is a minimal splitting field of some linear differential equation (or a system of such equations) over K (Theorem 3.11). Giving a system of linear differential equations over K is equivalent to giving a differential K-module of finite K-dimension. Giving an appropriate definition of a minimal splitting field of a possibly infinite-dimensional differential K-module, we extend the last result, characterizing PV extensions in general (Proposition 3.17).

There is another known approach to the PV theory proposed by Deligne [15] called the Tannaka approach. It is based on the result that a finite-dimensional differential K-module V *generates* the neutral Tannaka category, which corresponds to the PV group scheme $\mathbf{G}(L/K)$, where L is a minimal splitting field for V. We reformulate this result in the C-ferential context (Theorem 8.11) and give an elementary proof based on a fundamental result of Hopf–Galois theory. In this way, our approach links the traditional approach and the Tannaka approach.

Throughout this article, rings, fields, and algebras are supposed to be commutative unless otherwise stated.

Part I: PV theory in the differential context

1. Formalism of Picard–Vessiot extensions

DEFINITION 1.1. A *differential ring* (or *field*) is a ring (or field) A with a given derivation $' : A \to A$, which is by definition is an additive operation such that $(ab)' = a'b + ab'$, and then necessarily $1' = 0$. The subring (or subfield) of constants in a differential ring (or field) A is defined and denoted by

$$A_0 = \{a \in A \mid a' = 0\}.$$

The adjective "*differential*" will mean $'$-stable or $'$-preserving. Thus, a *differential subring* is a $'$-stable subring. A *differential ring map* is a $'$-preserving ring map. Given such a map $i : K \to A$, the pair (A, i) is called a *differential K-algebra*.

LEMMA 1.2 (Extension lemma). *Let A be a differential ring and let $S \subset A$ be a multiplicative subset. Then, the localization $S^{-1}A$ by S has a unique differential A-algebra structure. The derivation is*

$$\left(\frac{a}{s}\right)' = \frac{a's - as'}{s^2} \quad (a \in A,\ s \in S).$$

In particular, if A is an integral domain, its quotient field, which will be denoted by $Q(A)$, has a unique differential A-algebra structure.

PROPOSITION 1.3 (see [16, Lemma 10.2.2]). *Let K be a differential field, and let A be a differential K-algebra. Then, K and A_0 are K_0-linearly disjoint, that is, the multiplication $K \otimes_{K_0} A_0 \to A$, $x \otimes a \mapsto xa$ is injective.*

We will denote iterated derivations by $x^{(1)} = x'$, $x^{(2)} = x''$, $\ldots$, $x^{(i+1)} = (x^{(i)})'$.

PROPOSITION 1.4 (Wronskian criterion, [1, Theorem 3.7]). *Finitely many elements $x_1, x_2, \ldots, x_n$ in a differential field K are K_0-linearly independent if and only if the determinant*

$$w(x_1, x_2, \ldots, x_n) = \begin{vmatrix} x_1 & x_2 & \cdots & x_n \\ x_1' & x_2' & \cdots & x_n' \\ \vdots & \vdots & & \vdots \\ x_1^{(n-1)} & x_2^{(n-1)} & \cdots & x_n^{(n-1)} \end{vmatrix},$$

called the Wronskian*, is nonzero.*

Let K be a differential field, and let $A \neq 0$ be a differential K-algebra. The K-algebra $A \otimes_K A$ is regarded as differential with respect to $(a \otimes b)' = a' \otimes b + a \otimes b'$. Similarly, $A \otimes_K A \otimes_K A$ and $A \otimes_K A \otimes_K A \otimes_K A$ are regarded as differential K-algebras. The A-bimodule

$$\mathcal{C} = \mathcal{C}_{A/K} := A \otimes_K A \tag{1.5}$$

together with the A-bimodule maps

$$\begin{aligned} \boldsymbol{\Delta} : \mathcal{C} = A \otimes_K A &\to A \otimes_K A \otimes_K A = \mathcal{C} \otimes_A \mathcal{C}, \\ \boldsymbol{\Delta}(a \otimes_K b) &= a \otimes_K 1 \otimes_K b, \\ \boldsymbol{\varepsilon} : \mathcal{C} = A \otimes_K A &\to A, \quad \boldsymbol{\varepsilon}(a \otimes_K b) = ab \end{aligned} \tag{1.6}$$

forms an A-coring; this means that $\boldsymbol{\Delta}$ and $\boldsymbol{\varepsilon}$ satisfy the coalgebra axioms. Note that both $\boldsymbol{\Delta}$ and $\boldsymbol{\varepsilon}$ are maps of differential K-algebras, which therefore induce K_0-algebra maps $\boldsymbol{\Delta}_0$ and $\boldsymbol{\varepsilon}_0$ on the constants. Let

$$H = (A \otimes_K A)_0$$

denote the K_0-subalgebra of constants in $A \otimes_K A$.

PROPOSITION 1.7.

(1) *We have three natural K_0-algebra maps:*

$$(i) \quad K_0 \to A_0, \quad (ii) \quad H \otimes_{K_0} H \to (A \otimes_K A \otimes_K A)_0, \text{ and}$$
$$(iii) \quad H \otimes_{K_0} H \otimes_{K_0} H \to (A \otimes_K A \otimes_K A \otimes_K A)_0.$$

(2) *If both the maps* (*i*) *and* (*ii*) *are isomorphic,* $\boldsymbol{\Delta}_0$ *and* $\boldsymbol{\varepsilon}_0$ *can be transfered to* K_0*-algebra maps,*

$$\Delta : H \to H \otimes_{K_0} H, \quad \varepsilon : H \to K_0.$$

(3) *If in addition, the map* (*iii*) *is injective, then* (H, Δ, ε) *forms a commutative* K_0*-Hopf algebra. Its antipode* $S : H \to H$ *is induced by the flip*

$$\boldsymbol{\tau} : A \otimes_K A \to A \otimes_K A, \quad \boldsymbol{\tau}(a \otimes b) = b \otimes a.$$

In summary, $(H, \Delta, \varepsilon, S)$ *is a kind of* K_0*-form of* $(\mathcal{C}_{A/K}, \boldsymbol{\Delta}, \boldsymbol{\varepsilon}, \boldsymbol{\tau})$.[1]

PROOF. To see (3), draw first the commutative diagrams, which represent that $\mathcal{C}_{A/K}$ is an A-coring, and take constants $(\ )_0$. What is obtained are the commutative diagrams, which show that H is a K_0-coalgebra. To see then that $S = \boldsymbol{\tau}_0$ is a left, say, convolution-inverse of the identity map id_H, take constants in the two coinciding maps

$$A \otimes_K A \xrightarrow{\boldsymbol{\Delta}} A \otimes_K A \otimes_K A \xrightarrow{\boldsymbol{\tau}\otimes\mathrm{id}} A \otimes_K A \otimes_K A \xrightarrow{\mathrm{id}\otimes\boldsymbol{\varepsilon}} A \otimes_K A,$$

$$A \otimes_K A \xrightarrow{\boldsymbol{\varepsilon}} A \xrightarrow{1\otimes} A \otimes_K A. \qquad \square$$

DEFINITION 1.8. An inclusion $L \supset K$ of differential fields will be called an extension and will be denoted by L/K. Such an extension L/K is called a *PV (Picard–Vessiot) extension* if the following are satisfied:

(i) $L_0 = K_0$.
(ii) There exists an intermediate differential ring $L \supset A \supset K$ such that
  (a) $L = Q(A)$, the quotient field of A,
  (b) $H := (A \otimes_K A)_0$ generates the left A-module $A \otimes_K A$, that is, $(A \otimes_K K)\cdot H = A \otimes_K A$; this is equivalent to the statement that H generates the right A-module $A \otimes_K A$, as is seen by applying the flip $\boldsymbol{\tau}$.

Such an A as above is called a *principal differential ring* for L/K; it will be seen to be unique in 1.11. When L/K is PV, we will write

$$k := L_0 = K_0.$$

LEMMA 1.9. *Let* A *be a principal differential ring for a PV extension* L/K. *Let* $H := (A \otimes_K A)_0$.

(1) *The natural maps* (*i*)–(*iii*) *of 1.7 are all isomorphisms so that* H *forms a commutative* k*-Hopf algebra.*
(2) *We have a unique* K*-linear map*

$$\theta : A \to A \otimes_k H, \quad \theta(a) = \sum a_0 \otimes a_1$$

[1] This means that tensoring up $(H, \Delta, \varepsilon, S)$ with K over K_0 gives (up to isomorphism) $(\mathcal{C}_{A/K}, \boldsymbol{\Delta}, \boldsymbol{\varepsilon}, \boldsymbol{\tau})$

such that for every $a \in A$,

$$1 \otimes_K a = \sum (a_0 \otimes_K 1) \cdot a_1 \quad \textit{in} \quad A \otimes_K A.$$

(3) *We have*

(i) *The θ above is a k-algebra map, which makes A a right H-comodule; (A, θ) is thus a right H-comodule algebra.*

(ii) *The k-subalgebra $A^{\mathrm{co}\,H} = \{a \in A \mid \theta(a) = a \otimes 1\}$ of H-coinvariants in A coincides with K.*

(iii) *The left A-linearization of θ*

$$_A\theta : A \otimes_K A \to A \otimes_k H, \quad {}_A\theta(a \otimes_K b) = a\theta(b)$$

is an isomorphism.

In summary, A/K is a right H-Galois extension [14, Definition 8.1.1] in the sense of Kreimer and Takeuchi [17].

Proof. (1) We have $K_0 = A_0$, since $k = K_0 \subset A_0 \subset L_0 = k$. It then suffices to prove that the maps (ii) and (iii) of 1.7 are isomorphic. By applying 1.3 to the differential L-algebra $L \otimes_K A$ together with $(L \otimes_K A)_0 \supset H$, we see that the natural map $L \otimes_k H \to L \otimes_K A$ is injective. Hence,

$$\mu : A \otimes_k H \to A \otimes_K A, \quad \mu(a \otimes_k h) = (a \otimes_K 1) \cdot h, \tag{1.10}$$

being surjective by assumption (ii)(b) of 1.8, is isomorphic. As an iteration of μ, we have

$$A \otimes_k H \otimes_k H \xrightarrow{\simeq} A \otimes_K A \otimes_k H \xrightarrow{\simeq} A \otimes_K A \otimes_K A,$$

and similarly $A \otimes_k H \otimes_k H \otimes_k H \xrightarrow{\simeq} A \otimes_K A \otimes_K A \otimes_K A$. By taking $(\ )_0$ in these isomorphisms of differential K-algebras, the desired result follows.

(2) Let θ be the composite of $1\otimes : A \to A \otimes_K A$, $a \mapsto 1 \otimes a$ with μ^{-1}; this is the unique map as required earlier. To see that (A, θ) is a right H-comodule, transfer the commutative diagrams, which show that the map $1\otimes : A \to A \otimes_A \mathcal{C}_{A/K}$ above satisfies the axioms for right comodules.

(3) Note that ${}_A\theta$ is an isomorphism with inverse μ. □

Lemma 1.11. *A principal differential ring for a PV extension L/K is unique.*

Proof. Suppose we have two such rings, say A_1 and A_2. Since A_1A_2 is principal as well, we may suppose $A_1 \subset A_2$. Then, the corresponding $H_i := (A_i \otimes_K A_i)_0$ gives an extension $H_1 \subset H_2$ of commutative Hopf algebras, which is necessarily faithfully flat [18, Theorm 3.1]. The isomorphisms $L \otimes_k H_i \simeq L \otimes_K A_i$, which arise from the μ in (1.10), commute with the inclusions. It follows that $A_1 \subset A_2$ is faithfully flat. Since $A_2 \subset Q(A_1)$, we must have $A_1 = A_2$. □

In summary, given a PV extension L/K, we have a unique principal differential ring A, which naturally defines a commutative k-Hopf algebra $H = (A \otimes_K A)_0$.

DEFINITION 1.12. This H is called the *PV Hopf algebra* of L/K. We say that $(L/K, A, H)$ is a PV extension, indicating A, H too.

REMARK 1.13. Suppose that L/K is an extension of differential fields with the same constant field k. Let A be a differential intermediate ring in L/K, and let $J \subset (A \otimes_K A)_0$ be a k-subspace. As is seen from the proof of 1.9, if $(A \otimes_K K)\cdot J = A \otimes_K A$, then $J = (A \otimes_K A)_0$. Therefore, if $L = Q(A)$ in addition, $(L/K, A, J)$ is PV.

To give examples of PV extensions, recall from (1.5) that given a field extension L/K, we have an L-coring $\mathcal{C}_{L/K} = L \otimes_K L$. Just as with ordinary coalgebras, an element c in $\mathcal{C}_{L/K}$ is called *primitive* (resp. *grouplike*), if $\boldsymbol{\Delta}(c) = 1 \otimes_L c + c \otimes_L 1$, $\boldsymbol{\varepsilon}(c) = 0$ (resp., $\boldsymbol{\Delta}(c) = c \otimes_L c$, $\boldsymbol{\varepsilon}(c) = 1$).

LEMMA 1.14. *In* $\mathcal{C}_{L/K}$,

(1) *every primitive is of the form* $1 \otimes_K x - x \otimes_K 1$, *where* $x \in L$,
(2) *every grouplike is of the form* $x^{-1} \otimes_K x$, *where* $0 \neq x \in L$.

This is verified directly. One may note that the lemma can be reformulated as the vanishing of the Amitsur cohomology groups $H^1(L/K, \mathbf{G}_a)$, $H^1(L/K, \mathbf{G}_m)$ [19, Chapter 17]. The lemma yields the following.

EXAMPLE 1.15. Let L/K be an extension of differential fields with the same constant field k.

(1) An element $x \in L$ is said to be *primitive*[2] over K, if $x' \in K$ or, symbolically speaking, if x is the integral $\int a$ of some element a in K. This condition is equivalent to

$$l := 1 \otimes_K x - x \otimes_K 1 \in (L \otimes_K L)_0.$$

With x and l as given earlier, construct the subalgebras $A = K[x]$ in L and $H = k[l]$ in $(A \otimes_K A)_0$; note that $A' \subset A$. Since $1 \otimes_K x = x \otimes_K 1 + l$ in $A \otimes_K A$, we see that $(K(x)/K, A, H)$ is PV. Moreover, l is a primitive in H, and the map $\theta : A \to A \otimes_k H$ is given by $\theta(x) = x \otimes 1 + 1 \otimes l$. Any PV Hopf algebra that is generated by a primitive arises in this way.

(2) A nonzero element x in L is said to be *exponential* over K, if $x'x^{-1} \in K$ or, symbolically speaking, if x is the exponential $\exp \int a$ of some integral $\int a$ $(a \in K)$. This condition is equivalent to

$$g := x^{-1} \otimes_K x \in (L \otimes_K L)_0.$$

[2] Warning! One should distinguish the word "primitive" in this context of differential calculus, which comes from "primitive functions," from the (more familiar) usage of the same word for primitive elements in a Hopf algebra or a coring, especially because these two distinct (though related) concepts can appear at the same time. Recall this warning at Definition 5.3, Theorem 5.8 and just above Theorem 8.4.

With x and g as given earlier, construct $A = K[x, x^{-1}]$ in L and $H = k[g, g^{-1}]$ in $(A \otimes_K A)_0$; note that $A' \subset A$. Since $1 \otimes_K x^{\pm 1} = (x^{\pm 1} \otimes_K 1) \cdot g^{\pm 1}$ in $A \otimes_K A$, we see that $(K(x)/K, A, H)$ is PV. Moreover, g is a grouplike in H, and the map θ is given by $\theta(x^{\pm 1}) = x^{\pm 1} \otimes g^{\pm 1}$. Any PV Hopf algebra that is generated by a grouplike and its inverse arises in this way.

To generalize the last example, we will work in the matrix algebra $M_n(L \otimes_K L)$, which is a noncommutative differential ring with entrywise derivation.

EXAMPLE 1.16. Let L/K and k be as given earlier. Fix an integer $n > 0$. Suppose that an invertible matrix $X = (x_{ij})$ in $GL_n(L)$ is GL_n*-primitive* in the sense that $X'X^{-1} \in M_n(K)$, that is, $X' = CX$ for some $C \in M_n(K)$. Then, $(X^{-1})' = -X^{-1}C$. Hence the matrices in $M_n(L \otimes_K L)$

$$B := (X^{-1} \otimes_K 1)(1 \otimes_K X), \quad B^{-1} := (1 \otimes_K X^{-1})(X \otimes_K 1),$$

which are mutual inverses, are constants. Let $B = (b_{ij})$, and construct $A = K[x_{ij}, \frac{1}{|X|}]$ in L, and $H = k[b_{ij}, \frac{1}{|B|}]$ in $(A \otimes_K A)_0$; note that $A' \subset A$. It follows from $1 \otimes_K X = (X \otimes_K 1) \cdot B$ that $(K(x_{ij})/K, A, H)$ is *PV*. Moreover,

$$\Delta(B) = (B \otimes_k 1)(1 \otimes_k B), \quad \varepsilon(B) = I, \quad S(B) = B^{-1},$$
$$\theta(X) = (X \otimes_k 1)(1 \otimes_k B),$$

where the last identity, for example, should be read as $\theta(x_{ij}) = \sum_r x_{ir} \otimes_k b_{rj}$. We will see in Section 3 that any finitely generated PV extension arises in this way.

2. Galois correspondence

To establish a Galois correspondence for a PV extension, let L/K be an extension of differential fields. The L-coring $\mathcal{C}_{L/K} = L \otimes_K L$ with its natural derivation may be called a *differential L-coring*. If M is an intermediate differential field in L/K, then $L \otimes_M L$ is a quotient differential L-coring of $\mathcal{C}_{L/K}$, or in other words the kernel, say $\mathfrak{a}_M$, of the natural surjection $L \otimes_K L \twoheadrightarrow L \otimes_M L$ is a differential coideal in $\mathcal{C}_{L/K}$.

LEMMA 2.1. *$M \mapsto \mathfrak{a}_M$ gives a bijection from the set of all intermediate differential fields M in L/K onto the set of all differential coideals $\mathfrak{a}$ in $\mathcal{C}_{L/K}$. The inverse assigns to each $\mathfrak{a}$*

$$M_{\mathfrak{a}} := \{x \in L \mid 1 \otimes_K x - x \otimes_K 1 \in \mathfrak{a}\}. \tag{2.2}$$

This follows immediately from the correspondence due to Sweedler [20, Theorem 2.1] in the nondifferential situation.

PROPOSITION 2.3. *Let $(L/K, A, H)$ be a PV extension.*

(1) *There is a 1–1 correspondence between the ideals* $\mathfrak{a}_0$ *in* H *and the differential ideals* $\mathfrak{a}$ *in* $L \otimes_K L$*, given by*

$$\mathfrak{a}_0 = \mathfrak{a} \cap H, \quad \mathfrak{a} = \mathfrak{a}_0 \cdot (L \otimes_K L).$$

(2) *Suppose* $\mathfrak{a}_0 \leftrightarrow \mathfrak{a}$ *under this correspondence. Then* $\mathfrak{a}_0$ *is a Hopf ideal if and only if* $\mathfrak{a}$ *is a differential coideal.*

PROOF. (1) Let $\mathcal{I}$ (resp., $\mathcal{I}_{\mathrm{dif}}$) indicate the set of ordinary (resp., differential) ideals. It follows from the isomorphism $L \otimes_A \mu : L \otimes_k H \xrightarrow{\simeq} L \otimes_K A$ of differential L-algebras that $\mathfrak{a}_0 \mapsto \mathfrak{a}_0 \cdot (L \otimes_K A)$ gives a bijection $\mathcal{I}(H) \xrightarrow{\simeq} \mathcal{I}_{\mathrm{dif}}(L \otimes_K A)$. This means that $\mathfrak{a}_0 \mapsto \mathfrak{a}_0 \cdot (A \otimes_K A)$ gives an injection $\mathcal{I}(H) \hookrightarrow \mathcal{I}_{\mathrm{dif}}(A \otimes_K A)$ with image $\mathcal{I}_{\mathrm{dif}}(L \otimes_K A)$, which is contained in $\mathcal{I}_{\mathrm{dif}}(A \otimes_K A)$, since $L \otimes_K A$ is a localization of $A \otimes_K A$. By symmetry, the image above coincides with $\mathcal{I}_{\mathrm{dif}}(A \otimes_K L)$ and hence with the intersection $\mathcal{I}_{\mathrm{dif}}(L \otimes_K A) \cap \mathcal{I}_{\mathrm{dif}}(A \otimes_K L) = \mathcal{I}_{\mathrm{dif}}(L \otimes_K L)$.

(2) Similarly, we have a bijection $\mathcal{I}(H \otimes_k H) \simeq \mathcal{I}_{\mathrm{dif}}(L \otimes_K L \otimes_K L)$ under which $\mathfrak{b}_0 := H \otimes_k \mathfrak{a}_0 + \mathfrak{a}_0 \otimes_k H \leftrightarrow \mathfrak{b} := L \otimes_K \mathfrak{a} + \mathfrak{a} \otimes_K L$ when $\mathfrak{a}_0 \leftrightarrow \mathfrak{a}$ under $\mathcal{I}(H) \simeq \mathcal{I}_{\mathrm{dif}}(L \otimes_K L)$. Moreover, $\Delta(\mathfrak{a}_0) \subset \mathfrak{b}_0$, $\varepsilon(\mathfrak{a}_0) = 0$ if and only if $\mathbf{\Delta}(\mathfrak{a}) \subset \mathfrak{b}$, $\varepsilon(\mathfrak{a}) = 0$; this proves (2). Here, recall that every biideal in a commutative Hopf algebra is stable under the antipode, as is seen in our situation from the fact that every coideal in $L \otimes_K L$ is stable under $\boldsymbol{\tau}$. □

DEFINITION 2.4. A differential ring $A \neq 0$ is said to be *simple* if it includes no nontrivial differential ideal. In this case, A_0 is a field since every constant generates a differential ideal.

PROPOSITION 2.5. *The principal differential ring* A *for a PV extension* L/K *is simple.*

PROOF. Let $0 \neq \mathfrak{a} \subset A$ be a differential ideal. Suppose $\mathfrak{a}_0 \leftrightarrow L \otimes_K \mathfrak{a}$ under $\mathcal{I}(H) \simeq \mathcal{I}_{\mathrm{dif}}(L \otimes_K A)$. Then, $\mathfrak{a}_0 \cdot (L \otimes_K L) = L \otimes_K \mathfrak{a} L = L \otimes_K L$. By 2.3 (1), $\mathfrak{a}_0 = H$, whence $\mathfrak{a} = A$. □

THEOREM 2.6 (Galois correspondence I). *Let* $(L/K, A, H)$ *be a PV extension.*

(1) *There is a 1–1 correspondence between the Hopf ideals* I *in* H *and the intermediate differential fields* M *in* L/K*, given by*

$$\begin{aligned} M &= \{x \in L \mid 1 \otimes_K x - x \otimes_K 1 \in I \cdot (L \otimes_K L)\} \\ I &= H \cap \mathrm{Ker}(L \otimes_K L \twoheadrightarrow L \otimes_M L). \end{aligned} \tag{2.7}$$

(2) *If* $I \leftrightarrow M$ *under the correspondence,* $(L/M, AM, H/I)$ *is a PV extension.*

PROOF. (1) This follows from 2.1 and 2.3. (2) We see that $AM \cdot (H/I) = AM \otimes_M AM$, by considering the image of $A \otimes_k H \xrightarrow[\mu]{\simeq} A \otimes_K A \to L \otimes_M L$. □

In general, let H be a commutative Hopf algebra over a field k. Given a Hopf subalgebra $J \subset H$, the ideal HJ^+ generated by $J^+ = \mathrm{Ker}(\varepsilon : J \to k)$ is a normal Hopf ideal in H; recall that a Hopf ideal I in H is called *normal* if $\sum a_1 S(a_3) \otimes a_2 \in H \otimes_k I$ for all $a \in I$. It is known that $J \mapsto HJ^+$ gives a bijection from the set of all Hopf subalgebras in H onto the set of all normal Hopf ideals in H; see [18, Theorem 4.3].

Suppose $I \leftrightarrow M$ under the correspondence in 2.6. It has been proved that the Hopf ideal I is normal if and only if M/K is PV. In other words, we have the following.

THEOREM 2.8 (Galois correspondence II). *Let $(L/K, A, H)$ be a PV extension. There is a 1–1 correspondence between the Hopf subalgebras J in H and the intermediate PV extensions M/K in L/K, in which to each M/K its PV Hopf algebra is assigned.*

PROOF. Fix a normal Hopf ideal I in H. Given a right H-comodule V with structure map $\lambda : V \to V \otimes_k H$, define

$$V_1 := \{v \in V \mid \lambda(v) - v \otimes 1 \in H \otimes_k I\}.$$

When $V = H$, H_1 is precisely the Hopf subalgebra corresponding to I [18, Corollary 3.10]. In general, V_1 is the largest H_1-comodule in V or $V_1 = \lambda^{-1}(V \otimes_k H_1)$.

Since $\theta(A_1) \subset A_1 \otimes_k H_1$, we remark that the isomorphism ${}_A\theta : A \otimes_K A \xrightarrow{\simeq} A \otimes_k H$ maps $A_1 \otimes_K A_1$ into $A_1 \otimes_k H_1$. Regard $A \otimes_K A$ as a right H-comodule with respect to the structure arising from the right factor. Since the inclusion $H \hookrightarrow A \otimes_K A$ is right H-colinear, $H_1 \subset A \otimes_K A_1$. By applying the antipode S and the flip $\boldsymbol{\tau}$, we have $H_1 \subset A_1 \otimes_K A$, whence $H_1 \subset A_1 \otimes_K A_1$. This together with the last remark implies that the isomorphism $\mu : A \otimes_k H \xrightarrow{\simeq} A \otimes_K A$ restricts to

$$A_1 \otimes_k H_1 \xrightarrow{\simeq} A_1 \otimes_K A_1. \tag{2.9}$$

Hence, $(L_1/K, A_1, H_1)$ is PV, where $L_1 = Q(A_1)$; note that $A_1' \subset A_1$, $L_1' \subset L_1$. To see the correspondence $I \leftrightarrow L_1$, note from (2.9) that in $A_1 \otimes_K A_1$, $(A_1 \otimes_K K)\cdot H_1^+$ equals the ideal generated by $1 \otimes_K a - a \otimes_K 1$ $(a \in A_1)$, whence in $L \otimes_K L$,

$$I\cdot(L \otimes_K L) = H_1^+\cdot(L \otimes_K L) = \mathrm{Ker}(L \otimes_K L \twoheadrightarrow L \otimes_{L_1} L).$$

Conversely, let $(M/K, B, J)$ be an intermediate PV extension. Obviously, $B \subset A$, $J \subset H$. We see that ${}_B\theta : B \otimes_K B \xrightarrow{\simeq} B \otimes_k J$ induces $A \otimes_K B \xrightarrow{\simeq} A \otimes_k J$, which extends to ${}_A\theta : A \otimes_K A \xrightarrow{\simeq} A \otimes_k H$. It follows that $B = \theta^{-1}(A \otimes_k J)$. The argument in the preceding paragraph shows that $HJ^+ \leftrightarrow M$. □

Let $(L/K, A, H)$ be a PV extension. We denote by

$$\mathbf{G}(L/K) = \mathrm{Sp}\, H$$

the affine k-group scheme corresponding to H; let us call this the *PV group scheme* of L/K. By definition, this is the functor, which associates to each commutative k-algebra R the group $\mathrm{Alg}(H, R)$ of all k-algebra maps $H \to R$; see [19, Section 1.4].

Let $\mathbf{Aut}_{\mathrm{dif}}(A/K)$ denote the functor, which associates to each R the group $\mathrm{Aut}_{\mathrm{dif}}(A \otimes_k R/K \otimes_k R)$ of differential $K \otimes_k R$-algebra automorphisms of $A \otimes_k R$. It has been proved that the k-linear representation $\mathbf{G}(L/K) \to \mathbf{GL}(A)$ arising from $\theta : A \to A \otimes_k H$ gives an isomorphism $\phi : \mathbf{G}(L/K) \xrightarrow{\simeq} \mathbf{Aut}_{\mathrm{dif}}(A/K)$ of group-functors; see [10, Theorem (A.2)]. Consequently, the algebraic group $\mathbf{G}(L/K)(k)$ of the rational points in k is isomorphic to the group $\mathrm{Aut}_{\mathrm{dif}}(A/K)$ of differential K-algebra automorphisms of A, which is isomorphic to the group $\mathrm{Aut}_{\mathrm{dif}}(L/K)$ for L, by 1.11. This last group plays the role of Galois groups in standard PV theory (see 2.12 below).

EXAMPLE 2.10. Let the situation and the notation be as in 1.16, but we suppose $L = K(x_{ij})$. We have a closed embedding of affine k-group schemes,

$$\alpha : \mathbf{G}(L/K) \to \mathbf{GL}_n, \quad g \mapsto g(B) = (g(b_{ij})).$$

If $f \in \mathrm{Aut}_{\mathrm{dif}}(A \otimes_k R/K \otimes_k R)$, then

$$\beta(f) := X^{-1} \cdot f(X) \in GL_n(R),$$

since this equals the image ${}_A f(B)$ of the differential A-algebra map ${}_A f : A \otimes_K A \to A \otimes_k R$, ${}_A f(a \otimes_K b) = a \cdot f(b)$ applied entry-wise to $B = (X^{-1} \otimes_K 1) \cdot (1 \otimes_K X)$. Thus, we have a closed embedding $\beta : \mathbf{Aut}_{\mathrm{dif}}(A/K) \to \mathbf{GL}_n$. Since $\beta(\phi(\mathrm{id}_H)) = {}_A\theta(B) = B = \alpha(\mathrm{id}_H)$, the following diagram of group functors commutes by the Yoneda lemma:

$$\begin{array}{ccc} \mathbf{G}(L/K) & \xrightarrow[\simeq]{\phi} & \mathbf{Aut}_{\mathrm{dif}}(A/K) \\ & {\scriptstyle\alpha}\searrow \quad \swarrow{\scriptstyle\beta} & \\ & \mathbf{GL}_n & \end{array}$$

The Galois correspondence theorems 2.6 and 2.8 are interpreted in terms of group schemes as follows.

THEOREM 2.11. *Let L/K be a PV extension with PV group scheme $\mathbf{G}(L/K)$.*

(1) *For every intermediate differential field M in L/K, the extension L/M is PV. The map $M \mapsto \mathbf{G}(L/M)$ gives a bijection from the set of all intermediate differential fields in L/K onto the set of all closed subgroup schemes in $\mathbf{G}(L/K)$.*
(2) *An intermediate differential field M in L/K is PV over K if and only if $\mathbf{G}(L/M)$ is normal in $\mathbf{G}(L/K)$. In this case, we have $\mathbf{G}(M/K) \simeq \mathbf{G}(L/K)/\mathbf{G}(L/M)$.*

REMARK 2.12. In the situation of 2.10, $\mathrm{Aut}_{\mathrm{dif}}(L/K)$ ($\simeq \mathbf{G}(L/K)(k)$) is naturally a linear algebraic group over k. If k is an algebraically closed field of characteristic zero, we obtain from 2.11 a 1–1 correspondence between the intermediate differential fields in L/K and the closed subgroups in $\mathrm{Aut}_{\mathrm{dif}}(L/K)$, since the latter are in 1–1 correspondence with the closed affine group schemes in $\mathbf{G}(L/K)$. This is precisely the Galois correspondence in standard PV theory.

Lemma 1.9 (3) is interpreted as follows.

PROPOSITION 2.13. *Let* $(L/K, A, H)$ *be a PV extension. The* k*-scheme map*

$$\operatorname{Sp}\theta : \operatorname{Sp} A \times_k \mathbf{G}(L/K) \to \operatorname{Sp} A$$

corresponding to θ *makes* $\operatorname{Sp} A$ *a right* $\mathbf{G}(L/K)$*-torsor over* $\operatorname{Sp} K$.

3. Splitting fields

Let K be a differential field.

DEFINITION 3.1. A *differential* K*-module* is a K-module V together with an additive operation $' : V \to V$ such that $(av)' = a'v + av'$ $(a \in K, v \in V)$. For such V, we write as before,

$$V_0 := \{v \in V \mid v' = 0\}.$$

This is a K_0-subspace of V.

Proposition 1.3 can be directly generalized as follows.

PROPOSITION 3.2. *For a differential* K*-module* V*, the natural map* $K \otimes_{K_0} V_0 \to V$*,* $x \otimes v \mapsto xv$ *is injective.*

LEMMA 3.3. *Let* V, W *be differential* K*-modules. The* K*-module* $\operatorname{Hom}_K(V, W)$ *of all* K*-linear maps* $V \to W$ *becomes a differential* K*-module with respect to*

$$f'(v) := f(v)' - f(v') \quad (f \in \operatorname{Hom}_K(V, W),\ v \in V)$$

in which $\operatorname{Hom}_K(V, W)_0$ *coincides with the* K_0*-subspace*

$$\operatorname{Hom}_{K\,\mathrm{dif}}(V, W)$$

of differential K*-module maps. By 3.2, we have* $K \otimes_{K_0} \operatorname{Hom}_{K\,\mathrm{dif}}(V, W) \hookrightarrow \operatorname{Hom}_K(V, W)$.

To give a variation, let L/K be an extension of differential fields, and suppose that W is a differential L-module. Then, $\operatorname{Hom}_K(V, W)$ is a differential L-module so that we have $L \otimes_{L_0} \operatorname{Hom}_{K\,\mathrm{dif}}(V, W) \hookrightarrow \operatorname{Hom}_K(V, W)$. Important is the case when $W = L$. We then write

$$L^V := \operatorname{Hom}_{K\,\mathrm{dif}}(V, L).$$

If the K-dimension $\dim_K V = n < \infty$, then the L_0-dimension $\dim_{L_0} L^V \leq n$, since we have

$$L \otimes_{L_0} L^V \hookrightarrow \operatorname{Hom}_K(V, L). \tag{3.4}$$

EXAMPLE 3.5. A differential K-module V is said to be *cyclic* if it is generated by one element, say y; this means that V is K-spanned by $y, y', y'', \ldots$. If $\dim_K V < \infty$, such a pair (V, y) uniquely arises from a monic homogeneous linear differential equation with coefficients in K,

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y = 0 \quad (a_i \in K). \tag{3.6}$$

We recover V as $V = Ky \oplus Ky' \oplus \cdots \oplus Ky^{(n-1)}$ with relation (3.6). By assigning the value $f(y)$ to each $f \in L^V$, we can identify L^V with the L_0-space of the solutions of (3.6) in L, whose dimension is at most n.

EXAMPLE 3.7. The differential K-module $V = Ky_1 \oplus \cdots \oplus Ky_n$ of finite K-dimension, together with an ordered K-basis $y_1, \ldots, y_n$, uniquely arises from a homogeneous linear differential system with coefficients in K,

$$\begin{pmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad (a_{ij} \in K). \tag{3.8}$$

Here, (3.8) is regarded as giving the defining relations for V. By assigning the vector ${}^t(f(y_1), \ldots, f(y_n))$ to each $f \in L^V$, we can identify L^V with the L_0-space of solutions of (3.8) in L^n, whose dimension is at most n. As is well known, the differential equation (3.6) can be treated as a differential system with coefficient matrix

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ & 0 & 1 & \ddots & \vdots \\ & & \ddots & \ddots & 0 \\ & & & 0 & 1 \\ -a_{n-1} & \cdots & \cdots & -a_2 & -a_1 \end{pmatrix}.$$

Let L/K be an extension of differential fields. Given elements $u_1, \ldots, u_r \in L$, let $K\langle u_1, \ldots, u_r \rangle$ denote the smallest intermediate differential field in L/K containing $u_1, \ldots, u_r$. Given a differential K-module V, let $K\langle V \rangle$ denote the smallest intermediate differential field in L/K including all $f(V)$, where $f \in L^V$. We have $K\langle V \rangle^V = L^V$. In 3.7, $K\langle V \rangle$ equals the smallest intermediate differential field in L/K that contains all entries in the solutions of (3.8) in L^n.

LEMMA-DEFINITION 3.9. Let L/K be as stated earlier. Let V be a differential K-module with $\dim_K V = n < \infty$. Then the following are equivalent:

(i) $L \otimes_K V \simeq L^n$ as differential L-modules;
(ii) The natural injection $L \otimes_{L_0} L^V \hookrightarrow \mathrm{Hom}_K(V, L)$ is bijective;
(iii) $\dim_{L_0} L^V = n$.

If these hold, L is called a *splitting field* for V. If $L = K\langle V\rangle$ in addition, L is called a *minimal splitting field* for V. Note that (i) implies $K\langle V\rangle \otimes_K V \simeq K\langle V\rangle^n$, whence $K\langle V\rangle$ is a splitting field for V if L is.

PROOF. (ii) $\Leftrightarrow$ (iii) is obvious. (i) implies $L^V \simeq L_0^n$, (iii). Take the L-dual of the map in (ii). Then we have

$$L \otimes_K V \to \mathrm{Hom}_{L_0}(L^V, L), \quad a \otimes v \mapsto (f \mapsto af(v)). \tag{3.10}$$

Hence, (iii) implies (i). □

THEOREM 3.11 (Characterization). *Let L/K be an extension of differential fields with the same constant field k. Then the following are equivalent:*

(i) *L/K is a PV extension, which is finitely generated in the sense $L = K\langle u_1, \dots, u_n\rangle$ for certain finitely many elements $u_1, \dots, u_n$ in L.*
(ii) *L is a minimal splitting field of some cyclic differential K-module of finite K-dimension.*
(iii) *L is a minimal splitting field of some differential K-module of finite K-dimension.*
(iv) *There exists a GL_n-primitive matrix $X = (x_{ij}) \in GL_n(L)$ for some n (see 1.16), such that $L = K(x_{ij})$.*

PROOF. (ii) $\Rightarrow$ (iii) is obvious. (iv) $\Rightarrow$ (i) follows from 1.16.

(i) $\Rightarrow$ (ii). Suppose that $(L/K, A, H)$ is finitely generated PV. Suitable elements $u_1, \dots, u_n$ as in (i) can be taken so that they form a k-basis of an H-subcomodule of $A = (A, \theta)$. Then, $\theta(u_j) = \sum_r u_r \otimes_k b_{rj}$, where $b_{rj} \in H$. This means that in $L \otimes_K L$, $1 \otimes_K u_j = \sum_r (u_r \otimes_K 1)\cdot b_{rj}$, whence

$$1 \otimes_K u_j^{(i)} = \sum_r \left(u_r^{(i)} \otimes_K 1\right)\cdot b_{rj} \quad (i = 0, 1, \dots). \tag{3.12}$$

Let

$$W = \left(u_j^{(i)}\right) \in GL_n(L), \quad B = (b_{ij}) \in M_n(H), \quad \mathbf{u}^{(n)} = \left(u_1^{(n)}, \dots, u_n^{(n)}\right) \in L^n. \tag{3.13}$$

Note that $W \in GL_n(L)$ by 1.4. By (3.12),

$$1 \otimes_K W = (W \otimes_K 1)\cdot B, \quad 1 \otimes_K \mathbf{u}^{(n)} = (\mathbf{u}^{(n)} \otimes_K 1)\cdot B.$$

It follows that $B \in GL_n(H)$. We have

$$1 \otimes_K \mathbf{u}^{(n)} W^{-1} = (\mathbf{u}^{(n)} \otimes_K 1) B B^{-1} (W \otimes_K 1)^{-1} = \mathbf{u}^{(n)} W^{-1} \otimes_K 1,$$

whence $\mathbf{u}^{(n)} W^{-1} \in K^n$. Hence, there exist $a_i \in K$ such that $\mathbf{u}^{(n)} = (a_1, \dots, a_n) W$. Thus, $u_1, \dots, u_n$ are solutions of $y^{(n)} = a_1 y + \cdots + a_{n-1} y^{(n-1)}$; this implies (ii).

(iii) $\Rightarrow$ (iv). Suppose that L is a minimal splitting field of such a differential system as given by (3.8). Let

$$\mathbf{x}_1, \dots, \mathbf{x}_n \ (\in L^n) \tag{3.14}$$

be k-linearly independent solutions of (3.8). The crucial point is that we see from (3.4) that they are L-linearly independent. One concludes that the matrix $X = (\mathbf{x}_1 \ \dots \ \mathbf{x}_n)$ satisfies Condition (iv). □

COROLLARY 3.15.

(1) *Let* $(L/K, A, H)$ *be a PV extension. The following are equivalent:*
   (i) *L/K is finitely generated as an extension of differential fields (see 3.11(i));*
   (ii) *L/K is finitely generated as an extension of fields;*
   (iii) *A is finitely generated as a K-algebra;*
   (iv) *H is finitely generated as a k-algebra.*

(2) *Any intermediate PV extension M/K in a finitely generated PV extension L/K is finitely generated.*

(3) *If $(L/K, A, H)$ is a finitely generated PV extension, then the transcendence degree* $\mathrm{trans.deg}_K L$ *and the Krull dimensions* $\mathrm{Kdim}\, A$ *and* $\mathrm{Kdim}\, H$ *all coincide.*

PROOF. (1) We see from the last proof that (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii). The remaining equivalence (iii) $\Leftrightarrow$ (iv) follows from $L \otimes_K A \simeq L \otimes_k H$. (2) This follows since a Hopf subalgebra of a commutative finitely generated Hopf algebra is finitely generated [18, Corollary 3.11]. (3) This also follows from $L \otimes_K A \simeq L \otimes_k H$. □

Let us extend the notion of (minimal) splitting fields so that the differential K-module in question can be of infinite dimension.

LEMMA-DEFINITION 3.16. Let L/K be an extension of differential fields. Let V be a differential K-module of possibly infinite K-dimension. The following are equivalent:

(i) There exists a differential L-linear injection $L \otimes_K V \hookrightarrow \prod_{\lambda\in\Lambda} L_\lambda$ into the direct product of copies L_λ of L,
(ii) The natural injection $L \otimes_{L_0} L^V \hookrightarrow \mathrm{Hom}_K(V, L)$ has a dense image, or in other words, its L-dual given by (3.10) is injective.

If these conditions (and $L = K\langle V\rangle$) hold, L is called a *(minimal) splitting field* for V. Note from (i) that $K\langle V\rangle$ is a splitting field for V if L is.

PROOF. Modify the proof of 3.9. □

PROPOSITION 3.17. *Let L/K be an extension of differential fields with the same constant field. Then, L/K is PV if and only if L is a minimal splitting field for a differential*

K-module V, which is locally finite in the sense that it is a sum of differential K-submodules of finite K-dimension.

PROOF. "Only if". Suppose that H is the PV Hopf algebra of L/K. One can write $H = \bigcup_\lambda H_\lambda$ as a directed union of finitely generated Hopf subalgebras H_λ. We have $L = \bigcup_\lambda L_\lambda$, where L_λ/K is a finitely generated PV extension with Hopf algebra H_λ; see 2.8, 3.15. By 3.11, L_λ is a minimal splitting field of a differential K-module, say V_λ, of finite K-dimension, say n_λ. Let $V = \bigoplus_\lambda V_\lambda$; this is locally finite. We have $K\langle V\rangle = \bigcup_\lambda K\langle V_\lambda\rangle = L$ and $L \otimes_K V \simeq \bigoplus_\lambda L^{n_\lambda} \hookrightarrow \prod_\lambda L^{n_\lambda}$.

"If". Suppose that L is a minimal splitting field for V, where $V = \bigcup_\nu V_\nu$ is a directed union of differential K-submodules V_ν of finite K-dimension. Then, $L_\nu := K\langle V_\nu\rangle$ $(\subset L)$ is a minimal splitting field for V_ν such that $L = \bigcup_\nu L_\nu$. By 3.11, we have finitely generated PV extensions, $(L_\nu/K, A_\nu, H_\nu)$. Let $A := \bigcup_\nu A_\nu$; this is a directed union of intermediate differential rings in L/K, with $L = Q(A)$. By taking $\bigcup_\nu$ in $A_\nu \otimes_k H_\nu \simeq A_\nu \otimes_K A_\nu$, we obtain $A \otimes_k H \simeq A \otimes_K A$, where $H = \bigcup_\nu H_\nu$. It follows that (L, A, H) is PV. □

4. Unique existence of minimal splitting fields

Let A be a ring. Let $A[[t]]$ denote the ring of formal power series of one variable t with coefficients in A. This is regarded as a differential ring with respect to the standard derivation

$$\left(\sum_{n=0}^{\infty} a_n t^n\right)' = \sum_{n=1}^{\infty} n a_n t^{n-1} \quad (a_n \in A).$$

Given an ideal I of A, $I[[t]] = \left\{\sum_{n=0}^{\infty} a_n t^n \mid a_n \in I\right\}$ is a differential ideal of $A[[t]]$.

LEMMA 4.1 (cf. [1, Lemma 1.8, Theorem 2.1]). *Suppose that A is a differential ring including $\mathbb{Q}$ as a subfield.*

(1) *$\rho : A \to A[[t]]$, $\rho(a) = \sum_{n=0}^{\infty} \frac{a^{(n)}}{n!} t^n$ is an injection of differential rings.*
(2) *An ideal I of A is a differential ideal if and only if $I = \rho^{-1}(I[[t]])$.*
(3) *If I is a differential ideal of A, then the radical $\sqrt{I}$ of I is differential too. Moreover, $\sqrt{I}$ equals the intersection $\bigcap_\lambda Q_\lambda$ of suitable differential ideals Q_λ that are prime.*
(4) *If A is simple as a differential ring (see 2.4), it is an integral domain.*

PROOF. (1), (2) Easy. (3) The first assertion follows from (2), since $I = \rho^{-1}(I[[t]])$ implies $\sqrt{I} = \rho^{-1}(\sqrt{I}[[t]])$. For the second, let $Q_\lambda = \rho^{-1}(P_\lambda[[t]])$, where P_λ are prime ideals such that $\sqrt{I} = \bigcap_\lambda P_\lambda$. (4) This follows from (3). □

LEMMA 4.2 (Levelt [21, Appendix]). *Let K be a differential field of characteristic* ch $K = 0$*, and let A be a differential K-algebra. Suppose that A is finitely generated*

as a K-algebra and is simple as a differential ring, which is then necessarily an integral domain; see 4.1 (4). Let $L = Q(A)$; see 1.2. Then the field extension L_0/K_0 is algebraic.

PROOF. First, we claim $L_0 = A_0$. In fact, if $x \in L_0$, then $\{a \in A \mid ax \in A\}$ is a nonzero differential ideal in A and hence equals A, whence $x \in A_0$. Next, we see that every $y \in A_0$ is algebraic over K, since for any maximal ideal P of A, the constant field A_0 is included in A/P, which, being finitely generated as a K-algebra, is algebraic over K. Let $\varphi(t) = t^n + c_1 t^{n-1} + \cdots + c_n$ be the minimal polynomial of y ($\in A_0$) over K. Since $c_1' t^{n-1} + c_2' t^{n-2} + \cdots + c_n'$ has y as a root, the minimality of $\varphi(t)$ implies that all $c_i \in K_0$ so that y is algebraic over K_0. □

THEOREM 4.3 (Unique existence). *Suppose that K is a differential field of characteristic* 0 *whose constant field K_0 is algebraically closed. Let V be a differential K-module of finite K-dimension. Then, there exists a minimal splitting field L for V with $L_0 = K_0$ so that L/K is necessarily a finitely generated PV extension. Such an L is unique up to isomorphism of differential K-algebras.*

PROOF. *Existence*. Suppose that (3.8) is the differential system corresponding to V with respect to an ordered K-basis $y_1, \ldots, y_n$. Let $K[X_{ij}]$ be the polynomial algebra in X_{ij} ($1 \le i, j \le n$), and let $O := K[X_{ij}, \frac{1}{\det}]$ be its localization at the determinant $\det(X_{ij})$. Make these into differential K-algebras by defining $(X_{ij}') = (a_{ij})(X_{ij})$. Choose a maximal differential ideal P in O, and let $L = Q(O/P)$. By 4.2, $L_0 = K_0$, since K_0 is algebraically closed. Let x_{ij} denote the natural image of X_{ij} in L. Then, $L = K(x_{ij})$, and the matrix (x_{ij}) is invertible. It follows that ${}^t(x_{1j}, \ldots, x_{nj})$ ($1 \le j \le n$) are n K_0-linearly independent solutions of (3.8). Hence, L is what is required.

Uniqueness. Suppose we have two fields, L_i ($i = 1, 2$), such as required. Let A_i be the principal differential ring for L_i/K, and define $A = A_1 \otimes_K A_2$, and consider it as a differential K-algebra. Choose a maximal differential ideal I in A, and let $L = Q(A/I)$. Again by 4.2, $L_0 = K_0$. Since A_i is simple by 2.5, the natural maps $A_i \to A/I$ of differential K-algebras are both injective so that they extend to injections $L_i \hookrightarrow L$ of differential fields over K. Since we have $L_i^V \xrightarrow{\simeq} L^V$ by counting L_0-dimension, the images of the L_i are both $K\langle V\rangle$ in L so that $L_1 \simeq L_2$, as desired. □

5. Liouville extensions

Let k be a field with an algebraic closure $\bar{k}$. A *separable k-algebra* is a finite-dimensional k-algebra R such that the base extension $R \otimes_k \bar{k}$ is reduced; this is equivalent to the statement that R is isomorphic to a product $L_1 \times \cdots \times L_r$ of finitely many separable field extensions L_i/k of finite degree. Every finitely generated k-algebra A contains the largest separable subalgebra; it will be denoted by $\pi_0 A$, called the *separable part* of A. We have $\pi_0(A \otimes_k B) = \pi_0 A \otimes_k \pi_0 B$, $\pi_0(A \otimes_k k') = \pi_0 A \otimes_k k'$, where A and B are finitely generated k-algebras, and k'/k is a field extension.

If H is a finitely generated commutative Hopf algebra, then $\pi_0 H \subset H$ is a finite-dimensional Hopf subalgebra. Let $H^0 := H/(\pi_0 H)^+ H$ denote the corresponding quotient Hopf algebra. Note that $(\pi_0 H)^+ H$ is the smallest ideal I of H such that $\varepsilon(I) = 0$, and H/I is coconnected in the sense that the topological space Spec H/I of all prime ideals is connected; this means that H/I contains no nontrivial idempotents. If k is algebraically closed, $\pi_0 H$ is isomorphic to the dual $(k\Gamma)^*$ of the group algebra $k\Gamma$ of some finite group Γ.

PROPOSITION 5.1. *Let $(L/K, A, H)$ be a finitely generated PV extension with $k = K_0$ $(= L_0)$.*

(1) *The K-separable part $\pi_0 A$ of A is an intermediate differential field in L/K. It coincides with the separable algebraic closure K^{s} of K in L. In particular, K^{s}/K is a finite extension.*
(2) *$(\pi_0 A/K, \pi_0 A, \pi_0 H)$ is a PV extension. Suppose that k is algebraically closed so that $\pi_0 H = (k\Gamma)^*$ with Γ a finite group. Then, $\pi_0 A/K$ is a finite Galois extension with Galois group Γ.*
(3) *L/K is finite separable if and only if H is separable as a k-algebra.*
(4) *L/K is separably closed (i.e. $K^{\mathrm{s}} = K$) if and only if H is coconnected.*

PROOF. (1), (2) Taking the K-separable part in $A \otimes_K A \simeq A \otimes_k H = A \otimes_K (K \otimes_k H)$, we have

$$\pi_0 A \otimes_K \pi_0 A \simeq \pi_0 A \otimes \pi_0 H, \tag{5.2}$$

whence $A \otimes_K \pi_0 A \simeq A \otimes_k \pi_0 H$. This implies $\pi_0 A = \theta^{-1}(A \otimes_k \pi_0 H)$. Hence, $\pi_0 A$ is an intermediate differential field in L/K. Part 2 now follows by (5.2). Set $M = \pi_0 A$. It remains to prove that L/M is separably closed. Since $(L/M, A, H^0)$ is PV by 2.6 (2) and 2.8, we have $L \otimes_M A \simeq L \otimes_k H^0$. Since the base extension $L \otimes_k H^0$ is still coconnected, its spectrum contains the nilradical as a generic point. Therefore, its localization $L \otimes_M L$, whose spectrum necessarily contains such a point, is coconnected; this implies what should be proved. (3), (4) These follow from (1) and (2). □

DEFINITION 5.3. An extension L/K of differential fields is said to be *Liouville* if the following are satisfied.

(i) $L_0 = K_0$.
(ii) There is a sequence of intermediate differential fields in L/K

$$K = L_0 \subset L_1 \subset \cdots \subset L_r = L$$

in which each L_i/L_{i-1} $(0 < i \leq r)$ is either
(a) finite separable,
(b) of the form $L_i = L_{i-1}(x)$ with x primitive over L_{i-1} (1.15 (1)), or
(c) of the form $L_i = L_{i-1}(x)$ with x exponential over L_{i-1} (1.15 (2)).

Of course it follows that L/K is finitely generated as a field extension.

To characterize Liouville PV extensions, the language of group schemes seems more suitable than that of Hopf algebras. We will use some terminology and basic results on group schemes from [19]. Let $\mathbf{G} = \operatorname{Sp} H$ be an affine group scheme over a field k, which corresponds to a commutative Hopf algebra H. Suppose that H is finitely generated, or in other words, $\mathbf{G}$ is algebraic. The dimension $\dim \mathbf{G}$ is by definition the Krull dimension $\operatorname{Kdim} H$. We write

$$\pi_0\mathbf{G} = \operatorname{Sp}(\pi_0 H), \quad \mathbf{G}^0 = \operatorname{Sp} H^0.$$

The latter is called the connected component of the unit element in $\mathbf{G}$. We have a short exact sequence $\mathbf{G}^0 \rightarrowtail \mathbf{G} \twoheadrightarrow \pi_0\mathbf{G}$ of affine group schemes. $\mathbf{G}$ is said to be *finite etale* (resp., *connected*) if $\mathbf{G} = \pi_0\mathbf{G}$ (resp., $\mathbf{G}^0 = \mathbf{G}$). If k is algebraically closed, $\mathbf{G}^0 \rightarrowtail \mathbf{G}$ splits as a left or right $\mathbf{G}^0$-equivariant morphism of schemes.

EXAMPLE 5.4 (Two basic affine group schemes). (1) *The additive group* $\mathbf{G}_\mathrm{a} = \operatorname{Sp} k[X]$; X is a primitive in $k[X]$. The group $\mathbf{G}_\mathrm{a}(R)$ of the rational points in a k-algebra R is the additive group R. A closed subgroup $\mathbf{G}$ of $\mathbf{G}_\mathrm{a}$ corresponds to a Hopf algebra $H = k[l]$ that is generated by a primitive l, with a closed embedding given by $X \mapsto l$, $k[X] \twoheadrightarrow k[l]$. If $\mathbf{G} = \mathbf{G}_\mathrm{a}$, it is connected. Suppose that $\mathbf{G} \subsetneq \mathbf{G}_\mathrm{a}$. If $\operatorname{ch} k = 0$, then $\mathbf{G} = \{1\}$. If $\operatorname{ch} k = p > 0$, then $H = k[X]/(f(X^{p^r}))$, where $r \geq 0$ and $f(X) = X + c_1X^p + c_2X^{p^2} + \cdots + c_sX^{p^s}$ ($c_i \in k$). In this case,

$$H^0 = k[X]/(X^{p^r}), \quad \pi_0 H = k[X^{p^r}]/(f(X^{p^r})).$$

(2) *The multiplicative group* $\mathbf{G}_\mathrm{m} = \operatorname{Sp} k[X, X^{-1}]$; X is a grouplike in $k[X, X^{-1}]$. The group $\mathbf{G}_\mathrm{m}(R)$ is the group of units in R. A closed subgroup of $\mathbf{G}_\mathrm{m}$ corresponds to a Hopf algebra $H = k[g, g^{-1}]$ that is generated by a grouplike g and its inverse, with a closed embedding given by $X^{\pm 1} \mapsto g^{\pm 1}$, $k[X, X^{-1}] \twoheadrightarrow k[g, g^{-1}]$. If $\mathbf{G} = \mathbf{G}_\mathrm{m}$, it is connected. Suppose that $\mathbf{G} \subsetneq \mathbf{G}_\mathrm{m}$. Then, $\mathbf{G}$ is finite and g has a finite order, say n. If $\operatorname{ch} k = 0$, $\mathbf{G}$ is finite etale. Suppose $\operatorname{ch} k = p > 0$ and $n = p^e n'$ with $p \nmid n'$. In this case,

$$H^0 = k[X]/(X^{p^e} - 1), \quad \pi_0 H = k[X^{p^e}]/(X^n - 1).$$

DEFINITION 5.5. An algebraic affine group scheme $\mathbf{G}$ is said to be *Liouville*, if there is a normal chain of closed subgroups

$$\mathbf{G} = \mathbf{G}_0 \rhd \mathbf{G}_1 \rhd \cdots \rhd \mathbf{G}_r = \{1\}$$

in which each factor $\mathbf{G}_i/\mathbf{G}_{i+1}$ ($0 \leq i < r$) is either

(a) finite etale,
(b) a closed subgroup scheme of $\mathbf{G}_\mathrm{a}$, or
(c) a closed subgroup scheme of $\mathbf{G}_\mathrm{m}$.

Such a chain as above will be called a *Liouville normal chain* (*LNC*) of length r.

LEMMA 5.6. *Let* $\mathbf{G}$ *be an algebraic affine group scheme over a field* k.

(1) *If* $\mathbf{G}$ *is Liouville, every closed subgroup of* $\mathbf{G}$ *is Liouville.*
(2) *Given a normal closed subgroup* $\mathbf{N} \lhd \mathbf{G}$, $\mathbf{G}$ *is Liouville if and only if* $\mathbf{N}$ *and* $\mathbf{G}/\mathbf{N}$ *are both Liouville.*
(3) $\mathbf{G}$ *is Liouville if and only if* $\mathbf{G}^0$ *is Liouville.*

PROOF. (1), (2) Standard; see the proof of [13, Lemma 1.2]. (3) This follows from (2). □

PROPOSITION 5.7. *Let* $\mathbf{G}$ *be a connected algebraic affine group scheme over a field* k.

(1) *Suppose that* $\mathbf{G}$ *is Liouville. Then, it has an LNC in which each factor is a* connected *closed subgroup of* $\mathbf{G}_{\mathrm{a}}$ *or* $\mathbf{G}_{\mathrm{m}}$. *If* $\operatorname{ch} k = 0$, *this means that an LNC can be chosen so that each factor is isomorphic to* $\mathbf{G}_{\mathrm{a}}$ *or* $\mathbf{G}_{\mathrm{m}}$.
(2) *If* $\mathbf{G}$ *is Liouville, it is solvable* (*i.e. has a normal chain with Abelian factors*). *The converse holds if* k *is algebraically closed.*

PROOF. (1) Let $\mathbf{G} = \mathbf{G}_0 \rhd \cdots \rhd \mathbf{G}_r = \{1\}$ be an LNC of least length. We will proceed by induction on r. Since $\mathbf{G}$ is connected, the derived closed subgroup $\mathscr{D}\mathbf{G}$ is connected as well [19, Theorem 10.1]. The first factor $\mathbf{G}/\mathbf{G}_1$ cannot be finite etale so that $\mathscr{D}\mathbf{G} \subset \mathbf{G}_1$. Hence, the least length of LNC's for $\mathscr{D}\mathbf{G}$ is smaller than r; the induction hypothesis can apply to $\mathscr{D}\mathbf{G}$. By replacing $\mathbf{G}$ with $\mathbf{G}/\mathscr{D}\mathbf{G}$, we may suppose that $\mathbf{G}$ is Abelian. We then have a short exact sequence $\mathbf{G}_{\mathrm{s}} \rightarrowtail \mathbf{G} \twoheadrightarrow \mathbf{G}_{\mathrm{u}}$, where $\mathbf{G}_{\mathrm{u}}$ is unipotent and $\mathbf{G}_{\mathrm{s}}$ is of multiplicative type; this sequence splits if k is perfect [19, Theorem 9.5]. So it remains to prove (1) when $\mathbf{G} = \mathbf{G}_{\mathrm{u}}$ or $\mathbf{G}_{\mathrm{s}}$.

The case $\mathbf{G} = \mathbf{G}_{\mathrm{u}}$. If $\operatorname{ch} k = 0$, $\mathbf{G} \simeq \mathbf{G}_{\mathrm{a}} \times \cdots \times \mathbf{G}_{\mathrm{a}}$. In general, every (connected) unipotent algebraic affine group scheme has a normal chain in which every factor is a (connected) closed subgroup of $\mathbf{G}_{\mathrm{a}}$. Hence we are done.

The case $\mathbf{G} = \mathbf{G}_{\mathrm{s}}$. Since $\mathbf{G}$ is connected, it is a torus if $\operatorname{ch} k = 0$. If $\operatorname{ch} k = p > 0$, we have a short exact sequence $\mathbf{T} \rightarrowtail \mathbf{G} \twoheadrightarrow \mathbf{F}$, where $\mathbf{T}$ is a torus, and $\mathbf{F}$ $(= \operatorname{Sp} J)$ is a finite group scheme of multiplicative type whose order, i.e. the dimension $\dim J$, is a power of p. The desired result for $\mathbf{F}$ is proved by induction on its order. In any characteristic, we may suppose that $\mathbf{G}$ is a torus. The first factor $\mathbf{G}/\mathbf{G}_1$ then must be a closed subgroup of $\mathbf{G}_{\mathrm{m}}$. The result applied to the largest anisotropic subtorus $\mathbf{T}_{\mathrm{a}}$ in $\mathbf{G}$ implies that $\mathbf{T}_{\mathrm{a}}$ must be trivial. By [19, Theorem 7.4], this means that $\mathbf{G}$ splits or $\mathbf{G} \simeq \mathbf{G}_{\mathrm{m}} \times \cdots \times \mathbf{G}_{\mathrm{m}}$. The desired result is then obvious.

(2) The first assertion follows by (1). For the second, we can replace $\mathbf{G}$ with $\mathscr{D}^i\mathbf{G}/\mathscr{D}^{i+1}\mathbf{G}$, and may suppose that $\mathbf{G}$ is Abelian, in which case we are done using the argument stated earlier since $\mathbf{G}_{\mathrm{s}}$ is diagonalizable if k is algebraically closed. □

THEOREM 5.8. *Let* L/K *be a finitely generated PV extension with PV group scheme* $\mathbf{G} = \mathbf{G}(L/K)$. *Let* K^{s} *denote the separable algebraic closure of* K *in* L; *see 5.1* (*1*).

(1) *The following are equivalent:*
  (i) *L/K is Liouville.*
  (ii) *There is a sequence of intermediate differential fields in L/K^{s}*

$$K^{\mathrm{s}} = L_0 \subset L_1 \subset \cdots \subset L_r = L \tag{5.9}$$

  such that for each $0 < i \leq r$, L_{i-1} is separably closed in L_i, and $L_i = L_{i-1}(x_i)$, where x_i is either primitive or exponential over L_{i-1}.
  (iii) *L/K is embedded into some Liouville extension $\tilde{L}/K$.*
  (iv) *$\mathbf{G}$ is Liouville.*
  (v) *$\mathbf{G}^0$ is Liouville.*

  If the constant field k is algebraically closed, these are further equivalent to
  (vi) *$\mathbf{G}^0$ is solvable.*

(2) *Suppose* $\operatorname{ch} k = 0$*, and that L/K is Liouville. Then, such a sequence as in (ii) earlier can be chosen so that for each $0 < i \leq r$, x_i is transcendental over L_{i-1} or equivalently $\mathbf{G}(L_i/L_{i-1}) \simeq \mathbf{G}_{\mathrm{a}}$ or $\mathbf{G}_{\mathrm{m}}$. In this case,*

$$r = \operatorname{trans.deg}_K L = \dim \mathbf{G} = \text{the least length of LNC's for } \mathbf{G}^0.$$

Proof. (1) Recall from 1.15 the characterization of the two types of basic PV extensions. Then, Theorem 2.11, combined with 5.1 and 5.7, proves the equivalence among the conditions except (iii). Obviously, (i) $\Rightarrow$ (iii). Suppose (iii), and that $K = \tilde{L}_0 \subset \tilde{L}_1 \subset \cdots \subset \tilde{L}_r = \tilde{L}$ is a sequence satisfying the conditions in 5.3. We will prove (iv) and (v) by induction on r. If $r = 0$, this is obvious. Suppose $r > 0$, and let $L\tilde{L}_1$ be the composite field. By the next proposition, $L\tilde{L}_1/\tilde{L}_1$ is a finitely generated PV extension with $\mathbf{G}(L\tilde{L}_1/\tilde{L}_1) \simeq \mathbf{G}(L/L \cap \tilde{L}_1)$. With the induction hypothesis applied to $L\tilde{L}_1/\tilde{L}_1$, $\mathbf{G}(L/L \cap \tilde{L}_1)$ is Liouville. If $\tilde{L}_1/K$ is finite separable, $L/L \cap \tilde{L}_1$ too is, so that $\mathbf{G}(L/K)^0 \subset \mathbf{G}(L/L \cap \tilde{L}_1)$, and (v) follows. Otherwise, $L \cap \tilde{L}_1/K$ is a PV extension whose PV group scheme, a quotient of $\mathbf{G}(\tilde{L}_1/K)$, is a closed subgroup of $\mathbf{G}_{\mathrm{a}}$ or $\mathbf{G}_{\mathrm{m}}$; (iv) follows. (2) This follows from Part 1 just proven and 5.7 (1). □

Proposition 5.10. *Let $\tilde{L}/K$ be an extension of differential fields such that $\tilde{L}_0 = K_0$. Let L, M be an intermediate differential fields in $\tilde{L}/K$. Suppose that L/K is a finitely generated PV extension. Then, LM/M is a finitely generated PV extension whose PV group scheme $\mathbf{G}(LM/M)$ is naturally isomorphic to $\mathbf{G}(L/L \cap M)$.*

Proof. Set $F := L \cap M$. Let $X = (x_{ij}) \in GL_n(L)$ be a GL_n-primitive matrix such that $L = K(x_{ij})$; see 3.11 (iv). Set $B = (b_{ij}) := (X^{-1} \otimes_F 1)(1 \otimes_F X)$ in $M_n(L \otimes_F L)$, and $\tilde{B} = (\tilde{b}_{ij}) := (X^{-1} \otimes_M 1)(1 \otimes_M X)$ in $M_n(LM \otimes_M LM)$. Construct the subalgebras

$$H := k\left[b_{ij}, \frac{1}{|B|}\right] \subset L \otimes_F L, \quad \tilde{H} := k\left[\tilde{b}_{ij}, \frac{1}{|\tilde{B}|}\right] \subset LM \otimes_M LM$$

over the constant field $k = K_0$. By 3.11, L/F (resp., LM/M) is a finitely generated PV extension with PV Hopf algebra H (resp., $\tilde{H}$). It remains to prove that the canonical epimorphism $\varphi : H \to \tilde{H}$, $\varphi(b_{ij}) = \tilde{b}_{ij}$ is injective. Consider the correspondence of

2.1 applied to L/F. It then suffices to prove that the intermediate differential field $M_{\mathfrak{a}}$ which corresponds to the differential coideal $\mathfrak{a} := (\mathrm{Ker}\,\varphi) \cdot (L \otimes_F L)$ coincides with F. Since $\mathfrak{a} \subset \mathrm{Ker}\,(L \otimes_F L \to LM \otimes_M LM)$, it follows that if $x \epsilon M_{\mathfrak{a}}$, then $1 \otimes_M x = x \otimes_M 1$ in $LM \otimes_M LM$, whence $x \epsilon L \cap M = F$; this proves $M_{\mathfrak{a}} = F$, as desired. □

Part II: PV theory in the *C*-ferential context

Throughout this part, we work over a fixed ground field R. All vector spaces, (co)algebras, and linear maps are over R. Let $\otimes$, and Hom stand for the tensor product $\otimes_R$ over R, and the vector space Hom_R of linear maps, respectively. Let C denote a cocommutative coalgebra (over R) with a specified grouplike 1_C.

6. *C*-ferential algebras

We let

$$\Delta : C \to C \otimes C, \quad \Delta(c) = \sum c_1 \otimes c_2; \quad \varepsilon : C \to R$$

denote the structure maps of C. So, $\Delta(1_C) = 1_C \otimes 1_C, \varepsilon(1_C) = 1$.

Definition 6.1. A *C-ferential module* is a vector space V equipped with a *unital C-action* $\psi : C \otimes V \to V$ by which we mean that ψ is a linear map such that $\psi(1_C \otimes v) = v$ $(v \in V)$. We will simply write cv for $\psi(c \otimes v)$. Let ${}_C\mathcal{M}$ denote the category of C-ferential modules and C-ferential (i.e. C-action preserving) maps.

The category ${}_C\mathcal{M}$ forms an R-linear Abelian tensor category $({}_C\mathcal{M}, \otimes, R)$ with respect to the tensor product $V \otimes W$ $(V, W \in {}_C\mathcal{M})$ endowed with the C-action

$$c(v \otimes w) = \sum c_1 v \otimes c_2 w \quad (c \in C,\ v \otimes w \in V \otimes W)$$

and the unit object R with the trivial action $c1 = \varepsilon(c)1$. This is symmetric with respect to the obvious symmetry $v \otimes w \mapsto w \otimes v$, $V \otimes W \to W \otimes V$, since C is supposed to be cocommutative [or in notation, $\sum c_2 \otimes c_1 = \sum c_1 \otimes c_2$ $(c \in C)$]. We attach the adjective 'C-ferential' to the algebra- and the module-objects in ${}_C\mathcal{M}$, as follows.

Definition 6.2. A commutative algebra in the symmetric tensor category ${}_C\mathcal{M}$ is called a *C-ferential algebra*; we thus keep the assumption that algebras are commutative. For such an algebra A, an A-module in ${}_C\mathcal{M}$ is called a *C-ferential A-module*. Let ${}_{A,C}\mathcal{M}$ denote the category of C-ferential A-modules and C-ferential A-linear maps.

Explicitly, a C-ferential algebra is an ordinary commutative algebra A equipped with a unital C-action such that $c(ab) = \sum (c_1 a)(c_2 b), c1 = \varepsilon(c)1$ $(c \in C, a, b \in A)$.

It is called a *C-ferential field* if it is a field. A C-ferential A-module is an ordinary A-module M together with a unital C-action such that $c(am) = \sum (c_1 a)(c_2 m)$ $(c \in C, a \in A, m \in M)$.

REMARK 6.3. Let $C^+ = \mathrm{Ker}\varepsilon$. The tensor algebra $T(C^+)$ on C^+ has a unique (cocommutative) bialgebra structure such that the natural embedding $C \to T(C^+)$, which sends 1_C to the unit 1 of $T(C^+)$, and is the identity on C^+, is a coalgebra map. In the obvious way, a C-ferential module is identified with a left (or right) $T(C^+)$-module. Moreover, the symmetric tensor category ${}_C\mathcal{M}$ is identified with that of left $T(C^+)$-modules ${}_{T(C^+)}\mathcal{M}$. A C-ferential algebra A and a C-ferential A-module are identified with a $T(C^+)$-module algebra and a left module over the smash-product algebra $A\#T(C^+)$, respectively. Thus, ${}_{A,C}\mathcal{M} = {}_{A\#T(C^+)}\mathcal{M}$; see [16, Chapter VII].

EXAMPLE 6.4. Suppose $C = R1_C \oplus Rd$, where $d(\neq 0)$ is a primitive. Then, ${}_C\mathcal{M}$ is precisely the tensor category of differential R-modules, where R is regarded as a differential algebra with zero derivation. Hence, C-ferential objects (e.g. C-ferential algebras) equal differential objects over R (e.g. differential R-algebras). Note that any differential field K can be regarded as a C-ferential field, by choosing as R the prime field of K.

One will see that the results in this part specialize to those in the previous part, with C chosen just as above, and also specialize to the situation of [6, p. 3], [7], with C chosen as in the following two examples. In Sections 8 and 9, we will impose some additional assumptions on C, which are, however, fulfilled for these three examples at least in characteristic zero.

EXAMPLE 6.5. (1) Suppose that $C = R1_C \oplus \bigoplus_{\delta \in \Delta} R\delta$, where δ $(\neq \phi)$ is a family Δ of primitives[3]. Then, a C-ferential algebra is precisely an algebra A given a family Δ of linear derivations on A; this is a Δ-*algebra* in the sense of Buium [6].

(2) Suppose that $C = R1_C \oplus Rd_1 \oplus Rd_2 \oplus \cdots$, where $d_0 = 1_C, d_1, d_2, \ldots$ form a divided power sequence in the sense $\Delta(d_n) = \sum_{i=0}^{n} d_i \otimes d_{n-i}$ $(n = 1, 2, \ldots)$. Then, a C-ferential algebra is an algebra A provided with a *higher derivation*, that is, an infinite sequence of linear endomorphisms $\partial_0 = \mathrm{id}, \partial_1, \partial_2, \ldots$ such that $\partial_n(ab) = \sum_{i=0}^{n} (\partial_i a)(\partial_{n-i} b)$ $(a, b \in A)$.

Given a C-ferential module V, we define the subspace of *constants* by

$$V_0 = \{v \in V \mid cv = \varepsilon(c)v \quad (c \in C)\}.$$

If A is a C-ferential algebra, $A_0 \subset A$ is a subalgebra. More generally, suppose that A is a *C-ferential K-algebra*, that is, a C-ferential algebra equipped with a C-ferential

[3] Following a custum we have used the symbol Δ to denote families of primitives and of derivations; the same symbol always denotes the comultiplication of a coalgebra, elsewhere in the text.

algebra map $K \to A$. Then A_0 is a K_0-subalgebra of A. If $M \in {}_{A,C}\mathcal{M}$, then $M_0 \subset M$ is an A_0-submodule.

DEFINITION 6.6. A C-ferential algebra A ($\neq 0$) is said to be *simple* if it includes no nontrivial C-ferential ideal.

PROPOSITION 6.7 (see [16, Lemma 10.2.2]). *A C-ferential algebra A is simple if and only if*

(a) *A_0 is a field and*

(b) *for every $M \in {}_{A,C}\mathcal{M}$, the natural map $A \otimes_{A_0} M_0 \to M$, $a \otimes m \mapsto am$ is injective.*

In general, if D is a coalgebra and A an algebra, $\mathrm{Hom}(D, A)$ becomes an algebra with respect to the convolution product $(f * g)(d) = \sum f(d_1)g(d_2)$, where $f, g \in \mathrm{Hom}(D, A)$, $d \in D$, $\Delta(d) = \Sigma d_1 \otimes d_2$; see [16, Section 4.0].

LEMMA 6.8. *Suppose that A is a C-ferential algebra. Then* $\mathrm{Hom}(T(C^+), A)$ *is a C-ferential A-algebra with respect to the C-action*

$$(cf)(h) = f(hc) \qquad (c \in C,\ h \in T(C^+),\ f \in \mathrm{Hom}(T(C^+), A))$$

and the (*splitting*) *algebra map*

$$\tilde{\rho} : A \to \mathrm{Hom}(T(C^+), A), \quad a \mapsto (h \mapsto ha). \tag{6.9}$$

We have

$$\mathrm{Hom}(T(C^+), A)_0 = \mathrm{Hom}(T(C^+)/\mathrm{Ker}\varepsilon, A) = A.$$

COROLLARY 6.10. *If K is a C-ferential field, then the map*

$$K \otimes_{K_0} K \to \mathrm{Hom}(T(C^+), K), \quad a \otimes b \mapsto (h \mapsto a(hb))$$

is injective.

PROOF. This follows from property (b) in 6.7 which holds for the C-ferential K-module $\mathrm{Hom}(T(C^+), K)$. □

The following will play the role of the Wronskian criterion 1.4.

PROPOSITION 6.11. *In a C-ferential field K, finitely many, say n, elements $a_1, \ldots, a_n$ are K_0-linearly independent if and only if there exist n elements $h_1, \ldots, h_n$ in $T(C^+)$ such that the matrix $(h_i a_j)$ in $M_n(K)$ is invertible.*

PROOF. "If". If $\sum_j c_j a_j = 0$ with $c_i \in K_0$, $\sum_j c_j(ha_j) = 0$ ($h \in T(C^+)$), whence $(c_1, \ldots, c_n)^t(h_i a_j) = (0, \ldots, 0)$; this implies $c_1 = \cdots = c_n = 0$.

"Only if." Set $W = \sum_{j=1}^{n} K_0 a_j$. By 6.10, we have a K-linear injection

$$K \otimes_{K_0} W \hookrightarrow \mathrm{Hom}(T(C^+), K) \simeq \mathrm{Hom}_K(K \otimes T(C^+), K),$$

whose K-predual

$$K \otimes T(C^+) \to \mathrm{Hom}_{K_0}(W, K) = K^n, \quad a \otimes h \mapsto (a(ha_j))_j$$

must be surjective. Hence, there exist $h_1, \ldots, h_n \in T(C^+)$ such that their restrictions $h|_W, \ldots, h_n|_W$ form a K-basis of $\mathrm{Hom}_{K_0}(W, K)$, or in other words, $(h_i a_j)$ is invertible. □

7. Formalism of PV extensions and Galois correspondence I

Both the "PV extension formalism" and "Galois correspondence I", [see Theorem 2.6] can be directly generalized to the C-ferential context, without any additional assumptions on C.

Definition 7.1. An inclusion $L \supset K$, or an extension L/K, of C-ferential fields is called a *PV extension*, if

(i) $L_0 = K_0$, which will be denoted by k, and

(ii) there exists an intermediate C-ferential algebra $L \supset A \supset K$ such that
   (a) $L = Q(A)$, the quotient field of A,
   (b) $A \otimes_K A$ is generated by $(A \otimes_K A)_0$ as a left (or equivalently right) A-module.

Such an A as given earlier is called a *principal C-ferential algebra* for L/K.

With L/K, A as given earlier, $A \otimes_K A$ is naturally a C-ferential K-algebra so that $H := (A \otimes_K A)_0$ is an algebra over $k = K_0$. The A-bimodules in ${}_C\mathcal{M}$ form a tensor category, $({}_A({}_C\mathcal{M})_A, \otimes_A, A)$. The object $A \otimes_K A$ in ${}_A({}_C\mathcal{M})_A$ has the natural structure maps $\boldsymbol{\Delta}$ and $\boldsymbol{\varepsilon}$ given by (1.6), of a *C-ferential A-coring* (i.e. a coalgebra in the tensor category just mentioned). Just as in the proof of 1.9, the property (b) in 6.7 of the C-ferential L-module $L \otimes_K A$ gives the isomorphism $\mu : A \otimes_k H \xrightarrow{\simeq} A \otimes_K A$, $\mu(a \otimes_k h) = (a \otimes_K 1) \cdot h$, from which the next two results follow.

Lemma 7.2. *Let A be a principal C-ferential algebra in a PV extension L/K of C-ferential fields. Let $H := (A \otimes_K A)_0$.*

(1) *The C-ferential K-algebra maps $\boldsymbol{\Delta}, \boldsymbol{\varepsilon}$, and the flip $\boldsymbol{\tau}$ induce on H structure maps*

$$\Delta : H \to H \otimes_k H, \quad \varepsilon : H \to k, \quad S : H \to H$$

which make it a commutative k-Hopf algebra.

(2) *A/K is a right H-Galois extension with respect to the H-comodule structure $\theta : A \to A \otimes_k H$, $\theta(a) = \sum a_0 \otimes a_1$ uniquely determined by*

$$1 \otimes_K a = \sum (a_0 \otimes_K 1) \cdot a_1 \quad \text{in } A \otimes_K A.$$

LEMMA 7.3. *For each PV extension of C-ferential fields, its principal C-ferential algebra is unique.*

Given a PV extension L/K, we say that $(L/K, A, H)$ is a PV extension, to indicate that A is the (unique) principal C-ferential algebra and $H = (A \otimes_K A)_0$ is the associated Hopf algebra, which is called the *PV Hopf algebra* for L/K.

We can reprove 2.1 and 2.3 in the C-ferential context, so that the next two results follow.

PROPOSITION 7.4. *The principal C-ferential algebra for a PV extension of C-ferential fields is simple.*

THEOREM 7.5 (Galois correspondence I). *Let $(L/K, A, H)$ be a PV extension of C-ferential fields.*

(1) *There is a 1–1 correspondence, given by (2.7), between the Hopf ideals I in H and the intermediate C-ferential fields M in L/K.*
(2) *If $I \leftrightarrow M$ under the correspondence, $(L/M, AM, H/I)$ is a PV extension.*

8. Galois correspondence II and Characterization

Throughout this section, we assume the following:

ASSUMPTION 8.1. *C is pointed irreducible, or in other words, the coradical of C equals $R1_C$.*

This holds for the C in 6.4 and 6.5. The assumption ensures the following.

LEMMA 8.2 (Sweedler [16, Lemma 9.2.3]). *Let A be an algebra. An element f in the algebra* $\mathrm{Hom}(C, A)$ *is invertible if and only if $f(1_C)$ is invertible in A.*

The Extension Lemma 1.2 is generalized by the following.

PROPOSITION 8.3. *Let A be a C-ferential algebra, and let $T \subset A$ be a multiplicative subset. Then the C-ferential algebra structure on A uniquely extends via the canonical map $A \to T^{-1}A$ to such a structure on the localization $T^{-1}A$.*

PROOF. Quite generally, if A is an algebra, the C-ferential algebra structures on A are in 1–1 correspondence with the sections $A \to \mathrm{Hom}(C, A)$ of the algebra epimorphism

$\mathrm{Hom}(C, A) \twoheadrightarrow A$, $f \mapsto f(1_C)$. We see from 8.2 that the composite

$$A \xrightarrow{\rho} \mathrm{Hom}(C, A) \xrightarrow{\text{cano}} \mathrm{Hom}(C, T^{-1}A),$$

where ρ denotes the algebra map $a \mapsto (c \mapsto ca)$ corresponding to the given C-ferential algebra structure, sends each t ($\in T$) to an invertible map in $\mathrm{Hom}(C, T^{-1}A)$ with value $t/1$ at 1_C. Therefore, it uniquely extends to an algebra map $T^{-1}A \to \mathrm{Hom}(C, T^{-1}A)$, which, as is easily checked, is the desired section. □

Given an extension L/K of C-ferential fields, an element x in L is said to be *primitive* (resp., *exponential*) over K, if $cx \in K$ for all $c \in C^+$ [resp., if it is nonzero and $(cx)x^{-1} \in K$ for all $c \in C$]. By using the proposition stated earlier, we can restate word for word Examples 1.15, and therefore reprove the results in Section 5 (especially Theorem 5.8), all in the C-ferential context.

THEOREM 8.4 (Galois correspondence II). *Let $(L/K, A, H)$ be a PV extension of C-ferential fields. Then, there is a 1–1 correspondence between the Hopf subalgebras J in H and the intermediate PV extensions M/K in L/K, in which to each M/K its PV Hopf algebra is assigned.*

PROOF. Note from the proof of 8.3 that if $A_1 \subset L$ is a C-ferential subalgebra, the quotient field $L_1 = Q(A_1)$ of A_1 (realized in L) is C-stable. Then, one sees that the proof of 2.8 is valid; see especially the sentence following (2.9). □

Just as in 2.11 and 2.13, the results 7.5 and 8.4 can be interpreted in terms of group schemes. We can restate Example 2.10 in the C-ferential context.

Assumption 8.1 is equivalent to the property that the bialgebra $T(C^+)$ is pointed irreducible as a coalgebra. It follows that $T(C^+)$ is necessarily a Hopf algebra. It is easy to see the following, which generalizes 3.3.

LEMMA 8.5. *Let A be a C-ferential algebra. Let V and W be a C-ferential A-modules [or left $A\#T(C^+)$-modules]. Then the A-module $\mathrm{Hom}_A(V, W)$ of A-linear maps $V \to W$ is a C-ferential A-module with respect to the C-action*

$$(cf)(v) = \sum c_1 f(S(c_2)v) \tag{8.6}$$

[$c \in C$, $v \in V$, $f \in \mathrm{Hom}_A(V, W)$], where S denotes the antipode of $T(C^+)$. We have

$$\mathrm{Hom}_A(V, W)_0 = \mathrm{Hom}_{A,C}(V, W),$$

which denotes the A_0-module of C-ferential A-linear maps $V \to W$.

Let L/K be an extension of C-ferential fields, and let V be a C-ferential K-module. Let us write

$$L^V := \mathrm{Hom}_{K,C}(V, L).$$

By a slight modification of 8.5, $\mathrm{Hom}_K(V, L)$ is a C-ferential L-module with $\mathrm{Hom}_K(V, L)_0 = L^V$. By 6.7, we have a natural injection $L \otimes_{L_0} L^V \hookrightarrow \mathrm{Hom}_K(V, L)$ in ${}_{L,C}\mathcal{M}$ so that $\dim_{L_0} L^V \leq \dim_K V$.

DEFINITION 8.7. L is called a *splitting field* for V if the following mutually equivalent conditions are satisfied (see 3.16):

(i) There exists an injection $L \otimes_K V \hookrightarrow \prod_\lambda L_\lambda$ in ${}_{L,C}\mathcal{M}$, where $\prod_\lambda L_\lambda$ is the direct product of copies L_λ of L.
(ii) $L \otimes_{L_0} L^V \hookrightarrow \mathrm{Hom}_K(V, L)$ has a dense image.

If $\dim_K V = n < \infty$, Conditions (i) and (ii) are equivalent to each of the following (see 3.9):

(iii) $L \otimes_K V \simeq L^n$ in ${}_{L,C}\mathcal{M}$.
(iv) $L \otimes_{L_0} L^V \xrightarrow{\simeq} \mathrm{Hom}_K(V, L)$.
(v) $\dim_{L_0} L^V = n$.

For a subset $Z = \{z_1, z_2, \dots\}$ of L, let $K\langle Z\rangle = K\langle z_1, z_2, \dots\rangle$ denote the smallest intermediate C-ferential field in L/K that contains Z. By 8.3, this equals the quotient field of the K-subalgebra of L generated by $T(C^+)Z$. L/K is said to be *finitely generated* if $L = K\langle z_1, \dots, z_n\rangle$ for some finitely many elements $z_1, \dots, z_n$ in L.

Let $K\langle V\rangle$ stand for $K\langle \bigcup_{f \in L^V} f(V)\rangle$; this is the smallest intermediate C-ferential field in L/K such that $L^V = K\langle V\rangle^V$. A splitting C-ferential field L for V is said to be *minimal* if $L = K\langle V\rangle$.

THEOREM 8.8 (Characterization). *Let L/K be an extension of C-ferential fields with the same constant field k.*

(1) *The following are equivalent:*
  (i) *L/K is a finitely generated PV extension.*
  (ii) *L is a minimal splitting field for some C-ferential K-module of finite K-dimension that is cyclic, regarded as a $K\#T(C^+)$-module.*
  (iii) *L is a minimal splitting field for some C-ferential K-module V of finite K-dimension.*
  (iv) *There exists a GL_n-primitive matrix $X = (x_{ij}) \in GL_n(L)$ for some n such that $L = K(x_{ij})$; by saying that X is GL_n-primitive, we mean that for every $c \in C$, $(cX)X^{-1} \in M_n(K)$, where $cX = (cx_{ij})$.*
(2) *The following are equivalent:*
  (i) *L/K is a PV extension.*
  (ii) *L is a minimal splitting field for some C-ferential K-module that is locally finite in the sense that it is a sum of C-ferential K-submodules of finite K-dimension.*

PROOF. (1) This is a modification of the proof of 3.11. For (i) $\Rightarrow$ (ii), let $u_1, \dots, u_n$ and $\mathbf{u} = (u_1, \dots, u_n)$ be just as in that proof. By 6.11, we can choose $h_1, \dots, h_n$ in

$T(C^+)$ such that the matrix $(h_i u_j)$ in $M_n(L)$ is invertible. Replace the W and the $\mathbf{u}^{(n)}$ in (3.13) with

$$W = (h_i u_j) \text{ and } h\mathbf{u} = (hu_1, \dots, hu_n) \quad (h \in T(C^+)),$$

respectively. Then, we see that $(h\mathbf{u})W^{-1} \in K^n$. This means that the cyclic C-ferential K-submodule, say V, of L^n, which is generated by $\mathbf{u}$, is K-spanned by the L-basis $h_1\mathbf{u}, \dots, h_n\mathbf{u}$ of L^n, whence $L \otimes_K V = L^n$. Note in addition that $K\langle V\rangle = K\langle u_1, \dots, u_n\rangle$, since L^V is k-spanned by the n projections.

For (iii) $\Rightarrow$ (iv), we need to choose a K-basis $v_1, \dots, v_n$ of V, and replace (3.8) with

$$\begin{pmatrix} cv_1 \\ \vdots \\ cv_n \end{pmatrix} = (a_{ij}(c)) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \quad (c \in K,\ a_{ij}(c) \in K).$$

The $\mathbf{x}_i$ in (3.14) should be replaced by

$$\mathbf{x}_i = {}^t(f_i(v_1), \dots, f_i(v_n)), \tag{8.9}$$

where $f_1, \dots, f_n$ constitute a k-basis of L^V. We see that

$$X = (\mathbf{x}_1, \dots, \mathbf{x}_n) \tag{8.10}$$

is the desired GL_n-primitive matrix. The further modifications should be obvious.

(2) For this, the proof of 3.17 can be copied since 3.15 is directly generalizable in the C-ferential context. □

To state a result without counterpart in Part I, let K be a C-ferential field with $k = K_0$. There is the k-Abelian symmetric tensor category, $({}_{K,C}\mathcal{M}, \otimes_K, K)$. Let $V \in {}_{K,C}\mathcal{M}$ be of finite K-dimension. The K-dual $V^* := \mathrm{Hom}_K(V, K)$ is in ${}_{K,C}\mathcal{M}$; see 8.6. Therefore, the full tensor subcategory, ${}_{K,C}\mathcal{M}_{\mathrm{fin}}$, consisting of all finite K-dimensional objects in ${}_{K,C}\mathcal{M}$ is rigid. Let $\{\{V\}\}$ denote the Abelian, rigid full tensor subcategory of ${}_{K,C}\mathcal{M}_{\mathrm{fin}}$ generated by V, that is, the smallest full subcategory containing V that is closed under subquotients, finite direct sums, tensor products, and duals. Thus, an object in $\{\{V\}\}$ is precisely a subquotient of some $W_1 \oplus \cdots \oplus W_r$, where each W_i is the tensor product of copies of V, V^*.

THEOREM 8.11 (cf. [4, Theorem 2.33]). *Let $(L/K, A, H)$ be a finitely generated PV extension of C-ferential fields. By 8.8 (1), we have an object $V \in {}_{K,C}\mathcal{M}_{\mathrm{fin}}$ for which L is a minimal splitting field.*

(1) *Let $W \in \{\{V\}\}$. Regard $A \otimes_K W$ as a right H-comodule with respect to the structure induced by $A = (A, \theta)$; see 7.2 (2). Then, $(A \otimes_K W)_0$ is an H-subcomodule with* $\dim_k(A \otimes_K W)_0 = \dim_K W$.

(2) *$W \mapsto (A \otimes_K W)_0$ gives a k-linear equivalence*

$$\{\{V\}\} \approx \mathcal{M}^H_{\mathrm{fin}}$$

of symmetric tensor categories, where $\mathcal{M}_{\text{fin}}^H = (\mathcal{M}_{\text{fin}}^H, \otimes_k, k)$ *denotes the rigid symmetric tensor category of finite-dimensional right* H*-comodules; note that this is isomorphic to the category of the same kind* $\text{Rep}_{\mathbf{G}(L/K)}$*, which consists of the finite-dimensional* k*-linear representations of* $\mathbf{G}(L/K) = \text{Sp}\, H$*, the PV group scheme of* L/K*.*

PROOF. Let $({}_C\mathcal{M}^H, \otimes_k, k)$ denote the symmetric tensor category consisting of those C-ferential k-modules N, which have a C-ferential, right H-comodule structure $N \to N \otimes_k H$. Since A is an algebra in ${}_C\mathcal{M}^H$, we have the symmetric tensor category $({}_{A,C}\mathcal{M}^H, \otimes_A, A)$ of A-modules in ${}_C\mathcal{M}^H$. Since A/K is H-Galois by 7.2 (2), it follows from [22, Theorem 2.11] (or see [14, Theorem 8.5.6, (1) $\Rightarrow$ (3)]) that $W \mapsto A \otimes_K W$ gives an equivalence ${}_{K,C}\mathcal{M} \approx {}_{A,C}\mathcal{M}^H$; this is a k-linear equivalence of symmetric tensor categories. Define k-linear functors,

$$(\mathcal{M}^H, \otimes_k, k) \underset{\Xi}{\overset{\Theta}{\rightleftarrows}} {}_{A,C}\mathcal{M}^H \ (\approx {}_{K,C}\mathcal{M}),$$

$$\Xi(N) = N_0, \quad \Theta(U) = A \otimes_k U \quad \text{(codiagonal } H\text{-coaction)},$$

where the category of the left-hand side is that of right H-comodules. Obviously, Θ is a symmetric tensor functor. Since $A_0 = k$, $\Xi \circ \Theta \simeq \text{id}$. By 6.7 and 7.4, the natural morphism in ${}_{A,C}\mathcal{M}^H$

$$\mu_N : \Theta \circ \Xi(N) = A \otimes_k N_0 \to N, \quad \mu_N(a \otimes_k n) = an,$$

is injective. Let $\mathcal{N}$ denote the full subcategory in ${}_{A,C}\mathcal{M}^H$ consisting of those N for which μ_N is bijective. Since $\Theta(U) \in \mathcal{N}$, $\mathcal{N}$ is closed under tensor products and Θ gives an equivalence $\mathcal{M}^H \approx \mathcal{N}$ of symmetric tensor categories.

Part 2 of the theorem will follow if we prove

(a) $A \otimes_K V \in \mathcal{N}$,

(b) $(A \otimes_K V)_0 = \Xi(A \otimes_K V) \in \mathcal{M}_{\text{fin}}^H$, and

(c) $(A \otimes_K V)_0$ generates $\mathcal{M}_{\text{fin}}^H$.

Choose a K-basis $v_1, \ldots, v_n$ of V and set $\mathbf{v} = (v_1, \ldots, v_n)$. Let $\mathbf{x}_i = {}^t(f_i(v_1), \ldots, f_i(v_n))$ $(1 \le i \le n)$ be as in (8.9) and set $X = (x_{ij}) = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ as in (8.10). Recall from 1.16 that $A = K[x_{ij}, \frac{1}{|X|}]$. We see that $1 \otimes_K \mathbf{v} \mapsto {}^tX$, or $1 \otimes v_i \mapsto (f_1(v_i), \ldots, f_n(v_i))$ $(1 \le i \le n)$, gives an isomorphism $A \otimes_K V \overset{\simeq}{\to} A^n$ in ${}_{A,C}\mathcal{M}$, which implies (a) and (b). It also follows that $(A \otimes_K V)_0$ has the entries in $(1 \otimes_K \mathbf{v})({}^tX^{-1} \otimes_K 1)$ as a k-basis of it. Let $B = (b_{ij}) = (X^{-1} \otimes_K 1)(1 \otimes_K X)$; see 1.16. Since $\theta({}^tX^{-1}) = ({}^tX^{-1} \otimes_k 1)(1 \otimes_k {}^tB^{-1})$ in $M_n(A \otimes_k H)$, we see that the coefficient space of the right H-comodule $(A \otimes_K V)_0$ is k-spanned by $S(b_{ij})$, the entries of ${}^tB^{-1}$; this implies (c).

If $W \in \{\{V\}\}$, then $A \otimes_K W \in \mathcal{N}$. Hence, $\dim_k(A \otimes_K W)_0$ equals the free A-rank of $A \otimes_K W$ or $\dim_K W$; this proves Part 1. □

9. Uniqueness and existence of minimal splitting fields

In general, a coalgebra D is said to be *Birkhoff–Witt* if it is isomorphic to the cofree pointed irreducible cocommutative coalgebra $B(U)$ on some vector space U; see [16, p. 261]. Such a D is necessarily pointed irreducible and cocommutative. Suppose that D is an irreducible cocommutative bialgebra, which is necessarily a Hopf algebra. If ch $R = 0$, then D is necessarily Birkhoff–Witt. If ch $R = p > 0$, then D is Birkhoff–Witt if and only if the Verschiebung map $\mathscr{V}_D : D \to R^{1/p} \otimes D$ is surjective [23, Corollary 4.2.7 (b)]. Recall that the map $\mathscr{V}_D$ is defined for every cocommutative coalgebra D, as the dual notion of the Frobenius map for commutative algebras.

Throughout this section we assume the following:

ASSUMPTION 9.1. *$T(C^+)$ is Birkhoff-Witt.*

If ch $R = 0$, this is equivalent to Assumption 8.1 that C is pointed irreducible. If ch $R = p > 0$, the assumption stated afore holds if C is pointed irreducible and if $\mathscr{V}_C$ is surjective. These last two conditions are true if C is spanned by divided power sequences. Therefore, such C as in Example 6.5 (2) satisfy the assumption in arbitrary characteristic, while those C in Examples 6.4 and 6.5 (1) do so only when ch $R = 0$; see the following example below.

Let A be a C-ferential algebra. Recall from 6.8 that $\mathrm{Hom}(T(C^+), A)$ is a C-ferential A-algebra endowed with the C-ferential algebra map $\tilde{\rho} : A \to \mathrm{Hom}(T(C^+), A)$ given by (6.9).

EXAMPLE 9.2. Suppose ch $R = 0$ and $C = R1_C \oplus Rd$, as in 6.4. Then, $T(C^+) = R[d]$, the polynomial Hopf algebra, in which $1, d, \frac{d^2}{2!}, \frac{d^3}{3!}, \ldots$ form a divided power sequence; see 6.5 (2). We can identify $\mathrm{Hom}(T(C^+), A) = A[[t]]$, by means of the pairing $\langle d^i, t^j \rangle = \delta_{ij} i!$

The map $\tilde{\rho}$ above generalizes the map $\rho : A \to A[[t]]$ defined in Lemma 4.1 (1). Lemma 4.1 (4) generalizes to the following.

LEMMA 9.3. *If a C-ferential algebra A is simple, it is an integral domain.*

PROOF. Just as in the proof of 4.1 (4), it suffices to prove the following in the order given:

(a) An ideal I of A is C-ferential if and only if $I = \tilde{\rho}^{-1}(\mathrm{Hom}(T(C^+), I))$.
(b) If $I \subset A$ is a C-ferential ideal, then $\sqrt{I} = \tilde{\rho}^{-1}(\mathrm{Hom}(T(C^+), \sqrt{I}))$.
(c) If $P \subset A$ is a prime ideal, then $\mathrm{Hom}(T(C^+), P)$ is prime.

It is easy to prove (a). Assumption 9.1 implies that $\mathrm{Hom}(T(C^+), A)$ is a projective limit of A-algebras, $A[[t_1, \ldots, t_r]]$, of formal power series. Therefore, $\mathrm{Hom}(T(C^+), A)$ is reduced (resp., an integral domain) if A is. This implies (b) and (c). □

Using 8.3 and 9.3, we can now reprove Lemma 4.2 in the C-ferential context.

THEOREM 9.4 (Unique existence). *Suppose that K is a C-ferential field whose constant field K_0 is algebraically closed. Let V be a C-ferential K-module of finite K-dimension. There exists a minimal splitting field L for V with $L_0 = K_0$ so that L/K is necessarily a finitely generated PV extension of C-ferential fields. Such an L is unique up to isomorphism of C-ferential K-algebras.*

PROOF. Modify the proof of 4.3, by using the just reproved lemma together with 7.4 and 8.3. □

We end this part with an example in positive characteristic.

EXAMPLE 9.5. Let C be as in Example 6.5 (2), but write k for R. Recall that this satisfies Assumption 9.1. Both the algebra $k[[t]]$ of formal power series and its quotient field $k((t))$ are C-ferential algebra with respect to the action $d_i t^j = \binom{j}{i} t^{j-i}$ $(i \le j), d_i t^j = 0$ $(i > j)$, so that $k[[t]]_0 = k((t))_0 = k$. Note that C is a commutative Hopf algebra with respect to the product $d_i d_j = \binom{i+j}{i} d_{i+j}$ and that the C-action above coincides with the natural action on the dual algebra $C^* = k[[t]]$. Suppose $\operatorname{ch} k = p > 0$. In the dual Hopf algebra C° $(\subset k[[t]])$ of C [16, Section 6.2], pick the primitive t, and choose another primitive $f \neq 0$; this must be an element of the form $f = \sum_{i=0}^{\infty} a_i t^{p^i}$ $(a_i \in k)$ and all these quality. Now take the (Hopf) subalgebra, $J := k[t, f]$. The Hopf subalgebra $J_1 := k[t] \cap k[f]$ in the polynomial Hopf algebra $k[f]$ is generated by a primitive, say $\nu(f)$. Since $k[t] \otimes_{J_1} k[f] = J$, we see $J/(t) = k[f]/(\nu(f))$. Let $H = k[f]/(\nu(f))$, $K = k(t), A = K \otimes_{k[t]} J = K[f]$, and $L = Q(A) = k(t, f)$. The natural isomorphism $J \otimes_{k[t]} J \xrightarrow{\simeq} J \otimes H, x \otimes y \mapsto \sum xy_1 \otimes \bar{y}_2$ is C-ferential, where C acts on the factor J in $J \otimes H$. By applying $K \otimes_{k[t]}$, we have a C-ferential isomorphism $A \otimes_K A \xrightarrow{\simeq} A \otimes H$ so that $(L/K, A, H)$ is PV. For example, if $f = \sum_{i=0}^{\infty} t^{p^i}$, then $\nu(f) = f - f^p$ $(= t)$ so that $L = A = K[f]/(f^p - f + t)$ and $H = k[f]/(f^p - f) \simeq k(\mathbb{Z}/(p))^*$. If $f = \sum_{i=0}^{\infty} t^{p^{p^i}}$, then $\nu(f) = 0$ so that $A = K[f], H = k[f]$ are polynomial algebras.

Part III: Unified PV theory

In this part, we work over a fixed ground field R, keeping the conventions stated at the beginning of Part II, but C will be replaced by a special kind of D.

10. What's the difference in the difference case

To move from the differential world to the difference world, one has only to exchange "derivation" to "automorphism." A *difference field* is thus a field K together with a single (specified) automorphism; this is what is called an *inversive difference field* by Cohn [24]. A homogeneous linear *difference equation* with coefficients in K is an equation of the form

$$\phi^n y + a_1 \phi^{n-1} y + \cdots + a_{n-1} y = 0 \quad (a_i \in K),$$

where ϕ is understood to be an automorphism. This notion can be generalized to difference systems (or even to difference modules), just as linear differential equations are generalized to differential systems (and differential modules).

Linear recurrence equations with coefficients, say, in the complex number field $\mathbb{C}$ are regarded as difference equations over the difference field $(\mathbb{C}, \mathrm{id}_{\mathbb{C}})$, whose solutions lie in the $\mathbb{C}$-algebra $\mathbb{C}^{\mathbb{Z}} = \mathrm{Map}(\mathbb{Z}, \mathbb{C})$ of $\mathbb{C}$-valued maps on $\mathbb{Z}$; this is a difference ring with constants $\mathbb{C}$, with respect to the shift operator $\phi\{\alpha_n\}_n = \{\alpha_{n+1}\}_n$, where $\{\alpha_n\}_n \in \mathbb{C}^{\mathbb{Z}}$.

EXAMPLE 10.1. Consider the Fibonacci recurrence equation

$$\alpha_{n+2} - \alpha_{n+1} - \alpha_n = 0. \tag{10.2}$$

This is regarded as a difference equation by setting $\phi(\alpha_n) = \alpha_{n+1}$. As is well-known, its solutions in $\mathbb{C}^{\mathbb{Z}}$ form a $\mathbb{C}$-vector space with basis $\alpha = \{(\frac{1+\sqrt{5}}{2})^n\}_n$, $\beta = \{(-\frac{2}{1+\sqrt{5}})^n\}_n$. Let A be the difference $\mathbb{C}$-subalgebra of $\mathbb{C}^{\mathbb{Z}}$ generated by α, β. It contains the nontrivial idempotent $\frac{1}{2}(\alpha\beta + 1) = \{\ldots, 1, 0, 1, 0, \ldots\}$, as though we could say that (10.2) splits in A. This suggests that the difference *fields* could not be adequate to work with when we develop difference PV theory. We remark that $\alpha \mapsto (t, \frac{1}{t})$, $\beta \mapsto (\frac{1}{t}, -t)$ gives an embedding of A into the difference $\mathbb{C}$-algebra $\mathbb{C}(t) \times \mathbb{C}(t)$ with the automorphism $\phi(t, 0) = (0, \frac{1+\sqrt{5}}{2t})$, $\phi(0, t) = (\frac{2}{(1+\sqrt{5})t}, 0)$. Precisely as in this example, the product of copies of fields equipped with an automorphism that acts transitively on the primitive idempotents will turn out to be a desired object.

To extend the framework so as to include difference PV theory as well, we have to replace, on the one hand, the C [or rather $T(C^+)$] in Part II by some D that contains grouplikes, and on the other hand, replace dif- or C-ferential fields by some generalized object. These are precisely the *AS D-module algebras* as will be defined in 11.6, while D is as follows.

NOTATION 10.3. *D denotes a pointed cocommutative Hopf algebra, which may not be commutative. Let $G = G(D)$ denote the group of all grouplikes in D, and let D^1 denote the largest irreducible Hopf subalgebra in D. The multiplication $d \otimes g \mapsto dg$ gives an isomorphism $D^1 \otimes RG \xrightarrow{\simeq} D$ of coalgebras. Since D^1 is stable under G-conjugation, we have*

$$D = D^1 \# RG.$$

If the C as in Part II is pointed, then $T(C^+)$ can be D. We will later impose some additional assumption on D^1; see 11.4.

The left D-modules form a symmetric tensor category, $({}_D\mathcal{M}, \otimes, R)$. Given $V \in {}_D\mathcal{M}$, we define the subspace of *D-invariants* by

$$V^D = \{v \in V \mid dv = \varepsilon(d)v \quad (d \in D)\}.$$

This notation is used, instead of V_0, since D will possibly change in the sequel. Recall that a *D-module algebra* is a commutative algebra in $({}_D\mathcal{M}, \otimes, R)$. The result 8.3 generalizes as follows.

PROPOSITION 10.4. *Let A be a D-module algebra, and let $T \subset A$ be a G-stable multiplicative set. Then the D-module algebra structure on A uniquely extends to such a structure on $T^{-1}A$.*

PROOF. By [14, Lemma 5.2.10] due to Takeuchi, the algebra map $\rho : A \to \mathrm{Hom}(D, A)$, $a \mapsto (d \mapsto da)$ uniquely extends to an algebra map $\tilde{\rho} : T^{-1}A \to \mathrm{Hom}(D, T^{-1}A)$, since in $\mathrm{Hom}(D, T^{-1}A)$, each $\rho(t)$ $(t \in T)$ is invertible on the coradical RG of D. The associated D-action $d(a/t) = \tilde{\rho}(a/t)(d)$ makes $T^{-1}A$ into a D-module, since the two maps $D \otimes D \to T^{-1}A$ given by $c \otimes d \mapsto (cd)(1/t)$ and $c \otimes d \mapsto c(d(1/t))$ coincide, both being inverse to $c \otimes d \mapsto cdt$. □

A D-module algebra $A \neq 0$ is *simple* if it includes no nontrivial D-stable ideal or in other words if it is simple as an $A\#D$-module. By modifying 6.7 in the obvious way we have the simplicity criterion of a D-module algebra; see also Proposition 12.5. We have especially that if A is simple, A^D is a field.

11. AS D-module algebras

Let $G_1 \subset G$ be a subgroup of finite index. Define a Hopf subalgebra of $D = D^1\#RG$ by

$$D(G_1) = D^1\#RG_1.$$

We will identify the finite set G/G_1 of left cosets with some fixed system of their representatives in G. By convention, we choose 1 as one representative. Set $D_1 = D(G_1)$, and let V be a left D_1-module. We denote the left D-module $D \otimes_{D_1} V$ by $\Phi_{G_1}(V)$ or by $\Phi(V)$. This is identified as

$$\Phi(V) = \bigoplus_{g \in G/G_1} g \otimes V,$$

where by $g \in G/G_1$, we mean that g runs over the fixed representatives for G/G_1. D acts the right hand side so that if $d \in D^1$,

$$d(g \otimes v) = g \otimes g^{-1}dgv \qquad (v \in V)$$

and if $h \in G$,

$$h(g \otimes v) = g' \otimes h'v \qquad (v \in V),$$

where $g' \in G/G_1$, $h' \in G_1$, such that $hg = g'h'$.

LEMMA 11.1. *A natural isomorphism $V^{D_1} \xrightarrow{\simeq} \Phi(V)^D$ is given by $v \mapsto \sum_{g \in G/G_1} g \otimes v$.*

Keep the notation as above. Let A_1 be a D_1-module algebra. Then, $\Phi(A_1)$ is a D-module algebra with respect to the "component-wise" product

$$(g \otimes a)(g' \otimes a') = \delta_{g,g'} g \otimes aa' \qquad (a, a' \in A_1).$$

Thus the $g \otimes 1$ ($g \in G/G_1$) are orthogonal idempotents in $\Phi(A_1)$ whose sum is 1. Set $A = \Phi(A_1)$. Note that if V is a left $A_1 \# D_1$-module, then $\Phi(V)$ is a left $A \# D$-module with respect to the component-wise A-action.

PROPOSITION 11.2. *$\Phi_{G_1} : {}_{A_1\#D_1}\mathcal{M} \to {}_{A\#D}\mathcal{M}$ is a category equivalence.*

The A_1-bimodules in $({}_{D_1}\mathcal{M}, \otimes, R)$ form a tensor category, $({}_{A_1}({}_{D_1}\mathcal{M})_{A_1}, \otimes_{A_1}, A_1)$. Similarly, we have $({}_A({}_D\mathcal{M})_A, \otimes_A, A)$.

PROPOSITION 11.3. *$\Phi_{G_1} : ({}_{A_1}({}_{D_1}\mathcal{M})_{A_1}, \otimes_{A_1}, A_1) \to ({}_A({}_D\mathcal{M})_A, \otimes_A, A)$ gives a tensor equivalence onto the full tensor subcategory of ${}_A({}_D\mathcal{M})_A$, which consists of the objects M such that $(g \otimes 1)m = m(g \otimes 1) (m \in M)$ for every (or equivalently, an arbitrarily chosen) $g \in G/G_1$.*

Subsequently, we assume the following.

ASSUMPTION 11.4. *D^1 is Birkhoff-Witt.*

This is equivalent to saying that D is smooth as a cocommutative coalgebra. Since 11.4 is the same as Assumption 9.1 if $D = T(C^+)$, the results in Section 9 can be recovered from the forthcoming results by taking $D = D^1 = T(C^+)$.

Let K be a D-module algebra. Suppose that K is Noetherian as a ring. Let $\Omega(K)$ denote the (finite) set of all minimal prime ideals in K. Then G acts on $\Omega(K)$. For $P \in \Omega(K)$, let G_P denote the subgroup of stabilizers of P. Set $G_{\Omega(K)} = \bigcap_{P \in \Omega(K)} G_P$; this is a normal subgroup in G of finite index.

PROPOSITION 11.5. *Suppose in addition that K is simple.*

(1) *The action of G on $\Omega(K)$ is transitive so that the subgroups G_P [$P \in \Omega(K)$] are conjugate to each other.*
(2) *Every $P \in \Omega(K)$ is D^1-stable so that K/P is a $D(G_P)$-module domain. This is simple as a $D(G_{\Omega(K)})$-module algebra.*
(3) *Let $P \in \Omega(K)$ and set $K_1 = K/P$. Then, we have a natural isomorphism $K \simeq \Phi_{G_P}(K_1)$ of D-module algebras.*

PROOF. (2) Just as in the proof of 9.3, we see that $\mathrm{Hom}(D^1, P)$ is a prime in $\mathrm{Hom}(D^1, K)$, and $Q := \rho^{-1}(\mathrm{Hom}(D^1, P))$ is a D^1-stable prime in K, which is included in P. By its minimality, P must equal Q and be D^1-stable. For the second assertion, let $P \subset J \subsetneqq K$ be a $D(G_{\Omega(K)})$-stable ideal. Then, $\bigcap_{g \in G/G_{\Omega(K)}} gJ$ is D-stable and hence is zero. There exists a g such that $gJ \subset P$; this implies $P \subset J \subset g^{-1}P$, whence $P = J$ by the minimality of $g^{-1}P$.

(1) If $P \in \Omega(K)$, then $\bigcap_{g\in G} gP$, being D-stable, must be zero. Hence, $\{gP \mid g \in G\} = \Omega(K)$, as desired.

(3) By (1), $g \mapsto gP$ gives a bijection $G/G_P \xrightarrow{\sim} \Omega(K)$. If $g \neq g'$ in G/G_P, then $gP + g'P = K$ by (2). Hence, $K \xrightarrow{\sim} \prod_{g\in G/G_P} K/gP = \bigoplus_{g\in G/G_P} g \otimes K_1 = \Phi_{G_P}(K_1)$, as desired. □

A ring K is said to be *total* if every nonzero-divisor in K is invertible. In Part III, we let $Q(K)$ stand for the *total quotient ring* of K; this is by definition the localization of K by the multiplicative set of all nonzero-divisors. Hence, K is total if and only if $K = Q(K)$.

LEMMA-DEFINITION 11.6. Let $K \neq 0$ be a D-module algebra. The following are equivalent:

(i) K is Artinian and simple (as a D-module algebra);
(ii) K is Noetherian, total, and simple.

If these conditions are satisfied, K is said to be *AS*. In this case, every $K\#D$-module is free as a K-module.

PROOF. Suppose that K is Noetherian and simple. By 11.5, $K = \prod_{P\in\Omega(K)} K/P$. Both (i) and (ii) are equivalent to the statement that each K/P is a field. Choose $P \in \Omega(K)$ and set $K_1 = K/P$. By 11.3 and 11.5 (3), every $K\#D$-module is of the form $\Phi_{G_P}(V) = \bigoplus_{g\in G/G_P} g \otimes V$, where V is a $K_1\#D(G_P)$-module; this implies the last assertion. □

REMARK 11.7. A D-module that is a field is obviously AS; it will be called a *D-module field*. A C-ferential field as defined in Part II is precisely a $T(C^+)$-module field, where for Assumption 11.4, we need to assume 9.1. On the other hand, a D-module field is necessarily a D-ferential field.

LEMMA 11.8. *Let $G_1 \subset G$ be a subgroup of finite index. A $D(G_1)$-module algebra K_1 is AS if and only if the D-module algebra $\Phi_{G_1}(K_1)$ is AS.*

LEMMA 11.9. *Let L be an AS D-module algebra, and let $A \subset L$ be a D-module subalgebra.*

(1) *Every nonzero-divisor of A is a nonzero-divisor of L so that the total quotient ring $Q(A)$ is realized in L.*
(2) *This $Q(A)$ is a D-module subalgebra of L, which is AS.*

PROOF. (1) Given $0 \neq x \in L$, define its support by

$$\mathrm{Supp}(x) = \{P \in \Omega(L) \mid x \notin P\}.$$

Note that x is a nonzero-divisor of L if and only if $\mathrm{Supp}(x) = \Omega(L)$. Let x be a nonzero-divisor of A whose support is minimal among the nonzero-divisors of A. If

$\mathrm{Supp}(x) \neq \Omega(L)$, then by 11.5 (1) and the minimality, there exists a $g \in G$ such that $\mathrm{Supp}(x) \cap \mathrm{Supp}(gx) = \emptyset$; this implies $x(gx) = 0$, a contradiction. (2) By 10.4, $Q(A)$ is a D-module subalgebra of L. It is necessarily AS, since a total D-module subalgebra of an AS D-module algebra is AS, as is proved in [11, Lemma 2.8]. □

12. PV extensions of AS D-module algebras

Let $L \supset K$ be an inclusion of AS D-module algebras; it is also referred to as an extension L/K. There is the corresponding field extension L^D/K^D.

Definition 12.1. We say that L/K is a *PV extension* if

(i) $L^D = K^D$, which is a field and will be denoted by k, and

(ii) there exists an intermediate D-module algebra $L \supset A \supset K$ such that

(a) $L = Q(A)$ [see 11.9 (2)],

(b) $A \otimes_K A$ is generated by $(A \otimes_K A)^D$ as a left (or equivalently right) A-module.

Such an A as above is called a *principal D-module algebra* for L/K.

Let L/K and A be as above. Set $H = (A \otimes_K A)^D$; this is a k-subalgebra of $A \otimes_K A$. By using the simplicity of L, it follows that $\mu : A \otimes_k H \to A \otimes_K A$, $\mu(a \otimes_k h) = (a \otimes_K 1) \cdot h$ is an isomorphism. Then, we can now reprove Lemmas 7.2 and 7.3 in the present context. Thus, H has a natural structure of commutative k-Hopf algebra; it is called the *PV Hopf algebra* for L/K. A principal D-module algebra A for L/K is unique; it has a natural structure $\theta : A \to A \otimes_k H$ of a right H-comodule algebra for which A/K is an H-Galois extension. We say that $(L/K, A, H)$ is a PV extension, indicating A, H as well.

Example 12.2. Let $G_1 \subset G$ be a normal subgroup of finite index. Let K be a D-module field. Regarding this as a $D(G_1)$-module field, define $L := \Phi_{G_1}(K)$. There is an inclusion of D-module algebras,

$$K \hookrightarrow L = \bigoplus_{g \in G/G_1} g \otimes K, \quad x \mapsto \sum_g g \otimes g^{-1}x.$$

If $K^{D(G_1)} = K^D$, then $K^D = L^D$ (=: k) by 11.1. Moreover, $(L/K, L, H)$ is a PV extension, where $H = k(G/G_1)^*$. In fact, we see that the elements

$$e_g := \sum_{h \in G/G_1} (h \otimes 1) \otimes_K (hg \otimes 1) \qquad (g \in G/G_1)$$

in $L \otimes_K L$ are D-invariant and behave as the dual basis of the group elements g in $k(G/G_1)$. Thus, $\Delta(e_g) = \sum_h e_{gh^{-1}} \otimes_k e_h$, $\varepsilon(e_g) = \delta_{1,g}$, $S(e_g) = e_{g^{-1}}$. The H-comodule structure $\theta : L \to L \otimes_k H$ is given by $\theta(h \otimes x) = \sum_g (hg^{-1} \otimes gx) \otimes_k e_g$.

LEMMA 12.3. *Let $G_1 \subset G$ be a subgroup of finite index, and write $\Phi = \Phi_{G_1}$. Let L_1/K_1 be an extension of AS $D(G_1)$-module algebras; $\Phi(L_1)/\Phi(K_1)$ is then also an extension of AS D-module algebras. $(L_1/K_1, A_1, H)$ is a PV extension if and only if $(\Phi(L_1)/\Phi(K_1), \Phi(A_1), H)$ is a PV extension.*

PROOF. The natural coalgebra isomorphism $\Phi(A_1 \otimes_{K_1} A_1) \simeq \Phi(A_1) \otimes_{\Phi(K_1)} \Phi(A_1)$ (see 11.3) together with 11.1 proves the lemma. □

REMARK 12.4. Let L/K be an extension of AS D-module algebras. Choose $\mathfrak{p} \in \Omega(K)$, and let $P_1, \ldots, P_r$ be those primes in $\Omega(L)$, which lie over $\mathfrak{p}$. Define $K_1 = K/\mathfrak{p}$, $L_1 = \prod_{i=1}^{r} L/P_i$. Then, we have an extension L_1/K_1 of AS $D(G_\mathfrak{p})$-module algebras such that the induced $\Phi(L_1)/\Phi(K_1)$ is identified with L/K. Replacing L/K with L_1/K_1, we can often suppose that K is a field when discussing PV extensions; see 12.3.

To give a general result, which holds in an arbitrary Abelian category $\mathcal{A}$, fix an object X in $\mathcal{A}$, and set $E = \mathcal{A}(X, X)$, the endomorphism ring of X. Note that for every object Y in $\mathcal{A}$, $\mathcal{A}(X, Y)$ is naturally a right E-module.

PROPOSITION 12.5 (Simplicity criterion). *X is simple (i.e., is nonzero and includes no nontrivial subobject) if and only if*

(a) *E is a division ring, and*
(b) *for every object Y in $\mathcal{A}$, the morphism*

$$(f_1, \ldots, f_n) : X^n = X \oplus \cdots \oplus X \to Y$$

given by finitely many, E-linearly independent morphisms $f_1, \ldots, f_n$ in $\mathcal{A}(X, Y)$ is a monomorphism.

PROOF. "If". Let $Z \subsetneqq X$ be a proper subobject. The quotient map f in $\mathcal{A}(X, X/Z)$, being nonzero, is E-linearly independent by (a). By (b), $f : X \to X/Z$ is monic, and so $Z = 0$.

"Only if". The argument that proves Schur's lemma shows (a). We prove (b) by induction on the number n. The proof in case $n = 1$ is obvious. Let $n > 1$, and assume contrary to (b) that $(f_1, \ldots, f_n)$ is not monic, while $f_1, \ldots, f_n$ are E-linearly independent. Set $g = (f_1, \ldots, f_{n-1})$; this g as well as f_n is monic by the induction hypothesis. Let

$$\begin{array}{ccc} X^{n-1} & \xrightarrow{g} & Y \\ \uparrow & & \uparrow{\scriptstyle f_n} \\ Z & \longrightarrow & X \end{array}$$

be a pull-back diagram, which consists of monomorphisms. By the assumption above, $Z \neq 0$, and so $Z \xrightarrow{\sim} X$ by the simplicity of X. The induced morphism $X \simeq Z \to$

X^{n-1} is of the form ${}^t(a_1, \dots, a_{n-1})$ with $a_i \in E$. This implies $f_n = \sum_{i=1}^{n-1} f_i a_i$, contradicting the E-linear independence. □

Note that Propositions 1.3, 3.2, and 6.7 can all be obtained as special cases of the proposition just proved.

THEOREM 12.6 (Galois correspondence). *Let $(L/K, A, H)$ be a PV extension of AS D-module algebras.*

(1) *There is a 1–1 correspondence, given by (2.7), between the Hopf ideals I in H and the intermediate AS D-module algebras M in L/K.*
(2) *If $I \leftrightarrow M$ under the correspondence, $(L/M, AM, H/I)$ is a PV extension.*
(3) *If $I \leftrightarrow M$ under the correspondence, the Hopf ideal I is normal if and only if M/K is a PV extension. By [18, Theorem 4.3], this gives rise to a 1–1 correspondence between the Hopf subalgebras J in H and the intermediate PV extensions M/K in L/K, in which to each M/K its PV Hopf algebra is assigned.*

PROOF. Once Part 1 is proved, Parts 2 and 3 will follow in the same way as for 2.6 (2) and 2.8. For Part 1, we need to prove 2.1 and 2.3 in the present, generalized situation. For 2.3, this is easy. For 2.1, suppose that M is given. Since L, being an $M\#D$-module, is M-free, M can be recovered from the corresponding D-stable coideal $\mathrm{Ker}(L \otimes_K L \to L \otimes_M L)$ in the coalgebra $L \otimes_K L$ in $({}_L({}_D\mathcal{M})_L, \otimes_L, L)$. Conversely, let $\mathfrak{a} \subset L \otimes_K L$ be a D-stable coideal. Define $M = M_{\mathfrak{a}}$ by 2.2; this is obviously an intermediate D-module algebra, which is seen to be total, and so AS by 11.9 (2). Set $\mathcal{C} = L \otimes_K L/\mathfrak{a}$. One sees that the quotient map $L \otimes_K L \twoheadrightarrow \mathcal{C}$ factors through a coalgebra surjection $\gamma : L \otimes_M L \twoheadrightarrow \mathcal{C}$. We need to prove that γ is injective. By 11.3, we may suppose that M is a field. To apply 12.5, regard $\mathcal{C}$ just as an L-coring or a coalgebra in $({}_L\mathcal{M}_L, \otimes_L, L)$, and suppose that $\mathcal{A}$ is the category of right $\mathcal{C}$-comodules; an object Y in $\mathcal{A}$ is thus a right L-module together with an L-linear comodule structure map $Y \to Y \otimes_L \mathcal{C}$. Take L as the X in 12.5; it has the natural $\mathcal{C}$-comodule structure $\lambda : L \to L \otimes_L \mathcal{C} = \mathcal{C}$, $\lambda(x) \equiv 1 \otimes_K x$ mod $\mathfrak{a}$. Since $E = \mathcal{A}(L, L) \simeq M$, $\mathcal{A}(L, \mathcal{C}) \simeq L$, Condition (b) in 12.5, applied to $Y = \mathcal{C}$, is equivalent to injectivity of γ. Therefore, it suffices to verify the assumption of 12.5 that L is simple in $\mathcal{A}$. A subobject of L is of the form eL, where e is an idempotent. Since λ is D-linear, $g(eL)$, where $g \in G$, is also a subobject so that L is semisimple by 11.5 (1). However, L must be simple, since the endomorphism ring E is a field. □

Just as in the previous parts I and II, the result above can be interpreted in terms of group schemes.

COROLLARY 12.7. *Let $(L/K, A, H)$ be a PV extension of AS D-module algebras.*

(1) *A is simple as a D-module algebra.*
(2) *A contains all primitive idempotents in L so that $A = \prod_{P \in \Omega(L)} A/P \cap A$.*

PROOF. (1) See the proof of 2.5. (2) Since $A \hookrightarrow L$ is a localization, we have $\Omega(A) \supset \Omega(L)$. We need to prove that if $P \neq Q$ in $\Omega(L)$, the sum $J := P \cap A + Q \cap A$ equals A. If on the contrary $J \subsetneqq A$, one sees from (1), as in the proof of 11.5 (2), that $J = P \cap A = Q \cap A$, which implies $P = Q$. □

PROPOSITION 12.8. *Let $(L/K, A, H)$ be a PV extension of AS D-module algebras. Choose $P \in \Omega(L)$ arbitrarily, and write $\Phi = \Phi_{G_P}$. Set $\mathfrak{p} = P \cap K$ $(\in \Omega(K))$, and define $K_1 = K/\mathfrak{p}$, $A_1 = A/P \cap A$, $L_1 = L/P$. Then*

(1) $A \simeq \Phi(A_1)$.
(2) *$\Phi(K_1)$ identifies with the K-subalgebra of L, say $\hat{K}$, which is K-spanned by the primitive idempotents in L.*
(3) *$(L_1/K_1, A_1, \bar{H} := H/I)$ is a PV extension of $D(G_P)$-module fields, where $I = H \cap \mathrm{Ker}(L \otimes_K L \to L \otimes_{\hat{K}} L)$, the Hopf ideal corresponding to $\hat{K}$.*
(4) *The subalgebra of H*

$$B := \{h \in H \mid \Delta(h) - h \otimes 1 \in H \otimes I\}$$

is a separable k-algebra of dimension $[G_\mathfrak{p} : G_P]$. There is a right $\bar{H}$-colinear B-algebra isomorphism $H \simeq B \otimes_k \bar{H}$.
(5) *The subgroup $G_P \subset G_\mathfrak{p}$ is normal if and only if B splits in the sense of $B \simeq k \times \cdots \times k$. In this case, $B \subset H$ is a Hopf subalgebra isomorphic to $k(G_\mathfrak{p}/G_P)^*$ so that there is a short exact sequence of k-Hopf algebras, $k(G_\mathfrak{p}/G_P)^* \rightarrowtail H \twoheadrightarrow \bar{H}$.*

PROOF. (1) This follows by 12.7 (2). (2) This is easy. (3) By 12.6 (2), we have a PV extension, $(L/\hat{K}, A, \bar{H}) = (\Phi(L_1)/\Phi(K_1), \Phi(A_1), \bar{H})$. Part 3 now follows from 12.3.

(4) and (5). By 12.3, we may suppose that K is a field, and so $\mathfrak{p} = 0$, $G = G_\mathfrak{p}$. The obvious equalizer diagram $0 \to A \otimes_K \hat{K} \to A \otimes_K A \overset{\to}{\to} A \otimes_K A \otimes_{\hat{K}} A$ of D-module algebras is naturally identified with $0 \to A \otimes_k B \to A \otimes_k H \overset{\to}{\to} A \otimes_k H \otimes_k \bar{H}$. Induced is an isomorphism $A \otimes_k B \simeq A \otimes_K \hat{K}$. By applying $L_1 \otimes_A$, we obtain

$$L_1 \otimes_k B \simeq L_1 \otimes_K \hat{K} = \Phi(L_1) = L. \tag{12.1}$$

This implies the first assertion of (4). For the second, note that $\sigma : \Phi(A_1 \otimes_K A_1) = A \otimes_{\hat{K}} A \to A \otimes_K A$ given by $\sigma(g \otimes (a \otimes_K b)) = (g \otimes a) \otimes_K (g \otimes b)$ $(g \in G/G_P)$ is a D-linear, right $\bar{H}$-colinear k-algebra splitting of $A \otimes_K A \to A \otimes_{\hat{K}} A$. The restriction $\sigma^D : \bar{H} \to H$ to D-invariants is an $\bar{H}$-colinear k-algebra splitting of $H \to \bar{H}$. It follows from [25, Theorem 9] (or see [14, Theorem 7.2.2]) that $B \otimes_k \bar{H} \to H$, $b \otimes x \mapsto b\sigma^D(x)$ is the desired isomorphism.

If $G_P \lhd G$, then by 12.2, $\hat{K}/K$ is a PV extension with PV Hopf algebra $k(G/G_P)^*$. We must have $B = k(G/G_P)^*$, which splits. If B splits, it follows from the isomorphism (12.1), which preserves the G_P-action, that G_P stabilizes each component in L, whence $G_P \lhd G$. This proves the first assertion of (5). The second one is easy. □

THEOREM 12.10. *Let L/K be an extension of AS D-module algebras. Choose an arbitrary $P \in \Omega(L)$, and set $\mathfrak{p} = P \cap K$ $(\in \Omega(K))$. Set $K_1 = K/\mathfrak{p}$, $L_1 = L/P$. Then L/K is a PV extension if*

(i) *G_P is normal in $G_\mathfrak{p}$, and*
(ii) *the extension L_1/K_1 of $D(G_P)$-module fields is PV.*

The converse is true if the field K^D $(= L^D)$ is separably closed.

PROOF. If K^D is separably closed and L/K is PV, then the separable algebra B as given in 12.8 (4) splits, which implies (i) above, by 12.8 (5). This together with 12.8 (3) proves the last assertion.

For what remains, we can suppose that K is a field. Suppose that $(L_1/K_1, A_1, \bar{H})$ is PV. Let $\Phi = \Phi_{G_P}$, and define $A = \Phi(A_1)$. If L/K is PV, the principal D-module algebra must be A; see 12.8 (1). $A \otimes_K A$ is a right $\bar{H}$-comodule algebra with $\bar{H}$-coinvariants $A \otimes_K \Phi(K)$. We can define a D-linear, $\bar{H}$-colinear k-algebra map $\sigma : \Phi(A_1 \otimes_K A_1) \to A \otimes_K A$ as in the last proof. Induced is an $\bar{H}$-colinear k-algebra map $\sigma^D : \bar{H} \to (A \otimes_K A)^D$. Again by [25, Theorem 9], we have a D-linear, $\bar{H}$-colinear algebra isomorphism $A \otimes_K \Phi(K) \otimes_k \bar{H} \xrightarrow{\simeq} A \otimes_K A$ over $A \otimes_K \Phi(K)$. Hence, L/K is PV if and only if the natural injection $A \otimes_k (A \otimes_k \Phi(K))^D \to A \otimes_k \Phi(K)$ is surjective. Indeed, this is surjective if $G_P \lhd G$, since then by 12.2, $A \otimes_k (\Phi(K) \otimes_K \Phi(K))^D \to A \otimes_k \Phi(K)$ is already surjective. □

There is an example [11, Example 3.16] of a PV extension of the form $\Phi_{G_1}(K)/K$, where K is a D-module field, and $G_1 \subset G$ is a non-normal subgroup of finite index. This shows that, in 12.10, the converse does not necessarily hold without the assumption of separable closedness.

13. Characterization and unique existence

Let L/K be an extension of AS D-module algebras. Let V be a $K\#D$-module, which is necessarily K-free; see 11.6. Its K-free rank will be denoted by $\mathrm{rk}_K V$ and simply called the *K-rank*. By slightly modifying the argument in Section 8, we see that $\mathrm{Hom}_K(V, L)$ is naturally an $L\#D$-module [see (8.6)], whose D-invariants equal the L^D-vector space

$$L^V := \mathrm{Hom}_{K\#D}(V, L).$$

There is a natural injection $L \otimes_{L^D} L^V \hookrightarrow \mathrm{Hom}_K(V, L)$ in ${}_{L\#D}\mathcal{M}$, which necessarily splits L-linearly so that $\dim_{L^D} L^V \leq \mathrm{rk}_K V$.

DEFINITION 13.1. Let V be a $K\#D$-module. L is a *splitting algebra* for V if there exists an injection $L \otimes_K V \hookrightarrow \prod_\lambda L_\lambda$ in ${}_{L\#D}\mathcal{M}$, where $\prod_\lambda L_\lambda$ is the product of copies L_λ of L. If $\mathrm{rk}_K V = n < \infty$, the condition is equivalent to each of the following: $L \otimes_K V \simeq L^n$ in ${}_{L\#D}\mathcal{M}$; $L \otimes_{L^D} L^V \xrightarrow{\simeq} \mathrm{Hom}_K(V, L)$; $\dim_{L^D} L^V = n$.

For a subset $Z = \{z_1, z_2, \dots\}$ of L, let $K\langle Z\rangle = K\langle z_1, z_2, \dots\rangle$ denote the smallest intermediate AS D-module algebra in L/K that includes Z. This coincides with the localization $T^{-1}A$ of the K-subalgebra $A \subset L$ generated by the D-submodule DZ, where T is the multiplicative set consisting of those elements x in A with full support, i.e., $\mathrm{Supp}(x) = \Omega(L)$; see the proof of 11.9. L/K is said to be *finitely generated* if $L = K\langle z_1, \dots, z_n\rangle$ for some finite elements $z_1, \dots, z_n$ in L. If L/K is PV, this, as will be seen from 13.3, is equivalent to the following: L is the total quotient ring of some finitely generated K-subalgebra in L.

Let $K\langle V\rangle$ stand for $K\langle\bigcup_{f\in L^V} f(V)\rangle$. A splitting algebra L for V is said to be *minimal* if $L = K\langle V\rangle$.

LEMMA 13.2. *Let $G_1 \subset G$, $K_1 \subset L_1$ be as in 12.3. Write $\Phi = \Phi_{G_1}$. Then, L_1 is a (minimal) splitting algebra for a $K_1\#D(G_1)$-module V_1 if and only if $\Phi(L_1)$ is a (minimal) splitting algebra for the $\Phi(K_1)\#D$-module $\Phi(V_1)$.*

PROOF. This follows from 11.2; note that $\Phi(K_1\langle V_1\rangle) = \Phi(K_1)\langle\Phi(V_1)\rangle$. □

THEOREM 13.3 (Characterization). *Let L/K be an extension of AS D-module algebras such that $L^D = K^D$.*

(1) *The following are equivalent:*
   - (i) *L/K is a finitely generated PV extension.*
   - (ii) *L is a minimal splitting algebra for some cyclic $K\#D$-module of finite K-rank.*
   - (iii) *L is a minimal splitting algebra for some $K\#D$-module of finite K-rank.*
   - (iv) *There is a GL_n-primitive matrix $X = (x_{ij}) \in GL_n(L)$ for some n such that $L = K\langle x_{ij}\rangle$; by saying that X is GL_n-primitive, we mean that for every $d \in D$, $(dX)X^{-1} \in M_n(K)$, where $dX = (dx_{ij})$.*

(2) *The following are equivalent:*
   - (i) *L/K is a PV extension.*
   - (ii) *L is a minimal splitting algebra for some $K\#D$-module that is a sum of $K\#D$-submodules of finite K-rank.*

PROOF. (1) We prove only the implication (i) $\Rightarrow$ (ii). For the rest just adapt the proofs of 3.11 and 8.8. Suppose that $(L/K, A, H)$ is finitely generated PV. By 12.3 and 13.2, we can suppose that K is a field. By 12.8 (3), we have a finitely generated PV extension $(L_1/K_1, A_1, \tilde{H})$ of module fields over $D_1 = D(G_P)$ with $P \in \Omega(L)$ such that $L = \Phi_{G_P}(L_1)$, $A = \Phi_{G_P}(A_1)$. Write $k = K^D$ $(= L^D)$.

There exist finitely many elements $u_1, \dots, u_n$ in $A = (A, \theta)$ such that they k-span an H-subcomodule and satisfy $L = K\langle u_1, \dots, u_n\rangle$. Take the element $\mathbf{u} = (u_1, \dots, u_n)$ in A^n, and let $V = (K\#D)\mathbf{u}$, the cyclic $K\#D$-module generated by $\mathbf{u}$. Since $L\otimes_K A \simeq L\otimes_k H$, we see that L is a minimal splitting algebra for A^n and hence for V. It remains to prove $\dim_K V < \infty$. For this, it suffices to prove that the natural image $V(P)$, say, of V under the projection $A^n \to A_1^n$ has finite K-dimension since $V \hookrightarrow \prod_{P\in\Omega(L)} V(P)$. Let $g_1, \dots, g_s$ $(\in G)$ be a system of representatives for $G_P\backslash G$.

Then, $V = \sum_{i=1}^{s} (K\#D_1)g_i\mathbf{u}$. Fix $1 \leq i \leq s$, and let $\mathbf{w} = (w_1, \ldots, w_n) \in A_1^n$ be the natural image of $g_i\mathbf{u}$. It suffices to prove that $W := (K\#D_1)\mathbf{w}$ has finite K-dimension. By renumbering, we suppose that $w_1, \ldots, w_r$ ($r \leq n$) are a k-basis of the k-subspace of A_1 spanned by $w_1, \ldots, w_n$. Set $\mathbf{w}' = (w_1, \ldots, w_r)$. Then, $\mathbf{w} = \mathbf{w}'T$ for some rank r matrix T with entries in k. It suffices to prove that $W' := (K\#D_1)\mathbf{w}'$ has finite K-dimension, since $W' \xrightarrow{\simeq} W$ via the right multiplication by T, but this follows by an obvious modification of the proof of 8.8, (i) $\Rightarrow$ (ii), since $w_1, \ldots, w_r$ form a k-basis of an $\bar{H}$-subcomodule of A_1^n.

(2) We can reprove 3.15 (1) in the present context. The result, together with Part 1 above and 12.6 (3), proves Part 2; see the proof of 3.17. □

We can now reprove Theorem 8.11 in the present context; see [11, Theorem 4.10].

THEOREM 13.4 (Unique existence). *Suppose that K is an AS D-module algebra such that the field K^D is algebraically closed. Let V be a $K\#D$-module of finite K-rank. There exists a minimal splitting algebra L for V with $L^D = K^D$ so that L/K is necessarily a finitely generated PV extension of AS D-module algebras. Such an L is unique up to isomorphism of D-module algebras over K.*

PROOF. Modify the proof of 4.3 by using the following lemma, which generalizes 4.2. □

LEMMA 13.5 (cf. [21, Appendix]). *Let A/K be an extension of D-module algebra such that A is finitely generated as a K-algebra. Suppose that K is AS, and A is simple; then by 11.5 (3), $A \simeq \Phi_{G_P}(A/P)$, where $P \in \Omega(A)$. Let $L = Q(A)$; by 10.4, L is uniquely a D-module algebra, which is AS. Then the field extension L^D/K^D is algebraic.*

PROOF. As in the first part of the proof of 4.2, one sees $L^D = A^D$. The remaining parts are also valid, since by 11.1 and 11.5 (3), we can suppose that K is a field, and A is a domain, by replacing K, A with $K/P \cap K$, A/P, where $P \in \Omega(A)$. □

We can rewrite Section 5 in the present context. We only point out some aspects; see [13] for details. Let L/K be an extension of AS D-module algebras. Recall the remark on the Amitsur cohomology vanishing just after Lemma 1.14. One then sees that the lemma holds for L/K, since K is the product of fields, and L is K-free. Using the lemma we can restate Example 1.15. To proceed, understand K^{s} to denote $\Phi(K_1^{\mathrm{s}})$ with the notation of 12.8, where K_1^{s} denotes the separable algebraic closure of K_1 in the field extension L_1/K_1.

References

[1] I. Kaplansky, *"An introduction to differential algebra"*, Second ed., Hermann, Paris, 1976.

[2] E.R. Kolchin, *"Differential algebra and algebraic groups"*, Pure Appl. Math., vol. **54**, Academic Press, New York-London, 1973.

[3] A.R. Magid, *"Lectures on differential Galois theory"*, Univ. Lec. Series vol. **7**, AMS, Providence, 1994.
[4] M. van der Put, M.F. Singer, *"Galois theory of linear differential equations"*, Grundlehren Math. Wiss., vol. **328**, Springer, Berlin, 2003.
[5] E. R. Kolchin, Picard-Vessiot theory of partial differential fields, Proc. Amer. Math. Soc. **3** (1952) 596–603.
[6] A. Buium, *"Differential function fields and moduli of algebraic varieties"*, Lecture Notes in Math., vol. **1226**, Springer, Berlin, 1986.
[7] K. Okugawa, *Basic properties of differential fields of an arbitrary characteristic and the Picard-Vessiot theory*, J. Math. Kyoto Univ. **2–3** (1963) 295–322.
[8] C.H. Franke, Picard-Vessiot theory of linear homogeneous difference equations, Trans. Amer. Math. Soc. **108** (1963) 491–515.
[9] M. van der Put, M.F. Singer, *"Galois theory of difference equations"*, Lecture Notes in Math. vol. **1666**, Springer, Berlin, 1997.
[10] M. Takeuchi, *A Hopf algebraic approach to the Picard-Vessiot theory*, J. Algebra **122** (1989) 481–509.
[11] K. Amano, A. Masuoka, Picard-Vessiot extensions of artinian simple module algebras, J. Algebra **285** (2005) 743–767.
[12] Y. André, Différentielles non commutatives et théorie de Galois différentielle ou aux différences, Ann. Sci. Éc. Norm. Super. **34** (3) (2001) 685–739; arXiv:math.GM/0203274.
[13] K. Amano, Liouville extensions of artinian simple module algebras, Commun. Algebra **34** (2006) 1811–1823.
[14] S. Montgomery, *"Hopf algebras and their actions on rings"*, CBMS Reg. Conf. Series in Math., vol. **82**, AMS, 1993.
[15] P. Deligne, *Catégories Tannakiennes*, in P. Cartier, et al., (eds.), "Grothendieck Festschrift", vol. **2**, Progress in Math. **87**, Birkhäuser, Boston, 1990, pp. 111–195.
[16] M. Sweedler, *"Hopf algebras"*, Benjamin, New York, 1969.
[17] H.F. Kreimer, M. Takeuchi, Hopf algebras and Galois extensions of an algebra, Indiana Univ. Math. J. **30** (1981) 675–692.
[18] M. Takeuchi, *A correspondance between Hopf ideals and sub-Hopf algebras*, Manuscripta Math. **7** (1972) 251–270.
[19] W.C. Waterhouse, *"Introduction to affine group schemes"*, Grad. Texts in Math., vol. **66**, Springer, New York-Berlin, 1979.
[20] M. Sweedler, *The predual theorem to the Jacobson-Bourbaki theorem*, Trans. Amer. Math. Soc. **213** (1975) 391–406.
[21] A.H.M. Levelt, Differential Galois theory and tensor products, Indag. Math. (N.S.), **1** (4) (1990) 439–450.
[22] Y. Doi, M. Takeuchi, Hopf-Galois extensions of algebras, the Miyashita-Ulbrich action, and Azumaya algebras, J. Algebra **121** (1989) 488–516.
[23] R.G. Heyneman, M. Sweedler, Affine Hopf algebras, II, J. Algebra **16** (1970) 271–297.
[24] R.M. Cohn, *"Difference algebra"*, Interscience Tracts Pure Appl. Math., vol. **17**, Interscience Publ., New York-London-Sydney, 1965.
[25] Y. Doi, M. Takeuchi, Cleft comodule algebras for a bialgebra, Commun. Algebra **14** (1986) 801–818.

Hopf Algebroids

Gabriella Böhm

Research Institute for Particle and Nuclear Physics, Budapest,
H-1525 Budapest 114, P.O.B.49 Hungary
E-mail: G.Bohm@rmki.kfki.hu

Contents

HANDBOOK OF ALGEBRA, VOL. 6
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(08)00205-2

1. Introduction

If it should be formulated in one sentence what a Hopf algebroid is, it should be described as a generalization of a Hopf algebra to a noncommutative base algebra. More precisely, the best known examples of Hopf algebroids are Hopf algebras. Clearly, the notion of a Hopf algebroid turns out to be a successful generalization only if a convincing amount of results about Hopf algebras extend to Hopf algebroids. But one expects more: working with Hopf algebroids should be considered to be useful if in this way one could solve problems that could not be solved in terms of Hopf algebras. Hopf algebroids provide us with results of two types: one that extends known results about Hopf algebras and the other that is conceptually new.

Hopf algebras have been intensively studied and successfully applied in various fields of mathematics and even physics, for more than 50 years. Without aiming at a complete list, let us mention a few applications. Hopf algebras were used to construct invariants in topology and knot theory. In connection with solutions of the quantum Yang–Baxter equation, quantum groups, that is, certain Hopf algebras, play a central role. In (low-dimensional) quantum field theory, Hopf algebras are capable of describing internal symmetry of some models. In noncommutative differential geometry (faithfully flat), Galois extensions by a Hopf algebra are interpreted as noncommutative principal bundles. Although the theory of Hopf algebras was (is!) extremely successful, in the 1990s, there arose more and more motivations for a generalization.

Originally, the term *Hopf algebroid* was used for cogroupoid objects in the category of commutative algebras. These are examples of Hopf algebroids in the current chapter with commutative underlying algebra structure. They found an application, for example, in algebraic topology [1]. As a tool for a study of the geometry of principal fiber bundles with groupoid symmetry, recently, more general, noncommutative Hopf algebroids have been used but still over commutative base algebras [2]. For some applications, this is still not the necessary level of generality. In Poisson geometry, solutions of the dynamical Yang–Baxter equation correspond to dynamical quantum groups, which are not Hopf algebras [3–8]. In topology, invariants obtained in [9] do not fit the Hopf algebraic framework. In transverse geometry, extensions of Hopf algebras by noncommutative base algebras occurred [10]. In low-dimensional quantum field theories, nonintegral values of the statistical (also called quantum) dimensions in some models exclude a Hopf algebra symmetry [11]. Another field where important questions could not be answered in the framework of Hopf algebras is noncommutative geometry, that is, Hopf Galois theory. Thinking about classical Galois extensions of fields by a finite group, such an extension can be characterized without explicitly mentioning the Galois group. A (unique, up to isomorphism) Galois group is determined by a Galois field extension. In the case of Hopf Galois extensions, no such intrinsic characterization, without explicit use of a Hopf algebra, is known. Also, although the Hopf algebra describing the symmetry of a given Hopf Galois extension is known to be nonunique, the relation between the possible choices is not known. These questions have been handled by allowing for noncommutative base algebras [12, 13]. On the other hand, as the study of Hopf algebroids has a quite short

past, there are many aspects of Hopf algebras that have not yet been investigated as to how they extend to Hopf algebroids. It has to be admitted that almost nothing has been done yet toward a classification and structure theory of Hopf algebroids.

What does it mean that the base algebra $\mathbf{R}$ of a Hopf algebroid is noncommutative? Recall that a bialgebra over a commutative base ring k is a k-module, with compatible algebra and coalgebra structures. By analogy, in a bialgebroid, the coalgebra structure is replaced by a coring over any not necessarily commutative k-algebra $\mathbf{R}$. Also, the algebra structure is replaced by a ring over a noncommutative base algebra. However, in order to formulate the compatibility between the ring and coring structures, the base algebra of the ring has to be not $\mathbf{R}$ but $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$. A Hopf algebra is a bialgebra with an additional antipode map. In the Hopf algebroid case, the antipode relates two different bialgebroid structures over the base algebras $\mathbf{R}$ and $\mathbf{R}^{\mathrm{op}}$.

In these notes, we arrive at the notion of a Hopf algebroid after considering all constituent structures. In Section 2, $\mathbf{R}$-rings and $\mathbf{R}$-corings are introduced. They are seen to generalize algebras and coalgebras. Emphasis is put on their duality. Section 3 is devoted to a study of bialgebroids, generalizing bialgebras. Several equivalent descriptions are given and examples are collected. In particular, constructions of new bialgebroids from known ones are presented. Some of them change the base algebra of a bialgebroid, so they have no counterparts for bialgebras. Although bialgebroid axioms are not manifestly self-dual, duals of finitely generated and projective bialgebroids are shown to be bialgebroids. Key properties of a bialgebroid are monoidality of the categories of modules and comodules. This is explained in some detail. Section 3 is closed by a most important and most successful application, Galois theory of bialgebroids. Hopf algebroids are the subject of Section 4. After presenting the definition, listing some examples, and deriving some immediate consequences of the axioms, we discuss the theory of comodules. Since in a Hopf algebroid there are two bialgebroids (hence corings), comodules of the Hopf algebroid comprise comodule structures of both. The relation between the categories of comodules of a Hopf algebroid and comodules of the constituent bialgebroids is investigated. The category of comodules of a Hopf algebroid is proven to be monoidal, what is essential from the point of view of Galois theory. Next, we turn to a study of the theory of integrals. It is a good example of results that are obtained by using some new ideas but that extend analogous results for Hopf algebras in a reassuring way. The structure of Galois extensions by Hopf algebroids is investigated. Useful theorems are presented about situations when surjectivity of a canonical map implies Galois property. They extend known results about Hopf Galois extensions. While there seems to be an accord in the literature that the right generalization of a bialgebra to a noncommutative base is a bialgebroid, there is some discussion about the right generalization of a Hopf algebra. We close these notes by collecting and comparing notions suggested by various authors.

To keep the list of references perspicuous, we do not refer in these notes to papers containing classical results about Hopf algebras, which are generalized here. We believe it is more useful to give here a detailed bibliography of those papers that deal with structures over a noncommutative base. A very good and detailed bibliography of the literature of Hopf algebras can be found in chapter "Hopf Algebras" of *Handbook of Algebra* [14].

Notations and conventions. Throughout k is an associative and commutative unital ring. All algebras are associative and unital k-algebras. A k-algebra is denoted by $\mathbf{A}$, and the underlying k-module is denoted by A. On elements of A, multiplication is denoted by juxtaposition. The unit element is denoted by $1_{\mathbf{A}}$. For an algebra $\mathbf{A}$, with multiplication $(a, a') \mapsto aa'$, $\mathbf{A}^{\mathrm{op}}$ denotes the opposite of $\mathbf{A}$. As a k-module it is equal to A and multiplication is $(a, a') \mapsto a'a$. The category of right (resp. left) modules of an algebra $\mathbf{A}$ is denoted by $\mathcal{M}_{\mathbf{A}}$ (resp. ${}_{\mathbf{A}}\mathcal{M}$). Hom sets are denoted by $\mathrm{Hom}_{\mathbf{A}}(-,-)$ (resp. ${}_{\mathbf{A}}\mathrm{Hom}(-,-)$). The category of $\mathbf{A}$-bimodules is denoted by ${}_{\mathbf{A}}\mathcal{M}_{\mathbf{A}}$, and its hom sets by ${}_{\mathbf{A}}\mathrm{Hom}_{\mathbf{A}}(-,-)$. Often, we identify left $\mathbf{A}$-modules with right $\mathbf{A}^{\mathrm{op}}$-modules, but in every such case, this is explicitly said. Action on a (say right) module M of a k-algebra $\mathbf{A}$, if evaluated on elements $m \in M$ and $a \in A$, is denoted by $\varrho_M : m \otimes a \mapsto m \cdot a$.

For coproducts in a coalgebra and, more generally, in a coring C, Sweedler's index notation is used. That is, for an element $c \in C$, we write $c \mapsto c_{(1)} \otimes c_{(2)}$ (or sometimes $c \mapsto c^{(1)} \otimes c^{(2)}$) for the coproduct, where implicit summation is understood. Similarly, for a (say right) coaction on a comodule M of a coring, evaluated on an element $m \in M$, the notation $\varrho^M : m \mapsto m_{[0]} \otimes m_{[1]}$ (or $m \mapsto m^{[0]} \otimes m^{[1]}$) is used, where implicit summation is understood. The category of right (resp. left) comodules of a coring C is denoted by $\mathcal{M}^C$ (resp. ${}^C\mathcal{M}$). Hom sets are denoted by $\mathrm{Hom}^C(-,-)$ (resp. ${}^C\mathrm{Hom}(-,-)$).

In any category $\mathcal{A}$, the identity morphism at an object A is denoted by the same symbol A. Hom sets in $\mathcal{A}$ are denoted by $\mathrm{Hom}_{\mathcal{A}}(-,-)$.

In a monoidal category $(\mathcal{M}, \otimes, U)$, we allow for nontrivial coherence isomorphisms $(- \otimes -) \otimes - \cong - \otimes (- \otimes -)$ and $- \otimes U \cong - \cong U \otimes -$ but do not denote them explicitly. (Such monoidal categories are called in the literature, sometimes, *lax* monoidal.) The opposite of a monoidal category $(\mathcal{M}, \otimes, U)$, denoted by $(\mathcal{M}, \otimes, U)^{\mathrm{op}}$, means the same category $\mathcal{M}$ with opposite monoidal product. A functor $\mathbb{F}$ between monoidal categories $(\mathcal{M}, \otimes, U)$ and $(\mathcal{M}', \otimes', U')$ is said to be monoidal if there exist natural transformations $\mathbb{F}^2 : \mathbb{F}(-) \otimes' \mathbb{F}(-) \to \mathbb{F}(- \otimes -)$ and $\mathbb{F}^0 : U' \to \mathbb{F}(U)$, satisfying the usual compatibility conditions. $\mathbb{F}$ is said to be op-monoidal if there exist compatible natural transformations $\mathbb{F}_2 : \mathbb{F}(- \otimes -) \to \mathbb{F}(-) \otimes' \mathbb{F}(-)$ and $\mathbb{F}_0 : \mathbb{F}(U) \to U'$. A monoidal functor $(\mathbb{F}, \mathbb{F}^2, \mathbb{F}^0)$ is strong monoidal if $\mathbb{F}^2$ and $\mathbb{F}^0$ are isomorphisms, and it is strict monoidal if $\mathbb{F}^2$ and $\mathbb{F}^0$ are identity morphisms.

2. R-rings and R-corings

A *monoid* in a monoidal category $(\mathcal{M}, \otimes, U)$ is a triple (A, μ, η). Here, A is an object and $\mu : A \otimes A \to A$ and $\eta : U \to A$ are morphisms in $\mathcal{M}$, satisfying associativity and unitality conditions

$$\mu \circ (\mu \otimes A) = \mu \circ (A \otimes \mu) \qquad \text{and} \qquad \mu \circ (\eta \otimes A) = A = \mu \circ (A \otimes \eta). \quad (2.1)$$

The morphism μ is called a *multiplication* (or *product*) and η is called a *unit*. An algebra over a commutative ring k can be described as a monoid in the monoidal category $(\mathcal{M}_k, \otimes_k, k)$ of k-modules.

A *right module* of a monoid (A, μ, η) is a pair (V, ν), where V is an object and $\nu : V \otimes A \to V$ is a morphism in $\mathcal{M}$ such that

$$\nu \circ (V \otimes \mu) = \nu \circ (\nu \otimes A) \qquad \text{and} \qquad \nu \circ (V \otimes \eta) = V.$$

Left modules are defined symmetrically.

Reversing all arrows in the definition of a monoid, we arrive at the dual notion of a comonoid. A *comonoid* in a monoidal category $(\mathcal{M}, \otimes, U)$ is a triple (C, Δ, ϵ). Here, C is an object and $\Delta : C \to C \otimes C$ and $\epsilon : C \to U$ are morphisms in $\mathcal{M}$, satisfying coassociativity and counitality conditions

$$(\Delta \otimes C) \circ \Delta = (C \otimes \Delta) \circ \Delta \qquad \text{and} \qquad (\epsilon \otimes C) \circ \Delta = C = (C \otimes \epsilon) \circ \Delta. \quad (2.2)$$

The morphism Δ is called a *comultiplication* (or *coproduct*) and ϵ is called a *counit*. A coalgebra over a commutative ring k can be described as a comonoid in the monoidal category $(\mathcal{M}_k, \otimes_k, k)$ of k-modules. Dualizing the definition of a module of a monoid, one arrives at the notion of a comodule of a comonoid.

Many aspects of the theory of algebras and their modules, or coalgebras and their comodules, can be extended to monoids or comonoids in general monoidal categories. Here, we are interested in monoids and comonoids in a monoidal category $({}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}, \otimes_{\mathbf{R}}, \mathbf{R})$ of bimodules over a k-algebra $\mathbf{R}$. These monoids and comonoids are called $\mathbf{R}$-*rings* and $\mathbf{R}$-*corings*, respectively.

2.1. $\mathbf{R}$-*rings*

Generalizing algebras over commutative rings, we study monoids in bimodule categories.

Definition 2.1. For an algebra $\mathbf{R}$ over a commutative ring k, an $\mathbf{R}$-*ring* is a triple (A, μ, η). Here, A is an $\mathbf{R}$-bimodule and $\mu : A \otimes_{\mathbf{R}} A \to A$ and $\eta : R \to A$ are $\mathbf{R}$-bimodule maps, satisfying the associativity and unit conditions (2.1). A *morphism of* $\mathbf{R}$-*rings* $f : (A, \mu, \eta) \to (A', \mu', \eta')$ is an $\mathbf{R}$-bimodule map $f : A \to A'$ such that $f \circ \mu = \mu' \circ (f \otimes_{\mathbf{R}} f)$ and $f \circ \eta = \eta'$.

For an $\mathbf{R}$-ring (A, μ, η), the *opposite* means the $\mathbf{R}^{\mathrm{op}}$-ring $(A^{\mathrm{op}}, \mu^{\mathrm{op}}, \eta)$. Here, A^{op} is the same k-module A. It is understood to be a left (resp. right) $\mathbf{R}^{\mathrm{op}}$-module via the right (resp. left) $\mathbf{R}$-action. Multiplication is $\mu^{\mathrm{op}}(a \otimes_{\mathbf{R}^{\mathrm{op}}} a') := \mu(a' \otimes_{\mathbf{R}} a)$ and unit is η.

A most handy characterization of $\mathbf{R}$-rings comes from the following observation.

Lemma 2.2. *There is a bijective correspondence between* $\mathbf{R}$-*rings* (A, μ, η) *and* k-*algebra homomorphisms* $\eta : \mathbf{R} \to \mathbf{A}$.

Indeed, starting with an $\mathbf{R}$-ring (A, μ, η), a multiplication map $A \otimes_k A \to A$ is obtained by composing the canonical epimorphism $A \otimes_k A \to A \otimes_{\mathbf{R}} A$ with μ. Conversely, starting with an algebra map $\eta : \mathbf{R} \to \mathbf{A}$, an $\mathbf{R}$-bilinear associative

multiplication $A \otimes_{\mathbf{R}} A \to A$ is obtained by using the universality of the coequalizer $A \otimes_k A \to A \otimes_{\mathbf{R}} A$.

An $\mathbf{R}$-ring (A, μ, η) determines monads on the categories of right and left $\mathbf{R}$-modules (i.e. monoids in the monoidal categories of endofunctors on $\mathcal{M}_{\mathbf{R}}$ and ${}_{\mathbf{R}}\mathcal{M}$). They are given by $- \otimes_{\mathbf{R}} A : \mathcal{M}_{\mathbf{R}} \to \mathcal{M}_{\mathbf{R}}$ and $A \otimes_{\mathbf{R}} - : {}_{\mathbf{R}}\mathcal{M} \to {}_{\mathbf{R}}\mathcal{M}$, respectively.

DEFINITION 2.3. A *right module* for an $\mathbf{R}$-ring (A, μ, η) is an algebra for the monad $- \otimes_{\mathbf{R}} A$ on the category $\mathcal{M}_{\mathbf{R}}$. A *right module morphism* is a morphism of algebras for the monad $- \otimes_{\mathbf{R}} A$.

A *left module* for an $\mathbf{R}$-ring (A, μ, η) is an algebra for the monad $A \otimes_{\mathbf{R}} -$ on the category ${}_{\mathbf{R}}\mathcal{M}$. A *left module morphism* is a morphism of algebras for the monad $A \otimes_{\mathbf{R}} -$.

Left modules of an $\mathbf{R}$-ring are canonically identified with right modules for the opposite $\mathbf{R}^{\mathrm{op}}$-ring. Analogously to Lemma 2.2, modules for $\mathbf{R}$-rings can be characterized as follows.

LEMMA 2.4. *A k-module M is a (left or right) module of an $\mathbf{R}$-ring (A, μ, η) if and only if it is a (left or right) module of the corresponding k-algebra $\mathbf{A}$ in Lemma 2.2. Furthermore, a k-module map $f : M \to M'$ is a morphism of (left or right) modules of an $\mathbf{R}$-ring (A, μ, η) if and only if it is a morphism of (left or right) modules of the corresponding k-algebra $\mathbf{A}$ in Lemma 2.2.*

The situation when the (left or right) regular $\mathbf{R}$-module extends to a (left or right) module of an $\mathbf{R}$-ring (A, μ, η) is of particular interest.

LEMMA 2.5. *The right regular module of a k-algebra $\mathbf{R}$ extends to a right module of an $\mathbf{R}$-ring (A, μ, η) if and only if there exists a k-module map $\chi : A \to R$, obeying the following properties.*

(i) $\chi(a\eta(r)) = \chi(a)r$, *for* $a \in A$ *and* $r \in R$ (*right* $\mathbf{R}$-*linearity*),
(ii) $\chi(aa') = \chi((\eta \circ \chi)(a)a')$, *for* $a, a' \in A$ (*associativity*),
(iii) $\chi(1_{\mathbf{A}}) = 1_{\mathbf{R}}$ (*unitality*).

A map χ obeying these properties is called a right character *on the $\mathbf{R}$-ring (A, μ, η).*

In terms of a right character χ, a right $\mathbf{A}$-action on R is given by $r \cdot a := \chi(\eta(r)a)$. Conversely, in terms of a right $\mathbf{A}$-action on R, a right character is constructed as $\chi(a) := 1_{\mathbf{R}} \cdot a$. Symmetrically, one can define a *left character* on an $\mathbf{R}$-ring (A, μ, η) via the requirement that the left regular $\mathbf{R}$-module extends to a left module for (A, μ, η).

DEFINITION 2.6. Let (A, μ, η) be an $\mathbf{R}$-ring possessing a right character $\chi : A \to R$. The *invariants* of a right module (M, ϱ_M) with respect to χ are the elements of the

k-submodule

$$M_\chi := \{ m \in M \mid \varrho_M(m \underset{\mathbf{R}}{\otimes} a) = \varrho_M(m \underset{\mathbf{R}}{\otimes} (\eta \circ \chi)(a)), \quad \forall a \in A \} \cong \mathrm{Hom}_{\mathbf{A}}(R, M),$$

where the isomorphism $M_\chi \to \mathrm{Hom}_{\mathbf{A}}(R, M)$ is given by $m \mapsto (r \mapsto m \cdot \eta(r))$. In particular, the invariants of R are the elements of the subalgebra

$$\mathbf{B} := R_\chi = \{ b \in R \mid \chi(\eta(b)a) = b\chi(a), \quad \forall a \in A \}.$$

Associated to a character χ, there is a canonical map

$$A \to {}_{\mathbf{B}}\mathrm{End}(R), \qquad a \mapsto \big(r \mapsto \chi(\eta(r)a) \big). \tag{2.3}$$

The $\mathbf{R}$-ring (A, μ, η) is said to be a *Galois* $\mathbf{R}$-ring (with respect to the character χ) provided that the canonical map (2.3) is bijective.

2.2. $\mathbf{R}$-*corings*

The theory of $\mathbf{R}$-corings is dual to that of $\mathbf{R}$-rings. A detailed study can be found in the monograph [15].

Definition 2.7. For an algebra $\mathbf{R}$ over a commutative ring k, an $\mathbf{R}$-*coring* is a triple (C, Δ, ϵ). Here, C is an $\mathbf{R}$-bimodule and $\Delta : C \to C \otimes_{\mathbf{R}} C$ and $\epsilon : C \to R$ are $\mathbf{R}$-bimodule maps, satisfying the coassociativity and counit conditions (2.2). A *morphism of* $\mathbf{R}$-*corings* $f : (C, \Delta, \epsilon) \to (C', \Delta', \epsilon')$ is an $\mathbf{R}$-bimodule map $f : C \to C'$ such that $\Delta' \circ f = (f \otimes_{\mathbf{R}} f) \circ \Delta$ and $\epsilon' \circ f = \epsilon$.

For an $\mathbf{R}$-coring (C, Δ, ϵ), the *co-opposite* is the $\mathbf{R}^{\mathrm{op}}$-coring $(C_{\mathrm{cop}}, \Delta_{\mathrm{cop}}, \epsilon)$. Here C_{cop} is the same k-module C. It is understood to be a left (resp. right) $\mathbf{R}^{\mathrm{op}}$-module via the right (resp. left) $\mathbf{R}$-action. The comultiplication is $\Delta_{\mathrm{cop}}(c) := c_{(2)} \otimes_{\mathbf{R}^{\mathrm{op}}} c_{(1)}$ and the counit is ϵ.

An $\mathbf{R}$-coring (C, Δ, ϵ) determines comonads on the categories of right and left $\mathbf{R}$-modules (i.e. comonoids in the monoidal categories of endofunctors on $\mathcal{M}_{\mathbf{R}}$ and ${}_{\mathbf{R}}\mathcal{M}$). They are given by $- \otimes_{\mathbf{R}} C : \mathcal{M}_{\mathbf{R}} \to \mathcal{M}_{\mathbf{R}}$ and $C \otimes_{\mathbf{R}} - : {}_{\mathbf{R}}\mathcal{M} \to {}_{\mathbf{R}}\mathcal{M}$, respectively.

Definition 2.8. A *right comodule* for an $\mathbf{R}$-coring (C, Δ, ϵ) is a coalgebra for the comonad $- \otimes_{\mathbf{R}} C$ on the category $\mathcal{M}_{\mathbf{R}}$. That is, a pair (M, ϱ^M), where M is a right $\mathbf{R}$-module and $\varrho^M : M \to M \otimes_{\mathbf{R}} C$ is a right $\mathbf{R}$-module map, satisfying the coassociativity and counit conditions

$$(\varrho^M \underset{\mathbf{R}}{\otimes} C) \circ \varrho^M = (M \underset{\mathbf{R}}{\otimes} \Delta) \circ \varrho^M \qquad \text{and} \qquad (M \underset{\mathbf{R}}{\otimes} \epsilon) \circ \varrho^M = M. \tag{2.4}$$

A *right comodule morphism* $f : (M, \varrho^M) \to (M', \varrho^{M'})$ is a morphism of coalgebras for the comonad $- \otimes_{\mathbf{R}} C$. That is, a right $\mathbf{R}$-module map $f : M \to M'$, satisfying $\varrho^{M'} \circ f = (f \otimes_{\mathbf{R}} C) \circ \varrho^M$.

Symmetrically, a *left comodule* is a coalgebra for the comonad $C \otimes_{\mathbf{R}} -$ on the category ${}_{\mathbf{R}}\mathcal{M}$. A *left comodule morphism* is a morphism of coalgebras for the comonad $C \otimes_{\mathbf{R}} -$.

Left comodules of an $\mathbf{R}$-coring are canonically identified with right comodules for the co-opposite $\mathbf{R}^{\mathrm{op}}$-coring.

The situation when the (left or right) regular $\mathbf{R}$-module extends to a (left or right) comodule of an $\mathbf{R}$-coring (C, Δ, ϵ) is of particular interest (see [16, Lemma 5.1]).

LEMMA 2.9. *The (left or right) regular* $\mathbf{R}$*-module extends to a (left or right) comodule of an* $\mathbf{R}$*-coring* (C, Δ, ϵ) *if and only if there exists an element* $g \in C$*, obeying the following properties.*

(i) $\Delta(g) = g \otimes_{\mathbf{R}} g$.
(ii) $\epsilon(g) = 1_{\mathbf{R}}$.

An element g *obeying these properties is called a* grouplike element *in the* $\mathbf{R}$*-coring* (C, Δ, ϵ).

Having a grouplike element g in an $\mathbf{R}$-coring (C, Δ, ϵ), a right coaction on R is constructed as a map $R \to C$, $r \mapsto g \cdot r$. Conversely, a right coaction $\varrho^R : R \to C$ determines a grouplike element $\varrho^R(1_{\mathbf{R}})$.

DEFINITION 2.10. Let (C, Δ, ϵ) be an $\mathbf{R}$-coring possessing a grouplike element $g \in C$. The *coinvariants* of a right comodule (M, ϱ^M) with respect to g are the elements of the k-submodule

$$M^g := \{\, m \in M \mid \varrho^M(m) = m \underset{\mathbf{R}}{\otimes} g \,\} \cong \mathrm{Hom}^C(R, M),$$

where the isomorphism $M^g \to \mathrm{Hom}^C(R, M)$ is given by $m \mapsto (r \mapsto m \cdot r)$. In particular, the coinvariants of R are the elements of the subalgebra

$$\mathbf{B} := R^g = \{\, b \in R \mid b \cdot g = g \cdot b \,\}.$$

Associated to a grouplike element g, there is a canonical map

$$R \underset{\mathbf{B}}{\otimes} R \to C, \qquad r \underset{\mathbf{B}}{\otimes} r' \mapsto r \cdot g \cdot r'. \tag{2.5}$$

The $\mathbf{R}$-coring (C, Δ, ϵ) is said to be a *Galois* $\mathbf{R}$-coring (with respect to the grouplike element g) provided that the canonical map (2.5) is bijective.

Let (C, Δ, ϵ) be an $\mathbf{R}$-coring possessing a grouplike element g. Put $\mathbf{B} := R^g$. For any right C-comodule M, M^g is a right $\mathbf{B}$-module. Furthermore, any right C-comodule map $M \to M'$ restricts to a right $\mathbf{B}$-module map $M^g \to M'^g$. There is an adjoint pair of functors

$$- \underset{\mathbf{B}}{\otimes} R : \mathcal{M}_{\mathbf{B}} \to \mathcal{M}^C \qquad \text{and} \qquad (-)^g : \mathcal{M}^C \to \mathcal{M}_{\mathbf{B}}. \tag{2.6}$$

If (C, Δ, ϵ) is a Galois coring (with respect to g), then $\mathcal{M}^C$ is equivalent to the category of descent data for the extension $\mathbf{B} \subseteq \mathbf{R}$. Hence, the situation, when the functors (2.6) establish an equivalence, is interesting from the descent theory point of view.

2.3. *Duality*

Beyond the formal duality between algebras and coalgebras, it is well known that the k-dual of a coalgebra over a commutative ring k possesses a canonical algebra structure. The converse is true whenever a k-algebra is finitely generated and projective as a k-module. In what follows, we recall analogs of these facts for rings and corings over an arbitrary algebra $\mathbf{R}$.

PROPOSITION 2.11. *Let* $\mathbf{R}$ *be an algebra over a commutative ring* k.

(1) *For an* $\mathbf{R}$*-coring* (C, Δ, ϵ)*, the left dual* ${}^*C := {}_{\mathbf{R}}\mathrm{Hom}(C, R)$ *possesses a canonical* $\mathbf{R}$*-ring structure. Multiplication is given by* $(\phi\psi)(c) := \psi\big(c_{(1)} \cdot \phi(c_{(2)})\big)$*, for* $\phi, \psi \in {}^*C$ *and* $c \in C$*. Unit map is* $\mathbf{R} \to {}^*C$, $r \mapsto \epsilon(-)r$.

(2) *For an* $\mathbf{R}$*-ring* (A, μ, η)*, which is a finitely generated and projective right* $\mathbf{R}$*-module, the right dual* $A^* := \mathrm{Hom}_{\mathbf{R}}(A, R)$ *possesses a canonical* $\mathbf{R}$*-coring structure. In terms of a dual basis* $(\{a_i \in A\}, \{\alpha_i \in A^*\})$*, comultiplication is given by* $\xi \mapsto \sum_i \xi(a_i -) \otimes_{\mathbf{R}} \alpha_i$*, which is independent of the choice of a dual basis. Counit is* $A^* \to R$, $\xi \mapsto \xi(1_{\mathbf{A}})$.

(3) *For an* $\mathbf{R}$*-coring* (C, Δ, ϵ)*, which is a finitely generated and projective left* $\mathbf{R}$*-module, the second dual* $({}^*C)^*$ *is isomorphic to* C *as an* $\mathbf{R}$*-coring.*

(4) *For an* $\mathbf{R}$*-ring* (A, μ, η)*, which is a finitely generated and projective right* $\mathbf{R}$*-module, the second dual* ${}^*(A^*)$ *is isomorphic to* A *as an* $\mathbf{R}$*-ring.*

Applying Proposition 2.11 to the co-opposite coring and the opposite ring, analogous correspondences are found between right duals of corings and left duals of rings.

PROPOSITION 2.12. *Let* C *be a coring over an algebra* $\mathbf{R}$.

(1) *Any right* C*-comodule* (M, ϱ^M) *possesses a right module structure for the* $\mathbf{R}$*-ring* ${}^*C$,

$$m \cdot \phi := m_{[0]} \cdot \phi(m_{[1]}), \qquad \text{for } m \in M,\ \phi \in {}^*C. \tag{2.7}$$

Any right C*-comodule map becomes a* ${}^*C$*-module map. That is, there is a faithful functor* $\mathcal{M}^C \to \mathcal{M}_{{}^*C}$.

(2) *The functor* $\mathcal{M}^C \to \mathcal{M}_{{}^*C}$ *is an equivalence if and only if* C *is a finitely generated and projective left* $\mathbf{R}$*-module.*

There is a duality between Galois rings and Galois corings too.

PROPOSITION 2.13. *Let C be an $\mathbf{R}$-coring, which is a finitely generated and projective left $\mathbf{R}$-module. For an element $g \in C$, introduce the map $\chi_g : {}^*C \to R$, $\phi \mapsto \phi(g)$. The following statements hold.*

(1) *The element $g \in C$ is grouplike if and only if χ_g is a right character on the $\mathbf{R}$-ring ${}^*C$.*
(2) *An element $b \in R$ is a coinvariant of the right C-comodule R (with coaction induced by a grouplike element g) if and only if b is an invariant of the right ${}^*C$-module R (with respect to the right character χ_g).*
(3) *The $\mathbf{R}$-coring C is a Galois coring (with respect to a grouplike element g) if and only if the $\mathbf{R}$-ring ${}^*C$ is a Galois ring (with respect to the right character χ_g).*

3. Bialgebroids

In Section 2, we generalized algebras and coalgebras over commutative rings to monoids and comonoids in bimodule categories. We could easily do so, the category of bimodules over any k-algebra $\mathbf{R}$ is a monoidal category, just as the category of k-modules. If we try to generalize bialgebras to a noncommutative base algebra $\mathbf{R}$, however, we encounter difficulties. Recall that a k-bialgebra consists of an algebra (B, μ, η), and a coalgebra (B, Δ, ϵ) defined on the same k-module B. They are subject to compatibility conditions. Unit and multiplication must be coalgebra maps or equivalently counit, and comultiplication must be algebra maps. This means in particular that, for any elements b and b' in B, multiplication and comultiplication satisfy the condition

$$(bb')_{(1)} \underset{k}{\otimes} (bb')_{(2)} = b_{(1)}b'_{(1)} \underset{k}{\otimes} b_{(2)}b'_{(2)}. \tag{3.1}$$

Note that (3.1) is formulated in terms of the symmetry tw in $\mathcal{M}_k$. For any k-modules M and N, the twist map $\mathsf{tw}_{M,N} : M \otimes_k N \to N \otimes_k M$ maps $m \otimes_k n$ to $n \otimes_k m$. Written in terms of morphisms (3.1) is equivalent to

$$\Delta \circ \mu = (\mu \underset{k}{\otimes} \mu) \circ (B \underset{k}{\otimes} \mathsf{tw}_{B,B} \underset{k}{\otimes} B) \circ (\Delta \underset{k}{\otimes} \Delta).$$

In the literature, one can find generalizations when tw is replaced by a braiding [17–19]. (For an approach when tw is replaced by a *mixed distributive law* (see [20]).) However, general bimodule categories are neither symmetric nor braided. There is no natural way to formulate an analog of (3.1) in a bimodule category. Indeed, more sophisticated ideas are needed.

The notion that is known today as a (left) bialgebroid was introduced (independently) by several authors. The first definition is due to Takeuchi, who used the name $\times_R$*-bialgebra* in [21]. Some twenty years later, with motivation coming from Poisson geometry, in [6], Lu proposed an equivalent definition. The term *bialgebroid* is due to her. A third equivalent set of axioms was invented by Xu in [8]. The equivalence of the listed definitions is far from obvious. It was proven by Brzeziński and Militaru

in [22]. Symmetrical notions of left and right bialgebroids were formulated and studied by Kadison and Szlachányi in [23]. The definition presented here is a slightly reformulated version of the one in [6] or [23].

3.1. *Right and left bialgebroids*

In contrast to the definition of a bialgebra, in this section, a bialgebroid is not described as a compatible monoid and comonoid in some monoidal category (of bimodules). The ring and coring structures of a bialgebroid are defined over different base algebras; they are a monoid and a comonoid in different monoidal categories. Recall from Lemma 2.2 that an $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring A (for some algebra, $\mathbf{R}$ over a commutative ring k) is described by a k-algebra map $\eta : \mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}} \to \mathbf{A}$. Equivalently, instead of η, we can consider its restrictions

$$s := \eta(- \underset{k}{\otimes} 1_{\mathbf{R}}) : \mathbf{R} \to \mathbf{A} \qquad \text{and} \qquad t := \eta(1_{\mathbf{R}} \underset{k}{\otimes} -) : \mathbf{R}^{\mathrm{op}} \to \mathbf{A}, \quad (3.2)$$

which are k-algebra maps with commuting ranges in $\mathbf{A}$. The maps s and t in (3.2) are called the *source* and *target* maps of an $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring A, respectively. In what follows, an $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring will be given by a triple $(\mathbf{A}, s, t)$, where $\mathbf{A}$ is a k-algebra (with underlying k-module A) and s and t are algebra maps with commuting ranges as in (3.2).

Definition 3.1. Let $\mathbf{R}$ be an algebra over a commutative ring k. A *right* $\mathbf{R}$*-bialgebroid* $\mathcal{B}$ consists of an $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring $(\mathbf{B}, s, t)$ and an $\mathbf{R}$-coring (B, Δ, ϵ) on the same k-module B. They are subject to the following compatibility axioms.

(i) The bimodule structure in the $\mathbf{R}$-coring (B, Δ, ϵ) is related to the $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring $(\mathbf{B}, s, t)$ via

$$r \cdot b \cdot r' := bs(r')t(r), \quad \text{for } r, r' \in R,\ b \in B. \quad (3.3)$$

(ii) Considering B as an $\mathbf{R}$-bimodule as in (3.3), the coproduct Δ corestricts to a k-algebra map from $\mathbf{B}$ to

$$B \times_R B := \left\{ \sum_i b_i \underset{\mathbf{R}}{\otimes} b_i' \mid \sum_i s(r) b_i \underset{\mathbf{R}}{\otimes} b_i' = \sum_i b_i \underset{\mathbf{R}}{\otimes} t(r) b_i', \quad \forall r \in R \right\}, \quad (3.4)$$

where $B \times_R B$ is an algebra via factorwise multiplication.

(iii) The counit ϵ is a right character on the $\mathbf{R}$-ring $(\mathbf{B}, s)$.

Remark 3.2. The bialgebroid axioms in Definition 3.1 have some immediate consequences.

(1) Note that the k-submodule $B \times_R B$ of $B \otimes_{\mathbf{R}} B$ is defined in such a way that factorwise multiplication is well defined on it. $B \times_R B$ is called the *Takeuchi*

product. In fact, it has more structure than that of a k-algebra: it is an $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring with unit map $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}} \to B \times_R B$, $r \otimes_k r' \mapsto t(r') \otimes_{\mathbf{R}} s(r)$. The (corestriction of the) coproduct is easily checked to be an $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-bimodule map $B \to B \times_R B$.

(2) Axiom (iii) is equivalent to the requirement that the counit ϵ is a right character on the $\mathbf{R}^{\mathrm{op}}$-ring $(\mathbf{B}, t)$.

(3) Yet another equivalent formulation of axiom (iii) is the following. The map $\theta : \mathbf{B} \to \mathrm{End}_k(R)^{\mathrm{op}}$, $b \mapsto \big(r \mapsto \epsilon(s(r)b) \big)$ is a k-algebra map, where $\mathrm{End}_k(R)^{\mathrm{op}}$ is an algebra via opposite composition of endomorphisms. The map θ is called an *anchor map* in [8].

Recall that in a bialgebra over a commutative ring, replacing the algebra with the opposite one or replacing the coalgebra with the co-opposite one, one arrives at bialgebras again. Analogously, the *co-opposite* of a right $\mathbf{R}$-bialgebroid $\mathcal{B}$, with structure maps denoted as in Definition 3.1, is the following right $\mathbf{R}^{\mathrm{op}}$-bialgebroid $\mathcal{B}_{\mathrm{cop}}$. The $\mathbf{R}^{\mathrm{op}} \otimes_k \mathbf{R}$-ring structure is $(\mathbf{B}, t, s)$, and the $\mathbf{R}^{\mathrm{op}}$-coring structure is $(B_{\mathrm{cop}}, \Delta_{\mathrm{cop}}, \epsilon)$. However, the $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring $(\mathbf{B}^{\mathrm{op}}, t, s)$ and the $\mathbf{R}$-coring (B, Δ, ϵ) do not satisfy the same axioms in Definition 3.1. Instead, they are subject to a left- right symmetrical version of Definition 3.1.

Definition 3.3. Let $\mathbf{R}$ be an algebra over a commutative ring k. A *left* $\mathbf{R}$*-bialgebroid* $\mathcal{B}$ consists of an $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring $(\mathbf{B}, s, t)$ and an $\mathbf{R}$-coring (B, Δ, ϵ) on the same k-module B. They are subject to the following compatibility axioms.

(i) The bimodule structure in the $\mathbf{R}$-coring (B, Δ, ϵ) is related to the $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring $(\mathbf{B}, s, t)$ via

$$r \cdot b \cdot r' := s(r)t(r')b, \quad \text{for } r, r' \in R,\ b \in B. \tag{3.5}$$

(ii) Considering B as an $\mathbf{R}$-bimodule as in (3.5), the coproduct Δ corestricts to a k-algebra map from $\mathbf{B}$ to

$$B\,{}_R{\times}\,B := \left\{ \sum_i b_i \underset{\mathbf{R}}{\otimes} b'_i \mid \sum_i b_i t(r) \underset{\mathbf{R}}{\otimes} b'_i = \sum_i b_i \underset{\mathbf{R}}{\otimes} b'_i s(r), \quad \forall r \in R \right\}, \tag{3.6}$$

where $B\,{}_R{\times}\,B$ is an algebra via factorwise multiplication.

(iii) The counit ϵ is a left character on the $\mathbf{R}$-ring $(\mathbf{B}, s)$.

Since in this note left and right bialgebroids are considered simultaneously, we use two versions of Sweedler's index notation. In a left bialgebroid, we use lower indices to denote components of the coproduct, that is, we write $b \mapsto b_{(1)} \otimes_{\mathbf{R}} b_{(2)}$. In a right bialgebroid, we use upper indices to denote components of the coproduct, that is, we write $b \mapsto b^{(1)} \otimes_{\mathbf{R}} b^{(2)}$. In both cases, implicit summation is understood.

Recall that coalgebras over a commutative ring k form a monoidal category with respect to the k-module tensor product. Bialgebras over k can be described as monoids in the monoidal category of k-coalgebras. In [21], Takeuchi defined bialgebroids

($\times_R$-*bialgebras* in his terminology) as monoids in a monoidal category of certain corings too. By [24, Definition 3.5], for two k-algebras $\mathbf{R}$ and $\mathbf{S}$, an $\mathbf{S}|\mathbf{R}$-*coring* is an $\mathbf{S} \otimes_k \mathbf{R}$-bimodule C together with an $\mathbf{R}$-coring structure (C, Δ, ϵ) such that the following identities hold.

$$\Delta((s \underset{k}{\otimes} 1_{\mathbf{R}}) \cdot c \cdot (s' \underset{k}{\otimes} 1_{\mathbf{R}})) = c^{(1)} \cdot (s' \underset{k}{\otimes} 1_{\mathbf{R}}) \underset{\mathbf{R}}{\otimes} (s \underset{k}{\otimes} 1_{\mathbf{R}}) \cdot c^{(2)} \quad \text{and}$$

$$(s \underset{k}{\otimes} 1_{\mathbf{R}}) \cdot c^{(1)} \underset{\mathbf{R}}{\otimes} c^{(2)} = c^{(1)} \underset{\mathbf{R}}{\otimes} c^{(2)} \cdot (s \underset{k}{\otimes} 1_{\mathbf{R}}), \quad \text{for } s, s' \in S, c \in C.$$

Morphisms of $\mathbf{S}|\mathbf{R}$-corings are morphisms of $\mathbf{R}$-corings, which are in addition $\mathbf{S}$-bimodule maps. In particular, it can be shown by using the same methods as in [22] that the notion of an $\mathbf{R}|\mathbf{R}$-coring is equivalent to a $\times_R$-*coalgebra* in [21, Definition 4.1]. Part (1) of the following Theorem 3.4 is, thus, a reformulation of [21, Proposition 4.7]. Part (2) states an equivalence of a right $\mathbf{R}$-bialgebroid as in Definition 3.1 and a symmetrical version of a $\times_R$-bialgebra as in [21, Definition 4.5].

THEOREM 3.4. *For an algebra* $\mathbf{R}$ *over a commutative ring* k*, the following statements hold.*

(1) $\mathbf{R}|\mathbf{R}$-*corings form a monoidal category. The monoidal product of two objects* (C, Δ, ϵ) *and* (C', Δ', ϵ') *is the* $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-*module tensor product*

$$C \odot C' := C \underset{k}{\otimes} C' / \{ (1_{\mathbf{R}} \underset{k}{\otimes} r) \cdot c \cdot (1_{\mathbf{R}} \underset{k}{\otimes} r') \underset{k}{\otimes} c' - c \underset{k}{\otimes} (r' \underset{k}{\otimes} 1_{\mathbf{R}}) \cdot c' \cdot (r \underset{k}{\otimes} 1_{\mathbf{R}}) \mid r, r' \in R \}.$$

$C \odot C'$ is an $\mathbf{R} \otimes_k \mathbf{R}$-bimodule, via the actions

$$(r_1 \underset{k}{\otimes} r_2) \cdot (c \odot c') \cdot (r_1' \underset{k}{\otimes} r_2') := (r_1 \underset{k}{\otimes} 1_{\mathbf{R}}) \cdot c \cdot (r_1' \underset{k}{\otimes} 1_{\mathbf{R}}) \odot (1_{\mathbf{R}} \underset{k}{\otimes} r_2) \cdot c' \cdot (1_{\mathbf{R}} \underset{k}{\otimes} r_2').$$

The coproduct and counit in $C \odot C'$ are

$$c \odot c' \mapsto (c^{(1)} \odot c'^{(1)}) \underset{\mathbf{R}}{\otimes} (c^{(2)} \odot c'^{(2)}) \qquad \text{and} \qquad c \odot c' \mapsto \epsilon'\big((\epsilon(c) \underset{k}{\otimes} 1_{\mathbf{R}}) \cdot c'\big).$$

The monoidal unit is $R \otimes_k R$, with $\mathbf{R}|\mathbf{R}$-coring structure described in Section 3.2.3 below. The monoidal product of morphisms $\alpha_i : (C_i, \Delta_i, \epsilon_i) \to (C_i', \Delta_i', \epsilon_i')$, for $i = 1, 2$, is given by

$$(\alpha_1 \odot \alpha_2)(c_1 \odot c_2) := \alpha_1(c_1) \odot \alpha_2(c_2).$$

(2) *Monoids in the monoidal category of* $\mathbf{R}|\mathbf{R}$-*corings are the same as right* $\mathbf{R}$-*bialgebroids.*

(3) *Monoidal morphisms in the monoidal category of* $\mathbf{R}|\mathbf{R}$-*corings are the same as maps of* $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-*rings and of* $\mathbf{R}$-*corings.*

Considering a right $\mathbf{R}$-bialgebroid as an $\mathbf{R}|\mathbf{R}$-coring, the $\cdot|\mathbf{R}$ -bimodule structure is given by right multiplications by the source and target maps. The $\mathbf{R}|\cdot$ -bimodule structure is given by left multiplications by the source and target maps.

Theorem 3.4 was extended by Szlachányi in [25]. He has shown that $\mathbf{S}|\mathbf{R}$-corings form a bicategory. Bialgebroids can be described as monads in this bicategory. This

makes it possible to define (*base changing*) morphisms of bialgebroids as bimodules for the corresponding monads. The constructions described in Section 3.4.1 and also in Definition 3.15 provide examples of bialgebroid morphisms in this sense. For more details we refer to [25].

An $\mathbf{S}|\mathbf{R}$-coring C can be looked at as an $\mathbf{S} \otimes_k \mathbf{S}^{\mathrm{op}}$-$\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$ bimodule in a canonical way. Identifying $\mathbf{S}$-bimodules with right $\mathbf{S} \otimes_k \mathbf{S}^{\mathrm{op}}$-modules and $\mathbf{R}$-bimodules with right $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-modules, there is an op-monoidal left adjoint functor $- \otimes_{\mathbf{S} \otimes_k \mathbf{S}^{\mathrm{op}}} C : {}_{\mathbf{S}}\mathcal{M}_{\mathbf{S}} \to {}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$. As a matter of fact, this correspondence establishes a bifunctor from the bicategory of $\mathbf{S}|\mathbf{R}$-corings to the 2-category of op-monoidal left adjoint functors between bimodule categories. For any two 0-cells (i.e. algebras $\mathbf{S}$ and $\mathbf{R}$), it gives an equivalence of the vertical subcategories. So, in addition to a characterization of bialgebroids as monads in the bicategory of $\mathbf{S}|\mathbf{R}$-corings, they can be described as monads in the 2-category of op-monoidal left adjoint functors between bimodule categories. Monads in the 2-category of op-monoidal functors (i.e. op-monoidal monads such that multiplication and unit natural transformations of the monad are compatible with the op-monoidal structure) were termed *Hopf monads* in [26] and *bimonads* in [27]. Using the latter terminology, the following characterization of bialgebroids in [27, Theorem 4.5] is obtained.

THEOREM 3.5. *For an algebra* $\mathbf{R}$*, any right* $\mathbf{R}$*-bialgebroid induces a bimonad on* ${}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$*, which possesses a right adjoint. Conversely, every bimonad on* ${}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$ *possessing a right adjoint is naturally equivalent to a bimonad induced by a right* $\mathbf{R}$*-bialgebroid.*

Another aspect of the equivalence in Theorem 3.5 is explained in Section 3.5.

In the paper [28] by Day and Street, op-monoidal monads were studied in the more general framework of pseudomonoids in monoidal bicategories. Based on Theorem 3.5, a description of bialgebroids as *strong monoidal morphisms between pseudomonoids* in the monoidal bicategory of bimodules was obtained [28, Proposition 3.3].

3.2. *Examples*

In order to make the reader more familiar with the notion, in this section, we list some examples of bialgebroids.

3.2.1. Bialgebras Obviously, a bialgebra B over a commutative ring k determines a (left or right) k-bialgebroid in which both the source and target maps are equal to the unit map $k \to \mathbf{B}$. Note, however, that there are k-bialgebroids in which the source and target maps are different, or possibly they are equal but their range is not central in the total algebra, hence they are not bialgebras (see Section 4.1.4).

3.2.2. Weak bialgebras A weak bialgebra over a commutative ring k consists of an algebra and a coalgebra structure on the same k-module B, subject to compatibility axioms generalizing the axioms of a bialgebra [29, 30]. Explicitly, the coproduct Δ

is required to be multiplicative in the sense of (3.1). Unitality of the coproduct Δ and multiplicativity of the counit ϵ are weakened to the identities

$$\begin{aligned}(\Delta(1_{\mathbf{B}}) \underset{k}{\otimes} 1_{\mathbf{B}})(1_{\mathbf{B}} \underset{k}{\otimes} \Delta(1_{\mathbf{B}})) &= (\Delta \underset{k}{\otimes} B) \circ \Delta(1_{\mathbf{B}}) \\ &= (1_{\mathbf{B}} \underset{k}{\otimes} \Delta(1_{\mathbf{B}}))(\Delta(1_{\mathbf{B}}) \underset{k}{\otimes} 1_{\mathbf{B}}) \qquad \text{and} \\ \epsilon(b1_{(1)})\epsilon(1_{(2)}b') &= \epsilon(bb') = \epsilon(b1_{(2)})\epsilon(1_{(1)}b'), \qquad \text{for } b \text{ and } b' \in B,\end{aligned}$$

respectively. Here, $1_{(1)} \otimes_k 1_{(2)}$ denotes $\Delta(1_{\mathbf{B}})$ (which may differ from $1_{\mathbf{B}} \otimes_k 1_{\mathbf{B}}$). The map

$$\sqcap^R : B \to B, \qquad b \mapsto 1_{(1)}\epsilon(b1_{(2)})$$

is checked to be idempotent. Its range is a subalgebra $\mathbf{R}$ of $\mathbf{B}$. B is an $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring, with source map given by the inclusion $\mathbf{R} \to \mathbf{B}$ and target map given by the restriction to R of the map $B \to B, b \mapsto \epsilon(b1_{(1)})1_{(2)}$. Consider B as an $\mathbf{R}$-bimodule via right multiplication by these source and target maps. The coproduct in an $\mathbf{R}$-coring B is obtained by composing $\Delta : B \to B \otimes_k B$ with the canonical epimorphism $B \otimes_k B \to B \otimes_{\mathbf{R}} B$. It has a counit $\sqcap^R$. The $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring and $\mathbf{R}$-coring structures constructed on B in this way constitute a right $\mathbf{R}$-bialgebroid.

A left $\mathbf{R}^{\mathrm{op}}$-bialgebroid structure in a weak bialgebra B is constructed symmetrically. Its source map is the inclusion map into B of the range subalgebra of the idempotent map

$$\sqcap^L : B \to B, \qquad b \mapsto \epsilon(1_{(1)}b)1_{(2)}.$$

The coproduct is obtained by composing the weak coproduct $\Delta : B \to B \otimes_k B$ with an appropriate canonical epimorphism to an $\mathbf{R}^{\mathrm{op}}$-module tensor product too.

As it was observed by Schauenburg in [31] (see also [32] Sections 1.2 and 1.3), the base algebra $\mathbf{R}$ of a weak bialgebra B is *Frobenius separable*. This means the existence of a k-module map $R \to k$ (given by the counit ϵ), possessing a dual basis $\sum_i e_i \otimes_k f_i \in R \otimes_k R$ such that $\sum_i e_i f_i = 1_{\mathbf{R}}$. The dual-basis property means $\sum_i e_i\epsilon(f_i r) = r = \sum_i \epsilon(re_i) f_i$, for all $r \in R$. In a weak bialgebra, a dual basis is given by $1_{(1)} \otimes_k \sqcap^R(1_{(2)}) \in R \otimes_k R$. In [23, 31, 32] also the converse is proved: a k-module map $\epsilon : R \to k$ on the base algebra $\mathbf{R}$ of a bialgebroid $\mathcal{B}$, with normalized dual basis $\sum_i e_i \otimes_k f_i \in R \otimes_k R$, determines a weak bialgebra structure on the underlying k-module B. In [23, Proposition 9.3], any separable algebra over a field was proved to be Frobenius separable.

Consider a small category $\mathcal{C}$ with finitely many objects. For a commutative ring k, the free k-module generated by the morphisms in $\mathcal{C}$ carries a weak bialgebra structure. The product of two morphisms is equal to the composite morphism if they are composable, and zero otherwise. The unit element is the (finite) sum of the identity morphisms for all objects. Extending the product k-linearly, we obtain a k-algebra. The coproduct is diagonal on all morphisms, that is, $\Delta(f) = f \otimes_k f$. The counit maps every morphism to 1_k. Extending the coproduct and the counit k-linearly, we obtain a k-coalgebra. The algebra and coalgebra structures constructed in this way constitute a weak bialgebra.

3.2.3. The bialgebroid $\mathbf{R} \otimes \mathbf{R}^{\mathrm{op}}$ For any algebra $\mathbf{R}$ over a commutative ring k, a simplest possible right $\mathbf{R}$-bialgebroid is constructed on the algebra $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$. Source and target maps are given by the inclusions

$$\mathbf{R} \to \mathbf{R} \underset{k}{\otimes} \mathbf{R}^{\mathrm{op}}, \quad r \mapsto r \underset{k}{\otimes} 1_{\mathbf{R}} \qquad \text{and} \qquad \mathbf{R}^{\mathrm{op}} \to \mathbf{R} \underset{k}{\otimes} \mathbf{R}^{\mathrm{op}}, \quad r \mapsto 1_{\mathbf{R}} \underset{k}{\otimes} r,$$

respectively. Coproduct is

$$\mathbf{R} \underset{k}{\otimes} \mathbf{R}^{\mathrm{op}} \to (\mathbf{R} \underset{k}{\otimes} \mathbf{R}^{\mathrm{op}}) \underset{\mathbf{R}}{\otimes} (\mathbf{R} \underset{k}{\otimes} \mathbf{R}^{\mathrm{op}}), \qquad r \underset{k}{\otimes} r' \mapsto (1_{\mathbf{R}} \underset{k}{\otimes} r') \underset{\mathbf{R}}{\otimes} (r \underset{k}{\otimes} 1_{\mathbf{R}}).$$

Counit is $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}} \to R$, $r \otimes_k r' \mapsto r'r$. The corresponding $\mathbf{R}|\mathbf{R}$-coring occurred in Theorem 3.4 (1). The opposite co-opposite of the above construction yields a left $\mathbf{R}^{\mathrm{op}}$-bialgebroid structure on $\mathbf{R}^{\mathrm{op}} \otimes_k \mathbf{R} \cong \mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$.

3.3. *Duality*

In Section 2.3, the duality between $\mathbf{R}$-rings and $\mathbf{R}$-corings has been studied. Now, we shall see how it leads to a duality of bialgebroids. Recall that the axioms of a bialgebra over a commutative ring k are self-dual. That is, the diagrams (in the category $\mathcal{M}_k$ of k-modules), expressing the bialgebra axioms, remain unchanged if all arrows are reversed. As a consequence, if a bialgebra B is finitely generated and projective as a k-module (hence possesses a dual in $\mathcal{M}_k$), then the dual has a bialgebra structure too, which is the transpose of the bialgebra structure of B. In contrast to bialgebras, the axioms of a bialgebroid are not self-dual in the same sense. Although it follows by the considerations in Section 2.3 that the $\mathbf{R}$-dual of a finitely generated and projective bialgebroid possesses an $\mathbf{R}$-ring and an $\mathbf{R}$-coring structure, it is not obvious that these dual structures constitute a bialgebroid. The fact that they do indeed was shown first by Schauenburg in [33]. A detailed study can be found also in the paper [23] by Kadison and Szlachányi. Our presentation here is closer to [23, Proposition 2.5].

Proposition 3.6. *Let* $\mathbf{R}$ *be an algebra. Consider a left bialgebroid* $\mathcal{B}$ *over* $\mathbf{R}$*, which is a finitely generated and projective left* $\mathbf{R}$*-module (via left multiplication by the source map). Then the left* $\mathbf{R}$*-dual* ${}^*B := {}_{\mathbf{R}}\mathrm{Hom}(B, R)$ *possesses a (canonical) right* $\mathbf{R}$*-bialgebroid structure.*

Applying part (1) of Proposition 2.11 to the $\mathbf{R}$-coring (B, Δ, ϵ) underlying $\mathcal{B}$, we find the existence of an $\mathbf{R}$-ring structure on ${}^*B$. Its unit map is ${}^*s : \mathbf{R} \to {}^*B$, $r \mapsto \epsilon(-)r$. Multiplication is given by

$$(\beta\beta')(b) = \beta'\big(t(\beta(b_{(2)}))b_{(1)}\big), \quad \text{for } \beta, \beta' \in B,$$

where we used that the right $\mathbf{R}$-module structure in B is given via the target map t of $\mathcal{B}$. Applying (a symmetrical version of) part (2) of Proposition 2.11 to the $\mathbf{R}$-ring $(\mathbf{B}, s)$ underlying $\mathcal{B}$, we deduce the existence of an $\mathbf{R}$-coring structure in ${}^*B$. It has a bimodule structure

$$r \cdot \beta \cdot r' = \beta\big(-s(r)\big)r' \quad \text{for } r, r' \in R,\ \beta \in {}^*B.$$

In particular, $\epsilon \cdot r = {}^*s(r)$, for $r \in R$, as expected. A to-be-target-map is defined as ${}^*t(r) := r \cdot \epsilon$, for $r \in R$. The counit in the $\mathbf{R}$-coring ${}^*B$ is ${}^*\epsilon : {}^*B \to R$, $\beta \mapsto \beta(1_{\mathbf{B}})$. The coproduct is given in terms of a dual basis $(\{a_i \in B\}, \{\alpha_i \in {}^*B\})$ as

$$ {}^*\Delta : {}^*B \to {}^*B \underset{\mathbf{R}}{\otimes} {}^*B, \qquad \beta \mapsto \sum_i \alpha_i \underset{\mathbf{R}}{\otimes} \beta(-a_i). $$

The right bialgebroid axioms are verified by direct computations.

One can apply Proposition 3.6 to the co-opposite left bialgebroid, which is a finitely generated and projective right $\mathbf{R}^{\mathrm{op}}$-module via left multiplication by the target map. In this way, one verifies that the right dual $B^* := \mathrm{Hom}_{\mathbf{R}}(B, R)$ of a left $\mathbf{R}$-bialgebroid $\mathcal{B}$, which is a finitely generated and projective right $\mathbf{R}$-module, possesses a right $\mathbf{R}$-bialgebroid structure $\mathcal{B}^* := ({}^*(\mathcal{B}_{\mathrm{cop}}))_{\mathrm{cop}}$. Note that (conventionally), multiplication is chosen in such a way that it results in right bialgebroid structures on both duals of a left bialgebroid. Applying the constructions to the opposite bialgebroid, left and right duals of a right bialgebroid, which is a finitely generated and projective $\mathbf{R}$-module on the appropriate side, are concluded to be left bialgebroids. Our convention is to choose $(\mathcal{B}^{\mathrm{op}})^* := (\mathcal{B}^*)^{\mathrm{op}}$ and ${}^*(\mathcal{B}^{\mathrm{op}}) := ({}^*\mathcal{B})^{\mathrm{op}}$.

3.4. *Construction of new bialgebroids from known ones*

In addition to the examples in Section 3.2, further (somewhat implicit) examples of bialgebroids are provided by various constructions starting with given bialgebroids.

3.4.1. Drinfel'd twist A Drinfel'd twist of a bialgebra B over a commutative ring k is a bialgebra with the same algebra structure in B and coproduct deformed (or *twisted*) by an invertible normalized 2-cocycle in B (the so called *Drinfel'd element*). In this section, we recall analogous Drinfel'd twists of bialgebroids from [25, Section 6.3]. More general twists, which do not correspond to invertible Drinfel'd elements, are studied in [8]. Such generalized twists will not be considered here.

Definition 3.7. For an algebra $\mathbf{R}$, consider a right $\mathbf{R}$-bialgebroid $\mathcal{B}$, with structure maps denoted as in Definition 3.1. An (invertible) element J of the Takeuchi product $B \times_R B$ is called an *(invertible) normalized 2-cocycle in* $\mathcal{B}$ provided it satisfies the following conditions.

(i) $(t(r) \otimes_{\mathbf{R}} s(r'))J = J(t(r) \otimes_{\mathbf{R}} s(r'))$ for $r, r' \in R$ (*bilinearity*),
(ii) $(J \otimes_{\mathbf{R}} 1_{\mathbf{B}})(\Delta \otimes_{\mathbf{R}} B)(J) = (1_B \otimes_{\mathbf{R}} J)(B \otimes_{\mathbf{R}} \Delta)(J)$ (*cocycle condition*),
(iii) $(\epsilon \otimes_{\mathbf{R}} B)(J) = 1_{\mathbf{B}} = (B \otimes_{\mathbf{R}} \epsilon)(J)$ (*normalization*).

Proposition 3.8. *Let J be an invertible normalized 2-cocycle in a right $\mathbf{R}$-bialgebroid $\mathcal{B}$. The $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring $(\mathbf{B}, s, t)$ in $\mathcal{B}$, the counit ϵ of $\mathcal{B}$, and the* twisted form $\Delta_J := J\Delta(-)J^{-1}$ *of the coproduct Δ in $\mathcal{B}$ constitute a right $\mathbf{R}$-bialgebroid $\mathcal{B}_J$.*

3.4.2. Cocycle double twist Dually to the construction in Section 3.4.1, one can leave the coproduct in a bialgebroid $\mathcal{B}$ unchanged and *twist* multiplication by an

invertible normalized 2-cocycle *on* $\mathcal{B}$. For an algebra $\mathbf{R}$ over a commutative ring k, consider a left $\mathbf{R}$-bialgebroid $\mathcal{B}$, with structure maps denoted as in Definition 3.3. Recall from Theorem 3.4 that the $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-module tensor product $B \otimes_{\mathbf{R}\otimes_k\mathbf{R}^{\mathrm{op}}} B$ (with respect to the right (resp. left) actions given by right (resp. left) multiplications by s and t) is an R-coring. It has a bimodule structure $r \cdot (b \otimes b') \cdot r' := s(r)t(r')b \otimes b'$, coproduct $b \otimes b' \mapsto (b_{(1)} \otimes b'_{(1)}) \otimes_{\mathbf{R}} (b_{(2)} \otimes b'_{(2)})$, and counit $b \otimes b' \mapsto \epsilon(bb')$. Hence, there is a corresponding convolution algebra ${}_{\mathbf{R}}\mathrm{Hom}_{\mathbf{R}}(B \otimes_{\mathbf{R}\otimes_k\mathbf{R}^{\mathrm{op}}} B, R)$ with multiplication $(f \diamond g)(b \otimes b') := f(b_{(1)} \otimes b'_{(1)})g(b_{(2)} \otimes b'_{(2)})$.

Definition 3.9. Let $\mathbf{R}$ be an algebra over a commutative ring k and let $\mathcal{B}$ be a left $\mathbf{R}$-bialgebroid, with structure maps denoted as in Definition 3.3. An (invertible) element of the convolution algebra ${}_{\mathbf{R}}\mathrm{Hom}_{\mathbf{R}}(B \otimes_{\mathbf{R}\otimes_k\mathbf{R}^{\mathrm{op}}} B, R)$ is called an *(invertible) normalized 2-cocycle on* $\mathcal{B}$ provided it satisfies the following conditions, for $b, b', b'' \in B$ and $r, r' \in R$.

(i) $\sigma(s(r)t(r')b, b') = r\sigma(b, b')r'$ (*bilinearity*),
(ii) $\sigma(b, s(\sigma(b'_{(1)}, b''_{(1)}))b'_{(2)}b''_{(2)}) = \sigma(s(\sigma(b_{(1)}, b'_{(1)}))b_{(2)}b'_{(2)}, b'')$ (*cocycle condition*),
(iii) $\sigma(1_{\mathbf{B}}, b) = \epsilon(b) = \sigma(b, 1_{\mathbf{B}})$ (*normalization*).

Proposition 3.10. *Let σ be an invertible normalized 2-cocycle on a left $\mathbf{R}$-bialgebroid $\mathcal{B}$, with inverse $\widetilde{\sigma}$. The source and target maps s and t in $\mathcal{B}$, the $\mathbf{R}$-coring (B, Δ, ϵ) in $\mathcal{B}$, and the* twisted product $b \cdot_\sigma b' := s\big(\sigma(b_{(1)}, b'_{(1)})\big)t\big(\widetilde{\sigma}(b_{(3)}, b'_{(3)})\big)b_{(2)}b'_{(2)}$, *for $b, b' \in B$, constitute a left $\mathbf{R}$-bialgebroid $\mathcal{B}^\sigma$.*

3.4.3. Duality The constructions in Sections 3.4.1 and 3.4.2 are dual of each other in the following sense.

Proposition 3.11. *For an algebra $\mathbf{R}$, let $\mathcal{B}$ be a left $\mathbf{R}$-bialgebroid, which is a finitely generated and projective right $\mathbf{R}$-module via left multiplication by the target map t, and consider the right dual right $\mathbf{R}$-bialgebroid $\mathcal{B}^*$. The following statements hold.*

(1) *An element $J = \sum_k \xi_k \otimes_{\mathbf{R}} \zeta_k \in B^* \times_{\mathbf{R}} B^*$ is an invertible normalized 2-cocycle in $\mathcal{B}^*$ if and only if*

$$\sigma_J : B \underset{\mathbf{R}\otimes\mathbf{R}^{\mathrm{op}}}{\otimes} B \to R, \qquad \sigma_J(b, b') := \sum_k \xi_k\big(bt(\zeta_k(b'))\big) \tag{3.7}$$

is an invertible normalized 2-cocycle on $\mathcal{B}$.
(2) *Assume the equivalent properties in part (1). The right bialgebroid $(\mathcal{B}^*)_J$, obtained by twisting the coproduct of $\mathcal{B}^*$ by the cocycle J, is right dual of the left bialgebroid $\mathcal{B}^{\sigma_J}$, obtained by twisting the product in $\mathcal{B}$ by the cocycle σ_J in* (3.7).

The inverse of the construction in (3.7) is given by associating to an invertible normalized 2-cocycle σ on $\mathcal{B}$ an invertible normalized 2-cocycle in $\mathcal{B}^*$: $J_\sigma := \sum_i \sigma(-, a_i) \otimes_{\mathbf{R}} \alpha_i$, where $(\{a_i \in B\}, \{\alpha_i \in B^*\})$ is a dual basis.

3.4.4. Drinfel'd double For a Hopf algebra B over a commutative ring k such that B is a finitely generated and projective k-module, the k-module $\mathcal{D}(B) := B \otimes_k B^*$ has a bialgebra (in fact Hopf algebra) structure. It is known as the *Drinfel'd double* of B. The category of $\mathcal{D}(B)$-modules is isomorphic (as a monoidal category) to the category of Yetter–Drinfel'd modules for B and also to the monoidal center of the category of B-modules. These results were extended to certain bialgebroids by Schauenburg in [33].

In this section, let $\mathcal{B}$ be a left bialgebroid over a k-algebra $\mathbf{R}$, finitely generated and projective as a right $\mathbf{R}$-module. Assume in addition that the following map is bijective.

$$\vartheta : B \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} B \to B \underset{\mathbf{R}}{\otimes} B, \qquad b \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} b' \mapsto b_{(1)} \underset{\mathbf{R}}{\otimes} b_{(2)} b', \tag{3.8}$$

where in the domain of ϑ, the module structures are given by right and left multiplications by the target map, and in the codomain, the module structures are given by left multiplications by the target and source maps. Left bialgebroids, for which the map (3.8) is bijective, were named (*left*) $\times_R$*-Hopf algebras* in [33] and they are discussed in more detail in Section 4.6.2. In the following Proposition 3.12, Sweedler's index notation is used for the coproducts and also the index notation $\vartheta^{-1}(b \otimes_{\mathbf{R}} 1_{\mathbf{B}}) = b^{\langle 1 \rangle} \otimes_{\mathbf{R}^{\mathrm{op}}} b^{\langle 2 \rangle}$, where implicit summation is understood.

Proposition 3.12. *Let $\mathbf{R}$ be an algebra over a commutative ring k. Let $\mathcal{B}$ be a left $\times_R$-Hopf algebra, which is a finitely generated and projective right $\mathbf{R}$-module via left multiplication by the target map. Denote the structure maps in $\mathcal{B}$ as in Definition 3.3. Consider B as a right $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-module via right multiplications by the source and target maps s and t in the left bialgebroid $\mathcal{B}$. Consider the right dual B^* as a left $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-module via right multiplications by the target and source maps t^* and s^* in the right bialgebroid $\mathcal{B}^*$. The tensor product $\mathcal{D}(\mathcal{B}) := B \otimes_{\mathbf{R} \otimes \mathbf{R}^{\mathrm{op}}} B^*$ has a left $\mathbf{R}$-bialgebroid structure as follows. Multiplication is given by*

$$(b \bowtie \beta)(b' \bowtie \beta') := bs\big(\beta^{(1)}(b'_{(1)})\big) b'_{(2)}{}^{\langle 1 \rangle} \bowtie \beta' \beta^{(2)} s^*\big(\beta^{(3)}(b'_{(2)}{}^{\langle 2 \rangle})\big), \qquad \text{for } b \bowtie \beta,\ b' \bowtie \beta' \in \mathcal{D}(\mathcal{B}).$$

The source and target maps are

$$\mathbf{R} \to \mathcal{D}(\mathcal{B}), \quad r \mapsto 1_{\mathbf{B}} \bowtie t^*(r) \quad \text{and} \quad \mathbf{R}^{\mathrm{op}} \to \mathcal{D}(\mathcal{B}), \quad r \mapsto 1_{\mathbf{B}} \bowtie s^*(r),$$

respectively. The coproduct is

$$\mathcal{D}(\mathcal{B}) \to \mathcal{D}(\mathcal{B}) \underset{\mathbf{R}}{\otimes} \mathcal{D}(\mathcal{B}), \qquad b \bowtie \beta \mapsto (b_{(1)} \bowtie \beta^{(1)}) \underset{\mathbf{R}}{\otimes} (b_{(2)} \bowtie \beta^{(2)}),$$

and the counit is $\mathcal{D}(\mathcal{B}) \to R$, $b \bowtie \beta \mapsto \epsilon\big(bs(\beta(1_{\mathbf{B}}))\big)$.

It is not known if the Drinfel'd double (or the dual) of a $\times_R$-Hopf algebra is a $\times_R$-Hopf algebra as well.

3.4.5. Morita base change In contrast to the previous sections, constructions in this section and in the forthcoming ones *change the base algebra of a bialgebroid*.

Let $\mathbf{R}$ be an algebra over a commutative ring k and let $\mathcal{B}$ be a left $\mathbf{R}$-bialgebroid. Denote the structure maps of $\mathcal{B}$ as in Definition 3.3. Let $\widetilde{\mathbf{R}}$ be a k-algebra, which is Morita equivalent to $\mathbf{R}$. Fix a strict Morita context $(\mathbf{R}, \widetilde{\mathbf{R}}, P, Q, \bullet, \circ)$. Denote the inverse image of $1_{\mathbf{R}}$ under the map $\bullet$ by $\sum_i q_i \otimes_{\widetilde{\mathbf{R}}} p^i \in Q \otimes_{\widetilde{\mathbf{R}}} P$ and denote the inverse image of $1_{\widetilde{\mathbf{R}}}$ under the map $\circ$ by $\sum_j p_j \otimes_{\mathbf{R}} q^j \in P \otimes_{\mathbf{R}} Q$. The $\widetilde{\mathbf{R}}$-$\mathbf{R}$ bimodule P determines a canonical $\mathbf{R}^{\mathrm{op}}$-$\widetilde{\mathbf{R}}^{\mathrm{op}}$ bimodule $\overline{P}$. Similarly, there is a canonical $\widetilde{\mathbf{R}}^{\mathrm{op}}$-$\mathbf{R}^{\mathrm{op}}$ bimodule $\overline{Q}$. In [34, Section 5], a left $\widetilde{\mathbf{R}}$-bialgebroid structure was constructed on the k-module $\widetilde{B} := (P \otimes_k \overline{Q}) \otimes B \otimes (Q \otimes_k \overline{P})$, where an unadorned tensor product is meant to be one over $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$. Multiplication is given by

$$\begin{aligned}&[(p_1 \otimes q_1) \otimes b \otimes (q_2 \otimes p_2)][(p_1' \otimes q_1') \otimes b' \otimes (q_2' \otimes p_2')]\\ &\quad := (p_1 \otimes q_1) \otimes bs(q_2 \bullet p_1')t(q_1' \bullet p_2)b' \otimes (q_2' \otimes p_2').\end{aligned}$$

The source and target maps in $\widetilde{\mathcal{B}}$ are, for $\widetilde{r} \in \widetilde{R}$,

$$\widetilde{r} \mapsto \sum_{j,j'} (\widetilde{r} \cdot p_{j'} \otimes q^j) \otimes 1_{\mathbf{B}} \otimes (q^{j'} \otimes p_j) \quad \text{and} \quad \widetilde{r} \mapsto \sum_{j,j'} (p_{j'} \otimes q^j \cdot \widetilde{r}) \otimes 1_{\mathbf{B}} \otimes (q^{j'} \otimes p_j),$$

respectively. The coproduct and counit are given on an element $(p_1 \otimes q_1) \otimes b \otimes (q_2 \otimes p_2) \in \widetilde{B}$ as

$$\begin{aligned}(p_1 \otimes q_1) \otimes b \otimes (q_2 \otimes p_2) &\mapsto \sum_{i,j} [(p_1 \otimes q_i) \otimes b_{(1)} \otimes (q_2 \otimes p_j)] \underset{\mathbf{R}}{\otimes}\\ &\qquad [(p^i \otimes q_1) \otimes b_{(2)} \otimes (q^j \otimes p_2)] \quad \text{and}\\ (p_1 \otimes q_1) \otimes b \otimes (q_2 \otimes p_2) &\mapsto p_1 \cdot \epsilon\big(bs(q_2 \bullet p_2)\big) \circ q_1,\end{aligned}$$

respectively. A generalization and a more conceptual background of Morita base change in bialgebroids are presented in Section 3.5.

3.4.6. Connes–Moscovici bialgebroids The following bialgebroid was constructed in [10] in the framework of transverse geometry. Let H be a Hopf algebra over a commutative ring k, with coproduct $\delta : h \mapsto h_{(1)} \otimes_k h_{(2)}$ and counit ε. Let $\mathbf{R}$ be a left H-module algebra. Consider the k-module $B := R \otimes_k H \otimes_k R$. It can be equipped with an associative multiplication

$$(r_1 \underset{k}{\otimes} h \underset{k}{\otimes} r_2)(r_1' \underset{k}{\otimes} h' \underset{k}{\otimes} r_2') := r_1(h_{(1)} \cdot r_1') \underset{k}{\otimes} h_{(2)}h' \underset{k}{\otimes} (h_{(3)} \cdot r_2')r_2,$$

with unit $1_{\mathbf{R}} \otimes_k 1_{\mathbf{H}} \otimes_k 1_{\mathbf{R}}$. The algebra $\mathbf{B}$ can be made a left $\mathbf{R}$-bialgebroid with source and target maps

$$\mathbf{R} \to \mathbf{B}, \quad r \mapsto r \underset{k}{\otimes} 1_{\mathbf{H}} \underset{k}{\otimes} 1_{\mathbf{R}} \qquad \text{and} \qquad \mathbf{R}^{\mathrm{op}} \to \mathbf{B}, \quad r \mapsto 1_{\mathbf{R}} \underset{k}{\otimes} 1_{\mathbf{H}} \underset{k}{\otimes} r,$$

respectively. The coproduct and counit are

$$r \underset{k}{\otimes} h \underset{k}{\otimes} r' \mapsto (r \underset{k}{\otimes} h_{(1)} \underset{k}{\otimes} 1_{\mathbf{R}}) \underset{\mathbf{R}}{\otimes} (1_{\mathbf{R}} \underset{k}{\otimes} h_{(2)} \underset{k}{\otimes} r') \quad \text{and} \quad r \underset{k}{\otimes} h \underset{k}{\otimes} r' \mapsto r\varepsilon(h)r'.$$

3.4.7. Scalar extension Let B be a bialgebra over a commutative ring k and let A be an algebra in the category of right–right Yetter–Drinfel'd modules of B. Recall that this means A is a right B-module algebra and a right B-comodule algebra such that the following compatibility condition holds, for $a \in A$ and $b \in B$.

$$(a \cdot b_{(2)})_{[0]} \underset{k}{\otimes} b_{(1)}(a \cdot b_{(2)})_{[1]} = a_{[0]} \cdot b_{(1)} \underset{k}{\otimes} a_{[1]} b_{(2)}.$$

The category of right–right Yetter–Drinfel'd modules is prebraided. Assume that A is braided commutative, that is, for $a, a' \in A$, the identity $a'_{[0]}(a \cdot a'_{[1]}) = aa'$ holds. Under these assumptions, it follows by a symmetrical version of [22, Theorem 4.1] that the smash product algebra has a right $\mathbf{A}$-bialgebroid structure. Recall that the smash product algebra is the k-module $A\#B := A \otimes_k B$, with multiplication $(a\#b)(a'\#b') := a'(a \cdot b'_{(1)}) \underset{k}{\otimes} bb'_{(2)}$. The source and target maps are

$$s : \mathbf{A} \to A\#B, a \mapsto a_{[0]}\#a_{[1]} \quad \text{and} \quad t : \mathbf{A}^{\mathrm{op}} \to A\#B, a \mapsto a\#1_{\mathbf{B}},$$

respectively. The coproduct is

$$\Delta : A\#B \to (A\#B) \underset{\mathbf{A}}{\otimes} (A\#B), \qquad a\#b \mapsto (a\#b_{(1)}) \underset{\mathbf{A}}{\otimes} (1_{\mathbf{A}}\#b_{(2)})$$

and the counit is given in terms of the counit ε in B as $A\#B \to A$, $a\#b \mapsto a\varepsilon(b)$. The name *scalar extension* comes from the feature that the base algebra k of B (the subalgebra of "scalars") becomes replaced by the base algebra $\mathbf{A}$ of $A\#B$.

A solution $\mathcal{R}$ of the quantum Yang–Baxter equation on a finite-dimensional vector space determines a bialgebra $B(\mathcal{R})$ via the so called Faddeev-Reshetikhin-Takhtajan construction. In [22, Proposition 4.3], a braided commutative algebra in the category of Yetter–Drinfel'd modules for $B(\mathcal{R})$ was constructed; thus, a bialgebroid was associated to a finite-dimensional solution $\mathcal{R}$ of the quantum Yang–Baxter equation.

The construction above of a scalar extension was extended in [12, Theorem 4.6]. Following it, a smash product of a right bialgebroid $\mathcal{B}$ over an algebra $\mathbf{R}$, with a braided commutative algebra A in the category of right–right Yetter–Drinfel'd modules for $\mathcal{B}$, is shown to possess a right $\mathbf{A}$-bialgebroid structure. The fundamental importance of scalar extensions from the point of view of Galois extensions by bialgebroids is discussed in Section 3.7.

3.5. *The monoidal category of modules*

An algebra $\mathbf{B}$ over a commutative ring k is known to have a k-bialgebra structure if and only if the category of (left or right) $\mathbf{B}$-modules is monoidal such that the forgetful functor to $\mathcal{M}_k$ is strict monoidal [35]. A generalization [36, Theorem 5.1] of this fact to bialgebroids is due to Schauenburg. Recall that any (right) module of an $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring $(\mathbf{B}, s, t)$ is an $\mathbf{R}$-bimodule via the actions by $t(r)$ and $s(r)$, for $r \in R$.

THEOREM 3.13. *For an algebra* $\mathbf{R}$ *over a commutative ring* k*, the following data are equivalent on an* $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$*-ring* $(\mathbf{B}, s, t)$.

(1) *A right bialgebroid structure on* $(\mathbf{B}, s, t)$*;*
(2) *A monoidal structure on the category* $\mathcal{M}_{\mathbf{B}}$ *of right* $\mathbf{B}$*-modules such that the forgetful functor* $\mathcal{M}_{\mathbf{B}} \to {}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$ *is strict monoidal.*

Applying Theorem 3.13 to the opposite $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring, an analogous equivalence is obtained between left bialgebroid structures and monoidal structures on the category of left modules. At the heart of Theorem 3.13 lies the fact that the right regular $\mathbf{B}$-module is a generator in the category $\mathcal{M}_{\mathbf{B}}$. Hence, the $\mathbf{B}$-module structure on the $\mathbf{R}$-module tensor product of any two $\mathbf{B}$-modules can be expressed in terms of the action on $1_{\mathbf{B}} \otimes_{\mathbf{R}} 1_{\mathbf{B}} \in B \otimes_{\mathbf{R}} B$, which defines a coproduct. In terms of a coproduct $b \mapsto b^{(1)} \otimes_{\mathbf{R}} b^{(2)} := (1_{\mathbf{B}} \otimes_{\mathbf{R}} 1_{\mathbf{B}}) \cdot b$, the right $\mathbf{B}$-action in the $\mathbf{R}$-module tensor product of two right $\mathbf{B}$-modules M and N is written as

$$(m \underset{\mathbf{R}}{\otimes} n) \cdot b = m \cdot b^{(1)} \underset{\mathbf{R}}{\otimes} n \cdot b^{(2)}, \qquad \text{for } m \underset{\mathbf{R}}{\otimes} n \in M \underset{\mathbf{R}}{\otimes} N,\ b \in B. \tag{3.9}$$

A $\mathbf{B}$-module structure on the monoidal unit R is equivalent to a right character $\epsilon : B \to R$ by Lemma 2.5.

Recall that, for an $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring $(\mathbf{B}, s, t)$, the category $\mathcal{M}_{\mathbf{B}}$ is isomorphic to the category of algebras for the monad $- \otimes_{\mathbf{R}\otimes\mathbf{R}^{\mathrm{op}}} B$ on ${}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$ (where $\mathbf{R}$-bimodules are considered as right $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-modules). In light of this fact, Theorem 3.13 is a particular case of a question discussed by Moerdijk in [26]: Having a monad $\mathbb{B}$ on a monoidal category $\mathcal{M}$, the monoidal structure of $\mathcal{M}$ lifts to a monoidal structure on the category $\mathcal{M}_{\mathbb{B}}$ of $\mathbb{B}$-algebras (in the sense that the forgetful functor $\mathcal{M}_{\mathbb{B}} \to \mathcal{M}$ is strict monoidal, as in part (2) in Theorem 3.13) if and only if $\mathbb{B}$ is a monoid in the category of op-monoidal endofunctors on $\mathcal{M}$, that is, a *Hopf monad* in the terminology of [26] (called a *bimonad* in [27]). Comparing this result with Theorem 3.13, we obtain other evidence for a characterization of bialgebroids as bimonads as in Theorem 3.5.

In the paper [25] by Szlachányi, it was investigated what bialgebroids possess monoidally equivalent module categories. That is, a *monoidal Morita theory* of bialgebroids was developed. Based on Theorem 3.5, one main result in [25] can be reformulated as in Theorem 3.14 below. Recall that a bimodule, for two monads $\mathbb{B} : \mathcal{M} \to \mathcal{M}$ and $\mathbb{B}' : \mathcal{M}' \to \mathcal{M}'$, is a functor $\mathbb{M} : \mathcal{M} \to \mathcal{M}'$ together with natural transformations $\varrho : \mathbb{M}\mathbb{B} \to \mathbb{M}$ and $\lambda : \mathbb{B}'\mathbb{M} \to \mathbb{M}$, satisfying the usual compatibility conditions for the right and left actions in a bimodule. If any pair of parallel morphisms in $\mathcal{M}'$ possesses a coequalizer, and $\mathbb{B}'$ preserves coequalizers, then the bimodule $\mathbb{M}$ induces a functor $\overline{\mathbb{M}} : \mathcal{M}_{\mathbb{B}} \to \mathcal{M}'_{\mathbb{B}'}$, between the categories of algebras for $\mathbb{B}$ and $\mathbb{B}'$, respectively, with object map $(V, v) \mapsto \mathrm{Coeq}(\mathbb{M}(v), \varrho_V)$. If both categories $\mathcal{M}$ and $\mathcal{M}'$ possess coequalizers and both $\mathbb{B}$ and $\mathbb{B}'$ preserve them, then one can define the inverse of a bimodule as in Morita theory. A $\mathbb{B}'$-$\mathbb{B}$ bimodule $\mathbb{M}'$ is said to be the inverse of $\mathbb{M}$ provided that $\overline{\mathbb{M}'}\,\overline{\mathbb{M}}$ is naturally equivalent to the identity functor on $\mathcal{M}_{\mathbb{B}}$, and $\overline{\mathbb{M}}\,\overline{\mathbb{M}'}$ is naturally equivalent to the identity functor on $\mathcal{M}_{\mathbb{B}'}$.

THEOREM 3.14. *For two right bialgebroids* $\mathcal{B}$ *and* $\mathcal{B}'$*, over respective base algebras* $\mathbf{R}$ *and* $\mathbf{R}'$*, the right module categories* $\mathcal{M}_{\mathbf{B}}$ *and* $\mathcal{M}_{\mathbf{B}'}$ *are monoidally equivalent if*

and only if there exists an invertible bimodule in the 2-category of op-monoidal left adjoint functors, for the monads $-\otimes_{\mathbf{R}\otimes\mathbf{R}^{\mathrm{op}}} B : {}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}} \to {}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$ *and* $-\otimes_{\mathbf{R}'\otimes\mathbf{R}'^{op}} B' : {}_{\mathbf{R}'}\mathcal{M}_{\mathbf{R}'} \to {}_{\mathbf{R}'}\mathcal{M}_{\mathbf{R}'}$.

By standard Morita theory, an equivalence $\overline{\mathbb{M}} : \mathcal{M}_{\mathbf{B}} \to \mathcal{M}_{\mathbf{B}'}$ is of the form $\overline{\mathbb{M}} = -\otimes_{\mathbf{B}} M$, for some invertible $\mathbf{B}$-$\mathbf{B}'$ bimodule M. In [25], monoidality of the equivalence is translated to properties of the Morita equivalence bimodule M.

In [24, Definition 2.1], two algebras $\mathbf{R}$ and $\widetilde{\mathbf{R}}$ over a commutative ring k were said to be $\sqrt{\textit{Morita-equivalent}}$ whenever the bimodule categories ${}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$ and ${}_{\widetilde{\mathbf{R}}}\mathcal{M}_{\widetilde{\mathbf{R}}}$ are strictly equivalent as k-linear monoidal categories. This property implies that the algebras $\mathbf{R}\otimes_k \mathbf{R}^{\mathrm{op}}$ and $\widetilde{\mathbf{R}}\otimes_k \widetilde{\mathbf{R}}^{\mathrm{op}}$ are $\sqrt{\text{Morita}}$-equivalent (but not conversely). In this situation, any $\mathbf{R}\otimes_k \mathbf{R}^{\mathrm{op}}$-ring B (i.e. monoid in the category of $\mathbf{R}\otimes_k \mathbf{R}^{\mathrm{op}}$-bimodules) determines an $\widetilde{\mathbf{R}}\otimes_k \widetilde{\mathbf{R}}^{\mathrm{op}}$-ring $\widetilde{B}$, with underlying k-algebra $\widetilde{\mathbf{B}}$ Morita equivalent to $\mathbf{B}$. If $\mathcal{B}$ is a right $\mathbf{R}$-bialgebroid, then the forgetful functor $\mathcal{M}_{\mathbf{B}} \to {}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$ is strict monoidal by Theorem 3.13. Hence, the equivalence $\mathcal{M}_{\mathbf{B}} \cong \mathcal{M}_{\widetilde{\mathbf{B}}}$ can be used to induce a monoidal structure on $\mathcal{M}_{\widetilde{\mathbf{B}}}$ such that the forgetful functor $\mathcal{M}_{\widetilde{\mathbf{B}}} \to {}_{\widetilde{\mathbf{R}}}\mathcal{M}_{\widetilde{\mathbf{R}}}$ is strict monoidal. By Theorem 3.13, we conclude that there is a right $\widetilde{\mathbf{R}}$-bialgebroid structure on $\widetilde{B}$. In [34], the bialgebroid $\widetilde{\mathcal{B}}$ was said to be obtained from $\mathcal{B}$ via $\sqrt{\textit{Morita}}$-*base change*. Since Morita equivalent algebras are also $\sqrt{\text{Morita}}$-equivalent (but not conversely), the construction in Section 3.4.5 is a special instance of a $\sqrt{\text{Morita}}$-base change.

For a right $\mathbf{R}$-bialgebroid $\mathcal{B}$, with structure maps denoted as in Definition 3.1, consider B as an $\mathbf{R}$-bimodule (or $\mathbf{R}^{\mathrm{op}}$-bimodule) via right multiplications by the source map s and the target map t. Both $B\otimes_{\mathbf{R}} B$ and $B\otimes_{\mathbf{R}^{\mathrm{op}}} B$ are left modules for $\mathbf{B}\otimes_k \mathbf{B}$ via the regular actions on the two factors. Associated to $\mathcal{B}$, we construct a category $\mathcal{C}(\mathcal{B})$, with two objects $\circ$ and $\bullet$. The morphisms with source $\circ$ are defined to be elements of $B\otimes_{\mathbf{R}} B$ and the morphisms with source $\bullet$ are elements of $B\otimes_{\mathbf{R}^{\mathrm{op}}} B$. The morphisms F, with target $\circ$ and $\bullet$, are required to satisfy the following ($\mathbf{R}$-centralizing) conditions $(T\circ)$ and $(T\bullet)$, respectively, for all $r \in R$.

$$(T\circ) \qquad (s(r)\underset{k}{\otimes} 1_{\mathbf{B}})\cdot F = (1_{\mathbf{B}}\underset{k}{\otimes} t(r))\cdot F,$$

$$(T\bullet) \qquad (t(r)\underset{k}{\otimes} 1_{\mathbf{B}})\cdot F = (1_{\mathbf{B}}\underset{k}{\otimes} s(r))\cdot F.$$

Via composition given by factorwise multiplication, $\mathcal{C}(\mathcal{B})$ is a category. The unit morphisms at the objects $\circ$ and $\bullet$ are $1_{\mathbf{B}}\otimes_{\mathbf{R}} 1_{\mathbf{B}}$ and $1_{\mathbf{B}}\otimes_{\mathbf{R}^{\mathrm{op}}} 1_{\mathbf{B}}$, respectively. The range of the coproduct Δ lies in $\mathrm{Hom}(\circ,\circ) = B\times_R B$, and the range of the co-opposite coproduct Δ_{cop} lies in $\mathrm{Hom}(\bullet,\bullet) = B_{\mathrm{cop}}\times_{R^{\mathrm{op}}} B_{\mathrm{cop}}$. In terms of the category $\mathcal{C}(\mathcal{B})$, the definition of a quasi-triangular bialgebroid as formulated in [3, Proposition 3.13] can be described as follows.

Definition 3.15. For a right $\mathbf{R}$-bialgebroid $\mathcal{B}$, with structure maps denoted as in Definition 3.1, let $\mathcal{C}(\mathcal{B})$ be the category constructed above. An invertible morphism $\mathcal{R} = \mathcal{R}^1\otimes_{\mathbf{R}^{\mathrm{op}}}\mathcal{R}^2 \in \mathrm{Hom}(\bullet,\circ)$ (where implicit summation is understood)

is a *universal R-matrix* provided that for any $b \in B$, the following identity holds in $\mathrm{Hom}(\bullet, \circ)$:

$$\Delta(b)\mathcal{R} = \mathcal{R}\Delta_{\mathrm{cop}}(b)$$

and

$$(\Delta_{\mathrm{cop}} \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} B)(\mathcal{R}) = \mathcal{R} \triangleleft \mathcal{R} \qquad \text{and} \qquad (B \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} \Delta_{\mathrm{cop}})(\mathcal{R}) = \mathcal{R} \triangleright \mathcal{R},$$

where the (well defined) maps

$$- \triangleleft \mathcal{R} : B \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} B \to B \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} B \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} B, \qquad b \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} b' \mapsto b \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} \mathcal{R}^1 \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} b'\mathcal{R}^2 \quad \text{and}$$

$$\mathcal{R} \triangleright - : B \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} B \to B \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} B \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} B, \qquad b \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} b' \mapsto b\mathcal{R}^1 \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} \mathcal{R}^2 \underset{\mathbf{R}^{\mathrm{op}}}{\otimes} b'$$

are used. A right bialgebroid $\mathcal{B}$ with a given universal R-matrix $\mathcal{R}$ is called a *quasi-triangular* bialgebroid.

The following Theorem 3.16 was obtained in [3, Theorem 3.15], as a generalization of an analogous result for quasi-triangular bialgebras.

THEOREM 3.16. *Consider a quasi-triangular right bialgebroid $(\mathcal{B}, \mathcal{R})$ over a base algebra* $\mathbf{R}$. *The monoidal category of right B-modules is braided, with as braiding the natural isomorphism*

$$M \underset{\mathbf{R}}{\otimes} M' \to M' \underset{\mathbf{R}}{\otimes} M, \qquad m \underset{\mathbf{R}}{\otimes} m' \mapsto m' \cdot \mathcal{R}^2 \underset{\mathbf{R}}{\otimes} m \cdot \mathcal{R}^1.$$

3.6. *The monoidal category of comodules*

For a bialgebra over a commutative ring k, not only the category of modules but also the category of (left or right) comodules has a monoidal structure such that the forgetful functor to $\mathcal{M}_k$ is strict monoidal. In trying to prove an analogue of this result for bialgebroids, the first step is to find a forgetful functor. A right, say, comodule of (the constituent coring in) an $\mathbf{R}$-bialgebroid is by definition only a right $\mathbf{R}$-module. To obtain a forgetful functor to the *monoidal* category of $\mathbf{R}$-*bimodules*, the following lemma [75, Lemma 1.4.1] is needed.

LEMMA 3.17. *Let* $\mathbf{R}$ *and* $\mathbf{S}$ *be two algebras over a commutative ring k and let C be an* $\mathbf{S}|\mathbf{R}$*-coring. Any right comodule (M, ϱ^M) of the* $\mathbf{R}$*-coring C can be equipped with a unique left* $\mathbf{S}$*-module structure such that $\varrho^M(m)$ belongs to the center of the* $\mathbf{S}$*-bimodule $M \otimes_{\mathbf{R}} C$, for every $m \in M$. This unique left* $\mathbf{S}$*-action makes M an* $\mathbf{S}$-$\mathbf{R}$ *bimodule. Every C-comodule map becomes an* $\mathbf{S}$-$\mathbf{R}$ *bimodule map. That is, there is a forgetful functor $\mathcal{M}^C \to {}_{\mathbf{S}}\mathcal{M}_{\mathbf{R}}$.*

The left $\mathbf{S}$-action on a right comodule (M, ϱ^M) of the $\mathbf{R}$-coring C is constructed as

$$s \cdot m := m^{[0]} \cdot \epsilon(m^{[1]} \cdot (s \underset{k}{\otimes} 1_{\mathbf{R}})), \quad \text{for } s \in \mathbf{S},\ m \in M. \tag{3.10}$$

In particular, (3.10) can be used to equip a right comodule of a right $\mathbf{R}$-bialgebroid with an $\mathbf{R}$-$\mathbf{R}$ bimodule structure. Applying the construction to co-opposite and opposite bialgebroids, forgetful functors are obtained from categories of left and right comodules of left and right $\mathbf{R}$-bialgebroids to ${}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$.

THEOREM 3.18. *Let $\mathbf{R}$ be an algebra and let $\mathcal{B}$ be a right $\mathbf{R}$-bialgebroid. The category $\mathcal{M}^{\mathcal{B}}$ of right $\mathcal{B}$-comodules is monoidal such that the forgetful functor $\mathcal{M}^{\mathcal{B}} \to {}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$ is strict monoidal.*

The monoidal unit R in ${}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$ is a right $\mathcal{B}$-comodule via a coaction provided by the source map. One has to verify that, for any two right $\mathcal{B}$-comodules M and N, the diagonal coaction

$$M \underset{\mathbf{R}}{\otimes} N \to M \underset{\mathbf{R}}{\otimes} N \underset{\mathbf{R}}{\otimes} B, \qquad m \underset{\mathbf{R}}{\otimes} n \mapsto m^{[0]} \underset{\mathbf{R}}{\otimes} n^{[0]} \underset{\mathbf{R}}{\otimes} m^{[1]} n^{[1]} \tag{3.11}$$

is well defined and that the natural coherence isomorphisms in ${}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$ are $\mathcal{B}$-comodule maps.

We do not know a converse of Theorem 3.18, that is, an analog of the correspondence (2)⇒(1) in Theorem 3.13 for the category of comodules. A reason for this is that (in contrast to modules of an $\mathbf{R}$-bialgebroid $\mathcal{B}$, which are algebras for the monad $- \otimes_{\mathbf{R}\otimes\mathbf{R}^{\mathrm{op}}} B$ on the monoidal category of $\mathbf{R}$-bimodules) it is not known whether comodules can be described as coalgebras of a comonad on ${}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$.

However, the definition of a bialgebroid can be dualized in the sense of reversing all arrows in the diagrams in $\mathcal{M}_k$, expressing the axioms of a bialgebroid over a k-algebra. For a flat k-coalgebra C, (e.g. when k is a field), C-bicomodules constitute a monoidal category ${}^{C}\mathcal{M}^{C}$. A monoidal structure is given by *cotensor products* – a notion dual to a module tensor product. That is, for C-bicomodules M and N, the cotensor product $M\Box_C N$ is the equalizer of the maps $\varrho^M \otimes_k N$ and $M \otimes_k {}^{N}\varrho$, where ϱ^M is the right coaction on M and ${}^{N}\varrho$ is the left coaction on N. Flatness of the k-module C implies that $M\Box_C N$ is a C-bicomodule via the left C-coaction on M and right C-coaction on N. In this case, dualization of the bialgebroid axioms leads to the notion of a *bicoalgebroid* over the k-coalgebra C (see [22]). The relation between C-bicoalgebroid structures on a comonoid in ${}^{C}\mathcal{M}^{C}$ and strict monoidal structures on the forgetful functor from its comodule category to ${}^{C}\mathcal{M}^{C}$ is studied in [37] and [38].

Applying Theorem 3.18 to the co-opposite bialgebroid, strict monoidality follows for the forgetful functor ${}^{\mathcal{B}}\mathcal{M} \to {}_{\mathbf{R}^{\mathrm{op}}}\mathcal{M}_{\mathbf{R}^{\mathrm{op}}}$, for a right $\mathbf{R}$-bialgebroid $\mathcal{B}$. Applying Theorem 3.18 to opposite bialgebroids, it follows that the forgetful functors ${}^{\mathcal{B}}\mathcal{M} \to {}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$ and $\mathcal{M}^{\mathcal{B}} \to {}_{\mathbf{R}^{\mathrm{op}}}\mathcal{M}_{\mathbf{R}^{\mathrm{op}}}$ are strict monoidal, for a left $\mathbf{R}$-bialgebroid $\mathcal{B}$. Note that, for an $\mathbf{R}$-bialgebroid $\mathcal{B}$, which is a finitely generated and projective $\mathbf{R}$-module on the appropriate side, the equivalence in Proposition 2.12 between the categories of comodules for $\mathcal{B}$ and modules for its dual is strict (anti)monoidal.

The reader should be warned that in the paper [39], a different notion of a comodule is used. For an $\mathbf{S}|\mathbf{R}$-coring C, the coproduct and the counit of the $\mathbf{R}$-coring C project to a coproduct and a counit on the quotient $\mathbf{R}$-bimodule $C/\{(s\otimes_k 1_{\mathbf{R}})\cdot c - c\cdot(s\otimes_k 1_{\mathbf{R}}) \mid c \in C,\ s \in S\}$. Applying the definition of a comodule of a bimonad in [39, Section 4.1] to a bimonad induced by a right $\mathbf{R}$-bialgebroid $\mathcal{B}$, the resulting notion is a comodule

for the corresponding quotient coring $B/\{s(r)b - t(r)b \mid b \in B,\ r \in R\}$ (where s and t are the source and target maps of $\mathcal{B}$, respectively). The category of such comodules is not known to be monoidal.

3.7. *Algebra extensions by bialgebroids. Galois extensions*

In analogy with bialgebra extensions, there are two symmetrical notions of an algebra extension by a bialgebroid $\mathcal{B}$. In the *action picture*, one deals with a $\mathcal{B}$-module algebra M and its invariant subalgebra (with respect to a character defined by the counit). In this picture, *Galois property* means Galois property of an associated $\mathbf{M}$-ring. Dually, in the *coaction picture*, one deals with a $\mathcal{B}$-comodule algebra M and its coinvariant subalgebra (with respect to a grouplike element defined by the unit). In this picture, *Galois property* means Galois property of an associated $\mathbf{M}$-coring. Although the two approaches are symmetric (and equivalent for finitely generated and projective bialgebroids), the coaction picture is more popular and more developed. We present it in more detail but, for the sake of completeness, we shortly describe the action picture as well.

3.7.1. The action and coaction pictures By Theorem 3.13, the category $\mathcal{M}_{\mathbf{B}}$ of right modules of a right $\mathbf{R}$-bialgebroid $\mathcal{B}$ is monoidal. By definition, a right $\mathcal{B}$-module algebra is a monoid M in $\mathcal{M}_{\mathbf{B}}$. Denote the structure maps of $\mathcal{B}$ as in Definition 3.1. In view of Lemma 2.2, a right $\mathcal{B}$-module algebra is the same as an algebra and right $\mathbf{B}$-module $\mathbf{M}$ such that the multiplication in $\mathbf{M}$ is $\mathbf{R}$-balanced and

$$(mm') \cdot b = (m \cdot b^{(1)})(m' \cdot b^{(2)}) \ \text{ and } \ 1_{\mathbf{M}} \cdot b = 1_{\mathbf{M}} \cdot s\big(\epsilon(b)\big), \quad \text{for } m, m' \in M,\ b \in B,$$

cf. (3.9). Note in passing that, by strict monoidality of the forgetful functor $\mathcal{M}_{\mathbf{B}} \to {}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$, a right $\mathcal{B}$-module algebra M has a canonical $\mathbf{R}$-ring structure. Its unit is the map $\mathbf{R} \to \mathbf{M}$, $r \mapsto 1_{\mathbf{M}} \cdot s(r) = 1_{\mathbf{M}} \cdot t(r)$.

For a right $\mathcal{B}$-module algebra M, $B \otimes_{\mathbf{R}} M$ has an $\mathbf{M}$-ring structure. It is called the *smash product*. The multiplication is $(b \otimes_{\mathbf{R}} m)(b' \otimes_{\mathbf{R}} m') = bb'^{(1)} \otimes_{\mathbf{R}} (m \cdot b'^{(2)})m'$, and the unit is $m \mapsto 1_{\mathbf{B}} \otimes_{\mathbf{R}} m$. The right character ϵ on the $\mathbf{R}$-ring $(\mathbf{B}, s)$ determines a right character $\epsilon \otimes_{\mathbf{R}} M$ on the $\mathbf{M}$-ring $B \otimes_{\mathbf{R}} M$. Hence, we can consider the invariant subalgebra of the base algebra $\mathbf{M}$, with respect to the right character $\epsilon \otimes_{\mathbf{R}} M$. It coincides with the ϵ-invariants of the $(\mathbf{B}, s)$-module M,

$$N := M_\epsilon = \{n \in M \mid n \cdot b = n \cdot s\,(\epsilon(b)), \quad \forall b \in B\}.$$

In the *action picture*, the algebra $\mathbf{M}$ is said to be a *right $\mathcal{B}$-Galois extension* of the invariant subalgebra $\mathbf{N}$ provided that $B \otimes_{\mathbf{R}} M$ is a Galois $\mathbf{M}$-ring with respect to the right character $\epsilon \otimes_{\mathbf{R}} M$. That is, the *canonical map*

$$B \underset{\mathbf{R}}{\otimes} M \to {}_{\mathbf{N}}\mathrm{End}(M), \qquad b \underset{\mathbf{R}}{\otimes} m \mapsto \big(m' \mapsto (m' \cdot b)m\big)$$

is bijective. Left Galois extensions by a left bialgabroid $\mathcal{B}$ are defined symmetrically, referring to a left $\mathcal{B}$-module algebra and its ϵ-invariant subalgebra.

By Theorem 3.18, also the category $\mathcal{M}^B$ of right comodules of a right $\mathbf{R}$-bialgebroid $\mathcal{B}$ is monoidal. By definition, a *right $\mathcal{B}$-comodule algebra* is a monoid in $\mathcal{M}^B$. In view of Lemma 2.2, a right $\mathcal{B}$-comodule algebra is the same as an algebra and right $\mathcal{B}$-comodule $\mathbf{M}$, with coaction $m \mapsto m^{[0]} \otimes_{\mathbf{R}} m^{[1]}$ such that the multiplication in $\mathbf{M}$ is $\mathbf{R}$-balanced and, for $m, m' \in M$,

$$(mm')^{[0]} \underset{\mathbf{R}}{\otimes} (mm')^{[1]} = m^{[0]}m'^{[0]} \underset{\mathbf{R}}{\otimes} m^{[1]}m'^{[1]} \quad \text{and} \quad 1_{\mathbf{M}}{}^{[0]} \underset{\mathbf{R}}{\otimes} 1_{\mathbf{M}}{}^{[1]} = 1_{\mathbf{M}} \underset{\mathbf{R}}{\otimes} 1_{\mathbf{B}}. \tag{3.12}$$

Note in passing that, by strict monoidality of the forgetful functor $\mathcal{M}^B \to {}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$, a right $\mathcal{B}$-comodule algebra M has a canonical $\mathbf{R}$-ring structure. Its unit is the map $\mathbf{R} \to \mathbf{M}$, $r \mapsto 1_{\mathbf{M}} \cdot r = r \cdot 1_{\mathbf{M}}$.

For a right $\mathcal{B}$-comodule algebra M, $M \otimes_{\mathbf{R}} B$ has an $\mathbf{M}$-coring structure with left and right $\mathbf{M}$-actions

$$m_1 \cdot (m \underset{\mathbf{R}}{\otimes} b) \cdot m_2 = m_1 m m_2{}^{[0]} \underset{\mathbf{R}}{\otimes} b m_2{}^{[1]}, \quad \text{for } m_1, m_2 \in M,\ m \underset{\mathbf{R}}{\otimes} b \in M \underset{\mathbf{R}}{\otimes} B, \tag{3.13}$$

comultiplication $m \otimes_{\mathbf{R}} b \mapsto (m \otimes_{\mathbf{R}} b^{(1)}) \otimes_{\mathbf{M}} (1_{\mathbf{M}} \otimes_{\mathbf{R}} b^{(2)})$, and counit $m \otimes_{\mathbf{R}} b \mapsto m \cdot \epsilon(b)$. The grouplike element $1_{\mathbf{B}}$ in the $\mathbf{R}$-coring (B, Δ, ϵ) determines a grouplike element $1_{\mathbf{M}} \otimes_{\mathbf{R}} 1_{\mathbf{B}}$ in the $\mathbf{M}$-coring $M \otimes_{\mathbf{R}} B$. Hence, we can consider the coinvariant subalgebra of the base algebra $\mathbf{M}$, with respect to the grouplike element $1_{\mathbf{M}} \otimes_{\mathbf{R}} 1_{\mathbf{B}}$. It coincides with the $1_{\mathbf{B}}$-coinvariants of the $\mathcal{B}$-comodule M,

$$N := M^{1_{\mathbf{B}}} = \{n \in M \mid n^{[0]} \underset{\mathbf{R}}{\otimes} n^{[1]} = n \underset{\mathbf{R}}{\otimes} 1_{\mathbf{B}}\}.$$

Note that by right $\mathbf{R}$-linearity of the $\mathcal{B}$-coaction on M and (3.12), for $n \in N$ and $r \in R$,

$$(n \cdot r)^{[0]} \underset{\mathbf{R}}{\otimes} (n \cdot r)^{[1]} = n \underset{\mathbf{R}}{\otimes} s(r) = (r \cdot n)^{[0]} \underset{\mathbf{R}}{\otimes} (r \cdot n)^{[1]}.$$

Hence, for $n \in N$ and $r \in R$,

$$n(1_{\mathbf{M}} \cdot r) = n \cdot r = r \cdot n = (1_{\mathbf{M}} \cdot r)n. \tag{3.14}$$

In the *coaction picture*, the algebra $\mathbf{M}$ is said to be a *right $\mathcal{B}$-Galois extension* of the coinvariant subalgebra $\mathbf{N}$ provided that $M \otimes_{\mathbf{R}} B$ is a Galois $\mathbf{M}$-coring with respect to the grouplike element $1_{\mathbf{M}} \otimes_{\mathbf{R}} 1_{\mathbf{B}}$. That is, the *canonical map*

$$\text{can} : M \underset{\mathbf{N}}{\otimes} M \to M \underset{\mathbf{R}}{\otimes} B, \qquad m \underset{\mathbf{N}}{\otimes} m' \mapsto m m'^{[0]} \underset{\mathbf{R}}{\otimes} m'^{[1]} \tag{3.15}$$

is bijective. Since for a right $\mathbf{R}$-bialgebroid $\mathcal{B}$ also the category of left comodules is monoidal, there is a symmetrical notion of a left $\mathcal{B}$-Galois extension $\mathbf{N} \subseteq \mathbf{M}$. It is a left $\mathcal{B}$-comodule algebra M, with coinvariant subalgebra $\mathbf{N}$, such that an associated $\mathbf{M}$-coring $B \otimes_{\mathbf{R}} M$ is a Galois coring, with respect to the grouplike element $1_{\mathbf{B}} \otimes_{\mathbf{R}} 1_{\mathbf{M}}$. Left and right Galois extensions by left bialgebroids are treated symmetrically.

For a right comodule algebra M of a right $\mathbf{R}$-bialgebroid $\mathcal{B}$, a *right–right relative Hopf module* is a right M-module *in* $\mathcal{M}^B$. The category of right–right relative Hopf

modules is denoted by $\mathcal{M}_M^B$, and it turns out to be isomorphic to the category of right comodules for the $\mathbf{M}$-coring (3.13). Hence, the grouplike element $1_{\mathbf{M}} \otimes_{\mathbf{R}} 1_{\mathbf{B}} \in M \otimes_{\mathbf{R}} B$ determines an adjunction as in (2.6) between $\mathcal{M}_M^B$ and the category $\mathcal{M}_{\mathbf{N}}$ of right modules for the coinvariant subalgebra $\mathbf{N}$ of M. It will be denoted as

$$-\underset{\mathbf{N}}{\otimes} M : \mathcal{M}_{\mathbf{N}} \to \mathcal{M}_M^B \qquad \text{and} \qquad (-)^{coB} : \mathcal{M}_M^B \to \mathcal{M}_{\mathbf{N}}. \tag{3.16}$$

Recall from Section 2.2 that for a $\mathcal{B}$-Galois extension $\mathbf{N} \subseteq \mathbf{M}$, this adjunction is interesting from the descent theory point of view.

For a finitely generated and projective bialgebroid, the action and coaction pictures are equivalent in the sense of Proposition 3.19. This equivalence was observed (in a slightly more restricted context) in [12, Theorem and Definition 3.3].

PROPOSITION 3.19. *Let $\mathcal{B}$ be a right $\mathbf{R}$-bialgebroid, which is a finitely generated and projective right $\mathbf{R}$-module via right multiplication by the source map.*

(1) *There is a bijective correspondence between right $\mathcal{B}$-module algebra structures and right $(\mathcal{B}^*)^{\mathrm{op}}$-comodule algebra structures on a given algebra $\mathbf{M}$.*
(2) *The invariant subalgebra $\mathbf{N}$ of a right $\mathcal{B}$-module algebra M (with respect to the right character given by the counit) is the same as the coinvariant subalgebra of the corresponding right $(\mathcal{B}^*)^{\mathrm{op}}$-comodule algebra M (with respect to the grouplike element given by the unit).*
(3) *A right $\mathcal{B}$-module algebra $\mathbf{M}$ is a $\mathcal{B}$-Galois extension of its invariant subalgebra $\mathbf{N}$ in the action picture if and only if $\mathbf{M}$ is a $(\mathcal{B}^*)^{\mathrm{op}}$-Galois extension of $\mathbf{N}$ in the coaction picture.*

Part (1) of Proposition 3.19 follows by the strict monoidal equivalence $\mathcal{M}_{\mathbf{B}} \cong \mathcal{M}^{(B^*)^{\mathrm{op}}}$. Parts (2) and (3) follow by Proposition 2.13 since the $\mathbf{M}$-ring $B \otimes_{\mathbf{R}} M$, associated to a right $\mathcal{B}$-module algebra M, is the left $\mathbf{M}$-dual of the $\mathbf{M}$-coring $M \otimes_{\mathbf{R}} (B^*)^{\mathrm{op}}$, associated to the right $(\mathcal{B}^*)^{\mathrm{op}}$-comodule algebra M.

Consider a right $\mathbf{R}$-bialgebroid $\mathcal{B}$, which is a finitely generated and projective right $\mathbf{R}$-module via the source map. Then, for a right $\mathcal{B}$-module algebra M, the category of right–right $(M, (\mathcal{B}^*)^{\mathrm{op}})$ relative Hopf modules is equivalent also to the category of right modules for the smash product algebra $B \otimes_{\mathbf{R}} M$.

In the remainder of these notes, *only the coaction picture* of Galois extensions will be used.

3.7.2. Quantum torsors and bi-Galois extensions Following the work of Grunspan and Schauenburg [40–43], for a bialgebra B over a commutative ring k, a faithfully flat right B-Galois extension $\mathbf{T}$ of k can be described without explicit mention of the bialgebra B. Instead, a *quantum torsor* structure is introduced on $\mathbf{T}$, from which B can be reconstructed uniquely. What is more, a quantum torsor determines a second k-bialgebra B' for which $\mathbf{T}$ is a left B'-Galois extension of k. It is said that any faithfully flat Galois extension of k by a k-bialgebra is in fact a bi-Galois extension. The categories of (left) comodules for the bialgebras B and B' are monoidally equivalent.

Such a description of faithfully flat Galois extensions by bialgebroids was developed in the PhD thesis of Hobst [44] and in the paper [45].

DEFINITION 3.20. For two algebras $\mathbf{R}$ and $\mathbf{S}$ over a commutative ring k, an $\mathbf{R}$-$\mathbf{S}$ *torsor* is a pair (T, τ). Here, T is an $\mathbf{R} \otimes_k \mathbf{S}$-ring with underlying k-algebra $\mathbf{T}$ and unit maps $\alpha : \mathbf{R} \to \mathbf{T}$ and $\beta : \mathbf{S} \to \mathbf{T}$ (with commuting ranges in $\mathbf{T}$). Considering T as an $\mathbf{R}$-$\mathbf{S}$ bimodule and as an $\mathbf{S}$-$\mathbf{R}$ bimodule via the maps α and β, τ is an $\mathbf{S}$-$\mathbf{R}$ bimodule map $T \to T \otimes_{\mathbf{R}} T \otimes_{\mathbf{S}} T$, $t \mapsto t^{\langle 1\rangle} \otimes_{\mathbf{R}} t^{\langle 2\rangle} \otimes_{\mathbf{S}} t^{\langle 3\rangle}$ (where implicit summation is understood), satisfying the following axioms, for $t, t' \in T$, $r \in R$, and $s \in S$.

(i) $(\tau \otimes_{\mathbf{R}} T \otimes_{\mathbf{S}} T) \circ \tau = (T \otimes_{\mathbf{R}} T \otimes_{\mathbf{S}} \tau) \circ \tau$ (*coassociativity*),
(ii) $(\mu_{\mathbf{R}} \otimes_{\mathbf{S}} T) \circ \tau = \beta \otimes_{\mathbf{S}} T$ and $(T \otimes_{\mathbf{R}} \mu_{\mathbf{S}}) \circ \tau = T \otimes_{\mathbf{R}} \alpha$ (*left and right counitality*),
(iii) $\tau(1_{\mathbf{T}}) = 1_{\mathbf{T}} \otimes_{\mathbf{R}} 1_{\mathbf{T}} \otimes_{\mathbf{S}} 1_{\mathbf{T}}$ (*unitality*),
(iv) $\alpha(r)t^{\langle 1\rangle} \otimes_{\mathbf{R}} t^{\langle 2\rangle} \otimes_{\mathbf{S}} t^{\langle 3\rangle} = t^{\langle 1\rangle} \otimes_{\mathbf{R}} t^{\langle 2\rangle}\alpha(r) \otimes_{\mathbf{S}} t^{\langle 3\rangle}$ and
$t^{\langle 1\rangle} \otimes_{\mathbf{R}} \beta(s)t^{\langle 2\rangle} \otimes_{\mathbf{S}} t^{\langle 3\rangle} = t^{\langle 1\rangle} \otimes_{\mathbf{R}} t^{\langle 2\rangle} \otimes_{\mathbf{S}} t^{\langle 3\rangle}\beta(s)$, (*centrality conditions*)
(v) $\tau(tt') = t^{\langle 1\rangle}t'^{\langle 1\rangle} \otimes_{\mathbf{R}} t'^{\langle 2\rangle}t^{\langle 2\rangle} \otimes_{\mathbf{S}} t^{\langle 3\rangle}t'^{\langle 3\rangle}$ (*multiplicativity*),

where $\mu_{\mathbf{R}}$ and $\mu_{\mathbf{S}}$ denote multiplication in the $\mathbf{R}$-ring $(\mathbf{T}, \alpha)$ and the $\mathbf{S}$-ring $(\mathbf{T}, \beta)$, respectively.

An $\mathbf{R}$-$\mathbf{S}$ torsor (T, τ) is said to be faithfully flat if T is a faithfully flat right $\mathbf{R}$-module and a faithfully flat left $\mathbf{S}$-module.

Note that axiom (iv) in Definition 3.20 is needed in order for the multiplication in axiom (v) to be well defined.

THEOREM 3.21. *For two k-algebras $\mathbf{R}$ and $\mathbf{S}$, there is a bijective correspondence between the following sets of data.*

(i) *Faithfully flat $\mathbf{R}$-$\mathbf{S}$ torsors (T, τ).*
(ii) *Right $\mathbf{R}$-bialgebroids $\mathcal{B}$ and left faithfully flat right $\mathcal{B}$-Galois extensions $\mathbf{S} \subseteq \mathbf{T}$ such that T is a right faithfully flat $\mathbf{R}$-ring.*
(iii) *Left $\mathbf{S}$-bialgebroids $\mathcal{B}'$ and right faithfully flat left $\mathcal{B}'$-Galois extensions $\mathbf{R} \subseteq \mathbf{T}$ such that T is a left faithfully flat $\mathbf{S}$-ring.*

Furthermore, a faithfully flat $\mathbf{R}$-$\mathbf{S}$ torsor T is a $\mathcal{B}'$-$\mathcal{B}$ bicomodule, that is, the left $\mathcal{B}'$, and right $\mathcal{B}$-coactions on T do commute.

Starting with the data in part (ii) of Theorem 3.21, a torsor map on T is constructed in terms of the $\mathcal{B}$-coaction $\varrho^T : T \to T \otimes_{\mathbf{R}} B$, and the inverse of the canonical map (3.15) (with the role of the comodule algebra M in (3.15) played by T), as $\tau := (T \otimes_{\mathbf{R}} \mathrm{can}^{-1}(1_{\mathbf{T}} \otimes_{\mathbf{R}} -)) \circ \varrho^T$. Conversely, to a faithfully flat $\mathbf{R}$-$\mathbf{S}$ torsor (T, τ) (with multiplication $\mu_{\mathbf{S}}$ in the $\mathbf{S}$-ring $(\mathbf{T}, \beta)$), one associates a right $\mathbf{R}$-bialgebroid $\mathcal{B}$ defined on the $\mathbf{R}$-$\mathbf{R}$ bimodule given by the equalizer of the maps $(\mu_{\mathbf{S}} \otimes_{\mathbf{R}} T \otimes_{\mathbf{S}} T) \circ (T \otimes_{\mathbf{S}} \tau)$ and $\alpha \otimes_{\mathbf{R}} T \otimes_{\mathbf{S}} T : T \otimes_{\mathbf{S}} T \to T \otimes_{\mathbf{R}} T \otimes_{\mathbf{S}} T$.

THEOREM 3.22. *For two k-algebras* $\mathbf{R}$ *and* $\mathbf{S}$*, consider a faithfully flat* $\mathbf{R}$-$\mathbf{S}$ *torsor* (T, τ)*. Let* $\mathcal{B}$ *and* $\mathcal{B}'$ *be the associated bialgebroids in Theorem 3.21. Assume that* T *is a faithfully flat right* $\mathbf{S}$*-module and* B' *is a flat right* $\mathbf{S}$*-module. Then the categories of left* $\mathcal{B}$*- and* $\mathcal{B}'$*-comodules are monoidally equivalent.*

Note that the assumptions made about the right $\mathbf{S}$-modules T and B' in Theorem 3.22 become redundant when working with one commutative base ring $\mathbf{R} = \mathbf{S}$ and equal unit maps $\alpha = \beta$. The equivalence in Theorem 3.22 is given by $T\Box_{\mathcal{B}} - : {}^{B}\mathcal{M} \to {}^{B'}\mathcal{M}$, a cotensor product with the $\mathcal{B}'$-$\mathcal{B}$ bicomodule T. (Recall that the notion of a *cotensor product* is dual to the one of module tensor product. That is, for a right $\mathcal{B}$-comodule (M, ϱ^M) and a left $\mathcal{B}$-comodule $(N, {}^N\varrho)$, $M\Box_{\mathcal{B}}N$ is the equalizer of the maps $\varrho^M \otimes_{\mathbf{R}} N$ and $M \otimes_{\mathbf{R}} {}^N\varrho$.)

3.7.3. Galois extensions by finitely generated and projective bialgebroids The bialgebra B, for which a given algebra extension is B-Galois, is nonunique. Obviously, there are (potentially) even more possibilities for the choice of $\mathcal{B}$ if it is allowed to be a bialgebroid. Still, as a main advantage of studying Galois extensions by bialgebroids, in appropriate finite cases all possible bialgebroids $\mathcal{B}$ can be related to a canonical one. The following Theorem 3.24 is a mild generalization of [12, Proposition 4.12].

DEFINITION 3.23. Let $\mathcal{B}$ be a right bialgebroid over an algebra $\mathbf{R}$. A *right–right Yetter–Drinfel'd module* for $\mathcal{B}$ is a right $\mathcal{B}$-module and right $\mathcal{B}$-comodule M (with one and the same underlying $\mathbf{R}$-bimodule structure) such that the following compatibility condition holds,

$$(m \cdot b^{(2)})^{[0]} \underset{\mathbf{R}}{\otimes} b^{(1)} (m \cdot b^{(2)})^{[1]} = m^{[0]} \cdot b^{(1)} \underset{\mathbf{R}}{\otimes} m^{[1]} b^{(2)}, \qquad \text{for } m \in M,\ b \in B.$$

It follows by a symmetrical version of [33, Proposition 4.4] that the category of right–right Yetter–Drinfel'd modules of a right bialgebroid $\mathcal{B}$ is isomorphic to the weak center of the monoidal category of right $\mathbf{B}$-modules. Hence, it is monoidal and prebraided. Following the paper [12], the construction in Section 3.4.7 can be extended to a braided commutative algebra A in the category of right–right Yetter–Drinfel'd modules of a right $\mathbf{R}$-bialgebroid $\mathcal{B}$. That is, the smash product algebra $A\#B$ can be proven to carry the structure of a right $\mathbf{A}$-bialgebroid, which is called a *scalar extension* of $\mathcal{B}$ by A.

In Theorem 3.24, the center of a bimodule M of an algebra $\mathbf{R}$ is denoted by $M^{\mathbf{R}}$.

THEOREM 3.24. *For an algebra* $\mathbf{R}$*, consider a right* $\mathbf{R}$*-bialgebroid* $\mathcal{B}$*, which is a finitely generated and projective left* $\mathbf{R}$*-module via right multiplication by the target map. Let* $\mathbf{N} \subseteq \mathbf{M}$ *be a right* $\mathcal{B}$*-Galois extension. Then* $\mathbf{N} \subseteq \mathbf{M}$ *is a right Galois extension by a right bialgebroid* $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$ *over the base algebra* $M^{\mathbf{N}}$*. What is more, the bialgebroid* $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$ *is isomorphic to a scalar extension of* $\mathcal{B}$*.*

In proving Theorem 3.24, the following key ideas are used. First, a braided commutative algebra structure, in the category of right–right Yetter–Drinfel'd modules

for $\mathcal{B}$, is constructed on $M^{\mathbf{N}}$. The $\mathcal{B}$-coaction on $M^{\mathbf{N}}$ is given by restriction of the $\mathcal{B}$-coaction on M. The $\mathbf{B}$-action on $M^{\mathbf{N}}$ is of the Miyashita–Ulbrich type, that is, it is given in terms of the inverse of the canonical map (3.15). Introducing the index notation $\mathrm{can}^{-1}(1_{\mathbf{M}} \otimes_{\mathbf{R}} b) = b^{\{1\}} \otimes_{\mathbf{N}} b^{\{2\}}$, for $b \in B$ (implicit summation is understood), the right $\mathbf{B}$-action on $M^{\mathbf{N}}$ is $a \cdot b := b^{\{1\}} a b^{\{2\}}$, for $b \in B$ and $a \in M^{\mathbf{N}}$. Since in this way $M^{\mathbf{N}}$ is a braided commutative algebra in the category of right–right Yetter–Drinfel'd modules for $\mathcal{B}$, there exists a right $M^{\mathbf{N}}$-bialgebroid $M^{\mathbf{N}} \otimes_{\mathbf{R}} B$ (cf. Section 3.4.7). Restriction of the $\mathcal{B}$-canonical map (3.15) establishes a bijection $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}} \to M^{\mathbf{N}} \otimes_{\mathbf{R}} B$. Hence, it induces an $M^{\mathbf{N}}$-bialgebroid structure on $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$ and also an $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$-comodule algebra structure on M. After checking that the coinvariants of the $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$-comodule M are precisely the elements of $\mathbf{N}$, the $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$-Galois property of the extension $\mathbf{N} \subseteq \mathbf{M}$ becomes obvious: the $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$-canonical map differs from the $\mathcal{B}$-canonical map (3.15) by an isomorphism.

3.7.4. Depth 2 algebra extensions Classical finitary Galois extensions of fields can be characterized inherently by normality and separability properties, without referring to the Galois group G. That is, a (unique up to isomorphism) finite Galois group $G := \mathrm{Aut}_K(F)$ is determined by any normal and separable field extension F of K. While no such inherent characterization of Galois extensions by (finitely generated and projective) bialgebras is known, a most important achievement in the Galois theory of bialgebroids is a characterization of Galois extensions by finitely generated and projective bialgebroids. A first result in this direction was given in [12, Theorem 3.7]. At the level of generality presented here, it was proven in [13, Theorem 2.1].

The following definition in [23, Definition 3.1] was abstracted from depth 2 extensions of C^*-algebras.

DEFINITION 3.25. Consider an extension $\mathbf{N} \subseteq \mathbf{M}$ of algebras. It is said to satisfy the *right (resp. left) depth 2* condition if the $\mathbf{M}$-$\mathbf{N}$ bimodule (resp. $\mathbf{N}$-$\mathbf{M}$ bimodule) $M \otimes_{\mathbf{N}} M$ is a direct summand in a finite direct sum of copies of M.

Note that the right depth 2 property of an algebra extension $\mathbf{N} \subseteq \mathbf{M}$ is equivalent to the existence of finitely many elements $\gamma_k \in {}_{\mathbf{N}}\mathrm{End}_{\mathbf{N}}(M) \cong {}_{\mathbf{M}}\mathrm{Hom}_{\mathbf{N}}(M \otimes_{\mathbf{N}} M, M)$ and $c_k \in (M \otimes_{\mathbf{N}} M)^{\mathbf{N}} \cong {}_{\mathbf{M}}\mathrm{Hom}_{\mathbf{N}}(M, M \otimes_{\mathbf{N}} M)$, the so called *right depth 2 quasi-basis* satisfying the identity

$$\sum_k m\gamma_k(m')c_k = m \underset{\mathbf{N}}{\otimes} m' \quad \text{for } m, m' \in M. \tag{3.17}$$

DEFINITION 3.26. An extension $\mathbf{N} \subseteq \mathbf{M}$ of algebras is *balanced* if all endomorphisms of M, as a left module for the algebra $\mathcal{E} := \mathrm{End}_{\mathbf{N}}(M)$, are given by right multiplication by some element of $\mathbf{N}$.

THEOREM 3.27. *For an algebra extension $\mathbf{N} \subseteq \mathbf{M}$, the following properties are equivalent.*

(i) $\mathbf{N} \subseteq \mathbf{M}$ *is a right Galois extension by some right* $\mathbf{R}$*-bialgebroid* $\mathcal{B}$*, which is a finitely generated and projective left* $\mathbf{R}$*-module via right multiplication by the target map.*
(ii) *The algebra extension* $\mathbf{N} \subseteq \mathbf{M}$ *is balanced and satisfies the right depth 2 condition.*

If $\mathbf{N} \subseteq \mathbf{M}$ is a right Galois extension by a right $\mathbf{R}$-bialgebroid $\mathcal{B}$, then $M \otimes_{\mathbf{N}} M \cong M \otimes_{\mathbf{R}} B$ as $\mathbf{M}$-$\mathbf{N}$ bimodules. Hence, the right depth 2 condition follows by finitely generated projectivity of the left $\mathbf{R}$-module B. The map, given by left multiplication by an element $m \in M$, and the map, given by right multiplication by $r \in R$, are both right $\mathbf{N}$-linear. Therefore, a left $\mathcal{E} = \mathrm{End}_{\mathbf{N}}(M)$-module endomorphism of M is given by right multiplication by an element $x \in M^{\mathbf{R}}$. Since by (3.14) also the right action (2.7) on M by $\phi \in {}^*B$ is a right $\mathbf{N}$-module map, it follows that $x^{[0]}\phi(x^{[1]}) = x\phi(1_{\mathbf{B}})$ for all $\phi \in {}^*B$. Together with the finitely generated projectivity of the left $\mathbf{R}$-module B, this implies that x belongs to the coinvariant subalgebra $\mathbf{N}$, and hence the extension $\mathbf{N} \subseteq \mathbf{M}$ is balanced.

By Theorem 3.24, if $\mathbf{N} \subseteq \mathbf{M}$ is a right Galois extension by some finitely generated projective right $\mathbf{R}$-bialgebroid $\mathcal{B}$, then it is a Galois extension by the (finitely generated projective) right $M^{\mathbf{N}}$-bialgebroid $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$. Hence, in the converse direction, a balanced algebra extension $\mathbf{N} \subseteq \mathbf{M}$, satisfying the right depth 2 condition, is shown to be a right Galois extension by a right $M^{\mathbf{N}}$-bialgebroid $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$, as constructed in [23, Section 5]. (In fact, in [23], both the left and right depth 2 properties are assumed. It is proven in [13] that the construction works for one-sided depth 2 extensions as well.) The coproduct in $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$ and its coaction on M are constructed in terms of the right depth 2 quasi-basis (3.17). Let us mention that the only point in the proof where the balanced property is used is to show that the $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$-coinvariants in M are precisely the elements of N.

4. Hopf algebroids

A Hopf algebra is a bialgebra H equipped with an additional *antipode* map $H \to H$. The antipode is known to be a bialgebra map from H to the opposite co-opposite of H. It does not seem to be possible to define a Hopf algebroid based on this analogy. Starting with a, say left, bialgebroid $\mathcal{H}$, its opposite co-opposite $\mathcal{H}^{\mathrm{op}}_{\mathrm{cop}}$ is a right bialgebroid. There is no sensible notion of a bialgebroid map $\mathcal{H} \to \mathcal{H}^{\mathrm{op}}_{\mathrm{cop}}$. If we choose as a guiding principle the antipode of a Hopf algebroid $\mathcal{H}$ to be a bialgebroid map $\mathcal{H} \to \mathcal{H}^{\mathrm{op}}_{\mathrm{cop}}$, then $\mathcal{H}$ and $\mathcal{H}^{\mathrm{op}}_{\mathrm{cop}}$ need to carry the same, say left, bialgebroid structure. This means that the underlying algebra $\mathbf{H}$ must be equipped with both left and right bialgebroid structures. The first definition fulfilling this requirement was proposed in [46, Definition 4.1], where, however, the antipode was defined to be bijective. Bijectivity of the antipode was relaxed in [47, Definition 2.2]. Here, we present a set of axioms, which is equivalent to [47, Definition 2.2], as it was formulated in [48, Remark 2.1].

Definition 4.1. For two algebras $\mathbf{R}$ and $\mathbf{L}$ over a commutative ring k, a *Hopf algebroid* over the base algebras $\mathbf{R}$ and $\mathbf{L}$ is a triple $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$. Here $\mathcal{H}_L$ is a left

L-bialgebroid and $\mathcal{H}_R$ is a right **R**-bialgebroid such that their underlying k-algebra **H** is the same. The antipode S is a k-module map $H \to H$. Denote the $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring structure of $\mathcal{H}_R$ by $(\mathbf{H}, s_R, t_R)$ and its **R**-coring structure by $(H, \Delta_R, \epsilon_R)$. Similarly, denote the $\mathbf{L} \otimes_k \mathbf{L}^{\mathrm{op}}$-ring structure of $\mathcal{H}_L$ by $(\mathbf{H}, s_L, t_L)$ and its **L**-coring structure by $(H, \Delta_L, \epsilon_L)$. Denote the multiplication in the **R**-ring $(\mathbf{H}, s_R)$ by μ_R and denote the multiplication in the **L**-ring $(\mathbf{H}, s_L)$ by μ_L. These structures are subject to the following compatibility axioms.

(i) $s_L \circ \epsilon_L \circ t_R = t_R$, $\quad t_L \circ \epsilon_L \circ s_R = s_R$, $\quad s_R \circ \epsilon_R \circ t_L = t_L$, and $t_R \circ \epsilon_R \circ s_L = s_L$.
(ii) $(\Delta_L \otimes_{\mathbf{R}} H) \circ \Delta_R = (H \otimes_{\mathbf{L}} \Delta_R) \circ \Delta_L$ and $(\Delta_R \otimes_{\mathbf{L}} H) \circ \Delta_L = (H \otimes_{\mathbf{R}} \Delta_L) \circ \Delta_R$.
(iii) For $l \in L$, $r \in R$ and $h \in H$, $S(t_L(l) h t_R(r)) = s_R(r) S(h) s_L(l)$.
(iv) $\mu_L \circ (S \otimes_{\mathbf{L}} H) \circ \Delta_L = s_R \circ \epsilon_R$ and $\mu_R \circ (H \otimes_{\mathbf{R}} S) \circ \Delta_R = s_L \circ \epsilon_L$.

REMARK 4.2. The Hopf algebroid axioms in Definition 4.1 require some interpretation.

(1) By the bialgebroid axioms, all the maps $s_L \circ \epsilon_L$, $t_L \circ \epsilon_L$, $s_R \circ \epsilon_R$, and $t_R \circ \epsilon_R$ are idempotent maps $H \to H$. Hence, the message of axiom (i) is that the ranges of s_L and t_R and the ranges of s_R and t_L are coinciding subalgebras of **H**. These axioms imply that the coproduct Δ_L in $\mathcal{H}_L$ is not only an **L**-bimodule map but also an **R**-bimodule map. Symmetrically, Δ_R is an **L**-bimodule map so that axiom (ii) makes sense.
(2) The k-module H underlying a (left or right) bialgebroid is a left and right comodule via the coproduct. Hence, the k-module H underlying a Hopf algebroid is a left and right comodule for the constituent left and right bialgebroids $\mathcal{H}_L$ and $\mathcal{H}_R$, via the two coproducts Δ_L and Δ_R. Axiom (ii) expresses the property that these regular coactions commute, that is, H is an $\mathcal{H}_L$-$\mathcal{H}_R$ bicomodule and also an $\mathcal{H}_R$-$\mathcal{H}_L$ bicomodule.

    Alternatively, considering H and $H \otimes_{\mathbf{L}} H$ as right $\mathcal{H}_R$-comodules via the respective coactions Δ_R and $H \otimes_{\mathbf{L}} \Delta_R$, the first axiom in (ii) expresses that Δ_L is a right $\mathcal{H}_R$-comodule map. Symmetrically, this condition can be interpreted as a left $\mathcal{H}_L$-comodule map property of Δ_R. Similarly, the second axiom in (ii) can be read as right $\mathcal{H}_L$-colinearity of Δ_R or left $\mathcal{H}_R$-colinearity of Δ_L.
(3) Axiom (iii) formulates the **R**-**L** bimodule map property of the antipode, which is needed in order for axiom (iv) to make sense.
(4) In analogy with the Hopf algebra axioms, axiom (iv) says that the antipode is a convolution inverse of the identity map of H in some generalized sense. The notion of convolution products in the case of two different base algebras **L** and **R** is discussed in Section 4.5.2.

Since in a Hopf algebroid $\mathcal{H}$, there are two constituent bialgebroids $\mathcal{H}_L$ and $\mathcal{H}_R$, in these notes we use two versions of Sweedler's index notation in parallel to denote components of the coproducts Δ_L and Δ_R. We will use lower indices in the case of a left bialgebroid $\mathcal{H}_L$, that is, we write $\Delta_L(h) = h_{(1)} \otimes_{\mathbf{L}} h_{(2)}$, and upper indices in the case of a right bialgebroid $\mathcal{H}_R$, that is, we write $\Delta_R(h) = h^{(1)} \otimes_{\mathbf{R}} h^{(2)}$, for $h \in H$, where implicit summation is understood in both cases. Analogously, we use

upper indices to denote components of a coaction by $\mathcal{H}_R$ and lower indices to denote components of a coaction by $\mathcal{H}_L$.

4.1. *Examples and constructions*

Before turning to a study of the structure of Hopf algebroids, let us present some examples.

4.1.1. Hopf algebras A Hopf algebra H over a commutative ring k is an example of a Hopf algebroid over base algebras $\mathbf{R} = k = \mathbf{L}$. Both bialgebroids $\mathcal{H}_L$ and $\mathcal{H}_R$ are equal to the k-bialgebra H, and the Hopf algebra antipode of H satisfies the Hopf algebroid axioms.

Certainly, not every Hopf algebroid over base algebras $\mathbf{R} = k = \mathbf{L}$ is a Hopf algebra (see e.g. Section 4.1.4). Examples of this kind have been constructed in [49], as follows. Let H be a Hopf algebra over k, with coproduct $\Delta_L : h \mapsto h_{(1)} \otimes_k h_{(2)}$, counit ϵ_L, and antipode S. Let χ be a character on H, that is, a k-algebra map $H \to k$. The coproduct $\Delta_R : h \mapsto h_{(1)} \otimes_k \chi\left(S(h_{(2)})\right) h_{(3)}$ and the counit $\epsilon_R := \chi$ define a second bialgebra structure on the k-algebra $\mathbf{H}$. Looking at these two bialgebras as left and right k-bialgebroids respectively, we obtain a Hopf algebroid with a twisted antipode $h \mapsto \chi(h_{(1)})S(h_{(2)})$. This construction was extended in [50, Theorem 4.2], where new examples of Hopf algebroids were constructed by twisting a (bijective) antipode of a given Hopf algebroid.

4.1.2. Weak Hopf algebras A *weak Hopf algebra* over a commutative ring k is a weak bialgebra H equipped with a k-linear antipode map $S : H \to H$, subject to the following axioms [29]. For $h \in H$,

$$h_{(1)}S(h_{(2)}) = \sqcap^L(h), \quad S(h_{(1)})h_{(2)} = \sqcap^R(h), \quad \text{and} \quad S(h_{(1)})h_{(2)}S(h_{(3)}) = S(h),$$

where the maps $\sqcap^L$ and $\sqcap^R$ were introduced in Section 3.2.2.

The right $\mathbf{R}$-bialgebroid and the left $\mathbf{R}^{\mathrm{op}}$-bialgebroid constructed for a weak Hopf algebra H in Section 3.2.2, together with the antipode S, satisfy the Hopf algebroid axioms.

In particular, consider a small groupoid with finitely many objects. By Section 3.2.2, the free k-module spanned by its morphisms is a weak k-bialgebra. It can be equipped with an antipode by putting $S(f) := f^{-1}$ for every morphism f and extending it k-linearly. Motivated by this example, weak Hopf algebras and sometimes also Hopf algebroids are called *quantum groupoids* in the literature.

Weak Hopf algebras have a nice and well-understood representation theory. The category of finite-dimensional modules of a finite-dimensional semisimple weak Hopf algebra H over a field k is a k-linear semisimple category with finitely many inequivalent irreducible objects, with all hom spaces finite dimensional. It is a monoidal category with left and right duals. A category with the listed properties is termed a *fusion category*. Conversely, based on Tannaka–Krein type reconstruction theorems

in [32] and [51], it was proven in [52] that any fusion category is equivalent to the category of finite-dimensional modules of a (nonunique) finite-dimensional semisimple weak Hopf algebra.

4.1.3. $\mathbf{R} \otimes \mathbf{R}^{\mathrm{op}}$ The left and right bialgebroids on an algebra of the form $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$, constructed for any k-algebra $\mathbf{R}$ in Section 3.2.3, form a Hopf algebroid together with the antipode $r \otimes_k r' \mapsto r' \otimes_k r$.

4.1.4. The algebraic quantum torus Consider an algebra $\mathbf{T}_q$ over a commutative ring k generated by two invertible elements U and V subject to the relation $UV = qVU$, where q is an invertible element in k. $\mathbf{T}_q$ possesses a right bialgebroid structure over the commutative subalgebra $\mathbf{R}$ generated by U. Both the source and target maps are given by the inclusion $\mathbf{R} \to \mathbf{T}_q$. The coproduct and counit are defined by $\Delta_R : V^m U^n \mapsto V^m U^n \otimes_{\mathbf{R}} V^m$ and $\epsilon_R : V^m U^n \mapsto U^n$, respectively. Symmetrically, there is a left $\mathbf{R}$-bialgebroid structure given by the coproduct $\Delta_L : U^n V^m \mapsto U^n V^m \otimes_{\mathbf{R}} V^m$ and counit $\epsilon_L : U^n V^m \mapsto U^n$. Together with the antipode $S : U^n V^m \mapsto V^{-m} U^n$, they constitute a Hopf algebroid.

4.1.5. Scalar extension Consider a Hopf algebra H and a braided commutative algebra A in the category of right–right Yetter–Drinfel'd modules of H. As was seen in Section 3.4.7, the smash product algebra $A\#H$ carries a right $\mathbf{A}$-bialgebroid structure. If the antipode S of H is bijective, then the $\mathbf{A}$-bialgebroid structure of $A\#H$ extends to a Hopf algebroid. Indeed, $A\#H$ is a left $\mathbf{A}^{\mathrm{op}}$-bialgebroid via the source map $a \mapsto a_{[0]} \cdot S(a_{[1]})\#1_{\mathbf{B}}$, target map $a \mapsto a_{[0]}\#a_{[1]}$, coproduct $a\#h \mapsto (a\#h_{(1)}) \otimes_{\mathbf{A}^{\mathrm{op}}} (1_{\mathbf{A}}\#h_{(2)})$, and counit $a\#h \mapsto a_{[0]} \cdot S^{-1}\big(hS^{-1}(a_{[1]})\big)$. The (bijective) antipode is given by $a\#h \mapsto a_{[0]} \cdot S(h_{(2)})\#a_{[1]}S(h_{(1)})$.

4.2. *Basic properties of Hopf algebroids*

In this section, some consequences of Definition 4.1 of a Hopf algebroid are recalled from [47, Section 2].

The opposite co-opposite of a Hopf algebra H is a Hopf algebra, with the same antipode S of H. If S is bijective, then also the opposite and the co-opposite of H are Hopf algebras, with antipode S^{-1}. A generalization of these facts to Hopf algebroids is given below.

PROPOSITION 4.3. *Consider a Hopf algebroid* $(\mathcal{H}_L, \mathcal{H}_R, S)$ *over base algebras* $\mathbf{L}$ *and* $\mathbf{R}$. *Then*

(1) *The triple* $((\mathcal{H}_R)^{\mathrm{op}}_{\mathrm{cop}}, (\mathcal{H}_L)^{\mathrm{op}}_{\mathrm{cop}}, S)$ *is a Hopf algebroid over the base algebras* $\mathbf{R}^{\mathrm{op}}$ *and* $\mathbf{L}^{\mathrm{op}}$.
(2) *If the antipode S is bijective, then* $((\mathcal{H}_R)^{\mathrm{op}}, (\mathcal{H}_L)^{\mathrm{op}}, S^{-1})$ *is a Hopf algebroid over the base algebras* $\mathbf{R}$ *and* $\mathbf{L}$ *and* $((\mathcal{H}_L)_{\mathrm{cop}}, (\mathcal{H}_R)_{\mathrm{cop}}, S^{-1})$ *is a Hopf algebroid over the base algebras* $\mathbf{L}^{\mathrm{op}}$ *and* $\mathbf{R}^{\mathrm{op}}$.

Proposition 4.4 states the expected behavior of the antipode of a Hopf algebroid with respect to the underlying ring and coring structures. Consider a Hopf algebroid $\mathcal{H}$ over base algebras $\mathbf{L}$ and $\mathbf{R}$, with structure maps denoted as in Definition 4.1. It follows immediately by axiom (i) in Definition 4.1 that the base algebras $\mathbf{L}$ and $\mathbf{R}$ are anti-isomorphic. Indeed, there are inverse algebra isomorphisms

$$\epsilon_L \circ s_R : \mathbf{R}^{\mathrm{op}} \to \mathbf{L} \quad \text{and} \quad \epsilon_R \circ t_L : \mathbf{L} \to \mathbf{R}^{\mathrm{op}}. \tag{4.1}$$

Symmetrically, there are inverse algebra isomorphisms

$$\epsilon_R \circ s_L : \mathbf{L}^{\mathrm{op}} \to \mathbf{R} \quad \text{and} \quad \epsilon_L \circ t_R : \mathbf{R} \to \mathbf{L}^{\mathrm{op}}. \tag{4.2}$$

PROPOSITION 4.4. *Let $\mathcal{H}$ be a Hopf algebroid over base algebras $\mathbf{L}$ and $\mathbf{R}$ with structure maps denoted as in Definition 4.1. The following assertions hold.*

(1) *The antipode S is a homomorphism of $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-rings*

$$(\mathbf{H}, s_R, t_R) \to (\mathbf{H}^{\mathrm{op}}, s_L \circ (\epsilon_L \circ s_R), t_L \circ (\epsilon_L \circ s_R))$$

and also a homomorphism of $\mathbf{L} \otimes_k \mathbf{L}^{\mathrm{op}}$-rings

$$(\mathbf{H}, s_L, t_L) \to (\mathbf{H}^{\mathrm{op}}, s_R \circ (\epsilon_R \circ s_L), t_R \circ (\epsilon_R \circ s_L)).$$

In particular, S is a k-algebra antihomomorphism $\mathbf{H} \to \mathbf{H}$.

(2) *The antipode S is a homomorphism of $\mathbf{R}$-corings*

$$(H, \Delta_R, \epsilon_R) \to (H, \Delta_L^{\mathrm{cop}}, (\epsilon_R \circ s_L) \circ \epsilon_L),$$

where Δ_L^{cop} is considered as a map $H \to H \otimes_{\mathbf{L}^{\mathrm{op}}} H \cong H \otimes_{\mathbf{R}} H$, via the isomorphism induced by the algebra isomorphism (4.2)*. Symmetrically, S is a homomorphism of $\mathbf{L}$-corings*

$$(H, \Delta_L, \epsilon_L) \to (H, \Delta_R^{\mathrm{cop}}, (\epsilon_L \circ s_R) \circ \epsilon_R),$$

where Δ_R^{cop} is considered as a map $H \to H \otimes_{\mathbf{R}^{\mathrm{op}}} H \cong H \otimes_{\mathbf{L}} H$, via the isomorphism induced by the algebra isomorphism (4.1)*.*

The k-dual H^* of a finitely generated and projective Hopf algebra H over a commutative ring k is a Hopf algebra. The antipode in H^* is the transpose of the antipode of H. No generalization of this fact for Hopf algebroids is known. Although the dual $\mathcal{H}^*$ of a finitely generated and projective Hopf algebroid $\mathcal{H}$ has a bialgebroid structure (cf. Section 3.3), the transpose of the antipode in $\mathcal{H}$ is not an endomorphism of $\mathcal{H}^*$. Only duals of *Frobenius* Hopf algebroids are known to be (Frobenius) Hopf algebroids (see Section 4.4.2).

4.3. *Comodules of Hopf algebroids*

In a Hopf algebroid, the constituent left and right bialgebroids are defined on the same underlying algebra. Therefore, modules for the two bialgebroids coincide. This

is not the case with comodules: the two bialgebroids have different underlying corings (over anti-isomorphic base algebras, cf. (4.1) and (4.2)), which have a priori different categories of (say, right) comodules. We take the opportunity here to call the reader's attention to a regrettable error in the literature. Based on [53, Theorem 2.6], whose proof turned out to be *incorrect*, the categories of (right) comodules of two constituent bialgebroids in a Hopf algebroid were claimed to be strict monoidally isomorphic in [48, Theorem 2.2]. Since it turned out recently that in [53, Theorem 2.6] some assumptions are missing, the derived result [48, Theorem 2.2] need not hold either at the stated level of generality. (There is a similar error in [54, Proposition 3.1].). Regrettably, this error influences some results also in [55], [48], and [56]. In the current section and in 4.5, we present the corrected statements.

4.3.1. Comodules of a Hopf algebroid and of its constituent bialgebroids Since, as was explained in the first paragraph of Section 4.3, comodules of the two constituent bialgebroids in a Hopf algebroid are in general different notions, none of them can be expected to be a well-working definition of a comodule of a Hopf-algebroid. The following definition of a comodule for a Hopf algebroid, as a compatible comodule of both constituent bialgebroids, was suggested in [54, Definition 3.2] and [12, Section 2.2].

Definition 4.5. Let $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$ be a Hopf algebroid over base k-algebras $\mathbf{L}$ and $\mathbf{R}$ with structure maps denoted as in Definition 4.1. A *right $\mathcal{H}$-comodule* is a right $\mathbf{L}$-module as well as a right $\mathbf{R}$-module M, together with a right $\mathcal{H}_R$-coaction $\varrho_R : M \to M \otimes_{\mathbf{R}} H$ and a right $\mathcal{H}_L$-coaction $\varrho_L : M \to M \otimes_{\mathbf{L}} H$ such that ϱ_R is an $\mathcal{H}_L$-comodule map and ϱ_L is an $\mathcal{H}_R$-comodule map. Explicitly, $\varrho_R : m \mapsto m^{[0]} \otimes_{\mathbf{R}} m^{[1]}$ is a right $\mathbf{L}$-module map in the sense that

$$(m \cdot l)^{[0]} \underset{\mathbf{R}}{\otimes} (m \cdot l)^{[1]} = m^{[0]} \underset{\mathbf{R}}{\otimes} t_L(l) m^{[1]}, \quad \text{for } m \in M,\ l \in L,$$

$\varrho_L : m \mapsto m_{[0]} \otimes_{\mathbf{L}} m_{[1]}$ is a right $\mathbf{R}$-module map in the sense that

$$(m \cdot r)_{[0]} \underset{\mathbf{L}}{\otimes} (m \cdot r)_{[1]} = m_{[0]} \underset{\mathbf{L}}{\otimes} m_{[1]} s_R(r), \quad \text{for } m \in M,\ r \in R,$$

and the following compatibility conditions hold.

$$(\varrho_R \underset{\mathbf{L}}{\otimes} H) \circ \varrho_L = (M \underset{\mathbf{R}}{\otimes} \Delta_L) \circ \varrho_R \quad \text{and} \quad (\varrho_L \underset{\mathbf{R}}{\otimes} H) \circ \varrho_R = (M \underset{\mathbf{L}}{\otimes} \Delta_R) \circ \varrho_L.$$

Morphisms of $\mathcal{H}$-comodules are precisely those morphism that are $\mathcal{H}_L$- and $\mathcal{H}_R$-comodule maps.

The category of right comodules of a Hopf algebroid $\mathcal{H}$ is denoted by $\mathcal{M}^{\mathcal{H}}$.

It is not difficult to see that the right $\mathbf{R}$- and $\mathbf{L}$-actions on a right $\mathcal{H}$-comodule M necessarily commute. That is, M carries the structure of a right $\mathbf{L} \otimes_k \mathbf{R}$-module.

Note that by Definition 4.1 (i) and (ii), the right $\mathbf{R} \otimes_k \mathbf{L}$-module H, with $\mathbf{R}$-action via the right source map s_R and $\mathbf{L}$-action via the left target map t_L, is a right comodule of the Hopf algebroid $\mathcal{H}$, via the two coactions given by the two coproducts Δ_R and Δ_L.

Left comodules of a Hopf algebroid $\mathcal{H}$ are defined symmetrically, and their category is denoted by ${}^{\mathcal{H}}\mathcal{M}$.

REMARK 4.6. The antipode S in a Hopf algebroid $\mathcal{H}$ defines a functor $\mathcal{M}^{\mathcal{H}} \to {}^{\mathcal{H}}\mathcal{M}$. Indeed, if M is a right $\mathcal{H}$-comodule, with $\mathcal{H}_R$-coaction $m \mapsto m^{[0]} \otimes_{\mathbf{R}} m^{[1]}$ and $\mathcal{H}_L$-coaction $m \mapsto m_{[0]} \otimes_{\mathbf{L}} m_{[1]}$, then it is a left $\mathcal{H}$-comodule with left $\mathbf{R}$-action $r \cdot m = m \cdot \epsilon_L(t_R(r))$, left $\mathbf{L}$-action $l \cdot m = m \cdot \epsilon_R(t_L(l))$ (where the notations in Definition 4.1 are used), and the respective coactions

$$m \mapsto S(m_{[1]}) \underset{\mathbf{R}}{\otimes} m_{[0]} \quad \text{and} \quad m \mapsto S(m^{[1]}) \underset{\mathbf{L}}{\otimes} m^{[0]}.$$

Right $\mathcal{H}$-comodule maps are also left $\mathcal{H}$-comodule maps for these coactions.
A functor ${}^{\mathcal{H}}\mathcal{M} \to \mathcal{M}^{\mathcal{H}}$ is constructed symmetrically.

Although comodules of a Hopf algebroid $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$ cannot be described as comodules of a coring, the free-forgetful adjunction (cf. [15, 18.13(2)]), corresponding to the L-coring underlying $\mathcal{H}_L$, lifts to an adjunction between the categories $\mathcal{M}^{\mathcal{H}}$ and $\mathcal{M}_{\mathbf{L}}$. Indeed, the forgetful functor $\mathcal{M}^{\mathcal{H}} \to \mathcal{M}_{\mathbf{L}}$ has a right adjoint $- \otimes_{\mathbf{L}} H : \mathcal{M}_{\mathbf{L}} \to \mathcal{M}^{\mathcal{H}}$. The unit and counit of the adjunction are given for a right $\mathcal{H}$-comodule $(M, \varrho_L, \varrho_R)$ and a right $\mathbf{L}$-module $\mathbf{N}$ by the maps

$$\varrho_L : M \to M \underset{\mathbf{L}}{\otimes} H \quad \text{and} \quad N \underset{\mathbf{L}}{\otimes} \epsilon_L : N \underset{\mathbf{L}}{\otimes} H \to N, \tag{4.3}$$

respectively, where ϵ_L is the counit of $\mathcal{H}_L$. There is a similar adjunction between the categories $\mathcal{M}^{\mathcal{H}}$ and $\mathcal{M}_{\mathbf{R}}$.

Our next task is to look for situations when the category of comodules of a Hopf algebroid coincides with the comodule category of one the constituent bialgebroids. Recall from Remark 4.2 (2) that for a Hopf algebroid $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$, the underlying k-module H is an $\mathcal{H}_L$-$\mathcal{H}_R$ bicomodule and an $\mathcal{H}_R$-$\mathcal{H}_L$ bicomodule via the coactions given by the coproducts. In appropriate situations, taking *cotensor products* with a bicomodule induces a functor between the comodule categories of the two corings (see [15, 22.3] and its Erratum). In Theorem 4.7, functors of this type are considered.

Recall from Section 3.6 that any right comodule of a right bialgebroid over a k-algebra $\mathbf{R}$ possesses a unique $\mathbf{R}$-bimodule structure such that any comodule map is $\mathbf{R}$-bilinear. Thus, if $\mathcal{H}$ is a Hopf algebroid over the anti-isomorphic base k-algebras $\mathbf{L}$ and $\mathbf{R}$, then any right comodule of the constituent right (or left) bialgebroid can be regarded as a right $\mathbf{L} \otimes_k \mathbf{R}$-module and the coaction is a right $\mathbf{R} \otimes_k \mathbf{L}$-module map.

THEOREM 4.7. *Let $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$ be a Hopf algebroid over base k-algebras $\mathbf{L}$ and $\mathbf{R}$, with structure maps denoted as in Definition 4.1. Consider the $\mathbf{R}$- and $\mathbf{L}$-actions on H that define its coring structures (cf. Definitions 3.1 and 3.3). If the equalizer*

$$M \xrightarrow{\ \varrho\ } M \underset{\mathbf{R}}{\otimes} H \underset{M \otimes_{\mathbf{R}} \Delta_R}{\overset{\varrho \otimes_{\mathbf{R}} H}{\rightrightarrows}} M \underset{\mathbf{R}}{\otimes} H \underset{\mathbf{R}}{\otimes} H \tag{4.4}$$

in the category $\mathcal{M}_{\mathbf{L}}$ of right $\mathbf{L}$-modules is preserved by both functors $- \otimes_{\mathbf{L}} H \otimes_{\mathbf{L}} H$ and $- \otimes_{\mathbf{L}} H \otimes_{\mathbf{R}} H : \mathcal{M}_{\mathbf{L}} \to \mathcal{M}_k$, for any right $\mathcal{H}_R$-comodule (M, ϱ), then the forgetful functor $\mathcal{M}^{\mathcal{H}} \to \mathcal{M}^{\mathcal{H}_R}$ is an isomorphism.

By standard terminology, the conditions in Theorem 4.7 are phrased as the equalizer (4.4) in $\mathcal{M}_{\mathbf{L}}$ is $H \otimes_{\mathbf{L}} H$-*pure* and $H \otimes_{\mathbf{R}} H$-*pure*. Symmetrical conditions imply that the forgetful functor to the category of right $\mathcal{H}_L$-comodules is an isomorphism.

Theorem 4.7 is proved by constructing the inverse of the forgetful functor, that is, by equipping any right $\mathcal{H}_R$-comodule with an $\mathcal{H}$-comodule structure. Any right $\mathcal{H}_R$-comodule M is isomorphic, as a right $\mathbf{R}$-module, to the cotensor product $M\Box H$ of M with the left $\mathcal{H}_R$ comodule H. Since under the assumptions in Theorem 4.7, $M\Box H$ is a right $\mathcal{H}_L$-comodule via $M\Box\Delta_L : M\Box H \to M\Box(H \otimes_{\mathbf{L}} H) \cong (M\Box H) \otimes_{\mathbf{L}} H$, the isomorphism $M \cong M\Box H$ induces a right $\mathcal{H}_L$-coaction on M. Moreover, by the assumptions in Theorem 4.7, we have isomorphisms $M\Box(H \otimes_{\mathbf{L}} H) \cong M \otimes_{\mathbf{L}} H$, $M\Box(H \otimes_{\mathbf{R}} H) \cong M \otimes_{\mathbf{R}} H$, $M\Box(H \otimes_{\mathbf{R}} H \otimes_{\mathbf{L}} H) \cong M \otimes_{\mathbf{R}} H \otimes_{\mathbf{L}} H$, and $M\Box(H \otimes_{\mathbf{L}} H \otimes_{\mathbf{R}} H) \cong M \otimes_{\mathbf{L}} H \otimes_{\mathbf{R}} H$. The compatibility conditions between the $\mathcal{H}_R$- and $\mathcal{H}_L$-coactions follow by these isomorphisms, the Hopf algebroid axioms of Definition 4.1 (ii), and functoriality of the cotensor product. With similar methods, any $\mathcal{H}_R$-comodule map is checked to be also $\mathcal{H}_L$-colinear.

The purity conditions in Theorem 4.7 are checked to hold in all the examples in Section 4.1. Moreover, if a Hopf algebroid is a flat left $\mathbf{R}$-module (via right multiplication by the target map of the right bialgebroid) and a flat left $\mathbf{L}$-module (via left multiplication by the source map of the left bialgebroid) then it satisfies all purity conditions in Theorem 4.7 since taking tensor products with flat modules preserves any equalizer.

4.3.2. Coinvariants in a comodule of a Hopf algebroid By Definition 4.5, a comodule M of a Hopf algebroid $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$ is a comodule for both constituent bialgebroids $\mathcal{H}_L$ and $\mathcal{H}_R$. Since the unit element 1_H is grouplike for both corings underlying $\mathcal{H}_L$ and $\mathcal{H}_R$, one can speak about coinvariants

$$M^{\mathrm{co}\mathcal{H}_R} = \{m \in M \mid \varrho_R(M) = m \underset{\mathbf{R}}{\otimes} 1_{\mathbf{H}}\}$$

of M with respect to the $\mathcal{H}_R$-coaction ϱ_R or coinvariants

$$M^{\mathrm{co}\mathcal{H}_L} = \{m \in M \mid \varrho_L(M) = m \underset{\mathbf{L}}{\otimes} 1_{\mathbf{H}}\}$$

with respect to the $\mathcal{H}_L$-coaction ϱ_L. Proposition 4.8 relates these two notions. It is of crucial importance from the point of view of Galois theory (see Section 4.5).

PROPOSITION 4.8. *Let $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$ be a Hopf algebroid and $(M, \varrho_L, \varrho_R)$ be a right $\mathcal{H}$-comodule. Then, any coinvariant of the $\mathcal{H}_R$-comodule (M, ϱ_R) is coinvariant also for the $\mathcal{H}_L$-comodule (M, ϱ_L).*

If moreover the antipode S is bijective, then the coinvariants of the $\mathcal{H}_R$-comodule (M, ϱ_R) and the $\mathcal{H}_L$-comodule (M, ϱ_L) coincide.

For a right $\mathcal{H}$-comodule $(M, \varrho_L, \varrho_R)$, consider the map

$$\Phi_M : M \underset{\mathbf{R}}{\otimes} H \to M \underset{\mathbf{L}}{\otimes} H, \qquad m \otimes_{\mathbf{R}} h \mapsto \varrho_L(m) \cdot S(h), \tag{4.5}$$

where (using the notations in Definition 4.1) H is a left $\mathbf{L}$-module via the source map s_L and a left $\mathbf{R}$-module via the target map t_R, and $M \otimes_{\mathbf{L}} H$ is understood to be a right H-module via the second factor. Since $\Phi_M(\varrho_R(m)) = m \otimes_{\mathbf{L}} 1_H$ and $\Phi_M(m \otimes_{\mathbf{R}} 1_H) = \varrho_L(m)$, we have the first claim in Proposition 4.8 proved. In order to prove the second assertion, note that if S is an isomorphism, then so is Φ_M, with inverse $\Phi_M^{-1}(m \otimes_{\mathbf{L}} h) = S^{-1}(h) \cdot \varrho_R(m)$, where $M \otimes_{\mathbf{R}} H$ is understood to be a left H-module via the second factor.

4.3.3. Comodule algebras of a Hopf algebroid As it was explained in Section 3.7, from the point of view of Galois theory (in the coaction picture), monoidality of the category of comodules is of central importance. Theorem 4.9 replaces the *unjustified* theorem [48, Theorem 2.2] (cf. the first paragraph of Section 4.3).

By Definition 4.5, a right comodule of a Hopf algebroid $\mathcal{H}$ over base k-algebras $\mathbf{L}$ and $\mathbf{R}$ is a right $\mathbf{L} \otimes_k \mathbf{R}$-module. Since $\mathbf{L}$ and $\mathbf{R}$ are anti-isomorphic algebras, we may regard alternatively any $\mathcal{H}$-comodule as an $\mathbf{R}$-bimodule by translating the right $\mathbf{L}$-action to a left $\mathbf{R}$-action via the algebra anti-isomorphism (4.1).

THEOREM 4.9. *For a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$ over base k-algebras $\mathbf{L}$ and $\mathbf{R}$, the category $\mathcal{M}^{\mathcal{H}}$ of right $\mathcal{H}$-comodules is monoidal. Moreover, there are strict monoidal forgetful functors, rendering the following diagram commutative.*

$$\begin{array}{ccc} \mathcal{M}^{\mathcal{H}} & \longrightarrow & \mathcal{M}^{\mathcal{H}_R} \\ \downarrow & & \downarrow \\ \mathcal{M}^{\mathcal{H}_L} & \longrightarrow & {}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}. \end{array}$$

Commutativity of the diagram in Theorem 4.9 follows by comparing the unique $\mathbf{R}$-actions that make $\mathbf{R}$-bilinear the $\mathcal{H}_R$-coaction and the $\mathcal{H}_L$-coaction in an $\mathcal{H}$-comodule. Strict monoidality of the functor on the right-hand side was proved in Theorem 3.18. Strict monoidality of the functor in the bottom row follows by applying Theorem 3.18 to the opposite of the bialgebroid $\mathcal{H}_L$ and identifying $\mathbf{L}^{\mathrm{op}}$-bimodules and $\mathbf{R}$-bimodules via the algebra isomorphism (4.1). To see strict monoidality of the remaining two functors, recall that by Theorem 3.18 – applied to $\mathcal{H}_R$ and the opposite of $\mathcal{H}_L$ – the $\mathbf{R}$-module tensor product of any two $\mathcal{H}$-comodules is an $\mathcal{H}_R$-comodule and an $\mathcal{H}_L$-comodule, via the diagonal coactions, cf. (3.11). It is straightforward to check compatibility of these coactions in the sense of Definition 4.5. Similarly, $\mathbf{R}(\cong \mathbf{L}^{\mathrm{op}})$ is known to be an $\mathcal{H}_R$-comodule and an $\mathcal{H}_L$-comodule and compatibility of the coactions is obvious. Finally, the $\mathbf{R}$-module tensor product of $\mathcal{H}$-comodule maps is an $\mathcal{H}_R$-comodule map and an $\mathcal{H}_L$-comodule map by Theorem 3.18. Thus, it is an $\mathcal{H}$-comodule map. By Theorem 3.18 also the natural coherence transformations in ${}_{\mathbf{R}}\mathcal{M}_{\mathbf{R}}$ are $\mathcal{H}_R$- and $\mathcal{H}_L$-comodule maps, so $\mathcal{H}$-comodule maps, which proves Theorem 4.9.

Theorem 4.9 enables us to introduce comodule algebras of Hopf algebroids.

DEFINITION 4.10. A *right comodule algebra* of a Hopf algebroid $\mathcal{H}$ is a monoid in the monoidal category $\mathcal{M}^{\mathcal{H}}$ of right $\mathcal{H}$-comodules, explicitly, an $\mathbf{R}$-ring (M, μ, η),

such that M is a right $\mathcal{H}$-comodule and $\eta : R \to M$ and $\mu : M \otimes_{\mathbf{R}} M \to M$ are right $\mathcal{H}$-comodule maps. Using the notations $m \mapsto m^{[0]} \otimes_{\mathbf{R}} m^{[1]}$ and $m \mapsto m_{[0]} \otimes_{\mathbf{L}} m_{[1]}$ for the $\mathcal{H}_L$- and $\mathcal{H}_R$-coactions, respectively, $\mathcal{H}$-colinearity of η and μ means the identities, for all $m, m' \in M$,

$$1_{\mathbf{M}}{}^{[0]} \underset{\mathbf{R}}{\otimes} 1_{\mathbf{M}}{}^{[1]} = 1_{\mathbf{M}} \underset{\mathbf{R}}{\otimes} 1_{\mathbf{H}}, \qquad (mm')^{[0]} \underset{\mathbf{R}}{\otimes} (mm')^{[1]} = m^{[0]} m'^{[0]} \underset{\mathbf{R}}{\otimes} m^{[1]} m'^{[1]}$$

$$1_{\mathbf{M}[0]} \underset{\mathbf{L}}{\otimes} 1_{\mathbf{M}[1]} = 1_{\mathbf{M}} \underset{\mathbf{L}}{\otimes} 1_{\mathbf{H}}, \qquad (mm')_{[0]} \underset{\mathbf{L}}{\otimes} (mm')_{[1]} = m_{[0]} m'_{[0]} \underset{\mathbf{L}}{\otimes} m_{[1]} m'_{[1]}.$$

Symmetrically, a *left $\mathcal{H}$-comodule algebra* is a monoid in ${}^{\mathcal{H}}\mathcal{M}$.

The functors in Remark 4.6 induced by the antipode are checked to be strictly antimonoidal. Therefore, the opposite of a right $\mathcal{H}$-comodule algebra, with coactions in Remark 4.6, is a left $\mathcal{H}$-comodule algebra and vice versa. Thus, there are four different categories of modules of a comodule algebra of a Hopf algebroid.

DEFINITION 4.11. Let $\mathcal{H}$ be a Hopf algebroid and M be a right $\mathcal{H}$-comodule algebra. The left and right M-modules in $\mathcal{M}^{\mathcal{H}}$ are called left–right and right–right *relative Hopf modules*, respectively. Their categories are denoted by ${}_M\mathcal{M}^{\mathcal{H}}$ and $\mathcal{M}^{\mathcal{H}}_M$, respectively. Left and right M^{op}-modules in ${}^{\mathcal{H}}\mathcal{M}$ are called right–left and left–left relative Hopf modules, respectively, and their categories are denoted by ${}^{\mathcal{H}}\mathcal{M}_M$ and ${}^{\mathcal{H}}_M\mathcal{M}$, respectively.

Explicitly, for example, a right–right $(M, \mathcal{H})$-relative Hopf module is a right module W for the $\mathbf{R}$-ring M such that the action is a right $\mathcal{H}$-comodule map $W \otimes_{\mathbf{R}} M \to W$. Using index notations, with superscripts for the $\mathcal{H}_R$-coactions and subscripts for the $\mathcal{H}_L$-coactions, both on W and M, this amounts to the following identities, for $w \in W$ and $m \in M$,

$$(w \cdot m)^{[0]} \underset{\mathbf{R}}{\otimes} (w \cdot m)^{[1]} = w^{[0]} \cdot m^{[0]} \underset{\mathbf{R}}{\otimes} w^{[1]} m^{[1]} \quad \text{and}$$

$$(w \cdot m)_{[0]} \underset{\mathbf{L}}{\otimes} (w \cdot m)_{[1]} = w_{[0]} \cdot m_{[0]} \underset{\mathbf{L}}{\otimes} w_{[1]} m_{[1]}.$$

In contrast to relative Hopf modules of bialgebroids in Section 3.7.1, relative Hopf modules of Hopf algebroids cannot be identified with comodules of a coring. Still they determine an adjunction, very similar to (3.16). Consider a right comodule algebra M of a Hopf algebroid $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$ over base algebras $\mathbf{L}$ and $\mathbf{R}$. Denote the $\mathcal{H}_R$-coinvariant subalgebra of M by $\mathbf{N}$. It follows from Proposition 4.8 that for any right $\mathbf{N}$-module V, $V \otimes_{\mathbf{N}} M$ is a right–right relative Hopf module via the second factor. The resulting functor $- \otimes_{\mathbf{N}} M : \mathcal{M}_{\mathbf{N}} \to \mathcal{M}^{\mathcal{H}}_M$ turns out to have a right adjoint. Any object W in $\mathcal{M}^{\mathcal{H}}_M$ can be regarded as an object in $\mathcal{M}^{\mathcal{H}_R}_M$, so we can take its $\mathcal{H}_R$-coinvariants (cf. Section 4.3.2). These considerations lead to an adjoint pair of functors

$$- \underset{\mathbf{N}}{\otimes} M : \mathcal{M}_{\mathbf{N}} \to \mathcal{M}^{\mathcal{H}}_M \quad \text{and} \quad (-)^{\mathrm{co}\mathcal{H}_R} : \mathcal{M}^{\mathcal{H}}_M \to \mathcal{M}_{\mathbf{N}}. \tag{4.6}$$

The unit of the adjunction is given, for any right $\mathbf{N}$-module V, by the map

$$V \to (V \underset{\mathbf{N}}{\otimes} M)^{\mathrm{co}\mathcal{H}_R}, \qquad v \mapsto v \underset{\mathbf{N}}{\otimes} 1_{\mathbf{M}} \tag{4.7}$$

and the counit is given, for an $(M, \mathcal{H})$-relative Hopf module W, by

$$W^{\mathrm{co}\mathcal{H}_R} \underset{\mathbf{N}}{\otimes} M \to M, \qquad w \underset{\mathbf{N}}{\otimes} m \mapsto w \cdot m. \tag{4.8}$$

The message of this observation is that by studying *descent theory* of Galois extensions of a Hopf algebroid $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$, one can examine the adjunction in both (4.6), corresponding to an $\mathcal{H}$-comodule algebra M, and (3.16) determined by M regarded as an $\mathcal{H}_R$-comodule algebra. Proposition 4.12 is obtained by observing that the units of the two adjunctions coincide, and the counit of the adjunction in (4.8) is obtained by restricting to the objects in $\mathcal{M}_M^{\mathcal{H}}$ the counit of the adjunction (3.16).

PROPOSITION 4.12. *Consider a Hopf algebroid $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$ and a right $\mathcal{H}$-comodule algebra M. Denote the $\mathcal{H}_R$-coinvariant subalgebra of M by* $\mathbf{N}$.

(1) *The functor $- \otimes_{\mathbf{N}} M : \mathcal{M}_{\mathbf{N}} \to \mathcal{M}_M^{\mathcal{H}_R}$ is fully faithful if and only if the functor $- \otimes_{\mathbf{N}} M : \mathcal{M}_{\mathbf{N}} \to \mathcal{M}_M^{\mathcal{H}}$ is fully faithful.*

(2) *If the functor $- \otimes_{\mathbf{N}} M : \mathcal{M}_{\mathbf{N}} \to \mathcal{M}_M^{\mathcal{H}_R}$ is an equivalence, then also the functor $- \otimes_{\mathbf{N}} M : \mathcal{M}_{\mathbf{N}} \to \mathcal{M}_M^{\mathcal{H}}$ is an equivalence.*

4.3.4. The fundamental theorem of Hopf modules In this section, we investigate the adjunction (4.6) in a special case.

The coproducts Δ_L and Δ_R in a Hopf algebroid $\mathcal{H}$ make the underlying algebra $\mathbf{H}$ a right $\mathcal{H}$-comodule algebra. The corresponding right–right relative Hopf modules are called simply Hopf modules, and their category is denoted by $\mathcal{M}_H^H$. The coinvariants of the right $\mathcal{H}_R$-comodule algebra H are the elements $t_R(r)$, for $r \in R$, where t_R is the target map. If $\mathcal{H}_R$ is the underlying bialgebroid in a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$, then $t_R : \mathbf{R}^{\mathrm{op}} \to \mathbf{H}$, equivalently, $s_L : \mathbf{L} \to \mathbf{H}$, is a right $\mathcal{H}_R$-Galois extension (cf. Section 4.6.2). Hence, Theorem 4.13, known as the *Fundamental Theorem of Hopf modules*, can be interpreted as a descent Theorem for this Galois extension $\mathbf{L} \cong \mathbf{R}^{\mathrm{op}} \subseteq \mathbf{H}$.

THEOREM 4.13. *For a Hopf algebroid $\mathcal{H}$ over base algebras* $\mathbf{L}$ *and* $\mathbf{R}$*, the functor $- \otimes_{\mathbf{L}} H : \mathcal{M}_{\mathbf{L}} \to \mathcal{M}_H^H$ is an equivalence.*

Theorem 4.13 is proved by constructing the inverses of the unit (4.7) and the counit (4.8) of the relevant adjunction. Use the notations for the structure maps of a Hopf algebroid in Definition 4.1. For a right $\mathbf{L}$-module V, the inverse of (4.7) is the map $(V \otimes_{\mathbf{L}} H)^{\mathrm{co}\mathcal{H}_R} \to V$, $\sum_i v_i \otimes_{\mathbf{L}} h_i \mapsto \sum_i v_i \cdot s_L(\epsilon_L(h_i))$. For a Hopf module W, denote the $\mathcal{H}_R$-coaction by $w \mapsto w^{[0]} \otimes_{\mathbf{R}} w^{[1]}$ and for the $\mathcal{H}_L$-coaction, write $w \mapsto w_{[0]} \otimes_{\mathbf{L}} w_{[1]}$. Then, an epimorphism $W \to W^{\mathrm{co}\mathcal{H}_R}$ is given by $w \mapsto w^{[0]} \cdot S(w^{[1]})$. The inverse of (4.8) is the map $W \to W^{\mathrm{co}\mathcal{H}_R} \otimes_{\mathbf{L}} H$, $w \mapsto w_{[0]}{}^{[0]} \cdot S(w_{[0]}{}^{[1]}) \otimes_{\mathbf{L}} w_{[1]}$.

4.4. *Integral theory*

In a Hopf algebra H over a commutative ring k, integrals are invariants of the regular module of the underlying k-algebra $\mathbf{H}$, with respect to the character given by the

counit. The study of integrals provides a lot of information about the structure of the k-algebra $\mathbf{H}$. Most significantly, (k-relative) semisimplicity of $\mathbf{H}$ is equivalent to its separability over k and also to the existence of a normalized integral in H [57,58]. Since this extends Maschke's theorem about group algebras, it is known as a Maschke-type theorem. Its dual version relates cosemisimplicity and coseparability of the k-coalgebra underlying H to the existence of normalized cointegrals [57]. Another group of results concerns Frobenius property of $\mathbf{H}$, which is equivalent to the existence of a nondegenerate integral in H [59,60]. Integral theory of Hopf algebras was generalized to Hopf algebroids in [47].

DEFINITION 4.14. For an algebra $\mathbf{R}$, consider a right $\mathbf{R}$-bialgebroid $\mathcal{B}$, with structure maps denoted as in Definition 3.1. *Right integrals* in $\mathcal{B}$ are the invariants of the right regular module of the underlying $\mathbf{R}$-ring $(\mathbf{B}, s)$, with respect to the right character ϵ. Equivalently, invariants of the right regular module of the $\mathbf{R}^{\mathrm{op}}$-ring $(\mathbf{B}, t)$, with respect to ϵ. That is, the elements of

$$B_\epsilon = \{i \in B \mid ib = is(\epsilon(b)), \quad \forall b \in B\} = \{i \in B \mid ib = it(\epsilon(b)),$$

$$\forall b \in B\} \cong \mathrm{Hom}_{\mathbf{B}}(R, B).$$

A right integral i is normalized if $\epsilon(i) = 1_{\mathbf{R}}$.

Symmetrically, *left integrals* in a left $\mathbf{R}$-bialgebroid are defined as invariants of the left regular module of the underlying $\mathbf{R}$-ring or $\mathbf{R}^{\mathrm{op}}$-ring, with respect to the left character defined by the counit. Note that a left integral i in a left bialgebroid $\mathcal{B}$ is a left integral also in $\mathcal{B}_{\mathrm{cop}}$ and a right integral in the right bialgebroids $\mathcal{B}^{\mathrm{op}}$ and $\mathcal{B}^{\mathrm{op}}_{\mathrm{cop}}$.

Since the counit in a right (resp. left) bialgebroid is a right (resp. left) character, there is no way to consider left (resp. right) integrals in a right (resp. left) bialgebroid. On the contrary, the base algebra in a right bialgebroid $\mathcal{B}$ is both a right and a left $\mathcal{B}$-comodule via coactions given by the source and target maps, respectively. Hence, there are corresponding notions of left and right cointegrals.

DEFINITION 4.15. For an algebra $\mathbf{R}$, consider a right $\mathbf{R}$-bialgebroid $\mathcal{B}$, with structure maps denoted as in Definition 3.1. A *right cointegral* on $\mathcal{B}$ is an element of

$$\mathrm{Hom}^B(B, R) = \{\iota \in \mathrm{Hom}_{\mathbf{R}}(B, R) \mid (\iota \underset{\mathbf{R}}{\otimes} B) \circ \Delta = s \circ \iota\}.$$

A right cointegral ι is normalized if $\iota(1_{\mathbf{B}}) = 1_{\mathbf{R}}$.

Symmetrically, a *left cointegral* on $\mathcal{B}$ is an element of ${}^B\mathrm{Hom}(B, R)$.

Left and right cointegrals on a left $\mathbf{R}$-bialgebroid $\mathcal{B}$ are defined analogously as left and right comodule maps $B \to R$.

If a right $\mathbf{R}$-bialgebroid $\mathcal{B}$ is finitely generated and projective as a, say right, $\mathbf{R}$-module (via right multiplication by the source map), then the isomorphism $\mathrm{Hom}_{\mathbf{R}}(R, B^*) \cong \mathrm{Hom}_{\mathbf{R}}(B, R)$ induces an isomorphism ${}_{B^*}\mathrm{Hom}(R, B^*) \cong \mathrm{Hom}^B(B, R)$. Hence in this case, left integrals in the left $\mathbf{R}$-bialgebroid $\mathcal{B}^*$ are the

same as right cointegrals on $\mathcal{B}$. Similar statements hold for all other duals of left and right bialgebroids.

For a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$, left (resp. right) cointegrals on $\mathcal{H}_L$ and $\mathcal{H}_R$ can be shown to be left (resp. right) $\mathcal{H}$-comodule maps.

4.4.1. Maschke-type theorems Recall that an $\mathbf{R}$-ring B is said to be separable provided that the multiplication map $B \otimes_{\mathbf{R}} B \to B$ is a split epimorphism of B-bimodules. The $\mathbf{R}$-ring B is said to be left (resp. right) semisimple (or sometimes $\mathbf{R}$-relatively semisimple) if every left (resp. right) B-module is $\mathbf{R}$-relative projective. That is, every B-module epimorphism, which has an $\mathbf{R}$-module section, is a split epimorphism of B-modules. By a classical result due to Hirata and Sugano [61], a separable $\mathbf{R}$-ring is left and right semisimple. For Hopf algebroids also the converse can be proved.

THEOREM 4.16. *Let $(\mathcal{H}_L, \mathcal{H}_R, S)$ be a Hopf algebroid over base algebras $\mathbf{L}$ and $\mathbf{R}$ with structure maps as denoted in Definition 4.1. The following properties are equivalent.*

(i) *The $\mathbf{R}$-ring $(\mathbf{H}, s_R)$ underlying $\mathcal{H}_R$ is separable.*
(ii) *The $\mathbf{R}^{\mathrm{op}}$-ring $(\mathbf{H}, t_R)$ underlying $\mathcal{H}_R$ is separable.*
(iii) *The $\mathbf{L}$-ring $(\mathbf{H}, s_L)$ underlying $\mathcal{H}_L$ is separable.*
(iv) *The $\mathbf{L}^{\mathrm{op}}$-ring $(\mathbf{H}, t_L)$ underlying $\mathcal{H}_L$ is separable.*
(v) *The $\mathbf{R}$-ring $(\mathbf{H}, s_R)$ underlying $\mathcal{H}_R$ is right semisimple.*
(vi) *The $\mathbf{R}^{\mathrm{op}}$-ring $(\mathbf{H}, t_R)$ underlying $\mathcal{H}_R$ is right semisimple.*
(vii) *The $\mathbf{L}$-ring $(\mathbf{H}, s_L)$ underlying $\mathcal{H}_L$ is left semisimple.*
(viii) *The $\mathbf{L}^{\mathrm{op}}$-ring $(\mathbf{H}, t_L)$ underlying $\mathcal{H}_L$ is left semisimple.*
(ix) *There exists a normalized right integral in $\mathcal{H}_R$.*
(x) *There exists a normalized left integral in $\mathcal{H}_L$.*
(xi) *The counit ϵ_R in $\mathcal{H}_R$ is a split epimorphism of right $\mathbf{H}$-modules.*
(xii) *The counit ϵ_L in $\mathcal{H}_L$ is a split epimorphism of left $\mathbf{H}$-modules.*

Since the source map in a right $\mathbf{R}$-bialgebroid is a right $\mathbf{R}$-module section of the counit, implication (v)⇒(xi) is obvious. For a right $\mathbf{H}$-module section ν of the counit, $\nu(1_{\mathbf{R}})$ is a normalized right integral. Thus, (xi)⇒(ix). The antipode in a Hopf algebroid maps a normalized right integral in $\mathcal{H}_R$ to a normalized left integral in $\mathcal{H}_L$ and vice versa. So (ix)⇔(x). If i is a normalized left integral in $\mathcal{H}_L$, then the map $H \to H \otimes_{\mathbf{R}} H$, $h \mapsto hi^{(1)} \otimes_{\mathbf{R}} S(i^{(2)}) = i^{(1)} \otimes_{\mathbf{R}} S(i^{(2)})h$ is an $\mathbf{H}$-bimodule section of the multiplication in the $\mathbf{R}$-ring underlying $\mathcal{H}_R$ (where the index notation $\Delta_R(h) = h^{(1)} \otimes_{\mathbf{R}} h^{(2)}$ is used for $h \in H$). This proves (x)⇒(i). The remaining equivalences follow by symmetry. Note that equivalences (iv)⇔(viii)⇔(x) ⇔(xii) hold also for a $\times_L$-Hopf algebra $\mathcal{H}_L$ (as discussed in Section 4.6.2).

As an alternative to Theorem 4.16, one can ask about properties of the $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring, underlying a right bialgebroid $\mathcal{H}_R$, and the $\mathbf{L} \otimes_k \mathbf{L}^{\mathrm{op}}$-ring, underlying a left bialgebroid $\mathcal{H}_L$, in a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$. Theorem 4.17 is obtained by application of [39, Theorem 6.5]. For a k-algebra $\mathbf{L}$, consider a left $\mathbf{L}$-bialgebroid $\mathcal{H}_L$. Denote its $\mathbf{L} \otimes_k \mathbf{L}^{\mathrm{op}}$-ring structure by $(\mathbf{H}, s_L, t_L)$ and its $\mathbf{L}$-coring structure by

$(H, \Delta_L, \epsilon_L)$. Look at L as a left $\mathbf{L} \otimes_k \mathbf{L}^{\mathrm{op}}$-module, with action given by left and right multiplications. Look at H as a right $\mathbf{L} \otimes_k \mathbf{L}^{\mathrm{op}}$-module, with action given by right multiplications by s_L and t_L. Note that

$$H \otimes_{\mathbf{L}\otimes_k \mathbf{L}^{\mathrm{op}}} L \cong H/\{\, hs_L(l) - ht_L(l) \mid h \in H,\ l \in L \,\} \tag{4.9}$$

is an $\mathbf{L}$-coring (via quotient maps of Δ_L and ϵ_L) and a left $\mathbf{H}$-module. Hence, we can speak about the invariants of $H \otimes_{\mathbf{L}\otimes_k \mathbf{L}^{\mathrm{op}}} L$ with respect to ϵ_L. An invariant of $H \otimes_{\mathbf{L}\otimes_k \mathbf{L}^{\mathrm{op}}} L$ is said to be *normalized* if the quotient of ϵ_L maps it to $1_{\mathbf{L}}$.

THEOREM 4.17. *Consider a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$ over base k-algebras $\mathbf{L}$ and $\mathbf{R}$. The following assertions are equivalent.*

(i) *The $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring underlying $\mathcal{H}_R$ is separable.*
(ii) *The $\mathbf{L} \otimes_k \mathbf{L}^{\mathrm{op}}$-ring underlying $\mathcal{H}_L$ is separable.*
(iii) *The $\mathbf{R} \otimes_k \mathbf{R}^{\mathrm{op}}$-ring underlying $\mathcal{H}_R$ is right semisimple.*
(iv) *The $\mathbf{L} \otimes_k \mathbf{L}^{\mathrm{op}}$-ring underlying $\mathcal{H}_L$ is left semisimple.*
(v) *There is a normalized invariant in the right $\mathbf{H}$-module $R \otimes_{\mathbf{R}\otimes_k \mathbf{R}^{\mathrm{op}}} H$.*
(vi) *There is a normalized invariant in the left $\mathbf{H}$-module $H \otimes_{\mathbf{L}\otimes_k \mathbf{L}^{\mathrm{op}}} L$.*

For a $\times_L$-Hopf algebra $\mathcal{H}_L$ (as discussed in Section 4.6.2), equivalences (ii)⇔(iv)⇔(vi) in Theorem 4.17 also hold.

Recall that an $\mathbf{R}$-coring B is said to be *coseparable* provided that the comultiplication map $B \to B \otimes_{\mathbf{R}} B$ is a split monomorphism of B-bicomodules. The $\mathbf{R}$-coring B is said to be left (resp. right) *cosemisimple* (or sometimes $\mathbf{R}$-relatively cosemisimple) if every left (resp. right) B-comodule is $\mathbf{R}$-relative injective. That is, every B-comodule monomorphism, which has an $\mathbf{R}$-module retraction, is a split monomorphism of B-comodules. A coseparable $\mathbf{R}$-coring is left and right cosemisimple. For Hopf algebroids also the converse can be proved.

THEOREM 4.18. *For a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$ over base algebras $\mathbf{L}$ and $\mathbf{R}$, the following properties are equivalent.*

(i) *The $\mathbf{R}$-coring underlying $\mathcal{H}_R$ is coseparable.*
(ii) *The $\mathbf{L}$-coring underlying $\mathcal{H}_L$ is coseparable.*
(iii) *The $\mathbf{R}$-coring underlying $\mathcal{H}_R$ is right cosemisimple.*
(iv) *The $\mathbf{R}$-coring underlying $\mathcal{H}_R$ is left cosemisimple.*
(v) *The $\mathbf{L}$-coring underlying $\mathcal{H}_L$ is right cosemisimple.*
(vi) *The $\mathbf{L}$-coring underlying $\mathcal{H}_L$ is left cosemisimple.*
(vii) *There exists a normalized right cointegral on $\mathcal{H}_R$.*
(viii) *There exists a normalized left cointegral on $\mathcal{H}_R$.*
(ix) *There exists a normalized right cointegral on $\mathcal{H}_L$.*
(x) *There exists a normalized left cointegral on $\mathcal{H}_L$.*
(xi) *The source map in $\mathcal{H}_R$ is a split right $\mathcal{H}_R$-comodule monomorphism.*
(xii) *The target map in $\mathcal{H}_R$ is a split left $\mathcal{H}_R$-comodule monomorphism.*
(xiii) *The source map in $\mathcal{H}_L$ is a split left $\mathcal{H}_L$-comodule monomorphism.*
(xiv) *The target map in $\mathcal{H}_L$ is a split right $\mathcal{H}_L$-comodule monomorphism.*

4.4.2. Frobenius Hopf algebroids It was proven by Larson and Sweedler in [59] that every finite-dimensional Hopf algebra over a field is a Frobenius algebra. Although this is not believed to be true for any finitely generated projective Hopf algebra over a commutative ring, Frobenius Hopf algebras form a distinguished class. A Hopf algebra is known to be a Frobenius algebra if and only if it possesses a nondegenerate integral [60]. It is a self-dual property; a nondegenerate integral determines a nondegenerate cointegral, that is, a nondegenerate integral in the dual Hopf algebra.

In a Hopf algebroid, there are four algebra extensions: the ones given by the source and target maps of the two constituent bialgebroids. Among Hopf algebroids, those in which these are Frobenius extensions play an even more distinguished role. Although the dual of any finitely generated projective Hopf algebroid is not known to be a Hopf algebroid, duals of Frobenius Hopf algebroids are Frobenius Hopf algebroids.

While every finitely generated and projective Hopf algebra over a commutative ring k was proved by Pareigis to be a quasi-Frobenius k-algebra in [60], an analogous statement fails to hold for Hopf algebroids. In [47, Section 5], Hopf algebroids were constructed, which are finitely generated and projective over their base algebras (in all the four senses in (3.3) and (3.5)) but are not quasi-Frobenius extensions of the base algebra.

Recall (e.g. from [62]) that an $\mathbf{R}$-ring $(\mathbf{H}, s)$ is said to be *Frobenius* provided that H is a finitely generated and projective left $\mathbf{R}$-module, and ${}^*H := {}_{\mathbf{R}}\mathrm{Hom}(H, R)$ is isomorphic to H as an $\mathbf{H}$-$\mathbf{R}$ bimodule. Equivalently, $(\mathbf{H}, s)$ is a Frobenius $\mathbf{R}$-ring if and only if H is a finitely generated and projective right $\mathbf{R}$-module and $H^* := \mathrm{Hom}_{\mathbf{R}}(H, R)$ is isomorphic to H as an $\mathbf{R}$-$\mathbf{H}$ bimodule. These properties are equivalent also to the existence of an $\mathbf{R}$-bimodule map $\psi : \mathbf{H} \to \mathbf{R}$, the so called Frobenius functional, possessing a dual basis $\sum_i e_i \otimes_{\mathbf{R}} f_i \in H \otimes_{\mathbf{R}} H$, satisfying, for all $h \in H$, $\sum_i e_i \cdot \psi(f_i h) = h = \sum_i \psi(h e_i) \cdot f_i$. The following characterization of Frobenius Hopf algebroids was obtained in [47, Theorem 4.7], see the Corrigendum.

THEOREM 4.19. *Consider a Hopf algebroid $\mathcal{H}$ over base algebras $\mathbf{L}$ and $\mathbf{R}$ with structure maps denoted as in Definition 4.1. Assume that H is a finitely generated and projective left and right $\mathbf{R}$-module via the actions in* (3.3) *and a finitely generated and projective left and right $\mathbf{L}$-module via the actions in* (3.5)*. The following statements are equivalent.*

(i) *The $\mathbf{R}$-ring $(\mathbf{H}, s_R)$ is Frobenius.*
(ii) *The $\mathbf{R}^{\mathrm{op}}$-ring $(\mathbf{H}^{\mathrm{op}}, t_R)$ is Frobenius.*
(iii) *The $\mathbf{L}$-ring $(\mathbf{H}, s_L)$ is Frobenius.*
(iv) *The $\mathbf{L}^{\mathrm{op}}$-ring $(\mathbf{H}^{\mathrm{op}}, t_L)$ is Frobenius.*
(v) *There exists a right cointegral ι on $\mathcal{H}_R$ such that the map $\widetilde{\iota} : H \to \mathrm{Hom}_R(H, R)$, $h \mapsto \iota(h-)$ is bijective.*
(vi) *There exists a left cointegral υ on $\mathcal{H}_L$ such that the map $\widetilde{\upsilon} : H \to {}_{\mathbf{L}}\mathrm{Hom}(H, L)$, $h \mapsto \upsilon(-h)$ is bijective.*
(vii) *There exists a right integral i in $\mathcal{H}_R$ such that the map $\widetilde{i} : {}_{\mathbf{L}}\mathrm{Hom}(H, L) \to H$, $\psi \mapsto t_L(\psi(i_{(2)}))i_{(1)}$ is bijective.*
(viii) *There exists a left integral j in $\mathcal{H}_L$ such that the map $\widetilde{j} : \mathrm{Hom}_{\mathbf{R}}(H, R) \to H$, $\phi \mapsto j^{(2)} t_R(\phi(j^{(1)}))$ is bijective.*

A Hopf algebroid for which these equivalent conditions hold is said to be Frobenius*, and a right (resp. left) integral obeying property (vii) (resp. (viii)) is said to be* non-degenerate.

In a Frobenius Hopf algebroid the antipode S is bijective.

If j is a nondegenerate left integral in $\mathcal{H}_L$, then $\iota := (\widetilde{j})^{-1}(1_{\mathbf{H}})$ is a right cointegral on $\mathcal{H}_R$ and a Frobenius functional for the $\mathbf{R}$-ring $(\mathbf{H}, s_R)$ with dual basis $j^{(1)} \otimes_{\mathbf{R}} S(j^{(2)})$. Thus, (viii)⇒(i). If property (i) holds, then the right cointegrals on $\mathcal{H}_R$ are shown to form a free rank one left $\mathbf{R}$-module $\mathcal{I}$, via the action $r \cdot \iota = \iota(t_R(r)-)$. Using finitely generated projectivity of the right $\mathbf{R}$-module H, the dual $H^* := \mathrm{Hom}_{\mathbf{R}}(H, R)$ can be equipped with a Hopf module structure, with coinvariants $\mathcal{I}$. Hence, Theorem 4.13 implies an isomorphism $H^* \cong H \otimes_{\mathbf{R}} \mathcal{I}$. This isomorphism is used to show that the cyclic generator ι of the $\mathbf{R}$-module $\mathcal{I}$ satisfies condition (v). If there is a right cointegral ι as in part (v), then a nondegenerate left integral j as in part (viii) is constructed in terms of $(\widetilde{\iota})^{-1}$ and a dual basis for the finitely generated projective right $\mathbf{R}$-module H. It is shown to satisfy $(\widetilde{j})^{-1} = \widetilde{\iota} \circ S$, which implies bijectivity of S. The remaining equivalences follow by relations between the source and target maps in $\mathcal{H}_L$ and $\mathcal{H}_R$ and symmetrical versions of the arguments above.

For a Frobenius Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$ over base algebras $\mathbf{L}$ and $\mathbf{R}$, all the four duals $\mathrm{Hom}_{\mathbf{R}}(H, R)$, ${}_{\mathbf{R}}\mathrm{Hom}(H, R)$, $\mathrm{Hom}_{\mathbf{L}}(H, L)$, and ${}_{\mathbf{L}}\mathrm{Hom}(H, L)$ possess (left or right) bialgebroid structures. A left integral j in $\mathcal{H}_L$, such that the map $\widetilde{j}$ in part (viii) of Theorem 4.19 is bijective, determines further similar isomorphisms ${}_{\mathbf{R}}\mathrm{Hom}(H, R) \to H$, $\mathrm{Hom}_{\mathbf{L}}(H, L) \to H$ and ${}_{\mathbf{L}}\mathrm{Hom}(H, L) \to H$. What is more, putting $\iota := (\widetilde{j})^{-1}(1_{\mathbf{H}})$, there is an algebra automorphism of H,

$$\zeta : H \to H, \qquad h \mapsto h^{(2)} t_R(\iota(j h^{(1)})). \tag{4.10}$$

These isomorphisms combine to bialgebroid (anti-) isomorphisms between the four duals of H (cf. [46, Theorem 5.16]).

The Frobenius property of a Hopf algebroid was shown to be self-dual in [46, Theorem 5.17 and Proposition 5.19] in the following sense.

THEOREM 4.20. *Consider a Frobenius Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$ over base algebras $\mathbf{L}$ and $\mathbf{R}$. Let j be a left integral in $\mathcal{H}_L$ such that the map $\widetilde{j}$ in part (viii) of Theorem 4.19 is bijective. Then, the left $\mathbf{R}$-bialgebroid $H^* := \mathrm{Hom}_{\mathbf{R}}(H, R)$ extends to a Hopf algebroid. A bijective antipode is given in terms of the map* (4.10) *by $S^* := (\widetilde{j})^{-1} \circ S \circ \zeta \circ \widetilde{j}$. The right bialgebroid structure is determined by the requirement that S^* is a bialgebroid anti-isomorphism in the sense of Proposition 4.4. This dual Hopf algebroid is Frobenius, with nondegenerate left integral $(\widetilde{j})^{-1}(1_{\mathbf{H}}) \in \mathcal{H}^*$.*

In a paper [63] by Szlachányi, an equivalent description of a Frobenius Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$ was proposed, via the so-called *double algebras*. In this picture, the isomorphism $\widetilde{j}$ in part (viii) of Theorem 4.19 is used to transfer the multiplication in $H^* := \mathrm{Hom}_{\mathbf{R}}(H, R)$ to a second algebra structure in H, with unit j. In this way, four Frobenius ring structures are obtained on H. Note that the coproducts in $\mathcal{H}_L$ and $\mathcal{H}_R$ correspond canonically to the Frobenius ring structures transferred from H^*. In

this approach, the Hopf algebroid axioms are formulated as compatibility conditions between the two algebra structures on H.

4.5. *Galois theory of Hopf algebroids*

Galois extensions by Hopf algebras are the same as Galois extensions by the constituent bialgebra. Still, since the structure of a Hopf algebra is more complex than that of a bialgebra, it allows to derive stronger results. On the contrary, Galois theory of a Hopf algebroid is more conceptually different from the Galois theory of the constituent bialgebroids. As it is discussed in Section 4.3, comodules of a Hopf algebroid carry more structure than comodules of any constituent bialgebroid. Consequently, Galois theory of Hopf algebroids, discussed in this section, is significantly richer than the theory of bialgebroids. In particular, for a comodule algebra of a Hopf algebroid (cf. Section 4.3.3), several theorems concerning an equivalence between the category of relative Hopf modules and the category of modules of the coinvariant subalgebra, that is descent theorems, can be proved.

By Definition 4.10 and Theorem 4.9, a right comodule algebra M of a Hopf algebroid $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$ is both an $\mathcal{H}_L$-comodule algebra and an $\mathcal{H}_R$-comodule algebra. Denote the $\mathcal{H}_R$-coinvariant subalgebra of M by $\mathbf{N}$. In light of Proposition 4.8, there are two corresponding canonical maps

$$M \underset{\mathbf{N}}{\otimes} M \to M \underset{\mathbf{R}}{\otimes} H, \qquad m \underset{\mathbf{N}}{\otimes} m' \mapsto m m'^{[0]} \underset{\mathbf{R}}{\otimes} m'^{[1]} \qquad \text{and} \qquad (4.11)$$

$$M \underset{\mathbf{N}}{\otimes} M \to M \underset{\mathbf{L}}{\otimes} H, \qquad m \underset{\mathbf{N}}{\otimes} m' \mapsto m_{[0]} m' \underset{\mathbf{L}}{\otimes} m_{[1]}, \qquad (4.12)$$

where $m \mapsto m^{[0]} \otimes_R m^{[1]}$ and $m \mapsto m_{[0]} \otimes_L m_{[1]}$ denote the $\mathcal{H}_R$-coaction and the $\mathcal{H}_L$-coaction on M, respectively. In general, the bijectivity of the two canonical maps (4.11) and (4.12) are not known to be equivalent. Only a partial result [54, Lemma 3.3] is known.

PROPOSITION 4.21. *If the antipode S in a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$ is bijective, then the $\mathcal{H}_R$-canonical map* (4.11) *is bijective if and only if the $\mathcal{H}_L$-canonical map* (4.12) *is bijective.*

This follows by noting that the two canonical maps differ by the isomorphism Φ_A from (4.5).

By Propositions 4.8 and 4.21, for a right comodule algebra M of a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$ with a bijective antipode, an algebra extension $\mathbf{N} \subseteq \mathbf{M}$ is $\mathcal{H}_L$-Galois if and only if it is $\mathcal{H}_R$-Galois.

REMARK 4.22. Consider a Hopf algebroid $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$ over base algebras $\mathbf{L}$ and $\mathbf{R}$, which is finitely generated and projective as a left $\mathbf{L}$-module via the source map in $\mathcal{H}_L$ and as a left $\mathbf{R}$-module via the target map in $\mathcal{H}_R$. Then, H is in particular flat as a left $\mathbf{L}$-module and as a left $\mathbf{R}$-module. Therefore, by Theorem 4.7 and Theorem 4.9,

the category of right $\mathcal{H}$-comodules is strict monoidally isomorphic to the category of comodules for both the constituent left and right bialgebroids $\mathcal{H}_L$ and $\mathcal{H}_R$. Thus, comodule algebras for $\mathcal{H}$ coincide with comodule algebras for $\mathcal{H}_L$ or $\mathcal{H}_R$.

Furthermore, if the antipode is bijective, it follows from Propositions 4.8 and 4.21 that an algebra extension $\mathbf{N} \subseteq \mathbf{M}$ is a right $\mathcal{H}_R$-Galois extension if and only if it is a right $\mathcal{H}_L$-Galois extension.

4.5.1. Depth 2 Frobenius extensions An analog of Theorem 3.27 for Frobenius Hopf algebroids is [12, Theorem 3.6].

THEOREM 4.23. *An algebra extension $\mathbf{N} \subseteq \mathbf{M}$ is a right Galois extension by a Frobenius Hopf algebroid (i.e. by any of its constituent bialgebroids) if and only if it is a Frobenius extension, it is balanced, and it satisfies the (left and right) depth 2 conditions.*

In a case of a Frobenius extension $\mathbf{N} \subseteq \mathbf{M}$, the left and right depth 2 properties are equivalent. By Remark 4.22, for a Frobenius Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$, the $\mathcal{H}_L$- and $\mathcal{H}_R$-Galois properties of an extension are equivalent. By Theorem 3.24, a right depth 2 and balanced algebra extension $\mathbf{N} \subseteq \mathbf{M}$ are Galois extension by a right bialgebroid $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$. In addition, if $\mathbf{N} \subseteq \mathbf{M}$ is a Frobenius extension, then $(M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$ is shown to be a constituent right bialgebroid in a Frobenius Hopf algebroid. A nondegenerate (left and right) integral $\sum_i m_i \otimes_{\mathbf{N}} m_i' \in (M \otimes_{\mathbf{N}} M)^{\mathbf{N}}$ is provided by the dual basis of a Frobenius functional $\psi : M \to N$. A nondegenerate (left and right) integral in the dual Hopf algebroid ${}_{\mathbf{N}}\mathrm{End}_{\mathbf{N}}(M)$ is provided by ψ. In the converse direction, note that a right comodule M of a Hopf algebroid $\mathcal{H}$ is an $\mathcal{H}^*$-module. If $\mathcal{H}$ is a Frobenius Hopf algebroid and $\mathbf{N} \subseteq \mathbf{M}$ is an $\mathcal{H}$-Galois extension, a Frobenius functional $M \to N$ is given by the action of a nondegenerate integral in $\mathcal{H}^*$.

4.5.2. Cleft extensions by Hopf algebroids For an algebra $\mathbf{M}$ and a coalgebra C over a commutative ring k, $\mathrm{Hom}_k(C, M)$ is a k-algebra via the convolution product

$$(f \diamond g)(c) := f(c_{(1)})g(c_{(2)}), \qquad \text{for } f, g \in \mathrm{Hom}_k(C, M),\ c \in C. \tag{4.13}$$

A comodule algebra M of a k-Hopf algebra H is said to be a cleft extension of its coinvariant subalgebra $\mathbf{N}$ provided that there exists a convolution invertible map $j \in \mathrm{Hom}_k(H, M)$, which is an H-comodule map. The relevance of cleft extensions by Hopf algebras stems from Doi and Takeuchi's observation in [64] that $\mathbf{N} \subseteq \mathbf{M}$ is a cleft extension if and only if it is a Galois extension and an additional normal basis property holds, that is, $M \cong N \otimes_k H$ as a left $\mathbf{N}$-module right H-comodule. What is more (establishing an even stronger similarity with Galois extensions of fields), $\mathbf{N} \subseteq \mathbf{M}$ is a cleft extension if and only if $\mathbf{M}$ is isomorphic to a crossed product of $\mathbf{N}$ with H with respect to an invertible 2-cocycle [64, 65].

The results have been extended to Hopf algebroids in [48]. To formulate the definition of a cleft extension, as a first step, a generalized convolution product has to be introduced. Using notations as in Definition 4.1, in a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$

there is an $\mathbf{L}$-coring $(H, \Delta_L, \epsilon_L)$ and an $\mathbf{R}$-coring $(H, \Delta_R, \epsilon_R)$. Consider an $\mathbf{L}\otimes_k \mathbf{R}$-ring M, with multiplications $\mu_{\mathbf{L}} : M \otimes_{\mathbf{L}} M \to M$ and $\mu_{\mathbf{R}} : M \otimes_{\mathbf{R}} M \to M$. For these data, the convolution algebra (4.13) can be generalized to a *convolution category*. It has two objects, conveniently labelled by $\mathbf{L}$ and $\mathbf{R}$. For $P, Q \in \{\mathbf{L}, \mathbf{R}\}$, morphisms from P to Q are Q-P bimodule maps $H \to M$, where the bimodule structure of the domain is determined by the (P- and Q-) coring structures of H and the bimodule structure of the codomain is determined by the (P- and Q-) ring structures of M. For $P, Q, T \in \{\mathbf{L}, \mathbf{R}\}$, and morphisms $f : Q \to P$ and $g : T \to Q$, composition is given by a convolution product

$$f \diamond g := \mu_Q \circ (f \underset{Q}{\otimes} g) \circ \Delta_Q. \tag{4.14}$$

Recall from Theorem 4.9 that a right comodule algebra of a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$ has a canonical $\mathbf{R}$-ring structure over the base algebra $\mathbf{R}$ of $\mathcal{H}_R$.

Definition 4.24. Consider a Hopf algebroid $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$ over base algebras $\mathbf{L}$ and $\mathbf{R}$. A right $\mathcal{H}$-comodule algebra M is said to be a *cleft extension* of the $\mathcal{H}_R$-coinvariant subalgebra $\mathbf{N}$ provided that the following properties hold.

(i) The canonical $\mathbf{R}$-ring structure of M extends to an $\mathbf{L} \otimes_k \mathbf{R}$-ring structure.
(ii) There exists an invertible morphism $j : \mathbf{R} \to \mathbf{L}$ in the convolution category (4.14), which is a right $\mathcal{H}$-comodule map.

In an $\mathcal{H}$-cleft extension $\mathbf{N} \subseteq \mathbf{M}$, $\mathbf{N}$ can be proved to be an $\mathbf{L}$-subring of $\mathbf{M}$.

In a Hopf algebroid $\mathcal{H}$, using the notations introduced in Definition 4.1, an $\mathbf{L}\otimes_k \mathbf{R}$-ring is given by $(\mathbf{H}, s_L, s_R)$. The identity map of H is a morphism $\mathbf{R} \to \mathbf{L}$ in the corresponding convolution category (4.14). It is obviously right $\mathcal{H}$-colinear. What is more, the antipode is its inverse by axioms (iii) and (iv) in Definition 4.1. Hence, the right regular comodule algebra of a Hopf algebroid is a cleft extension of the coinvariant subalgebra $t_R(\mathbf{R})$. This extends a well-known fact that the right regular comodule algebra of a Hopf algebra, over a commutative ring k, is a cleft extension of k. A further similarity between cleft extensions by Hopf algebras and Hopf algebroids is expressed by the following Theorem 4.25.

Theorem 4.25. *Consider a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$ over base algebras* $\mathbf{L}$ *and* $\mathbf{R}$. *Its right comodule algebra M is a cleft extension of the $\mathcal{H}_R$-coinvariant subalgebra* $\mathbf{N}$ *if and only if the following properties hold.*

(i) $\mathbf{N} \subseteq \mathbf{M}$ *is a Galois extension by $\mathcal{H}_R$;*
(ii) *the* normal basis condition *holds, that is, $M \cong N \otimes_{\mathbf{L}} H$ as left* $\mathbf{N}$*-modules right $\mathcal{H}$-comodules.*

Note the appearance of the two base algebras $\mathbf{L}$ and $\mathbf{R}$ in conditions (i) and (ii) in Theorem 4.25.

Another characterization of a cleft extension by a Hopf algebroid can be given by using the construction of a crossed product.

DEFINITION 4.26. Consider a left bialgebroid $\mathcal{B}$ over a base k-algebra $\mathbf{L}$. Denote its structure maps as in Definition 3.3. We say that $\mathcal{B}$ measures an $\mathbf{L}$-ring N with unit map $\iota : \mathbf{L} \to \mathbf{N}$ if there exists a k-module map: $\cdot B \otimes_k N \to N$, the so-called *measuring*, such that, for $b \in B$, $l \in L$ and $n, n' \in N$, the following axioms are satisfied.

(i) $b \cdot 1_{\mathbf{N}} = \iota(\epsilon(b))$,
(ii) $(t(l)b) \cdot n = (b \cdot n)\iota(l)$ and $(s(l)b) \cdot n = \iota(l)(b \cdot n)$,
(iii) $b \cdot (nn') = (b_{(1)} \cdot n)(b_{(2)} \cdot n')$.

Note that in Definition 4.26, condition (iii) makes sense in view of (ii).

Consider a left bialgebroid $\mathcal{B}$ over a k-algebra $\mathbf{L}$ and denote its structure maps as in Definition 3.3. Let N be an $\mathbf{L}$-ring with unit $\iota : \mathbf{L} \to \mathbf{N}$, which is measured by $\mathcal{B}$. These data determine a category $\mathcal{C}(\mathcal{B}, N)$ as follows. Consider $B \otimes_k B$ as an $\mathbf{L}$-bimodule via left multiplication by s and t in the first factor. For an element f in ${}_{\mathbf{L}}\mathrm{Hom}_{\mathbf{L}}(B \otimes_k B, N)$, consider the following ($\mathbf{L}$-balancing) conditions. For $a, b \in B$, and $l \in L$,

$$(T\circ) \qquad f(a \underset{k}{\otimes} s(l)b) = f(as(l) \underset{k}{\otimes} b)$$
$$(S\circ) \qquad f(a \underset{k}{\otimes} t(l)b) = f(at(l) \underset{k}{\otimes} b)$$
$$(T\bullet) \qquad f(a \underset{k}{\otimes} s(l)b) = (a_{(1)} \cdot \iota(l))\, f(a_{(2)} \underset{k}{\otimes} b)$$
$$(S\bullet) \qquad f(a \underset{k}{\otimes} t(l)b) = f(a_{(1)} \underset{k}{\otimes} b)(a_{(2)} \cdot \iota(l)).$$

Define a category $\mathcal{C}(\mathcal{B}, N)$ of two objects $\circ$ and $\bullet$. For two objects $X, Y \in \{\circ, \bullet\}$, the morphisms $X \to Y$ are elements of ${}_{\mathbf{L}}\mathrm{Hom}_{\mathbf{L}}(B \otimes_k B, N)$, satisfying conditions (SX) and (TY). The composition of the morphisms $g : X \to Y$ and $f : Y \to Z$ is given by

$$(f \diamond g)(a \underset{k}{\otimes} b) := f(a_{(1)} \underset{k}{\otimes} b_{(1)})g(a_{(2)} \underset{k}{\otimes} b_{(2)}).$$

The unit morphism at the object $\circ$ is the map $a \otimes_k b \mapsto (ab) \cdot 1_{\mathbf{N}} = \iota(\epsilon(ab))$ and the unit morphism at the object $\bullet$ is the map $a \otimes_k b \mapsto a \cdot (b \cdot 1_{\mathbf{N}})$.

DEFINITION 4.27. Consider a left bialgebroid $\mathcal{B}$ over a k-algebra $\mathbf{L}$ with structure maps as in Definition 3.3. Let N be a $\mathcal{B}$-measured $\mathbf{L}$-ring, with unit $\iota : \mathbf{L} \to \mathbf{N}$. An N-valued *2-cocycle* on $\mathcal{B}$ is a morphism $\circ \to \bullet$ in the category $\mathcal{C}(\mathcal{B}, N)$ given in the previous paragraph such that, for all $a, b, c \in B$,

(i) $\sigma(1_{\mathbf{B}}, b) = \iota\big(\epsilon(b)\big) = \sigma(b, 1_{\mathbf{B}})$,
(ii) $\big(a_{(1)} \cdot \sigma(b_{(1)}, c_{(1)})\big)\sigma(a_{(2)}, b_{(2)}c_{(2)}) = \sigma(a_{(1)}, b_{(1)})\sigma(a_{(2)}b_{(2)}, c)$.

The $\mathcal{B}$-measured L-ring N is called a *σ-twisted $\mathcal{B}$-module* if in addition, for $n \in N$ and $a, b \in B$,

(iii) $1_{\mathbf{B}} \cdot n = n$,
(iv) $\big(a_{(1)} \cdot (b_{(1)} \cdot n)\big)\sigma(a_{(2)}, b_{(2)}) = \sigma(a_{(1)}, b_{(1)})(a_{(2)}b_{(2)} \cdot n)$.

An N-valued 2-cocycle σ on $\mathcal{B}$ is said to be *invertible* if it is invertible as a morphism in $\mathcal{C}(\mathcal{B}, N)$.

Note that in Definition 4.27, conditions (ii) and (iv) make sense in view of the module map and balanced properties of σ. If an $\mathbf{L}$-ring N is measured by a left $\mathbf{L}$-bialgebroid $\mathcal{B}$, then the $\mathbf{L}^{\mathrm{op}}$-ring N^{op} is measured by the co-opposite left $\mathbf{L}^{\mathrm{op}}$-bialgebroid $\mathcal{B}_{\mathrm{cop}}$. The inverse of an N-valued 2-cocycle σ on $\mathcal{B}$ turns out to be an N^{op}-valued 2-cocycle on $\mathcal{B}_{\mathrm{cop}}$.

For a left bialgebroid $\mathcal{B}$ over an algebra $\mathbf{L}$, with structure maps denoted as in Definition 3.3, the base algebra $\mathbf{L}$ is measured by $\mathcal{B}$ via the left action $b \cdot l := \epsilon(bs(l))$. For this measuring, conditions $(S\circ)$ and $(S\bullet)$ are equivalent and also conditions $(T\circ)$ and $(T\bullet)$ are equivalent. Consequently, an $\mathbf{L}$-valued 2-cocycle on $\mathcal{B}$ in the sense of Definition 4.27 is equivalent to a cocycle as considered in Section 3.4.2. Extending the cocycle double twists of Section 3.4.2, one can consider more general deformations of a Hopf algebroid $\mathcal{H}$ (or a $\times_L$-Hopf algebra $\mathcal{B}$, discussed in Section 4.6.2) by an N-valued invertible 2-cocycle σ in Definition 4.27 (cf. [45, Appendix]). In that construction, the base algebra $\mathbf{L}$ of $\mathcal{H}_L$ is replaced by an $\mathcal{H}_L$-measured $\mathbf{L}$-ring N. In particular, the Connes and Moscovici's bialgebroids in Section 3.4.6 arise in this way.

A *crossed product* $N\#_\sigma B$ of a left $\mathbf{L}$-bialgebroid $\mathcal{B}$ with a σ-twisted $\mathcal{B}$-module N, with respect to an N-valued 2-cocycle σ, is the the $\mathbf{L}$-module tensor product $N \otimes_{\mathbf{L}} B$ (where B is a left $\mathbf{L}$-module via the source map s), with associative and unital multiplication

$$(n\#b)(n'\#b') = n(b_{(1)} \cdot n')\sigma(b_{(2)}, b'_{(1)})\#b_{(3)}b'_{(2)}, \quad \text{for } n\#b, n'\#b' \in N \underset{\mathbf{L}}{\otimes} B.$$

Equivalence classes of crossed products with a bialgebroid were classified in [48, Section 4].

THEOREM 4.28. *A right comodule algebra M of a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$ is a cleft extension of the $\mathcal{H}_R$-coinvariant subalgebra $\mathbf{N}$ if and only if M is isomorphic, as a left $\mathbf{N}$-module and right $\mathcal{H}$-comodule algebra, to a crossed product algebra $N\#_\sigma\mathcal{H}_L$, with respect to some invertible N-valued 2-cocycle σ on $\mathcal{H}_L$.*

4.5.3. The structure of Galois extensions by Hopf algebroids In the theory of Galois extensions by Hopf algebras (with a bijective antipode), important tools are provided by theorems, which state that in appropriate situations, *surjectivity* of the canonical map implies its bijectivity, that is, the Galois property of an algebra extension $\mathbf{N} \subseteq \mathbf{M}$. There are two big groups of such theorems. In the first group, a Hopf algebra H is assumed to be a flat module over its commutative base ring k and its regular comodule algebra is assumed to be a projective H-comodule. These properties hold in particular if H is a finitely generated and projective k-module, in which case such a theorem was proven first by Kreimer and Takeuchi [66]. In the second group of such results, due to Schneider, H is assumed to be a projective k-module and its comodule algebra M is assumed to be a k-relative injective H-comodule [67].

Analogous results for extensions by a Hopf algebroid $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$ were obtained in the papers [12, 54, 55], respectively. A common philosophy behind such theorems originates from a paper [43] of Schauenburg (on the Hopf algebra case).

The key idea is to investigate a lifting (4.16) of the canonical map (4.11) introduced for an $\mathcal{H}$-comodule algebra M in the following. By a general result [68, Theorem 2.1] about Galois comodules, split surjectivity of the lifted canonical map (4.16) as a morphism of relative $(M, \mathcal{H}_R)$-Hopf modules implies the $\mathcal{H}_R$-Galois property, that is, bijectivity of (4.11), whenever $(M \otimes_{\mathbf{T}} M)^{\mathrm{co}\mathcal{H}_R} = M \otimes_{\mathbf{T}} M^{\mathrm{co}\mathcal{H}_R}$, where $(-)^{\mathrm{co}\mathcal{H}_R}$ denotes the $\mathcal{H}_R$-coinvariants functor.

For a Hopf algebra H over a commutative ring k and a right H-comodule algebra M with coinvariant subalgebra $\mathbf{N}$, the canonical map $M \otimes_{\mathbf{N}} M \to M \otimes_k H$ can be lifted to a map

$$M \underset{k}{\otimes} M \longrightarrow\!\!\!\!\!\rightarrow M \underset{\mathbf{N}}{\otimes} M \xrightarrow{\text{can}} M \underset{k}{\otimes} H \ . \tag{4.15}$$

More generally, consider a Hopf algebroid $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$ over base k-algebras $\mathbf{L}$ and $\mathbf{R}$. For a right $\mathcal{H}$-comodule algebra M, the canonical map (4.11) can be lifted to

$$M \underset{\mathbf{T}}{\otimes} M \to M \underset{\mathbf{R}}{\otimes} H, \qquad m \underset{\mathbf{T}}{\otimes} m' \mapsto mm'^{[0]} \underset{\mathbf{R}}{\otimes} m'^{[1]} \tag{4.16}$$

for any k-algebra $\mathbf{T}$ such that the $\mathcal{H}_R$-coinvariant subalgebra $\mathbf{N}$ of M is a $\mathbf{T}$-ring. The map (4.16) is a morphism of right–right $(M, \mathcal{H})$-relative Hopf modules.

THEOREM 4.29. *Consider a Hopf algebroid $\mathcal{H}$ over base k-algebras $\mathbf{L}$ and $\mathbf{R}$, with a bijective antipode. Denote its structure maps as in Definition 4.1. Assume that H is a flat left $\mathbf{R}$-module (via right multiplication by t_R) and a projective right $\mathcal{H}_R$-comodule (via Δ_R). Let M be a right $\mathcal{H}$-comodule algebra with $\mathcal{H}_R$-coinvariant subalgebra $\mathbf{N}$. Under these assumptions the following statements hold.*

(1) *If the $\mathcal{H}_R$-coinvariants of the right $\mathcal{H}$-comodule $M \otimes_k M$ (with coactions given via the second factor) are precisely the elements of $M \otimes_k N$, then the canonical map* (4.11) *is bijective if and only if it is surjective.*
(2) *If the canonical map* (4.11) *is bijective, then M is a projective right $\mathbf{N}$-module.*

Since coinvariants are defined as a kernel, the coinvariants of $M \otimes_k M$ are precisely the elements of $M \otimes_k N$ if, for example, M is a flat k-module. To gain an impression of the proof of part (1) of Theorem 4.29, note that flatness of the left $\mathbf{R}$-module H and projectivity of the right regular $\mathcal{H}_R$-comodule together imply that $M \otimes_{\mathbf{R}} H$ is projective as a right–right $(M, \mathcal{H}_R)$-relative Hopf module. Hence, if the canonical map (4.11) is surjective, then the (surjective) lifted canonical map (4.16) is a split epimorphism of right–right $(M, \mathcal{H}_R)$-relative Hopf modules, for any possible k-algebra $\mathbf{T}$. Thus, bijectivity of the canonical map (4.11) follows by [68, Theorem 2.1]. Part (2) of Theorem 4.29 follows by exactness of the naturally equivalent functors $\mathrm{Hom}^{\mathcal{H}_R}(H, -) \cong \mathrm{Hom}_{\mathbf{N}}(M, (-)^{\mathrm{co}\mathcal{H}_R})$, which is a consequence of the projectivity of the right regular $\mathcal{H}_R$-comodule.

If in a Hopf algebroid $\mathcal{H}$ with a bijective antipode, H is a finitely generated and projective left and right $\mathbf{R}$-module, and a finitely generated and projective left and right $\mathbf{L}$-module, with respect to the respective actions in (3.3) and (3.5),

then it is obviously a flat left $\mathbf{R}$-module. Furthermore, in this case, the right regular $\mathcal{H}_R$-comodule can be shown to be projective by using Theorem 4.13. The following Weak Structure Theorem 4.30 is, thus, based on Remark 4.22, part (2) of Theorem 4.29 and its application to the right comodule algebra M^{op} of the opposite Hopf algebroid $\mathcal{H}^{\mathrm{op}}$, and a theorem [69, Theorem 3.5] about Galois corings.

THEOREM 4.30. *Consider a Hopf algebroid $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$ over base algebras $\mathbf{L}$ and $\mathbf{R}$, with a bijective antipode S. Assume that H is a finitely generated and projective left and right $\mathbf{R}$-module, and a finitely generated and projective left and right $\mathbf{L}$-module, with respect to the respective actions in* (3.3) *and* (3.5). *For a right $\mathcal{H}$-comodule algebra M, with $\mathcal{H}_R$-coinvariant subalgebra $\mathbf{N}$, the following statements are equivalent.*

(i) *The extension $\mathbf{N} \subseteq \mathbf{M}$ is $\mathcal{H}_R$-Galois.*
(ii) *M is a generator in the category $\mathcal{M}_M^{\mathcal{H}} \cong \mathcal{M}_M^{\mathcal{H}_R} \cong \mathcal{M}_M^{\mathcal{H}_L}$.*
(iii) *The $\mathcal{H}_R$-coinvariants functor $\mathcal{M}_M^{\mathcal{H}} \to \mathcal{M}_{\mathbf{N}}$ is fully faithful.*
(iv) *The extension $\mathbf{N} \subseteq \mathbf{M}$ is $\mathcal{H}_L$-Galois.*
(v) *M is a generator in the category ${}_M\mathcal{M}^{\mathcal{H}} \cong {}_M\mathcal{M}^{\mathcal{H}_R} \cong {}_M\mathcal{M}^{\mathcal{H}_L}$.*
(vi) *The $\mathcal{H}_L$-coinvariants functor ${}_M\mathcal{M}^{\mathcal{H}} \to {}_{\mathbf{N}}\mathcal{M}$ is fully faithful.*

Furthermore, if these equivalent conditions hold, then M is a projective left and right $\mathbf{N}$-module.

It was a key observation by Doi that relative injectivity of a comodule algebra M of a Hopf algebra H is equivalent to the existence of the so-called total integral – meaning an H-comodule map $j : H \to M$ such that $j(1_{\mathbf{H}}) = 1_{\mathbf{M}}$ [70]. This fact extends to Hopf algebroids. Recall (e.g. from [71]) that, for any functor $\mathbb{U} : \mathcal{A} \to \mathcal{B}$, between any categories $\mathcal{A}$ and $\mathcal{B}$, an object $A \in \mathcal{A}$ is said to be $\mathbb{U}$-*injective* if the map $\mathrm{Hom}_{\mathcal{A}}(g, A) : \mathrm{Hom}_{\mathcal{A}}(Y, A) \to \mathrm{Hom}_{\mathcal{A}}(X, A)$ is surjective, for any objects $X, Y \in \mathcal{A}$, and all morphisms $g \in \mathrm{Hom}_{\mathcal{A}}(X, Y)$ such that $\mathbb{U}(g)$ is a split monomorphism in $\mathcal{B}$. If $\mathbb{U}$ has a right adjoint, then $\mathbb{U}$-injectivity of an object A is equivalent to the unit of the adjunction, evaluated at A, being a split monomorphism in $\mathcal{A}$ (see [71, Proposition 1]). For example, for a Hopf algebra H over a commutative ring k, injective objects with respect to the forgetful functor $\mathcal{M}^H \to \mathcal{M}_k$ are precisely relative injective H-comodules.

A version of Theorem 4.31 below was proved in [55, Theorem 4.1], using the notion of *relative separability* of a forgetful functor. Recall from Remark 4.6 that the opposite of a right comodule algebra M of a Hopf algebroid $\mathcal{H}$ has a canonical structure of a left $\mathcal{H}$-comodule algebra.

THEOREM 4.31. *For a right comodule algebra M of a Hopf algebroid $\mathcal{H} = (\mathcal{H}_L, \mathcal{H}_R, S)$, the following statements are equivalent.*

(i) *There exists a right $\mathcal{H}$-comodule map (resp. right $\mathcal{H}_L$-comodule map) $j : H \to M$, such that $j(1_{\mathbf{H}}) = 1_{\mathbf{M}}$.*

(ii) *M is injective with respect to the forgetful functor $\mathcal{M}^{\mathcal{H}} \to \mathcal{M}_{\mathbf{L}}$ (resp. with respect to the forgetful functor $\mathcal{M}^{\mathcal{H}_L} \to \mathcal{M}_{\mathbf{L}}$, i.e. M is a relative injective $\mathcal{H}_L$-comodule).*

(iii) *Any object in the category $\mathcal{M}_M^{\mathcal{H}}$ of right–right relative Hopf modules is injective with respect to the forgetful functor $\mathcal{M}^{\mathcal{H}} \to \mathcal{M}_{\mathbf{L}}$ (resp. with respect to the forgetful functor $\mathcal{M}^{\mathcal{H}_L} \to \mathcal{M}_{\mathbf{L}}$).*

If in addition the antipode S is bijective, then assertions (i)–(iii) are also equivalent to

(iv) *There exists a left $\mathcal{H}$-comodule map (resp. left $\mathcal{H}_R$-comodule map) $j' : H \to M$ such that $j'(1_{\mathbf{H}}) = 1_{\mathbf{M}}$.*

Hence, assertions (i)–(iv) are also equivalent to the symmetrical versions of (ii) and (iii).

The key idea behind Theorem 4.31 is the observation that both forgetful functors $\mathcal{M}^{\mathcal{H}} \to \mathcal{M}_{\mathbf{L}}$ and $\mathcal{M}^{\mathcal{H}_L} \to \mathcal{M}_{\mathbf{L}}$ possess left adjoints $- \otimes_{\mathbf{L}} H$ (cf. (4.3)). A correspondence can be established between comodule maps j as in part (i) and natural retractions of the counit of the adjunction, that is, of the $\mathcal{H}_L$-coaction.

Based on [72, Theorem 4.7] and Theorem 4.31, also the following Strong Structure Theorem holds.

THEOREM 4.32. *Consider a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$ over base algebras $\mathbf{L}$ and $\mathbf{R}$, with a bijective antipode S. Assume that H is a finitely generated and projective left and right $\mathbf{R}$ and $\mathbf{L}$-module with respect to the actions in* (3.3) *and* (3.5). *For a right $\mathcal{H}_R$- (equivalently, right $\mathcal{H}_L$-), Galois extension $\mathbf{N} \subseteq \mathbf{M}$, the following statements are equivalent.*

(i) *M is a faithfully flat right $\mathbf{N}$-module.*
(ii) *The inclusion $\mathbf{N} \to \mathbf{M}$ splits in $\mathcal{M}_{\mathbf{N}}$.*
(iii) *M is a generator of right $\mathbf{N}$-modules.*
(iv) *M is a faithfully flat left $\mathbf{N}$-module.*
(v) *The inclusion $\mathbf{N} \to \mathbf{M}$ splits in ${}_{\mathbf{N}}\mathcal{M}$.*
(vi) *M is a generator of left $\mathbf{N}$-modules.*
(vii) *The functor $- \otimes_{\mathbf{N}} M : \mathcal{M}_{\mathbf{N}} \to \mathcal{M}_M^{\mathcal{H}}$ is an equivalence.*
(viii) *M is a projective generator in $\mathcal{M}_M^{\mathcal{H}}$.*
(ix) *There exists a right $\mathcal{H}$-comodule map $j : M \to H$ such that $j(1_{\mathbf{H}}) = 1_{\mathbf{M}}$.*

Note that if a Hopf algebra H over a commutative ring k is a projective k-module, then $M \otimes_k H$ is a projective left $\mathbf{M}$-module for any right H-comodule algebra M. Thus, denoting the subalgebra of coinvariants in M by $\mathbf{N}$, surjectivity of the canonical map $M \otimes_{\mathbf{N}} M \to M \otimes_k H$ implies that its lifted version (4.15) is a split epimorphism of left $\mathbf{M}$-modules, so in particular of k-modules. By Schneider's result [67, Theorem I], if the antipode of H is bijective and M is a k-relative injective right H-comodule, then bijectivity of the canonical map follows from the k-module splitting of its lifted version (4.15). To formulate the following generalization, Theorem 4.33, of this

result, note that the lifted canonical map (4.16) is an $\mathbf{L}$-bimodule map, with respect to the $\mathbf{L}$-actions $l\cdot(m\otimes_{\mathbf{T}} m')\cdot l' := m\cdot\epsilon_R(s_L(l))\otimes_{\mathbf{T}}\epsilon_R(t_L(l'))\cdot m'$ (recall that the $\mathbf{R}$-, and $\mathbf{T}$-actions on M commute by virtue of (3.14)) and $l\cdot(m\otimes_{\mathbf{R}} h)\cdot l' := m\otimes_{\mathbf{R}} s_L(l)t_L(l')h$, on its domain and codomain, respectively, where the notations from Definition 4.1 are used.

THEOREM 4.33. *Consider a Hopf algebroid $\mathcal{H}$ with a bijective antipode over base algebras $\mathbf{L}$ and $\mathbf{R}$. Denote its structure maps as in Definition 4.1. Let M be a right $\mathcal{H}$-comodule algebra with $\mathcal{H}_R$-coinvariants $\mathbf{N}$. Let $\mathbf{T}$ be a k-algebra such that $\mathbf{N}$ is a $\mathbf{T}$-ring. In this setting, if the lifted canonical map* (4.16) *is a split epimorphism of right $\mathbf{L}$-modules, then the following statements are equivalent.*

(i) $\mathbf{N}\subseteq\mathbf{M}$ *is an $\mathcal{H}_R$-Galois extension and the inclusion $\mathbf{N}\to\mathbf{M}$ splits in $\mathcal{M}_{\mathbf{N}}$.*

(ii) $\mathbf{N}\subseteq\mathbf{M}$ *is an $\mathcal{H}_R$-Galois extension and the inclusion $\mathbf{N}\to\mathbf{M}$ splits in ${}_{\mathbf{N}}\mathcal{M}$.*

(iii) *There exists a right $\mathcal{H}$-comodule map $j: H\to M$ such that $j(1_{\mathbf{H}})=1_{\mathbf{M}}$.*

(iv) $M\otimes_{\mathbf{N}} -: {}_{\mathbf{N}}\mathcal{M}\to{}^{\mathcal{H}}_{M}\mathcal{M}$ *is an equivalence and the inclusion $\mathbf{N}\to\mathbf{M}$ splits in $\mathcal{M}_{\mathbf{N}}$.*

Furthermore, if the equivalent properties (i)–(iii) hold, then M is a $\mathbf{T}$-relative projective right $\mathbf{N}$-module.

Note that by Propositions 4.8 and 4.21, in parts (i) and (ii) the $\mathcal{H}_R$-Galois property can be replaced equivalently by the $\mathcal{H}_L$-Galois property. Also, by Theorem 4.31, the existence of a unit preserving right $\mathcal{H}$-comodule map in part (iii) can be replaced equivalently by the existence of a unit preserving left $\mathcal{H}$-comodule map.

The most interesting part of Theorem 4.33 is perhaps the claim that if property (iii) holds, then right $\mathbf{L}$-module splitting of the lifted canonical map (4.16) implies the $\mathcal{H}_R$-Galois property. The proof of this fact is based on an observation, originally in [55], that assertion (iii) is equivalent to relative separability of the forgetful functor $\mathcal{M}^{\mathcal{H}}\to\mathcal{M}_{\mathbf{L}}$, with respect to the forgetful functor $\mathcal{M}^{\mathcal{H}}_{M}\to\mathcal{M}^{\mathcal{H}}$. A relative separable functor reflects split epimorphisms in the sense that if f is a morphism in $\mathcal{M}^{\mathcal{H}}_{M}$, which is a split epimorphism of right $\mathbf{L}$-modules, then it is a split epimorphism of right $\mathcal{H}$-comodules. This proves that, under the assumptions made, the lifted canonical map (4.16) is a split epimorphism of right $\mathcal{H}$-comodules. Furthermore, the forgetful functor $\mathcal{M}^{\mathcal{H}}_{M}\to\mathcal{M}^{\mathcal{H}}$ possesses a left adjoint $-\otimes_{\mathbf{R}} M$. Hence, the right-right $(M,\mathcal{H})$-relative Hopf module $M\otimes_{\mathbf{R}} H$, which is isomorphic to $H\otimes_{\mathbf{R}} M$ by bijectivity of the antipode, is relative projective in the sense that a split epimorphism g in $\mathcal{M}^{\mathcal{H}}$ of codomain $H\otimes_{\mathbf{R}} M\cong M\otimes_{\mathbf{R}} H$ is a split epimorphism in $\mathcal{M}^{\mathcal{H}}_{M}$. This proves that in the situation considered the lifted canonical map (4.16) is a split epimorphism of right-right $(M,\mathcal{H})$-relative Hopf modules. Then it is a split epimorphism of right-right $(M,\mathcal{H}_R)$-relative Hopf modules. Moreover, in terms of a unit preserving right $\mathcal{H}$-comodule map $H\to M$, a left $\mathbf{N}$-module splitting of the equalizer of the $\mathcal{H}_R$-coaction on M and the map $m\mapsto m\otimes_{\mathbf{R}} 1_{\mathbf{H}}$ can be constructed. This implies that the equalizer is preserved by the functor $M\otimes_{\mathbf{T}} -$, that is, $(M\otimes_{\mathbf{T}} M)^{\mathrm{co}\mathcal{H}_R}=M\otimes_{\mathbf{T}} N$. Thus, the canonical map (4.11) is bijective by [68, Theorem 2.1].

Recall that a left module V of a k-algebra $\mathbf{N}$ is k-relative projective if and only if the left action $N \otimes_k V \to V$ is a split epimorphism of left $\mathbf{N}$-modules. If V has an additional structure of a right comodule for a k-coalgebra C such that the $\mathbf{N}$-action is a right C-comodule map, then it can be asked if the action $N \otimes_k V \to V$ splits as a map of left $\mathbf{N}$-modules and right C-comodules as well. In the case when it does, V is said to be a C-equivariantly projective left $\mathbf{N}$-module. For a Galois extension $\mathbf{N} \subseteq \mathbf{M}$ by a k-Hopf algebra H, H-equivariant projectivity of the left $\mathbf{N}$-module M was shown by Hajac to be equivalent to the existence of a strong connection [73]. That is, interpreting a Hopf Galois extension as a noncommutative principal bundle, equivariant projectivity means its local triviality. In case of a Galois extension $\mathbf{N} \subseteq \mathbf{M}$ by a k-Hopf algebra H, the equivalent conditions (i)–(iii) in Theorem 4.33 are known to imply H-equivariant projectivity of the left $\mathbf{N}$-module M. To obtain an analogous result for a Galois extension by a Hopf algebroid, slightly stronger assumptions are needed, see Theorem 4.35 below.

Definition 4.34. Consider a Hopf algebroid $\mathcal{H}$ and consider a $\mathbf{T}$-ring N for some algebra $\mathbf{T}$. Let V be a left $\mathbf{N}$-module and right $\mathcal{H}$-comodule such that the left $\mathbf{N}$-action on V is a right $\mathcal{H}$-comodule map. V is said to be a $\mathbf{T}$-*relative* $\mathcal{H}$-*equivariantly projective* left $\mathbf{N}$-module provided that the left action $N \otimes_{\mathbf{T}} V \to V$ is an epimorphism split by a left $\mathbf{N}$-module, right $\mathcal{H}$-comodule map.

Theorem 4.35. *Consider a Hopf algebroid $\mathcal{H}$ with a bijective antipode. Let M be a right $\mathcal{H}$-comodule algebra with $\mathcal{H}_R$-coinvariants* $\mathbf{N}$. *Let* $\mathbf{T}$ *be an algebra such that* $\mathbf{N}$ *is a* $\mathbf{T}$*-ring. Assume that there exists a unit preserving right $\mathcal{H}$-comodule map $H \to M$ and that the lifted canonical map* (4.16) *is a split epimorphism of* $\mathbf{L}$*-bimodules. Then* $\mathbf{N} \subseteq \mathbf{M}$ *is a right $\mathcal{H}_R$-, and right $\mathcal{H}_L$-Galois extension and M is a* $\mathbf{T}$*-relative $\mathcal{H}$-equivariantly projective left* $\mathbf{N}$*-module.*

Under the assumptions of Theorem 4.35, the $\mathcal{H}_R$-Galois property holds by Theorem 4.33 and the $\mathcal{H}_L$-Galois property follows by Propositions 4.8 and 4.21. A proof of Theorem 4.35 is completed by constructing the required left $\mathbf{N}$-module right $\mathcal{H}$-comodule section of the action $N \otimes_{\mathbf{T}} M \to M$. The construction makes use of the relative Hopf module section of the lifted canonical map (4.16). The existence of such a section was proved in the paragraph following Theorem 4.33. Examples of $\mathbf{L}$-relative $\mathcal{H}$-equivariantly projective Galois extensions are provided by cleft extensions by a Hopf algebroid $\mathcal{H}$ with a bijective antipode over base algebras $\mathbf{L}$ and $\mathbf{R}$.

4.6. *Alternative notions*

In the literature there is an agreement that the right generalization of a bialgebra to the case of a noncommutative base algebra is a bialgebroid. On the contrary, there is some discussion about the structure that should replace a Hopf algebra. In this current final section we revisit and compare the various suggestions.

4.6.1. Lu Hopf algebroid In Definition 4.1, the antipode axioms are formulated for a compatible pair of a left and a right bialgebroid. In following Definition 4.36, quoted from [6], only a left bialgebroid is used. While the first one of the antipode axioms in Definition 4.1 (iv) is easily formulated also in this case, to formulate the second one, some additional assumption is needed.

DEFINITION 4.36. Consider a left bialgebroid $\mathcal{B}$, over a k-algebra $\mathbf{L}$, with structure maps denoted as in Definition 3.3. $\mathcal{B}$ is a *Lu Hopf algebroid* provided that there exists an antialgebra map $S : \mathbf{B} \to \mathbf{B}$ and a k-module section ξ of the canonical epimorphism $B \otimes_k B \to B \otimes_{\mathbf{L}} B$ such that the following axioms are satisfied.

(i) $S \circ t = s$,
(ii) $\mu_B \circ (S \otimes_{\mathbf{L}} B) \circ \Delta = t \circ \epsilon \circ S$,
(iii) $\mu_{\mathbf{B}} \circ (B \otimes_k S) \circ \xi \circ \Delta = s \circ \epsilon$,

where μ_B denotes multiplication in the $\mathbf{L}$-ring (B, s) and $\mu_{\mathbf{B}}$ denotes multiplication in the underlying k-algebra $\mathbf{B}$.

None of the notions of a Hopf algebroid in Definition 4.1 or 4.36 seem to be more general than the others. Indeed, a Hopf algebroid in the sense of Definition 4.1, which does not satisfy the axioms in Definition 4.36, is constructed as follows. Let k be a commutative ring in which 2 is invertible. For the order 2 cyclic group Z_2, consider the group bialgebra kZ_2 as a left bialgebroid over k. Equip it with the twisted (bijective) antipode S, mapping the order 2 generator t of Z_2 to $S(t) := -t$. Together with the unique right bialgebroid, determined by the requirement that S is a bialgebroid anti-isomorphism in the sense of Proposition 4.4, they constitute a Hopf algebroid as in Definition 4.1. However, for this Hopf algebroid there exists no section ξ as in Definition 4.36.

4.6.2. On $\times_L$-Hopf algebras The coinvariants of the (left or right) regular comodule of a bialgebra H over a commutative ring k are precisely the multiples of the unit element $1_{\mathbf{H}}$. H is known to be a Hopf algebra if and only if $\mathbf{H}$ is an H-Galois extension of k. Indeed, the hom-tensor relation ${}_H\mathrm{Hom}^H(H \otimes_k H, H \otimes_k H) \cong \mathrm{Hom}_k(H, H)$ relates the inverse of the canonical map to the antipode. Motivated by this characterization of a Hopf algebra, in [33] Schauenburg proposed the following definition.

DEFINITION 4.37. Let $\mathcal{B}$ be a left bialgebroid over an algebra $\mathbf{L}$, with structure maps denoted as in Definition 3.3. Consider the left regular $\mathcal{B}$-comodule, whose coinvariant subalgebra is $t(\mathbf{L}^{\mathrm{op}})$. $\mathcal{B}$ is said to be a *$\times_L$-Hopf algebra* provided that the algebra extension $t : \mathbf{L}^{\mathrm{op}} \to \mathbf{B}$ is left $\mathcal{B}$-Galois.

The notion of a $\times_L$-Hopf algebra in Definition 4.37 is more general than that of a Hopf algebroid as in Definition 4.1. Indeed, consider a Hopf algebroid $(\mathcal{H}_L, \mathcal{H}_R, S)$, over the base algebras $\mathbf{L}$ and $\mathbf{R}$, with structure maps denoted as in Definition 4.1. The canonical map $H \otimes_{\mathbf{L}^{\mathrm{op}}} H \to H \otimes_{\mathbf{L}} H$, $h \otimes h' \mapsto h_{(1)} \otimes h_{(2)}h'$ is bijective, with inverse $h \otimes h' \mapsto h^{(1)} \otimes S(h^{(2)})h'$. Hence $\mathcal{H}_L$ is a $\times_L$-Hopf algebra. An example

of a $\times_L$-Hopf algebra, that admits no Hopf algebroid structure, was proposed in [76, Example 8].

Extending a result [74] by Schauenburg about Hopf algebras, the following proposition was proved in [44].

PROPOSITION 4.38. *Consider a left bialgebroid $\mathcal{B}$ over a base algebra $\mathbf{L}$. If there is a left $\mathcal{B}$-Galois extension $\mathbf{N} \subseteq \mathbf{M}$ such that M is a faithfully flat left $\mathbf{L}$-module, then $\mathcal{B}$ is a $\times_L$-Hopf algebra.*

Indeed, the canonical maps $\mathrm{can} : M \otimes_{\mathbf{N}} M \to B \otimes_{\mathbf{L}} M$ and $\vartheta : B \otimes_{\mathbf{L}^{\mathrm{op}}} B \to B \otimes_{\mathbf{L}} B$ satisfy the pentagonal identity $(B \otimes_{\mathbf{L}} \mathrm{can}) \circ (\mathrm{can} \otimes_{\mathbf{N}} M) = (\vartheta \otimes_{\mathbf{L}} M) \circ \mathrm{can}_{13} \circ (M \otimes_{\mathbf{N}} \mathrm{can})$, where the (well defined) map $\mathrm{can}_{13} : M \otimes_{\mathbf{N}} (B \otimes_{\mathbf{L}} M) \to (B \otimes_{\mathbf{L}^{\mathrm{op}}} B) \otimes_{\mathbf{L}} M$ is obtained by applying 'can' to the first and third factors.

In Theorem 3.13, bialgebroids were characterized via strict monoidality of a forgetful functor. A characterization of a similar flavor for $\times_L$-Hopf algebras was given in [33, Theorem and Definition 3.5]. Recall that a monoidal category $(\mathcal{M}, \otimes, U)$ is said to be *right closed* if the endofunctor $- \otimes X$ on $\mathcal{M}$ possesses a right adjoint denoted by $\hom(X, -)$ for any object X in $\mathcal{M}$. The monoidal category of bimodules of an algebra $\mathbf{L}$ is right closed with $\hom(X, Y) = \mathrm{Hom}_{\mathbf{L}}(X, Y)$. It is slightly more involved to see that so is the category of left $\mathbf{B}$-modules, for a left $\mathbf{L}$-bialgebroid $\mathcal{B}$, with $\hom(X, Y) = {}_{\mathbf{B}}\mathrm{Hom}(B \otimes_{\mathbf{L}} X, Y)$ (where for a left $\mathbf{B}$-module X, $B \otimes_{\mathbf{L}} X$ is a left $\mathbf{B}$-module via the diagonal action and a right $\mathbf{B}$-module via the first factor). A strict monoidal functor $\mathbb{F} : \mathcal{M} \to \mathcal{M}'$ between right-closed categories is called *strong right closed* provided that the canonical morphism $\mathbb{F}(\hom(X, Y)) \to \hom(\mathbb{F}(X), \mathbb{F}(Y))$ is an isomorphism, for all objects $X, Y \in \mathcal{M}$.

THEOREM 4.39. *A left bialgebroid $\mathcal{B}$ over a base algebra $\mathbf{L}$ is a $\times_L$-Hopf algebra if and only if the (strict monoidal) forgetful functor ${}_{\mathbf{B}}\mathcal{M} \to {}_{\mathbf{L}}\mathcal{M}_{\mathbf{L}}$ is strong right closed.*

For a Hopf algebra H over a commutative ring k, those (left or right) H-modules, which are finitely generated and projective k-modules, possess (right or left) duals in the monoidal category of (left or right) H-modules. This property extends to $\times_L$-Hopf algebras (hence to Hopf algebroids!) as follows.

PROPOSITION 4.40. *Let $\mathcal{B}$ be a $\times_{\mathbf{L}}$-Hopf algebra over a base algebra $\mathbf{L}$. Denote its structure maps as in Definition 3.3. For a left $\mathbf{B}$-module M, the dual $M^* := \mathrm{Hom}_{\mathbf{L}}(M, L)$ is a left $\mathbf{B}$-module via the action*

$$(b \cdot \phi)(m) := \epsilon\big(b^{\langle 1 \rangle} t(\phi(b^{\langle 2 \rangle} \cdot m))\big), \qquad \text{for } b \in B,\ \phi \in M^*,\ m \in M,$$

where for the inverse of the (right $\mathbf{B}$-linear) canonical map $B \otimes_{\mathbf{L}^{\mathrm{op}}} B \to B \otimes_{\mathbf{L}} B$, the index notation $b \otimes b' \mapsto b^{\langle 1 \rangle} \otimes b^{\langle 2 \rangle} b'$ is used. Furthermore, if M is a finitely generated and projective right $\mathbf{L}$-module, then M^* is a right dual of M in the monoidal category ${}_{\mathbf{B}}\mathcal{M}$.

4.6.3. Hopf monad In Theorem 3.5, bialgebroids were related to bimonads on a bimodule category that possess a right adjoint. In the paper [39], special bimonads, the so-called Hopf monads, on autonomous categories were studied. Recall that a monoidal category is said to be left (resp. right) autonomous provided that every object possesses a left (resp. right) dual. In particular, a category of finitely generated and projective bimodules is autonomous. Instead of the somewhat technical definition in [39, 3.3], we adopt an equivalent description from [39, Theorem 3.8] as a definition.

DEFINITION 4.41. A *left (resp. right) Hopf monad* is a bimonad $\mathbb{B}$ on a left (resp. right) autonomous monoidal category $\mathcal{M}$ such that the left (resp. right) autonomous structure of $\mathcal{M}$ lifts to the category of $\mathbb{B}$-algebras.

The reader should be warned that although the same term *Hopf monad* is used in the papers [39] and [26], they have different meanings (and a further totally different meaning of the same term is used in [20]). Also, the notions of a comodule and a corresponding (co)integral in [39] are different from the notions used in these notes.

4.6.4. A *-autonomous structure on a strong monoidal special opmorphism between pseudomonoids in a monoidal bicategory In the paper [28], strong monoidal special opmorphisms h in monoidal bicategories, from a canonical pseudomonoid $R^{\mathrm{op}} \otimes R$ to some pseudomonoid B, were studied. The opmorphism h was called *Hopf* if in addition there is a *-autonomous structure on B and h is strong *-autonomous (where $R^{\mathrm{op}} \otimes R$ is meant to be *-autonomous in a canonical way). In [28, Section 3], a bialgebroid was described as a strong monoidal special opmorphism h of pseudomonoids in the monoidal bicategory of `[Algebras; Bimodules; Bimodule maps]`. This opmorphism h is strong *-autonomous if and only if the corresponding bialgebroid constitutes a Hopf algebroid with a bijective antipode (see [46, Section 4.2]).

Acknowledgment

The author's work is supported by the Bolyai János Fellowship and the Hungarian Scientific Research Fund OTKA F67910. She is grateful for their helpful comments to Dénes Bajnok, Imre Bálint, Tomasz Brzeziński, and Kornél Szlachányi.

References

[1] J. Morava, *Noetherian localisations of categories of cobordism comodules*, Ann. of Math. **121** (2) (1985) 1–39.
[2] J. Mrčun, *The Hopf algebroids of functions on étale groupoids and their principal Morita equivalence*, J. Pure Appl. Algebra **160** (2001) 249–262.
[3] J. Donin, A. Mudrov, *Quantum groupoids and dynamical categories*, J. Algebra **296** (2006) 348–384.
[4] P. Etingof, D. Nikshych, *Dynamical quantum groups at roots of 1*, Duke Math J. **108** (2001) 135–168.
[5] G. Karaali, *On Hopf Algebras and their Generalizations*, preprint arXiv:math.QA/0703441.

[6] J.H. Lu, *Hopf algebroids and quantum groupoids*, Int. J. Math. **7** (1996) 47–70.

[7] D. Nikshych, L. Vainerman, *Finite quantum groupoids and their applications*, in: "*New directions in Hopf algebras*," Math. Sci. Res. Inst. Publ., **43**, Cambridge Univ. Press, Cambridge, 2002, pp. 211–262.

[8] P. Xu, *Quantum groupoids*, Comm. Math. Phys. **216** (2001) 539–581.

[9] D. Nikshych, V. Turaev, L. Vainerman, *Invariants of knots and 3-manifolds from quantum groupoids*, Proceedings of the Pacific Institute for the Mathematical Sciences Workshop "Invariants of Three-Manifolds" (Calgary, AB, 1999). Topology Appl. **127** (2003) 91–123.

[10] A. Connes, H. Moscovici, *Rankin-Cohen brackets and the Hopf algebra of transverse geometry*, Mosc. Math. J. **4** (2004) 111–130.

[11] G. Böhm, K. Szlachányi, *Weak Hopf algebras II: Representation theory, dimensions, and the Markov trace*, J. Algebra **233** (2000) 156–212.

[12] I. Bálint, K. Szlachányi, *Finitary Galois extensions over noncommutative bases*, J. Algebra **296** (2006) 520–560.

[13] L. Kadison, *Galois theory for bialgebroids, depth two and normal Hopf subalgebras*, Ann. Univ. Ferrara Sez. VII (NS) **51** (2005) 209–231.

[14] M. Cohen, S. Gelaki, S. Westreich, *Hopf Algebras*, in: M. Hazewinkel (Ed.), Handbook of Algebra. vol. **4**, Elsevier, 2006, pp. 173–239.

[15] T. Brzeziński, R. Wisbauer, *Corings and comodules,* Cambridge University Press, Cambridge 2003. Erratum: `http://www-maths.swan.ac.uk/staff/tb/Corings.htm`.

[16] T. Brzeziński, *The structure of corings: Induction functors, Maschke-type theorem, and Frobenius and Galois-type properties*, Algebr. Represent. Theory **5** (2002) 389–410.

[17] S. Majid, *Algebras and Hopf algebras in braided categories*, in: Advances in Hopf algebras (Chicago, IL, 1992), pp. 55–105, Lecture Notes in Pure and Appl. Math., vol. **158**, Dekker, New York, 1994.

[18] P. Schauenburg, *On the braiding of a Hopf algebra in a braided category*, New York J. Math. **4** (1998) 259–263.

[19] M. Takeuchi, *Survey of braided Hopf algebras*, in: *New trends in Hopf algebra theory* (La Falda, 1999), in: Contemp. Math. vol. **267**, Amer. Math. Soc., Providence, RI, 2000, pp. 301–323.

[20] B. Mesablishvili, R. Wisbauer, *Bimonads and Hopf monads on categories*, preprint arXiv:0710.1163.

[21] M. Takeuchi, *Groups of algebras over* $A \otimes \overline{A}$, J. Math. Soc. Jpn. **29** (1977) 459–492.

[22] T. Brzeziński, G. Militaru, *Bialgebroids,* $\times_A$*-bialgebras and duality*, J. Algebra **251** (2002) 279–294.

[23] L. Kadison, K. Szlachányi, *Bialgebroid actions on depth two extensions and duality*, Adv. Math. **179** (2003) 75–121.

[24] M. Takeuchi, $\sqrt{\textit{Morita}}$ *theory - formal ring laws and monoidal equivalences of categories of bimodules*, J. Math. Soc. Japan **39** (1987) 301–336.

[25] K. Szlachányi, *Monoidal Morita equivalence*, AMS Contemp. Math. vol. **391**, AMS, Providence, RI, 2005, pp. 353–369.

[26] I. Moerdijk, *Monads on tensor categories*, J. Pure Appl. Algebra **168** (2002) 189–208.

[27] K. Szlachányi, *The monoidal Eilenberg-Moore construction and bialgebroids*, J. Pure Appl. Algebra **182** (2003) 287–315.

[28] B. Day, R. Street, *Quantum category, star autonomy and quantum groupoids*, Fields Inst. Comm. vol. **43**, AMS, Providence, RI, 2004, pp. 187–225.

[29] G. Böhm, F. Nill, K. Szlachányi, *Weak Hopf algebras I: Integral theory and* C^**-structure*, J. Algebra **221** (1999) 385–438.

[30] F. Nill, *Weak bialgebras*, Preprint arXiv:math.QA/9805104.

[31] P. Schauenburg, *Weak Hopf algebras and quantum groupoids*, Banach Center Publications **61**, Polish Acad. Sci., Warsaw, 2003, pp. 171–188.

[32] K. Szlachányi, *Finite quantum groupoids and inclusions of finite type*, Fields Inst. Comm. vol. **30**, AMS, Providence, RI, 2001, pp. 393–407.

[33] P. Schauenburg, *Duals and doubles of quantum groupoids (*$\times_R$*-Hopf algebras)*, AMS Contemp. Math. vol. **267**, AMS, Providence, RI, 2000, pp. 273–299.

[34] P. Schauenburg, *Morita base change in quantum groupoids*, in: *Locally compact quantum groups and groupoids* (Strasbourg, 2002), IRMA Lect. Math. Theor. Phys. vol. **2**, de Gruyter, Berlin, 2003, pp. 79–103.
[35] B. Pareigis, *A noncommutative noncocommutative Hopf algebra in "nature,"* J. Algebra **70** (1981) 356–374.
[36] P. Schauenburg, *Bialgebras over noncommutative rings and a structure theorem for Hopf bimodules*, Appl. Categor. Struct. **6** (1998) 193–222.
[37] I. Bálint, *Scalar extension of bicoalgebroids*, Appl. Categ. Str. **16** (2008) 29–55.
[38] I. Bálint, *Quantum groupoid symmetry in low dimensional algebraic quantum field theory*, PhD thesis, Eötvös University Budapest, 2008.
[39] A. Bruguiéres, A. Virelizier, *Hopf monads*, Adv. Math. **215** (2007) 679–733.
[40] C. Grunspan, *Quantum torsors*, J. Pure Appl. Algebra **184** (2003) 229–255.
[41] P. Schauenburg, *Hopf bi-Galois extensions*, Comm. Algebra **24** (1996) 3797–3825.
[42] P. Schauenburg, *Quantum torsors with fewer axioms*, Preprint arXiv math.QA/0302003.
[43] P. Schauenburg, *Hopf-Galois and bi-Galois extensions*, in: Galois theory, Hopf algebras, and semiabelian categories, Fields Inst. Commun., vol. **43**, Amer. Math. Soc., Providence, RI, 2004, pp. 469–515.
[44] D. Hobst, *Antipodes in the theory of noncommutative torsors*. PhD thesis, Ludwig-Maximilians Universität München 2004, Logos Verlag Berlin, 2004.
[45] G. Böhm, T. Brzeziński, *Pre-torsors and equivalences*, J. Algebra **317** (2007) 544–580, *Corrigendum*, J. Algebra **319** (2008) 1339–1340.
[46] G. Böhm, K. Szlachányi, *Hopf algebroids with bijective antipodes: axioms integrals and duals*, J. Algebra **274** (2004) 708–750.
[47] G. Böhm, *Integral theory of Hopf algebroids*, Algebra Represent. Theory **8** (2005) 563–599 *Corrigendum*, to be published.
[48] G. Böhm, T. Brzeziński, *Cleft extensions of Hopf algebroids*, Appl. Categ. Struct. **14** (2006) 431–469 *Corrigendum*, DOI 10.1007/s10485-008-9168-x.
[49] A. Connes, H. Moscovici, *Cyclic cohomology and Hopf algebras*, Lett. Math. Phys. **48** (1999) 97–108.
[50] G. Böhm, *An alternative notion of Hopf algebroid*, in: S. Caenepeel, F. Van Oystaeyen (Eds.), *Hopf algebras in noncommutative geometry and physics*. Lecture Notes in Pure and Appl. Math. vol. **239**, Dekker, New York, 2005, pp. 31–53.
[51] T. Hayashi, *Face algebras and tensor categories* (Japanese), in: "Representation theory and noncommutative harmonic analysis" pp. 34–47, Kyoto, 1998. An English version is the preprint T. Hayashi, *A canonical Tannaka duality for finite semisimple tensor categories*, arXiv:math.QA/9904073.
[52] P. Etingof, D. Nikshych, V. Ostrik, *On fusion categories*, Ann. Math. **162** (2) (2005) 581–642.
[53] T. Brzeziński, *A note on coring extensions*, Ann. Univ. Ferrara Sez. VII (NS) **51** (2005) 15–27. A corrected version is available at arXiv:math/0410020v3.
[54] G. Böhm, *Galois theory for Hopf algebroids*, Ann. Univ. Ferrara Sez. VII (NS) **51** (2005) 233–262. A corrected version is available at arXiv:math/0409513v3.
[55] A. Ardizzoni, G. Böhm, C. Menini, *A Schneider type theorem for Hopf algebroids*, J. Algebra **318** (2007) 225–269 *Corrigendum*, J. Algebra **321** (2009) 1786–1796.
[56] G. Böhm, J. Vercruysse, *Morita theory for coring extensions and cleft bicomodules*, Adv. Math. **209** (2007) 611–648 *Corrigendum*, DOI 10.1016/j.aim.2008.11.018.
[57] S. Caenepeel, G. Militaru, *Maschke functors, semisimple functors and separable functors of the second kind: applications*, J. Pure Appl. Algebra **178** (2003) 131–157.
[58] C. Lomp, *Integrals in Hopf algebras over rings*, Commun. Algebra **32** (2004) 4687–4711.
[59] R.G. Larson, M.E. Sweedler, *An associative orthogonal bilinear form for Hopf algebras*, Amer. J. Math. **91** (1969) 75–94.
[60] B. Pareigis, *When Hopf algebras are Frobenius algebras*, J. Algebra **18** (1971) 588–596.
[61] K. Hirata, K. Sugano, *On semisimple extensions and separable extensions over non commutative rings*, J. Math. Soc. Jpn. **18** (1966) 360–373.

[62] L. Kadison, *New examples of Frobenius extensions*, University Lecture Series vol. **14**, American Mathematical Society, Providence, RI, 1999. x+84 pp.
[63] K. Szlachányi, *The double algebraic view of finite quantum groupoids*, J. Algebra **280** (2004) 249–294.
[64] Y. Doi, M. Takeuchi, *Cleft comodule algebras for a bialgebra*, Commun. Algebra **14** (1986) 801–817.
[65] R. Blattner, M. Cohen, S. Montgomery, *Crossed products and inner actions of Hopf algebras*, Trans. Amer. Math. Soc. **298** (1986) 671–711.
[66] H.F. Kreimer, M. Takeuchi, *Hopf algebras and Galois extensions of an algebra*, Indiana Univ. Math. J. **30** (1981) 675–692.
[67] H-J. Schneider, *Principal homogeneous spaces for arbitrary Hopf algebras*, Isr. J. Math. **72** (1990) 167–195.
[68] T. Brzeziński, R.B. Turner, A.P. Wrightson, *The structure of weak coalgebra Galois extensions*, Commun. Algebra **34** (2006) 1489–1519.
[69] S. Caenepeel, J. Vercrusse, S. Wang, *Morita theory for corings and cleft entwining structures*, J. Algebra **276** (2004) 210–235.
[70] Y. Doi, *Algebras with total integrals*, Commun. Algebra **13** (1985) 2137–2159.
[71] R. Guitart, J. Riguet, *Enveloppe Karoubienne des catégories de Kleisli*, Cah. Topologie Géom. Différ. Catég. **33** (1992) 261–266.
[72] S. Caenepeel, *Galois corings from the descent theory point of view*, Fields. Inst. Comm. vol. **43**, AMS, Providence, RI, 2004, pp. 163–186.
[73] P.M. Hajac, *Strong connections on quantum principal bundles*, Commun. Math. Phys. **182** (1996) 579–617.
[74] P. Schauenburg, *A bialgebra that admits a Hopf-Galois extension is a Hopf algebra*, Proc. Amer. Math. Soc. **125** (1997) 83–85.
[75] Phung Ho Hai, *Tannaka-Krein duality for Hopf algebroids*, Israel J. Math. **167** (2008) 193–225.
[76] U. Krähmer and N. Kowalzig, *Duality and products in algebraic (co) homology theories*, Preprint arXiv:0812.4312v1.

Comodules and Corings

Tomasz Brzeziński

Department of Mathematics, Swansea University, Singleton Park, Swansea SA2 8PP, U.K.
E-mail: T.Brzezinski@swansea.ac.uk

Contents

HANDBOOK OF ALGEBRA, VOL. 6
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(08)00206-4

1. Introduction

A coring is one of the most basic algebraic structures (formally) dual to that of a ring (but also a ring itself can be viewed as a coring). The notion of a *coring* appeared first in the algebra literature in 1975 in Sweedler's paper [178] on a predual version of the Jacobson–Bourbaki correspondence. As corings are coalgebras in the category of bimodules of an algebra, they were implicitly studied earlier, for example, in [115]. Also, in the late 1970s, they appeared under the name BOCS in representation theory and were used in studies of matrix problems in works of Drozd, Kleiner, and Roiter [161]. Between the late 1970s and the end of the 1990s corings attracted rather limited attention, with, to the best of our knowledge, only a handful of papers [106, 122, 134, 135, 183, 184] and a monograph [20] published.

In the meantime, there was a resurgence of interest in Hopf algebras, triggered by the appearance of quantum groups as symmetries of integrable systems in quantum and statistical mechanics in works of Drinfeld and Jimbo. This led to intensive studies of Hopf algebras also from a purely algebraic point of view and to the development of more general categories of Hopf-type modules. These serve as representations of Hopf algebras and related structures, such as those described by solutions to braid or Yang-Baxter equations. By the end of the 1990s, all such Hopf-type modules were unified in terms of *entwined modules* associated to *entwining structures* (cf. [36, 49]). It was an observation of Takeuchi (reported in [136]) that entwined modules are comodules of a certain coring that instigated renewal of interest in corings in Hopf algebra theory. This allowed one to view more conceptually and directly a connection between properties of various types of Hopf modules [37]. More or less at the same time corings appeared as one of the main structures in an approach to noncommutative algebraic geometry advocated by Kontsevich and Rosenberg [124].

The aim of this chapter is two-fold. First, we would like to give a summary of basic definitions and basic properties of corings and comodules, focussing on recent developments. Rather than giving a complete picture, we give an overview of the coring and comodule theory, describe basic examples, and discuss applications which prompted recent resurgence of interest in this theory. Since a coalgebra is an example of a coring, a summary of properties of comodules of a coring provides also the reader with the knowledge of properties of comodules of a coalgebra. Furthermore, many types of comodules of coalgebras considered in Hopf algebra theory (Hopf-type modules) turn out to be comodules of specific corings. In this way, this chapter should supplement [140], which is our second aim.

For fuller and more detailed description of corings and comodules, the reader should consult [53]. Inevitably, there is an overlap between this chapter and [53], in particular, in these parts which deal with a pure or "older" coring theory. On the other hand, we have tried to include as much as possible results, which have been obtained or about which we have learnt since [53] was written. The reader should be aware that since coring theory went through a significant transformation over the last few years and is still being intensively developed, this chapter is a "state-of-the-art" presentation of these developments rather than a definitive account. It is still

too early to determine the significance of particular developments. The selection of topics reflects the author's taste more than their actual "objective" value (whatever this might be).

2. Categorical preliminaries

Throughout this chapter we use the language of category theory. On the other hand, to understand the main elements of the theory developed in this chapter, very little knowledge of categorical language is needed. The aim of this section is to introduce the reader who is not familiar with this language to the basic concepts and notions. We hope that in this way such a reader can freely read the later parts of this chapter. The reader familiar with category theory can skip this section, and only come back to it when notation or conventions used later on will raise doubts. A standard reference for category theory is [132].

2.1. Categories. A category $\mathbf{A}$ consists of a class of *objects*, and, for any two objects A, B, a set of *morphisms* or *arrows* from A to B, denoted by $\mathbf{A}(A, B)$. The sets of arrows satisfy the following axioms:

(a) $\mathbf{A}(A, B) \cap \mathbf{A}(C, D) = \emptyset$, if either $A \neq C$ or $B \neq D$.
(b) For any objects A, B, C and any morphisms $f \in \mathbf{A}(A, B)$ and $g \in \mathbf{A}(B, C)$, there exists a unique morphism $g \circ f \in \mathbf{A}(A, C)$ known as a *composition of f with g* and such that for any object D and any $h \in \mathbf{A}(C, D)$, $(h \circ g) \circ f = h \circ (g \circ f)$.
(c) For any object A, there is a unique *identity morphism* $\mathrm{id}_A \in \mathbf{A}(A, A)$ such that, for any object B and any $f \in \mathbf{A}(A, B)$, $g \in \mathbf{A}(B, A)$, $\mathrm{id}_A \circ g = g$, $f \circ \mathrm{id}_A = f$.

In this chapter, we adopt the convention whereby for any object A the identity morphism id_A is denoted by A. For any $f \in \mathbf{A}(A, B)$, we write $f : A \to B$. A composition is also denoted as a sequence of arrows $A \to B \to C$.

The *opposite category* $\mathbf{A}^{\mathrm{op}}$ of $\mathbf{A}$ is a category with the same collection of objects and arrows as $\mathbf{A}$, but any arrow $A \to B$ in $\mathbf{A}$ is understood as an arrow $B \to A$, that is, $\mathbf{A}^{\mathrm{op}}(A, B) = \mathbf{A}(B, A)$.

Given two categories $\mathbf{A}$ and $\mathbf{B}$, the *product category* $\mathbf{A} \times \mathbf{B}$ has ordered pairs (A, B), with A an object in $\mathbf{A}$ and B an object in $\mathbf{B}$, as objects and with morphism sets:

$$\mathbf{A} \times \mathbf{B}((A, B), (A', B')) = \mathbf{A}(A, A') \times \mathbf{B}(B, B').$$

A *subcategory* $\mathbf{B}$ of $\mathbf{A}$ is a category whose objects are a subclass of objects of $\mathbf{A}$, and, for each pair of objects in $\mathbf{B}$, the morphism set is a subset of the morphism set of corresponding objects in $\mathbf{A}$. A subcategory $\mathbf{B}$ of $\mathbf{A}$ is said to be *full* if $\mathbf{B}(A, B) = \mathbf{A}(A, B)$, for all objects A, B of $\mathbf{B}$.

If k is a commutative ring with unit, by a *k-linear category*, we understand a category $\mathbf{A}$ in which all morphism sets are k-modules and composition of morphisms is k-linear. In case $k = \mathbb{Z}$, a k-linear category is called a *preadditive* category. All the categories described in the main body of this chapter are k-linear.

2.2. Examples of categories. Here are some elementary examples of categories:

(1) **Set** (the category of sets): objects are sets and morphisms are functions.
(2) ${}_R\mathbf{M}$, $\mathbf{M}_R$, ${}_R\mathbf{M}_S$ (the category of left, resp. right, R-modules, over a ring R, resp. (R, S)-bimodules over rings R and S): objects are left, resp. right, R-modules, resp. (R, S)-bimodules, morphisms are R-module, resp. (R, S)-bimodule maps. Morphisms in these categories are denoted by ${}_R\mathrm{Hom}(-, -)$, resp. $\mathrm{Hom}_R(-, -)$, and ${}_R\mathrm{Hom}_S(-, -)$.
(3) Any monoid G can be understood as a category $\mathbf{G}$ with a single object $*$. The set of morphisms $\mathbf{G}(*, *)$ equals G; the composition of morphisms is simply the multiplication in G. Similarly, any associative, unital k-algebra can be understood as a k-linear category with one object.

2.3. Types of morphisms. A morphism $f : A \to B$ in $\mathbf{A}$ is called

(a) a *monomorphism* if, for all $g, h \in \mathbf{A}(C, A)$, $f \circ g = f \circ h$ implies $g = h$;
(b) an *epimorphism* if, for all $g, h \in \mathbf{A}(B, D)$, $g \circ f = h \circ f$ implies $g = h$;
(c) a *retraction* if there exists $g \in \mathbf{A}(B, A)$ with $f \circ g = B$; a retraction is necessarily an epimorphism and hence is also called a *split epimorphism*;
(d) a *section* if there exists $g \in \mathbf{A}(B, A)$ with $g \circ f = A$; a section is necessarily a monomorphism and hence is also called a *split monomorphism*;
(e) an *isomorphism* if f is both a retraction and a section. The splitting morphism g as in (c) and (d) is called the *inverse* of f and denoted by f^{-1}.

2.4. Functors. A *covariant functor* or simply a *functor* F from $\mathbf{A}$ to $\mathbf{B}$ is a map assigning to each object A of $\mathbf{A}$ an object $F(A)$ of $\mathbf{B}$ and to any arrow $f \in \mathbf{A}(A, B)$ an arrow $F(f) \in \mathbf{B}(F(A), F(B))$ such that

(a) for any object A of $\mathbf{A}$, the morphism $F(\mathrm{id}_A)$ is the identity morphism for $F(A)$;
(b) for any $f \in \mathbf{A}(A, B)$, $g \in \mathbf{A}(B, C)$, $F(g \circ f) = F(g) \circ F(f)$.

A *contravariant functor* F from $\mathbf{A}$ to $\mathbf{B}$ means a functor from $\mathbf{A}^{\mathrm{op}}$ to $\mathbf{B}$.

A functor $F : \mathbf{A} \to \mathbf{B}$ between k-linear categories is assumed to be a k-linear operation in the sense that for any two objects A and B in $\mathbf{A}$, $\mathbf{A}(A, B) \to \mathbf{B}(F(A), F(B))$ is a homomorphism of k-modules. In case $k = \mathbb{Z}$, a k-linear functor is called an *additive functor*.

Main examples of functors include the identity functor $\mathbf{A} : \mathbf{A} \to \mathbf{A}$, which sends objects A to A and morphisms f to f, and two hom-functors, defined for any object A of $\mathbf{A}$ as follows. The covariant hom-functor

$$\mathbf{A}(A, -) : \mathbf{A} \to \mathbf{Set}, \quad B \mapsto \mathbf{A}(A, B), \quad \mathbf{A}(A, f) : u \mapsto f \circ u,$$

and the contravariant hom-functor

$$\mathbf{A}(-, A) : \mathbf{A}^{\mathrm{op}} \to \mathbf{Set}, \quad B \to \mathbf{A}(B, A), \quad \mathbf{A}(f, A) : u \mapsto u \circ f.$$

2.5. Full and faithful functors. A functor is said to be *faithful* if it is injective when restricted to each set of morphisms and is said to be *full* if it is surjective when

restricted to each set of morphisms. A full and faithful functor is referred to as a *fully faithful* functor.

2.6. Tensor (induction) functors. Given rings R and S, and an (R, S)-bimodule M, one can define tensor functors $- \otimes_R M : \mathbf{M}_R \to \mathbf{M}_S$ and $M \otimes_S - : {}_S\mathbf{M} \to {}_R\mathbf{M}$. On objects $N \in \mathbf{M}_R$. $N \otimes_R M$ is a right S-module through the right multiplication in M, $(n \otimes_R m)s = n \otimes_R ms$. For any morphism $f : N \to N'$ in $\mathbf{M}_R$, $f \otimes_R M$ is a morphism in $\mathbf{M}_S$ by $n \otimes_R m \mapsto f(n) \otimes_R m$. The other tensor functor is defined symmetrically.

2.7. Forgetful functors. A forgetful functor is a functor that "forgets" part of the structure of objects. For example, since every right R-module is a set, there is a functor from $\mathbf{M}_R \to \mathbf{Set}$, which views modules as sets and module maps as functions. If A is a k-algebra, then there is a forgetful functor from the category $\mathbf{M}_A$ of right A-modules to the category of k-modules, which views objects in $\mathbf{M}_A$ purely as k-modules.

2.8. Natural transformations. Given functors $F, G : \mathbf{A} \to \mathbf{B}$, a *natural transformation* $\varphi : F \to G$ is a collection of morphisms in $\mathbf{B}$, $\varphi_A : F(A) \to G(A)$ labeled by all objects $A \in \mathbf{A}$ and such that for any morphism $f : A \to B$,

$$\varphi_B \circ F(f) = G(f) \circ \varphi_A.$$

In case all the φ_A are isomorphisms, one calls φ a *natural isomorphism*.

Functors from $\mathbf{A}$ to $\mathbf{B}$ and natural transformations form a category denoted by $\mathbf{Fun}(\mathbf{A}, \mathbf{B})$.

2.9. Adjoint functors. A pair of functors $L : \mathbf{A} \to \mathbf{B}$, $R : \mathbf{B} \to \mathbf{A}$ is said to be an *adjoint pair* provided that there exists a pair of natural transformations, $\eta : \mathrm{id}_{\mathbf{A}} \to RL$ and $\psi : LR \to \mathrm{id}_{\mathbf{B}}$ such that

$$\psi_{L(A)} \circ L(\eta_A) = L(A), \qquad R(\psi_B) \circ \eta_{R(B)} = R(B).$$

L is called a *left adjoint* of R and R is called a *right adjoint* of L. The transformation η is called a *unit* of the adjunction and ψ is termed a *counit* of the adjunction.

L, R is a pair of adjoint functors if and only if there is a natural isomorphism

$$\Omega : \mathbf{B}(L(-), -) \to \mathbf{A}(-, R(-))$$

of functors $\mathbf{A}^{\mathrm{op}} \times \mathbf{B} \to \mathbf{Set}$. In terms of the unit and counit of the adjunction, Ω is given by

$$\Omega_{A,B} = \mathbf{B}(L(A), B) \to \mathbf{A}(A, R(B)), \quad g \mapsto R(g) \circ \eta_A,$$

and the inverse of Ω is $\Omega^{-1}_{A,B} : \mathbf{A}(A, R(B)) \to \mathbf{B}(L(A), B)$, $f \mapsto \psi_B \circ L(f)$.

In an adjoint pair (L, R), the functor L sends epimorphisms to epimorphisms (i.e. L *preserves epimorphisms*), while R sends monomorphisms to monomorphisms (i.e. R *preserves monomorphisms*). Furthermore, L is fully faithful if and only if the unit of the adjunction is a natural isomorphism, while R is fully faithful if and only if the counit is an isomorphism.

The hom-tensor adjunction is one of the key adjoint pairs of functors in this chapter. Take R, S to be rings and M to be an (R, S)-bimodule. Then, the tensor functor $- \otimes_R M : \mathbf{M}_R \to \mathbf{M}_S$ (cf. 2.6) is the left adjoint of the hom-functor $\mathrm{Hom}_S(M, -) : \mathbf{M}_S \to \mathbf{M}_R$. Here, $\mathrm{Hom}_S(M, V)$ is a right R-module by $(fr)(m) = f(rm)$, for all $f \in \mathrm{Hom}_S(M, V)$, $m \in M$, and $r \in R$. The counit and unit of this adjunction are

$$\psi_V : \mathrm{Hom}_S(M, V) \otimes_R M \to V, \quad f \otimes_R m \mapsto f(m),$$
$$\eta_X : X \to \mathrm{Hom}_S(M, X \otimes_R M), \quad x \mapsto [m \mapsto x \otimes_R m].$$

The isomorphisms $\Omega_{X,V} : \mathrm{Hom}_S(X \otimes_R M, V) \to \mathrm{Hom}_R(X, \mathrm{Hom}_S(M, V))$ come out as

$$\Omega_{X,V}(g)(x) : m \mapsto g(x \otimes_R m),$$

for all right R-modules X, right S-modules V, $g \in \mathrm{Hom}_S(X \otimes_R M, V)$, $x \in X$, and $m \in M$.

2.10. Equivalence of categories. An adjoint pair of functors $L : \mathbf{A} \to \mathbf{B}$, $R : \mathbf{B} \to \mathbf{A}$ with counit and unit, which are natural isomorphisms, is called a pair of *inverse equivalences*. Each of L and R is called an *equivalence*, and the categories $\mathbf{A}$ and $\mathbf{B}$ are said to be *equivalent*.

2.11. Comonads and monads. Let $\mathbf{A}$ be a category. A functor $G : \mathbf{A} \to \mathbf{A}$ is called a *comonad* if there exist a natural transformation $\delta : G \to GG$ and a natural transformation $\sigma : G \to \mathrm{id}_{\mathbf{A}}$ such that for all objects $A \in \mathbf{A}$, the following diagrams

$$\begin{array}{ccc} G(A) & \xrightarrow{\delta_A} & GG(A) \\ {\scriptstyle \delta_A}\downarrow & & \downarrow{\scriptstyle G(\delta_A)} \\ GG(A) & \xrightarrow{\delta_{G(A)}} & GGG(A), \end{array} \qquad \begin{array}{ccc} G(A) & \xrightarrow{\delta_A} & GG(A) \\ {\scriptstyle \delta_A}\downarrow & \searrow^{G(A)} & \downarrow{\scriptstyle G(\sigma_A)} \\ GG(A) & \xrightarrow{\sigma_{G(A)}} & G(A) \end{array}$$

are commutative.

A *coalgebra* or *comodule* of a comonad (G, δ, σ) is a pair (A, ϱ^A), where A is an object in $\mathbf{A}$ and ϱ^A is a morphism $\varrho^A : A \to G(A)$ rendering commutative the following diagrams

$$\begin{array}{ccc} A & \xrightarrow{\varrho^A} & G(A) \\ {\scriptstyle \varrho^A}\downarrow & & \downarrow{\scriptstyle \delta_A} \\ G(A) & \xrightarrow{G(\varrho^A)} & GG(A), \end{array} \qquad \begin{array}{ccc} A & \xrightarrow{\varrho^A} & G(A) \\ & \searrow_{A} & \downarrow{\scriptstyle \sigma_A} \\ & & A. \end{array}$$

A morphism of G-coalgebras (A, ϱ^A), (B, ϱ^B) is an element $f \in \mathbf{A}(A, B)$ such that

$$\varrho^B \circ f = G(f) \circ \varrho^A.$$

G-coalgebras with their morphisms form a category denoted by $\mathbf{A}_G$.

Dual to comonads one defines *monads* as functors $F : \mathbf{A} \to \mathbf{A}$ together with natural transformations $\mu : FF \to F$ and $\eta : \mathrm{id}_{\mathbf{A}} \to F$ such that for all objects $A \in \mathbf{A}$, the following diagrams

$$\begin{array}{ccc} FFF(A) & \xrightarrow{F(\mu_A)} & FF(A) \\ {\scriptstyle \mu_{F(A)}}\downarrow & & \downarrow{\scriptstyle \mu_A} \\ FF(A) & \xrightarrow{\mu_A} & A\,, \end{array} \qquad \begin{array}{ccc} F(A) & \xrightarrow{\eta_{F(A)}} & FF(A) \\ {\scriptstyle F(\eta_A)}\downarrow & \searrow^{F(A)} & \downarrow{\scriptstyle \mu_A} \\ FF(A) & \xrightarrow{\mu_A} & F(A) \end{array}$$

are commutative. To monads one associates the category of *algebras* or *modules*, denoted by $\mathbf{A}^F$.[1] In this chapter, we are primarily interested in comonads and coalgebras.

2.12. Comonads and adjoint functors. Any adjoint pair of functors $L : \mathbf{A} \to \mathbf{B}$, $R : \mathbf{B} \to \mathbf{A}$ gives rise to a comonad (G, δ, σ) on $\mathbf{B}$, where $G = LR$, and for all objects B in $\mathbf{B}$, $\delta_B = L(\eta_{R(B)})$, $\sigma_B = \psi_B$. Here, η is the unit and ψ is the counit for the adjunction (L, R).

Conversely, any comonad (G, δ, σ) on $\mathbf{A}$ gives rise to a pair of adjoint functors $L : \mathbf{A}_G \to \mathbf{A}$, $R : \mathbf{A} \to \mathbf{A}_G$. The functor L is the forgetful functor, while R is defined by $R(A) = (G(A), \delta_A)$. The unit and counit of adjunction are $\eta_{(A,\varrho^A)} = \varrho^A$ and $\psi_B = \sigma_B$, for every object B of $\mathbf{A}$ and a G-coalgebra (A, ϱ^A).

Dually, a pair of adjoint functors $L : \mathbf{A} \to \mathbf{B}$, $R : \mathbf{B} \to \mathbf{A}$ gives rise to a monad $F = RL : \mathbf{A} \to \mathbf{A}$. Conversely, a monad F induces a pair of adjoint functors $\mathbf{A} \to \mathbf{A}^F$, $\mathbf{A}^F \to \mathbf{A}$.

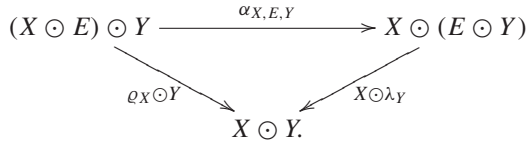
2.13. Monoidal categories and monoidal functors. A category $\mathbf{A}$ is called a *monoidal category* if there exist a functor $- \odot - : \mathbf{A} \times \mathbf{A} \to \mathbf{A}$, a distinguished object E in $\mathbf{A}$ and natural isomorphisms

$$\alpha : (-\odot-)\odot- \to -\odot(-\odot-), \quad \lambda : E\odot- \to \mathrm{id}_{\mathbf{A}}, \quad \varrho : -\odot E \to \mathrm{id}_{\mathbf{A}}$$

such that for all objects W, X, Y, Z in $\mathbf{A}$, the following two diagrams commute:

$$\begin{array}{ccccc} ((W\odot X)\odot Y)\odot Z & \xrightarrow{\alpha_{W,X,Y}\odot Z} & (W\odot(X\odot Y))\odot Z & & \\ & & & \searrow^{\alpha_{W,X\odot Y,Z}} & \\ {\scriptstyle \alpha_{W\odot X,Y,Z}}\downarrow & & & & W\odot((X\odot Y)\odot Z) \\ & & & \swarrow_{W\odot\alpha_{X,Y,Z}} & \\ (W\odot X)\odot(Y\odot Z) & \xrightarrow{\alpha_{W,X,Y\odot Z}} & W\odot(X\odot(Y\odot Z))\,, & & \end{array}$$

[1] The convention to denote coalgebras (comodules) of a comonad G by $\mathbf{A}_G$ while algebras of a monad by $\mathbf{A}^F$ is a standard category theory convention, which, unfortunately, is opposite to the module-theoretic conventions. We hope this will not cause any confusion later on.



A monoidal category is denoted by $(\mathbf{A}, \odot, E)$. The functor $\odot$ is referred to as a *multiplication* and E is called a *unit*. For example, if R is a ring, then the category of R-bimodules is monoidal with the usual tensor product of R-bimodules $\otimes_R$ as a multiplication and unit object R.

A *monoidal functor* between categories $(\mathbf{A}, \odot, E)$ and $(\mathbf{A}', \odot', E')$ is a triple (F, F^2, F^0), where $F : \mathbf{A} \to \mathbf{A}'$ is a functor, $F^2 : F(-) \odot' F(-) \to F(- \odot -)$ is a natural transformation, and $F^0 : E' \to F(E)$ is a morphism in $\mathbf{A}'$. F^2 and F^0 are compatible with the structure maps α, λ, ϱ and α', λ', ϱ' in an obvious (natural) way; see [132, Section 11.2] for more details. (F, F^2, F^0) is said to be *strong* if F^2 and F^0 are isomorphisms and *strict* if they are identities.

3. Corings

From now on, A is an associative unital algebra over an associative commutative ring k with unit. The product in A is denoted by μ_A and the unit, understood both as a map $k \to A$ or as an element of A, by 1_A. Unless explicitly specified otherwise, whenever we say "an algebra" we mean an associative k-algebra with unit. The unadorned tensor product is over k, that is, $\otimes = \otimes_k$.

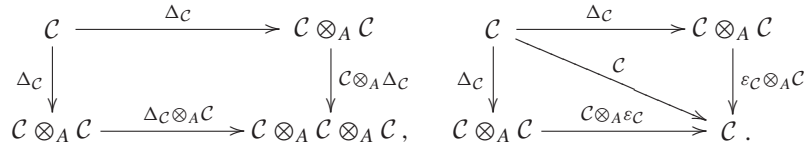
3.1. *Definition*

In this section, we give the definition and categorical interpretation of corings.

3.1. Definition of a coring [178]. An *A-coring* is an A-bimodule $\mathcal{C}$ with A-bilinear maps

$$\Delta_{\mathcal{C}} : \mathcal{C} \to \mathcal{C} \otimes_A \mathcal{C} \quad \text{and} \quad \varepsilon_{\mathcal{C}} : \mathcal{C} \to A,$$

called *coproduct* and *counit*, rendering commutative the following diagrams



The first of these diagrams is usually referred to as the *coassociative law* or it is said that it expresses the *coassociativity* of the coproduct. The diagrams involving $\varepsilon_{\mathcal{C}}$ express the *counitality* of the coproduct. In these diagrams, the obvious isomorphisms $A \otimes_A \mathcal{C} \simeq \mathcal{C} \simeq \mathcal{C} \otimes_A A$ are suppressed.

3.2. Coassociativity and the Sweedler–Heyneman notation. The *Sweedler–Heyneman notation* allows one to express the action of the coproduct on elements. Given $c \in \mathcal{C}$, one writes

$$\Delta_{\mathcal{C}}(c) = \sum c_{(1)} \otimes_A c_{(2)}$$

(with an implicit summation; the summation sign reminds about it). The coassociative law means that

$$\sum c_{(1)} \otimes_A c_{(2)(1)} \otimes_A c_{(2)(2)} = \sum c_{(1)(1)} \otimes_A c_{(1)(2)} \otimes_A c_{(2)},$$

and therefore one can relabel the Sweedler indices and write

$$(\mathcal{C} \otimes_A \Delta_{\mathcal{C}}) \circ \Delta_{\mathcal{C}}(c) = \sum c_{(1)} \otimes_A c_{(2)} \otimes_A c_{(3)}.$$

For a detailed discussion of the Sweedler–Heyneman notation, we refer to [140, pp. 610–611].

3.3. Morphisms of corings. Let $\mathcal{C}$ be an A-coring with coproduct $\Delta_{\mathcal{C}}$ and counit $\varepsilon_{\mathcal{C}}$, and let $\mathcal{D}$ be an A-coring with coproduct $\Delta_{\mathcal{D}}$ and counit $\varepsilon_{\mathcal{D}}$. An A-bimodule map $f : \mathcal{C} \to \mathcal{D}$ is said to be a *morphism of A-corings* or an *A-coring morphism* provided

$$\Delta_{\mathcal{D}} \circ f = (f \otimes_A f) \circ \Delta_{\mathcal{C}}, \qquad \varepsilon_{\mathcal{D}} \circ f = \varepsilon_{\mathcal{C}}.$$

In Sweedler's notation, one thus requires that for all $c \in C$,

$$\sum f(c)_{(1)} \otimes_A f(c)_{(2)} = \sum f(c_{(1)}) \otimes_A f(c_{(2)}), \qquad \varepsilon_{\mathcal{D}}(f(c)) = \varepsilon_{\mathcal{C}}(c).$$

3.4. Coideals. The kernel of a surjective A-coring map $p : \mathcal{C} \to \mathcal{D}$ is called a *coideal* in $\mathcal{C}$. Thus, an A-subbimodule $K \subset \mathcal{C}$ is a coideal if and only if $\mathcal{D} = \mathcal{C}/K$ is a coring and $p : \mathcal{C} \to \mathcal{C}/K$ is a coring map. Equivalently, $\Delta_{\mathcal{C}}(K) \subseteq \ker(p \otimes_A p)$ and $\varepsilon_{\mathcal{C}}(K) = 0$. If the inclusion $K \subset \mathcal{C}$ is left and right $\mathcal{C}$-pure (i.e. if it is a monomorphism after tensoring with $\mathcal{C}$ over A), the former condition is equivalent to $\Delta_{\mathcal{C}}(K) = \mathcal{C} \otimes_A K + K \otimes_A \mathcal{C}$.

3.5. Corings and comonads. For an A-bimodule $\mathcal{C}$, the following statements are equivalent:

(a) *$\mathcal{C}$ is an A-coring with comultiplication $\Delta_{\mathcal{C}}$ and counit $\varepsilon_{\mathcal{C}}$.*
(b) *The functor $- \otimes_A \mathcal{C} : \mathbf{M}_A \to \mathbf{M}_A$ is a comonad with comultiplication δ such that $\delta_A = \Delta_{\mathcal{C}}$ and counit σ such that $\sigma_A = \varepsilon_{\mathcal{C}}$.*
(c) *The functor $\mathcal{C} \otimes_A - : {}_A\mathbf{M} \to {}_A\mathbf{M}$ is a comonad with comultiplication δ such that $\delta_A = \Delta_{\mathcal{C}}$ and counit σ such that $\sigma_A = \varepsilon_{\mathcal{C}}$.*

Sketch of proof. If $\Delta_{\mathcal{C}}$ is a comultiplication for a coring $\mathcal{C}$, the comultiplication δ for $- \otimes_A \mathcal{C}$ (resp. $\mathcal{C} \otimes_A -$) is given by $\delta_M = M \otimes_A \Delta_{\mathcal{C}}$ (resp. $\delta_M = \Delta_{\mathcal{C}} \otimes_A M$). If $\varepsilon_{\mathcal{C}}$ is the counit for $\mathcal{C}$, then the counit σ for the comonad $- \otimes_A \mathcal{C}$ (resp. $\mathcal{C} \otimes_A -$) is $\sigma_M = M \otimes_A \varepsilon_{\mathcal{C}}$ (resp. $\sigma_M = \varepsilon_{\mathcal{C}} \otimes_A M$). □

3.6. The base algebra extension coring. Let $\mathcal{D}$ be a B-coring and let $j : B \to A$ be an algebra map. Thus A is a B-bimodule by $bab' = j(b)aj(b')$. Then, $\mathcal{C} = A \otimes_B \mathcal{D} \otimes_B A$ is an A-coring with the coproduct

$$\Delta_{\mathcal{C}} : A \otimes_B \mathcal{D} \otimes_B A \to \mathcal{C} \otimes_A \mathcal{C} \simeq A \otimes_B \mathcal{D} \otimes_B A \otimes_B \mathcal{D} \otimes_B A,$$
$$a \otimes_B d \otimes_B a' \mapsto \sum a \otimes_B d_{(1)} \otimes_B 1_A \otimes_B d_{(2)} \otimes_B a',$$

and counit $\varepsilon_{\mathcal{C}} : A \otimes_B \mathcal{D} \otimes_B A \to A$, $a \otimes_B d \otimes_B a' \mapsto aj(\varepsilon_{\mathcal{D}}(d))a'$.

3.7. The category of corings. Corings over different rings can be grouped in the *category of corings* **Crg**. Objects in **Crg** are pairs $(\mathcal{C}, A)$, where A is a k-algebra and $\mathcal{C}$ is an A-coring. A morphism between corings $(\mathcal{C}, A)$ and $(\mathcal{D}, B)$ (cf. [101]) is a pair of maps (γ, α) satisfying the following conditions:

(1) $\alpha : A \to B$ is an algebra map. Thus one can view $\mathcal{D}$ as an A-bimodule via α. Explicitly, $ada' = \alpha(a)d\alpha(a')$ for all $a, a' \in A$ and $d \in \mathcal{D}$.
(2) $\gamma : \mathcal{C} \to \mathcal{D}$ is an A-bimodule map such that the induced map,

$$B \otimes_A \mathcal{C} \otimes_A B \to \mathcal{D}, \qquad b \otimes_A c \otimes_A b' \mapsto b\gamma(c)b',$$

is a morphism of B-corings, where $B \otimes_A \mathcal{C} \otimes_A B$ is the base algebra extension of $\mathcal{C}$; see 3.6.

3.2. *Examples*

In this section, we list a number of examples of corings, in particular those which appeared in the literature in recent years and which are related to other fields of algebra.

3.8. Coalgebras. If A is a commutative algebra and $\mathcal{C}$ is an A-module (i.e. both the left and right A-multiplications coincide, i.e. $ac = ca$, for all $a \in A$ and $c \in \mathcal{C}$), then an A-coring $\mathcal{C}$ is called a *coalgebra*. Thus any k-coalgebra is a k-coring (but not every k-coring is a k-coalgebra). Whatever can be said about a coring, can also be said about a k-coalgebra.

3.9. The trivial A-coring. An algebra A is itself an A-coring with both the coproduct and counit given by the identity map A.

3.10. The Sweedler coring [178]. Given an algebra map $j : B \to A$, one defines the *canonical Sweedler coring* as follows. First, one views A as a B-bimodule via the algebra map j, i.e. $bab' = j(b)aj(b')$, for all $a \in A$ and $b, b' \in B$. Then, one forms the A-bimodule $\mathcal{C} = A \otimes_B A$, with the A-multiplications $a(a' \otimes_B a'')a''' = aa' \otimes_B a''a'''$. The coproduct is defined by

$$\Delta_{\mathcal{C}} : \mathcal{C} \to \mathcal{C} \otimes_A \mathcal{C} \simeq A \otimes_B A \otimes_B A, \qquad a \otimes_B a' \mapsto a \otimes_B 1_A \otimes_B a',$$

and the counit is $\varepsilon_{\mathcal{C}} : \mathcal{C} \to A$, $a \otimes_B a' \mapsto aa'$.

Even if both A and B are commutative algebras, the associated Sweedler coring need not be an A-coalgebra. The Sweedler coring can be understood as a base algebra extension coring 3.6, where $\mathcal{D} = B$ is a trivial coring.

3.11. Comatrix corings [92]. Given algebras B and A, let P be a (B, A)-bimodule that is finitely generated and projective as a right A-module. Let $e_i \in P$ and $\pi_i \in P^* = \mathrm{Hom}_A(P, A)$ be a finite dual basis for P. Write $e = \sum_i e_i \otimes_A \pi_i$ for the dual basis (canonical) element. The dual module P^* is an (A, B)-bimodule with multiplications $(afb)(x) = af(bx)$, for all $a \in A$, $b \in B$, $x \in P$ and $f \in P^*$. Thus $\mathcal{C} = P^* \otimes_B P$ is an A-bimodule, $a(f \otimes_B x)a' = af \otimes_B xa'$. Similarly, $P \otimes_A P^*$ is a B-bimodule, and for all $b \in B$, $be = eb$. $\mathcal{C}$ is an A-coring with the coproduct and counit,

$$\Delta_{\mathcal{C}} : P^* \otimes_B P \to P^* \otimes_B P \otimes_A P^* \otimes_B P, \qquad x \otimes_B f \mapsto x \otimes_B e \otimes_B f,$$

and $\varepsilon_{\mathcal{C}} : P^* \otimes_B P \to A$, $f \otimes_B x \mapsto f(x)$. The coring $\mathcal{C}$ is called a *(finite) comatrix coring* associated to a (B, A)-bimodule P.

One way of understanding a comatrix coring is to observe that since P is a finitely generated projective right A-module, for all right A-modules N, $\mathrm{Hom}_A(P, N) \simeq N \otimes_A P^*$. Hence $- \otimes_B P : \mathbf{M}_B \to \mathbf{M}_A$ is a left adjoint of $- \otimes_A P^* : \mathbf{M}_A \to \mathbf{M}_B$ (this is the hom-tensor adjunction 2.9), with the unit given by the dual basis element e and the counit given by the evaluation map. Thus $- \otimes_A P^* \otimes_B P : \mathbf{M}_A \to \mathbf{M}_A$ is a comonad (see 2.12), and therefore, by 3.5, $P^* \otimes_B P$ is an A-coring with the structure given above.

In case of an algebra map $B \to A$, A is a (B, A)-bimodule that is (obviously) finitely generated and projective as an A-module. The dual module A^* can be naturally identified with A. In this way, one can understand the Sweedler coring as a (finite) comatrix coring associated to A.

3.12. Infinite comatrix corings I [93]. Let A be a (unital) k-algebra, and let $\mathcal{P}$ be a set of finitely generated and projective right A-modules. For any $P \in \mathcal{P}$, write $e_P = \sum_i e_i^P \otimes_A \pi_i^P$ for the dual basis element of P. Set $M = \bigoplus_{P \in \mathcal{P}} P$ and let $\{u_P : M \to M \mid P \in \mathcal{P}\}$ be the corresponding set of orthogonal idempotents. For a subalgebra T of the endomorphism algebra $\mathrm{End}_A(M)$, define $B = \sum_{P,Q \in \mathcal{P}} u_P T u_Q$. Then B is a firm (nonunital) algebra and M is a (B, A)-bimodule. Define an (A, B)-bimodule

$$M^\dagger = \bigoplus_{P \in \mathcal{P}} \mathrm{Hom}_A(P, A).$$

$\mathcal{C} = M^\dagger \otimes_B M$ is an A-coring with the coproduct and counit determined by

$$\Delta_{\mathcal{C}}(f \otimes_B x) = f \otimes_B e_P \otimes_B x, \qquad \varepsilon_{\mathcal{C}}(f \otimes_B x) = f(x),$$

for all $f \in \mathrm{Hom}_A(P, A)$ and $x \in P$. A construction of $\mathcal{C}$ from colimits is described in [58]. If $\mathcal{P}$ is a one-element or finite set, then $\mathcal{C}$ is a finite comatrix coring as defined in 3.11.

3.13. Comatrix coring contexts [45, 103]. Let A be a unital algebra, and let B be a nonunital algebra that is *firm*[2] or *regular* in the sense that the multiplication map

[2] The terminologies *firm ring* and *firm module* are taken from [149].

$B \otimes_B B \to B$ is an isomorphism (of left or right B-modules). A *comatrix coring context* consists of

(i) an (A, B)-bimodule N, which is firm as a right B-module in the sense that the product map $N \otimes_B B \to N$ is an isomorphism of right B-modules;
(ii) a (B, A)-bimodule M, which is firm as a left B-module in the sense that the product map $B \otimes_B M \to M$ is an isomorphism of left B-modules;
(iii) a pair of A- resp. B-bimodule maps

$$\sigma : N \otimes_B M \to A, \qquad \tau : B \to M \otimes_A N,$$

such that the following diagrams

$$\begin{array}{ccc} N \otimes_B M \otimes_A N & \xrightarrow{\sigma \otimes_A N} & A \otimes_A N \\ {\scriptstyle N \otimes_B \tau}\uparrow & & \downarrow{\scriptstyle \simeq} \\ N \otimes_B B & \xrightarrow{\simeq} & N \end{array} \qquad , \qquad \begin{array}{ccc} M \otimes_A N \otimes_B M & \xleftarrow{\tau \otimes_B M} & B \otimes_B M \\ {\scriptstyle M \otimes_A \sigma}\downarrow & & \downarrow{\scriptstyle \simeq} \\ M \otimes_A A & \xrightarrow{\simeq} & M \end{array}$$

commute. A comatrix coring context is denoted by $(A, B, N, M, \sigma, \tau)$. By [189, Theorem 2.51], the existence of a comatrix coring context $(A, B, N, M, \sigma, \tau)$ is equivalent to the statement that $- \otimes_B M : \mathbf{M}_B \to \mathbf{M}_A$ and $- \otimes_A N : \mathbf{M}_A \to \mathbf{M}_B$ is an adjoint pair of functors between the categories of firm right modules.

In case B is a unital algebra (and then, by convention, M, N are unital B-modules), a comatrix coring context exists if and only if M is finitely generated and projective as a right A-module. The module N can be identified with the dual module M^*. With this identification, σ is the evaluation map and τ is the map $b \mapsto be$, where $e \in M \otimes_A M^*$ is the dual basis element for M.

3.14. Corings from comatrix contexts [45,103]. Let $(A, B, N, M, \sigma, \tau)$ be a comatrix coring context. Write $\nabla_N : N \to N \otimes_B B$ for the inverse of the product map $N \otimes_B B \to N$ and $\nabla_M : M \to B \otimes_B M$ for the inverse of the product map $B \otimes_B M \to M$. Then the A-bimodule $\mathcal{C} = N \otimes_B M$ is an A-coring with the coproduct

$$\Delta_{\mathcal{C}} = (N \otimes_B \tau \otimes_B M) \circ (\nabla_N \otimes_B M) = (N \otimes_B \tau \otimes_B M) \circ (N \otimes_B \nabla_M)$$

and counit $\varepsilon_{\mathcal{C}} = \sigma$. This is a consequence of the fact that $- \otimes_B M : \mathbf{M}_B \to \mathbf{M}_A$ and $- \otimes_A N : \mathbf{M}_A \to \mathbf{M}_B$ are an adjoint pair (compare the discussion in 3.11).

In case of a unital ring B, the coring $\mathcal{C}$ constructed here is isomorphic to a (finite) comatrix coring 3.11.

3.15. Infinite comatrix corings II [103]. Let A and B be (unital) k-algebras. Consider an (A, B)-bimodule N and a (B, A)-bimodule M. Let $\sigma : N \otimes_B M \to A$ be an A-bimodule map so that $M \otimes_A N$ can be made into a nonunital k-algebra with product

$$M \otimes_A \sigma \otimes_A N : M \otimes_A N \otimes_B M \otimes_A N \to M \otimes_A N.$$

Let R be a firm ring and suppose that there exists a (nonunital) algebra map $\tau : R \to M \otimes_A N$. Then M is a left R-module with multiplication $rm = (M \otimes_A \sigma)(\tau(r) \otimes_B m)$ and N is a right R-module with multiplication $nr = (\sigma \otimes_A N)(n \otimes_B \tau(r))$. If M is

a firm left R-module (with the isomorphism denoted by $\nabla_M : M \to R \otimes_R M$), then $\mathcal{C} = N \otimes_R M$ is an A-coring with the counit $\varepsilon_{\mathcal{C}}(n \otimes_R m) = \sigma(n \otimes_B m)$ and the coproduct

$$\Delta_{\mathcal{C}} : N \otimes_R M \to N \otimes_R M \otimes_A N \otimes_R M, \quad \Delta_{\mathcal{C}} = (N \otimes_R \tau \otimes_R M) \circ (N \otimes_R \nabla_M).$$

The coring $\mathcal{C}$ is known as an *infinite comatrix coring*. $\mathcal{C}$ is isomorphic to the coring constructed from the comatrix context $(A, R, N \otimes_R R, M, \sigma, \tau)$.

In the setup of 3.12, an infinite comatrix coring is constructed by taking $N = \mathrm{Hom}_A(M, A)$; see [103, Section 6] for a detailed discussion of the relationship of infinite comatrix corings to the corings described in 3.12.

3.16. Corings and semifree differential graded algebras [162]. A *differential (non-negatively) graded algebra* is an $\mathbb{N} \cup \{0\}$-graded k-algebra $\Omega = \bigoplus_{n=0}^{\infty} \Omega^n$ together with a k-linear degree one operation $d : \Omega^{\bullet} \to \Omega^{\bullet+1}$ such that $d \circ d = 0$ and, for all elements ω' and all degree n elements ω, $d(\omega\omega') = d(\omega)\omega' + (-1)^n \omega d(\omega')$. The degree zero part Ω^0 is a subalgebra of Ω; denote it by A. Ω^1 is an A-bimodule. Ω is said to be *semifree*, provided $\Omega^n = \Omega^{n-1} \otimes_A \Omega^1$, for all $n = 1, 2, \ldots$.

Starting with a semifree differential graded algebra Ω ($A = \Omega^0$), define

$$\mathcal{C} = Ag \oplus \Omega^1,$$

where g is an indeterminate. In other words $\mathcal{C}$ is a direct sum of A and Ω^1 as a left A-module. A compatible right A-module structure on $\mathcal{C}$ is defined by

$$(ag + \omega)a' = aa'g + ada' + \omega a',$$

for all $a, a' \in A$, $\omega \in \Omega^1$. $\mathcal{C}$ is an A-coring with the coproduct specified by

$$\Delta_{\mathcal{C}}(ag) = ag \otimes_A g, \qquad \Delta_{\mathcal{C}}(\omega) = g \otimes_A \omega + \omega \otimes_A g - d(\omega),$$

and the counit $\varepsilon_{\mathcal{C}}(ag + \omega) = a$, for all $a \in A$ and $\omega \in \Omega^1$.

3.17. Frobenius extensions. The studies of Frobenius extensions of noncommutative rings were initiated in [119]. According to [120, 143], an algebra map $A \to B$ is called a *Frobenius extension* (of the first kind) if and only if B is a finitely generated projective right A-module and $B \simeq \mathrm{Hom}_A(B, A)$ as (A, B)-bimodules. The (A, B)-bimodule structure of $\mathrm{Hom}_A(B, A)$ is given by $(afb)(b') = af(bb')$, for all $a \in A, b, b' \in B$ and $f \in \mathrm{Hom}_A(B, A)$. The statement that $A \to B$ is a Frobenius extension is equivalent to the existence of an A-bimodule map $E : B \to A$ and an element $\beta = \sum_i b_i \otimes \bar{b}^i \in B \otimes_A B$ such that, for all $b \in B$,

$$\sum\nolimits_i E(bb_i)\bar{b}^i = \sum\nolimits_i b_i E(\bar{b}^i b) = b.$$

E is called a *Frobenius homomorphism* and β is known as a *Frobenius element*. One can easily show that $\beta \in (B \otimes_A B)^B := \{x \in B \otimes_A B \mid \text{for all } b \in B, bx = xb\}$.

3.18. Frobenius extensions and corings [116, Proposition 4.6]. *An algebra map $A \to B$ is a Frobenius extension if and only if B is an A-coring such that the coproduct is a B-bimodule map.*

Sketch of proof. If $A \to B$ is a Frobenius extension with a Frobenius homomorphism E and a Frobenius element β, then B is an A-coring with the counit E and the B-bimodule coproduct $\Delta_B : B \to B \otimes_A B$, $b \mapsto b\beta = \beta b$. Conversely, if B is an A-coring with a B-bimodule coproduct Δ_B, then $\beta = \Delta_B(1_B)$ is a Frobenius element and the counit of B is a Frobenius homomorphism. □

This description of Frobenius extensions as corings generalizes the description of Frobenius algebras in terms of coalgebras [1]. The fact that a Frobenius extension induces a coring has been proved already in [191, Proposition 1.8.3].

3.19. Depth-2 extensions [118]. An extension $B \subseteq R$ of k-algebras is called a depth-2 extension if $R \otimes_B R$ is a direct summand in a finite direct sum of copies of R both as an (R, B)-bimodule and as a (B, R) bimodule. By [118, Lemma 3.7], this is equivalent to the existence of right and left *D2 quasi-bases*. A right D2 quasi-basis consists of finite sets $\{\gamma_j\}_{j\in J}$ in ${}_B\mathrm{End}_B(R)$ and $\{\sum_{m\in M_j} r^j{}_m \otimes_B r'^j{}_m\}_{j\in J}$ in $(R \otimes_B R)^B = \{x \in R \otimes_B R \mid \forall b \in B,\ bx = xb\}$ such that

$$\sum_{j\in J, m\in M_j} r\gamma_j(r')c^j{}_m \otimes_R c'^j{}_m = r \otimes_B r',$$

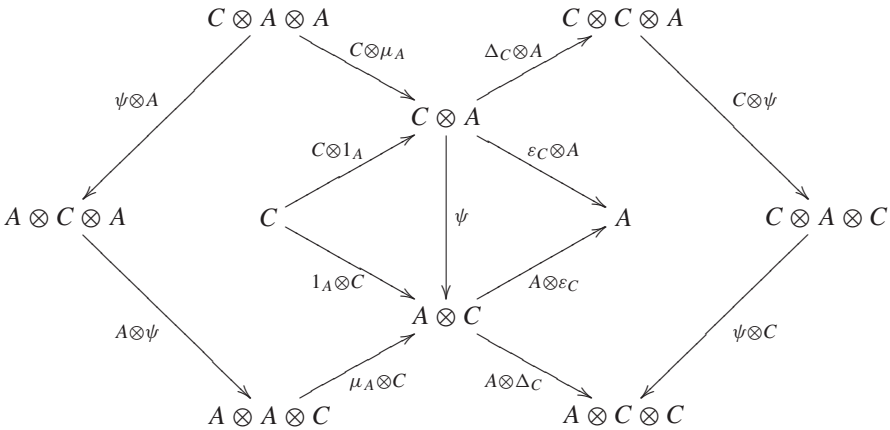
for all $r \otimes_B r' \in R \otimes_B R$. A left D2 quasi-basis is defined symmetrically.

Given a depth-2 extension $B \subseteq R$, define $A = \{a \in R \mid \forall b \in B,\ ab = ba\}$. Then $\mathcal{C} = {}_B\mathrm{End}_B(R)$ is an A-coring with A-multiplications $aca' : r \mapsto ac(r)a'$, counit $\varepsilon_{\mathcal{C}}(c) = c(1_R)$, and coproduct

$$\Delta_{\mathcal{C}}(c) = \sum_{j\in J} \gamma_j \otimes_A \sum_{m\in M_j} r^j{}_m c(r'^j{}_m -).$$

Furthermore, $\mathcal{D} = (R \otimes_B R)^B$ is an A-coring with coproduct and counit obtained by restricting the structure maps of the Sweedler coring $R \otimes_B R$.

3.20. Entwining structures [49]. A (right–right) *entwining structure* (over k) $(A, C)_\psi$ consists of a k-algebra $(A, \mu_A, 1_A)$, a k-coalgebra $(C, \Delta_C, \varepsilon_C)$, and a k-module map $\psi : C \otimes A \to A \otimes C$, rendering commutative the following *bow-tie diagram*:



The map ψ is known as an *entwining map*. C and A are said to be *entwined* by ψ.

An entwining structure is a special case of a *mixed distributive law* between a comonad and a monad; see [18]. The comonad is $- \otimes C$ and the monad is $- \otimes A$.

3.21. Corings associated to entwining structures [37, Proposition 2.2]. Let A be a k-algebra and C be a k-coalgebra with coproduct Δ_C and counit ε_C.

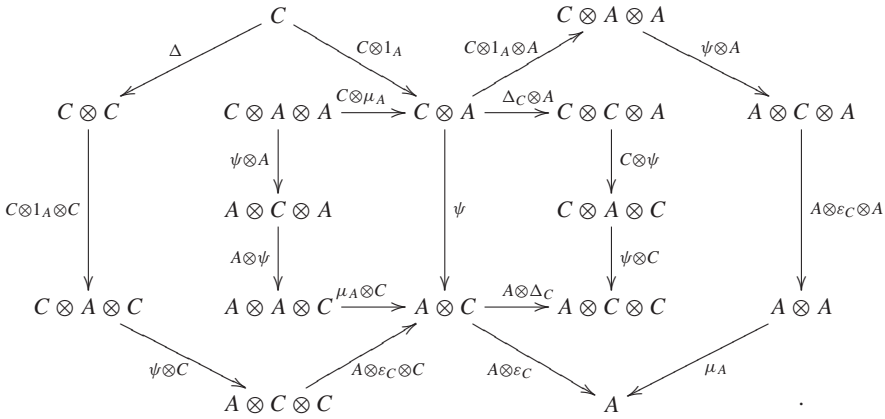
For an entwining structure $(A, C)_\psi$, view $A \otimes C$ as an (A, A)-bimodule with the left action $a(a' \otimes c) = aa' \otimes c$ and the right action $(a' \otimes c)a = a'\psi(c \otimes a)$, for all $a, a' \in A, c \in C$. Then $\mathcal{C} = A \otimes C$ is an A-coring with the coproduct

$$\Delta_{\mathcal{C}} : \mathcal{C} \to \mathcal{C} \otimes_A \mathcal{C} \simeq A \otimes C \otimes C, \qquad \Delta_{\mathcal{C}} = A \otimes \Delta_C,$$

and the counit $\varepsilon_{\mathcal{C}} = A \otimes \varepsilon_C$.

Conversely, if $A \otimes C$ is an A-coring with the coproduct, counit, and left A-module structure above, then $(A, C)_\psi$ is an entwining structure, where $\psi : c \otimes a \mapsto (1 \otimes c)a$.

3.22. Weak entwining structures [56]. A (right–right) *weak entwining structure* (over k) is a triple (A, C, ψ) consisting of a k-algebra $(A, \mu_A, 1_A)$, a k-coalgebra $(C, \Delta_C, \varepsilon_C)$, and a k-module map $\psi : C \otimes A \to A \otimes C$, rendering commutative the following bow-tie diagram:



One easily checks that any entwining structure is a weak entwining structure.

3.23. Corings associated to weak entwining structures [37, Proposition 2.3]. Let A be an algebra and C be a coalgebra with coproduct Δ_C and counit ε_C.

Given a weak entwining structure (A, C, ψ), define

$$p : A \otimes C \to A \otimes C, \qquad p = (\mu_A \otimes C) \circ (A \otimes \psi) \circ (A \otimes C \otimes 1_A),$$

and

$$\mathcal{C} = \operatorname{Im} p = \Big\{ \sum_i a_i \psi(c_i \otimes 1_A) \mid \sum_i a_i \otimes c_i \in A \otimes C \Big\}.$$

$\mathcal{C}$ is an A-bimodule with the A-multiplications $a' \left(\sum_i a_i \otimes c_i\right) a'' = \sum_i a' a_i \psi(c_i \otimes a'')$. Furthermore, $\mathcal{C}$ is an A-coring with coproduct $\Delta_{\mathcal{C}} = (A \otimes \Delta_C)|_{\mathcal{C}}$ and counit $\varepsilon_{\mathcal{C}} = (A \otimes \varepsilon_C)|_{\mathcal{C}}$.

3.3. *Duality*

The notion of an A-coring is dual to that of an A-ring or an *algebra over an algebra* or an *algebra extension*. In this section, we explain this duality.

3.24. Rings over algebras. Given an algebra A, an *A-ring* is an A-bimodule B together with A-bimodule maps $\mu_B : B \otimes_A B \to B$ and $\iota_B : A \to B$ rendering commutative the following diagrams:

$$\begin{array}{ccc} B \otimes_A B \otimes_A B & \xrightarrow{\mu_B \otimes_A B} & B \otimes_A B \\ {\scriptstyle B\otimes_A \mu_B}\downarrow & & \downarrow{\scriptstyle \mu_B} \\ B \otimes_A B & \xrightarrow{\mu_B} & B, \end{array} \qquad \begin{array}{ccc} B & \xrightarrow{B\otimes_A \iota_B} & B \otimes_A B \\ {\scriptstyle \iota_B\otimes_A B}\downarrow & \searrow^{B} & \downarrow{\scriptstyle \mu_B} \\ B \otimes_A B & \xrightarrow{\mu_B} & B \end{array}$$

The map μ_B is called a *product* and ι_B is called a *unit*. The product map μ_B composed with the canonical surjection $B \otimes B \to B \otimes_A B$ induces a k-algebra structure on B. The unit for the k-algebra B is $\iota_B(1_A)$. Thus the notion of an A-ring is equivalent to the existence of an algebra map $\iota_B : A \to B$.

3.25. Dual rings [178, Proposition 3.2]. Let $\mathcal{C}$ be an A-coring.

(1) $\mathcal{C}^* = \mathrm{Hom}_A(\mathcal{C}, A)$ is an algebra with unit $\varepsilon_{\mathcal{C}}$ and the product

$$f *^r g \; : \; \mathcal{C} \xrightarrow{\Delta_{\mathcal{C}}} \mathcal{C} \otimes_A \mathcal{C} \xrightarrow{f\otimes_A \mathcal{C}} \mathcal{C} \xrightarrow{g} A, \quad f *^r g(c) = \sum g(f(c_{(1)})c_{(2)}),$$

and there is an antialgebra map $\iota_R : A \to \mathcal{C}^*,\ a \mapsto \varepsilon_{\mathcal{C}}(a-)$. Thus $\mathcal{C}^*$ is an A^{op}-ring.

(2) ${}^*\mathcal{C} = {}_A\mathrm{Hom}(\mathcal{C}, A)$ is an algebra with unit $\varepsilon_{\mathcal{C}}$ and the product

$$f *^l g \; : \; \mathcal{C} \xrightarrow{\Delta_{\mathcal{C}}} \mathcal{C} \otimes_A \mathcal{C} \xrightarrow{\mathcal{C}\otimes_A g} \mathcal{C} \xrightarrow{f} A, \quad f *^l g(c) = \sum f(c_{(1)}g(c_{(2)})),$$

and there is an antialgebra map $\iota_L : A \to {}^*\mathcal{C},\ a \mapsto \varepsilon_{\mathcal{C}}(-a)$. Thus ${}^*\mathcal{C}$ is an A^{op}-ring.

(3) ${}^*\mathcal{C}^* = {}_A\mathrm{Hom}_A(\mathcal{C}, A)$ is an algebra with unit $\varepsilon_{\mathcal{C}}$ and the product

$$f * g(c) = \sum f(c_{(1)})g(c_{(2)}),$$

and there is a ring morphism $Z(A) \to Z({}^*\mathcal{C}^*),\ a \mapsto \varepsilon_{\mathcal{C}}(a-) = \varepsilon_{\mathcal{C}}(-a)$, where $Z(A)$ is the center of A and $Z({}^*\mathcal{C}^*)$ is the center of ${}^*\mathcal{C}^*$. Thus ${}^*\mathcal{C}^*$ is a $Z(A)$-algebra.

3.26. Dual rings as natural endomorphism rings.

(1) The algebra $\mathcal{C}^*$ in 3.25 is isomorphic to the algebra of natural endomorphisms of the forgetful functor ${}_A(-) : {}^{\mathcal{C}}\mathbf{M} \to {}_A\mathbf{M}$, with the product given by $\varphi\varphi' = \varphi \circ \varphi'$.

(2) The algebra ${}^*\mathcal{C}$ in 3.25 is isomorphic to the algebra of natural endomorphisms of the forgetful functor $(-)_A : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}_A$.

3.27. Monoidal functors and $B|A$-corings [184]. Corings and dual rings are closely related to monoidal functors between bimodule categories; see 2.13 for the definition of a monoidal category and a monoidal functor.

Let A and B be k-algebras. Following [184, Definition 3.5], a $B|A$-*coring* is a pair $(\mathcal{C}, \iota)$, where $\mathcal{C}$ is an A-coring and $\iota : B \to {}^*\mathcal{C}^* = {}_A\mathrm{Hom}_A(\mathcal{C}, A)$ is a morphism of algebras. Here ${}^*\mathcal{C}^*$ is an algebra as in 3.25(3).

As explained in [184, Theorem 3.6], there is a one-to-one correspondence between $B|A$-corings and representable monoidal functors between categories of bimodules $({}_A\mathbf{M}_A, \otimes_A, A) \to ({}_B\mathbf{M}_B, \otimes_B, B)$, that is, monoidal functors (F, F^2, F^0) with $F : {}_A\mathbf{M}_A \to {}_B\mathbf{M}_B$, of the form $M \to {}_A\mathrm{Hom}_A(\mathcal{C}, M)$, where $\mathcal{C}$ is an $(A^{\mathrm{e}}, B^{\mathrm{e}})$-bimodule ($A^{\mathrm{e}} = A \otimes A^{\mathrm{op}}$ is the enveloping algebra of A, and B^{e} is the enveloping algebra of B). Explicitly, given such an (F, F^2, F^0), the coproduct for $\mathcal{C}$ is $\Delta_{\mathcal{C}} = F^2_{\mathcal{C},\mathcal{C}}(\mathrm{id}_{\mathcal{C}} \otimes_B \mathrm{id}_{\mathcal{C}})$, the counit is $\varepsilon_{\mathcal{C}} = F^0(1_B)$, and the map $\iota : B \to {}^*\mathcal{C}^*$ is $\iota = F^0$.

3.28. Dual ring of the Sweedler coring. Let $B \to A$ be an algebra map and let $\mathcal{C}$ be the associated Sweedler coring 3.10. Using the isomorphism $\mathrm{Hom}_A(A \otimes_B A, A) \simeq \mathrm{End}_B(A)$, the right dual ring $\mathcal{C}^*$ is identified with the opposite of the endomorphism ring $\mathrm{End}_B(A)$. Similarly, ${}^*\mathcal{C}$ is identified with the endomorphism ring ${}_B\mathrm{End}(A)$ (both endomorphism rings with product given by the composition).

3.29. Dual ring of the comatrix coring. Let P be a (B, A)-bimodule such that P is finitely generated and projective as a right A-module, and let $\mathcal{C} = P^* \otimes_B P$ be the associated comatrix coring 3.11. Using the isomorphisms

$$ {}_A\mathrm{Hom}(P^* \otimes_B P, A) \simeq {}_B\mathrm{Hom}(P, {}_A\mathrm{Hom}(P^*, A)) \simeq {}_B\mathrm{End}(P), $$

the left dual ring ${}^*\mathcal{C}$ is identified with the endomorphism ring ${}_B\mathrm{End}(P)$. Similarly, $\mathcal{C}^*$ is identified with the opposite of the endomorphism ring $\mathrm{End}_B(P^*)$.

3.30. The ψ-twisted convolution algebra.

Let $(A, C)_\psi$ be an entwining structure and let $\mathcal{C} = A \otimes C$ be the associated A-coring; see 3.20, 3.21. With the isomorphism

$$ {}_A\mathrm{Hom}(A \otimes C, A) \simeq \mathrm{Hom}_k(C, A), $$

the algebra ${}^*\mathcal{C}$ can be identified with the k-module of k-linear maps $C \to A$ with the product

$$ f *_\psi g(c) = \sum_\psi g(c_{(2)})_\psi f(c_{(1)}{}^\psi), $$

where $\psi(c \otimes a) = \sum_\psi a_\psi \otimes c^\psi$ is the notation for the entwining map. The algebra $\mathrm{Hom}_k(C, A)$ with the above product (and unit $1_A \circ \varepsilon_C$) is known as a ψ-*twisted convolution algebra*.

3.31. Dual coring theorem [178, Theorem 3.7]. *Let $\phi : A \to S$ be an antialgebra map. View S as an A-bimodule with multiplications given by $asb := \phi(b)s\phi(a)$, for all $s \in S$, $a, b \in A$. Then*

(*1*) *If S is a finitely generated projective right A-module, then there is a unique A-coring structure on $S^* = \mathrm{Hom}_A(S, A)$ whereby the evaluation map*

$$\beta : S \to {}^*(S^*) = {}_A\mathrm{Hom}(\mathrm{Hom}_A(S, A), A), \quad s \mapsto [\sigma \mapsto \sigma(s)],$$

is an algebra isomorphism. Here, ${}^*(S^*)$ has an algebra structure as in 3.25. Furthermore, $\beta \circ \phi = \iota_L$, where ι_L is as in 3.25(2).

(*2*) *If S is a finitely generated projective left A-module, then there is a unique A-coring structure on ${}^*S = {}_A\mathrm{Hom}(S, A)$ whereby the evaluation map*

$$\beta : S \to ({}^*S)^* = \mathrm{Hom}_A({}_A\mathrm{Hom}(S, A), A), \quad s \mapsto [\sigma \mapsto \sigma(s)],$$

is an algebra isomorphism. Here, $({}^*S)^*$ has an algebra structure as in 3.25. Furthermore, $\beta \circ \phi = \iota_R$, where ι_R is as in 3.25(1).

Sketch of proof. The coproduct in $\mathcal{C} = S^*$ is defined as the unique map $\Delta_{\mathcal{C}} : \mathcal{C} \to \mathcal{C} \otimes_A \mathcal{C}$ such that, for all $s \otimes_A s' \in S \otimes_A S$ and $c \in \mathcal{C}$, $\Delta_{\mathcal{C}}(c)(s \otimes_A s') = c(s's)$. Explicitly, if $\{s^i \in S, c^i \in S^*\}$ is a finite dual basis for S, then $\Delta_{\mathcal{C}}(c) = \sum_i c(s^i -) \otimes_A c^i$. The counit of $\mathcal{C}$ is defined as $\varepsilon_{\mathcal{C}} : c \mapsto c(1_S)$. □

4. Comodules

Modules represent rings. Similarly, to represent corings one studies *comodules*. In this section, we describe the main properties and examples of comodules.

4.1. *Definition*

Here, we give the definition and basic properties of comodules.

4.1. Comodules of a coring. Let $\mathcal{C}$ be an A-coring and let $\mathcal{D}$ be a B-coring.

(1) A *right $\mathcal{C}$-comodule* is a pair (M, ϱ^M), where M is a right A-module, and $\varrho^M : M \to M \otimes_A \mathcal{C}$ is a right A-module map making the following diagrams

$$\begin{array}{ccc} M & \xrightarrow{\varrho^M} & M \otimes_A \mathcal{C} \\ {\scriptstyle \varrho^M}\downarrow & & \downarrow{\scriptstyle M \otimes_A \Delta_{\mathcal{C}}} \\ M \otimes_A \mathcal{C} & \xrightarrow{\varrho^M \otimes_A \mathcal{C}} & M \otimes_A \mathcal{C} \otimes_A \mathcal{C}, \end{array} \qquad \begin{array}{ccc} M & \xrightarrow{\varrho^M} & M \otimes_A \mathcal{C} \\ & \searrow^{\simeq} & \downarrow{\scriptstyle M \otimes_A \varepsilon_{\mathcal{C}}} \\ & & M \otimes_A A \end{array}$$

commute. The map ϱ^M is called a *right coaction*. A morphism of right $\mathcal{C}$-comodules (or a *right $\mathcal{C}$-colinear map*) $(M, \varrho^M) \to (N, \varrho^N)$ is a right A-module map $f : M \to N$ such that

$$(f \otimes_A \mathcal{C}) \circ \varrho^M = \varrho^N \circ f.$$

The category with objects right $\mathcal{C}$-comodules and morphisms right $\mathcal{C}$-colinear maps is denoted by $\mathbf{M}^{\mathcal{C}}$; k-modules of morphisms in this category are denoted by $\mathrm{Hom}^{\mathcal{C}}(M, N)$.

(2) A *left $\mathcal{C}$-comodule* is a pair $(M, {}^M\varrho)$, where M is a left A-module, and ${}^M\varrho : M \to \mathcal{C} \otimes_A M$ is a left A-module map making the following diagrams

$$\begin{array}{ccc} M & \xrightarrow{{}^M\varrho} & \mathcal{C} \otimes_A M \\ {\scriptstyle {}^M\varrho}\downarrow & & \downarrow{\scriptstyle \Delta_{\mathcal{C}} \otimes_A M} \\ \mathcal{C} \otimes_A M & \xrightarrow{\mathcal{C} \otimes_A {}^M\varrho} & \mathcal{C} \otimes_A \mathcal{C} \otimes_A M \end{array} , \qquad \begin{array}{ccc} M & \xrightarrow{{}^M\varrho} & \mathcal{C} \otimes_A M \\ & \searrow^{\simeq} & \downarrow{\scriptstyle \varepsilon_{\mathcal{C}} \otimes_A M} \\ & & A \otimes_A M \end{array}$$

commute. The map ${}^M\varrho$ is called a *left coaction*. A morphism of left $\mathcal{C}$-comodules (or a *left $\mathcal{C}$-colinear map*) $(M, {}^M\varrho) \to (N, {}^N\varrho)$ is a left A-module map $f : M \to N$ such that

$$(\mathcal{C} \otimes_A f) \circ {}^M\varrho = {}^N\varrho \circ f.$$

The category with objects left $\mathcal{C}$-comodules and morphisms left $\mathcal{C}$-colinear maps is denoted by ${}^{\mathcal{C}}\mathbf{M}$; k-modules of morphisms in this category are denoted by ${}^{\mathcal{C}}\mathrm{Hom}(M, N)$.

(3) A $(\mathcal{C}, \mathcal{D})$-bicomodule is a triple $(M, \varrho^M, {}^M\varrho)$, where M is an (A, B)-bimodule, (M, ϱ^M) is a right $\mathcal{D}$-comodule, and $(M, {}^M\varrho)$ is a left $\mathcal{C}$-comodule such that the right coaction ϱ^M is left $\mathcal{C}$-colinear or, equivalently, the left coaction ${}^M\varrho$ is right $\mathcal{D}$-colinear. Morphisms or $(\mathcal{C}, \mathcal{D})$-bicolinear maps are (A, B)-bimodule maps, which are at the same time left $\mathcal{C}$-colinear and right $\mathcal{D}$-colinear. The category of $(\mathcal{C}, \mathcal{D})$-bicomodules is denoted by ${}^{\mathcal{C}}\mathbf{M}^{\mathcal{D}}$; morphisms are denoted by ${}^{\mathcal{C}}\mathrm{Hom}^{\mathcal{D}}(M, N)$.

In case $\mathcal{D} = \mathcal{C}$ one simply talks about $\mathcal{C}$-bicomodules.

Whatever is said about right comodules can be equivalently said about left comodules (this is known as a *left-right symmetry*). Thus, in this chapter, we restrict our attention mainly to right comodules.

4.2. The Sweedler–Heyneman notation. For a right $\mathcal{C}$-comodule (M, ϱ^M), the coaction on elements $m \in M$ is denoted by the Sweedler–Heyneman notation:

$$\varrho^M(m) = \sum m_{(0)} \otimes_A m_{(1)} \in M \otimes_A \mathcal{C}.$$

Similar to coproducts, the coassociative law implies that

$$\sum m_{(0)} \otimes_A m_{(1)(1)} \otimes_A m_{(1)(2)} = \sum m_{(0)(0)} \otimes_A m_{(1)(1)} \otimes_A m_{(1)},$$

and therefore one can relabel the Sweedler–Heyneman indices and write

$$(M \otimes_A \Delta_{\mathcal{C}}) \circ \varrho^M(m) = (\varrho^M \otimes_A \mathcal{C}) \circ \varrho^M(m) = \sum m_{(0)} \otimes_A m_{(1)} \otimes_A m_{(2)}.$$

4.3. The image of the coaction. If (M, ϱ^M) is a right $\mathcal{C}$-comodule, then $(M \otimes_A \mathcal{C}, M \otimes_A \Delta_{\mathcal{C}})$ is also a right $\mathcal{C}$-comodule. Furthermore, ϱ^M is a $\mathcal{C}$-comodule map. Since ϱ^M has a retraction $M \otimes_A \varepsilon_{\mathcal{C}}$ (i.e. $(M \otimes_A \varepsilon_{\mathcal{C}}) \circ \varrho^M = M$), ϱ^M is a monomorphism of right $\mathcal{C}$-comodules. Thus

$$(M, \varrho^M) \simeq (\varrho^M(M), M \otimes_A \Delta_{\mathcal{C}} \mid_{\varrho^M(M)})$$

as $\mathcal{C}$-comodules.

4.4. Factorizable morphisms of comodules. Let $(M, \varrho^M), (N, \varrho^N), (P, \varrho^P)$ be right $\mathcal{C}$-comodules and let $f : M \to N$, $g : N \to P$ be A-module maps such that $g \circ f$ is a comodule morphism. If f is colinear, then

$$\varrho^P \circ g \circ f = (g \circ f \otimes_A \mathcal{C}) \circ \varrho^M = (g \otimes_A \mathcal{C}) \circ \varrho^N \circ f.$$

Thus if f is an epimorphism of comodules, then g is a right $\mathcal{C}$-comodule map. In particular, if a $\mathcal{C}$-colinear map f has a k-linear (hence A-linear) inverse f^{-1}, then f^{-1} is necessarily a $\mathcal{C}$-colinear map (i.e. f is an isomorphism of comodules).

4.5. The regular bicomodule. $(\mathcal{C}, \Delta_\mathcal{C}, \Delta_\mathcal{C})$ is a $\mathcal{C}$-bicomodule, that is, $\mathcal{C}$ can be viewed as a right $\mathcal{C}$-comodule and as a left $\mathcal{C}$-comodule with coaction given by the coproduct $\Delta_\mathcal{C}$. This is known as a *regular $\mathcal{C}$-comodule*.

4.6. Comodules finitely generated and projective as modules. Let (P, ϱ^P) be a right $\mathcal{C}$-comodule. Assume that P is finitely generated and projective as a right A-module, and let $\{e^i \in P,\ \pi^i \in P^* = \mathrm{Hom}_A(P, A)\}$ be its finite dual basis. View P^* as a left A-module with the multiplication $(af)(x) = af(x)$. Then P^* is a left $\mathcal{C}$-comodule with the coaction

$${}^{P^*}\varrho(f) = \sum_i f(e^i{}_{(0)}) e^i{}_{(1)} \otimes_A \pi^i.$$

4.7. Comodules as coalgebras. In view of 3.5, $\mathcal{C}$ is a coring if and only if $- \otimes_A \mathcal{C} : \mathbf{M}_A \to \mathbf{M}_A$ is a comonad. The category of $- \otimes_A \mathcal{C}$-coalgebras is isomorphic to the category of right $\mathcal{C}$-comodules. Similarly, the category of left $\mathcal{C}$-comodules is isomorphic to the category of coalgebras of the comonad $\mathcal{C} \otimes_A - : {}_A\mathbf{M} \to {}_A\mathbf{M}$.

4.8. The defining adjunctions [106, Proposition 3.1]. Since an A-coring $\mathcal{C}$ is given by comonads, whose coalgebras coincide with comodules, with every A-coring $\mathcal{C}$, one can associate two pairs of adjoint functors between categories of A-modules and $\mathcal{C}$-comodules.

(1) The forgetful functor $(-)_A : \mathbf{M}^\mathcal{C} \to \mathbf{M}_A$ is a left adjoint of the tensor functor

$$- \otimes_A \mathcal{C} : \mathbf{M}_A \to \mathbf{M}^\mathcal{C}, \qquad N \mapsto (N \otimes_A \mathcal{C}, N \otimes_A \Delta_\mathcal{C}).$$

The unit and the counit of this adjunction are

$$\eta_{(M,\varrho^M)} = \varrho^M : M \to M \otimes_A \mathcal{C}, \quad \psi_N = N \otimes_A \varepsilon_\mathcal{C} : N \otimes_A \mathcal{C} \to N,$$

for all right $\mathcal{C}$-comodules (M, ϱ^M) and right A-modules N. Consequently, there is an isomorphism (natural in $N \in \mathbf{M}_A$, $(M, \varrho^M) \in \mathbf{M}^\mathcal{C}$),

$$\Omega_{M,N} : \mathrm{Hom}_A(M, N) \to \mathrm{Hom}^\mathcal{C}(M, N \otimes_A \mathcal{C}), \quad \Omega_{M,N}(g) = (g \otimes_A \mathcal{C}) \circ \varrho^M.$$

The inverse of $\Omega_{M,N}$ is $\Omega^{-1}_{M,N}(f) = (N \otimes_A \varepsilon_\mathcal{C}) \circ f$.

(2) The forgetful functor ${}_A(-) : {}^\mathcal{C}\mathbf{M} \to {}_A\mathbf{M}$ is a left adjoint of the tensor functor

$$\mathcal{C} \otimes_A - : {}_A\mathbf{M} \to {}^\mathcal{C}\mathbf{M}, \qquad N \mapsto (\mathcal{C} \otimes_A N, \Delta_\mathcal{C} \otimes_A N).$$

The unit and the counit of this adjunction are

$$\eta_{(M,{}^M\varrho)} = {}^M\varrho : M \to \mathcal{C} \otimes_A M, \quad \psi_N = \varepsilon_{\mathcal{C}} \otimes_A N : \mathcal{C} \otimes_A N \to N,$$

for all left $\mathcal{C}$-comodules $(M, {}^M\varrho)$ and left A-modules N. Consequently, there is an isomorphism (natural in $N \in {}_A\mathbf{M}$, $(M, {}^M\varrho) \in {}^{\mathcal{C}}\mathbf{M}$),

$$\Omega_{M,N} : {}_A\mathrm{Hom}(M, N) \to {}^{\mathcal{C}}\mathrm{Hom}(M, \mathcal{C} \otimes_A N), \quad \Omega_{M,N}(g) = (\mathcal{C} \otimes_A g) \circ {}^M\varrho.$$

The inverse of $\Omega_{M,N}$ is $\Omega^{-1}_{M,N}(f) = (\varepsilon_{\mathcal{C}} \otimes_A N) \circ f$.

4.9. Coinvariants. Let (M, ϱ^M) be a right $\mathcal{C}$-comodule. Given any right $\mathcal{C}$-comodule (N, ϱ^N), the k-module $\mathrm{Hom}^{\mathcal{C}}(M, N)$ is called the *coinvariants of N with respect to M* or *M-coinvariants of N*.

The M-coinvariants or, simply, *coinvariants* of M are thus equal to the endomorphism ring

$$S = \mathrm{Hom}^{\mathcal{C}}(M, M) = \mathrm{End}^{\mathcal{C}}(M).$$

The product in S is given by the composition of maps, that is, $ss'(m) = s(s'(m))$, for all $s, s' \in S$, $m \in M$. The endomorphism ring S acts on M from the left by the evaluation, that is, for all $s \in S$ and $m \in M$, $sm := s(m)$. By the construction of S, the coaction ϱ^M is left S-linear. Obviously, S is a subalgebra of the right A-module endomorphism ring $\mathrm{End}_A(M)$.

4.10. The coinvariants of the regular comodule. Let (M, ϱ^M) be a right $\mathcal{C}$-comodule. In view of the hom-tensor relations 4.8, the M-coinvariants of the regular $\mathcal{C}$-comodule $(\mathcal{C}, \Delta_{\mathcal{C}})$ are isomorphic to the right dual module M^*, that is,

$$\mathrm{Hom}^{\mathcal{C}}(M, \mathcal{C}) \simeq \mathrm{Hom}^{\mathcal{C}}(M, A \otimes_A \mathcal{C}) \simeq \mathrm{Hom}_A(M, A) = M^*.$$

In particular, the endomorphisms of the right regular $\mathcal{C}$-comodule $(\mathcal{C}, \Delta_{\mathcal{C}})$ are isomorphic to the right dual module $\mathcal{C}^*$. When the algebra structure of $\mathrm{End}^{\mathcal{C}}(\mathcal{C})$ is ported through this isomorphism to $\mathcal{C}^*$, it coincides with the *opposite* algebra structure of the dual ring $\mathcal{C}^*$ as described in 3.25.

4.11. The coinvariants adjunction. For a right $\mathcal{C}$-comodule (M, ϱ^M), let $S = \mathrm{End}^{\mathcal{C}}(M)$ be the ring of M-coinvariants of M; see 4.9. For any right $\mathcal{C}$-comodule (N, ϱ^N), the M-coinvariants of N are a right S-module with the multiplication $fs := f \circ s$. Thus the M-coinvariants define a functor

$$\mathrm{Hom}^{\mathcal{C}}(M, -) : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}_S.$$

Since this is the hom-functor, it has the left adjoint, the tensor functor

$$- \otimes_S M : \mathbf{M}_S \to \mathbf{M}^{\mathcal{C}}, \qquad X \mapsto (X \otimes_S M, X \otimes_S \varrho^M).$$

Explicitly, the counit and unit of adjunction are

$$\psi_N : \mathrm{Hom}^{\mathcal{C}}(M, N) \otimes_S M \to N, \qquad f \otimes_S m \mapsto f(m),$$
$$\eta_X : X \to \mathrm{Hom}^{\mathcal{C}}(M, X \otimes_S M), \qquad x \mapsto [m \mapsto x \otimes_S m];$$

see 2.9. In other words, for all $(N, \varrho^N) \in \mathbf{M}^{\mathcal{C}}$ and $X \in \mathbf{M}_S$, there is an isomorphism [natural in $(N, {}^N\varrho)$ and X]

$$\mathrm{Hom}^{\mathcal{C}}(X \otimes_S M, N) \to \mathrm{Hom}_S(X, \mathrm{Hom}^{\mathcal{C}}(M, N)), \ \delta \mapsto [x \mapsto \delta(x \otimes_S -)]\,,$$

with the inverse $\varphi \mapsto [x \otimes_S m \mapsto \varphi(x)(m)]$.

This adjoint pair of functors has been one of the main objects of studies in the coring theory in recent years. We refer to it as the *coinvariants adjunction*.

4.12. Group-like elements. The (regular) right (or left) A-module A is a right (or left) comodule of an A-coring if and only if there exists an element $g \in \mathcal{C}$ such that

$$\Delta_{\mathcal{C}}(g) = g \otimes_A g, \qquad \varepsilon_{\mathcal{C}}(g) = 1_A.$$

An element g satisfying the above properties is called a *group-like element*. Given a group-like element g, a right $\mathcal{C}$-coaction on A is defined by

$$\varrho^A : A \to A \otimes_A \mathcal{C} \simeq \mathcal{C}, \qquad a \mapsto ga,$$

while a left $\mathcal{C}$-coaction is ${}^A\varrho : A \to \mathcal{C} \otimes_A A \simeq \mathcal{C}$, $a \mapsto ag$.

For example, the Sweedler coring $\mathcal{C} = A \otimes_B A$ associated to an algebra map $B \to A$ (cf. 3.10) has a group-like element $1_A \otimes_B 1_A$. Thus, A is a right $\mathcal{C}$-comodule with the coaction $a \mapsto 1_A \otimes_B a$ and a left $\mathcal{C}$-comodule with the coaction $a \mapsto a \otimes_B 1_A$.

4.13. Group-like elements and differential graded algebras. Any group-like element g in an A-coring $\mathcal{C}$ gives rise to a semifree differential graded algebra. This construction provides one with the inverse to the construction of corings in 3.16. Since a coring associated to a semifree differential graded algebra has a group-like element (the "indeterminate" g in 3.16), this establishes a bijective correspondence between corings with a group-like element and semifree differential graded algebras [162].

Given an A-coring $\mathcal{C}$ with a group-like element g, one defines a semifree differential graded algebra Ω as follows:

$$\Omega^1 = \ker \varepsilon_{\mathcal{C}}, \qquad \Omega^{n+1} = \Omega^1 \otimes_A \Omega^n,$$

and the multiplication is given by the tensor product [i.e. Ω is the tensor algebra $\Omega = T_A(\ker \varepsilon_{\mathcal{C}})$]. The differential is defined by

$$d(a) = ga - ag,$$

for all $a \in A$, and for all $c^1 \otimes_A \cdots \otimes_A c^n \in (\ker \varepsilon_{\mathcal{C}})^{\otimes_A n}$,

$$d(c^1 \otimes_A \cdots \otimes_A c^n) = g \otimes_A c^1 \otimes_A \cdots \otimes_A c^n + (-1)^{n+1} c^1 \otimes_A \cdots \otimes_A c^n \otimes_A g + \sum_{i=1}^{n} (-1)^i c^1 \otimes_A \cdots \otimes_A c^{i-1} \otimes_A \Delta_{\mathcal{C}}(c^i) \otimes_A c^{i+1} \otimes_A \cdots \otimes_A c^n.$$

The differential algebra associated to the Sweedler coring $A \otimes_B A$ and a group-like element $1_A \otimes_B 1_A$ coincides with the algebra of *B-relative universal differential forms* on A (cf. [80]).

4.14. The g-coinvariants of A. Suppose that A is a right (or left) $\mathcal{C}$-comodule with the coaction given by a group-like element g. Let (M, ϱ^M) be a right $\mathcal{C}$-comodule. Then the isomorphism of k-modules

$$M \simeq \mathrm{Hom}_A(A, M), \qquad m \mapsto [a \mapsto ma],$$

allows one to identify A-coinvariants of M with the k-module

$$M_g = \{m \in M \mid \varrho^M(m) = m \otimes_A g\}.$$

M_g is called the *g-coinvariants of* M. In particular, the g-coinvariants of A are the subalgebra A_g of A defined by $A_g = \{b \in A \mid gb = bg\}$. As A_g is isomorphic to $\mathrm{End}^{\mathcal{C}}(A)$ and M_g is isomorphic to $\mathrm{Hom}^{\mathcal{C}}(A, M)$, the g-coinvariants M_g form a right A_g-module. M is a left A_g-module and ϱ^M is a left A_g-linear map.

4.15. The cotensor product. Let (M, ϱ^M) be a right comodule and let $(N, {}^N\varrho)$ be a left comodule over an A-coring $\mathcal{C}$. The *cotensor product* $M\Box_{\mathcal{C}}N$ is the k-module defined by the equalizer

$$M\Box_{\mathcal{C}}N \longrightarrow M \otimes_A N \underset{M \otimes_A {}^N\varrho}{\overset{\varrho^M \otimes_A N}{\rightrightarrows}} M \otimes_A \mathcal{C} \otimes_A N .$$

Put differently, $M\Box_{\mathcal{C}}N$ is the kernel of the k-linear map $\omega_{M,N} = \varrho^M \otimes_A N - M \otimes_A {}^N\varrho$. Given a right A-module M', left A-module N', and module maps $f : M \to M'$, $g : N \to N'$, we write $f\Box_{\mathcal{C}}g$ for the restriction of $f \otimes_A g$ to $M\Box_{\mathcal{C}}N$. In case M', N' are comodules and f, g are comodule maps, $f\Box_{\mathcal{C}}g$ factors through the canonical inclusion $M'\Box_{\mathcal{C}}N' \subseteq M' \otimes_A N'$ and hence is considered as a map $M\Box_{\mathcal{C}}N \to M'\Box_{\mathcal{C}}N'$.

If either (M, ϱ^M) or $(N, {}^N\varrho)$ is the regular comodule $(\mathcal{C}, \Delta_{\mathcal{C}})$, then

$$M\Box_{\mathcal{C}}\mathcal{C} = \varrho^M(M) \simeq M, \quad \mathcal{C}\Box_{\mathcal{C}}N = {}^N\varrho(N) \simeq N,$$

with the isomorphisms given by $M\Box_{\mathcal{C}}\varepsilon_{\mathcal{C}}$ and $\varepsilon_{\mathcal{C}}\Box_{\mathcal{C}}N$, respectively.

4.16. Coinvariants and the cotensor product. Take any right $\mathcal{C}$-comodules (M, ϱ^M) and (N, ϱ^N) and assume that M is finitely generated and projective as a right A-module. Let $M^* = \mathrm{Hom}_A(M, A)$ be the dual module, and let $(M^*, {}^{M^*}\varrho)$ be the

induced left $\mathcal{C}$-comodule as in 4.6. Then the k-module isomorphism

$$N \otimes_A M^* \simeq \mathrm{Hom}_A(M, N), \qquad n \otimes_A f \mapsto [m \mapsto nf(m)],$$

restricts to the isomorphism

$$N\Box_{\mathcal{C}} M^* \simeq \mathrm{Hom}^{\mathcal{C}}(M, N).$$

Thus, in particular, in case $(N, \varrho^N) = (M, \varrho^M)$, the coinvariants $S = \mathrm{End}^{\mathcal{C}}(M)$ of M can be identified with the cotensor product $M\Box_{\mathcal{C}} M^*$.

4.17. The tensor–cotensor relations. In general, if (M, ϱ^M) is a right $\mathcal{C}$-comodule and $(N, \varrho^N, {}^N\varrho)$ is a $(\mathcal{C}, \mathcal{D})$-bicomodule, the coaction $M \otimes_A \varrho^N$ on $M \otimes_A N$ does not restrict to a coaction in $M\Box_{\mathcal{C}} N$. It does restrict if the map $\omega_{M,N} = \varrho^M \otimes_A N - M \otimes_A {}^N\varrho$ is $\mathcal{D} \otimes_B \mathcal{D}$-pure as a right B-module map.

The cotensor product is not an associative operation and it does not commute with the tensor product. Given algebras A and B and an A-coring $\mathcal{C}$, if (M, ϱ^M) is a right $\mathcal{C}$-comodule, $(N, {}^N\varrho)$ is a $(\mathcal{C}, B)$-bicomodule, and V is a left B-module, then there is a k-linear map

$$\tau_{M,N,V} : \quad (M\Box_{\mathcal{C}} N) \otimes_B V \to M\Box_{\mathcal{C}}(N \otimes_B V),$$

which is not an isomorphism unless $\omega_{M,N} = \varrho^M \otimes_A N - M \otimes_A {}^N\varrho$ is V-pure. Here, $N \otimes_B V$ is a left $\mathcal{C}$-comodule by ${}^N\varrho \otimes_B V$. Thus

$$(M\Box_{\mathcal{C}} N) \otimes_B V \simeq M\Box_{\mathcal{C}}(N \otimes_B V)$$

if, for example, V is a flat left B-module or (M, ϱ^M) is a $(\mathcal{C}, A)$-injective comodule [see 5.1 for the definition of $(\mathcal{C}, A)$-injectivity].

For a detailed discussion of the cotensor product see, [101] and [53] (including erratum).

4.18. Coinvariants of a coring morphism [133]. Let $\phi : \mathcal{D} \to \mathcal{C}$ be a morphism of A-corings. Then $(\mathcal{D}, (\mathcal{D} \otimes_A \phi) \circ \Delta_{\mathcal{D}}, (\phi \otimes_A \mathcal{D}) \circ \Delta_{\mathcal{D}})$ is a $\mathcal{C}$-bicomodule. Set $B = {}^{\mathcal{C}}\mathrm{Hom}^{\mathcal{C}}(\mathcal{D}, \mathcal{C})$. B is an algebra with unit ϕ and product, for all $b, b' \in B, d \in \mathcal{D}$,

$$bb'(d) = \sum \varepsilon_{\mathcal{C}}(b(d_{(1)}))b'(d_{(2)}) = \sum b(d_{(1)})\varepsilon_{\mathcal{C}}(b'(d_{(2)})).$$

The algebra B is called an *algebra of coinvariants of the A-coring morphism ϕ*.

Given an A-coring $\mathcal{C}$ and a right $\mathcal{C}$-comodule (P, ϱ^P) that is finitely generated projective as a right A-module, consider the comatrix coring $\mathcal{D} = P^* \otimes P$; see 3.11. Then

$$\phi : \mathcal{D} = P^* \otimes P \to \mathcal{C}, \qquad f \otimes x \mapsto \sum f(x_{(0)})x_{(1)},$$

is an A-coring morphism. The induced (by ϕ) $\mathcal{C}$-bicomodule structure on $\mathcal{D}$ comes out as $(\mathcal{D}, P^* \otimes \varrho^P, {}^{P^*}\varrho \otimes P)$, where ${}^{P^*}\varrho$ is a left $\mathcal{C}$-coaction as in 4.6. Since P is a finitely generated and projective A-module, ${}_A\mathrm{Hom}_A(P^* \otimes P, A) \simeq \mathrm{End}_A(P)$. This restricts to an isomorphism of algebras

$${}^{\mathcal{C}}\mathrm{Hom}^{\mathcal{C}}(P^* \otimes P, \mathcal{C}) \simeq \mathrm{End}^{\mathcal{C}}(P),$$

(note that ${}^{\mathcal{C}}\mathrm{Hom}^{\mathcal{C}}(P^* \otimes P, \mathcal{C}) \subseteq {}_A\mathrm{Hom}_A(P^* \otimes P, A)$ by the defining adjunction). Thus, the coinvariants of the A-coring morphism $\phi : P^* \otimes P \to \mathcal{C}$ coincide with the P-coinvariants of P.

4.2. *Examples*

Here, we describe comodules of examples of corings discussed in Section 3.2. In this way, we discover objects (and categories) studied in the precoring literature by different means.

4.19. Comodules of the trivial A-coring. Since both coproduct and counit of the trivial A-coring A are identity maps to be a right comodule of a trivial A-coring, A is the same as to be a right A-module. Thus $\mathbf{M}^A = \mathbf{M}_A$ and ${}^A\mathbf{M} = {}_A\mathbf{M}$.

In particular, as observed in 4.9, given an A-coring $\mathcal{C}$ and a right $\mathcal{C}$-comodule (M, ϱ^M), the coaction ϱ^M is left linear over the coinvariants $S = \mathrm{End}^{\mathcal{C}}(M)$. Viewing S as a trivial S-coring, we can thus say that M is an $(S, \mathcal{C})$-bicomodule.

4.20. Descent data. Let $\mathcal{C} = A \otimes_B A$ be the Sweedler coring associated to an algebra map $B \to A$; see 3.10. Right $\mathcal{C}$-comodules are in bijective correspondence with *descent data* (cf. [104]). The latter are defined as pairs (M, f), where M is a right A-module and $f : M \to M \otimes_B A$ is a right A-module morphism such that writing $f(m) = \sum_i m_i \otimes_B a_i$, for any $m \in M$,

$$\sum_i f(m_i) \otimes_B a_i = \sum_i m_i \otimes_B 1_A \otimes_B a_i, \qquad \sum_i m_i a_i = m.$$

This bijective correspondence is obtained simply by identifying $M \otimes_A A \otimes_B A$ with $M \otimes_B A$.

A morphism of descent data $(M, f) \to (M', f')$ is a right A-module map $\phi : M \to M'$ such that $f' \circ \phi = (\phi \otimes_B A) \circ f$. The category of descent data is denoted by $\mathbf{Desc}(A|B)$. This category has been introduced in [125] (for commutative rings) and [74] (for general rings). The identification of right $\mathcal{C}$-comodules with descent data is compatible with the definitions of morphisms in respective categories. Thus there is an isomorphism of categories $\mathbf{M}^{A\otimes_B A} \equiv \mathbf{Desc}(A|B)$; see [37], [68, Section 4.8].

4.21. Comodules in a comatrix coring context. Consider a comatrix coring context $(A, B, N, M, \sigma, \tau)$ and the associated coring $\mathcal{C} = N \otimes_B M$; see 3.13, 3.14. Recall that B is a firm (nonunital) algebra, and write $\nabla_M : M \to B \otimes_B M$ for the inverse of the multiplication map $B \otimes_B M \to M$ and $\nabla_N : N \to N \otimes_B B$ for the inverse of the multiplication map $N \otimes_B B \to B$. Then M is a right $\mathcal{C}$-comodule with the coaction

$$\varrho^M : M \to M \otimes_A N \otimes_B M, \qquad \varrho^M = (\tau \otimes_B M) \circ \nabla_M,$$

and N is a left $\mathcal{C}$-comodule with the coaction

$${}^N\varrho : N \to N \otimes_B M \otimes_A N, \qquad {}^N\varrho = (N \otimes_B \tau) \circ \nabla_N.$$

4.22. Comodules and (finite) comatrix corings. A comatrix A-coring associated to a (B, A)-bimodule P, finitely generated and projective as a right A-module (cf. 3.11), is a special case of a coring associated to a comatrix coring context. Thus both P and P^* are $\mathcal{C}$-comodules, with the right and left coactions, respectively,

$$\varrho^P : P \mapsto P \otimes_A P^* \otimes_B P, \qquad x \mapsto e \otimes_B x,$$

$$^{P^*}\varrho : P^* \mapsto P^* \otimes_B P \otimes_A P^*, \qquad \pi \mapsto \pi \otimes_A e,$$

where e is the dual basis element.

4.23. Flat connections [38]. Let $\Omega = \bigoplus_{n=0}^{\infty} \Omega^n$ be a semifree differential algebra with a differential $d : \Omega^\bullet \to \Omega^{\bullet+1}$. Set $A = \Omega^0$, and let $\mathcal{C} = Ag \oplus \Omega^1$ be the associated A-coring; see 3.16. Take a right A-module M. Any k-linear map $\varrho^M : M \to M \otimes_A \mathcal{C} = M \otimes_A Ag \oplus M \otimes_A \Omega^1$ such that $(M \otimes_A \varepsilon_{\mathcal{C}}) \circ \varrho^M = M$ can be identified with a map $\nabla : M \to M \otimes_A \Omega^1$ by

$$\varrho^M(m) = m \otimes_A g + \nabla(m).$$

The map ϱ^M is a right A-module map if and only if, for all $m \in M$, $a \in A$,

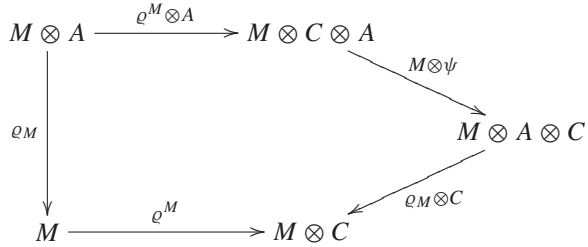
$$\nabla(ma) = \nabla(m)a + md(a).$$

A map $\nabla : M \to M \otimes_A \Omega^1$ satisfying this condition is known as the *connection* in M with respect to Ω (cf. [21, 2.8 Definition], [76]). The map ϱ^M is coassociative if and only if

$$(\nabla \otimes_A \Omega^1) \circ \nabla + (M \otimes_A d) \circ \nabla = 0.$$

The map $F := (\nabla \otimes_A \Omega^1) \circ \nabla + (M \otimes_A d) \circ \nabla$ is known as a *curvature* of ∇, and a connection with vanishing curvature is called a *flat connection*. Thus, right $\mathcal{C}$-comodules are in bijective correspondence with modules with flat connections with respect to Ω.

4.24. Entwined modules [36] [37, Proposition 2.2]. Let $(A, C)_\psi$ be an entwining structure and let $\mathcal{C} = A \otimes C$ be the associated A-coring; see 3.20, 3.21. Using the identification $M \otimes_A \mathcal{C} = M \otimes_A A \otimes C \simeq M \otimes C$, one identifies right $\mathcal{C}$-comodules with triples $(M, \varrho_M, \varrho^M)$, where (M, ϱ_M) is a right A-module (with A-multiplication $\varrho_M : M \otimes A \to M$) and (M, ϱ^M) is a right C-comodule such that



is a commutative diagram. Any such triple $(M, \varrho_M, \varrho^M)$ is called an *entwined module*. A morphism of entwined modules is a right A-linear map, which is also right C-colinear. Entwined modules and their morphisms form a category denoted by $\mathbf{M}_A^C(\psi)$. Again, through the identification $M \otimes_A \mathcal{C} \simeq M \otimes C$, morphisms of entwined modules can be identified with right $\mathcal{C}$-colinear maps. Hence $\mathbf{M}^{\mathcal{C}} \equiv \mathbf{M}_A^C(\psi)$.

4.25. Weak entwined modules [56], [37, Proposition 2.3]. Let (A, C, ψ) be a weak entwining structure, and let $\mathcal{C}$ be the associated A-coring; see 3.22, 3.23. By similar arguments to those in 4.24, one identifies right $\mathcal{C}$-comodules with right weak entwined modules. The latter are defined as right A-modules and right C-comodules satisfying the same compatibility condition as in the diagram in 4.24.

4.3. *The category of comodules*

In this section, we discuss properties of the category of comodules, and in particular, we describe how properties of $\mathcal{C}$ as an A-module affect the structure of the category $\mathbf{M}^{\mathcal{C}}$. Although many notions appearing in this section are explained in Section 2, some additional background in category theory might be required to understand some of the topics discussed here.

4.26. Comodules form a k-linear category. Since every hom-set $\mathrm{Hom}^{\mathcal{C}}(M, N)$ is a k-module and the composition of morphisms in $\mathbf{M}^{\mathcal{C}}$ is k-linear, the category $\mathbf{M}^{\mathcal{C}}$ is a k-linear (hence also preadditive) category.

4.27. Cokernels. Consider any morphism $f : (M, \varrho^M) \to (N, \varrho^N)$ in $\mathbf{M}^{\mathcal{C}}$. Since f is a morphism of right A-modules, it has a cokernel $g : N \to L$ in the category of right A-modules. Furthermore, f is a morphism of $\mathcal{C}$-comodules; hence, $g \circ f = 0$ implies $(g \otimes_A \mathcal{C}) \circ \varrho^N \circ f = 0$. By the universality of cokernels, there is a unique A-module map $\varrho^L : L \to L \otimes_A \mathcal{C}$ such that $\varrho^L \circ g = (g \otimes_A \mathcal{C}) \circ \varrho^N$. The coassociativity and counitality of ϱ^N (together with the fact that g is surjective) imply the coassociativity and counitality of ϱ^L. Hence (L, ϱ^L) is a right $\mathcal{C}$-comodule. By construction, g is a comodule map. By the universality of cokernels in $\mathbf{M}_A$, for any $h : (N, \varrho^N) \to (P, \varrho^P)$ such that $h \circ f = 0$, there is a unique A-module map $h' : L \to P$ such that $h' \circ g = h$. Since both h and g are comodule maps and g is an epimorphism, also h' is a comodule map by 4.4. This proves the universality of (L, ϱ^L). Hence (L, ϱ^L) is a cokernel of the right $\mathcal{C}$-comodule morphism f. Thus the category $\mathbf{M}^{\mathcal{C}}$ has cokernels.

4.28. Kernels. Consider any morphism $f : (M, \varrho^M) \to (N, \varrho^N)$ in $\mathbf{M}^{\mathcal{C}}$. Since f is a morphism of right A-modules, it has a kernel $h : K \to M$ in the category of right A-modules. By tensoring the defining sequence of the kernel with $\mathcal{C}$, one obtains the following commutative diagram

$$\begin{array}{ccccccc}
0 & \longrightarrow & K & \xrightarrow{h} & M & \xrightarrow{f} & N \\
 & & & & \downarrow{\scriptstyle \varrho^M} & & \downarrow{\scriptstyle \varrho^N} \\
0 & \longrightarrow & K \otimes_A \mathcal{C} & \xrightarrow{h \otimes_A \mathcal{C}} & M \otimes_A \mathcal{C} & \xrightarrow{f \otimes_A \mathcal{C}} & N \otimes_A \mathcal{C}\,.
\end{array}$$

The top sequence is exact by the definition of a kernel of an A-module morphism. However, since the tensor functor does not necessarily preserve monomorphisms, the bottom sequence need not be exact. The exactness of the bottom sequence is equivalent to saying that f is $\mathcal{C}$-*pure* as a right A-module morphism. Thus, if the bottom sequence is exact, the diagram can be extended commutatively to give a map $\varrho^K : K \to K \otimes_A \mathcal{C}$; see 4.27. To check that the map ϱ^K is coassociative, one needs the exactness of the following sequence:

$$0 \longrightarrow K \otimes_A \mathcal{C} \otimes_A \mathcal{C} \xrightarrow{h\otimes_A \mathcal{C}\otimes_A \mathcal{C}} M \otimes_A \mathcal{C} \otimes_A \mathcal{C} \xrightarrow{f\otimes_A \mathcal{C}\otimes_A \mathcal{C}} N \otimes_A \mathcal{C} \otimes_A \mathcal{C}.$$

This means that one needs to assume that f is $\mathcal{C} \otimes_A \mathcal{C}$-pure as a right A-module map. By the counitality of the coproduct, this implies that f is $\mathcal{C}$-pure. Thus, if f is $\mathcal{C} \otimes_A \mathcal{C}$-pure, then (K, ϱ^K) is a right $\mathcal{C}$-comodule. The pair (K, ϱ^K) is the kernel of the comodule morphism f.

If $\mathcal{C}$ is a flat left A-module, then the tensor functor $- \otimes_A \mathcal{C} : \mathbf{M}_A \to \mathbf{M}_k$ is exact. This implies that every right A-module map is $\mathcal{C} \otimes_A \mathcal{C}$-pure. Consequently, every $\mathcal{C}$-comodule morphism has a kernel. Thus, the category of *right* $\mathcal{C}$-comodules has kernels provided $\mathcal{C}$ is flat as a *left* A-module.

4.29. Coproducts in the category of comodules. Let $\{(M_\lambda, \varrho^{M_\lambda})\}_{\lambda\in\Lambda}$ be a family of right $\mathcal{C}$-comodules. Consider the right A-module $M = \bigoplus_{\lambda\in\Lambda} M_\lambda$ in $\mathbf{M}_A$ and define the right A-module map

$$\varrho^M : \bigoplus_{\lambda\in\Lambda} M_\lambda = M \to M \otimes_A \mathcal{C} = \bigoplus_{\lambda\in\Lambda} (M_\lambda \otimes_A \mathcal{C}), \qquad \varrho^M \mid_{M_\lambda} = \varrho^{M_\lambda}.$$

Since each of the ϱ^{M_λ} is coassociative, so is their coproduct $\varrho^M : M \to M \otimes_A \mathcal{C}$. Hence, (M, ϱ^M) is a right $\mathcal{C}$-comodule such that the inclusions $M_\lambda \to M$ are right $\mathcal{C}$-comodule morphisms. This is the coproduct of the family of right $\mathcal{C}$-comodules $\{(M_\lambda, \varrho^{M_\lambda})\}_{\lambda\in\Lambda}$.

4.30. Comodules of a flat coring [96, Proposition 1.2]. As a consequence of the defining adjunctions 4.8, the forgetful functor $(-)_A : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}_A$ has a right adjoint; hence it preserves epimorphisms, that is, every epimorphism of right $\mathcal{C}$-comodules is a right A-linear epimorphism. On the other hand, $(-)_A : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}_A$ preserves monomorphisms if and only if $\mathcal{C}$ is flat as a left A-module. This leads to the equivalence of the following statements:

(a) $\mathcal{C}$ is flat as a left A-module;
(b) $\mathbf{M}^{\mathcal{C}}$ is an Abelian category and the forgetful functor $(-)_A : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}_A$ is exact (i.e. it preserves exact sequences);
(c) $\mathbf{M}^{\mathcal{C}}$ is a Grothendieck category and $(-)_A : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}_A$ is an exact functor.

4.31. A Grothendieck category of comodules over a nonflat coring [96, Example 1.1]. The following two examples show that flatness of $\mathcal{C}$ as a left A-module is not necessary for the category of comodules to be Abelian (Grothendieck).

(1) Take k-algebras R, S, and an (R, S)-bimodule M, and let

$$A = \begin{pmatrix} R & M \\ 0 & S \end{pmatrix}, \qquad \mathcal{C} = \begin{pmatrix} R & M \\ 0 & 0 \end{pmatrix}.$$

A is a (matrix) algebra, and $\mathcal{C}$ is an ideal in A. $\mathcal{C}$ has a coproduct given by the isomorphism $\mathcal{C} \simeq \mathcal{C} \otimes_A \mathcal{C}$ and a counit given by the inclusion $\mathcal{C} \subseteq A$. Right $\mathcal{C}$-comodules can be identified with triples (X, Y, θ), where X is a right R-module, Y is a right S-module, and $\theta : X \otimes_R M \to Y$ is an isomorphism of right S-modules. The functor $(X, Y, \theta) \to X$ establishes an equivalence of categories $\mathbf{M}^{\mathcal{C}}$ and $\mathbf{M}_R$. The latter is a Grothendieck category hence so is the former. Yet $\mathcal{C}$ is not necessarily a flat left A-module.

(2) Even more explicit example of a nonflat coring, whose category of comodules is Abelian (Grothendieck), is constructed in [81, Remark 2.2.9] (cf. [34, Exercise 4.8.13]), by using the correspondence between comodules of the Sweedler coring and the category of descent data; see 4.20. Consider a monomorphism of commutative rings $f : B \to A$. By the Joyal–Tierney theorem (see [137, Theorem] for the commutative case and [55, Proposition 2.3], [59, Theorem 2.7] for noncommutative versions), the functor $- \otimes_B A : \mathbf{M}_B \to \mathbf{Desc}(A|B) \equiv \mathbf{M}^{A \otimes_B A}$ is an equivalence if and only if f is a pure morphism of B-modules (i.e. $N \otimes_B f$ is a monomorphism for all B-modules N). Take $B = \mathbb{Z}$, $A = \mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, and $f : n \mapsto (n, [n])$. Then A is not a flat B-module (the map $i \otimes_B A$, where $i : 2\mathbb{Z} \to \mathbb{Z}$ is the inclusion, has a nontrivial kernel), and hence $A \otimes_B A$ is not a flat left A-module. Yet, f is pure as a $\mathbb{Z}$-module map, as it has a $\mathbb{Z}$-module retraction $(n, [m]) \mapsto n$. Therefore, $\mathbf{M}^{A \otimes_B A}$ is equivalent to a Grothendieck category $\mathbf{M}_{\mathbb{Z}}$, hence, $\mathbf{M}^{A \otimes_B A}$ is a Grothendieck category.

4.32. Monomorphisms and products. As a consequence of the defining adjunctions 4.8, the functor $- \otimes_A \mathcal{C} : \mathbf{M}_A \to \mathbf{M}^{\mathcal{C}}$ has a left adjoint; hence, it preserves monomorphisms, that is, if f is a monomorphism of right A-modules, then $f \otimes_A \mathcal{C}$ is a monomorphism of right $\mathcal{C}$-comodules.

For the same reason, given a family of right A-modules $\{M_\lambda\}_{\lambda \in \Lambda}$, the comodule $((\prod_\Lambda M_\lambda) \otimes_A \mathcal{C}, (\prod_\Lambda M_\lambda) \otimes_A \Delta_{\mathcal{C}})$ is the product of the comodules $(M_\lambda \otimes_A \mathcal{C}, M_\lambda \otimes_A \Delta_{\mathcal{C}})$.

4.33. Comodules and modules of the dual ring. Let $\mathcal{C}$ be an A-coring and let $B = {}^*\mathcal{C}$ be the dual ring defined in 3.25(2). Every right $\mathcal{C}$-comodule (M, ϱ^M) is a left B-module with the multiplication, for all $b \in B$, $m \in M$,

$$bm = \sum m_{(0)} b(m_{(1)}).$$

Furthermore, every right $\mathcal{C}$-comodule map is a left B-module map. This defines a functor $\mathbf{M}^{\mathcal{C}} \to {}_B\mathbf{M}$.

Clearly, if two right $\mathcal{C}$-comodule maps differ when viewed as left B-module maps, they differ as right $\mathcal{C}$-comodule maps. Hence the functor $\mathbf{M}^{\mathcal{C}} \to {}_B\mathbf{M}$ is faithful.

4.34. Comodules of a locally projective coring. Following [99, 204], a left A-module P is said to be *locally projective* if every finitely generated submodule of P is a finitely generated projective module. Equivalently, P is a locally projective left A-module if, for any diagram of left A-modules with exact rows,

$$\begin{array}{ccccccc} 0 & \longrightarrow & F & \xrightarrow{i} & P & & \\ & & & & \downarrow{\scriptstyle g} & & \\ & & M & \xrightarrow{f} & N & \longrightarrow & 0, \end{array}$$

where F is finitely generated, there exists $h : P \to M$ such that $g \circ i = f \circ h \circ i$. Equivalently, P is a locally projective left A-module if and only if, for any left A-module N, the map

$$\alpha_N : N \otimes_A P \to \mathrm{Hom}_A({}_A\mathrm{Hom}(P, A), N), \qquad n \otimes_A x \mapsto [f \mapsto nf(x)],$$

is injective; see [204, Theorem 2.1]. If P is a locally projective module, then P is a flat module. Clearly, if P is projective as a left A-module, then P is locally projective. Thus, local projectivity is a property of a module located in-between flatness and projectivity. For a detailed discussion of local projectivity, the reader is referred to [189]. In relation to corings, one refers to an A-coring $\mathcal{C}$ that is a locally projective left A-module as an A-coring satisfying *the left α-condition*; see [2] (these notions arise from studies of rationality of coalgebras over commutative rings [7, 100]).

An A-coring $\mathcal{C}$ satisfies the left α-condition if and only if the category of right $\mathcal{C}$-comodules is a full subcategory of the category of left modules over the dual ring $B = {}^*\mathcal{C}$; see [197] (cf. [53, 19.3]). In other words, $\mathcal{C}$ is a locally projective module if and only if the functor $\mathbf{M}^{\mathcal{C}} \to {}_B\mathbf{M}$ is full, that is, $\mathrm{Hom}^{\mathcal{C}}(M, N) = {}_B\mathrm{Hom}(M, N)$, for any comodules $(M, \varrho^M), (N, \varrho^N) \in \mathbf{M}^{\mathcal{C}}$.

4.35. Rational modules and the rational functor. Let $\mathcal{C}$ be an A-coring, which is locally projective as a left A-module, and let $B = {}^*\mathcal{C}$. For any left B-module M, $r \in M$ is called a *rational element* if there exists an element $\sum_i m_i \otimes_A c_i \in M \otimes_A \mathcal{C}$ such that for all $b \in B$, $br = \sum_i m_i b(c_i)$. Since $\alpha_M : M \otimes_A \mathcal{C} \to \mathrm{Hom}_A(B, M)$ in 4.34 is assumed to be injective, such an element $\sum_i m_i \otimes_A c_i$ is uniquely determined. The set of all rational elements of M is a (left B-, hence right A-) submodule of M, which is called the *rational submodule* of M, and is denoted by $\mathrm{Rat}^{\mathcal{C}}(M)$. Since, for any r the corresponding $\sum_i m_i \otimes_A c_i$ is uniquely determined, there is a right A-linear map

$$\mathrm{Rat}^{\mathcal{C}}(M) \to M \otimes_A \mathcal{C}, \qquad r \mapsto \sum_i m_i \otimes_A c_i.$$

One can check that all the m_i are in $\mathrm{Rat}^{\mathcal{C}}(M)$ and that this map is coassociative and counital, thus giving rise to a coaction $\varrho^{\mathrm{Rat}^{\mathcal{C}}(M)}$ on $\mathrm{Rat}^{\mathcal{C}}(M)$. This defines a functor:

$$\begin{aligned} \mathrm{Rat}^{\mathcal{C}} : {}_B\mathbf{M} \to \mathbf{M}^{\mathcal{C}}, \qquad M &\mapsto (\mathrm{Rat}^{\mathcal{C}}(M), \varrho^{\mathrm{Rat}^{\mathcal{C}}(M)}), \\ f : M \to N &\mapsto f|_{\mathrm{Rat}^{\mathcal{C}}(M)} : \mathrm{Rat}^{\mathcal{C}}(M) \to \mathrm{Rat}^{\mathcal{C}}(N), \end{aligned}$$

known as the *rational functor*.

$\mathrm{Rat}^{\mathcal{C}}(M) = M$ if and only if the left B action on M is obtained from a $\mathcal{C}$-coaction ϱ^M, that is, $(M, \varrho^M) \in \mathbf{M}^{\mathcal{C}}$, and $bm = \sum m_{(0)} b(m_{(1)})$. The equality $\mathrm{Rat}^{\mathcal{C}}(M) = M$ holds for all left B-modules M if and only if $\mathcal{C}$ is finitely generated as a left A-module (and thus is finitely generated and projective).

For more details on rational modules including their description in terms of *pairings* and their topological aspects, the reader is referred to [2, 62, 69].

4.36. Comodules as the Wisbauer σ-category [53, 19.3]. An A-coring $\mathcal{C}$ is locally projective as a left A-module if and only if $\mathbf{M}^{\mathcal{C}}$ is the smallest full Grothendieck subcategory of ${}_B\mathbf{M}$, which contains $\mathcal{C}$. Such a category is denoted by $\sigma[\mathcal{C}]$ and is called *the Wisbauer σ-category*. Equivalently, $\sigma[\mathcal{C}]$ is characterized as the full subcategory of ${}_B\mathbf{M}$ consisting of all submodules of modules M, for which there is an epimorphism $\mathcal{C}^{(\Lambda)} \to M$ (i.e. $\sigma[\mathcal{C}]$ consists of $\mathcal{C}$-*subgenerated modules*). The Wisbauer σ-category $\sigma[\mathcal{C}]$ contains large amount of module-theoretic information about $\mathcal{C}$ as a left B-module. For detailed studies of σ-categories, we refer to [195].

4.37. Comodules of a finitely generated and projective coring. Let $\mathcal{C}$ be an A-coring and $B = {}^*\mathcal{C}$ the left-dual ring. The functor $\mathbf{M}^{\mathcal{C}} \to {}_B\mathbf{M}$ described in 4.33 can be combined with the defining functor $- \otimes_A \mathcal{C} : \mathbf{M}_A \to \mathbf{M}^{\mathcal{C}}$ to give the functor $- \otimes_A \mathcal{C} : \mathbf{M}_A \to {}_B\mathbf{M}$. On the other hand, since B is an A^{op}-ring, there is a restriction of scalar functors $(-)_{B/A} : \mathbf{M}_B \to \mathbf{M}_A$.

LEMMA. *The functor $- \otimes_A \mathcal{C} : \mathbf{M}_A \to {}_B\mathbf{M}$ is a right adjoint of the restriction of scalar functors $(-)_{B/A} : {}_B\mathbf{M} \to \mathbf{M}_A$ if and only if $\mathcal{C}$ is a finitely generated and projective left A-module.*

Sketch of proof. If $- \otimes_A \mathcal{C}$ is a right adjoint of $(-)_{B/A}$, then the dual basis element for $\mathcal{C}$ is $\eta_B(1_B) \in B \otimes_A \mathcal{C}$, where η is the unit of the adjunction. Conversely, if $\{c_i \in C, b_i \in B\}$ is a dual basis for $\mathcal{C}$, then the unit of the adjunction is defined as $\eta_M : M \to M \otimes_A \mathcal{C}, m \mapsto \sum_i b_i m \otimes_A c_i$, and the counit is $\psi_N = N \otimes_A \varepsilon_{\mathcal{C}}$, for all $M \in {}_B\mathbf{M}$, $N \in \mathbf{M}_A$. □

If $\mathcal{C}$ is finitely generated and projective as a left A-module, then one can identify the category $\mathbf{M}^{\mathcal{C}}$ with the category ${}_B\mathbf{M}$. More precisely, the functor $\mathbf{M}_{\mathcal{C}} \to {}_B\mathbf{M}$ is an isomorphism of categories with the inverse given on objects as follows. Let $\{c^i \in \mathcal{C}, b_i \in B\}$ be a finite dual basis for $\mathcal{C}$. Any left B-module M is mapped to a right $\mathcal{C}$-comodule (M, ϱ^M), with coaction

$$\varrho^M : M \to M \otimes_A \mathcal{C}, \qquad m \mapsto \sum_i b_i m \otimes_A c^i,$$

where M is a right A-module by the left multiplication with ι_L, that is,

$$ma = \iota_L(a)m = \varepsilon_{\mathcal{C}}(-a)m.$$

Conversely, if $\mathbf{M}^{\mathcal{C}} \equiv {}_B\mathbf{M}$, then, since there is the defining adjoint pair of functors $((-)_A = (-)_{B/A}, - \otimes_A \mathcal{C})$, Lemma implies that $\mathcal{C}$ is a finitely generated and projective left A-module.

4.38. Module properties of a coring and the structure of comodules. Summary. The impact of module properties of a coring on the category of its comodules can be summarized as the following three statements about an A-coring $\mathcal{C}$ and its dual ring $B = {}^*\mathcal{C}$:

(1) If $\mathcal{C}$ is flat as a left A-module, then $\mathbf{M}^{\mathcal{C}}$ is a Grothendieck category.
(2) $\mathbf{M}^{\mathcal{C}}$ is the smallest full Grothendieck subcategory of ${}_B\mathbf{M}$ containing $\mathcal{C}$ if and only if $\mathcal{C}$ is a locally projective left A-module (hence a flat left A-module).
(3) $\mathbf{M}^{\mathcal{C}}$ coincides with the category ${}_B\mathbf{M}$ if and only if $\mathcal{C}$ is a finitely generated and projective left A-module (hence a locally projective left A-module).

For example, taking $\mathcal{C} = A \otimes_D A$ to be the Sweedler coring associated to an algebra map $D \to A$, we obtain the following information about the category of descent data $\mathbf{Desc}(A|D)$ (cf. 4.20):

(1) $\mathbf{Desc}(A|D)$ is a Grothendieck category provided A is flat as a left D-module.
(2) If A is locally projective as a left D-module, then $\mathbf{Desc}(A|D)$ is the smallest full Grothendieck subcategory of the category of left modules over the endomorphism ring $B = {}_D\mathrm{End}(A)$ that contains $A \otimes_D A$ [where $A \otimes_D A$ is a left B-module via $b(a \otimes_D a') = a \otimes_D b(a')$].
(3) If A is a finitely generated and projective as a left D-module, then $\mathbf{Desc}(A|D) \equiv {}_{D\mathrm{End}(A)}\mathbf{M}$.

5. Special types of corings and comodules

In this section, we describe several classes of corings, which satisfy additional properties. Such corings arise in a natural way in discussions of derived functors of the cotensor product or from the analysis of ring extensions (coseparable corings, cosplit corings, and Frobenius corings) or from descent theory (Galois corings). The richness of the structure of a coring has a bearing on its category of comodules. We discuss this matter here.

5.1. *Coseparable and cosplit corings*

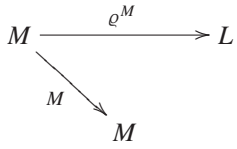
5.1. Injective comodules. A right $\mathcal{C}$-comodule (M, ϱ^M) is called a $(\mathcal{C}, A)$-*injective comodule* if, for every right $\mathcal{C}$-comodule map $i : N \to L$ that is a section in $\mathbf{M}_A$, every diagram

$$\begin{array}{ccc} N & \xrightarrow{\quad i \quad} & L \\ & \searrow^{f} & \\ & & M \end{array} \qquad (*)$$

of right $\mathcal{C}$-comodule maps can be completed commutatively by a right $\mathcal{C}$-comodule map $g : L \to M$.

Equivalently, (M, ϱ^M) is a $(\mathcal{C}, A)$-injective comodule if and only if there exists a right $\mathcal{C}$-colinear map $\sigma^M : M \otimes_A \mathcal{C} \to M$ such that $\sigma^M \circ \varrho^M = M$. Indeed, since

$\varrho^M : M \to M \otimes_A \mathcal{C}$ is a section in $\mathbf{M}_A$ with retraction $M \otimes_A \varepsilon_{\mathcal{C}} : M \otimes_A \mathcal{C} \to M$, the diagram



meets all the above requirements. Hence, if (M, ϱ^M) is $(\mathcal{C}, A)$-injective, then there is a map σ^M as required. Conversely, if there is such a map, then the diagram $(*)$ can be completed by $g = \sigma^M \circ (f \otimes_A \mathcal{C}) \circ (p \otimes_A \mathcal{C}) \circ \varrho^L$, where $p : L \to N$ is a right A-linear retraction of i.

Still equivalently (M, ϱ^M) is a $(\mathcal{C}, A)$-injective comodule if and only if every right $\mathcal{C}$-comodule map $M \to L$, which is a section in $\mathbf{M}_A$, is a section in $\mathbf{M}^{\mathcal{C}}$. To see this, simply set $N = M$ and $f = M$ in the diagram $(*)$.

Since $\Delta_{\mathcal{C}}$ has a right $\mathcal{C}$-colinear retraction $\varepsilon_{\mathcal{C}} \otimes_A \mathcal{C}$, the regular right $\mathcal{C}$-comodule $(\mathcal{C}, \Delta_{\mathcal{C}})$ is $(\mathcal{C}, A)$-injective. More generally and by the same token, for any right A-module X, $(X \otimes_A \mathcal{C}, X \otimes_A \Delta_{\mathcal{C}})$ is a $(\mathcal{C}, A)$-injective comodule.

5.2. Right $(\mathcal{C}, A)$-semisimple corings. An A-coring $\mathcal{C}$ whose all right comodules are $(\mathcal{C}, A)$-injective is called a *right $(\mathcal{C}, A)$-semisimple coring*.

5.3. Coseparable corings. Coseparable corings are an example of right (and left) $(\mathcal{C}, A)$-semisimple corings.

An A-coring $\mathcal{C}$ is called a *coseparable coring* [106] if the coaction $\Delta_{\mathcal{C}}$ has a $\mathcal{C}$-bicolinear retraction, that is, if there exists an A-bimodule map $\pi : \mathcal{C} \otimes_A \mathcal{C} \to \mathcal{C}$ such that $\pi \circ \Delta_{\mathcal{C}} = \mathcal{C}$, and the following diagrams are commutative

$$\begin{array}{ccc} \mathcal{C} \otimes_A \mathcal{C} & \xrightarrow{\mathcal{C} \otimes_A \Delta_{\mathcal{C}}} & \mathcal{C} \otimes_A \mathcal{C} \otimes_A \mathcal{C} \\ {\scriptstyle \pi}\downarrow & & \downarrow{\scriptstyle \pi \otimes_A \mathcal{C}} \\ \mathcal{C} & \xrightarrow{\Delta_{\mathcal{C}}} & \mathcal{C} \otimes_A \mathcal{C}\,, \end{array} \qquad \begin{array}{ccc} \mathcal{C} \otimes_A \mathcal{C} & \xrightarrow{\Delta_{\mathcal{C}} \otimes_A \mathcal{C}} & \mathcal{C} \otimes_A \mathcal{C} \otimes_A \mathcal{C} \\ {\scriptstyle \pi}\downarrow & & \downarrow{\scriptstyle \mathcal{C} \otimes_A \pi} \\ \mathcal{C} & \xrightarrow{\Delta_{\mathcal{C}}} & \mathcal{C} \otimes_A \mathcal{C}\,. \end{array}$$

The existence of such a map π is equivalent to the existence of an A-bimodule map $\delta : \mathcal{C} \otimes_A \mathcal{C} \to A$ such that for all $c, c' \in \mathcal{C}$,

$$\sum c_{(1)} \delta(c_{(2)} \otimes_A c') = \sum \delta(c \otimes_A c'_{(1)}) c'_{(2)}, \qquad \sum \delta(c_{(1)} \otimes_A c_{(2)}) = \varepsilon_{\mathcal{C}}(c).$$

The correspondence is given by

$$\pi \mapsto \varepsilon_{\mathcal{C}} \circ \pi, \qquad \delta \mapsto (\delta \otimes_A \mathcal{C}) \circ (\mathcal{C} \otimes_A \Delta_{\mathcal{C}}) = (\mathcal{C} \otimes_A \delta) \circ (\Delta_{\mathcal{C}} \otimes_A \mathcal{C}).$$

A map δ satisfying the above properties is called a *cointegral*.

Given a right comodule (M, ϱ^M) of a coseparable A-coring $\mathcal{C}$, the right $\mathcal{C}$-colinear map $\sigma^M = (M \otimes_A \delta) \circ (\varrho^M \otimes_A \mathcal{C})$ is a retraction of ϱ^M. This means that every right $\mathcal{C}$-comodule of a coseparable coring $\mathcal{C}$ is $(\mathcal{C}, A)$-injective, that is, a coseparable coring is $(\mathcal{C}, A)$-semisimple.

5.4. Coseparable corings as firm algebras [48]. Let $\mathcal{C}$ be a coseparable coring (with π and δ as in 5.3). Then $\mathcal{C}$ is a firm algebra with the product

$$\mu_{\mathcal{C}} : \mathcal{C} \otimes \mathcal{C} \to \mathcal{C}, \qquad c \otimes c' \mapsto \pi(c \otimes_A c').$$

The inverse of the product map $\mathcal{C} \otimes_{\mathcal{C}} \mathcal{C} \to \mathcal{C}$ is given by the coproduct $\Delta_{\mathcal{C}}$ (canonically projected to $\mathcal{C} \otimes_{\mathcal{C}} \mathcal{C}$). Furthermore, every right $\mathcal{C}$-comodule (M, ϱ^M) is a firm right $\mathcal{C}$-module with the multiplication, for all $m \in M$, $c \in \mathcal{C}$,

$$mc = \sum m_{(0)} \delta(m_{(1)} \otimes_A c).$$

The inverse $\nabla_M : M \to M \otimes_{\mathcal{C}} \mathcal{C}$ of the multiplication map is given by ϱ^M (projected canonically to $M \otimes_{\mathcal{C}} \mathcal{C}$).

5.5. Cosplit corings. An A-coring $\mathcal{C}$ is called a *cosplit coring* if there exists an element

$$e \in \mathcal{C}^A := \{c \in \mathcal{C} \mid \forall a \in A,\ ca = ac\}$$

such that $\varepsilon_{\mathcal{C}}(e) = 1_A$. Equivalently, one requires that the counit $\varepsilon_{\mathcal{C}}$ has an A-bimodule section. Such an element e is called a *normalized integral* in $\mathcal{C}$.

The name "cosplit" is motivated by the observation that a dual ring of a cosplit coring is a split extension. More precisely, an algebra map $i : A \to B$ is said to be a *split extension* [154] if it is an A-bimodule section. Let $B = ({}^*\mathcal{C})^{\mathrm{op}}$ be the opposite left dual ring of a cosplit A-coring $\mathcal{C}$. Then, the algebra map $\iota_L : A \to B, a \mapsto \varepsilon_{\mathcal{C}}(-a)$ is a split extension. Indeed, if $e \in \mathcal{C}^A$ is a normalized integral, then the map $E : B \to A$, $f \mapsto f(e)$, is an A-bilinear retraction of ι_L.

5.6. Separable bimodules [176] (cf. [65]). Let M be a (B, A)-bimodule. M is called a *separable bimodule* or B is said to be *M-separable over A*, provided the evaluation map

$$M \otimes_A {}_B\mathrm{Hom}(M, B) \to B, \qquad m \otimes_A \varphi \mapsto \varphi(m),$$

is a retraction (i.e. a split epimorphism) of B-bimodules.

The notion of a separable bimodule includes the notion of a separable extension and a split extension. Recall that an algebra map $A \to B$ is *separable* if there exists an invariant $e = \sum_i b_i \otimes_A b'_i \in (B \otimes_A B)^B$ such that $\sum_i b_i b'_i = 1_B$; see [109, Definition 2]. This is equivalent to the requirement that the product map $B \otimes_A B \to B$ is a retraction of B-bimodules. In case $M = B$, the dual module ${}_B\mathrm{Hom}(B, B)$ can be identified with B, and then the evaluation map coincides with the product map. Hence, in the case of an algebra map $A \to B$, B is a separable (B, A)-bimodule if and only if $A \to B$ is a separable extension.

Furthermore, as shown in [176] (cf. [117, Theorem 3.1]), if M is a separable bimodule, then $B \to S := \mathrm{End}_A(M)$ is a split extension. Conversely, if a (B, A)-bimodule M is such that M is finitely generated and projective as a right A-module, and $B \to S$ is a split extension, then M is a separable (B, A)-bimodule.

5.7. Separable bimodules and comatrix corings [45]. Let P be a (B, A)-bimodule that is finitely generated and projective as a right A-module, and let $\mathcal{C} = P^* \otimes_B P$ be the corresponding comatrix coring.

(1) The (A, B)-bimodule P^* is a separable bimodule if and only if $\mathcal{C}$ is a cosplit A-coring.
(2) If P is a separable bimodule, then $\mathcal{C}$ is a coseparable A-coring.
(3) Let $S = \mathrm{End}_A(P)$ and suppose that S is a faithfully flat left or right B-module. If $\mathcal{C}$ is a coseparable A-coring, then P is a separable bimodule.

5.8. Separable functors. The notions of coseparable and cosplit corings are complementary to each other in the sense that the coseparablity and the cosplit property correspond to the same property of functors in the defining adjunction. This common property is that of *separability*.

The notion of a separable functor was introduced in [144]. Following the formulation in [159], a functor $F : \mathbf{A} \to \mathbf{B}$ is said to be *separable* if the transformation $\mathbf{A}(-, -) \to \mathbf{B}(F(-), F(-)),\ f \mapsto F(f)$ is a split natural monomorphism.

In case the functor F has a right adjoint G, the Rafael theorem [159, Theorem 1.2] asserts that F is separable if and only if the unit of adjunction is a natural section, while G is separable if and only if the counit of adjunction is a natural retraction.

5.9. Coseparable functors and corings [37, Theorem 3.3, Corollary 3.6]. Let $\mathcal{C}$ be an A-coring.

(1) The forgetful functor $(-)_A : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}_A$ is a separable functor if and only if $\mathcal{C}$ is a coseparable coring.
(2) The functor $- \otimes_A \mathcal{C} : \mathbf{M}_A \to \mathbf{M}^{\mathcal{C}}$ is a separable functor if and only if $\mathcal{C}$ is a cosplit coring.

5.2. *Frobenius corings*

5.10. Frobenius corings. An A-coring $\mathcal{C}$ is called a *Frobenius coring* if there exist an A-bimodule map $\eta : A \to \mathcal{C}$ and a $\mathcal{C}$-bicomodule map $\pi : \mathcal{C} \otimes_A \mathcal{C} \to \mathcal{C}$ yielding a commutative diagram

$$\begin{array}{ccc} \mathcal{C} & \xrightarrow{\mathcal{C}\otimes_A\eta} & \mathcal{C}\otimes_A\mathcal{C} \\ {\scriptstyle \eta\otimes_A\mathcal{C}}\downarrow & \searrow^{\mathcal{C}} & \downarrow{\scriptstyle \pi} \\ \mathcal{C}\otimes_A\mathcal{C} & \xrightarrow{\pi} & \mathcal{C}\,. \end{array}$$

Equivalently, identifying π with the corresponding A-bimodule map $\delta = \varepsilon_{\mathcal{C}} \circ \pi$ (cf. 5.3) and η with an element $e \in \mathcal{C}^A$, $\mathcal{C}$ is a Frobenius coring if and only if there exist $e \in \mathcal{C}^A$ and an A-bimodule map $\delta : \mathcal{C} \otimes_A \mathcal{C} \to A$ such that for all $c, c' \in \mathcal{C}$,

$$\sum c_{(1)}\delta(c_{(2)}\otimes_A c') = \sum \delta(c\otimes_A c'_{(1)})c'_{(2)}, \quad \delta(c\otimes_A e) = \delta(e\otimes_A c) = \varepsilon_{\mathcal{C}}(c).$$

The pair (π, e) is called a *Frobenius system*, while the pair (δ, e) is called a *reduced Frobenius system*.

Note that any Frobenius A-coring $\mathcal{C}$ is a finitely generated projective left and right A-module. In terms of the reduced Frobenius system (δ, e), and writing $\Delta_{\mathcal{C}}(e) = \sum_i e_i \otimes_A \bar{e}_i$, a dual basis for the right A-module $\mathcal{C}$ is $\{e_i, \delta(\bar{e}_i \otimes_A -)\}$ and for the left A-module is $\{\bar{e}_i, \delta(- \otimes_A e_i)\}$.

5.11. Frobenius functors and Frobenius corings. In [141, Theorem 2.1], it has been observed that an algebra extension is a Frobenius extension if and only if the extension of scalars functor has the same left and right adjoint. Following [67], a functor F is said to be a *Frobenius functor* if and only if it has the same left and right adjoint (such functors were termed *strongly adjoint pairs* in [141]).

$\mathcal{C}$ is a Frobenius A-coring if and only if $- \otimes_A \mathcal{C} : \mathbf{M}_A \to \mathbf{M}^{\mathcal{C}}$ is a Frobenius functor and if and only if $(-)_A : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}_A$ is a Frobenius functor.

5.12. Frobenius corings and Frobenius extensions [39, Proposition 2.5, Theorem 2.7]. In 3.18, we have described a relationship between corings and Frobenius extensions. This description is completed by the following observations:

(1) Let $A \to B$ be a Frobenius extension with a Frobenius element β and a Frobenius homomorphism E. Then B is a Frobenius A-coring with coproduct β (viewed as a B-bimodule map $B \to B \otimes_A B$), counit E, and a Frobenius system $(\pi, 1_B)$, where $\pi : B \otimes_A B \to B$, $b \otimes_A b' \mapsto bb'$.
(2) Let $\mathcal{C}$ be a Frobenius A-coring with a Frobenius system (π, e). Then,
   (a) $\mathcal{C}$ is a k-algebra with product $cc' = \pi(c \otimes_A c')$ and unit $1_{\mathcal{C}} = e$.
   (b) $A \to \mathcal{C}$, $a \mapsto ae = ea$ is a Frobenius extension with a Frobenius element $\beta = \Delta_{\mathcal{C}}(e)$ and Frobenius homomorphism $E = \varepsilon_{\mathcal{C}}$.

Thus, the notions of a Frobenius coring and a Frobenius extension are dual to each other descriptions of the same situation.

5.13. Frobenius bimodules. A (B, A)-bimodule M is said to be *Frobenius* if M is finitely generated and projective as a left B- and right A-module and ${}_B\mathrm{Hom}(M, B) \simeq \mathrm{Hom}_A(M, A)$ as (A, B)-bimodules; see [9, p. 261] or [116, Definition 2.1]. The endomorphism ring theorem (cf. [116, Theorem 2.5]) asserts that if M is a Frobenius (B, A)-bimodule, then $B \to \mathrm{End}_A(M)$ is a Frobenius extension.

5.14. Frobenius bimodules and comatrix corings [45, Theorem 3.7, Corollary 3.8]. Let P be a (B, A)-bimodule that is finitely generated and projective as a right A-module, and let $\mathcal{C} = P^* \otimes_B P$ be the corresponding comatrix coring.

(1) If P is a Frobenius bimodule, then $\mathcal{C}$ is a Frobenius A-coring.
(2) Assume that P is finitely generated and projective as a left B-module. Let $S = \mathrm{End}_A(P)$ and suppose that S is a faithfully flat left or right B-module and that ${}_S\mathrm{Hom}(P, S) \cong \mathrm{Hom}_A(P, A)$ as (A, B)-bimodules. If $\mathcal{C}$ is a Frobenius A-coring, then P is a Frobenius bimodule.

In particular, if there is a Frobenius algebra map $B \to A$, then $P = A$ is a Frobenius (B, A)-bimodule, and hence, the corresponding Sweedler coring $A \otimes_B A$ is a Frobenius coring.

5.15. Towers of Frobenius extensions [39, Theorem 3.8]. Suppose that $\mathcal{C}$ is a Frobenius A-coring with a Frobenius system (π, e). Then by 5.12, e viewed as a map $A \to \mathcal{C}$ is a Frobenius extension with the Frobenius element $\Delta_{\mathcal{C}}(e)$ and the Frobenius homomorphism $\varepsilon_{\mathcal{C}}$. Now 5.14 implies that the Sweedler $\mathcal{C}$-coring $\mathcal{C} \otimes_A \mathcal{C}$ is Frobenius with the Frobenius system $(\mathcal{C} \otimes_A \varepsilon_{\mathcal{C}} \otimes_A \mathcal{C}, \Delta_{\mathcal{C}}(e))$. Then $\mathcal{C} \otimes_A \mathcal{C}$ is a ring with unit $\Delta_{\mathcal{C}}(e)$ and the product

$$(c \otimes_A c')(c'' \otimes_A c''') = c \otimes_A \varepsilon_{\mathcal{C}}(\pi(c' \otimes_A c''))c''',$$

and the extension $\Delta_{\mathcal{C}} : \mathcal{C} \to \mathcal{C} \otimes_A \mathcal{C}$ is Frobenius by 5.12. The Frobenius element explicitly reads $\sum e_{(1)} \otimes_A e \otimes_A e_{(2)}$, and the Frobenius homomorphism is π. Apply 5.12 to deduce that Sweedler's $\mathcal{C} \otimes_A \mathcal{C}$-coring $(\mathcal{C} \otimes_A \mathcal{C}) \otimes_{\mathcal{C}} (\mathcal{C} \otimes_A \mathcal{C}) \cong \mathcal{C} \otimes_A \mathcal{C} \otimes_A \mathcal{C}$ is Frobenius. Iterating this procedure we obtain the following Jones-like tower of corings and algebra maps

$$A \xrightarrow{e} \mathcal{C} \xrightarrow{\Delta_{\mathcal{C}}} \mathcal{C} \otimes_A \mathcal{C} \xrightarrow{\mathcal{C} \otimes_A e \otimes_A \mathcal{C}} \mathcal{C} \otimes_A \mathcal{C} \otimes_A \mathcal{C} \longrightarrow \ldots,$$

in which for all $k = 1, 2, \ldots$, $\mathcal{C}^{\otimes_A k-1} \to \mathcal{C}^{\otimes_A k}$ is a Frobenius extension and $\mathcal{C}^{\otimes_A k}$ is a Frobenius $\mathcal{C}^{\otimes_A k-1}$-coring.

5.16. Categorical interpretation. Many of the properties of Frobenius corings can be derived by purely categorical methods. Given any monoidal category $(\mathbf{M}, \odot, E)$, a *Frobenius algebra in* $\mathbf{M}$ is an object F such that $(F, \mu_F, 1_F)$ is an algebra in $(\mathbf{M}, \odot, E)$ and $(F, \Delta_F, \varepsilon_F)$ is a coalgebra in $\mathbf{M}$ such that

$$(\mu_F \odot F) \circ (F \odot \Delta_F) = \Delta_F \circ \mu_F = (F \odot \mu_F) \circ (\Delta_F \odot F);$$

see [142]. In particular, for any category $\mathbf{A}$, the category $\mathbf{End}\ \mathbf{A} := \mathbf{Fun}(\mathbf{A}, \mathbf{A})$ of functors $\mathbf{A} \to \mathbf{A}$ is a monoidal category (with tensor product given by the composition of functors). A Frobenius algebra in $\mathbf{End}\ \mathbf{A}$ is simply a pair consisting of a monad (F, ν, η) and a comonad (F, δ, σ) such that

$$\nu_F \circ F(\delta) = \delta \circ \nu = F(\nu) \circ \delta_F.$$

If $R : \mathbf{A} \to \mathbf{B}$ is a Frobenius functor with the same left and right adjoint L, $F = LR$ is both a monad and comonad. Both structures are compatible making F a Frobenius algebra in $\mathbf{End}\ \mathbf{A}$. One easily finds that, for a Frobenius algebra F in $\mathbf{End}\ \mathbf{A}$, the categories of algebras of the monad (F, ν, η) and coalgebras of the comonad (F, δ, σ) are mutually isomorphic.

On the other hand, a Frobenius algebra F in the monoidal category of A-bimodules (with tensor product $\otimes_A$) is the same as a Frobenius A-coring or a Frobenius extension $A \to F$. By arguments similar to those in 3.5, this is equivalent to the statement that $F \otimes_A -$ is a Frobenius algebra in $\mathbf{End}\ {}_A\mathbf{M}$. The results presented in this section

are just an interplay between these different descriptions of a Frobenius algebra (in monoidal categories).

For example, a (B, A)-bimodule P is Frobenius if and only if $P\otimes_A - : {}_A\mathbf{M} \to {}_B\mathbf{M}$ is a Frobenius functor with (left = right) adjoint $P^* \otimes_B -$; see [9, Exercise 21.8]. Thus, $P^*\otimes_B P\otimes_A -$ is a Frobenius algebra in $\mathbf{End}\ {}_A\mathbf{M}$; hence $P^*\otimes_B P$ is a Frobenius algebra in the category of A-bimodules, that is, it is a Frobenius A-coring.

We refer the reader to [142] for more details. A very accessible expanded review of [142] can also be found in [12]. Finally, we would like to mention that Frobenius algebras (hence also Frobenius corings) play an important role in description of certain quantum field theories. A review of recent developments in this area can be found in [168].

5.17. Co-Frobenius corings [112]. An A-coring $\mathcal{C}$ is said to be a *left co-Frobenius coring* in case there is an injective morphism $\mathcal{C} \to {}^*\mathcal{C}$ of left $A \otimes {}^*\mathcal{C}$-modules [i.e. $({}^*\mathcal{C}, A^{\mathrm{op}})$-bimodules, where ${}^*\mathcal{C}$ is a right A^{op}-module via ι_L; see 3.25]. $\mathcal{C}$ is a *right co-Frobenius coring* in case there is an injective morphism $\mathcal{C} \to \mathcal{C}^*$ of right $A \otimes \mathcal{C}^*$-modules; see [130] for the case of coalgebras. $\mathcal{C}$ is called a *left quasi-co-Frobenius coring* provided that there are a set Λ and an injective morphism $\mathcal{C} \to ({}^*\mathcal{C})^\Lambda$ of left $A \otimes {}^*\mathcal{C}$-modules.

A Frobenius A-coring $\mathcal{C}$ is isomorphic to ${}^*\mathcal{C}$ as a left $A \otimes {}^*\mathcal{C}$-module and is isomorphic to $\mathcal{C}^*$ as a right $A \otimes \mathcal{C}^*$-module. Explicitly, if (δ, e) is a reduced Frobenius system for $\mathcal{C}$, then the left $A \otimes {}^*\mathcal{C}$-module isomorphism is $\theta_L : \mathcal{C} \to {}^*\mathcal{C}$, $c \mapsto [c' \mapsto \delta(c' \otimes_A c)]$, with inverse $\xi \mapsto \sum e_{(1)}\xi(e_{(2)})$. The right $A \otimes \mathcal{C}^*$-module isomorphism is $\theta_R : \mathcal{C} \to \mathcal{C}^*$, $c \mapsto [c' \mapsto \delta(c \otimes_A c')]$, with inverse $\xi \mapsto \sum \xi(e_{(1)})e_{(2)}$. Thus, in particular, a Frobenius coring is left and right co-Frobenius.

Recall that $\mathcal{C}$ is a Frobenius coring if and only if the functor $- \otimes_A \mathcal{C}$ is Frobenius. Similar characterization exists also for (quasi)-co-Frobenius corings. A detailed discussion of (quasi)-co-Frobenius corings and related functors can be found in [112].

5.18. Generalizations of Frobenius corings. *Quasi-Frobenius corings* are defined in [72] in terms of *quasi-Frobenius functors* [141]. Let $\mathbf{A}$, $\mathbf{B}$ be categories with finite direct sums (finite coproducts, e.g. Grothendieck categories). A functor $L : \mathbf{B} \to \mathbf{A}$ is said to be *similar* to $R : \mathbf{B} \to \mathbf{A}$ if there exist positive integers m, n and natural transformations $\alpha : L \to R^n$, $\widetilde{\alpha} : R^n \to L$, $\beta : R \to L^m$, $\widetilde{\beta} : L^m \to R$ such that

$$\widetilde{\alpha} \circ \alpha = L, \qquad \widetilde{\beta} \circ \beta = R.$$

Here, R^n (resp., L^n) is a functor defined on objects B as the direct sum of n-copies of $R(B)$ [resp., $L(B)$]. A functor $F : \mathbf{A} \to \mathbf{B}$ is said to be *quasi-Frobenius* if it has a left adjoint L and right adjoint R such that L is similar to R. The notion of quasi-Frobenius functors is a symmetrized version of a *left quasi-Frobenius pair of functors* studied in [105]. $\mathcal{C}$ is called a *quasi-Frobenius A-coring* if the forgetful functor $(-)_A : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}_A$ is quasi-Frobenius. Properties of such corings as well as quasi-Frobenius functors between comodule categories are studied in [71, 72].

Frobenius properties of functors associated to morphisms of corings (or *Frobenius coring extensions*) are studied in [111, 203].

5.3. *Galois corings (comodules) and generalized descent*

Galois corings or *Galois comodules* are possibly the most studied class of corings in recent years. Galois corings can be understood as an approach to generalized Galois theory with its classical origins in field extensions (see [198] for a review of the classical Galois theory in the context of comodules), which is still not quite well-understood from the categorical point of view developed in [35]. Furthermore, Galois corings lead to a reformulation of noncommutative descent theory [55]. Probably the strongest motivation of the recent interest in Galois comodules comes from noncommutative geometry; see Sections 6.2 and 6.3.

5.19. Galois comodules. Let $\mathcal{C}$ be an A-coring and (M, ϱ^M) a right $\mathcal{C}$-comodule. Set $S = \mathrm{End}^{\mathcal{C}}(M)$. For any right A-module N, define a right $\mathcal{C}$-comodule map $\widetilde{\psi}_N$ as a composition

$$\widetilde{\psi}_N : \mathrm{Hom}_A(M, N) \otimes_S M \to N \otimes_A \mathcal{C}, \qquad f \otimes_S m \mapsto \sum f(m_{(0)}) \otimes_A m_{(1)}.$$

(M, ϱ^M) is called a *Galois comodule* if and only if, for any right A-module N, the map $\widetilde{\psi}_N$ is an isomorphism of right $\mathcal{C}$-comodules [199].

If M is finitely generated and projective as a right A-module, then $\mathrm{Hom}_A(M, N) \simeq N \otimes_A M^*$, where $M^* = \mathrm{Hom}_A(M, A)$. Thus, in this case, (M, ϱ^M) is a Galois comodule if and only if the *canonical* A-coring morphism

$$\mathsf{can}_M : M^* \otimes_S M \to \mathcal{C}, \qquad \xi \otimes m \mapsto \sum \xi(m_{(0)}) m_{(1)},$$

where $M^* \otimes_S M$ is the comatrix coring (cf. 3.11), is an isomorphism. A Galois comodule (M, ϱ^M) that is finitely generated and projective as an A-module is called a *finite Galois comodule* or the pair $(\mathcal{C}, M)$ is called a *Galois coring* [92]. The bijectivity of can_M is referred to as a *finite Galois condition*. Furthermore, identifying M^* with $\mathrm{Hom}^{\mathcal{C}}(M, \mathcal{C})$ by 4.10, one finds that the finite Galois condition is satisfied if and only if the evaluation map $\psi_{\mathcal{C}} : \mathrm{Hom}^{\mathcal{C}}(M, \mathcal{C}) \otimes_S M \to \mathcal{C}$, $f \otimes m \mapsto f(m)$, is an isomorphism of right $\mathcal{C}$-comodules. Note that the evaluation map is a counit of the coinvariants adjunction 4.11.

In this chapter, we focus on finite Galois comodules, as these have been more thoroughly studied so far. For more information on general Galois comodules, the reader is referred to [199].

5.20. A as a Galois comodule. The case in which A itself is a Galois $\mathcal{C}$-comodule is of prime interest. As explained in 4.12, A is a right $\mathcal{C}$-comodule if and only if there exists a group-like element g. In this case, the coinvariants $S = \mathrm{End}^{\mathcal{C}}(A)$ coincide with the subalgebra $A_g = \{s \in A \mid sg = gs\} \subseteq A$; see 4.14. Obviously, A is a finitely generated projective right A-module, $A^* \simeq A$, and $A \otimes_S A$ is the Sweedler A-coring; see 3.10 and 3.11. The canonical map comes out as

$$\mathsf{can}_A : A \otimes_S A \to \mathcal{C}, \qquad a \otimes a' \mapsto aga'.$$

Thus, A is a finite Galois comodule if and only if $\mathcal{C}$ is a *Galois coring* with respect to g, a notion introduced in [37].

5.21. Galois connections associated to a finite Galois comodule. [79] Let M be a (B, A)-bimodule that is finitely generated and projective as a right A-module, and let $\mathcal{C} = M^* \otimes_B M$ be the associated comatrix coring. Set $T = \mathrm{End}_A(M)$ and consider the set $\mathrm{Subext}(T \mid B)$ of all algebras S such that $B \subseteq S \subseteq T$. $\mathrm{Subext}(T \mid B)$ is a partially ordered set with the order given by inclusion. Next, consider the set $\mathrm{Coideal}(\mathcal{C})$ of all coideals of $\mathcal{C}$, cf. 3.4, and order it by $K \leq K' \Leftrightarrow K' \subseteq K$. For any coideal $K \in \mathrm{Coideal}(\mathcal{C})$, view M as a right $\mathcal{C}/K$-comodule with the coaction $\varrho^M : M \to M \otimes_A \mathcal{C}/K$, $m \mapsto (M \otimes_A p)(e \otimes_A m)$, where $e \in M \otimes_A M^*$ is a dual basis element and $p : \mathcal{C} \to \mathcal{C}/K$ is the defining epimorphism of K.

Given any $S \in \mathrm{Subext}(T \mid B)$, define a coideal $\mathfrak{K}(S) \in \mathrm{Coideal}(\mathcal{C})$ as the kernel of the canonical map $M^* \otimes_B M \to M^* \otimes_S M$. This defines an order-reversing operation

$$\mathfrak{K} : \mathrm{Subext}(T \mid B) \to \mathrm{Coideal}(\mathcal{C}).$$

Starting with a coideal $K \in \mathrm{Coideal}(\mathcal{C})$, define $\mathfrak{S}(K) = \mathrm{End}^{\mathcal{C}/K}(P) \subseteq T$. In this way, one obtains an order-reversing operation

$$\mathfrak{S} : \mathrm{Coideal}(\mathcal{C}) \to \mathrm{Subext}(T \mid B).$$

Then, for all $K \in \mathrm{Coideal}(\mathcal{C})$ and $S \in \mathrm{Subext}(T \mid B)$,

$$\mathfrak{S}(\mathfrak{K}(S)) \supseteq S, \qquad \mathfrak{K}(\mathfrak{S}(K)) \subseteq K,$$

that is, the pair $(\mathfrak{K}, \mathfrak{S})$ is a Galois connection between the partially ordered sets $\mathrm{Subext}(T \mid B)$ and $\mathrm{Coideal}(\mathcal{C})$. Furthermore, $\mathfrak{S}(\mathfrak{K}(S)) = S$, if and only if $\mathrm{End}^{M^* \otimes_S M}(M) = S$, and $\mathfrak{K}(\mathfrak{S}(K)) = K$, if and only if (M, ϱ^M) is a finite Galois $\mathcal{C}/K$-comodule.

5.22. The finite Galois comodule structure theorem. This theorem describes properties of the M-coinvariants functor associated to a finite Galois comodule. It arose as a generalization of the structure theorem for Hopf–Galois extensions [167, Theorem 3.7] and was first stated in [92, Theorem 3.2]. The present formulation includes further additions from [40, Theorem 2.1], [53, 18.27], [59].

THEOREM. *Let $\mathcal{C}$ be an A-coring and M be a right $\mathcal{C}$-comodule that is finitely generated and projective as a right A-module. Set $S = \mathrm{End}^{\mathcal{C}}(M)$, and denote the M-coinvariants functor $\mathrm{Hom}^{\mathcal{C}}(M, -) : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}_S$ by G.*

(1) *The following statements are equivalent:*
   (a) *M is a finite Galois comodule that is flat as a left S-module.*
   (b) *$\mathcal{C}$ is a flat left A-module and M is a generator in $\mathbf{M}^{\mathcal{C}}$.*
   (c) *$\mathcal{C}$ is a flat left A-module and G is fully faithful, that is, for any $N \in \mathbf{M}^{\mathcal{C}}$, the counit of the coinvariants adjunction $\psi_N : \mathrm{Hom}^{\mathcal{C}}(M, N) \otimes_S M \to N$ is an isomorphism of right $\mathcal{C}$-comodules.*
   (d) *M is a flat left S-module and G is fully faithful.*

(2) *The following statements are equivalent:*
   (a) *M is a finite Galois comodule that is faithfully flat as a left S-module.*
   (b) *$\mathcal{C}$ is a flat left A-module and M is a projective generator in $\mathbf{M}^{\mathcal{C}}$.*
   (c) *$\mathcal{C}$ is a flat left A-module and G is an equivalence with the inverse $-\otimes_S M : \mathbf{M}_S \to \mathbf{M}^{\mathcal{C}}$.*
   (d) *M is a flat left S-module and G is an equivalence.*

Part (1) is often referred to as the *weak structure theorem*, while part (2) is known as the *strong structure theorem*. It is worth pointing out that equivalences (b) $\Leftrightarrow$ (c) $\Leftrightarrow$ (d) are standard properties of Grothendieck categories and only (a) is special for comodules.

5.23. Semisimple and simple finite Galois comodules. A right $\mathcal{C}$-comodule (M, ϱ^M) is called a *semisimple comodule* if, for any right $\mathcal{C}$-comodule (N, ϱ^N), every right $\mathcal{C}$-colinear monomorphism $N \to M$ is a section (of $\mathcal{C}$-comodules). If any such monomorphism is an isomorphism, then (M, ϱ^M) is called a *simple comodule*. In case $\mathcal{C}$ is flat as a left A-module, and hence $\mathbf{M}^{\mathcal{C}}$ is an Abelian category, (M, ϱ^M) is simple if and only if every monomorphism $N \to M$ is either 0 or an isomorphism, and it is semisimple if it is a direct sum of simple comodules.

Following [199, 3.5], if $\mathcal{C}$ is a flat left A-module, then (M, ϱ^M) is a semisimple comodule if and only if, for any set Λ, the endomorphism ring of the direct sum comodule, $\mathrm{End}^{\mathcal{C}}(M^{(\Lambda)})$, is a von Neumann regular ring and for any right $\mathcal{C}$-comodule (N, ϱ^N), the counit of the M-coinvariants adjunction

$$\psi_N : \mathrm{Hom}^{\mathcal{C}}(M, N) \otimes_S M \to N, \qquad f \otimes m \mapsto f(m),$$

is injective. Here, as before, $S = \mathrm{End}^{\mathcal{C}}(M)$. If, in addition, M is a finitely generated and projective right A-module, then (M, ϱ^M) is semisimple if and only if S is a (right) semisimple ring and, for all (N, ϱ^N), ψ_N is injective.

Finally, by [40, Theorem 3.1], if $\mathcal{C}$ is flat as a left A-module, and (M, ϱ^M) is finitely generated and projective as a right A-module, then (M, ϱ^M) is a simple comodule if and only if S is a division ring, and the counit of the M-coinvariants adjunction ψ_N is injective for any right $\mathcal{C}$-comodule (N, ϱ^N). Therefore, every finite Galois comodule, whose endomorphism ring, is a division ring is a simple comodule. A (semi)simple comodule (M, ϱ^M) that is finitely generated and projective as a right A-module is a finite Galois comodule if and only if the evaluation map $\psi_{\mathcal{C}}$ is surjective and if and only if the canonical map can_M is surjective.

The definitions of a simple and semisimple comodules we use here in general differ from the definitions in [96] but coincide provided $\mathcal{C}$ is a flat left A-module.

5.24. Simple cosemisimple corings and the Jacobson–Bourbaki correspondence. An A-coring is said to be *cosemisimple* if $(\mathcal{C}, \Delta_{\mathcal{C}})$ is a semisimple right $\mathcal{C}$-comodule. This is a left–right symmetric notion [i.e. $\mathcal{C}$ is cosemisimple if and only if $(\mathcal{C}, \Delta_{\mathcal{C}})$ is a semisimple left $\mathcal{C}$-comodule]. If $\mathcal{C}$ is cosemisimple, then $\mathcal{C}$ is projective (hence also flat) as a left and right A-module; see [53, 19.14]. $\mathcal{C}$ is said to be *simple* if $(\mathcal{C}, \Delta_{\mathcal{C}}, \Delta_{\mathcal{C}})$ is a simple $\mathcal{C}$-bicomodule.

Cosemisimple corings were first studied in [96]. In both [53, 96], they are termed *semisimple corings*. The term *cosemisimple* was introduced in [92] to be in-line with the accepted coalgebra terminology.

A coring $\mathcal{C}$ is simple and cosemisimple if and only if there is a finite Galois comodule (M, ϱ^M) such that the coinvariants algebra $\mathrm{End}^{\mathcal{C}}(M)$ is a division algebra. Furthermore, $\mathcal{C}$ is a simple and cosemisimple A-coring if and only if $\mathcal{C}$ is (isomorphic to) a comatrix coring associated to a (B, A)-bimodule P, where B is a division algebra. If B' is another division algebra and P' is a (B', A)-bimodule such that $\mathcal{C}$ is isomorphic to the comatrix coring associated to P', then P and P' are isomorphic A-modules and B' is conjugate to B by this isomorphism; see [92, Theorem 4.3]. Finally, any (nonzero) finitely generated right comodule over a simple cosemisimple coring is a finite Galois comodule, and its coinvariants algebra is simple Artinian [92, Proposition 4.2].

If $\mathcal{C} = M^* \otimes_B M$ is a comatrix coring, then the Galois connection described in 5.21 restricts to mutually inverse maps on sets of simple Artinian subrings S of $\mathrm{End}_A(M)$ [i.e. subextensions $B \subseteq S \subseteq \mathrm{End}_A(M)$, where S is a simple Artinian algebra] and coideals K in $\mathcal{C}$ such that $\mathcal{C}/K$ is a simple cosemisimple A-coring [79, Theorem 2.4]. In case of an algebra map $B \to A$, where A is a division algebra, setting $M = A$ one obtains Sweedler's predual version of the Jacobson–Bourbaki correspondence [178, Theorem 2.1]. It is worth noting here that some coring-type ideas appear already in Hochschild's formulation of the Jacobson–Bourbaki theorem [110].

5.25. Principal Galois comodules. Let (M, ϱ^M) be a right $\mathcal{C}$-comodule; set $S = \mathrm{End}^{\mathcal{C}}(M)$ and consider any algebra map $B \to S$. (M, ϱ^M) is called a *B-relative principal Galois comodule* if it is a finite Galois comodule that is B-relatively projective as a left S-module. Explicitly, for (M, ϱ^M) to be a B-relative principal Galois comodule, it is required that the canonical map

$$\mathsf{can}_M : M^* \otimes_S M \to \mathcal{C}, \qquad \xi \otimes_S m \mapsto \sum \xi(m_{(0)}) m_{(1)},$$

be bijective (the Galois condition) and that the S-multiplication map

$$S \otimes_B M \to M, \qquad s \otimes_B m \mapsto s(m),$$

have a left S-linear section (the relative projectivity).

(M, ϱ^M) is said to be a *principal Galois comodule* if it is Galois and projective as a left S-module.

5.26. Structure theorem for principal Galois comodules. *Let $\mathcal{C}$ be an A-coring and (M, ϱ^M) a right $\mathcal{C}$-comodule that is finitely generated and projective as a right A-module. View $M^* = \mathrm{Hom}_A(M, A)$ as a left $\mathcal{C}$-comodule as in 4.6. Let $S = \mathrm{End}^{\mathcal{C}}(M)$, and let $B \to S$ be an algebra map such that*

$$S \otimes_B M \to {}^{\mathcal{C}}\mathrm{Hom}(M^*, M^* \otimes_B M), \qquad s \otimes_B m \mapsto [\xi \mapsto \xi \circ s \otimes_B m],$$

is an isomorphism of left S-modules. Then, the following statements are equivalent:

(*a*) *The map*

$$\widetilde{\mathsf{can}}_M : M^* \otimes_B M \to \mathcal{C}, \qquad \xi \otimes_B m \mapsto \sum \xi(m_{(0)}) m_{(1)},$$

is a split epimorphism of left $\mathcal{C}$-comodules.

(*b*) (M, ϱ^M) *is a B-relative principal Galois comodule.*

PROOF. The proof of implication (b) $\Rightarrow$ (a) is straightforward; thus, we only comment on the proof (a) $\Rightarrow$ (b). By assumption (a), the coring $\mathcal{C}$ is a direct summand of $M^* \otimes_B M$ as a left $\mathcal{C}$-comodule; hence, in view of the hypothesis; $M \simeq {}_A\mathrm{Hom}(M^*, A) \simeq {}^{\mathcal{C}}\mathrm{Hom}(M^*, \mathcal{C})$ is a direct summand of a left S-module $S \otimes_B M$, that is, M is a B-relatively projective left S-module.

The evaluation map $\widehat{\varphi}_M : M^* \otimes_S {}^{\mathcal{C}}\mathrm{Hom}(M^*, M^* \otimes_B M) \to M^* \otimes_B M$ factorizes through the isomorphism $S \otimes_B M \to {}^{\mathcal{C}}\mathrm{Hom}(M^*, M^* \otimes_B M)$ tensored with M^* and through the obvious isomorphism $M^* \otimes_S S \otimes_B M \to M^* \otimes_B M$. Hence $\widehat{\varphi}_M$ is an isomorphism. Consider the following diagram, which is commutative in all possible directions,

$$\begin{array}{ccc}
M^* \otimes_S {}^{\mathcal{C}}\mathrm{Hom}(M^*, M^* \otimes_B M) & \xrightarrow{\widehat{\varphi}_M} & M^* \otimes_B M \\
\uparrow \downarrow {\scriptstyle M^* \otimes_S {}^{\mathcal{C}}\mathrm{Hom}(M^*, \widetilde{\mathsf{can}}_M)} & & \uparrow \downarrow {\scriptstyle \widetilde{\mathsf{can}}_M} \\
M^* \otimes_S {}^{\mathcal{C}}\mathrm{Hom}(M^*, \mathcal{C}) & \xrightarrow{\widehat{\varphi}_{\mathcal{C}}} & \mathcal{C}
\end{array}$$

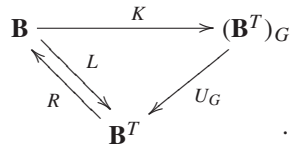
where $\widehat{\varphi}_{\mathcal{C}}$ is the evaluation map. The upward pointing arrows are sections of $M^* \otimes_S {}^{\mathcal{C}}\mathrm{Hom}(M^*, \widetilde{\mathsf{can}}_M)$ and $\widetilde{\mathsf{can}}_M$, respectively. Since $\widehat{\varphi}_M$ is an isomorphism, the map $\widehat{\varphi}_{\mathcal{C}}$ is bijective. The identifications ${}^{\mathcal{C}}\mathrm{Hom}(M^*, \mathcal{C}) \simeq M$ and $M^* \simeq \mathrm{Hom}^{\mathcal{C}}(M, \mathcal{C})$ lead to a k-linear isomorphism $M^* \otimes_S {}^{\mathcal{C}}\mathrm{Hom}(M^*, \mathcal{C}) \simeq \mathrm{Hom}^{\mathcal{C}}(M, \mathcal{C}) \otimes_S M$. In view of this isomorphism, the fact that $\widehat{\varphi}_{\mathcal{C}}$ is bijective implies that the counit of adjunction $\psi_{\mathcal{C}}$ is bijective. Thus, M is a finite Galois right $\mathcal{C}$-comodule. □

The above theorem is essentially the same as [40, Theorem 4.4]. The latter is formulated for algebras over a field, that is, for the case when $B = k$ is a field, and the isomorphism $S \otimes_B M \simeq {}^{\mathcal{C}}\mathrm{Hom}(M^*, M^* \otimes_B M)$ is proved rather than assumed. The case when $B = k$ is a commutative ring is described in [52, Theorem 2.1]. The case of a general B with a list of sufficient conditions for the isomorphism $S \otimes_B M \simeq {}^{\mathcal{C}}\mathrm{Hom}(M^*, M^* \otimes_B M)$ is studied in [199, Theorem 5.9] using the correspondence between Galois and ad-static comodules. The formulation given above follows that of [10, Theorem 5.1], with left–right conventions interchanged.

Note that if M is a projective left B-module, then there is the required isomorphism $S \otimes_B M \simeq {}^{\mathcal{C}}\mathrm{Hom}(M^*, M^* \otimes_B M)$, and the resulting finite Galois module in part (b) is principal (not only B-relative principal).

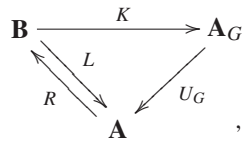
5.27. Comonadicity and generalized descent. Categorical descent theory (cf. [138]) considers the following situation. With a monad T on a category **B**, one

can associate the following (commutative, up to a natural isomorphism) triangle of categories and functors



Here $\mathbf{B}^T$ is the category of T-algebras (or T-modules), R and U_G are forgetful functors, L is a left adjoint of R (the induction functor), and G is the comonad LR. The functor K is the comparison functor defined by $K(B) = (L(B), L(\eta_B))$, where η is the unit of the adjunction (L, R). The category $(\mathbf{B}^T)_G$ is called the *category of descent data*. In descent theory, one is interested in studying when K is a full and faithful functor (one then says that T is of *descent type*) and when it is an equivalence (one then says that T is of *effective descent type*). This is a categorical formulation of the descent for an algebra map, discussed in 4.20. To see this, take an algebra map $B \to A$, take $\mathbf{B}$ to be the category of right B-modules and T to be the monad $(-\otimes_B A, -\otimes_B \mu_A, -\otimes_B 1_A)$. Then, $\mathbf{B}^T$ is the category of right A-modules, G is the comonad on $\mathbf{M}_A$ associated to the Sweedler coring $A \otimes_B A$, and $(\mathbf{B}^T)_G$ is indeed the category of descent data $\mathbf{Desc}(A \mid B)$.

Categorical descent theory fits in a (yet) more general framework. Given categories $\mathbf{A}$ and $\mathbf{B}$, a comonad G on $\mathbf{A}$, a diagram of categories and functors



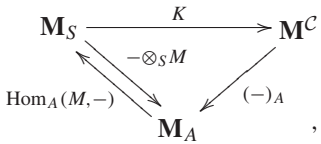
in which L is the left adjoint of R, U_G is the forgetful functor, and $L = U_G K$, is called a *G-comonadic triangle* on $\mathbf{A}$ and $\mathbf{B}$. We denote it by $(L \dashv R, K)_G$.

Start with an adjoint pair of functors (L, R), $L : \mathbf{B} \to \mathbf{A}$, and write ψ and η for the counit and the unit, respectively. Then $(LR, L(\eta_R), \psi)$ is a comonad on $\mathbf{A}$. By an argument dual to [88, Theorem II.1.1] (cf. [102, Theorem 2.2]), one can show that G-comonadic triangles $(L \dashv R, K)_G$ are in bijective correspondence with comonad morphisms $\varphi : LR \to G$. Explicitly, given such φ, define $\beta = \varphi_L \circ L(\eta)$, and then, the functor K is constructed by setting $K(B) = (L(B), \beta_B)$, for all objects B of $\mathbf{B}$. Conversely, given a G-comonadic triangle $(L \dashv R, K)_G$, the natural transformation φ is defined as the composition $\varphi = G(\psi) \circ \beta_R$, where $\beta : L \to GL$ is the natural transformation defined on objects B of $\mathbf{B}$ by the structure maps of $K(B)$. In this general situation, one can study when the functor K is fully faithful or an equivalence. A complete solution to this problem, which is partly based on Beck's theorem [17], can be found in [102, Theorems 2.6 and 2.7].

As explained in [102], the fact that the functor K in a comonadic triangle $(L \dashv R, K)_G$ is an equivalence is closely related to the fact that φ is an isomorphism. More precisely, given $(L \dashv R, K)_G$ and the corresponding φ, assume that, for all

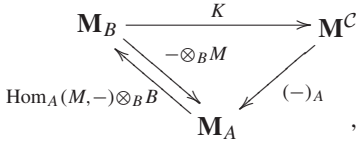
G-coalgebras, (M, ϱ^M), there exists in $\mathbf{B}$ an equalizer of $R(\varrho^M)$ and $R(\varphi_M) \circ \eta_{R(M)}$. Then K is an equivalence if and only if the functor L preserves all such equalizers and the map φ is an isomorphism of comonads.

The relevance of comonadic triangles to Galois properties of comodules of an A-coring $\mathcal{C}$ can be explained as follows. Take $\mathbf{A} = \mathbf{M}_A$ and $G = - \otimes_A \mathcal{C}$, the comonad associated to $\mathcal{C}$. Thus $\mathbf{A}_G = \mathbf{M}^{\mathcal{C}}$. Fix a right $\mathcal{C}$-comodule (M, ϱ^M) and let $S = \mathrm{End}^{\mathcal{C}}(M)$. Let $\mathbf{B} = \mathbf{M}_S$. Since the homomorphism functor is the right adjoint of the tensor functor, this can be recast as the following comonadic triangle



where $K(N) = (N \otimes_S M, N \otimes_S \varrho^M)$. The natural transformation φ associated to this triangle comes out as $\tilde{\psi}$ as defined in 5.19. Thus, the (finite) Galois property of (M, ϱ^M) is a necessary condition for the effectiveness of generalized descent associated with (M, ϱ^M) (note that M is a finitely generated and projective as a right A-module by [187, Corollary 3.7]).

5.28. Strong structure theorem over nonunital rings [102, Theorem 3.5]. Let $\mathcal{C}$ be an A-coring, B be a nonunital firm algebra. To any $(B, \mathcal{C})$-bicomodule (M, ϱ^M) (M is firm as a left B-module), one can associate the following comonadic triangle



where $K(N) = (N \otimes_B M, N \otimes_B \varrho^M)$. Applying 5.27 to this triangle, one concludes that K is an equivalence of categories if and only if M is a faithfully flat left B-module and, for all right A-modules N, the map

$$\mathrm{Hom}_A(M, N) \otimes_B B \otimes_B M \to N \otimes_A \mathcal{C}, \quad f \otimes_B b \otimes_B m \mapsto \sum f(bm_{(0)}) \otimes_A m_{(1)},$$

is an isomorphism.

5.4. *Morita theory and corings*

There are several Morita contexts, which can be associated to corings and comodules. They can be used to formulate structure theorems for Galois comodules or to study equivalences between module categories. In this section, we describe Morita contexts associated to corings.

5.29. Morita contexts. A *Morita context* or a set of *pre-equivalence data* is a sextuple $(A, B, N, M, \sigma, \tau)$, where

(i) A and B are algebras;
(ii) M is a (B, A)-bimodule and N is an (A, B)-bimodule;
(iii) $\tau : M \otimes_A N \to B$ is a B-bimodule map and $\sigma : N \otimes_B M \to A$ is an A-bimodule map rendering the following diagrams commutative

$$\begin{array}{ccc} N \otimes_B M \otimes_A N & \xrightarrow{\sigma \otimes_A N} & A \otimes_A N \\ {\scriptstyle N \otimes_B \tau}\downarrow & & \downarrow{\scriptstyle \simeq} \\ N \otimes_B B & \xrightarrow{\simeq} & N \end{array}, \qquad \begin{array}{ccc} M \otimes_A N \otimes_B M & \xrightarrow{\tau \otimes_B M} & B \otimes_B M \\ {\scriptstyle M \otimes_A \sigma}\downarrow & & \downarrow{\scriptstyle \simeq} \\ M \otimes_A A & \xrightarrow{\simeq} & M \end{array}.$$

A Morita context $(A, B, N, M, \sigma, \tau)$ is said to be *strict* provided both σ and τ are isomorphisms. Any strict Morita context induces and arises from an equivalence of categories of modules $\mathbf{M}_A$ and $\mathbf{M}_B$ (in which case one says that the algebras A and B are *Morita equivalent*). For a detailed discussion of Morita contexts or pre-equivalence data, the reader is referred to [14, Chapter II.3].

5.30. Morita context as a two-object category [27, Remark 3.2(1)]. A k-linear category $\mathbf{A}$ with two objects a and b gives rise to a Morita context as follows. The composition of morphisms makes k-modules $M = \mathbf{A}(a, b)$ and $N = \mathbf{A}(b, a)$ bimodules for k-algebras (with product given by composition) $A = \mathbf{A}(a, a)$ and $B = \mathbf{A}(b, b)$. Furthermore, the composition of maps defines bimodule maps

$$\tau : \mathbf{A}(a, b) \otimes_A \mathbf{A}(b, a) \to B, \qquad \sigma : \mathbf{A}(b, a) \otimes_B \mathbf{A}(a, b) \to A.$$

One easily checks that $(A, B, \mathbf{A}(b, a), \mathbf{A}(a, b), \sigma, \tau)$ is a Morita context. Clearly, there is a category of this kind behind any Morita context.

5.31. Morita context associated to a comodule I [59, Section 4]. The first Morita context connects the endomorphism algebra of a comodule with the right dual ring of a coring. Take a right comodule (M, ϱ^M) over an A-coring $\mathcal{C}$.

(i) Let $S = \mathrm{End}^{\mathcal{C}}(M)$ be the endomorphism algebra of M and B be the opposite algebra to the right dual ring $\mathcal{C}^*$; see 3.25.
(ii) Let $Q = \mathrm{Hom}^{\mathcal{C}}(\mathcal{C}, M)$ and $M^* = \mathrm{Hom}_A(M, A)$. Then Q is an (S, B)-bimodule with the multiplications given by, for all $s \in S$, $b \in B$, $q \in Q$, and $c \in \mathcal{C}$,

$$(qb)(c) = \sum q(b(c_{(1)})c_{(2)}), \qquad sq = s \circ q,$$

and M^* is a (B, S)-bimodule with the multiplications given by, for all $s \in S$, $b \in B$, $n \in M^*$, and $m \in M$,

$$(bn)(m) = \sum b(n(m_{(0)})m_{(1)}), \qquad ns = n \circ s.$$

(iii) Define an S-bimodule map

$$\sigma : Q \otimes_B M^* \to S, \qquad q \otimes_B n \mapsto \Big[m \mapsto \sum q(n(m_{(0)})m_{(1)})\Big],$$

and a B-bimodule map

$$\tau : M^* \otimes_S Q \to B, \qquad n \otimes_S q \mapsto n \circ q.$$

Then $\mathbb{M}^* = (S, B, Q, M^*, \sigma, \tau)$ is a Morita context. In view of 4.10, $B \simeq \mathrm{End}^{\mathcal{C}}(M)$ and $M^* \simeq \mathrm{Hom}^{\mathcal{C}}(M, \mathcal{C})$. Hence $\mathbb{M}^*$ is a context associated to the full subcategory of $\mathbf{M}^{\mathcal{C}}$ consisting of the two objects (M, ϱ^M) and $(\mathcal{C}, \Delta_{\mathcal{C}})$; cf. [189, Proposition 5.10].

A similar context can be constructed for a left $\mathcal{C}$-comodule.

5.32. Morita context associated to a comodule II [33, Section 2]. This Morita context connects the endomorphism algebra of a comodule with the left dual ring of a coring. Take a right comodule (M, ϱ^M) of an A-coring $\mathcal{C}$.

(i) Let $S = \mathrm{End}^{\mathcal{C}}(M)$ be the endomorphism algebra of M and B be the opposite algebra to the left dual ring ${}^*\mathcal{C}$, that is, $bb' = b' *^l b$; see 3.25.

(ii) Let

$$Q = \Big\{ q \in \mathrm{Hom}_A(M, B) \mid \forall m \in M, c \in \mathcal{C}, \ \sum q(m_{(0)})(c) m_{(1)} = \sum c_{(1)} q(m)(c_{(2)}) \Big\}.$$

Then Q is a (B, S)-bimodule with the multiplications $(bqs)(m) = bq(s(m))$, for all $s \in S$, $b \in B$, $q \in Q$, and $m \in M$. M is an (S, B)-bimodule with the multiplications $smb = \sum s(m_{(0)} b(m_{(1)}))$; cf. 4.33.

(iii) Define an S-bimodule map

$$\sigma : M \otimes_B Q \to S, \qquad m \otimes_B q \mapsto [m' \mapsto mq(m')],$$

and a B-bimodule map

$$\tau : Q \otimes_S M \to B, \qquad q \otimes_S m \mapsto q(m).$$

Then $\mathbb{M} = (S, B, M, Q, \sigma, \tau)$ is a Morita context.

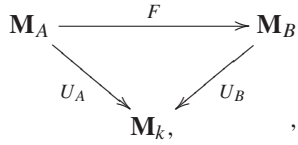
Recall from 4.6 that if M is finitely generated projective as a right A-module, then $(M^*, {}^{M^*}\mathcal{Q})$ is a left $\mathcal{C}$-comodule. In this case, the Morita context $\mathbb{M}$ is the same as the Morita context $\mathbb{M}^*$ in 5.31 associated to the *left* comodule $(M^*, {}^{M^*}\mathcal{Q})$; see [33, Remark 2.3].

In the case, $\mathcal{C}$ is a locally projective left A-module, the bimodule Q is isomorphic to the B-dual module $\mathrm{Hom}_B(M, B)$; see [33, Remark 2.2]. In this case, $\mathbb{M}$ is a Morita context associated to the full subcategory of $\mathbf{M}_B$ consisting of the two objects M and B (cf. 5.30).

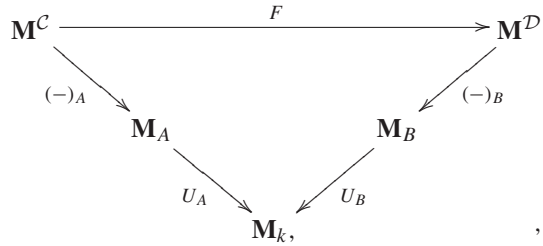
The context $\mathbb{M}$ controls the behavior of the functor $- \otimes_S M : \mathbf{M}_S \to \mathbf{M}^{\mathcal{C}}$. If the map σ in $\mathbb{M}$ is surjective, then $- \otimes_S M$ is a fully faithful functor [33, Theorem 2.6]. If $\mathcal{C}$ is finitely generated and projective as a left A-module, then $\mathbb{M}$ is strict if and only if (M, ϱ^M) is a finite Galois comodule and M is faithfully flat as a left S-module [59, Theorem 4.12], [33, Proposition 2.7] (and hence $- \otimes_S M$ is an equivalence by the finite Galois comodule structure theorem 5.22).

5.33. Coring extensions [41]. In case of algebras, an algebra map is the same as an *algebra extension*. More explicitly, given two algebras A and B, an algebra extension

or an algebra map $B \to A$ can be equivalently characterized as a k-linear functor $F : \mathbf{M}_A \to \mathbf{M}_B$ with the factorization property



where U_A and U_B are forgetful functors (cf. [152]). Extending this functorial characterization of algebra extensions, one says that a B-coring $\mathcal{D}$ is a *(right) functorial coring extension* of an A-coring $\mathcal{C}$ if there exists a k-linear functor $F : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}^{\mathcal{D}}$ with the factorization property



where $(-)_A$, and $(-)_B$ are forgetful functors. If the above diagram can be completed commutatively by a functor $\mathbf{M}_A \to \mathbf{M}_B$, then functorial coring extension is also known as a *cofunctor* (between internal cocategories); see [8, Section 4.2].

Following [41, Definition 2.1], $\mathcal{D}$ is called a *(right) coring extension* of $\mathcal{C}$ if there exists a right coaction $\kappa : \mathcal{C} \to \mathcal{C} \otimes_B \mathcal{D}$ such that $(\mathcal{C}, \kappa, \Delta_{\mathcal{C}})$ is a $(\mathcal{C}, \mathcal{D})$-bicomodule. A coring extension is denoted by $(\mathcal{D} \,|\, \mathcal{C}; \kappa)$. Every functorial coring extension gives rise to a coring extension (cf. [41, (2) $\Rightarrow$ (1) Theorem 2.6][3]). Given a functor $F : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}^{\mathcal{D}}$, the coaction κ is induced from the right regular coaction $\Delta_{\mathcal{C}}$, that is, $\kappa = F(\Delta_{\mathcal{C}})$. Conversely, given a right coring extension $(\mathcal{D} \,|\, \mathcal{C}; \kappa)$, the functor $F : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}^{\mathcal{D}}$ can be defined by

$$F : (M, \varrho^M) \mapsto (M, F(\varrho^M)),$$

$$F(\varrho^M) : \; M \xrightarrow{\varrho^M} M\Box_{\mathcal{C}}\mathcal{C} \xrightarrow{M\Box_{\mathcal{C}}\kappa} M\Box_{\mathcal{C}}(\mathcal{C} \otimes_B \mathcal{D}) \simeq M \otimes_B \mathcal{D},$$

provided that, for all $\mathcal{C}$-comodules M, the right B-module map $\varrho^M \otimes_A \mathcal{C} - M \otimes_A \Delta_{\mathcal{C}}$ is $\mathcal{D} \otimes_B \mathcal{D}$-pure. An extension $(\mathcal{D} \,|\, \mathcal{C}; \kappa)$ satisfying this purity condition is called a *pure (functorial) coring extension*. For example, if $\mathcal{D}$ is a flat left B-module, or $\mathcal{D}$ is a coring associated to an entwining structure as in 3.21, then any coring extension $(\mathcal{D} \,|\, \mathcal{C}; \kappa)$ is pure (hence functorial).

[3] Unfortunately, the proof of the claim (1) $\Rightarrow$ (2) of Theorem 2.6 in [41] is incorrect without further assumptions discussed in this section.

If $B = A$, then a coring map $\gamma : \mathcal{C} \to \mathcal{D}$ induces a coring extension $(\mathcal{D} \,|\, \mathcal{C}; (\mathcal{C} \otimes_A \gamma) \circ \Delta_{\mathcal{C}})$. Every coring extension $(\mathcal{D} \,|\, \mathcal{C}; \kappa)$, in which $\mathcal{C}$ is a right $\mathcal{D}$-comodule with its original A-multiplication, induces an A-coring morphism $\gamma : \mathcal{C} \to \mathcal{D}$ by $\gamma = (\varepsilon_{\mathcal{C}} \otimes_A \mathcal{D}) \circ \kappa$; see [28, Corrigendum, Lemma 1].

If $\mathcal{D}$ and $\mathcal{C}$ are corings associated to entwining structures (see 3.20 and 3.21), then extensions $(\mathcal{D} \,|\, \mathcal{C}; \kappa)$ correspond to morphisms of entwining structures in the sense of [153].

Take an entwining structure $(A, C)_\psi$ and let $\mathcal{C} = A \otimes C$ be the associated A-coring. Then C is a pure coring extension of $\mathcal{C}$, $(C \,|\, \mathcal{C}; A \otimes \Delta_C)$.

5.34. Morita context for a pure coring extension [33, Proposition 3.1]. Let $\mathcal{C}$ be an A-coring and $\mathcal{D}$ be a B-coring, and let $(\mathcal{D} \,|\, \mathcal{C}; \kappa)$ be a pure coring extension. Take a $(B, \mathcal{C})$-bicomodule (M, ϱ^M), and set $S = \mathrm{End}^{\mathcal{C}}(M)$. Since ϱ^M is left B-linear, there is an algebra map $j : B \to S$; hence S is a B-bimodule.

(i) View $R = {}_B\mathrm{Hom}_B(\mathcal{D}, S)$ as an algebra with the convolution product, $rr'(d) = \sum r(d_{(1)})r'(d_{(2)})$ and unit $j \circ \varepsilon_{\mathcal{D}}$, and set T to be the opposite of the endomorphism ring ${}^{\mathcal{C}}\mathrm{End}^{\mathcal{D}}(\mathcal{C})$ (i.e. the product in T is $tt' = t' \circ t$).

(ii) Let $N = {}_B\mathrm{Hom}^{\mathcal{D}}(\mathcal{D}, M)$ and

$$Q = \Big\{ q \in {}_A\mathrm{Hom}_B(\mathcal{C}, M^*) \mid \forall m \in M, c \in \mathcal{C}, \sum c_{(1)} q(c_{(2)})(m) = \sum q(c)(m_{(0)}) m_{(1)} \Big\},$$

where $M^* = \mathrm{Hom}_A(M, A)$ is an (A, B)-bimodule by $(afb)(m) = af(bm)$. Then N is an (R, T)-bimodule with multiplications, for all $n \in N$, $r \in R$, $t \in T$, and $d \in \mathcal{D}$,

$$(rn)(d) = \sum r(d_{(1)})(n(d_{(2)})), \quad (nt)(d) = \sum n(d)_{(0)} \varepsilon_{\mathcal{C}} \left(t \left(n \left(d \right)_{(1)} \right) \right),$$

and Q is a (T, R)-bimodule with multiplications, for all $q \in N$, $r \in R$, $t \in T$, and $c \in \mathcal{C}$,

$$qt(c) = \sum q(c^{[0]}) t(c^{[1]}), \qquad tq = q \circ t,$$

where $\sum c^{[0]} \otimes_A c^{[1]} = \kappa(c)$.

(iii) Define a T-bimodule map

$$\tau : Q \otimes_R N \to T, \quad q \otimes_R n \mapsto \Big[c \mapsto \sum c_{(1)} q({c_{(2)}}^{[0]}) \left(n \left({c_{(2)}}^{[1]} \right) \right) \Big],$$

and an R-bimodule map

$$\sigma : N \otimes_T Q \to R, \qquad n \otimes_T q \mapsto \Big[d \mapsto \sum n(d)_{(0)} q \left(n \left(d \right)_{(1)} \right) (-) \Big].$$

Then $\mathbb{BV} = (R, T, N, Q, \sigma, \tau)$ is a Morita context.

When interpreted as a category with two objects as in 5.30, the Morita context $\mathbb{BV}$ arises from a category whose objects are functors: the functor $F : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}^{\mathcal{D}}$, which

defines the pure coring extension $(\mathcal{D} \,|\, \mathcal{C}; \kappa)$, and the functor $\mathrm{Hom}^{\mathcal{C}}(M, -) \otimes_B \mathcal{D} : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}^{\mathcal{D}}$. Morphisms in this category are natural transformations.

5.35. Morita contexts and structure theorems for Galois comodules [33, Theorems 3.6, 4.1, 4.6]. Let $\mathbb{BV} = (R, T, N, Q, \sigma, \tau)$ be the Morita context constructed in 5.34.

(1) The map τ is surjective if and only if M is a Galois comodule, and, as an $(S, \mathcal{D})$-bicomodule, M is a direct summand of a finite number of copies of $S \otimes_B \mathcal{D}$. In particular, if τ is surjective, then the M-coinvariants functor $\mathrm{Hom}^{\mathcal{C}}(M, -) : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}_S$ is fully faithful.

(2) The M-coinvariants functor is an equivalence, provided the context $\mathbb{BV}$ is strict, and there exists a finite number of elements $r_i \in {}_B\mathrm{Hom}_B(\mathcal{D}, S)$, $d_i \in \mathcal{D}$, such that $\sum_i r_i(d_i) = 1_S$. In this case, M is a finitely generated and projective right A-module (hence a finite Galois comodule, but not necessarily faithfully flat as a left S-module).

5.36. Cleft bicomodules and structure theorems [33, Section 5]. Let $\mathcal{C}$ be an A-coring and $\mathcal{D}$ be a B-coring, and let $(\mathcal{D} \,|\, \mathcal{C}; \kappa)$ be a pure coring extension. Take a $(B, \mathcal{C})$-bicomodule (M, ϱ^M) and consider the associated Morita context $\mathbb{BV} = (R, T, N, Q, \sigma, \tau)$. If there exist $q \in Q$ and $n \in N$ such that

$$\tau(q \otimes_R n) = \mathcal{C}, \qquad \sigma(n \otimes_T q) = \varepsilon_{\mathcal{D}}(-)1_S,$$

then (M, ϱ^M) is said to be a *cleft bicomodule*.

Following [3, Definition 4.6], an entwining structure $(A, C)_\psi$ is said to be *cleft* if A is an entwined module, and there exists a convolution invertible right C-comodule map $C \to A$. Set $\mathcal{C} = A \otimes C$ to be the coring associated to $(A, C)_\psi$. Then A is a cleft bicomodule, with respect to the coring extension $(C \,|\, \mathcal{C}; A \otimes \Delta_C)$, if and only if $(A, C)_\psi$ is a cleft entwining structure.

Any cleft bicomodule is a Galois comodule. Furthermore, the M-coinvariants functor of a cleft bicomodule (M, ϱ^M) is fully faithful, and it is an equivalence, provided there exists a finite number of elements $r_i \in {}_B\mathrm{Hom}_B(\mathcal{D}, S)$, $d_i \in \mathcal{D}$, such that $\sum_i r_i(d_i) = 1_S$.

5.5. *Bialgebroids*

A *bialgebroid* is a generalization of the notion of a bialgebra to the case of a noncommutative base ring. Thus, this is a coring with a compatible algebra structure. The structure of bialgebroids and Hopf algebroids is treated in full in [25]. Here, we only give the most basic definitions.

5.37. A^{e}-rings. Let $A^{\mathrm{e}} = A \otimes A^{\mathrm{op}}$ be the enveloping algebra of an algebra A. An algebra H is an A^{e}-ring if and only if there exist an algebra map $s : A \to H$ and an antialgebra map $t : A \to H$ such that $s(a)t(b) = t(b)s(a)$, for all $a, b \in A$. A is called a *base algebra*, H a *total algebra*, s the *source* map and t the *target* map. To indicate explicitly the source and target maps, we write (H, s, t).

5.38. The Takeuchi product [179, 181]. Take a pair of A^{e}-rings (U, s_U, t_U) and (V, s_V, t_V), view U as a right A-module via the left multiplication by the target map t_U and view V as a left A-module via the left multiplication by the source map s_V. A *Takeuchi* $\times_A$*-product* is then defined as

$$U \times_A V := \left\{ \sum_i u_i \otimes_A v_i \in M \otimes_A N \mid \forall\, b \in A,\ \sum_i u_i t_U(b) \otimes_A v_i = \sum_i u_i \otimes_A v_i s_V(b) \right\}.$$

The importance of the notion of the Takeuchi $\times_A$-product is a direct consequence of the following observation ([179, Proposition 3.1], [181, Proposition 3.1]).

LEMMA. *For any pair of* A^{e}*-rings* (U, s_U, t_U) *and* (V, s_V, t_V)*, the* A^{e}*-bimodule* $U \times_A V$ *is an* A^{e}*-ring with source* $a \mapsto s_U(a) \otimes_A 1_V$ *and target* $a \mapsto 1_U \otimes_A t_V(a)$*, associative product*

$$\left(\sum\nolimits_i u^i \otimes_A v^i\right)\left(\sum\nolimits_j \tilde{u}^j \otimes_A \tilde{v}^j\right) = \sum\nolimits_{i,j} u^i \tilde{u}^j \otimes_A v^i \tilde{v}^j,$$

and unit $1_U \otimes_A 1_V$.

5.39. Bialgebroids [131, 181]. Let $(\mathcal{H}, s, t)$ be an A^{e}-ring. View $\mathcal{H}$ as an A-bimodule by $ah = s(a)h$, $ha = t(a)h$, for all $a \in A$, $h \in \mathcal{H}$. We say that $(\mathcal{H}, s, t, \Delta_{\mathcal{H}}, \varepsilon_{\mathcal{H}})$ is a *left* A*-bialgebroid* if

(1) $(\mathcal{H}, \Delta_{\mathcal{H}}, \varepsilon_{\mathcal{H}})$ is an A-coring;
(2) $\mathrm{Im}(\Delta_{\mathcal{H}}) \subseteq \mathcal{H} \times_A \mathcal{H}$ and the corestriction of $\Delta_{\mathcal{H}}$ to $\Delta_{\mathcal{H}} : \mathcal{H} \to \mathcal{H} \times_A \mathcal{H}$ is an algebra map;
(3) $\varepsilon_{\mathcal{H}}(1_{\mathcal{H}}) = 1_A$, and, for all $g, h \in \mathcal{H}$,

$$\varepsilon_{\mathcal{H}}(gh) = \varepsilon_{\mathcal{H}}\left(gs(\varepsilon_{\mathcal{H}}(h))\right) = \varepsilon_{\mathcal{H}}\left(gt(\varepsilon_{\mathcal{H}}(h))\right).$$

5.40. The monoidal structure of the category of (co) modules [165]. Let $(\mathcal{H}, s, t)$ be an A^{e}-ring. Then $\mathcal{H}$ is an A-bialgebroid if and only if ${}_{\mathcal{H}}\mathbf{M}$ is a monoidal category such that the forgetful functor $F : {}_{\mathcal{H}}\mathbf{M} \to {}_A\mathbf{M}_A$ is strict monoidal.

Furthermore, if $\mathcal{H}$ is an A-bialgebroid, the category ${}^{\mathcal{H}}\mathbf{M}$ of left comodules of the A-coring $\mathcal{H}$ is monoidal with the tensor product defined by

$$(M, {}^M\varrho) \otimes (N, {}^N\varrho) = (M \otimes_A N, {}^{M\otimes_A N}\varrho),$$

where, writing ${}^M\varrho(m) = \sum m_{(-1)} \otimes_A m_{(0)}$, ${}^N\varrho(n) = \sum n_{(-1)} \otimes_A n_{(0)}$, for the left $\mathcal{H}$-coactions,

$${}^{M\otimes_A N}\varrho : m \otimes_A n \mapsto \sum m_{(-1)} n_{(-1)} \otimes_A m_{(0)} \otimes_A n_{(0)}.$$

The right A-multiplication on a left $\mathcal{H}$-comodule $(M, {}^M\varrho)$ is given by

$$ma = \sum \varepsilon_{\mathcal{H}}(m_{(-1)} s(a)) m_{(0)}.$$

This is a unique right A-module structure which makes $\mathcal{H}$ an A-bimodule and such that the image of the coaction ${}^{M}\varrho$ is contained in the Takeuchi product $\mathcal{H} \times_A M$; see [23, Section 2.2].

6. Applications

6.1. *Hopf modules*

Recent interest in the theory of corings has arisen from their connection with Hopf-type modules, studied in Hopf algebra theory since the 1960s. This connection is provided through the observation that all such Hopf-type modules can be interpreted as entwined modules and hence as comodules of corings. In this section, we merely list the known types of Hopf modules and describe the corresponding entwining structures. More details can be found in [53, 68].

6.1. Bialgebras and Hopf algebras. A *bialgebra* can be defined as a special case of an A-bialgebroid by specifying A to be k. Explicitly, a bialgebra is a k-module B such that $(B, \mu_B, 1_B)$ is a k-algebra, $(B, \Delta_B, \varepsilon_B)$ is a k-coalgebra, and both Δ_B and ε_B are algebra maps, that is, for all $a, b \in B$,

$$\varepsilon_B(ab) = \varepsilon_B(a)\varepsilon_B(b), \qquad \varepsilon_B(1_B) = 1,$$

and

$$\Delta_B(ab) = \sum a_{(1)}b_{(1)} \otimes a_{(2)}b_{(2)}, \qquad \Delta_B(1_B) = 1_B \otimes 1_B.$$

A bialgebra B is called a *Hopf algebra* if there exists a k-linear map $S : B \to B$ such that

$$\mu_B \circ (B \otimes S) \circ \Delta_B = 1_B \circ \varepsilon_B = \mu_B \circ (S \otimes B) \circ \Delta_B.$$

The map S is called an *antipode*. For more details about bialgebras and Hopf algebras, the reader is referred, for example, to [75].

6.2. Comodule algebras and module coalgebras. Let B be a bialgebra. An algebra A is called a *right B-comodule algebra* if there exists an algebra map $\varrho^A : A \to A \otimes B$ such that (A, ϱ^A) is a right B-comodule. The algebra map property of ϱ^A means that, for all $a, b \in A$,

$$\varrho^A(ab) = \sum a_{(0)}b_{(0)} \otimes a_{(1)}b_{(1)}, \qquad \varrho^A(1_B) = 1_A \otimes 1_B.$$

Dually, a coalgebra C is called a *right B-module coalgebra* if C is a right B-module and, for all $b \in B$ and $c \in C$,

$$\Delta_C(cb) = \sum c_{(1)}b_{(1)} \otimes c_{(2)}b_{(2)}, \qquad \varepsilon_C(cb) = \varepsilon_C(c)\varepsilon_B(b).$$

6.3. Hopf modules. Let B be an algebra and a coalgebra. The k-linear map

$$\psi : B \otimes B \to B \otimes B, \quad b' \otimes b \mapsto \sum b_{(1)} \otimes b'b_{(2)},$$

entwines B with B if and only if B is a bialgebra. An entwined module corresponding to $(B, B)_\psi$ is a right B-comodule M such that for all $m \in M$ and $b \in B$,

$$\varrho^M(mb) = \sum m_{(0)} b_{(1)} \otimes m_{(1)} b_{(2)}.$$

Any such M is known as a *Hopf module*, a notion introduced and studied already in [129, 177].

Thus if B is a bialgebra, then $\mathcal{C} = B \otimes B$ is a B-coring with coproduct $\Delta_\mathcal{C} = B \otimes \Delta_B$, counit $\varepsilon_\mathcal{C} = B \otimes \varepsilon_B$, and the B-bimodule structure

$$b(b' \otimes b'')c = \sum bb'c_{(1)} \otimes b''c_{(2)},$$

for $b, b', b'', c \in B$. The category of right $\mathcal{C}$-comodules is the same as the category of (right) B-Hopf modules.

6.4. Relative Hopf modules. Let B be a bialgebra and let A be an algebra together with a k-linear map $\varrho^A : A \to A \otimes B$, $a \mapsto \sum a_{(0)} \otimes a_{(1)}$. Then the map

$$\psi : B \otimes A \to A \otimes B, \quad b \otimes a \mapsto \sum a_{(0)} \otimes ba_{(1)},$$

entwines A with B if and only if (A, ϱ^A) is a right B-comodule algebra (cf. [63, Lemma 4.1]). The corresponding entwined modules are right A-modules and right B-comodules (M, ϱ^M) such that, for all $m \in M$ and $a \in A$,

$$\varrho^M(ma) = \sum m_{(0)} a_{(0)} \otimes m_{(1)} a_{(1)}.$$

Such a module is known as a *relative* (B, A)*-Hopf module*, a notion introduced in [182] (cf. [83]).

Thus if A is a right B-comodule algebra, then $\mathcal{C} = A \otimes B$ is an A-coring with coproduct $\Delta_\mathcal{C} = A \otimes \Delta_B$, counit $\varepsilon_\mathcal{C} = A \otimes \varepsilon_B$, and the A-bimodule structure, for $a, a', a'' \in A$ and $b \in B$,

$$a''(a \otimes b)a' = \sum a''aa'_{(0)} \otimes ba'_{(1)}.$$

The category of right $\mathcal{C}$-comodules is the same as the category of relative (B, A)-Hopf modules.

Dualizing the above construction, one entwines a bialgebra B and a right B-module coalgebra C by the map

$$\psi : C \otimes B \to B \otimes C, \quad c \otimes b \mapsto \sum b_{(1)} \otimes cb_{(2)}.$$

The corresponding B-coring $\mathcal{C} = B \otimes C$ has a coproduct $\Delta_\mathcal{C} = B \otimes \Delta_C$, counit $\varepsilon_\mathcal{C} = B \otimes \varepsilon_C$, and the (B, B)-bimodule structure, for $b, b', b'' \in B$, and $c \in C$,

$$b''(b \otimes c)b' = \sum b''bb'_{(1)} \otimes cb'_{(2)}.$$

Comodules over $\mathcal{C}$ are identified with right C-comodules right B-modules M such that the coaction is compatible with the action via

$$\varrho^M(mb) = \sum m_{(0)}b_{(1)} \otimes m_{(1)}b_{(2)}.$$

Such an M is called a $[B, C]$-*Hopf module* [83].

6.5. Doi–Koppinen Hopf modules. All the above examples of Hopf modules are special cases of *Doi–Koppinen Hopf modules* introduced in [84, 126]. In this case, one starts with a triple (B, A, C), where B is a bialgebra, A is a right B-comodule algebra, and C is a right B-module coalgebra. A is then entwined with C by the map

$$\psi : C \otimes A \to A \otimes C, \qquad c \otimes a \mapsto \sum a_{(0)} \otimes ca_{(1)},$$

where $\varrho^A(a) = \sum a_{(0)} \otimes a_{(1)} \in A \otimes B$ is the B-coaction on A. Thus the corresponding A-coring $\mathcal{C} = A \otimes C$ has the coproduct $\Delta_{\mathcal{C}} = A \otimes \Delta_C$, counit $\varepsilon_{\mathcal{C}} = A \otimes \varepsilon_C$, and the A-bimodule structure, for $a, a', a'' \in A$, and $c \in C$,

$$a'(a'' \otimes c)a = \sum a'a''a_{(0)} \otimes ca_{(1)}.$$

Right $\mathcal{C}$-comodules are identified with right A-modules and right C-comodules M with the action-coaction compatibility, for all $a \in A$ and $m \in M$,

$$\varrho^M(ma) = \sum m_{(0)}a_{(0)} \otimes m_{(1)}a_{(1)}.$$

The triple (B, A, C) satisfying above conditions is known as a *Doi–Koppinen datum*. Doi–Koppinen data were considered in [126, Definition 2.1], where they were termed *opposite smash data*. This was the first (recorded) step toward understanding Doi–Koppinen Hopf modules as representations of an algebraic structure (hence, toward viewing Hopf modules as comodules of a coring). Finally, the entwining structure associated to a Doi–Koppinen datum was first introduced in [36]. The left dual algebra of the coring $A \otimes C$ is isomorphic to the Koppinen (opposite) smash product algebra [126]. Many properties and applications of Doi–Koppinen structures are studied in the monograph [68].

6.6. Yetter–Drinfeld modules. Given a Hopf algebra H with antipode S, a *(right–right) Yetter–Drinfeld module* is a right H-module and a right H-comodule (M, ϱ^M) such that, for all $m \in M$ and $h \in H$,

$$\varrho^M(mh) = \sum m_{(0)}h_{(2)} \otimes (Sh_{(1)})m_{(1)}h_{(3)}$$

[158, 201]. As explained in [66, 67], Yetter–Drinfeld modules are an example of Doi–Koppinen Hopf modules, hence entwined modules. The entwining map is

$$\psi : H \otimes H \to H \otimes H, \quad h' \otimes h \mapsto \sum h_{(2)} \otimes (Sh_{(1)})h'h_{(3)}.$$

Thus Yetter–Drinfeld modules are right comodules over the H-coring $\mathcal{C} = H \otimes H$ with coproduct $\Delta_{\mathcal{C}} = H \otimes \Delta_H$, counit $\varepsilon_{\mathcal{C}} = H \otimes \varepsilon_H$, and the H-bimodule structure, for all $h, h', h'', k \in H$,

$$k(h'' \otimes h')h = \sum kh''h_{(2)} \otimes (Sh_{(1)})h'h_{(3)}.$$

6.7. Anti-Yetter–Drinfeld modules. This most recent addition to Hopf modules appeared in [107, 108, 113] as coefficients for Hopf-cyclic homology [77]. Let H be a Hopf algebra with a bijective antipode S. An *anti-Yetter–Drinfeld module* is a right H-module and H-comodule M with the action-coaction compatibility condition, for all $a \in H, m \in M$,

$$\varrho^M(ma) = \sum m_{(0)}a_{(2)} \otimes S^{-1}(a_{(1)})m_{(1)}a_{(3)}.$$

As explicitly explained in [43], anti-Yetter–Drinfeld modules are entwined modules for the entwining map $\psi : H \otimes H \to H \otimes H$ given by

$$\psi(c \otimes a) = \sum a_{(2)} \otimes S^{-1}(a_{(1)})ca_{(3)},$$

for all $a, c \in H$. The existence of entwining means that $\mathcal{C} = H \otimes H$ is an H-coring with coproduct $\Delta_{\mathcal{C}} = H \otimes \Delta_H$, counit $\varepsilon_{\mathcal{C}} = H \otimes \varepsilon_H$, and the H-bimodule structure, for all $h, h', h'', k \in H$,

$$k(h'' \otimes h')h = \sum kh''h_{(2)} \otimes S^{-1}(h_{(1)})h'h_{(3)}.$$

Anti-Yetter–Drinfeld modules are right comodules over the coring $\mathcal{C}$.

In the definition of both Yetter–Drinfeld and anti-Yetter–Drinfeld modules, the important property is that the antipode is an antibialgebra map and the identity map is a bialgebra map. Following this observation, the notion of (α, β)-*equivariant* C-*comodules* was introduced in [150]. The defining data comprise a bialgebra A, A-bimodule coalgebra C, a bialgebra map $\alpha : A \to A$, and an antibialgebra map $\beta : A \to A$ (i.e. β is both an antialgebra and anticoalgebra map). All these data give rise to an entwining map $\psi : C \otimes A \to A \otimes C$,

$$\psi(c \otimes a) = \sum a_{(2)} \otimes \beta(a_{(1)})c\alpha(a_{(3)}),$$

and thus to an A-coring $\mathcal{C} = A \otimes C$ with coproduct $\Delta_{\mathcal{C}} = A \otimes \Delta_C$, counit $\varepsilon_{\mathcal{C}} = A \otimes \varepsilon_C$, and the A-multiplication

$$a'(a'' \otimes c)a = \sum a'a''a_{(2)} \otimes \beta(a_{(1)})c\alpha(a_{(3)}).$$

The (α, β)-equivariant C-comodules are simply comodules over $\mathcal{C}$.

6.8. Weak Doi–Koppinen modules. The notion of a *weak bialgebra* was introduced in [32, 145] (cf. [30]), by weakening the axioms of a bialgebra. A *weak* k-*bialgebra* H is a k-module with a k-algebra structure $(\mu_H, 1_H)$ and a k-coalgebra structure

$(\Delta_H, \varepsilon_H)$ such that Δ_H is a multiplicative map and

$$(\Delta_H \otimes H) \circ \Delta_H(1_H) = \sum 1_{H(1)} \otimes 1_{H(2)} 1_{H(1')} \otimes 1_{H(2')}$$
$$= \sum 1_{H(1)} \otimes 1_{H(1')} 1_{H(2)} \otimes 1_{H(2')},$$

$$\varepsilon_H(hkl) = \sum \varepsilon_H(hk_{(1)})\varepsilon_H(k_{(2)}l) = \sum \varepsilon_H(hk_{(2)})\varepsilon_H(k_{(1)}l)$$

for all $h, k, l \in H$. Here $\Delta_H(1_H) = \sum 1_{H(1)} \otimes 1_{H(2)} = \sum 1_{H(1')} \otimes 1_{H(2')}$.

Given a weak bialgebra H, a *right H-comodule algebra* is defined in [22, Definition 2.1] (cf. [56, Proposition 4.12]) as a k-algebra A with a right H coaction $\varrho^A : A \to A \otimes H$ such that, for all $a, b \in A$,

$$\varrho^A(ab) = \varrho^A(a)\varrho^A(b), \quad \sum a_{(0)} \otimes \varepsilon_H(1_{H(1)} a_{(1)}) 1_{H(2)} = \sum 1_{A(0)} a \otimes 1_{A(1)}.$$

Dually, a *right H-module coalgebra* is defined as a coalgebra $(C, \Delta_C, \varepsilon_C)$ and a right H-module such that, for all $h, k \in H$, and $c \in C$,

$$\Delta_C(ch) = \sum c_{(1)} h_{(1)} \otimes c_{(2)} h_{(2)}, \qquad \varepsilon_C(chk) = \sum \varepsilon_C(ch_{(2)})\varepsilon(h_{(1)}k).$$

Given a right H-comodule algebra A and a right H-module coalgebra C, one defines a k-linear map

$$\psi : C \otimes A \to A \otimes C, \quad c \otimes a \mapsto \sum a_{(0)} \otimes c a_{(1)}.$$

The triple (A, C, ψ) is a right-right weak entwining structure (cf. [56, Theorem 4.14]), and hence there is an associated A-coring $\mathcal{C}$ as described in 3.23. The comodules over this coring are identified with right A-modules and right C-comodules M that satisfy the action–coaction compatibility condition

$$\varrho^M(ma) = \sum m_{(0)} a_{(0)} \otimes m_{(1)} a_{(1)}.$$

Such modules were first introduced and studied in [22] and are termed *weak Doi–Koppinen Hopf modules*. A triple comprising a weak bialgebra H, a right H-comodule algebra A, and a right H-module coalgebra C is known as a *weak Doi–Koppinen datum*.

6.9. Hopf modules as comodules of a coring. Once it is realized that all known types of Hopf modules can be seen as comodules of specific corings, the general results of the coring theory can be used to gain information about Hopf modules. In all cases, the coring arises from an entwining structure $(A, C)_\psi$ hence is of the type $\mathcal{C} = A \otimes C$. The defining adjunction $(-)_A, - \otimes_A \mathcal{C}$ can be identified with the adjunction

$$(-)_A : \mathbf{M}_A^C(\psi) \to \mathbf{M}_A, \qquad - \otimes C : \mathbf{M}_A \to \mathbf{M}_A^C(\psi),$$

where, for any right A-module V, $V \otimes C$ is a right C-comodule by $V \otimes \Delta_C$ and a right A-module by $(v \otimes c)a = v\psi(c \otimes a)$. Using knowledge about the category of comodules over a coring, one thus derives conditions for a category of specific Hopf

modules to be Grothendieck or to be a full subcategory of the category of modules over the corresponding smash (i.e. twisted convolution) algebra. Furthermore, properties of the functors $-\otimes C$ and $(-)_A$ such as separability or the Frobenius property are derived from properties of the defining adjunction of the coring $\mathcal{C} = A \otimes C$ and thus correspond to $\mathcal{C}$ being a cosplit, coseparable, or Frobenius coring. More geometric properties such as the interpretation of Hopf modules as modules with flat connections discussed in [121] are a direct consequence of the interpretation of comodules over a coring as flat connections. Similar comments apply to weak Doi–Koppinen Hopf modules.

There are, however, limitations as to this use of corings in description of Hopf modules. First, every class of Hopf modules has its specific properties, which are not captured by the definition of a coring. For example, the category of Yetter–Drinfeld modules is a braided monoidal category. This property and its consequences cannot be derived from the general properties of corings (without specifying the coring precisely). Second, the notion of an entwining structure is self-dual in the sense that replacing A by C, μ_A by Δ_C, 1_A by ε_C, and reversing all the arrows in the Diagram 3.20 produces exactly the same conditions. Using this self-duality, one can translate properties of functors $(-)_A$, $-\otimes C$ into corresponding properties of functors

$$(-)^C : \mathbf{M}_A^C(\psi) \to \mathbf{M}^C, \qquad -\otimes A : \mathbf{M}^C \to \mathbf{M}_A^C(\psi),$$

where $(-)^C$ is the forgetful functor and, for any right C-comodule V, the C-coaction for $V \otimes A$ is $v \otimes c \mapsto \sum v_{(0)}\psi(v_{(1)} \otimes a)$. Obviously, there can be no functor such as $(-)^C$ for a general A-coring $\mathcal{C}$. Thus to have a more unifying point of view on these properties of Hopf modules, one needs to study algebras in the category of bicomodules over a coalgebra C or C-rings; see Section 7.2.

6.2. *Noncommutative differential geometry*

We have already seen in 3.16 that corings with a group-like element give rise to (semifree) differential graded algebras. Furthermore, as explained in 4.23, comodules over a coring with a group-like element are the same as modules with a flat connection (with respect to the induced differential graded algebra). This is one of the situations, where corings relate to noncommutative geometry. The second is the appearance of comodules over a coring (of the anti-Yetter–Drinfeld type) as coefficients for the Hopf-cyclic homology; see 6.7. Another one is the geometry of noncommutative principal bundles. As this is not a text about geometry or addressed to geometers, we recall briefly the basic notions, and we provide an algebraic description of a geometric situation that makes it possible to relate *noncommutative principal bundles* to Galois type extensions. For a detailed and easy accessible noncommutative description of classical bundles, we refer to [15].

6.10. Free actions and principal bundles. Let X be a compact Hausdorff space and G a compact group. The action of G on X is said to be *free* if the property $xg = x$ for any $x \in X$ implies that $g = u$, the group identity.

Consider the map $\widetilde{F}_X^G : X \times G \to X \times X$, $(x, g) \mapsto (x, xg)$. The freeness of the G-action on X means, equivalently, that the map $\widetilde{F}_X^G$ is injective. Write $Y = X/G$ and let $\pi : X \to Y$, $x \mapsto xG$, be the canonical surjection. Let $X \times_Y X := \{(x, y) \in X \times X \mid \pi(x) = \pi(y)\}$. Clearly, $X \times_Y X = \mathrm{Im}\, \widetilde{F}_X^G$; so, the action of G on X is free if and only if

$$\tilde{F}_X^G : X \times G \to X \times_Y X, \qquad (x, g) \mapsto (x, xg),$$

is bijective.

Given a free action of G on X, the function $\hat{\tau} : X \otimes_Y X \to G$, $(x, xg) \mapsto g$, is called a *translation function*. X is called a *G-principal bundle over Y* if and only if $\hat{\tau}$ is continuous. X is called the *total space*, Y is called the *base space*, and G is called the *structure group* or the *fiber*.

6.11. Algebras of functions on X, G, Y and $X \times_Y X$. Consider now algebras of (complex, polynomial – thus we assume that all spaces are embedded in $\mathbb{C}^n$) functions $A = \mathcal{O}(X)$ and $H = \mathcal{O}(G)$. Use the identification $\mathcal{O}(G \times G) \simeq \mathcal{O}(G) \otimes \mathcal{O}(G)$ to see that H is a Hopf algebra with the coproduct $\Delta_H(f)(g, h) = f(gh)$ and counit $\varepsilon_H(f) = f(u)$, where u is the neutral element of G. The action of G on X translates into the coaction of H on A $\varrho^A : A \to A \otimes H$, $\varrho^A(a)(x, g) = a(xg)$. Obviously, since both H and A are commutative algebras, ϱ^A is an algebra map; hence A is a right H-comodule algebra.

Let $B = \mathcal{O}(Y)$, the algebra of (polynomial) functions on $Y = X/G$. We can view B as a subalgebra of A via the pullback of $\pi : X \to Y$, that is, via the function $\pi^* : B \to A, b \mapsto b \circ \pi$. The map π^* is injective, since $b \neq b'$ means that there exists at least one coset xG such that $b(xG) \neq b'(xG)$. But $xG = \pi(x)$; so we conclude that $\pi^*(b)(x) \neq \pi^*(b')(x)$. Furthermore, $a \in \pi^*(B)$ if and only if $a(xg) = a(x)$, for all $x \in X$, $g \in G$. This is the same as $\varrho^A(a)(x, g) = (a \otimes 1)(x, g)$, for all $x \in X$, $g \in G$, where $1 : G \to k$ is the unit function $1(g) = 1$ [the unit of the Hopf algebra $H = \mathcal{O}(G)$]. Thus, we can identify B with the coinvariants:

$$B = A^{\mathrm{coH}} := \{a \in A \mid \varrho^A(a) = a \otimes 1_H\}.$$

Since ϱ^A is an algebra map, A^{coH} can be described equivalently as

$$A^{\mathrm{coH}} = \{b \in A \mid \forall a \in A, \quad \varrho^A(ba) = b\varrho^A(a)\}.$$

B is a subalgebra of A (via the map π^*), hence it acts on A through the inclusion map, $(bab')(x) = b(\pi(x))a(x)b'(\pi(x))$. We can identify $\mathcal{O}(X \times_Y X)$ with $\mathcal{O}(X) \otimes_{\mathcal{O}(Y)} \mathcal{O}(X) = A \otimes_B A$, via the map $\theta(a \otimes_B a')(x, y) = a(x)a'(y)$, with $\pi(x) = \pi(y)$.

6.12. Principal bundles as Hopf–Galois extensions. With X, G, Y as before, the action of G on X is free if and only if the map

$$F_X^{G*} : \mathcal{O}(X \times_Y X) \to \mathcal{O}(X \times G), \qquad f \mapsto f \circ F_X^G,$$

is bijective. In view of the identifications in 6.11, F_X^{G*} is the canonical map

$$\mathsf{can}_A : a \otimes_B a' \mapsto [(x, g) \mapsto a(x)a'(xg)] = a\varrho^A(a').$$

Thus the action of G on X is free if and only if the map can_A is bijective, that is, the extension of algebras $B \subseteq A$ is a *Hopf-Galois extension by* H. Here $A = \mathcal{O}(X)$, $H = \mathcal{O}(G)$, and $B = \mathcal{O}(X/G)$.

To deal with the topological aspects, one needs to consider C^*-algebras of continuous functions (in which, by the Stone-Weierstrass theorem, the polynomial algebras are dense subalgebras) and complete tensor products.

6.13. Coalgebra-Galois extensions. The algebraic formulation of the notion of a principal bundle leads to the following definition. Let C be a k-coalgebra and A a k-algebra and a right C-comodule with coaction $\varrho^A : A \to A \otimes C$. Let B be the subalgebra of *coinvariants* of A,

$$B := \{b \in A \mid \forall a \in A,\ \varrho^A(ba) = b\varrho^A(a)\}.$$

We say that $B \subseteq A$ is a *coalgebra extension by* C or a C*-extension*. The extension $B \subseteq A$ is called a *coalgebra-Galois extension* (or a C*-Galois extension*) if the canonical left A-module, right C-comodule map

$$\mathsf{can}_A :\ A \otimes_B A \to A \otimes C, \qquad a \otimes_B a' \mapsto a\varrho^A(a'),$$

is bijective.

The notion of a C-Galois extension in this generality and in this form was introduced in [46], extending the definition of a *coalgebra principal bundle* in [49]. In the case when C is a Hopf algebra and A is a C-comodule algebra (i.e. the coaction ϱ^A is an algebra map), a coalgebra-Galois extension is known as a *Hopf–Galois extension*, a notion introduced in [127] (and rooted in an algebraic approach to Galois theory [73]). The discovery of examples of noncommutative spaces such as quantum spheres [155] made it clear that the notion of a Hopf–Galois extension is too rigid to capture all the geometrically interesting algebras, hence the introduction of a more general Galois-type extensions in [49].

6.14. Canonical entwining for a coalgebra-Galois extension [46, Theorem 2.7]. *Let A be a C-Galois extension of B. Then, there exists a unique entwining map $\psi : C \otimes A \to A \otimes C$ such that $A \in \mathbf{M}_A^C(\psi)$ by the structure maps μ_A and ϱ^A. The map ψ is called the* canonical entwining map *associated to a C-Galois extension $B \subseteq A$.*

Sketch of proof. The entwining map ψ is given by

$$\psi : c \otimes a \mapsto \mathsf{can}_A\left(\mathsf{can}_A^{-1}(1_A \otimes c)\, a\right),$$

where can_A is the canonical map in 6.13. □

6.15. Coalgebra-Galois extension as a Galois coring. By 6.14 and 3.21, if $B \subseteq A$ is a C-Galois extension, then $\mathcal{C} = A \otimes C$ is an A-coring with coproduct $\Delta_{\mathcal{C}} = A \otimes \Delta_C$, counit $\varepsilon_{\mathcal{C}} = A \otimes \varepsilon_C$, and A-multiplications $a''(a' \otimes c)a = a''a'$ $\mathsf{can}_A\left(\mathsf{can}_A^{-1}(1_A \otimes c)\, a\right)$. Since A is an entwined module, it is a right $\mathcal{C}$-comodule; hence there exists a group-like element g; see 4.12. Since the $\mathcal{C}$-coaction (induced

by g) on A is the same as the C-coaction ϱ^A, necessarily, $g = \varrho^A(1_A)$. Again, since A is an entwined module, one finds that C-coinvariants B can be described differently as $B = \{b \in A, \mid \varrho^A(b) = b\varrho^A(1_A)\}$. Remember that the $\mathcal{C}$-coaction on A is $\varrho^A(a) = ga$. Hence b is an element of B if and only if

$$gb = \varrho^A(b) = b\varrho^A(1_A) = bg.$$

Thus B is the same subalgebra of A as g-coinvariants A_g in A. The canonical map can_A in 6.13 coincides with the canonical map can_A in 5.20. Therefore, if $B \subseteq A$ is a C-Galois extension, then A is a Galois $\mathcal{C}$-comodule [with respect to the group-like element $\varrho^A(1_A)$].

Conversely, if we are starting from an entwining structure $(A, C)_\psi$ such that $\mathcal{C} = A \otimes C$ is a Galois A-coring, that is, A is a Galois comodule, then A is a C-Galois extension (of C-coinvariants).

Thus the general theory of coalgebra-Galois extensions (understood as rudimentary noncommutative principal bundles) is a special case of the theory of (finite) Galois comodules.

6.16. Principal extensions. Although a coalgebra-Galois extension is a first approximation to a noncommutative principal bundle, it does not have all desired geometric properties. One of missing properties is the existence of a connection or local triviality. The definition of a noncommutative principal bundle, which appears to fill these gaps, was proposed in [47]. Since in differential geometry one is interested in algebras over a field (typically the field of complex numbers) in this item, we assume that k is a field.

A C-Galois extension $B \subseteq A$ is called a *principal extension* if

(a) the canonical entwining map ψ is bijective,
(b) the multiplication map $B \otimes A \to A$ has a left B-linear right C-colinear section (i.e. A is a C-equivariantly projective left B-module),
(c) there exists a group-like element $x \in C$ such that $\varrho^A(1_A) = 1_A \otimes x$.

From the algebraic point of view, (c) is a minor condition and is needed only for differential-geometric reasons. Since A is an entwined module, it is equivalent to the requirement that $\varrho^A(a) = \psi(x \otimes a)$.

If $B \subseteq A$ is a principal extension, then A is a principal Galois comodule of the associated A-coring $\mathcal{C} = A \otimes C$ (this is the original motivation for introducing principal Galois comodules over a coring).

In geometrically interesting cases (e.g. when C is a compact quantum group), the coalgebra C is a coseparable coalgebra (i.e. it has a cointegral; cf. 5.3). Applying the structure theorem for principal comodules 5.26 to this particular setup, one finds that for an extension to be principal, it is sufficient to assume that the canonical map can_A is surjective. More precisely, there is:

THEOREM [40, Theorem 4.6], [166, Theorem 5.9]. *Let k be a field and $(A, C)_\psi$ an entwining structure such that the map ψ is bijective. Suppose that $x \in C$ is a group-like element and view A as a right C-comodule with the coaction $\varrho^A : A \to A \otimes C$,*

$a \mapsto \psi(x \otimes a)$. *If C is a coseparable coalgebra and the (lifted) canonical map*

$$\widetilde{\mathsf{can}}_A : A \otimes A \to A \otimes C, \qquad a \otimes a' \mapsto a\varrho^A(a'),$$

is surjective, then A is a principal C-extension of the coinvariants B.

Principal extensions allow one to make a connection between the corepresentation theory of coalgebras C and the cyclic homology of B. This connection is provided by the *Chern–Galois character*. Its description goes beyond the scope of this chapter, and we refer the interested reader to the note [47] and to paper [26] for more detailed and general formulation.

6.3. *Noncommutative algebraic geometry*

To the best of our knowledge, there are at least three approaches to noncommutative algebraic geometry in which corings play some role. The terminology is specific to each one of these approaches.

6.3.1. Approach through Grothendieck categories. This is an approach to noncommutative algebraic geometry advocated by Rosenberg [163, 164] and Van den Bergh [186]. Although corings do not seem to play as fundamental role as in two other approaches described in subsequent sections, the category of comodules appears to have a geometric meaning. Also, specific properties of the category of comodules have geometric interpretation. Here, we simply translate some of the properties of comodules into the language of noncommutative algebraic geometry. On the algebraic geometry side, we follow the terminology of [172, 173].

6.17. Spaces as Grothendieck categories. A *noncommutative space* X is a Grothendieck category, denoted $\mathbf{Mod}X$. A standard classical (motivating) example of a space is the category of quasi-coherent sheaves on a quasi-compact, quasi-separated scheme.

Since a category of right modules over an algebra A is a Grothendieck category, it is a noncommutative space. A space X equivalent (as a category) to a module category $\mathbf{M}_A$ is called a *noncommutative affine space*. The algebra A is known as a *coordinate ring of X*. Making use of the Gabriel–Popescu theorem [156], one concludes that X is an affine non-commutative space if and only if $\mathbf{Mod}X$ has a finitely generated projective generator U. In this case $A \simeq \mathbf{Mod}X(U, U)$; see [174, p. 223].

6.18. Maps between noncommutative spaces. A map $X \to Y$ between two noncommutative spaces is defined as an adjoint pair of functors $f = (f^*, f_*)$, $f^* : \mathbf{Mod}Y \to \mathbf{Mod}X$, $f_* : \mathbf{Mod}X \to \mathbf{Mod}Y$. The functor f^* is referred to as the *inverse image* and f_* is termed the *direct image*.

The map $f = (f^*, f_*)$ is said to be *affine* if f_* is faithful and has a right adjoint.

6.19. Comodules as a noncommutative space. Let $\mathcal{C}$ be an A-coring. In view of 4.30, if $\mathcal{C}$ is a flat left A-module, then the category of right $\mathcal{C}$-comodules $\mathbf{M}^{\mathcal{C}}$

is a Grothendieck category. Hence, in this case, $\mathbf{M}^{\mathcal{C}}$ is a noncommutative space. Furthermore, since the forgetful functor $(-)_A : \mathbf{M}^{\mathcal{C}} \to \mathbf{M}_A$ has the right adjoint $-\otimes_A \mathcal{C} : \mathbf{M}_A \to \mathbf{M}^{\mathcal{C}}$, there is a morphism of noncommutative spaces $((-)_A, -\otimes_A \mathcal{C})$ from $\mathbf{M}^{\mathcal{C}}$ to $\mathbf{M}_A$.

If $\mathcal{C}$ is finitely generated and projective as a left A-module, then, in view of 4.38, $\mathbf{M}^{\mathcal{C}}$ is an affine noncommutative space with coordinate ring $B = ({}^*\mathcal{C})^{\mathrm{op}}$.

In view of the finite Galois comodule structure theorem 5.22, if $\mathcal{C}$ is flat as a left A-module and there exists a finite Galois comodule (M, ϱ^M) which is faithfully flat as a left module over its coinvariant ring $\mathrm{End}^{\mathcal{C}}(M)$, then $\mathbf{M}^{\mathcal{C}}$ is an affine noncommutative space with the coordiante ring $\mathrm{End}^{\mathcal{C}}(M)$.

6.20. Weakly closed subspaces. A *subspace* Y of a noncommutative space X is a full subcategory $\mathbf{Mod}Y$ of $\mathbf{Mod}X$, which is closed under direct sums, kernels, and isomorphisms. A subspace Y is said to be *weakly closed* if $\mathbf{Mod}Y$ is closed under subquotients and the inclusion functor $\mathbf{Mod}Y \to \mathbf{Mod}X$ has a right adjoint; see [173, Definition 2.4].

For an A-coring $\mathcal{C}$, $\mathbf{M}^{\mathcal{C}} = \sigma[\mathcal{C}]$ if and only if $\mathcal{C}$ is a locally projective left A-module. Wisbauer's σ-category is closed under direct sums, kernels, isomorphisms, and subquotients. Thus $\mathbf{M}^{\mathcal{C}}$ is the smallest weakly closed subspace of the affine noncommutative space with the coordinate ring, $B = ({}^*\mathcal{C})^{\mathrm{op}}$ containing $\mathcal{C}$ if and only if $\mathcal{C}$ is a locally projective left A-module.

6.21. Integral spaces. A noncommutative space X is said to be *integral* if $\mathbf{Mod}X = \sigma[E]$ for an indecomposable injective object E of $\mathbf{Mod}X$ whose endomorphism ring is a division ring; see [172, Definition 3.1], [151, Definition 7.1].

Thus, if $\mathcal{C}$ is an A-coring, locally projective as a left A-module, indecomposable and injective as a $\mathcal{C}$-comodule and such that the right dual algebra $\mathcal{C}^*$ is a division ring, then $\mathbf{M}^{\mathcal{C}}$ is an integral noncommutative space. Note that if $\mathcal{C}$ is a simple $\mathcal{C}$-comodule, then it is indecomposable and $\mathcal{C}^*$ is a division ring (by the Schur lemma). On the other hand, any (locally projective as a left A-module) coring $\mathcal{C}$ over an injective ring A is injective in $\mathbf{M}^{\mathcal{C}}$; [53, 18.19].

In view of [151, Theorem 3.7], if $\mathcal{C}$ is finitely generated as a left A-module, then $\mathbf{M}^{\mathcal{C}}$ is an integral (affine) space if and only if the left dual algebra ${}^*\mathcal{C}$ is a prime ring, that is $\mathcal{C}$ is a prime coring in the sense of [98].

6.3.2. Approach through covers. This is an approach to noncommutative algebraic geometry in which corings play a central role. Initiated in [123] it is fully developed in the series of papers [124].

6.22. Finite covers and quasi-coherent sheaves. An A-coring $\mathcal{C}$ that is faithfully flat as a left A-module is called a *finite cover*. Covers are denoted by $(A, \mathcal{C})$. Right $\mathcal{C}$-comodules are called *quasi-coherent sheaves*. In the context of noncommutative geometry, $\mathbf{M}^{\mathcal{C}}$ is denoted by $\mathbf{Qcoh}(A, \mathcal{C})$. The category of covers **Covers** is defined as a full subcategory of the opposite of the category of corings **Crg** consisting of all corings, which are faithfully flat as left modules (over their base rings); compare 3.7. Thus a morphism of covers $(A, \mathcal{C}) \to (B, \mathcal{D})$ is a pair of maps (α, γ), $\alpha : B \to A$, $\gamma : \mathcal{D} \to \mathcal{C}$ such that α is algebra map and $A \otimes_B \gamma \otimes_B A$ is an A-coring map.

6.23. Refinements and space covers. Given an inclusion of algebras $\alpha : B \to A$ and a B-coring $\mathcal{D}$, consider the base algebra extension coring $\mathcal{C} = A \otimes_B \mathcal{D} \otimes_B A$; see 3.6. There is a coring morphism $(\gamma, \alpha) : \mathcal{D} \to \mathcal{C}$, where $\gamma : d \mapsto 1_A \otimes_B d \otimes_B 1_A$. If A and $\mathcal{D}$ are faithfully flat as left B-modules, then $\mathcal{C}$ is faithfully flat as a left A-module; hence it is a cover. In this case, the coring morphism (γ, α) induces a morphism of covers $(A, \mathcal{C}) \to (B, \mathcal{D})$ (note the direction of the arrow). Any morphism of this type is called a *refinement morphism*. The class of refinement morphisms in **Covers** is denoted by Ref.

A cover $(A, \mathcal{C})$ is called a *space cover* if there is an A-coring epimorphism $s_{\mathcal{C}} : A \otimes A \to \mathcal{C}$ (where $A \otimes A$ is the Sweedler coring corresponding to $1_A : k \to A$). In other words, a space cover is an A-coring $\mathcal{C}$ generated as a bimodule by a group-like element and faithfully flat as a left A-module. Space covers are denoted by $(\mathcal{C}, s_{\mathcal{C}})$. Morphisms of space covers are morphisms of covers compatible with structure maps $s_{\mathcal{C}}$, $s_{\mathcal{D}}$ [i.e. sending the group-like element $s_{\mathcal{C}}(1_A \otimes 1_A)$ to the group-like element $s_{\mathcal{D}}(1_B \otimes 1_B)$]. The resulting category of space covers is denoted by $\mathbf{Covers}^{\mathbf{sp}}$.

6.24. Equivalent morphisms of covers and noncommutative spaces. Two morphisms $(\alpha, \gamma), (\alpha', \gamma') : (\mathcal{C}, s_{\mathcal{C}}) \to (\mathcal{D}, s_{\mathcal{D}})$ of space covers are said to be *equivalent* if, for any element $\sum_i x_i \otimes y_i \in \ker\, s_{\mathcal{D}}$,

$$\sum_i \alpha(x_i)\alpha'(y_i) = \sum_i \alpha'(x_i)\alpha(y_i) = 0.$$

A quotient category of $\mathbf{Covers}^{\mathbf{sp}}$ by this equivalence relation is denoted by $\widetilde{\mathbf{Covers}}^{\mathbf{sp}}$. A *noncommutative space* is defined as an object of the localization of $\widetilde{\mathbf{Covers}}^{\mathbf{sp}}$ with respect to the class of refinements Ref.

6.3.3. Approach through monoidal categories. This, the most recent proposal for noncommutative algebraic geometry, was made in [133] and is based on the premise that noncommutative schemes should be understood as monoidal categories; see 2.13.

6.25. Noncommutative schemes. A *noncommutative scheme* is an Abelian monoidal category $(\mathbf{Qcoh}(X), \odot_X, \mathcal{O}_X)$ ($\odot_X$ is multiplication, $\mathcal{O}_X$ is a unit). A morphism of noncommutative schemes is an isoclass of an additive monoidal functor $G : (\mathbf{Qcoh}(X), \odot_X, \mathcal{O}_X) \to (\mathbf{Qcoh}(Y), \odot_Y, \mathcal{O}_Y)$, which has a left adjoint. If A is a k-algebra, then $({}_A\mathbf{M}_A, \otimes_A, A)$ is called a *non-commutative affine scheme*. A noncommutative affine scheme corresponding to A is denoted by $\mathbf{Spec}(A)$.

6.26. Morphisms of affine schemes and representations. A $B|A$-coring $\mathcal{C}$ (cf. 3.27) is termed a *representation* of an algebra B over an algebra A. By the Eilenberg-Watts theorem (see [192, Theorem 6] or e.g. [174, Chapter IV, Proposition 10.1]), every additive functor between bimodule categories ${}_A\mathbf{M}_A \to {}_B\mathbf{M}_B$, which has a left adjoint, can be represented as $G = {}_A\mathrm{Hom}_A(\mathcal{C}, -)$. Thus, by 3.27, there is a one-to-one correspondence between morphisms of affine schemes and representations (or $B|A$-corings).

6.27. Flat covers. In the set-up of 6.26, the left adjoint of G is the tensor functor $F = - \otimes_{B^{\mathrm{op}}\otimes B} \mathcal{C}$. The A-coring structure on $\mathcal{C}$ induces an A-coring structure on

$$\mathcal{D} = F(B) = B \otimes_{B^{\mathrm{op}}\otimes B} \mathcal{C} = \mathcal{C}/[\mathcal{C}, B],$$

where $[\mathcal{C}, B]$ is the commutator submodule of $\mathcal{C}$. For any B-bimodules M and N, the unit η and counit ψ of the adjunction (F, G) induce a morphism (natural in M and N)

$$F^2_{M,N} \quad : \quad F(M \otimes_B N) \xrightarrow{F(\eta_M \otimes_B \eta_N)} F(GF(M) \otimes_B GF(N))$$
$$\xrightarrow{F(G^2_{F(M),F(N)})} FG(F(M) \otimes_A F(N)) \xrightarrow{\psi_{F(M)\otimes_A F(N)}} F(M)\Box_{\mathcal{D}} F(N).$$

A morphism of affine schemes G is said to be a *flat cover* if the left adjoint F of G is faithful and the maps $F^2_{M,N}$ are isomorphisms for all B-bimodules M, N.

6.28. Galois extensions. Let $(\mathcal{D}, \iota)$ be a $B|A$-coring. The map $\iota : B \to {}^*\mathcal{D}^*$ is called a *noncommutative Galois ring extension* if there exists an A-coring $\mathcal{C}$ and an A-coring morphism $\phi : \mathcal{D} \to \mathcal{C}$ such that

(a) the category of $\mathcal{C}$-bicomodules is an Abelian monoidal category with multiplication $\Box_{\mathcal{C}}$;
(b) $(\mathcal{D} \otimes_A \phi \otimes_A \mathcal{D}) \circ ((\Delta_{\mathcal{D}} \otimes_A \mathcal{D}) - (\mathcal{D} \otimes_A \Delta_{\mathcal{D}}))$ is a pure morphism of B-bimodules;
(c) B is isomorphic to the algebra of coinvariants ${}^{\mathcal{C}}\mathrm{Hom}^{\mathcal{C}}(\mathcal{D}, \mathcal{C})$ (cf. 4.18);
(d) $B \otimes \mathcal{D} \to \mathcal{D}\Box_{\mathcal{C}}\mathcal{D}$, $b \otimes d \mapsto \sum d_{(1)}\iota(b)(d_{(2)}) \otimes_A d_{(3)}$, is an isomorphism;
(e) $\mathcal{D}$ is a faithfully flat left $B^{\mathrm{op}} \otimes B$-module;
(f) the map of A-corings

$$\bar{\phi} : \mathcal{D}/[\mathcal{D}, B] \to \mathcal{C},$$

defined by $\phi = \bar{\phi} \circ p$, where $p : \mathcal{D} \to \mathcal{D}/[\mathcal{D}, B]$ is the canonical epimorphism, is an isomorphism;
(g) for all A-bimodules M, the map

$$\theta_M : {}_A\mathrm{Hom}_A(\mathcal{D}, M) \otimes_{B^{\mathrm{op}}\otimes B} \mathcal{D} \to \mathcal{C} \otimes_A M \otimes_A \mathcal{C},$$
$$f \otimes d \mapsto \sum \phi(d_{(1)}) \otimes_A f(d_{(2)}) \otimes_A \phi(d_{(3)}),$$

is an isomorphism (natural in M).

From the point of view of Galois comodules studied earlier, the key (Galois) properties are (c) and (f). More specifically, let $\mathcal{C}$ be an A-coring and (M, ϱ^M) be a right $\mathcal{C}$-comodule that is finitely generated and projective as an A-module. Set $B = \mathrm{End}^{\mathcal{C}}(M)$, $T = \mathrm{End}_A(M)$, and $\mathcal{D} = M^* \otimes M$ [the (finite) comatrix coring]. Since M is a finitely generated and projective right A-module,

$${}^*\mathcal{D}^* = {}_A\mathrm{Hom}_A(M^* \otimes M, A) \simeq \mathrm{End}_A(M) = T,$$

so that there is an algebra inclusion $\iota : B \hookrightarrow {}^*\mathcal{D}^*$. Thus $(\mathcal{D}, \iota)$ is a $B|A$-coring. Let

$$\phi : \mathcal{D} = M^* \otimes M \to \mathcal{C}, \qquad \xi \otimes m \mapsto \sum \xi(m_{(0)})m_{(1)},$$

be the (lifted) canonical map; see 5.26, 4.18. Then $B = {}^{\mathcal{C}}\mathrm{Hom}^{\mathcal{C}}(\mathcal{D}, \mathcal{C})$ (cf. 4.18); so condition (c) is automatically satisfied. Finally, $\mathcal{D}/[\mathcal{D}, B] = M^* \otimes_B M$, $p : M^* \otimes M \to M^* \otimes_B M$ is the canonical surjection, hence $\bar{\phi} = \mathsf{can}_M$. Thus, (f) is the same as the finite Galois condition in 5.19. For a finite Galois comodule (M, ϱ^M), all other conditions follow, provided the functor $M^* \otimes_B - \otimes_B M : {}_B\mathbf{M}_B \to {}_A\mathbf{M}_A$ is faithful and exact, and the map $T \to T \otimes_B T, t \mapsto 1_T \otimes_B t - t \otimes_B 1_T$ is pure in ${}_B\mathbf{M}_B$.

From the noncommutative algebraic geometry point of view, all the conditions (a)–(g) are needed, since as proved in [133, Theorem 3], there is a one-to-one correspondence between noncommutative Galois ring extensions and flat covers in the category of noncommutative affine schemes.

7. Extensions and dualizations

7.1. *Weak and lax corings*

As explained in 3.21, if $(A, C)_\psi$ is an entwining structure, then $A \otimes C$ is an A-coring. On the other hand, if (A, C, ψ) is a weak entwining structure not all of $A \otimes C$ is a coring, but only a direct summand of it. Weak corings were introduced to answer the following question: what is a coring-like structure of $A \otimes C$? On the other hand, recent interest in partial Galois theory led to the introduction of *lax* and *partial entwining structures*, and, consequently, *lax corings*. In this section, we briefly describe these notions.

7.1. Weak corings. Let A be a k-algebra, and let $\mathcal{C}$ be an A-bimodule, which can be *nonunital* as both left and right A-module, that is, there can exist $c \in \mathcal{C}$ such that $1_A c \neq c$ or $c 1_A \neq c$. $\mathcal{C}$ is called a *weak A-coring* [196] if there exist A-bimodule maps $\Delta_{\mathcal{C}} : \mathcal{C} \to \mathcal{C} \otimes_A A \otimes_A \mathcal{C}$ and $\varepsilon_{\mathcal{C}} : \mathcal{C} \to A$ such that the following diagrams

$$\begin{array}{ccc}
\mathcal{C} & \xrightarrow{\Delta_{\mathcal{C}}} & \mathcal{C} \otimes_A A \otimes_A \mathcal{C} \\
{\scriptstyle \Delta_{\mathcal{C}}}\downarrow & & \downarrow{\scriptstyle \mathcal{C}\otimes_A A\otimes_A \Delta_{\mathcal{C}}} \\
\mathcal{C} \otimes_A A \otimes_A \mathcal{C} & \xrightarrow{\Delta_{\mathcal{C}}\otimes_A A\otimes_A \mathcal{C}} & \mathcal{C} \otimes_A A \otimes_A \mathcal{C} \otimes_A A \otimes_A \mathcal{C},
\end{array}$$

$$\begin{array}{ccccc}
 & & \mathcal{C} & & \\
 & {\scriptstyle \Delta_{\mathcal{C}}}\swarrow & \downarrow & \searrow{\scriptstyle \Delta_{\mathcal{C}}} & \\
\mathcal{C} \otimes_A A \otimes_A \mathcal{C} & & {\scriptstyle 1_A\otimes_A -} \quad {\scriptstyle -\otimes_A 1_A} & & \mathcal{C} \otimes_A A \otimes_A \mathcal{C} \\
 & {\scriptstyle \varepsilon_{\mathcal{C}}\otimes_A A\mathcal{C}}\searrow & \downarrow & \swarrow{\scriptstyle \mathcal{C}A\otimes_A \varepsilon_{\mathcal{C}}} & \\
 & & \mathcal{C} & &
\end{array}$$

are commutative. Here $1_A \otimes_A -$ denotes the map $c \mapsto 1_A c$ and $- \otimes_A 1_A$ is the map $c \mapsto c 1_A$. If $\mathcal{C}$ is a left (resp. right) unital A-bimodule, then we say that $\mathcal{C}$ is a *left* (resp. *right*) *unital weak A-coring*.

For any weak A-coring $\mathcal{C}$, $A\mathcal{C}A$ is a unital A-bimodule, and it is an A-coring. Similarly, if $\mathcal{C}$ is a left (resp. right) unital weak A-coring, then $\mathcal{C}A$ (resp. $A\mathcal{C}$) is an A-coring.

As for corings, the coproduct on $\mathcal{C}$ induces an associative multiplication on the dual modules $\mathcal{C}^* = \mathrm{Hom}_A(\mathcal{C}, A)$, ${}^*\mathcal{C} = {}_A\mathrm{Hom}(\mathcal{C}, A)$, ${}^*\mathcal{C}^* = {}_A\mathrm{Hom}_A(\mathcal{C}, A)$. ${}^*\mathcal{C}^*$ is a unital algebra with unit $\varepsilon_{\mathcal{C}}$, and it coincides with the two-sided dual of the coring $A\mathcal{C}A$. $\mathcal{C}^*$ and ${}^*\mathcal{C}$ are nonunital, but $\varepsilon_{\mathcal{C}}$ is a central idempotent in each one of them (hence they are firm rings). The dual rings of the coring $A\mathcal{C}A$ coincide with the unital subrings of dual rings of $\mathcal{C}$ generated (as modules) by $\varepsilon_{\mathcal{C}}$.

7.2. Weak corings and weak entwining structures. Weak corings were introduced in [196] to understand fully the correspondence between weak entwining structures (see 3.22) and corings, in particular to obtain a statement similar to that in 3.21. This aim is achieved in the following

PROPOSITION. *Let A be an algebra and $(C, \Delta_C, \varepsilon_C)$ be a coalgebra. Set $\mathcal{C} = A \otimes C$ and view it as a (unital) left A-module with the product $a(a' \otimes c) = aa' \otimes c$.*

If (A, C, ψ) is a weak entwining structure, then $A \otimes C$ is a (nonunital) right A-module by $(a' \otimes c)a = a'\,\psi(c \otimes a)$, and $\mathcal{C}$ is a left unital weak A-coring with coproduct and counit

$$\begin{array}{rccccc} \Delta_{\mathcal{C}} : A \otimes C & \to & (A \otimes C) \otimes_A (A \otimes C) & \simeq & (A \otimes C)1_A \otimes C, \\ a \otimes c & \mapsto & \sum (a \otimes c_{(1)}) \otimes_A (1_A \otimes c_{(2)}) & \mapsto & \sum (a \otimes c_{(1)})1_A \otimes c_{(2)}, \end{array}$$

$$\begin{array}{rccccc} \varepsilon_{\mathcal{C}} : A \otimes C & \to & (A \otimes C)1_A & \to & A\,, \\ a \otimes c & \mapsto & (a \otimes c)1_A & \mapsto & (A \otimes \varepsilon_C)((a \otimes c)1_A). \end{array}$$

Conversely, if $\mathcal{C}$ is a left unital weak A-coring with $\Delta_{\mathcal{C}}$ and $\varepsilon_{\mathcal{C}}$ as defined above, then $C \otimes A \to A \otimes C$, $c \otimes a \mapsto (1_A \otimes c)a$, is a weak entwining map for A and C.

7.3. Comments on weak comodules. Similarly as for corings, one defines comodules over a weak A-coring $\mathcal{C}$ or *weak comodules*. These are defined as pairs (M, ϱ^M), where M is a nonunital right A-module, and $\varrho^M : M \to M \otimes_A A\mathcal{C}$ is a right A-module map such that the following diagrams

$$\begin{array}{ccc} M & \xrightarrow{\varrho^M} & M \otimes_A A\mathcal{C} \\ {\scriptstyle \varrho^M}\downarrow & & \downarrow{\scriptstyle M \otimes_A \Delta_{\mathcal{C}}} \\ M \otimes_A A\mathcal{C} & \xrightarrow{\varrho^M \otimes_A A\mathcal{C}} & M \otimes_A A\mathcal{C} \otimes_A A\mathcal{C}, \end{array} \qquad \begin{array}{ccc} M & \xrightarrow{\varrho^M} & M \otimes_A A\mathcal{C} \\ & {\scriptstyle -\otimes_A 1_A}\searrow & \downarrow{\scriptstyle M \otimes_A \varepsilon_{\mathcal{C}}} \\ & & M \otimes_A A, \end{array}$$

commute. Morphisms between comodules over weak corings are defined in the same way as for comodules over corings. If (M, ϱ^M) is a right $\mathcal{C}$-comodule, then $(MA, \varrho^M|_{MA})$ is a right comodule over the A-coring $A\mathcal{C}A$. As a consequence of this, the category of weak comodules has similar properties to the category of comodules over a coring. For example, it always has cokernels and direct sums, and it has kernels provided that $A\mathcal{C}$ is a flat left A-module. The tensor functor $- \otimes_A \mathcal{C}A$ from the

category of A-modules to $\mathcal{C}$-comodules is the right adjoint of the functor $-\otimes_A A$ (the defining adjunctions). For more detailed studies of weak corings and comodules, the reader is referred to [53, Chapter 6] and [196].

7.4. Lax corings. Let A be a k-algebra, and let $\mathcal{C}$ be an A-bimodule, which is *nonunital* as a right A-module, that is, although it is required that, for all c, $1_A c = c$, it is not required that $c1_A = c$. Write $\underline{\mathcal{C}}$ for $\mathcal{C}1_A = \mathcal{C}A$. $\underline{\mathcal{C}}$ is a unital A-bimodule, and it is a direct summand of $\mathcal{C}$ as an A-bimodule. Denote the inclusion $\underline{\mathcal{C}} \subseteq \mathcal{C}$ by ι. Since $\mathcal{C}$ is a unital left A-module, any A-bimodule map $\Delta_{\mathcal{C}} : \mathcal{C} \to \mathcal{C} \otimes_A \mathcal{C}$ restricts to the map $\Delta_{\underline{\mathcal{C}}} : \underline{\mathcal{C}} \to \underline{\mathcal{C}} \otimes_A \underline{\mathcal{C}}$.

$\mathcal{C}$ is called a left unital *lax A-coring* [63] if there exist a coassociative A-bimodule map $\Delta_{\mathcal{C}} : \mathcal{C} \to \mathcal{C} \otimes_A \mathcal{C}$ and an A-bimodule map $\varepsilon_{\mathcal{C}} : \mathcal{C} \to A$ such that $\underline{\mathcal{C}}$ is an A-coring with coproduct $\Delta_{\underline{\mathcal{C}}}$ and counit $\varepsilon_{\mathcal{C}} \circ \iota$.

A left unital weak A-coring is an example of a left unital lax A-coring.

7.5. Lax and partial entwining structures and corings. Let A be an algebra and C be a coalgebra. Consider a map $\psi : C \otimes A \to A \otimes C$, and write, for all $c \in C$, $a \in A$, $\psi(c \otimes a) = \sum_\psi a_\psi \otimes c^\psi = \sum_\Psi a_\Psi \otimes c^\Psi$. Assume that for all $a, b \in A$, and $c \in C$,

$$\sum_\psi (ab)_\psi \otimes c^\psi = \sum_{\psi,\Psi} a_\psi b_\Psi \otimes c^{\psi\Psi},$$

$$\sum_{\psi,\Psi} a_\psi 1_{A\Psi} \otimes {c^\psi}_{(1)}{}^\Psi \otimes {c^\psi}_{(2)} = \sum_{\psi,\Psi} a_{\psi\Psi} \otimes {c_{(1)}}^\Psi \otimes {c_{(2)}}^\psi.$$

The first of these conditions means that $A \otimes C$ is a nonunital right A-module with multiplication $(a' \otimes c)a = a'\psi(c \otimes a)$. Write $\mathcal{C} = A \otimes C$ and define

$$\Delta_{\mathcal{C}} : A \otimes C \to A \otimes C \otimes_A A \otimes C \simeq (A \otimes C)1_A \otimes C,$$

$$a \otimes c \mapsto \sum a\psi(c_{(1)} \otimes 1_A) \otimes c_{(2)}.$$

If the map ψ satisfies the following additional conditions, for all $a \in A$, $c \in C$,

$$\sum_\psi \varepsilon_C(c^\psi)1_{A\psi}a = \sum_\psi \varepsilon_C(c^\psi)a_\psi, \quad \sum_\psi 1_{A\psi} \otimes c^\psi = \sum_{\psi,\Psi} \varepsilon_C({c_{(1)}}^\Psi)1_{A\psi\Psi} \otimes {c_{(2)}}^\psi,$$

then the triple (A, C, ψ) is called a *lax entwining structure*. The existence of a lax entwining structure is equivalent to the statement that $\mathcal{C}$ is a left unital lax A-coring with coproduct $\Delta_{\mathcal{C}}$ and counit $\varepsilon_{\mathcal{C}} : A \otimes C \mapsto A, a \otimes c \mapsto \sum_\psi a1_{A\psi}\varepsilon_C(c^\psi)$ (and the canonical left A-multiplication).

If the map ψ satisfies the following additional condition,

$$(\varepsilon_C \otimes A) = (A \otimes \varepsilon_C) \circ \psi,$$

then the triple (A, C, ψ) is called a *partial entwining structure*. The existence of a partial entwining structure is equivalent to the statement that $\mathcal{C}$ is a left unital lax A-coring with coproduct $\Delta_{\mathcal{C}}$ and counit $A \otimes \varepsilon_C$.

Given a lax entwining structure (A, C, ψ), a *lax entwined module* is a right A-module M together with a k-linear map $\varrho^M : M \to M \otimes C$, written on elements as $\varrho^M(m) = \sum m^{[0]} \otimes m^{[1]}$, such that, for all $m \in M$, $a \in A$,

(a) $(M \otimes \varepsilon_C) \circ \varrho^M = M$,
(b) $(\varrho^M \otimes C) \circ \varrho^M(m) = \sum_{\psi,\Psi} m^{[0]} 1_{A\psi\Psi} \otimes m^{[1]}{}_{(1)}{}^{\Psi} \otimes m^{[1]}{}_{(2)}{}^{\psi}$,
(c) $\varrho^M(ma) = \sum m^{[0]} \psi(m^{[1]} \otimes a)$.

The category with objects lax entwined modules and morphisms A-linear, C-colinear maps is denoted by $\mathbf{M}_A^C(\psi)^{\mathsf{lax}}$. This category is isomorphic to the category $\mathbf{M}^{\underline{\mathcal{C}}}$, where $\underline{\mathcal{C}} = \underline{A \otimes C} := (A \otimes C)1_A$ is the A-coring associated to a left unital lax A-coring $\mathcal{C} = A \otimes C$; see 7.4.

7.6. Partial entwining structures and partial (co)actions. Let H be a bialgebra.

(1) An algebra A is said to be a *partial right H-comodule algebra* if there exists a k-linear map $\varrho^A : A \to A \otimes H$, written on elements as $\varrho^A(a) = \sum a^{[0]} \otimes a^{[1]}$, such that, for all $a, b \in A$,
   (a) $\varrho^A(ab) = \sum a^{[0]} b^{[0]} \otimes a^{[1]} b^{[1]}$,
   (b) $\sum a^{[0]} \varepsilon_H(a^{[1]}) = a$,
   (c) $\sum \varrho^A(a^{[0]}) \otimes a^{[1]} = a^{[0]} 1_A{}^{[0]} \otimes a^{[1]}{}_{(1)} 1_A{}^{[1]} \otimes a^{[1]}{}_{(2)}$.
(2) A coalgebra C is said to be a *partial right H-module coalgebra* if there exists a nonunital right action of H on C, $\varrho_C : C \otimes H \to C$, $c \otimes h \mapsto ch$, such that, for all $c \in C$ and $h \in H$,
   (a) $\varepsilon_C(ch) = \varepsilon_C(c)\varepsilon_H(h)$,
   (b) $\sum (ch)_{(1)} 1_H \otimes (ch)_{(2)} = \sum c_{(1)} h_{(1)} \otimes c_{(2)} h_{(2)}$.

If A is a partial right H-comodule algebra and C is a partial right H-module coalgebra, then there is a partial entwining structure (A, C, ψ), where

$$\psi : C \otimes A \to A \otimes C, \qquad c \otimes a \mapsto \sum a^{[0]} \otimes ca^{[1]}.$$

Partial entwining structures of this (or its dual) kind arise in the context of partial group actions; see [63, Section 5], [57, 85].

7.7. Comments on the structure and usage of lax corings. Since lax corings are defined through corings, that is, $\mathcal{C}$ is a left unital lax coring if and only if $\underline{\mathcal{C}}$ is an A-coring, they share many properties with the latter. For example, their dual modules are (nonunital) rings. Using this correspondence, one can study Frobenius or Galois properties of lax corings. In the case of lax corings associated to partial and lax entwining structures, where the category of lax entwined modules is isomorphic to the category of comodules of the associated A-coring $\underline{A \otimes C}$, such properties are a consequence of the (nonlax) coring theory. For further details, the reader is referred to [63].

7.2. *C-rings*

A notion of a *C-ring* or a *C-algebra* arises as a dualization of that of coring; corings are coalgebras in the category of bimodules, C-rings are algebras in the category of

bicomodules. To avoid complications arising from the nonassociativity of cotensor product (for coalgebras over rings), in this section, we assume that k is a field. In this case, given a coalgebra C, for all C-bicomodules $(M, \varrho^M, {}^M\varrho)$, $(N, \varrho^N, {}^N\varrho)$, the cotensor product $M\Box_C N$ is a C-bicomodule with coactions $M\Box_C\varrho^N$ and ${}^M\varrho\Box_C N$. Furthermore, for any bicomodule morphisms f, g, also $f\Box_C g$ is a bicomodule morphism (i.e. $-\Box_C-$ is a functor ${}^C\mathbf{M}^C \times {}^C\mathbf{M}^C \to {}^C\mathbf{M}^C$), and the category of C-bicomodules is a monoidal category, $({}^C\mathbf{M}^C, \Box_C, C)$; see 2.13.

7.8. C-rings [37, Section 6]. Let C be a k-coalgebra and let $\mathcal{A}$ be a C-bicomodule with coactions ${}^{\mathcal{A}}\varrho : \mathcal{A} \to C \otimes \mathcal{A}$ and $\varrho^{\mathcal{A}} : \mathcal{A} \to \mathcal{A} \otimes C$. $\mathcal{A}$ is called a *C-ring* if there are two bicomodule maps $\mu_{\mathcal{A}} : \mathcal{A}\Box_C\mathcal{A} \to \mathcal{A}$ and $\eta_{\mathcal{A}} : C \to \mathcal{A}$ rendering the following diagrams commutative

$$\begin{array}{ccc} \mathcal{A}\Box_C\mathcal{A}\Box_C\mathcal{A} & \xrightarrow{\mu_{\mathcal{A}}\Box_C\mathcal{A}} & \mathcal{A}\Box_C\mathcal{A} \\ {\scriptstyle \mathcal{A}\Box_C\mu_{\mathcal{A}}}\downarrow & & \downarrow{\scriptstyle \mu_{\mathcal{A}}} \\ \mathcal{A}\Box_C\mathcal{A} & \xrightarrow{\mu_{\mathcal{A}}} & \mathcal{A}, \end{array}$$

$$\begin{array}{ccc} \mathcal{A} & \xrightarrow{\varrho^{\mathcal{A}}} & \mathcal{A}\Box_C C \\ {\scriptstyle \mathcal{A}}\downarrow & & \downarrow{\scriptstyle \mathcal{A}\Box_C\eta_{\mathcal{A}}} \\ \mathcal{A} & \xleftarrow{\mu_{\mathcal{A}}} & \mathcal{A}\Box_C\mathcal{A}, \end{array} \qquad \begin{array}{ccc} \mathcal{A} & \xrightarrow{{}^{\mathcal{A}}\varrho} & C\Box_C\mathcal{A} \\ {\scriptstyle \mathcal{A}}\downarrow & & \downarrow{\scriptstyle \eta_{\mathcal{A}}\Box_C\mathcal{A}} \\ \mathcal{A} & \xleftarrow{\mu_{\mathcal{A}}} & \mathcal{A}\Box_C\mathcal{A}. \end{array}$$

Equivalently, a C-bicomodule $\mathcal{A}$ is a C-ring if and only if the cotensor functor $-\Box_C\mathcal{A} : \mathbf{M}^C \to \mathbf{M}^C$ (or, equivalently, $\mathcal{A}\Box_C- : {}^C\mathbf{M} \to {}^C\mathbf{M}$) is a monad. Still equivalently, a C-ring is a monoid (algebra) in the monoidal category $({}^C\mathbf{M}^C, \Box_C, C)$.

C-rings in this form were introduced in [37] to retain properties of entwining structures, which are not captured by corings, and in [157] as a natural algebraic structure underlying semi-infinite cohomology [190]. In [157], C-rings are termed *C-algebras*. However, since in parallel to the k-algebra case, the term "C-algebra" might suggest that $\mathcal{A}$ is C-(co)central, the term "C-ring" is our preferred name.

7.9. Modules over C-rings. A right module over a C-ring $\mathcal{A}$ is a triple $(M, \varrho^M, \varrho_M)$, where (M, ϱ^M) is a right C-comodule M with coaction $\varrho^M : M \to M \otimes C$, and $\varrho_M : M\Box_C\mathcal{A} \to M$ is a right C-comodule map such that

$$\begin{array}{ccc} M\Box_C\mathcal{A}\Box_C\mathcal{A} & \xrightarrow{\varrho_M\Box_C\mathcal{A}} & M\Box_C\mathcal{A} \\ {\scriptstyle M\Box_C\mu_{\mathcal{A}}}\downarrow & & \downarrow{\scriptstyle \varrho_M} \\ M\Box_C\mathcal{A} & \xrightarrow{\varrho_M} & M, \end{array} \qquad \begin{array}{ccc} M & \xrightarrow{\varrho^M} & M\Box_C C \\ {\scriptstyle M}\downarrow & & \downarrow{\scriptstyle M\Box_C\eta_{\mathcal{A}}} \\ M & \xleftarrow{\varrho_M} & M\Box_C\mathcal{A}. \end{array}$$

The map ϱ_M is called a *right $\mathcal{A}$-action*. A map of $\mathcal{A}$-modules is a right C-colinear map, which respects the actions, that is, a map of $\mathcal{A}$-modules $(M, \varrho^M, \varrho_M) \to (N, \varrho^N, \varrho_N)$

is a C-comodule map $f : M \to N$ such that

$$\varrho_N \circ (f\Box_C\mathcal{A}) = f \circ \varrho_M.$$

The category of right $\mathcal{A}$-modules is denoted by $\mathbf{M}_\mathcal{A}$.

The category of right $\mathcal{A}$-modules coincides with the category of algebras of the monad $-\Box_C\mathcal{A} : \mathbf{M}^C \to \mathbf{M}^C$, that is, $\mathbf{M}_\mathcal{A} = (\mathbf{M}^C)^{-\Box_C\mathcal{A}}$. Consequently the functor

$$-\Box_C\mathcal{A} : \mathbf{M}^C \to \mathbf{M}_\mathcal{A}, \qquad (M, \varrho^M) \mapsto (M\Box_C\mathcal{A}, M\Box_C\varrho^\mathcal{A}, M\Box_C\mu_\mathcal{A}),$$

is a left adjoint of the forgetful functor $(-)^C : \mathbf{M}_\mathcal{A} \to \mathbf{M}^C$.

Left $\mathcal{A}$-modules are defined symmetrically as triples $(N, {}^N\varrho, {}_N\varrho)$, where $(N, {}^N\varrho)$ is a left C-comodule and ${}_N\varrho : \mathcal{A}\Box_C N \to N$ is an associative, unital left C-comodule map.

7.10. Separable and split C-rings [37, Theorem 6.3]. The defining adjunction $-\Box_C\mathcal{A}, (-)^C$ retains large amount of information about the structure of C-rings. For example, the functor $-\Box_C\mathcal{A} : \mathbf{M}^C \to \mathbf{M}_\mathcal{A}$ is separable if and only if there exists a map $e : \mathcal{A} \to k$ such that $(e \otimes C) \circ \rho^\mathcal{A} = (C \otimes e) \circ {}^\mathcal{A}\varrho$ and $e \circ \eta_\mathcal{A} = \varepsilon_C$. A C-ring with such a map is called a *split C-ring*.

The forgetful functor $(-)^C : \mathbf{M}_\mathcal{A} \to \mathbf{M}^C$ is separable if and only if there exists $\gamma \in {}^C\mathrm{Hom}^C(C, \mathcal{A}\Box_C\mathcal{A})$ such that

$$(\mu_\mathcal{A}\Box_C\mathcal{A}) \circ (\mathcal{A}\Box_C\gamma) \circ \varrho^\mathcal{A} = (\mathcal{A}\Box_C\mu_\mathcal{A}) \circ (\gamma\Box_C\mathcal{A}) \circ {}^\mathcal{A}\varrho$$

and $\mu_\mathcal{A} \circ \gamma = \eta_\mathcal{A}$. In this case, $\mathcal{A}$ is called a *separable C-ring*.

A right $\mathcal{A}$-module $(P, \varrho^P, \varrho_P)$ is said to be $(\mathcal{A}, C)$-*projective* if the action ϱ_P has an $\mathcal{A}$-module section. Every module of a separable C-ring $\mathcal{A}$ is $(\mathcal{A}, C)$-projective.

7.11. Characters. A *character* in $\mathcal{A}$ is a k-linear map $\kappa : \mathcal{A} \to k$ such that

$$\kappa \circ \mu_\mathcal{A} = \kappa\Box_C\kappa, \qquad \kappa \circ \eta_\mathcal{A} = \varepsilon_C.$$

The regular right C-comodule (C, Δ_C) is a right $\mathcal{A}$-module if and only if there is a character in $\mathcal{A}$.

As explained in [43, Section 8], to every C-ring $\mathcal{A}$ with a character, one can associate a comodule with an extended coderivation (cf. [82]) and then interpret right $\mathcal{A}$-modules as comodules with flat connections. This is parallel to the description of comodules over a coring with a group-like element; see 4.23.

7.12. Trivial and dual-Sweedler C-rings. A coalgebra C is itself a C-ring with both the unit and multiplication provided by the identity map C (combined with the isomorphism $C\Box_C C \simeq C$ in case of the multiplication). Right C-modules are the same as right C-comodules, so $\mathbf{M}^C \equiv \mathbf{M}_C$.

Let $\pi : C \to B$ be a morphism of coalgebras, then C is a B-bicomodule via $(C \otimes \pi) \circ \Delta$ and $(\pi \otimes C) \circ \Delta$. Define $\mathcal{A} = C\Box_B C$. Then, $\mathcal{A}$ is a C-ring with product $\mu_\mathcal{A} : \mathcal{A}\Box_C\mathcal{A} \cong C\Box_B C\Box_B C \to \mathcal{A}$, $\mu_\mathcal{A} = C\Box_B\varepsilon_C\Box_B C$, and unit $\nu_A = \Delta_C$. This is a dualization of the Sweedler coring. The objects in the category of right $\mathcal{A}$-modules are triples (M, ϱ^M, f), where (M, ϱ^M) is a right C-comodule and $f : M\Box_B C \to M$

is a right C-comodule map such that

$$f \circ (f \Box_B C) = f \circ (M \Box_B \varepsilon_C \Box_B C), \qquad f \circ \varrho^M = M.$$

By analogy with Sweedler's coring and descent data, it appears quite natural to call such triples (M, ϱ^M, f) *ascent data*.

7.13. C-rings and entwining structures. The main motivation for introducing C-rings in [37] was that their defining adjunction involves the forgetful functor $(-)^C : \mathbf{M}_{\mathcal{A}} \to \mathbf{M}^C$. If, therefore, there is a C-ring associated to an entwining structure $(A, C)_\psi$ and $\mathbf{M}_{\mathcal{A}} \equiv \mathbf{M}_A^C(\psi)$ (the category of entwined modules), then, by studying general properties of C-rings, one derives these properties of entwined modules, which arise from the properties of $(-)^C$. This is, indeed, the case.

PROPOSITION. *For an entwining structure $(A, C)_\psi$, view $\mathcal{A} = C \otimes A$ as a C-bicomodule with the left coaction ${}^{\mathcal{A}}\varrho = \Delta_C \otimes C$ and the right coaction $\varrho^{\mathcal{A}} = (C \otimes \psi) \circ (\Delta_C \otimes A)$. Then $\mathcal{A}$ is a C-ring with the product $\mu_{\mathcal{A}} : \mathcal{A} \Box_C \mathcal{A} \cong C \otimes A \otimes A \to \mathcal{A}$, $\mu_{\mathcal{A}} = C \otimes \mu_A$, and the unit $\eta_{\mathcal{A}} = C \otimes 1_A$.*

Conversely, if $C \otimes A$ is a C-ring with the product and the unit above and the natural left C-comodule structure $\Delta_C \otimes A$, then $(A, C)_\psi$ is an entwining structure, where $\psi = (\varepsilon_C \otimes A \otimes C) \circ \varrho^{C \otimes A}$.

Under this bijective correspondence, $\mathbf{M}_{\mathcal{A}} \equiv \mathbf{M}_A^C(\psi)$.

In a similar way, one associates a C-ring to a weak entwining structure (A, C, ψ). In this case, $\mathcal{A} = \operatorname{Im} \overline{p_R}$, where

$$\overline{p_R} : C \otimes A \to C \otimes A, \qquad \overline{p_R} = (C \otimes A \otimes \varepsilon_C) \circ (C \otimes \psi) \circ (\Delta_C \otimes A),$$

and the unit of $\mathcal{A}$ is $c \mapsto \overline{p_R}(1_A \otimes c)$; see [51, Section 4].

7.14. Matrix ring contexts [51]. A *matrix ring context*, $(C, D, N, M, \sigma, \tau)$, consists of a pair of coalgebras C and D, a (C, D)-bicomodule $(N, \varrho^N, {}^N\varrho)$, a (D, C)-bicomodule $(M, \varrho^M, {}^M\varrho)$, and a pair of bicomodule maps

$$\sigma : C \to N \Box_D M, \qquad \tau : M \Box_C N \to D,$$

such that the diagrams

$$\begin{array}{ccc} N \Box_D M \Box_C N & \xleftarrow{\sigma \Box_C N} & C \Box_C N \\ {\scriptstyle N \Box_D \tau}\downarrow & & \uparrow{\scriptstyle {}^N\varrho} \\ N \Box_D D & \xleftarrow{\varrho^N} & N, \end{array} \qquad \begin{array}{ccc} M \Box_C N \Box_D M & \xrightarrow{\tau \Box_D M} & D \Box_D M \\ {\scriptstyle M \Box_C \sigma}\uparrow & & \uparrow{\scriptstyle {}^M\varrho} \\ M \Box_C C & \xleftarrow{\varrho^M} & M \end{array}$$

commute. The map σ is called a *unit* and τ is called a *counit* of a matrix context.

Since a counit τ of a matrix ring context is a D-bicomodule map, it is fully determined by its *reduced form* $\widehat{\tau} = \varepsilon_D \circ \tau$. The map $\widehat{\tau}$ is called a *reduced counit* of a matrix context. The D-bicolinearity of τ is equivalent to the following property of $\widehat{\tau}$,

$$(D \otimes \widehat{\tau}) \circ ({}^M\varrho \otimes N) = (\widehat{\tau} \otimes D) \circ (M \otimes \varrho^N).$$

In terms of the reduced counit, the above commutative diagrams read

$$(N \otimes \widehat{\tau}\,) \circ (\sigma \Box_C N\,) \circ {}^N\varrho = N, \qquad (\,\widehat{\tau} \otimes M) \circ (M \Box_C \sigma) \circ \varrho^M = M.$$

The notion of a matrix context is closely related to that of *pre-equivalence data* or a *Morita-Takeuchi context* introduced in [180, Definition 2.3]. In particular, in view of [180, Theorem 2.5], if one of the maps in a Morita-Takeuchi context is injective, then there is a corresponding matrix ring context. Furthermore, all equivalence data give rise to a matrix ring context.

Matrix ring contexts are in one-to-one correspondence with adjoint pairs of functors $-\Box_C N : \mathbf{M}^C \to \mathbf{M}^D$, $-\Box_D M : \mathbf{M}^D \to \mathbf{M}^C$ (or adjunctions in a bicategory: coalgebras, bicomodules, bicolinear maps; for bicategories see [19]). In particular, if $(C, D, N, M, \sigma, \tau)$ is a matrix ring context, then M is a *quasi-finite injector*, that is, the cotensor functor $-\Box_C N : \mathbf{M}^C \to \mathbf{M}_k$ is a left adoint of the tensor functor $G = - \otimes M : \mathbf{M}_k \to \mathbf{M}^C$ ([180, Proposition 1.3], [53, Section 12.8]). Conversely, (M, ϱ^M) can be a part of a matrix ring context only if it is a quasi-finite injector.

7.15. Matrix C-rings. Let $(C, D, N, M, \sigma, \tau)$ be a matrix ring context. Then $\mathcal{A} := N\Box_D M$ is a C-ring with the product $\mu_{\mathcal{A}} = N\Box_D\widehat{\tau}\,\Box_D M$ and unit $\eta_{\mathcal{A}} = \sigma$, where $\widehat{\tau}$ is the reduced counit. Furthermore, M is a right $\mathcal{A}$-module with the action $\widehat{\tau}\,\Box_D M$ and N is a left $\mathcal{A}$-module with the action $N\Box_D\widehat{\tau}$. The C-ring $\mathcal{A}$ is called a *matrix C-ring*. The monad $-\Box_C\mathcal{A} : \mathbf{M}^C \to \mathbf{M}^C$ is simply the monad corresponding to the adjoint pair of functors $-\Box_C N : \mathbf{M}^C \to \mathbf{M}^D$, $-\Box_D M : \mathbf{M}^D \to \mathbf{M}^C$; see 7.14.

7.16. Tensor product of modules and the coendomorphism C-ring. Let $\mathcal{A}$ be a C-ring, $(M, \varrho^M, \varrho_M)$ a right $\mathcal{A}$-module, and $(N, {}^N\varrho, {}_N\varrho)$ a left $\mathcal{A}$-module. The *tensor product* of $(M, \varrho^M, \varrho_M)$ with $(N, {}^N\varrho, {}_N\varrho)$ is defined as the coequalizer of k-linear maps

$$M\Box_C\mathcal{A}\Box_C N \underset{M\Box_C\,{}_N\varrho}{\overset{\varrho_M\Box_C N}{\rightrightarrows}} M\Box_C N \xrightarrow{\ \pi_{M,\mathcal{A},N}\ } M \otimes_{\mathcal{A}} N,$$

that is, $M \otimes_{\mathcal{A}} N$ is the cokernel of $\varrho_M\Box_C N - M\Box_C\,{}_N\varrho$. As argued in [157], *semi-infinite cohomology* of [11, 190] can be understood in terms of derived functors of such a tensor product for a specific C-ring.

Let (M, ϱ^M) be a quasi-finite injector, and let $(N, {}^N\varrho)$ be a left C-comodule such that $-\Box_C N$ is the left adjoint of the tensor functor $- \otimes M : \mathbf{M}_k \to \mathbf{M}^C$. Then $E = M\Box_C N$ is a coalgebra, known as a *C-coendomorphism coalgebra* of M. (E arises as a comonad $- \otimes M\Box_C N : \mathbf{M}_k \to \mathbf{M}_k$ corresponding to the adjoint pair of functors $-\Box_C N$, $- \otimes M$.) Furthermore, there exists a unique coalgebra structure on $M \otimes_{\mathcal{A}} N$, whereby $\pi_{M,\mathcal{A},N}$ is a coalgebra map. The coalgebra $M \otimes_{\mathcal{A}} N$ is called an *$\mathcal{A}$-coendomorphism coalgebra of M* and is denoted by $E_{\mathcal{A}}(M)$.

To any right $\mathcal{A}$-module $(M, \varrho^M, \varrho_M)$, which is a quasi-finite injector as a right C-comodule, one can associate the matrix ring context $(C, E_{\mathcal{A}}(M), N, M, \sigma, \tau)$ with

$$\sigma : C \xrightarrow{\eta_C} N\Box_E M \to N\Box_{E_{\mathcal{A}}(M)} M, \qquad \tau : M\Box_C N \xrightarrow{\psi_E} E \to E_{\mathcal{A}}(M),$$

where η is the unit and ψ is the counit of adjunction $-\Box_C N : \mathbf{M}^C \to \mathbf{M}^E$, $-\Box_E M : \mathbf{M}^E \to \mathbf{M}^C$. We refer to this context as an *$\mathcal{A}$-coendomorphism context* associated to M. The corresponding matrix C-ring is called an *$\mathcal{A}$-coendomorphism ring of M*.

7.17. Galois theory. In parallel to (finite) Galois comodules of a coring, one can study Galois comodules of a C-ring $\mathcal{A}$. Let $(M, \varrho^M, \varrho_M)$ be a right $\mathcal{A}$-module such that (M, ϱ^M) is a quasi-finite injector. Set $(N, {}^N\varrho)$ to be a left C-comodule such that $-\Box_C N$ is the left adjoint of the tensor functor $- \otimes M : \mathbf{M}_k \to \mathbf{M}^C$. Let $S = E_{\mathcal{A}}(M)$ and let σ be the unit of the $\mathcal{A}$-coendomorphism context associated to M. Then M is called a *Galois $\mathcal{A}$-module* if the map

$$\beta : \mathcal{A} \to N\Box_S M, \qquad \beta := (N \otimes \varrho_M) \circ (\sigma \Box_C \mathcal{A}) \circ {}^{\mathcal{A}}\varrho,$$

is an isomorphism.

As shown in [51, Theorem 3.11], M is a Galois $\mathcal{A}$-module if and only if β is a split monomorphism of left $\mathcal{A}$-modules.

7.18. Bicoalgebroids. Bialgebroids are defined as corings with a compatible algebra structure. In a dual way, one can define *bicoalgebroids* as C-rings with compatible coproduct and counit.

Let $(C, \Delta_C, \varepsilon_C)$ and $(\mathcal{H}, \Delta_{\mathcal{H}}, \varepsilon_{\mathcal{H}})$ be coalgebras, and assume that there is a coalgebra map $\alpha : \mathcal{H} \to C$ and an anticoalgebra map $\beta : \mathcal{H} \to C$ such that for all $h \in \mathcal{H}$, $\sum \alpha(h_{(1)}) \otimes \beta(h_{(2)}) = \sum \alpha(h_{(2)}) \otimes \beta(h_{(1)})$. This allows one to view $\mathcal{H}$ as a C-bicomodule via left coaction ${}^{\mathcal{H}}\varrho(h) = \sum \alpha(h_{(1)}) \otimes h_{(2)}$ and the right coaction $\varrho^{\mathcal{H}}(h) = \sum h_{(2)} \otimes \beta(h_{(1)})$. Following [50], $\mathcal{H}$ is called a *bicoalgebroid* if

(a) $(\mathcal{H}, \mu_{\mathcal{H}}, \eta_{\mathcal{H}})$ is a C-ring;
(b) for all $\sum_i g^i \otimes h^i \in \mathcal{H}\Box_C\mathcal{H}$,

$$\sum_i \mu_{\mathcal{H}}(g^i \otimes h^i{}_{(1)}) \otimes \alpha(h^i{}_{(2)}) = \sum_i \mu_{\mathcal{H}}(g^i{}_{(1)} \otimes h^i) \otimes \beta(g^i{}_{(2)}),$$

$$\Delta_{\mathcal{H}}(\mu_{\mathcal{H}}(\sum_i g^i \otimes h^i)) = \sum_i \mu_{\mathcal{H}}(g^i{}_{(1)} \otimes h^i{}_{(1)}) \otimes \mu_{\mathcal{H}}(g^i{}_{(2)} \otimes h^i{}_{(2)}),$$

and $\varepsilon_{\mathcal{H}}\Box_C\varepsilon_{\mathcal{H}} = \varepsilon_{\mathcal{H}} \circ \mu_{\mathcal{H}}$;
(c) for all $c \in C$, $\varepsilon_{\mathcal{H}}(\eta_{\mathcal{H}}(c)) = \varepsilon_C(c)$, and

$$\Delta_{\mathcal{H}}(\eta_{\mathcal{H}}(c)) = \sum \eta_{\mathcal{H}}(c)_{(1)} \otimes \eta_{\mathcal{H}}(\alpha(\eta_{\mathcal{H}}(c)_{(2)})) = \sum \eta_{\mathcal{H}}(c)_{(1)} \otimes \eta_{\mathcal{H}}(\beta(\eta_{\mathcal{H}}(c)_{(2)})).$$

Similarly to bialgebroids, the category of left comodules of a C-bicoalgebroid $\mathcal{H}$ is monoidal and the forgetful functor to the category of C-bicomodules is strict monoidal [13].

7.3. *Other topics*

As time, space, and the author's expertise are all limited, there are several topics in coring and comodule theory, which are not covered in this chapter. In this final section, we would like to list some of these topics and supply the reader with references.

7.19. Contramodules. While we concentrated on comodules, there is another, much less-studied but seemingly equally natural, category of representations of corings. Let $\mathcal{C}$ be an A-bimodule. Since the tensor functor $-\otimes_A \mathcal{C} : \mathbf{M}_A \to \mathbf{M}_A$ has a right adjoint, the hom-functor $\mathrm{Hom}_A(\mathcal{C}, -)$, the characterization of corings as comonads $-\otimes_A \mathcal{C}$ in 3.5 can be extended to the characterization of corings as monads $\mathrm{Hom}_A(\mathcal{C}, -)$. This has been observed already in [90]. As explained in 4.7, right comodules over $\mathcal{C}$ are the same as coalgebras of the comonad $-\otimes_A \mathcal{C}$. Algebras of the monad $\mathrm{Hom}_A(\mathcal{C}, -)$ are known as *contramodules* over $\mathcal{C}$ [89]. Various properties of contramodules are discussed in [29, 157].

7.20. Module theoretic and topological aspects. Several module theoretic and topological aspects of corings are described in [53]. Zariski topology for corings and comodules and top bicomodules are studied in [6]. Rational modules, pairings, and related topologies are described in [2]. Further rationality properties and comodules over semiperfect corings are studied in [62, 69]. (Co)prime and (co)semiprime corings and comodules are discussed in [5, 98, 193, 194]. Symmetric corings are considered in [16].

7.21. Applications to representation theory. Some of the earliest applications of corings were in representation theory, in particular to *matrix problems* (cf. [161]). In representation theory, corings are known as BOCSs, and among others, they were used in settling the tame-wild conjecture [86]. Possibly the earliest reference to BOCSs is [162], and their role in the tame-wild conjecture is explained in [78, 87].

7.22. Index theory for C^*-algebras. Corings appeared in [191] as an algebraic structure describing (finite) index theory for operator algebras [114]. Motivated by this, coring structures on Hilbert C^*-modules are studied in [147, 148]. The theory of depth-2 extensions initiated as a special case of the Jones index theory in [118] makes extensive use of corings (bialgebroids); see also 3.19.

7.23. Categorical aspects. Throughout the chapter we have not discussed in great detail the categorical aspects of corings. An algebraist oriented overview of categorical aspects of corings can be found in [200]. One of the main modern examples of corings is a coring associated to an entwining structure. An entwining structure itself is an example of a distributive law [18]. For recent developments in the categorical approach to distributive laws, we refer to [171, Section 6], [170, 200]. Entwining structures over noncommutative rings and associated corings are studied in [24]. For a description of functors (and equivalences) between categories of comodules, see, for example, [4, 95, 101, 187]. Since, from the categorical point of view, corings are comonads (in a specific category), they fit well into the *formal theory of* (*co*)*monads* [128, 175]. The bicategories arising from this formal theory are discussed in [44]. Monoidal aspects of the category of corings are discussed in [91].

7.24. The Picard and Brauer groups. The Picard and Brauer groups for corings and Azumaya corings are studied in [60, 202]. Corings are used in studies of Picard and Brauer groups for rings in [61, 94, 134, 183].

7.25. Cohomology. Cohomology of corings and comodules is discussed in [97, 106]. Cyclic cohomology of corings is studied in [31, 160]. Čech cohomology for corings is proposed in [169]. This construction is motivated by the interpretation of complete coverings of algebras [70] in terms of Galois corings [54]. Non-Abelian (descent) cohomology that extends non-Abelian cohomology of Hopf algebras [146] is proposed in [42] and is given a categorical treatment in [139]. Some cohomological aspects of corings are also described in [53, Section 30].

7.26. Generalizations of corings. Corings with local structure maps are introduced in [188] and group corings in [64].

7.27. Universal algebra and computer science. Coalgebras in universal algebra, and therefore also corings, are used in theoretical computer science. For example, entwining structures appear in [185] in the context of operational semantics. The interested reader can find some comments and references in [200].

Acknowledgments

I would like to thank Gabriella Böhm, Stefaan Caenepeel, José Gómez-Torrecillas, Joost Vercruysse, and Robert Wisbauer for many valuable comments and suggestions.

References

[1] Abrams, L., Two-dimensional topological quantum field theories and Frobenius algebras, J. Knot Theory Ramifi. **5** (1996), 569–587.
[2] Abuhlail, J., Rational modules for corings, Comm. Algebra **31** (2003), 5793–5840.
[3] Abuhlail, J., Morita contexts for corings and equivalences, in: S. Caenepeel, F. Van Oystaeyen eds., *Hopf Algebras in Noncommutative Geometry and Physics*. Lecture Notes in Pure and Appl. Math., 239, Marcel Dekker, New York, 2005, pp. 1–19.
[4] Abuhlail, J., A note on coinduction functors between categories of comodules for corings, Annali dell'Universitá di Ferrara, Sezione VII.-Sci. Mat. **51** (2005), 151–172.
[5] Abuhlail, J., Fully coprime comodules and fully coprime corings, Appl. Categ. Str. **14** (2006), 379–409.
[6] Abuhlail, J., A Zariski topology for bicomodules and corings, Appl. Categ. Str. **16** (2008), 13–28.
[7] Abuhlail, J., Gómez-Torrecillas, J., Lobillo, F.J., Duality and rational modules in Hopf algebras over commutative rings, J. Algebra **240** (2001), 165–184.
[8] Aguiar, M., *Internal Categories and Quantum Groups*, PhD Thesis, Cornell University, 1997.
[9] Anderson, F., Fuller, K., *Rings and Categories of Modules*, Springer, Berlin, 1974.
[10] Ardizzoni, A., Böhm, G., Menini, C., A Schneider type theorem for Hopf algebroids, J. Algebra **318** (2007), 225–269.
[11] Arkhipov, S.M., Semi-infinite cohomology of associative algebras and bar duality, Internat. Math. Res. Not. **17** (1997), 833–863.
[12] Baez, J. *This week's finds in Mathematical Physics (week 174)*, http://math.ucr.edu/home/baez/week174.html (2001), last accessed on 25 February, 2009.
[13] Bálint, I., Scalar extension of bicoalgebroids, Appl. Categ. Str. **16** (2008), 29–55.
[14] Bass, H., *Algebraic K-Theory*, W.A. Benjamin, Inc., New York, 1968.
[15] Baum, P.F., Hajac, P.M., Matthes, R., Szymański, W., Noncommutative geometry approach to principal and associated bundles, arXiv:math/0701033 (2007).

[16] Beattie, M., Bulacu, D., Raianu, Ş., Balanced bilinear forms for corings, in: T. Brzeziński, J.L. Gómez Pardo, I. Shestakov, P.F. Smith, eds., *Modules and Comodules*, Birkhäuser, Basel, 2008, pp. 87–100.
[17] Beck, J. *Triples, Algebras and Cohomology,* PhD Thesis, Columbia University (1967); Reprinted in Theory and Applications of Categories; **2** (2003), 1–59.
[18] Beck, J., *Distributive laws*, in: B. Eckmann, ed., *Seminar on Triples and Categorical Homology Theory*, Springer Lecture Notes in Mathematics, vol. **80**, Springer-Verlag, Berlin, 1969, pp. 119–140.
[19] Bénabou, J., Introduction to bicategories, Reports of the Midwest Category Seminar, Lecture Notes in Mathematics. vol. **106**, Springer-Verlag, Berlin, 1967, 1–77.
[20] Bergman G.M., Hausknecht, A.O., Cogroups and Co-rings in Categories of Associative Rings, Amer. Math. Soc. 1996.
[21] Berthelot, P., Ogus, A., *Notes on Crystalline Cohomology*, Princeton University Press, Princeton, 1978.
[22] Böhm, G., Doi-Hopf modules over weak Hopf algebras, Comm. Algebra **28** (2000), 4687–4698.
[23] Böhm, G., Galois theory for Hopf algebroids, Annali dell'Universitá di Ferrara, Sezione VII.-Sci. Mat. **51** (2005), 233–262.
[24] Böhm, G., Internal bialgebroids, entwining structures and corings, in: *Algebraic Structures and Their Representations*, in: J.A. de la Peña, E. Vallejo and N. Atakishiyev, eds., Contemp. Math., vol. **376**, Amer. Math. Soc., Providence, RI, 2005, pp. 207–226.
[25] Böhm, G., Hopf algebroids, in: M. Hazewinkel, ed., *Handbook of Algebra*, vol. **6**, Elsevier, The Netherlands, 2009, pp. 173–236.
[26] Böhm, G., Brzeziński, T., Strong connections and the relative Chern-Galois character for corings, Int. Math. Res. Not. **42** (2005), 2579–2625.
[27] Böhm, G., Brzeziński, T., Cleft extensions of Hopf algebroids, Appl. Categ. Str. **14** (2006), 431–469.
[28] Böhm, G., Brzeziński, T., Pre-torsors and equivalences, J. Algebra **317**, 544–580 (2007); Corrigendum, J. Algebra **319** (2007), 1339–40.
[29] Böhm, G., Brzeziński, T., Wisbauer, R., Monads and comonads in module categories, arxiv: 0804.1460, 2008.
[30] Böhm, G., Nill, F., Szlachányi, K., Weak Hopf algebras I. Integral theory and C^*-structure, J. Algebra **221** (1999), 385–438.
[31] Böhm, G., Ştefan, D., (Co)cyclic (co)homology of bialgebroids: an approach via (co)monads, arXiv:0705.3190. Commun. Math. Phys., **282** (2008), 239–286.
[32] Böhm, G., Szlachányi, K., A coassociative C^*-quantum group with nonintegral dimension, Lett. Math. Phys. **35** (1996), 437–456.
[33] Böhm, G., Vercruysse, J., Morita theory for coring extensions and cleft bicomodules, Adv. Math. **209** (2007), 611–648.
[34] Borceux, F., *Handbook of Categorical Algebra 2. Categories and Structures*, Cambridge University Press, Cambridge, 1994.
[35] Borceux, F., Janelidze, G., *Galois Theories*, Cambridge University Press, Cambridge, 2001.
[36] Brzeziński, T., On modules associated to coalgebra-Galois extensions, J. Algebra **215** (1999), 290–317.
[37] Brzeziński, T., The structure of corings. Induction functors, Maschke-type theorem, and Frobenius and Galois-type properties, Algebras Rep. Theory **5** (2002), 389–410.
[38] Brzeziński, T., The structure of corings with a group-like element, in: P.M. Hajac and W. Pusz, eds., *Noncommutative Geometry and Quantum Groups*, vol. **61**, Banach Center Publications, Warsaw, 2003, pp. 21–35.
[39] Brzeziński, T., Towers of corings, Comm. Algebra **31** (2003), 2015–2026.
[40] Brzeziński, T., Galois comodules, J. Algebra **290** (2005), 503–537.
[41] Brzeziński, T., A note on coring extensions, Annali dell'Universitá di Ferrara, Sezione VII.-Sci. Mat. **51** (2005), 15–27.
[42] Brzeziński, T., Descent cohomology and corings, arXiv:math.RA/0601491v3. Comm. Algebra **36** (2008), 1894–1900.

[43] Brzeziński, T., Flat connections and (co)modules, in: S. Caenepeel, F. Van Oystaeyen, eds., *New Techniques in Hopf Algebras and Graded Ring Theory*, Universa Press, Wetteren, 2007, pp. 35–52.
[44] Brzeziński, T., El Kaoutit, L., Gómez-Torrecillas, J., The bicategories of corings, J. Pure Appl. Algebra **205** (2006), 510–541.
[45] Brzeziński, T., Gómez-Torrecillas, J., On comatrix corings and bimodules, K-Theory **29** (2003), 101–115.
[46] Brzeziński, T., Hajac, P.M., Coalgebra extensions and algebra coextensions of Galois type, Comm. Algebra **27** (1999), 1347–1367.
[47] Brzeziński, T., Hajac, P.M., The Chern-Galois character, C. R. Acad. Sci. Paris, Ser. I **338** (2004), 113–116.
[48] Brzeziński, T., Kadison, L., Wisbauer, R., On coseparable and biseparable corings, in: S. Caenepeel, F. Van Oystaeyen, eds., *Hopf Algebras in Noncommutative Geometry and Physics*, Marcel Dekker, New York, 2005, pp. 71–88.
[49] Brzeziński, T., Majid, S., Coalgebra bundles, Comm. Math. Phys. **191** (1998), 467–492.
[50] Brzeziński, T., Militaru, G., Bialgebroids, $\times_A$-bialgebras and duality, J. Algebra **251** (2002), 279–294.
[51] Brzeziński, T., Turner, R.B., The Galois theory of C-rings, Appl. Categ. Str. **14** (2006), 471–501.
[52] Brzeziński, T., Turner, R.B., Wrightson, A.P., The structure of weak coalgebra-Galois extensions, Comm. Algebra **34** (2006), 1489–1509.
[53] Brzeziński, T., Wisbauer, R., *Corings and Comodules*, Cambridge University Press, Cambridge, 2003. Erratum: http://www-maths.swan.ac.uk/staff/tb/Corings.htm, last accessed on 25 February, 2009.
[54] Brzeziński, T., Wrightson, A.P., Complete coverings and Galois corings, Int. J. Geom. Methods Mod. Phys. **2** (2005), 751–758.
[55] Caenepeel, S., Galois corings from the descent theory point of view, in: *Galois Theory, Hopf Algebras, and Semiabelian Categories*, in: G. Janelidze, B. Pareigis and W. Tholen, eds., Fields Inst. Commun., vol. **43**, Amer. Math. Soc., Providence, RI, 2004, pp. 163–186.
[56] Caenepeel, S., De Groot, E., Modules over weak entwining structures, AMS Contemp. Math. **267** (2000), 31–54.
[57] Caenepeel, S., De Groot, E., Galois corings applied to partial Galois theory, in: S.L. Kalla, M.M. Chawla eds., *Proceedings of the International Conference on Mathematics and Applications, ICMA 2004*, Kuwait University, Kuwait, 2005, pp. 117–134.
[58] Caenepeel, S., De Groot, E., Vercruysse, J., Constructing infinite comatrix corings from comodules, Appl. Categ. Str. **14** (2006), 539–565.
[59] Caenepeel, S., De Groot, E., Vercruysse, J., Galois theory for comatrix corings: Descent theory, Morita theory, Frobenius and separability properties, Trans. Amer. Math. Soc. **359** (2007), 185–226.
[60] Caenepeel, S., Femić, B., The Brauer group of Azumaya corings and the second cohomology group, K-Theory **34** (2005), 361–393.
[61] Caenepeel, S., Guédénon, T., The relative Picard group of a comodule algebra and Harrison cohomology, Proc. Edinb. Math. Soc. **48** (2005), 557–569.
[62] Caenepeel, S., Iovanov, M., Comodules over semiperfect corings, in: S.L. Kalla, M.M. Chawla, eds., *Proceedings of the International Conference on Mathematics and Applications, ICMA 2004*, Kuwait University, Kuwait, 2005, pp. 135–160.
[63] Caenepeel, S., Janssen, K., Lax entwining structures, arXiv:math/0610524 (2006), to appear in Comm. Alg.
[64] Caenepeel, S., Janssen, K., Wang, S.H., Group corings, Appl. Categ. Str. **16** (2008), 65–96.
[65] Caenepeel, S., Kadison, L., Are biseparable extensions Frobenius? K-Theory **24** (2001), 361–383.
[66] Caenepeel, S., Militaru, G., Zhu, S., Crossed modules and Doi-Hopf modules, Israel J. Math. **100** (1997), 221–247.
[67] Caenepeel, S., Militaru, G., Zhu, S., Doi-Hopf modules, Yetter-Drinfel'd modules and Frobenius type properties, Trans. Amer. Math. Soc. **349** (1997), 4311–4342.

[68] Caenepeel, S., Militaru, G., Zhu, S., *Frobenius and Separable Functors for Generalized Hopf Modules and Nonlinear Equations*, Lecture Notes in Mathematics 1787, Springer, Berlin, 2002.
[69] Caenepeel, S., Vercruysse, J., Wang, S.H., Rationality properties for Morita contexts associated to corings, in: S. Caenepeel, F. Van Oystaeyen, eds., *Hopf Algebras in Noncommutative Geometry and Physics*, Marcel Dekker, New York, 2005, pp. 113–136.
[70] Calow, D., Matthes R., Covering and gluing of algebras and differential algebras, J. Geom. Phys. **32** (2000), 364–396.
[71] Castaño Iglesias, F., On quasi-Frobenius bimodules and corings, arXiv:math/0612663, (2006).
[72] Castaño Iglesias, F., Năstăsescu, C., Quasi-Frobenius functors with applications to corings, arXiv:math/0612662, (2006).
[73] Chase, S.U., Harrison, D.K., Rosenberg, A., Galois theory and cohomology of commutative rings, Mem. Amer. Math. Soc. **52** (1965).
[74] Cipolla, M., Discesa fedelemente piatta dei moduli, Rend. Circ. Mat. Palermo **25** (1976), 43–46.
[75] Cohen, M., Gelaki, S., Westreich, S., Hopf algebras, Handb Algebra **4** (2006), 173–239.
[76] Connes, A., Non-commutative differential geometry, Inst. Hautes Études Sci. Publ. Math. **62** (1985), 257–360.
[77] Connes, A., Moscovici, H., Cyclic cohomology and Hopf algebra symmetry, Lett. Math. Phys. **52** (2000), 1–28.
[78] Crawley-Boevey, W.W., On tame algebras and bocses, Proc. London Math. Soc. **56** (1988), 451–483.
[79] Cuadra, J., Gómez-Torrecillas, J., Galois corings and a Jacobson-Bourbaki type correspondence, J. Algebra **308** (2007), 178–198.
[80] Cuntz, J., and Quillen, D., Algebra extensions and nonsingularity, J. Amer. Math. Soc. **8** (1995), 251–289.
[81] De Groot, E., *Comatrix Corings Applied to Weak and Partial Galois Theory,* PhD thesis, Vrije Universiteit Brussel, Brussels, 2005.
[82] Doi, Y., Homological coalgebra, J. Math. Soc. Japan **33** (1981), 31–50.
[83] Doi, Y., On the structure of relative Hopf modules, Comm. Algebra **11** (1983), 243–253.
[84] Doi, Y., Unifying Hopf modules, J. Algebra **153** (1992), 373–385.
[85] Dokuchaev, M., Ferrero, M., Pacques, A., Partial actions and Galois theory, J. Pure Appl. Algebra **208** (2007), 77–87.
[86] Drozd, Y.A., Tame and wild matrix problems, in: Yu. A. Mitropolskii, ed., *Representations and Quadratic Forms,* Institute of Mathematics, Kiev, 1979, pp. 39–74. (English translation: Amer. Math. Soc. Transl. **128** (1986), 31–55.)
[87] Drozd, Y.A., Reduction algorithm and representations of boxes and algebras, Comptes Rendue Math. Acad. Sci. Canada **23** (2001), 97–125.
[88] Dubuc, E., *Kan extensions in enriched category theory,* Lecture Notes in Mathematics vol. 145, Springer, Berlin, 1970.
[89] Eilenberg, S. and Moore, J.C., Foundations of relative homological algebra, Mem. Amer. Math. Soc. **55** (1965).
[90] Eilenberg, S. and Moore, J.C., Adjoint functors and triples, Ill. J. Math. **9** (1965), 381–398.
[91] El Kaoutit, L., Monoidal categories of corings, Annali dell'Universitá di Ferrara, Sezione VII.-Sci. Mat. **51** (2005), 197–208.
[92] El Kaoutit, L., Gómez-Torrecillas, J., Comatrix corings: Galois corings, descent theory, and a structure theorem for cosemisimple corings, Math. Z. **244** (2003), 887–906.
[93] El Kaoutit, L., Gómez-Torrecillas, J., Infinite comatrix corings, Int. Math. Res. Notices **39** (2004), 2017–2037.
[94] El Kaoutit, L., Gómez-Torrecillas, J., Comatrix corings and invertible bimodules, Annali dell'Universitá di Ferrara, Sezione VII.-Sci. Mat. **51** (2005), 263–280.
[95] El Kaoutit, L., Gómez-Torrecillas, J., Morita duality for corings over quasi-Frobenius rings, in: S. Caenepeel, F. Van Oystaeyen, eds., *Hopf Algebras in Noncommutative Geometry and Physics*, Marcel Dekker, New York, 2005, pp. 137–153.

[96] El Kaoutit, L., Gómez-Torrecillas, J., Lobillo, F.J., Semisimple corings, Algebra Colloq. **11** (2004), 427–442.
[97] El Kaoutit, L., Vercruysse, J., Cohomology for bicomodules. Separable and Maschke functors, J. *K*-Theory to appear, arXiv:math/0608195 (2006).
[98] Ferrero, M.A., Rodrigues, V., On prime and semiprime modules and comodules, J. Algebra Appl. **5** (2006), 681–694.
[99] Garfinkel, G.S., Universally torsionless and trace modules, Trans. Amer. Math. Soc. **215** (1976), 119–144.
[100] Gómez-Torrecillas, J., Coalgebras and comodules over a commutative ring, Rev. Roumaine Math. Pures Appl. **43** (1998), 591–603.
[101] Gómez-Torrecillas, J., Separable functors in corings, Int. J. Math. Math. Sci. **30** (2002), 203–225.
[102] Gómez-Torrecillas, J., Comonads and Galois corings, Appl. Categ. Str. **14** (2006), 579–598.
[103] Gómez-Torrecillas, J., Vercruysse, J., Comatrix corings and Galois comodules over firm rings, Algebras Rep. Theory **10** (2007), 271–306.
[104] Grothendieck, A., Technique de descente et théorèmes d'existence en géométrie algébrique, I. Généralités, descente par morphismes fidèlement plats. Séminaire Bourbaki. **12** (1959/1960), 190.
[105] Guo, J., Quasi-Frobenius corings and quasi-Frobenius extensions, Comm. Algebra **34** (2006), 2269–2280.
[106] Guzman, F., Cointegrations, relative cohomology for comodules, and coseparable corings, J. Algebra **126** (1989), 211–224.
[107] Hajac, P.M., Khalkhali, M., Rangipour, B., Sommerhäuser, Y., Stable anti-Yetter-Drinfeld modules, C. R. Acad. Sci. Paris, Ser. I **338** (2004), 587–590.
[108] Hajac, P.M., Khalkhali, M., Rangipour, B., Sommerhäuser, Y., Hopf-cyclic homology and cohomology with coefficients, C. R. Acad. Sci. Paris, Ser. I **338** (2004), 667–672.
[109] Hirata, K., Sugano, K., On semisimple extensions and separable extensions over noncommutative rings, J. Math. Soc. Japan **18** (1966), 360–374.
[110] Hochschild, G., Double vector spaces over division rings, Amer. J. Math. **71** (1949), 443–460.
[111] Iovanov, M., Frobenius extensions of corings, arXiv:math/0612447 (2006).
[112] Iovanov, M., Vercruysse, J., CoFrobenius corings and related functors, arXiv:math/0610853 (2006), to appear in J. Pure Appl. Algebra.
[113] Jara, P., Ştefan, D., Cyclic homology of Hopf-Galois extensions and Hopf algebras, Proc. London Math. Soc. **93** (2006), 138–174.
[114] Jones, V.F.R., Index for subfactors, Invent. Math. **72** (1983), 1–25.
[115] Jonah, D.W., Cohomology of Coalgebras, Mem. Amer. Math. Soc. **82** (1968).
[116] Kadison, L., *New Examples of Frobenius Extensions, AMS*, Providence, RI, 1999.
[117] Kadison, L., Separability and the twisted Frobenius bimodule, Algebras Rep. Theory **2** (1999), 397–414.
[118] Kadison, L., Szlachányi, K., Bialgebroid actions on depth two extensions and duality, Adv. Math. **179** (2003), 75–121.
[119] Kasch, F., Grundlagen einer Theorie der Frobeniuserweiterungen, Math. Ann. **127** (1954), 453–474.
[120] Kasch, F., Projektive Frobenius-Erweiterungen, Sitzungsber. Heidelberger Akad. Wiss. Math.-Natur. Kl. **4** (1960/61), 89–109.
[121] Kaygun, A., Khalkhali, M., Hopf modules and noncommutative differential geometry, Lett. Math. Phys. **76** (2006), 77–91.
[122] Kleiner, M., The dual ring to a coring with a grouplike, Proc. Amer. Math. Soc. **91** (1984), 540–542.
[123] Kontsevich M., Rosenberg, A.L., Noncommutative smooth spaces, in: I.M. Gelfand and V.S. Retakh, eds., *The Gelfand Mathematical Seminars, 1996–1999*, Gelfand Math. Sem., Birkhäuser, Boston, MA, 2000, pp. 85–108.
[124] Kontsevich M., Rosenberg, A.L., Noncommutative spaces; Noncommutative spaces and flat descent; Noncommutative stacks, Preprints MPIM2004-35–37, 2004.
[125] Knus, M.A., Ojanguren, M., *Théorie de la descente et algèbres d'Azumaya*, Lecture Notes in Mathematics, vol. **389**, Springer, Berlin, 1974.

[126] Koppinen, M., Variations on the smash product with applications to group-graded rings, J. Pure Appl. Algebra **104** (1994), 61–80.
[127] Kreimer, H.F., Takeuchi, M., Hopf algebras and Galois extensions of an algebra, Indiana Univ. Math. J. **30** (1981), 675–691.
[128] Lack, S., Street, R., The formal theory of monads II, J. Pure Appl. Algebra **175** (2002), 243–265.
[129] Larsson, R.G., Sweedler, M.E., An associative orthogonal bilinear form for Hopf algebras, Amer. J. Math. **91** (1969), 75–93.
[130] Lin, B.I., Semiperfect coalgebras, J. Algebra **49** (1977), 357–373.
[131] Lu, J.H., Hopf algebroids and quantum groupoids, Int. J. Math. **7** (1996), 47–70.
[132] Mac Lane, S., *Categories for the Working Mathematician*, 2nd ed., Springer, New York, 1998.
[133] Maszczyk, T. Noncommutative geometry through monoidal categories, arXiv:math.QA/0611806v2, 2007.
[134] Masuoka, A., Corings and invertible bimodules, Tsukuba J. Math. **13** (1989), 353–362.
[135] Masuoka, A., Co-Galois theory for field extensions, J. Math. Soc. Japan **41** (1989), 577–592.
[136] Masuoka, A. Review of reference [36], MathSciNet MR1684158 (2000c:16047), 2000.
[137] Mesablishvili, B., Pure morphisms of commutative rings are effective descent morphisms for modules – a new proof, Theory Appl. Categ. **7** (2000), 38–42.
[138] Mesablishvili, B., Monads of effective descent type and comonadicity, Theory Appl. Categ. **16** (2006), 1–45.
[139] Mesablishvili, B., On nonabelian cohomology of comonads, Preprint, 2007.
[140] Michaelis, W., Coassociative coalgebras, in: M. Hazewinkel, ed., *Handbook of Algebra*, vol. **3**, North-Holland, Amsterdam, 2005, pp. 587–788.
[141] Morita, K., Adjoint pairs of functors and Frobenius extensions, Sci. Rep. Tokyo Kyoiku Daigaku Sect. A **9** (1965), 40–71.
[142] Müger, M., From subfactors to categories and topology I. Frobenius algebras in and Morita equivalence of tensor categories, J. Pure Appl. Algebra **180** (2003), 81–157.
[143] Nakayama, T., Tsuzuku, T., On Frobenius extensions I, Nagoya. Math. J. **17** (1960), 89–110.
[144] Năstăsescu, C., Van den Bergh, M., Van Oystaeyen, F., Separable functors applied to graded rings, J. Algebra **123** (1989), 397–413.
[145] Nill, F., Axioms for a weak bialgebra, arXiv:math.QA/9805104, 1998.
[146] Nuss, P., Wambst, M., Non-Abelian Hopf cohomology, J. Algebra **312** (2007), 733–754.
[147] O'uchi, M., Coring structures associated with multiplicative unitary operators on Hilbert C^*-modules, Far East J. Math. Sci. **11** (2003), 121–136.
[148] O'uchi, M., Coring structures on a Hilbert C^*-module of compact operators, Far East J. Math. Sci. **15** (2004), 193–201.
[149] Quillen, D., Module theory over non-unital rings, Preprint, 1997.
[150] Panaite, F., Staic, M.D., Generalized (anti) Yetter-Drinfeld modules as components of a braided T-category, Israel J. Math. **158** (2007), 349–365.
[151] Pappacena, C.J., The injective spectrum of a noncommutative space, J. Algebra **250** (2002), 559–602.
[152] Pareigis, B., Vergessende Funktoren und Ringhomomorphismen, Math. Z. **93** (1966), 265–275.
[153] Pareigis, B., Tensor products and forgetful functors of entwined modules, in: A. Krieg, S. Walcher eds., *The Pumplün 70 Festschrift*, RWTH Aachen University, Aachen, 2003, pp. 1–12.
[154] Pierce, R., *Associative Algebras*, Springer, Berlin, 1982.
[155] Podleś, P. Quantum spheres, Lett. Math. Phys., **14** (1987), 193–202.
[156] Popesco, N., Gabriel, P., Caractérisation des catégories abéliennes avec générateurs et limites inductives exactes, C. R. Acad. Sci. Paris **258** (1964), 4188–4190.
[157] Positselski, L., Homological algebra of semimodules and semicontramodules, arXiv:0708.3398, 2007.
[158] Radford, D.E., Towber, J., Yetter-Drinfeld categories associated to an arbitrary algebra, J. Pure Appl. Algebra **87** (1993), 259–279.
[159] Rafael, M.D., Separable functors revisited, Comm. Algebra **18** (1990), 1445–1459.
[160] Rangipour, B., Cyclic cohomology of corings, arXiv:math/0607248, 2006.

[161] Roiter, A.V., Matrix problems, in: Acad. Sci. Fennica, Helsinki, *Proceedings of the International Congress of Mathematicians (Helsinki, 1978)*, 1980, pp. 319–322.
[162] Roiter, A.V., Matrix problems and representations of BOCS's, in: Lecture Notes in Mathematics, vol. **831**, Springer-Verlag, Berlin and New York, 1980, pp. 288–324.
[163] Rosenberg, A.L, *Non-commutatie Algebraic Geometry and Representations of Quantized Algebras,* Kluwer Acadmic Publishers, Dordrecht, 1995.
[164] Rosenberg, A.L, Noncommutative schemes, Compositio Math. **112** (1998), 93–125.
[165] Schauenburg, P., Bialgebras over noncommutative rings and structure theorems for Hopf bimodules, Appl. Categ. Str. **6** (1998), 193–222.
[166] Schauenburg, P., Schneider, H.-J., On generalised Hopf-Galois extensions, J. Pure Appl. Algebra **202** (2005), 168–194.
[167] Schneider, H.-J., Principal homogeneous spaces for arbitrary Hopf algebras, Israel J. Math. **72** (1990), 167–195.
[168] Schweigert, C., Fuchs, J., Runkel, I., Categorification and correlation functions in conformal field theory, in: M. Sanz-Solé, J. Soria, J.L. Varona, J. Verdera, eds., Proceedings of the International Congress of Mathematicians 2006, vol. **III**, EMS, Zürich, 2006, pp. 443–458.
[169] Sitarz, A., On matrix type corings, algebra coverings and Čech cohomology, Int. J. Geom. Methods Mod. Phys. **4** (2007), 1099–1105.
[170] Škoda, Z., Distributive laws for actions of monoidal categories, arXiv:math/0406310v2, 2004.
[171] Škoda, Z., Noncommutative localization in noncommutative geometry, in: A. Ranicki, ed., *Noncommutative Localization in Algebra and Topology*, Cambridge University Press, Cambridge, 2006, pp. 220–313.
[172] Smith, S.P., Integral non-commutative spaces, J. Algebra **246** (2001), 793–810.
[173] Smith, S.P., Subspaces of non-commutative spaces, Trans. Amer. Math. Soc. **354** (2002), 2131–2171.
[174] Stenström, B., *Rings of Quotients*, Springer, Berlin, 1975.
[175] Street, R., The formal theory of monads, J. Pure Appl. Algebra **2** (1972), 149–168.
[176] Sugano, K., Note on separability of endomorphism rings, Hokkaido Math. J. **11** (1982), 111–115.
[177] Sweedler, M.E., Integrals for Hopf algebras, Ann. Math. **89** (1969), 323–335.
[178] Sweedler, M.E., The predual theorem to the Jacobson-Bourbaki theorem, Trans. Amer. Math. Soc. **213** (1975), 391–406.
[179] Sweedler, M.E., Groups of simple algebras, Inst. Hautes Études Sci. Publ. Math. **44** (1975), 79–189.
[180] Takeuchi, M., Morita theorems for categories of comodules, J. Fac. Sci. Univ. Tokyo Sect. IA Math. **24** (1977), 629–644.
[181] Takeuchi, M., Groups of algebras over $A \otimes \bar{A}$, J. Math. Soc. Japan **29** (1977), 459–492.
[182] Takeuchi, M., Relative Hopf modules – equivalences and freeness criteria, J. Algebra **60** (1979), 452–471.
[183] Takeuchi, M., $\mathrm{Ext}_{\mathrm{ad}}(\mathrm{Sp}R, \mu^A) \simeq \hat{\mathrm{Br}}(A/k)$, J. Algebra **67** (1980), 436–475.
[184] Takeuchi, M., $\sqrt{Morita}$ theory: Formal ring laws and monoidal equivalences of categories of bimodules, J. Math. Soc. Japan **39** (1987), 301–336.
[185] Turi, D., Plotkin, G., Towards a mathematical operational semantics, *Proceedings 12th Ann. IEEE Symp. on Logic in Computer Science, LICS'97,* Warsaw, Poland, 1997.
[186] Van den Bergh, M., Blowing-up of non-commutative smooth surfaces, Mem. Amer. Math. Soc. **734** (2001).
[187] Vercruysse, J., Equivalences between categories of modules and categories of comodules, arXiv:math/0604423, 2006, to appear in Acta Math. Sin. (Engl. Ser.).
[188] Vercruysse, J., Local units versus local projectivity dualisations: corings with local structure maps, Comm. Algebra **34** (2006), 2079–2103.
[189] Vercruysse, J., *Galois Theory for Corings and Comodules,* PhD thesis, Vrije Universiteit Brussel, Brussels, 2007.
[190] Voronov, A., Semi-infinite homological algebra, Invent. Math. **113** (1993), 121–154.
[191] Watatani, Y., Index for C^*-subalgebras, Mem. Amer. Math. Soc. **424** (1990).

[192] Watts, C.E., Intrinsic characterizations of some additive functors, Proc. Amer. Math. Soc. **11** (1960), 5–8.
[193] Wijayanti, I.E., Coprime Modules and Comodules, PhD Thesis, University of Düsseldorf, 2006.
[194] Wijayanti, I.E., Wisbauer, R., On coprime modules and comodules, Preprint, 2007.
[195] Wisbauer, R., *Foundations of Module and Ring Theory*, Gordon and Breach, Reading-Paris, 1991.
[196] Wisbauer, R., Weak corings, J. Algebra **245** (2001), 123–160.
[197] Wisbauer, R., On the category of comodules over corings, in: S. Elaydi, E.S. Titi, M. Saleh, S.K. Jain and R. Abu-Saris, eds., *Mathematics and mathematics Education (Bethlehem, 2000)*, World Sci. Publ., River Edge, NJ, 2002, pp. 325–336.
[198] Wisbauer, R., From Galois field extensions to Galois comodules, in: J.L. Chen, N.Q. Ding and H. Marubayashi, *Advances in ring theory*, World Sci. Publ., Hackensack, NJ, 2005, pp. 263–281.
[199] Wisbauer, R., On Galois comodules, Comm. Algebra **34** (2006), 2683–2711.
[200] Wisbauer, R., Algebras versus coalgebras, Appl. Categ. Str. **16** (2008), 255–295.
[201] Yetter, D.N., Quantum groups and representations of monoidal categories, Math. Proc. Camb. Phil. Soc. **108** (1990), 261–290.
[202] Zarouali-Darkaoui, M., The Picard group for corings, Comm. Algebra **35** (2007), 4018–4031.
[203] Zarouali-Darkaoui, M., *Adjoint and Frobenius Pairs of Functors, Equivalences, and the Picard Group for Corings,* PhD Thesis, University of Granada, 2007 (arXiv:math/0703763).
[204] Zimmermann-Huisgen, B., Pure submodules of direct products of free modules, Math. Ann. **224** (1976), 233–245.

Witt vectors. Part 1[1]

Michiel Hazewinkel

Burg. s'Jacob laan 18, 1401BR Bussum, The Netherlands
E-mail: michhaz@xs4all.nl

Contents

[1] Part 2 (and last) is planned for volume 7 of the Handbook of Algebra.

HANDBOOK OF ALGEBRA, VOL. 6
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(08)00207-6

Part 2 of this survey chapter will include discussions of the following topics: **Symm** vs **BU**; **Symm** vs Exp; **Symm** vs $R_{rat}(GL_\infty)$; Formal groups; Classification of affine commutative algebraic groups. Cartier duality; Covectors and bivectors; The K-theory of endomorphisms; de Rham-Witt complex. Witt vector cohomology; The lifting game; Witt vectors and deformations; Ramified Witt vectors; Noncommutative Witt vectors; The elementary particle view of **Symm**; Brief comments on more results on and applications of Witt vectors.

1. Introduction and delimitation

Let **CRing** be the category of commutative rings with unit element. Most of this chapter is about the big (or large or generalized) Witt vectors; that is about a certain functor

$$W\colon \mathbf{CRing} \longrightarrow \mathbf{CRing} \tag{1.1}$$

that (therefore) assigns to each unital commutative ring A a new unital commutative ring $W(A)$ (and to a unit element preserving ring morphism $A \longrightarrow B$ a unital ring morphism $W(A) \longrightarrow W(B)$). It is also about the many quotient functors of W of which the most important are the p-adic Witt vectors W_{p^∞}, the truncated big Witt vectors W_n, and the truncated p-adic Witt vectors W_{p^n}. Instead of 'truncated' one also finds 'of finite length' in the literature.

The p-adic Witt vectors functor gets defined via the socalled p-adic Witt polynomials

$$Y_0,\ Y_0^p + pY_1,\quad Y_0^{p^2} + pY_1^p + p^2Y_2,\quad Y_0^{p^3} + pY_1^{p^2} + p^2Y_2^p + p^3Y_3,\ \cdots \tag{1.2}$$

and their generalizations, the big Witt polynomials

$$\begin{aligned} w_1(X) &= X_1 \\ w_2(X) &= X_1^2 + 2X_2 \\ &\cdots \qquad\quad \cdots \\ w_n(X) &= \sum_{d|n} dX_d^{n/d} \\ &\cdots \qquad\quad \cdots \end{aligned} \tag{1.3}$$

can[2] be used to define the big Witt vectors. To see that (1.3) specializes to (1.2) relabel X_{p^i} in $w_{p^n}(X)$ as Y_i.

Here and there in the published literature one finds the Witt polynomials referred to as "mysterious polynomials that come out of nowhere." It is one of my aims in the present screed to try to argue that that is simply not true. There is something inevitable about the Witt polynomials and they turn up naturally in various contexts and with some frequency.

1.4. Historically, this is also how things started. The setting is that of the investigations by Helmut Hasse and his students and collaborators into the structure of complete discrete valuation rings A with residue field k. Oswald Teichmüller discovered that in such a situation there is a multiplicative system of representatives, i.e. a multiplicative section of the natural projection $A \longrightarrow k$. Such a system is unique if k is perfect.[3] These are now called Teichmüller representatives. Assuming that A

[2] But there are other ways. One of these will be used below (most of the time).

[3] If k is not perfect a multiplicative system of representatives still exists but is not unique.

is unramified, every element of A can be written as a power series in T, a generator of the maximal ideal of A, with coefficients from any chosen system of representatives. Given the nice properties of the Teichmüller system it was and is natural to use these. As they are already multiplicative the first problem was of course to figure out how Teichmüller representatives should be added in the arithmetic of the ring A. In the equal characteristic case, $\mathrm{char}(A) = \mathrm{char}(k) = p > 0$ that is no problem at all. The Teichmüller system is then additive as well. But in the unequal characteristic case, $\mathrm{char}(A) = 0$, $\mathrm{char}(k) = p > 0$ things are very different. In this case one can and does choose $T = p$. In 1936 Teichmüller [382], found a formula for doing precisely this, i.e. adding Teichmüller representatives. The formula is, see loc. cit. p. 156

$$a + b = \sum_{n=0}^{\infty} c_n p^n \tag{1.5}$$

Here a and b are Teichmüller representatives of, say, $\overline{a}$, $\overline{b}$, and the c_n are the Teichmüller representatives of elements $\overline{c}_n$ in k that satisfy

$$\overline{c}_n^{p^n} = r_n(\overline{a}, \overline{b}) \tag{1.6}$$

where the $r_n(X, Y)$ are the integer valued polynomials (recursively) determined by

$$r_0(X, Y)^{p^n} + pr_1(X, Y)^{p^{n-1}} + \cdots + p^n r_n(X, Y) = X^{p^n} + Y^{p^n} \tag{1.7}$$

And here they surface, on the left hand side of (1.7), the p-adic Witt polynomials.

1.8. Some 30+ years later, in another context, a very similar picture emerged. As a functor the big Witt vectors are isomorphic to the functor Λ that assigns to a commutative unital ring A the Abelian group $\Lambda(A)$ of power series with constant term 1 and coeffients in A (under multiplication of power series). Much of Witt vector theory can be developed from this point of view without ever mentioning the generalized Witt polynomials (1.3). Except that it is far from obvious how to get the p-adic Witt vectors in this picture. In the main, this power series treatment is what is given below. But the Witt vectors themselves and the Witt polynomials will not be denied, as will be seen now.

For any functor one is (or at least should be) interested in the operations on it, i.e. its functorial endomorphisms. In the present case a number of obvious and easy operations are the homothety operators $\langle u\rangle f(t) = f(ut)$ and the Verschiebung operators $\mathbf{V}_n f(t) = f(t^n)$. Rather less obvious are the Frobenius operators $\mathbf{f}_n$ which are as much like raising to the n-th power as a morphism of Abelian groups can be. Now what is the sum (pointwise sum of operations) of two homothety operators? The answer is

$$\langle u\rangle + \langle v\rangle = \sum_{n=1}^{\infty} \mathbf{V}_n \langle r_n(u, v)\rangle \mathbf{f}_n \tag{1.9}$$

where this time the polynomials $r_n(X, Y)$ are the integer valued polynomials determined recursively by

$$\sum_{d|n} d r_d(X, Y)^{n/d} = X^n + Y^n \tag{1.10}$$

and there they are, the generalized Witt polynomials (1.3).[4]

To appreciate the near perfect fit of the picture of formulas (1.5)–(1.7) with that of (1.9)–(1.10), reflect that in a field of characteristic p raising to the p-th power is the Frobenius morphism and that multiplication by p (in a characteristic zero ring, say, the p-adic integers) has a shift built into it. Further, the $\langle u \rangle$, which are multiplicative, can be seen, in a very real sense, as Teichmüller representatives.

1.11. The underlying set of $W(A)$ or, better (for the moment), $\Lambda(A)$, is the set of power series over A with constant term 1 (the unit element of A):

$$\Lambda(A) = \{1 + a_1 t + a_2 t^2 + a_3 t^3 + \cdots : a_i \in A \text{ for all } i \in \mathbf{N}\} \tag{1.12}$$

(where, as usual, $\mathbf{N}$ denotes the natural numbers $\mathbf{N} = \{1, 2, 3, \cdots\}$). This functor (to **Set**, the category of sets), is obviously representable by the ring

$$\mathbf{Symm} = \mathbf{Z}[h_1, h_2, h_3, \cdots] \tag{1.13}$$

of polynomials in an infinity of indeterminates over the integers $\mathbf{Z}$.

As will rapidly become clear one really should see the h_i as the complete symmetric functions (polynomials)[5] in an infinity of indeterminates[6] $\xi_1, \xi_2, \xi_3, \cdots$. Whence the notation used in (1.13). Also it turns out that many constructions, results, ... for the Witt vectors have their natural counterparts in the theory of symmetric functions (as should be). Thus a chapter on the Witt vectors could very properly include most of symmetric function theory, that very large subject that could fill several volumes. This will, of course, not be done. Also, there will be a separate chapter on the symmetric functions in this Handbook of Algebra.

1.14. Symm, the Hopf algebra of the symmetric functions is a truly amazing and rich object. It turns up everywhere and carries more extra structure than one would believe possible. For instance it turns up as the homology of the classifying space **BU** and also as the the cohomology of that space, illustrating its selfduality. It turns up as the direct sum of the representation spaces of the symmetric group and as

[4] But in this case, I believe, the generalized Witt polynomials were known before formula (1.8) was written down.

[5] Of course **Symm** is also equal to the polynomials in the elementary symmetric functions in the same infinity of indeterminates and one could also work with those (and that is often done). But using the complete symmetric functions works out just a bit more elegantly, especially in connection with the autoduality of the Hopf algebra **Symm**. See sections 11.37 and 12.

[6] Some of the readers hoped for may not be familiar with working with symmetric polynomials in an infinity of indeterminates. There is really nothing to it. But for those that do not feel comfortable about it (and for the pernickety) there is a short appendix on the matter.

the ring of rational representations of the infinite general linear group. This time it is Schur duality that is involved. It is the free λ-ring on one generator. It has a nondegenerate inner product which makes it selfdual and the associated orthonormal basis of the Schur symmetric functions is such that coproduct and product are positive with respect to these basis functions. In this setting, positive, self dual, Hopf algebras with distinguished basis, there is a uniqueness theorem due to Andrei Zelevinsky [425]. But this is not yet entirely satisfactory unless it can be shown that the Schur symmetric functions are in some algebraic way canonical (which seems very likely). **Symm** is also the representing ring of the functor of the big Witt vectors and the covariant bialgebra of the formal group of the big Witt vectors (another manisfestation of its autoduality). Most of these things will be at least touched on below.

As the free λ-ring on one generator it of course carries a λ-ring structure. In addition it carries ring endomorphisms which define a functorial λ-ring structure on the rings $W(A) = \mathbf{CRing}(\mathbf{Symm}, A)$ for all unital commutative rings A. A sort of higher λ-ring structure. Being selfdual there are also co-λ-ring structures and higher co-λ-ring structures (whatever those may be).

Of course, **Symm** carries still more structure: it has a second multiplication and a second comultiplication (dual to each other) that make it a coring object in the category of algebras and, dually, (almost) a ring object in the category of coalgebras.

The functor represented by **Symm**, i.e. the big Witt vector functor, has a comonad structure and the associated coalgebras are precisely the λ-rings.

All this by no means exhausts the manifestations of and structures carried by **Symm**. It seems unlikely that there is any object in mathematics richer and/or more beautiful than this one, and many more uniqueness theorems are needed.

In this chapter I will only touch upon the aspects of **Symm** which relate directly to the Witt vector constructions and their properties.

To conclude this introduction let me remark that the Witt vector construction is a very beautiful one. But is takes one out of the more traditional realms of algebra very quickly. For instance the ring of p-adic Witt vectors of a field is Noetherian if and only if that field is perfect, [63], Chapter IX, page 43, exercise 9. Also, much related, the p-adic Witt vector ring $W_{p^\infty}(k[T])$ of one dimensional affine space over a characteristic $p > 0$ field is not Noetherian, which would appear to rule out any kind of systematic use of the Witt vectors in algebraic geometry as currently practiced.[7] It is perhaps because of this that the Witt vector functors and the rings and algebras thus arising have not really been much studied,[8] except in so far as needed for their applications (which are many and varied).

1.15. *Initial sources of information.* There are a number of (electronic) encyclopedia articles on the Witt vectors, mostly the p-adic Witt vectors, see [4, 5, 6, 7]. These can

[7] This is not really true; witness crystalline cohomologyy and the de Rham-Witt complex (of which cristalline cohomology is the hyper homology) which is made up of W(k) modules.

[8] Should a reader inadvertently get really interested in the Witt vector ring functors he/she is recommended to work through the 57 exercises on the subject in [63], 26 pp worth for the statements only, mostly contributed, I have been told, by Pierre Cartier.

serve to obtain a first sketchy impression. There are also a number of (introductory) lecture notes and chapters in books on the subject, see e.g. [57, 74, 86, 188], [192 Chapter 3], [212, 263], [271, §26] [326 Lectures 6–12], [336]. However, the full story of the Witt vectors is a long and varied one. Here, in the present chapter. I try to present a first outline.

2. Terminology

The big and p-adic and truncated Witt vectors carry ring and algebra structures, and hence, naturally, are sometimes referred to as a ring or algebra of Witt vectors and then (by erosion) as Witt ring and Witt algebra. This is a bit unfortunate and potentially confusing because these phrases mostly carry other totally unrelated meanings.

Mostly, 'Witt ring' refers to a ring of equivalence classes of quadratic forms (with addition and multiplication induced by direct product and tensor product respectively), see e.g. [96, 296].

Further 'Witt algebra' mostly refers to something like the Lie algebra of differential operators spanned by the $d_i = x^{i+1}\frac{d}{dx}$, $i \in \mathbf{Z}$ under the commutator product $[d_i, d_j] = (j-i)d_{i+j}$, i.e. the centerless Virasoro algebra, much studied in physics; see e.g. [182, 410, 426]. The fact that the sub Lie algebra spanned by the d_i for $i \geq -1$ is usually denoted W_1 adds a bit more potential confusion, and this becomes worse when one encounters W_n for the more variable version of this algebra of differential operators.

On the other hand the term 'Witt group' mostly indeed refers to a group of Witt vectors.[9]

3. The p-adic Witt vectors. More historical motivation

It has become customary to 'motivate' the introduction of the Witt vectors by looking at unramified complete discrete valuation fields, or, more precisely their rings of integers. To start with, consider the ring of p-adic integers $\mathbf{Z}_p$, i.e. the completion of the integers with respect to the norm $||n|| = p^{-v_p(n)}$ where the valuation $v_p(n)$ of an integer n is the largest power of p that divides n. Here p is a prime number.

Every p-adic integer α, i.e. element of $\mathbf{Z}_p$, can be uniquely written as a convergent sum

$$\alpha = a_0 + a_1 p + a_2 p^2 + a_3 p^3 + \cdots, a_i \in \{0, 1, \cdots, p-1\} \tag{3.1}$$

More generally, instead of the set $\{0, 1, \cdots, p-1\}$ one can choose any set of representatives of the residue field $\mathbf{Z}/(p) = \mathbf{F}_p = \mathbf{GF}(p)$.[10] Now what happens if two such expressions are added or multiplied; i.e. what are the coefficients of a sum or product of p-adic integers? As in the case of the familiar decimal notation for arithmetic

[9] But not always, cf e.g. [82], which is again about quadratic form related matters.

[10] To use three of the standard notations for the prime field of p elements.

with integers this involves carry-overs. That is perfectly acceptable for a calculating machine but irritating to algebraists who would like universal formulas which always work. Things get much worse if the residue field is not a prime field where it is not even clear (a priori) what set of representatives to choose. Of course at this stage it is far from obvious whether a universal formula can exist; that even seems unlikely at first sight.

As it turns out there are universal formulas: the addition and multiplication polynomials of the p-adic Witt vectors.

However, judging from the introduction in Witt's seminal paper [420] the arithmetic of p-adic fields was not Witt's primary motivation. This seems also indicated by the fact that this arithmetic is not mentioned in the title of the paper, but only in the subtitle. Instead Witt seems to have been mainly motivated by a desire to obtain a theory of cyclic Galois extensions for general charateristic p fields similar to the theory he had himself obtained for power series fields in [415, 416] and, especially, to understand some mysterious formulas of Hermann Ludwig Schmid, [355] concerning central cyclic algebras in characteristic p. All this more or less in the framework of a class field theory for function fields that was being vigorously developed at the time by H Hasse, F K Schmidt, O Teichmüller, H L Schmid, C Chevalley, E Witt and others. See [338] for a thorough and very readable account of this part of the history of class field theory. Some of the important original papers are [16, 23, 87, 191, 190, 355, 356, 357, 382, 415, 416, 420].

The situation as regards these cyclic central algebras was as follows. For $\alpha, \beta \in k$ (of characteristic $p > 0$) let $(\alpha, \beta]$ denote the simple central cyclic[11] algebra of degree p with two generators u, θ subject to the relations

$$u^p = \alpha, \quad \theta^p - \theta = \beta, \quad u\theta u^{-1} = \theta + 1$$

In the Brauer group there are the relations

$$(\alpha, \beta] \cdot (\alpha', \beta] = (\alpha\alpha', \beta] \quad \text{and} \quad (\alpha, \beta] \cdot (\alpha, \beta'] = (\alpha, \beta + \beta']$$

H L Schmid in [355] succeeded in defining simple central cyclic algebras of degree p^n denoted $(\alpha|\beta] = (\alpha|\beta_0, \beta_1, \ldots, \beta_{n-1}]$ and by dint of some heroic formula manipulation found rules of the form

$$(\alpha|\beta] \cdot (\alpha'|\beta] = (\alpha\alpha'|\beta] \text{ and } (\alpha|\beta] \cdot (\alpha|\beta'] = (\alpha|s_0(\beta, \beta'), \cdots, s_{n-1}(\beta, \beta')]$$

with, as Witt writes, "certain polynomials $s_i(x, y)$ that are initially defined in characteristic 0 and only after they have been proved to be integral are taken modulo p". In his fundamental paper [420] Witt proves that these formulas are in fact the addition formulas of the p-adic Witt vectors; see also section 5 below. I will also briefly return to these algebras in section 8 below.

[11] The word 'cyclic' here refers to the fact that the algebras considered are cross products involving a cyclic Galois group; see [96], chapter 7, especially §7.2 and §7.5 for more detail. See also section 8 below.

So, here is already a third way in which the Witt vector polynomials turn up naturally and unavoidably.

4. Teichmüller representatives

Let A be a ring with an ideal $\mathfrak{m}$ such that the quotient ring $k = A/\mathfrak{m}$ of characteristic $p > 0$ is perfect, which means that the ring morphism $\sigma\colon k \longrightarrow k, x \mapsto x^p$, the Frobenius morphism, is supposed to be bijective.[12] Suppose, moreover, that A is complete in the $\mathfrak{m}$-adic topology. For instance A can be the ring of p-adic numbers $\mathbf{Z}_p$, the quotient ring $\mathbf{Z}/(p^m)$, or the ring of power series $k((T))$ over a perfect field k of characteristic p. Here the corresponding ideals $\mathfrak{m}$ are respectively the principal ideals $p\mathbf{Z}$, $p\mathbf{Z}/(p^m)$, and (T).

There is now the following simple observation:

$$\text{For all } a, b \in A, \text{ if } a \equiv b \bmod \mathfrak{m}^r, r \geq 1, \text{ then } a^p \equiv b^p \bmod \mathfrak{m}^{r+1} \quad (4.1)$$

Now for any $x \in k$ take lifts $y_r \in A$ of $\sigma^{-r}(x), r = 0, 1, 2, \cdots$; i.e. $q(y_r) = \sigma^{-r}(x)$ where $q\colon A \longrightarrow k$ is the canonical projection, so that all the $y_r^{p^r}$ are lifts of x. Now consider the sequence $y_0, y_1^p, y_2^{p^2}, \cdots, y_r^{p^r}, \cdots$. It follows from (4.1) and the completeness hypothesis on A that this sequence converges to a limit that is a lift of x. Moreover, again by (4.1), this limit does not depend on the choice of the y_r. This limit is denoted $t(x)$ or $t_A(x)$ and called the Teichmüller representative of x. This Teichmüller system of representatives has the following properties

$$t(0) = 0, \quad t(1) = 1, \quad t(xx') = t(x)t(x') \quad (4.2)$$

i.e. it is multiplicative. (And if A is also of characteristic p it is also additive: $t_A(x + x') = t_A(x) + t_A(x')$. But in general this is not the case.)

The Teichmüller system is also the unique multiplicative one and the unique one which commutes with p-th powers.

5. Construction of the functor of the *p*-adic Witt vectors

Let p be a prime number and consider the following polynomials in a countable infinity of commuting indeterminates $X_0, X_1, X_2, \cdots$

$$\begin{aligned} w_0(X) &= X_0 \\ w_1(X) &= X_0^p + pX_1 \\ &\cdots \quad \cdots \\ w_n(X) &= X_0^{p^n} + pX_1^{p^{n-1}} + \cdots + p^{n-1}X_{n-1}^p + p^n X_n \\ &\cdots \quad \cdots \end{aligned} \quad (5.1)$$

[12] For instance the finite fields $\mathbf{GF}(p^f)$ are perfect. There are also perfect rings that are not fields, for instance, the ring $k[T^{p^{-i}}, i = 0, 1, 2, \cdots]$ for a perfect field of constants k.

These are called the (p-adic) Witt polynomials. Now there occurs what has been called 'the miracle of the Witt polynomials'. This is the following integrality statement.

5.2. *Theorem.* Let $\varphi(X, Y, Z)$ be a polynomial over the integers in three (or less, or more) commuting indeterminates. Then there are unique polynomials $\varphi_n(X_0, X_1, \cdots, X_n; Y_0, Y_1, \cdots, Y_n; Z_0, Z_1, \cdots, Z_n) = \varphi_n(X; Y; Z), n = 0, 1, 2, \cdots$ such that for all n

$$w_n(\varphi_0(X; Y; Z), \cdots, \varphi_n(X; Y; Z)) = \varphi(w_n(X), w_n(Y), w_n(Z)) \tag{5.3}$$

The proof is really quite simple. For instance by induction using the following simple observation.

5.4. *Lemma.* Let $\psi(X) = \psi(X_i, i \in I)$ be a polynomial (or power series for that matter) in any set of commuting indeterminates with integer coefficients. Write $\psi(X^p)$ for the polynomial obtained from $\psi(X)$ by replacing each indeterminate by its p-th power. Then

$$\psi(X^p) \equiv \psi(X)^p \mod p \quad \text{and} \quad \psi(X^p)^{p^j} \equiv \psi(X)^{p^{j+1}} \mod p^{j+1} \tag{5.5}$$

Note the similarity with (4.1).

5.6. *Proof of theorem* 5.2. Obviously the polynomials φ_n are unique and can be recursively calculated from (5.3) over the rationals, starting with $\varphi_0(X_0; Y_0; Z_0) = \varphi(X_0; Y_0; Z_0)$. This also provides the start of the induction. So suppose with induction that the φ_i have been proved integral for $i = 0, 1, \cdots, n-1$. Now observe that

$$w_n(X) \equiv w_{n-1}(X^p) \mod p$$

and hence

$$\begin{aligned} \varphi(w_n(X), w_n(Y), w_n(Z)) &\equiv \varphi(w_{n-1}(X^p), w_{n-1}(Y^p), w_{n-1}(Z^p)) \\ &= w_{n-1}(\varphi_0(X^p; Y^p; Z^p), \cdots, \varphi_{n-1}(X^p; Y^p; Z^p)) \end{aligned} \tag{5.7}$$

Using (5.5) one has

$$p^{n-i}\varphi_{n-i}^{p^i}(X; Y; Z) \equiv p^{n-i}\varphi_{n-i}^{p^{i-1}}(X^p; Y^p; Z^p) \mod p^n$$

and so the n terms of the last expression in (5.7) are term for term equal $\bmod p^n$ to the first n terms of

$$w_n(\varphi_0(X; Y; Z), \cdots, \varphi_n(X; Y; Z)) = \varphi_0^{p^n} + p\varphi_1^{p^{n-1}} + \cdots + p^{n-1}\varphi_{n-1}^p + p^n\varphi_n.$$

Hence $p^n\varphi_n(X; Y; Z) \equiv 0 \bmod p^n$, i.e. $\varphi_n(X; Y; Z)$ is integral.

There are also other proofs; for instance a very elegant one due to Lazard, [258], and reproduced in [367].

5.8. It is obvious from the proof that the theorem holds for polynomials φ in any number of variables or power series in any number of variables, but things tend to get a bit messy notationally. It is also obvious that there are all kinds of versions for other

rings of coefficients; see also subsection 9.77–9.98 below for some more remarks on this theme.

Actually, as will be seen later, see section 9 on the big Witt vectors, theorem 5.2 is not at all necessary for the construction of the various functors of Witt vectors and to work with them; nor, for that matter the Witt polynomials. There are several other ways of doing things.

5.9. *The p-adic Witt addition and multiplication polynomials.* The addition polynomials $s_n(X_0, \cdots, X_n; Y_0, \cdots, Y_n)$ and multiplication polynomials $m_n(X_0, \cdots, X_n; Y_0, \cdots, Y_n)$ of the p-adic Witt vectors are now defined by

$$\begin{aligned} w_n(s_0, s_1, \cdots, s_n) &= w_n(X) + w_n(Y) \text{ and} \\ w_n(m_0, m_1, \cdots, m_n) &= w_n(X)w_n(Y) \end{aligned} \tag{5.10}$$

5.11. And the functor of the p-adic Witt vectors itself is defined as

$$W_{p^\infty}(A) = \left\{ (a_0, a_1, \cdots a_n, \cdots) \colon a_i \in A \right\} = A^{\mathbf{N}\cup\{0\}} \tag{5.12}$$

as a set and with multiplication and addition defined by

$$\begin{aligned} a +_W b &= (a_0, a_1, \cdots, a_n, \cdots) +_W (b_0, b_1, \cdots, b_n, \cdots) \\ &= (s_0(a;b), s_1(a;b), \cdots, s_n(a;b_0), \cdots) \\ a \cdot_W b &= (a_0, a_1, \cdots, a_n, \cdots) \cdot_W (b_0, b_1, \cdots, b_n, \cdots) \\ &= (m_0(a;b), m_1(a;b), \cdots, m_n(a;b_0), \cdots) \end{aligned} \tag{5.13}$$

There is a zero element, viz $(0, 0, 0, \cdots)$ and a unit element, viz $(1, 0, 0, \cdots)$ and the claim is:

5.14. *Theorem.* The sets $W_{p^\infty}(A)$ together with the addition and multiplication defined by (5.9) and the unit and zero element as specified define a commutative unital ring valued functor, where the ring morphism corresponding to a ring morphism $\alpha\colon A \longrightarrow B$ is component wise, i.e. $W_{p^\infty}(\alpha)(a_0, a_1, \cdots) = (\alpha(a_0), \alpha(a_1), \cdots)$. Moreover, the Witt polynomials w_n define functor morphisms $w_n\colon W_{p^\infty}(A) \longrightarrow A$, $a = (a_0, a_1, a_2, \cdots) \mapsto w_n(a)$. Finally for $\mathbf{Q}$-algebras $w\colon W_{p^\infty}(A) \longrightarrow A^{\mathbf{N}\cup\{0\}}$, $a \mapsto w(a) = (w_0(a), w_1(a), \cdots)$ is an isomorphism onto the ring $A^{\mathbf{N}\cup\{0\}}$ with componentwise addition and multiplication.

W_{p^∞} is also (obviously) the unique functor which setwise looks like $A \mapsto A^{\mathbf{N}\cup\{0\}}$ and for which the w_n are functorial ring morphisms.

The elements $w_n(a)$ for a Witt vector a are often called the ghost components of that Witt vector (originally: 'Nebenkomponente').

5.15. *Proofs of theorem* 5.14. To prove commutativity, associativity, distributivity and that the zero and unit element have the required properties there are several methods. One is to use defining polynomials as in (5.3). For instance the sequences of polynomials $s(m(X;Y); m(X;Z))$ and $m(X; s(Y;Z))$ both satisfy (5.3) for the

polynomial $\varphi(X, Y, Z) = X(Y + Z) = XY + XZ$ and so they are equal proving distributivity on the left.

Another way is to rely on functoriality. First, by the last line in the theorem (which is obvious) distributivity etc. hold for $\mathbf{Q}$-algebras. Then, because a commutative integral domain embeds injectively into its ring of quotients, the required properties hold for integral domains. Finally for every unital commutative ring there is an integral domain that surjects onto it and so it follows that the required properties hold for every unital commutative ring.

5.16. *Ghost component equations. Universal calculations* (= '*Calculating with universal polynomials*'). The polynomials $s_n(X;Y), m_n(X; Y)$ are solutions of what I call 'ghost component equations'. For instance the ghost component equations for addition can be written

$$w_n(\text{addition}) = w_n(X) + w_n(Y), \quad n = 0, 1, 2, \cdots \tag{5.17}$$

and call for a sequence of polynomials $s(X;Y) = (s_0(X;Y), s_1(X;Y), \cdots)$ such that $w_n(s) = w_n(X) + w_n(Y)$.

For the unit Witt vector one has the ghost component equations

$$w_n(\text{unitvector}) = 1, \quad n = 0, 1, 2, \cdots \tag{5.18}$$

which call for a series of polynomials u (which turn out to be constants) such that $w_n(u) = 1$ for all n. (So that $u_0 = 1, u_n = 0$ for $n \geq 1$.)

There will be many 'ghost component equations' below. They constitute a most useful and elegant tool, though, as has already been remarked on, one can perfectly well do without them. Not all of these ghost component equations fall within the scope of theorem 5.2, see e.g. subsections 5.25 and 5.27 below.

Here are some more general properties of the p-adic Witt vector functor.

5.19. *Ideals and topology.* For each $n = 1, 2, 3, \cdots$

$$\mathfrak{m}_n = \{(0, \cdots, 0, a_n, a_{n+1}, a_{n+2}, \cdots): a_j \in A\} \tag{5.20}$$

is a (functorial) ideal in $W_{p^\infty}(A)$; further $\mathfrak{m}_i\mathfrak{m}_j \subset \mathfrak{m}_{\max(i,j)}$ (obviously, and no more can be expected in general;[13] for instance the multiplication polynomial m_1 has the term pX_1Y_1 and so in general $\mathfrak{m}_1\mathfrak{m}_1 \not\subset \mathfrak{m}_2$). The ring $W_{p^\infty}(A)$ is complete and Hausdorff in the topology defined by these ideals. The quotients $W_{p^n}(A) = W_{p^\infty}(A)/m_{n+1}$ are the rings of p-adic Witt vectors of finite length $n + 1$.

5.21. *Teichmüller representatives.* For each $x \in A$ let $t(x) = t_{W_{p^\infty}(A)}(x) = (x, 0, 0, \cdots)$. This is the Teichmüller representative of x (for the natural projection w_0: $W_{p^\infty}(A) \longrightarrow A = W_{p^\infty}(A)/\mathfrak{m}_1$). This system of representatives is indeed multiplicative because, as is easily checked, $m_0(X_0; Y_0) = X_0Y_0$ and $m_n(X_0, 0, \cdots, 0; Y_0, 0, \cdots, 0) = 0$ for $n \geq 1$.

[13] But things change very much for the case of p-adic Witt vectors over a ring of characteristic p; see section 6.

5.22. *Multiplication with a Teichmüller representative.* It is an easy exercise to check that for the multiplication polynomials

$$m_n(X_0, 0, \cdots, 0; Y_0, Y_1, \cdots, Y_n) = X_0^{p^n} Y_n \tag{5.23}$$

and so in every $W_{p^\infty}(A)$

$$(a_0, 0, \cdots) \cdot_W (b_0, b_1, \cdots, b_n, \cdots) = (a_0 b_0, a_0^p b_1, \cdots, a_0^{p^n} b_n, \cdots) \tag{5.24}$$

5.25. *Verschiebung.* Consider the ghost component equations

$$w_0(\mathbf{V}_p) = 0, w_n(\mathbf{V}_p) = p w_{n-1} \quad \text{for } n \geq 1 \tag{5.26}$$

These call for a series of polynomials $v = (v_0, v_1, v_2, \cdots)$ such that $w_0(v) = 0$, $w_n(v) = p w_{n-1}(v)$ for $n \geq 1$ and do not fall within the scope of theorem 5.2. The immediate and obvious solution is

$$v_0(X) = 0, v_n(X) = X_{n-1} \quad \text{for } n \geq 1$$

and so for each $a = (a_0, a_1, \cdots) \in W_{p^\infty}(A)$ the operation $\mathbf{V}_p$ acts like $\mathbf{V}_p a = (0, a_0, a_1, a_2, \cdots)$. Then because $w_n(\mathbf{V}_p a) = p w_{n-1}(a)$, $w_n(\mathbf{V}_p(a +_W b)) = w_n(\mathbf{V}_p a) + w_n(\mathbf{V}_p b)$, and it follows that $\mathbf{V}_p$ defines a functorial group endomorphism of the Witt vectors.[14] It does not respect the multiplication. It is called Verschiebung. Note that, see (5.19)

$$\mathfrak{m}_i(A) = \mathbf{V}_p^i(W_{p^\infty}(A))$$

5.27. *Frobenius.* The ghost component equations for the Frobenius operation are

$$w_n(\mathbf{f}_p) = w_{n+1}, \quad n = 0, 1, 2, \cdots \tag{5.28}$$

This calls for a sequence of polynomials $f = (f_0, f_1, f_2, \cdots)$ such that $w_n(f) = w_{n+1}$, a set of equations that also falls outside the scope of theorem 5.2. For one thing the polynomial f_n involves the indeterminate X_{n+1}; for instance

$$f_0 = X_0^p + pX_1, f_1 = X_1^p + pX_2 - \sum_{i=0}^{p-1} p^{p-i-1} \binom{p}{i} X_0^{ip} X_1^{p-i} \equiv X_1^p \mod p \tag{5.29}$$

It is easy to show that the f_n are integral and that, moreover,

$$f_n(X) \equiv X_n^p \mod p, \quad n = 0, 1, 2, 3, \cdots \tag{5.30}$$

[14] There are pitfalls in calculating with ghost components as is done here. Such a calculation gives a valid proof of an identity or something else only if it is a universal calculation; that is, makes no use of any properties beyond those that follow from the axioms for a unital commutative ring only. That is the case here. See also the second proof of theorem 5.14 in 5.15.

so that for a ring k of characteristic p the Frobenius operations on the p-adic Witt vectors over k are given by

$$\mathbf{f}_p(a_0, a_1, a_2, \cdots) = (a_0^p, a_1^p, a_2^p, \cdots) \tag{5.31}$$

It follows from the ghost component equations (5.28) that the Frobenius operation

$$\mathbf{f}_p(a_0, a_1, a_2, \cdots) = (f_0(a), f_1(a), f_2(a), \cdots) \tag{5.32}$$

is a functorial endomorphism of unital rings of the functor of the p-adic Witt vectors.

5.33. *Multiplication by p for the p-adic Witt vectors.* The operation of taking p-fold sums on the p-adic Witt vectors, i.e. $a \mapsto p \cdot_W a = a +_W a +_W \cdots +_W a$ is (of course) given by the polynomials $P_n = s(s(\cdots(s(X; X); X); \cdots; X)$ which satisfy the ghost component equations

$$w_n(P_0, P_1, P_2, \cdots) = pw_n(X) \tag{5.34}$$

The first two polynomials are $P_0 = pX_0$, $P_1 = X_0^p + pX_1 - p^{p-1}X_0^{p^2}$. With induction one sees

$$P_n(X) \equiv X_{n-1}^p \bmod p \quad \text{for } n \geq 1 \quad (\text{and } P_0(X) \equiv 0 \bmod p) \tag{5.35}$$

For a change, here are the details. According to (5.34) the polynomials P_n are recursively given by the formulas

$$\begin{aligned} &P_0^{p^n} + pP_1^{p^{n-1}} + \cdots + p^{n-1}P_{n-1}^p + p^n P_n \\ &\quad = pX_0^{p^n} + p^2 X_1^{p^{n-1}} + \cdots + p^n X_{n-1}^p + p^{n+1}X_n \end{aligned} \tag{5.36}$$

With induction, assume $P_i \equiv X_{i-1}^p \bmod p$ for $i \geq 1$. This gives $P_i^{p^{n-i}} \equiv X_{i-1}^{p^{n-i+1}} \bmod p^{n-i+1}$ and $p^i P_i^{p^{n-i}} \equiv p^i X_{i-1}^{p^{n-i+1}} \bmod p^{n+1}$. So the middle $n-1$ terms of the left hand side of (5.36) are term for term congruent to the first $n-1$ terms of the right hand side leaving

$$P_0^{p^n} + p^n P_n \equiv p^n X_{n-1}^p + p^{n+1}X_n \bmod p^{n+1}$$

so that indeed $P_n \equiv X_{n-1}^p \bmod p$ because $P_0 = pX_0$ and $p^n > n$.

5.37. *Taking p-th powers in the Witt vectors.* This operation is governed by the ghost component equations

$$w_n(M_0, M_1, M_2, \cdots) = w_n(X)^p \tag{5.38}$$

These polynomials M_i are of course integral as they can be obtained by repeated substitution of the p-adic Witt vector multiplication polynomials into themselves. For the first few polynomials M_i one finds (by direct calculation)

$$M_0(X) = X_0^p, M_1(X) \equiv pX_0^{p-1}X_1 \bmod p^2, \quad M_2(X) \equiv X_0^{p^2-1}X_1^p \bmod p \tag{5.39}$$

5.40. *Adding 'disjoint' Witt vectors.* Finally, suppose for two p-adic Witt vectors $a = (a_0, a_1, a_2, \cdots)$ and $b = (b_0, b_1, b_2, \cdots)$ it is the case that for every n at least one of a_n or b_n is zero. Then $(a_i + b_i)^{p^j} = a_i^{p^j} + b_i^{p^j}$ for each i, j and it follows that $w_n(a_0 + b_0, a_1 + b_1, a_2 + b_2, \cdots) = w_n(a) + w_n(b)$ so that in such a case $a +_W b = (a_0 + b_0, a_1 + b_1, a_2 + b_2, \cdots)$. More generally it now follows that each Witt vector $a = (a_0, a_1, a_2, \cdots)$ is equal to the unique convergent sum

$$a = \sum_{i=0}^{\infty} \mathbf{V}_p^i(t(a_i)) \tag{5.41}$$

5.42. *Product formula.* The Frobenius operator, Verschiebung operator, and multiplication for Witt vectors are related in various ways. One of these is the product formula, [192], page 126, formula (17.3.17):

$$\mathbf{V}_p(a \cdot (\mathbf{f}_p b)) = \mathbf{V}_p(a) \cdot b \tag{5.43}$$

This is a kind of formula that shows up in various parts of mathematics, such as when dealing with direct and inverse images of sheaves in algebraic geometry, in (algebraic) K-theory and in (abstract representation theory when dealing with restriction and induction of representations.[15]

For instance in representation theory such a formula holds with 'induction' in place of $\mathbf{V}_p$ and 'restriction' in place of $\mathbf{f}_p$. It is important enough in representation theory that it has become an axiom in the part of abstract representation theory known as the theory of Green functors. There it takes the form, [409], page 809:

$$I_K^H(a \cdot R_K^H(b)) = I_K^H(a) \cdot b$$

Sometimes this axiom is called the Frobenius axiom. Further in algebraic geometry one has a natural isomorphism $f_*(\mathcal{F} \otimes_{\mathcal{O}_X} f^*\mathcal{E}) \cong f_*\mathcal{F} \otimes_{O_X} \mathcal{E}$ for a morphism of ringed spaces f: $(X, \mathcal{O}_X) \longrightarrow (Y, \mathcal{O}_Y)$ and suitable sheaves $\mathcal{E}$ and $\mathcal{F}$ where f_* and f^* stand for taking direct and inverse images under f, see [189], exercise 5.1(d) on page 124, while in (étale) cohomology one finds in [295], Chapter VI, §6, proposition 6.5, page 250, a cupproduct formula $i_*(i^*(x) \cup y) = x \cup i_*(y)$. These types of formulae are variously called both product formulae and projection formulae.

It is absolutely not an accident that the same kind of formula turns up in Witt vector theory.

To prove the product formula (5.43) one does a universal calculation. Apply the ghost component morphism w_n to both sides of (5.43) to find respectively (for $n \geq 1$):

$$w_n(\mathbf{V}_p(a \cdot \mathbf{f}_p b) = p w_{n-1}(a \cdot \mathbf{f}_p b) = p w_{n-1}(a) w_{n-1}(\mathbf{f}_p b) = p w_{n-1}(a) w_n(b)$$
$$w_n(\mathbf{V}_p a \cdot b) = w_n(\mathbf{V}_p a) w_n(b) = p w_{n-1}(a) w_n(b)$$

[15] The terminology 'product formula' is not particularly fortunate. There are many things called 'product formula' in many part of mathematics. Mostly they refer to formulae that assert that an object associated to a product is the product of the objects associated to the constituents or to such formulae where a global object for e.g. a global field is the product of corresponding local objects. This happens e.g. for norm residue symbols in class field theory. The terminology 'projection formula' (which is also used for formulas like (5.43) is also not particularly fortunate and suffers from similar defects.

and these are equal. As also applying w_0 to the two sides gives the same result, the product formula is proved.

5.44. Here is another interrelation between the Frobenius and Verschiebung operators and multiplication:

$$\mathbf{f}_p\mathbf{V}_p = [p] \tag{5.45}$$

where $[p]$ stands for the operation that takes a p-adic Witt vector into the p-fold sum of itself, $a \mapsto a +_W a +_W \cdots +_W a$.

But in general $\mathbf{V}_p\mathbf{f}_p$ is not equal to $[p]$.

To prove (5.45) one again does a universal calculation: $w_n(\mathbf{f}_p\mathbf{V}_p a) = w_{n+1}(\mathbf{V}_p a) = pw_n(a)$ and also $w_n([p]a) = pw_n(a)$.

6. The ring of p-adic Witt vectors over a perfect ring of characteristic p

In this section k will finally be a perfect ring of characteristic $p > 0$. In this case the ring of p-adic Witt vectors over k has a number of additional nice properties including a first nice universality property. Recall that 'k is perfect' means that the Frobenius ring morphism $\sigma\colon k \longrightarrow k, x \mapsto x^p$ is bijective. One of the aims of this section to show that if k is a perfect field of characteristic p then the p-adic Witt vectors form a characteristic zero complete discrete valuation ring with residue field k.

6.1. *p-adic Witt vectors over a ring of characteristic p.* First, just assume only that k is of characteristic p. Then calculations with p-adic Witt vectors simplify quite a bit. For instance:

$$\begin{aligned}
&\mathbf{f}_p(a_0, a_1, a_2, \cdots) = (a_0^p, a_1^p, a_2^p, \cdots)\\
&[p](a_0, a_1, a_2, \cdots) = p \cdot (a_0, a_1, a_2, \cdots) = (0, a_0^p, a_1^p, a_2^p, \cdots)\\
&p = (0, 1, 0, 0, 0, \cdots)\\
&\mathbf{f}_p\mathbf{V}_p = [p] = \mathbf{V}_p\mathbf{f}_p\\
&\mathbf{V}_p^i a \cdot \mathbf{V}_p^j b = \mathbf{V}_p^{i+j}(a_0^{p^j} b_0^{p^i}, ??, ??, \cdots)
\end{aligned} \tag{6.2}$$

where the ??'s in the last line stand for some not specified polynomial expressions in the coordinates (components) of a and b. The first four of these formulas follow directly from (5.25), (5.30) and (5.35). The last one follows from the fourth one, repeated application of the product formula, and the first formula of (6.2) (and commutativity of the multiplication). Note that the last formula implies that

$$\mathfrak{m}_i\mathfrak{m}_j \subset \mathfrak{m}_{i+j} \tag{6.3}$$

(but equality need not hold). Also there is the corrollary

6.4. *Corollary.* If k is an integral domain of characteristic p then the ring of p-adic Witt vectors over k is an integral domain of characteristic zero.

6.5. *Valuation rings.* A (normalized) discrete valuation[16] on a (commutative) field K is a surjective function $v\colon K \longrightarrow \mathbf{Z} \cup \{\infty\}$ such that

$$\begin{aligned} &v(x) = \infty \quad \text{if and only if} \quad x = 0, \\ &v(xy) = v(x) + v(y), \\ &v(x+y) \geq \min\{v(x), v(y)\} \end{aligned} \tag{6.6}$$

for all $x, y \in K$. Such a valuation can be used to define a norm, and hence a topology, on K by setting $||x|| = r^{-v(x)}$ where r is e.g. an integer > 1. This defines an ultrametric or non-Archimedean metric, $d(x, y) = ||x - y||$, on K where 'ultra' means that instead of the familiar triangle inequality one has the stronger statement $d(x, y) \leq \max\{d(x, z), d(y, z)\}$.

The valuation ring of the valued field (K, v) is the ring

$$A = \{x \in K\colon v(x) \geq 0\} \tag{6.7}$$

Note that such a ring automatically has two properties

$$A \text{ is local with maximal ideal } \mathfrak{m} = \{x \in K\colon v(x) \geq 1\} \tag{6.8}$$

$$\text{The maximal ideal } \mathfrak{m} \text{ is principal} \tag{6.9}$$

One can also start with a valuation on an integral domain A, which now is a surjective function $v\colon A \longrightarrow \mathbf{N} \cup \{0, \infty\}$ such that the properties (6.6) hold. A valuation on the field of fractions $Q(A)$ is then defined by $v(xy^{-1}) = v(x) - v(y)$ (which is independent of how an element of the field of fractions is written as a fraction). Such an integral domain with a valuation is not necessarily a valuation ring. In particular, one or both of the properties (6.8), (6.9) may fail.

6.10. *Example. The p-adic valuation on the integers.* Take a prime number p. Define on the integers the function $v_p(n) = r$ if and only if p^r is the largest power of the prime number p that divides n. This is obviously a valuation, but $\mathbf{Z}$ is not its corresponding valuation ring. That valuation ring is the localization $\mathbf{Z}_{(p)}$ consisting of all rational numbers that can be written as fractions with a denominator that is prime to p.

6.11. However, if A is an integral domain with a valuation and the ring A satisfies properties (6.8), (6.9), then A is a discrete valuation ring. See e.g. [136], page 50, theorem 7.7; [62], Chapter 5, §3, no 6, proposition 9, page 100.

After this intermezzo on discrete valuation ringss let's us return to rings of p-adic Witt vectors.

---

[16] Such valuations as here are also sometimes called 'exponential valuations'. More generally one considers such functions with values in the real numbers (or more general ordered groups) with infinity adjoined, cf e.g. [136], page 20. Then the value group $v(K\backslash\{0\})$ is a subgroup of the additive group of real numbers. Such a group is either discrete (when there is a smallest positive real number in it) or dense. In the discrete case one may as well assume that the value subgroup is the group of integers. Whence the terminology 'normalized discrete'.

6.12. *Units in the ring of p-adic Witt vectors.* Let k be any ring of characteristic p, then a p-adic Witt vector $a = (a_0, a_1, a_2, \cdots)$ is invertible in $W_{p^\infty}(A)$ if and only if a_0 is invertible in k.

Indeed, multiplying with $(a_0^{-1}, 0, 0, 0, \cdots)$ it can be assumed by (5.24) that $a_0 = 1$ (multiplication with a Teichmüller representative). Then $a = 1 + \mathbf{V}_p b$ with $b = (a_1, a_2, \cdots)$. The series $1 - \mathbf{V}_p b + (\mathbf{V}_p b)^2 - (\mathbf{V}_p b)^3 + \cdots$ converges in $W_{p^\infty}(k)$ because of (6.2) or (6.3), say, to an element c. Then $a \cdot c = 1$.

It follows in particular that if k is a field of characteristic p then $W_{p^\infty}(k)$ is a local ring with maximal ideal $\mathfrak{m} = \mathfrak{m}_1 = \mathbf{V}_p W_{p^\infty}(k)$.

6.13. *Rings of p-adic Witt vectors over a perfect ring of characteristic p.* Now let k be a ring of characteristic p that is perfect. There are of course such rings that are not fields, for instance a ring $k'[T_1, T_2, \cdots]$ with the relations $T_i^p = T_{i-1}$ for $i \geq 2$ and k' a perfect field. This field can be suggestively written $k'[T, T^{p^{-1}}, T^{p^{-2}}, \cdots]$.

Now first note that if k is perfect of characteristic p then the ideal $\mathfrak{m} = \mathfrak{m}_1 = \mathbf{V}_p W_{p^\infty}(k)$ is principal and generated by the element $p \in W_{p^\infty}(k)$. This is immediate from the second formula in (6.2) above. Second, observe that in this case $\mathfrak{m}_r = p^r W_{p^\infty}(k) = (\mathfrak{m}_1)^r = \mathbf{V}_p^r W_{p^\infty}(k)$.

In this setting there is the first nice universality property of the p-adic Witt vectors.

6.14. *Theorem.* Let k be a ring of characteristic p that is perfect. Let A be a ring with an ideal $\mathfrak{a}$ such that $A/\mathfrak{a} = k$ and such that A is complete and separated in the topology defined by $\mathfrak{a}$. Then there is a unique morphism of rings φ such that the following diagram commutes

$$\begin{array}{ccc} W_{p^\infty}(k) & \xrightarrow{\varphi} & A \\ \downarrow w_0 & & \downarrow q \\ k & = & k \end{array}$$

Here the topology defined by $\mathfrak{a}$ is the topology defined by the sequence of ideals $\mathfrak{a}^n$, $n = 1, 2, \cdots$ and 'separated' means that $\bigcap_n \mathfrak{a}^n = \{0\}$. Further q is the canonical quotient mapping $A \longrightarrow A/\mathfrak{a} = k$. It is perfectly OK for A to be a ring of characteristic p or for it to be such that $\mathfrak{a}^r = 0$ for some r.

6.15. *Proof of theorem.* 6.14. First, let's prove uniqueness of φ (if it exists at all). To this end observe that by the remarks in (6.13) and (5.41) every element in $W_{p^\infty}(k)$ can be uniquely written in the form of a convergent sum

$$\sum_{n=0}^{\infty} t(a_n^{p^{-n}})\, p^n, \quad a_n \in k$$

and thus (by the completeness and separatedness of A) a ring morphism $\varphi : W_{p^\infty}(k) \longrightarrow A$ is uniquely determined by what it does to the Teichmüller representives. But, as φ is a ring morphsim and hence multiplicative, the $\varphi(t(a))$, $a \in k$ from a system of Teichmüller representatives for k in A. As such systems are unique it follows that there can be at most one φ that does the job.

For existence consider the ghost component ring morphism

$$w_n\colon W_{p^n}(A) = W_{p^\infty}(A)/\mathfrak{m}_{n+1}(A) \longrightarrow A$$

If $r_0, r_1, \cdots, r_n \in \mathfrak{a}$ then $w_n(r_1, r_2, \cdots, r_n) \in \mathfrak{a}^{n+1}$ and so there is an induced ring morphism ψ_n that makes the following diagram commutative

$$\begin{array}{ccc} W_{p^n}(A) & \xrightarrow{w_n} & A \\ \downarrow W_{p^n}(q) & & \downarrow \\ W_{p^n}(k) & \xrightarrow{\psi_n} & A/\mathfrak{a}^{n+1} \end{array}$$

Now define

$$\varphi_n\colon W_{p^n}(k) \longrightarrow A/\mathfrak{a}^{n+1}$$

as the composite

$$W_{p^n}(k) \xrightarrow{W_{p^n}(\sigma^{-n})} W_{p^n}(k) \xrightarrow{\psi_n} A/\mathfrak{a}^{n+1}$$

Then for any $a_0, a_1, \cdots, a_n \in A$ it follows that

$$\varphi_n(q(a_0^{p^n}, q(a_1^{p^n}), \cdots, q(a_n^{p^n})) = a_0^{p^n} + pa_1^{p^{n-1}} + \cdots + p^n a_n \tag{6.16}$$

Now let $x_0, x_1, \cdots, x_n \in k$ and substitute in this formula (6.16) the Teichmüller representatives (in A) $t_A(x_i^{p^{-n}})$ to find that

$$\varphi_n(x_0, x_1, \cdots, x_n) = t_A(x_0) + pt_A\left(x_1^{p^{-1}}\right) + p^2 t_A\left(x_2^{p^{-2}}\right) + \cdots + p^n t_A(x_n^{p^{-n}}) \mod \mathfrak{a}^{n+1}$$

and so the projective limit φ of the φ_n exists, is a ring morphism, and is given by the formula

$$\varphi(x_0, x_1, x_2, \cdots) = t_A(x_0) + pt_A(x_1^{p^{-1}}) + p^2 t_A(x_2^{p^{-2}}) + \cdots \tag{6.17}$$

Of course one can write down this formula directly but then one is faced with proving that it is a ring morphism which comes down to much the same calculations as were carried out just now.

Now suppose in addition that $\mathfrak{a} = (p)$ and that A is of characteristic zero. Then the ring morphism

$$W_{p^\infty}(k) \xrightarrow{\varphi} A, (x_0, x_1, x_2, \cdots) \mapsto x_0 + pt_A(x_1)^p + p^2 t_A(x_2)^{p^2} + \cdots \tag{6.18}$$

is clearly bijective and hence an isomorphism. So there is a (strong) uniqueness result as follows

6.19. *Theorem.* Let k be a perfect ring of characteristic $p > 0$. Then, up to isomorphism, there is precisely one ring A of characteristic zero with residue ring $A/(p) = k$ and complete in the p-adic topology. Moreover A is rigid in the sense that the only ring automorphism of A that induces the identity on k is the identity on A.

6.20. *The p-adic Witt vectors over a perfect field of characteristic $p > 0$.* Finally let k be a perfect field of characteristic p. Then by 6.13, $\mathfrak{m} = \mathfrak{m}_1 = \mathbf{V}_p W_{p^\infty}(k)$ is generated by p. Define v as follows

$$v\colon W_{p^\infty}(k) \longrightarrow \mathbf{N} \cup \{0, \infty\}, \tag{6.21}$$
$$v(0) = \infty,\ v(x_0, x_1, x_2, \cdots) = n \text{ if } x_n \text{ is the first coordinate unequal to zero}$$

Then v is a valuation on the ring of p-adic Witt vectors and makes this ring into a complete discrete valuation ring by (6.11) and (6.12). And this ring is unramified as the maximal ideal is generated by p.

Theorem 6.19 now translates into an existence and uniqueness (and rigidity) theorem for complete discrete unramified valuation rings of characteristic zero and residue characteristic $p > 0$.

Thus, for a perfect residue field there are indeed universal formulae that govern the addition and multiplication of p-adic expansions as in section 3 above. The treatment here has a bit of a 'deus ex machina' flavor in that the p-adic Witt vectors are constructed first and subsequently proved to do the job. Motivationally speaking one can do better. Once it is accepted that it is a good idea to write p-adic expansions using Teichmüller representatives one can calculate and come up with addition and multiplication formulae. This is nicely done in [188]. It was also already known in 1936, [382], that there should be integral coefficient formulae that could do the job.[17] Witt's major contribution was finding a way to describe them nicely and recursively.

6.22. *Complete discrete valuation rings with nonperfect residue field.* Immediately after Witt's paper, in the same[18] issue of 'Crelle', there is a paper by Teichmüller, [380], in which he proves existence and uniqueness of unramified complete discrete valuation rings with given residue field, thus completing his own arguments from [382]. For the unequal characteristic case he uses the p-adic Witt vectors.

Existence and uniqueness of unramified complete discrete valuation rings had been treated before by Friedrich Carl Schmidt and Helmut Hasse in [191], but that paper had an error, which was later, in 1940, pointed out by Saunders MacLane, [277], and still later corrected by him [278]. For more see [339].

6.23. *Cohen rings.* The unique characteristic zero complete unramified discrete valuation ring with a given residue field k of characteristic $p > 0$ is nowadays often

[17] Indeed in [382], the theorem at the bottom of page 58 says the following. Define polynomials $h_0, h_1, h_2, \cdots$ in two variables by the formulae $x^{p^n} + y^{p^n} = h_0^{p^n} + ph_1^{p^{n-1}} + \cdots + p^{n-1}h_{n-1}^p + p^n h_n$. Then these have integer coefficients. Given a, b in the perfect residue field k, let c_n, $n = 0, 1, 2, \cdots$ be given by $c_n^{p^n} = h_n(a, b)$. Then the p-adic expansion of the sum to the Teichmüller representatives of a and b is $t(a) + t(b) = t(c_0) + pt(c_1) + p^2 t(c_2) + \cdots$, which of course nicely agrees with the addition of p-adic Witt vectors via the isomorphism (6.17). Thus the addition formula for Teichmüller representatives was known, and given by universal formulae, and involved what were to be called Witt polynomials. The multiplication of Teichmüller representatives was of course also known (as these are multiplicative). Distributivity then determines things recursively. There was still much to be done and no prediction that things would come out as nicely as they did, but the seeds were there.

[18] The two papers carry the same date of receipt.

called the Cohen ring[19] of k; see [63] Chapter IX, §2 for a treatment of Cohen rings.[20] If k is perfect the Cohen ring of k is the ring of p-adic Witt vectors. If k is not perfect it is a 'much smaller' ring.

The technical definition of Cohen ring in loc. cit. is as follows. A p-ring is (by definition) a unital commutative ring such that the ideal pC is maximal and the ring is complete and separated in the p-adic topology defined by the powers of this ideal. Given a local separated complete ring A a Cohen subring of it is a p-ring that satisfies $A = \mathfrak{m}_A + C$ (where $\mathfrak{m}_A$ is the maximal ideal of A.

The existence and uniqueness of such a subring of the p-adic Witt vectors is implicit in the work of Teichmüller, [380] and Nagata, [304], Chapter V, §31, p. 106ff, but a nice functorial description had to wait till the work of Colette Schoeller, [358].[21]

6.24. *p-basis.* Let k be a field of characteristic $p > 0$. The field is perfect if and only if $k = k^p = \{x^p : x \in k\}$. If it is not perfect there exists a set of elements $\mathcal{B} = \{b_i : i \in I\}$ of k such that the monomials

$$\prod_{i \in I} b_i^{j_i},\ j_i \in \{0, 1, \cdots, p-1\},\ j_i \text{ equal to zero for all but finitely many } i$$

form a basis (as a vector space) for k over the subfield k^p. The cardinality of I, which can be anything, is an invariant of k and sometimes called the imperfection degree of k. Such a set is called a p-basis. The notion is again due to Teichmüller and first appeared in [381] (for other purposes than valuation theory). For some theory of p-bases see [64], Chapter 5, §8, exercise 1, and [304], p. 107ff. If $\mathcal{B}$ is a p-basis for k the monomials

$$\prod_{i \in I} b_i^{j_i},\ j_i \in \{0, 1, \cdots, p^n - 1\},\ j_i \text{ equal to zero for all but finitely many } i$$

form a vector space basis for k over the subfield k^{p^n}.

6.25. *Cohen functor.* Given a characteristic $p > 0$ field k and a chosen p-basis let $\mathbf{Z}[\mathcal{B}]$ be the ring of polynomials over the integers in the symbols from $\mathcal{B}$. Then Schoeller defines functors C_n on the category of commutative unital $\mathbf{Z}[\mathcal{B}]$-algebras

[19] There are other rings that are called 'Cohen rings' in the published literature. First there are certain twisted power series rings which are a kind of noncommutative complete discrete valuation rings, see e.g. [131, 342, 406]. Further in [137], section 18.34B, p. 48, a Cohen ring is defined as a ring R such that R/P is right Artinian for each nonzero prime ideal P; see also [138, 413]. This last bit of terminology has to do with the paper [93].

[20] This terminology would appear to come from the so-called Cohen structure theorems for complete local rings, [94].

[21] Existence of a multiplicative system of representatives is again crucial. In the general case, i.e., when the residue field k is not necessarily perfect there is still existence (as noticed by Teichmüller, [382]), but no uniqueness. A very nice way of seeing existence is due to Kaplansky, [235] section 26, p. 84ff, who notes that the group of 1-units (Einseinheiten) of A, i.e. the invertible elements that are mapped to 1 under the residue mapping, form a direct summand of the group of all invertible elements of A. This remark seems to have been of some importance for Schoeller, and reference 4 in her paper, a paper that never appeared, should probably be taken as referring to this.

to the category of unital commutative rings. These functors depend on the choice of the p-basis. They define affine group schemes.

Now if the characteristic of the base field k is p and perfect the finite length p-adic Witt vector group schemes serve to classify unipotent affine algebraic groups over k, a topic which will get some attention in a later section of this chapter. For a non-perfect base field the Cohen functors defined in [358], see also [248] are a well-working substitute.[22]

6.26. *Cohen ring of k.* It follows from (6.25) that $\mathbf{F}_p[\mathcal{B}]$, where $\mathbf{F}_p$ is the field of p elements, naturally embeds in k; combined with the canonical projection $\mathbf{Z}[\mathcal{B}] \longrightarrow \mathbf{F}_P[\mathcal{B}]$ this defines a $\mathbf{Z}[\mathcal{B}]$-algebra structure on k. The value of the Cohen functors on k can now be described as follows, see [358], section 3.2, p. 260ff; see also [63], Chapter IX, §2, exercise 10, p. 72.

$$C_n(k) \text{ is the subring of } W_{n+1}(k) \text{ generated by } W_{n+1}(k^{p^n}) \subset W_{n+1}(k)$$
$$\text{and the Teichmüller representatives } (b, 0, 0, \cdots, 0) \in W_{n+1}(k),\ b \in \mathcal{B}$$

These form a projective system (with surjective morphisms of rings $\mathcal{C}_{n+1}(k) \to \mathcal{C}_n(k)$ induced by the projections $W_{n+1}(k) \to W_n(k)$) and taking the projective limit one finds a subring $\mathcal{C}(k)$ of the ring of p-adic Witt vectors which is a discrete complete unramified valuation ring. It is unique but one loses the rigidity property that holds for the Witt vectors, see theorem 6.19 and 6.20.

7. Cyclic Galois extension of degree p^n over a field of characteristic p

As already indicated in the introduction (section 1) for Witt himself one of the most important aspects of the Witt vectors was that they could be used to extend and complete his own results from [416, 415] to obtain a Kummer theory (class field theory) for Abelian extensions of a field of characteristic p. This section is a brief outline of this theory. In this whole section p is a fixed prime number.

7.1. *Construction of some Abelian extensions of a field of characteristic $p > 0$.* Consider the functor of p-adic Witt vectors W_{p^∞}. It is an easy observation that for each $n \geq 1$, the multiplication and addition polynomials $s_i(X; Y), m_i(X; Y), i = 0, \cdots, n-1$ only depend on $X_0, \cdots, X_{n-1}; Y_0, \cdots, Y_{n-1}$. It follows immediately that the sets

$$\mathbf{V}_p^n W_{p^\infty}(A) = \{(\underbrace{0, \cdots, 0}_{n}, a_0, a_1, \cdots, a_m, \cdots)\} \tag{7.2}$$

are (functorial) ideals in the rings $W_{p^\infty}(A)$. And hence, for each n, there is a quotient functor

$$W_{p^{n-1}}(A) = W_{p^\infty}(A)/\mathbf{V}_p^n W_{p^\infty}(A) = \{(a_0, a_1, \cdots, a_{n-1}) \colon a_i \in A\} \tag{7.3}$$

[22] There may be some difficulties with the constructions in [358] in case the cardinality of the p-basis is not finite as is stated in [378], where also an alternative is proposed.

These are called the p-adic Witt vectors of length n. The multiplication and addition of these vectors of length n are given by the multiplication and addition polynomials $s_i(X;Y), m_i(X;Y), i = 0.\cdots, n-1$.

In other words $W_{p^{n-1}}$ is the functor on **CRing** to **CRing** represented by $\mathbf{Z}[X_0, X_1, \cdots, X_{n-1}]$ provided with the coring object structure given by the polynomials $s_i(X;Y), m_i(X;Y), i = 0.\cdots, n-1$.

If the ring A is of characteristic p the Frobenius endomorphism on $W_{p^\infty}(A)$ is given by (see (6.2) above)

$$\mathbf{f}_p(a_0, a_1, a_2, \cdots) = (a_0^p, a_1^p, a_2^p, \cdots) \tag{7.4}$$

and thus manifestly takes the ideal $\mathbf{V}_p^n W_{p^\infty}(A)$ into itself. In general this is not the case. Thus for rings of characteristic p there is an induced (functorial) Frobenius endomorphism of rings

$$\mathbf{f}_p\colon W_{p^{n-1}}(A) \to W_{p^{n-1}}(A) \tag{7.5}$$

From now on in this section k is a field of characteristic $p > 0$. Consider the additive operator

$$\mathfrak{p}\colon W_{p^{n-1}}(k) \to W_{p^{n-1}}(k), \mathfrak{p}(\alpha) = \mathbf{f}_p(\alpha) - \alpha \tag{7.6}$$

(Witt vector subtraction, of course.) For $n = 1$, so that $W_{p^{n-1}}(k) = k$ this is the well-known Artin-Schreier operator $x^p - x$ which governs the theory of cyclic extensions of degree p of the field k.

For $n > 1$ the operator (7.6) is very similar, particularly if one reflects that the Frobenius operator is 'as much like raising to the power p as an additive operator on the Witt vectors can be' (as has been remarked before).

Moreover, consider the kernel of the operator $\mathfrak{p}$. For a given n his kernel consists of all p-adic Witt vectors $\alpha = (a_0, a_1, \cdots, a_{n-1})$ of length n such that $a_i^p = a_i$ for all $i = 0, 1, \cdots, n-1$. Thus for all these i, $a_i \in \mathbf{F}_p \subset k$ the prime subfield of p elements of k. Thus

$$\mathrm{Ker}(\mathfrak{p}) = W_{p^{n-1}}(\mathbf{F}_p) = \mathbf{Z}/p^n\mathbf{Z} \tag{7.7}$$

the cyclic group of order p^n. (The last equality in (7.7) is immediate from the considerations of the previous chapter.) This is most encouraging in that it suggests that the study of equations

$$\mathbf{f}_p(\alpha) - \alpha = \mathfrak{p}(\alpha) = \beta \tag{7.8}$$

in $W_{p^{n-1}}(\bar{k})$ with $\beta \in W_{p^{n-1}}(k)$ and $\bar{k}$ an algebraic closure of k could lead to cyclic extensions of degree p^n just like in standard Artin–Schreier theory as described e.g. in [96], pp. 205–206. This is indeed substantially the case and things turn out even better.

For any $\beta \in W_{p^{n-1}}(k)$ consider a solution $\alpha = (a_0, a_1, \cdots, a_{n-1})$ of (7.8). Then $k(\mathfrak{p}^{-1}\beta)$ denotes the extension field of k generated by the components $a_0, a_1, \cdots, a_{n-1}$. Because of (7.7) if $\alpha' = (a_0', a_1', \cdots, a_{n-1}')$ is another solution $k(a_0, a_1, \cdots, a_{n-1}) = k(a_0', a_1', \cdots, a_{n-1}')$ so that the extension $k(\mathfrak{p}^{-1}\beta)$ of k does

not depend on the choice of a solution. More generally if Ω is any subset of $W_{p^{n-1}}(k)$ then $k(\mathfrak{p}^{-1}\Omega)$ denotes the union of all the $k(\mathfrak{p}^{-1}\beta)$ for $\beta \in \Omega$.

7.9. *The main theorems for Abelian extension of exponent p^n of fields of characteristic p.* There is now sufficient notation to formulate the main theorems of Kummer theory for Abelian extensions of exponent p^n of a field of characteristic p.

7.10. *Theorem (Kummer theory).* Let $\Omega \subset W_{p^{n-1}}(k)$ be a subgroup which contains $\mathfrak{p}W_{p^{n-1}}(k)$ and such that $\Omega/\mathfrak{p}W_{p^{n-1}}(k)$ is finite. Then the Galois group of the extension $k(\mathfrak{p}^{-1}\Omega)$ is isomorphic to $\Omega/\mathfrak{p}W_{p^{n-1}}(k)$.

For each Abelian field extension K/k of exponent a divisor of p^n there is precisely one group $\Omega/\mathfrak{p}W_{p^{n-1}}(k)$ such that $K = k(\mathfrak{p}^{-1}\Omega)$.

7.11. *Theorem (on cyclic extensions of degree p^n).* Let $\beta = (b_0, b_1, \cdots, b_{n-1})$ be a p-adic Witt vector of length n over k such that $b_0 \notin \mathfrak{p}W_{p^{n-1}}(k)$. Then $k(\mathfrak{p}^{-1}\beta)$ is a cyclic extension of degree p^n of k.

A generator of the Galois group is given by $\sigma(\alpha) = \alpha + \mathbf{1}$ (Witt vector addition, and $\mathbf{1}$ the unit of the ring $W_{p^{n-1}}(k)$).

All cyclic extensions of degree p^n can be obtained in this way.

7.12. *On the proofs.* The theorems above already occur in [420]. But the proofs there are very terse. They consists of brief instructions to the reader to first prove a kind of (additve) "Hilbert 90 theorem" for Witt vectors by redoing the proof of theorem 1.2 in [415]. This says that a first Galois cohomology group with coefficients in Witt vectors is zero and is 'Satz 11' in [420]. The further instructions are to redo the arguments of [415] using vectors instead of numbers and using 'Satz 11' instead of the usual 'Hilbert 90' as the occasions demand.

For a good complete treatment of this Kummer theory for Abelian extensions of fields of characteristic $p > 0$ see [271], pp. 146–150. The statements there are slightly more general and a bit more elegant than in [420] in that the group $\Omega/pW_{p^{n-1}}(k)$ is not required to be finite. The isomorphism statement of theorem 7.10 now becomes a duality statement to the effect that the group $\Omega/pW_{p^{n-1}}(k)$ and the Galois group of the extension are dual to each other under a natural pairing.

8. Cyclic central simple algebras of degree p^n over a field of characteristic p

The second main application of the p-adic Witt vectors in [420] is to cyclic central simple algebras of prime power degree p^n over a field of characteristic $p > 0$. This is a topic in the theory of simple central algebras over and the Brauer group of a field, [1, 2], [96] Chapter 7, [102] Chapter III.

8.1. *Central simple algebras.* A central simple algebra over a field k is what the name indicates: it is a finite dimensional algebra over k, it is simple, i.e. no nontrivial ideals, and it is central, i.e. its centre, $\{a \in A\colon ar = ra \text{ for all } r \in A\}$ coincides with k. A

central division algebra over k is a central simple algebra in which every nonzero element is invertible. The classic example is the algebra of quaternions over the reals.

For every central simple algebra A there is a (unique up to isomorphism) central division algebra D such that A is isomorphic to a full matrix ring over D.

For every central simple algebra over k there is a field extension[23] K/k such the tensor product $A_k = A \otimes_k K$ is a full matrix ring over K. Such a field is called a splitting field.

The tensor product (over k) of two central simple algebras over k is again a central simple algebra over k.

8.2. *Brauer group.* Two central simple algebras A, B, over k are called (Brauer) equivalent if for suitable natural numbers m and n

$$A \otimes_k M_m(k) \cong B \otimes_k M_n(k) \tag{8.3}$$

Here $M_m(k)$ is the full matrix algebra over k of all $m \times m$ matrices with entries in k.

Equivalently, A and B are equivalent if they have isomorphic associated division algebras.

The tensor product is compatible with this equivalence notion and defines (hence) a group structure on the equivalence classes. This group is called the Brauer group, $\mathrm{Br}(k)$, of k. It can be interpreted as a second Galois cohomology group.

8.4. *Crossed product central simple algebras,* [102] §III-2, [96] §7.5. A central simple algebra over k is called a cross product if it contains a maximal subfield K such that K/k is a Galois extension.

That K is then a splitting field.

This is the 'abstract' definition of a crossed product. There is also an explicit description/construction which is important and explains the terminology. This goes as follows. Let U be the group of invertible elements of a crossed product central simple algebra A. Let N be the centralizer of $K^\times = K\backslash\{0\}$ in U:

$$N = \{u \in U\colon u^{-1}Ku \subseteq K\} \tag{8.5}$$

Then $K^\times$ is a normal subgroup of N and hence gives rise to a short exact sequence of groups

$$1 \longrightarrow K^\times \longrightarrow N \longrightarrow \Gamma \longrightarrow 1 \tag{8.6}$$

a group extension of Γ by $K^\times$. One easily shows that $\Gamma \cong \mathrm{Gal}(K/k)$.

Being a group extension, it is determined by what is called a factor system and this same factor system can be used to recover (up to isomorphism) the crossed product central simple algebra A.

Not every central simple algebra is a crossed product, but every Brauer equivalence class contains one.

[23] As always the term 'field' implies commutativity.

When the Galois extension involved is cyclic, i.e. the group $\Gamma = \mathrm{Gal}(K/k)$ is cyclic, one speaks of a cyclic central simple algebra.

8.7. *Cyclic central algebra associated to a finite length p-adic Witt vector.* From now on in this section k is a fixed field of characteristic $p > 0$. Take an element $\alpha \in k$ and a finite length Witt vector $\beta = (b_0, b_1, \cdots, b_{n-1}) \in W_{p^{n-1}}(k)$. To these data associate the algebra generated over k by an indeterminate u and commuting indeterminates $\theta_0, \theta_1, \cdots, \theta_{n-1}$ subject to the relations

$$u^{p^n} = \alpha, \quad \mathfrak{p}\theta = \beta, \quad u\theta u^{-1} = \theta + 1 \tag{8.8}$$

Here, $\theta = (\theta_0, \theta_1, \cdots, \theta_{n-1})$ (as a Witt vector) and $u\theta u^{-1} = (u\theta_0 u^{-1}, u\theta_1 u^{-1}, \cdots, u\theta_{n-1}u^{-1})$. This algebra will be denoted $(\alpha|\beta]$. These algebras are central simple algebras of 'degree' p^n (meaning that their dimension over k is p^{2n}).

Note that conjugation by u is an operation of order p^n. Also, by the results of Section 7 above, if $b_0 \notin \mathfrak{p}W_{p^{n-1}}(k)$ the subalgebra $k(\theta_0, \theta_1, \cdots, \theta_{n-1})$ is a cyclic field extension of degree p^n and conjugation by u is the action of a generator of the Galois group. This is just about sufficient to prove that $(\alpha|\beta]$ is a cyclic central simple (crossed product type) algebra. In case $b_0 \in \mathfrak{p}W_{p^{n-1}}(k)$ there is a 'reduction theorem', Satz 15 from [420], which says the following.

8.9. *Theorem.* The algebra $(\alpha|0, b_1, b_2, \cdots, b_{n-1}]$ is Brauer equivalent to $(\alpha|b_1, b_2, \cdots, b_{n-1}]$

The next theorem is rather remarkable and shows once more that there is no escaping the Witt vectors.

8.10. *Theorem (Brauer group and Witt vectors).* In the Brauer group of the field k there are the (calculating) rules

$$\begin{aligned} (\alpha|\beta] \cdot (\alpha'|\beta] &= (\alpha\alpha'|\beta] \\ (\alpha|\beta] \cdot (\alpha|\beta'] &= (\alpha|\beta + \beta'] \end{aligned} \tag{8.11}$$

(where in the second line of (8.11) the '+' sign means Witt vector addition.

For the time being the sections above conclude the discussions on p-adic Witt vectors. There will be more about the p-adic Witt vectors in various sections below. But first it is time to say something about that truly universal object the functor of the big Witt vectors; and that will be the subject in the next few sections.

9. The functor of the big Witt vectors

In the early 1960's, probably independent of each other, several people, notably Ernst Witt himself and Pierre Cartier, noticed that the p-adic Witt polynomials (5.1) are but part of a more general family and that these polynomials can be used in a similar manner to define a ring valued functor W: **CRing** $\longrightarrow$ **CRing** of which the p-adic Witt vectors are a quotient, see [105, 149, 192, 255, 419]. This functor W is called

the functor of the big Witt vectors.[24] (And also, over $\mathbf{Z}_{(p)}$-algebras the canonical projection $W(A) \longrightarrow W_{p^\infty}(A)$ has an Abelian group section, making in this case the p-adic Witt vectors also a sub functor of the big Witt vectors.)

Here, I will first construct the big Witt vectors in another way, before describing those 'generalized' Witt polynomials alluded to, and how the p-adic Witt vectors fit with the big ones.

9.1. *The functor of one power series.* For each ring (unital and commutative) let

$$\Lambda(A) = 1 + tA[[t]] = \{f = f(t) = 1 + a_1 t + a_2 t^2 + \cdots : a_i \in A\} \tag{9.2}$$

be the set of power series over A with constant term 1. Under the usual multiplication of power series this is an Abelian group (for which the power series 1 acts as the zero element). Thus (9.2) defines an Abelian group valued functor Λ: **CRing** $\longrightarrow$ **AbGroup**. The morphism of Abelian groups associated to a morphism $\alpha: A \longrightarrow B$ is coefficient-wise, i.e.

$$\Lambda(\alpha)(1 + a_1 t + a_2 t^2 + a_3 t^3 + \cdots) = 1 + \alpha(a_1)t + \alpha(a_2)t^2 + \alpha(a_3)t^3 + \cdots \tag{9.3}$$

Giving a power series like (9.2) is of course the same thing as giving an infinite length vector $(a_1, a_2, a_3, \cdots)$ and in turn such a vector is the same as a morphism of commutative rings

$$\mathbf{Symm} = \mathbf{Z}[h_1, h_2, h_3, \cdots] = \mathbf{Z}[h] \xrightarrow{\varphi_f} A, \quad \varphi_f: h_i \mapsto a_i \tag{9.4}$$

where, as indicated in the notation, **Symm** is the ring of polynomials in an infinity of commuting indeterminates $h_1, h_2, h_3, \cdots$. Thus the functor Λ is representable by **Symm**. The functorial addition on the Abelian group $\Lambda(A)$ then defines a co-multiplication on **Symm**

$$\mu_S = \mu_{Sum}: \mathbf{Symm} \longrightarrow \mathbf{Symm} \otimes \mathbf{Symm}, h_n \mapsto \sum_{i+j=n} h_i \otimes h_j \tag{9.5}$$

where by definition and for ease of notation $h_0 = 1$. This makes **Symm** a Hopf algebra. The comultiplication formula of course encodes the 'universal' formula (that is 'recipe') for multiplying power series which is

$$(1 + a_1 + a_2 + a_3 + \cdots)(1 + b_1 + b_2 + b_3 + \cdots) = 1 + c_1 + c_2 + c_3 + \cdots$$
$$\Leftrightarrow c_n = \sum_{i+j=n} a_i b_j$$

with, again, $a_0 = b_0 = c_0 = 1$. Which, in turn, is the same as saying that the addition on **CRing**(**Symm**, A) is given by the convolution product

$$\varphi_{fg} = \mathbf{Symm} \xrightarrow{\mu_s} \mathbf{Symm} \otimes \mathbf{Symm} \xrightarrow{\varphi_f \otimes \varphi_g} A \otimes A \xrightarrow{m_A} A \tag{9.6}$$

where m_A is the multiplication on the ring A.

[24] In earlier writings also 'generalized Witt vectors'.

As in the case of the p-adic Witt vectors of the previous 7 sections it is often convenient (but never, strictly speaking, necessary) to work with ghost components. These are defined by the formula

$$s(f) = s(a_1, a_2, a_3, \cdots) = s_1 t + s_2 t^2 + s_3 t^3 + \cdots = t\frac{d}{dt}\log(f(t)) = \frac{tf'(t)}{f(t)} \tag{9.7}$$

so that for example the first three ghost components are given by the universal formulas

$$s_1(a) = -a_1, \quad s_2(a) = a_1^2 - a_2, \quad s_3(a) = -a_1^3 + 2a_1a_2 - a_3 \tag{9.8}$$

Because of the properties of 'log' as defined formally by

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^{n+1}\frac{x^n}{n} + \cdots \tag{9.9}$$

so that $\log(fg) = \log(f) + \log(g)$, this implies that addition in $\Lambda(A)$ (which is multiplication of power series) corresponds to coordinate-wise addition of ghost components (just as in the case of the p-adic Witt vectors).

9.10. *The symmetric function point of view.* Now imagine that the a_i are really the complete symmetric functions in a set of elements $\xi_1, \xi_2, \xi_3, \cdots$ (living as it were in some larger ring containing A); i.e.

$$a_n = h_n(\xi_1, \cdots, \xi_n, \cdots) \tag{9.11}$$

where the $h_n(X)$ are the familiar complete symmetric functions in the commuting indeterminates $X_1, X_2, X_3, \cdots$, viz

$$h_n(X) = \sum_{j_1 \le j_2 \le \cdots \le j_n} X_{j_1} X_{j_2} \cdots X_{j_n} \tag{9.12}$$

The relation (9.11) can be conveniently written down in power series terms as

$$\prod_{i=1}^{\infty} \frac{1}{(1-\xi_i)} = 1 + a_1 t + a_2 t^2 + a_3 t^3 + \cdots = f(t) = f \tag{9.13}$$

The ghost components of f are given by (9.7). In terms of the ξ_i that means that

$$\begin{aligned} s_1 t + s_2 t^2 + s_3 t^3 + \cdots &= t\frac{d}{dt}\log f(t) = t\frac{d}{dt}\sum_i \log\left((1-\xi_i t)^{-1}\right) \\ &= -t\frac{d}{dt}\sum_i \left(-\xi_i t - \frac{\xi_i^2 t^2}{2} - \frac{\xi_i^3 t^3}{3} - \cdots\right) \\ &= \sum_{j=1}^{\infty}\left(\xi_1^j + \xi_2^j + \xi_3^j + \cdots\right)t^j \end{aligned} \tag{9.14}$$

so that in terms of the ξ's the ghost components are given by the power sums.

9.15. *Multiplication on the Abelian groups* $\Lambda(A)$. If there is to be a ring multiplication on $\Lambda(A)$ then in particular there should be a multiplication of the very special power series $(1-xt)^{-1}$ and $(1-yt)^{-1}$. Just about the simplest thing one can imagine

is that the product of these very special power series is

$$(1 - xt)^{-1} * (1 - yt)^{-1} = (1 - xyt)^{-1} \tag{9.15}$$

Something which fits with what has been seen in the theory of the p-adic Witt vectors: the $(1 - xt)^{-1}$ are the Teichmüller representatives of the $x \in A$ which should be multiplicative. Distributivity and functoriality together now force that the multiplication of power series in $\Lambda(A)$ should be given by

$$\begin{aligned} f(t) = \prod_i (1 - \xi_i t)^{-1}, \quad g(t) = \prod_i (1 - \eta_i t)^{-1} \\ \Rightarrow (f * g)(t) = \prod_{i,j} (1 - \xi_i \eta_j t)^{-1} \end{aligned} \tag{9.16}$$

Note that this formula makes perfectly good sense. The right most expression is symmetric in the ξ's and also symmetric in the η's and so, when written out, gives a power series with coefficients that are complete symmetric functions in the ξ's and in the η's. And that means that they are universal polynomials in the coefficients of f and g.

Here is how this multiplication works out on the ghost components. As in (9.14)

$$\begin{aligned} t\frac{d}{dt} \log\Big(\prod_{i,j} (1 - \xi_i \eta_j t)^{-1} \Big) &= \sum_{i,j} ((\xi_i \eta_j)t + (\xi_i \eta_j)^2 t^2 + (\xi_i \eta_j)^3 t^3 + \cdots) \\ &= \sum_{n=1}^{\infty} p_n(\xi)\, p_n(\eta) t^n \end{aligned} \tag{9.17}$$

Thus, multiplication according to (9.16) translates into component-wise multiplication for the ghost components. Actually because of distributivity it would have been sufficient to do this calculation for (9.15).

Associativity of the multiplication and distributivity follow by functoriality (or directly for that matter).

All in all, what has been defined is a unital-commutative-ring-valued functor

$$\Lambda\colon \mathbf{CRing} \longrightarrow \mathbf{CRing} \tag{9.18}$$

together with a set of 'ghost component' functorial ring morphisms

$$s_n\colon \Lambda(A) \longrightarrow A \tag{9.19}$$

which determine the ring structure on $\Lambda(A)$ by the very requirement that they be functorial ring morphisms.

The unit element of the ring $\Lambda(A)$ is the power series $1 + t + t^2 + t^3 + \cdots = (1 - t)^{-1}$.

9.20. *The functor of the big Witt vectors.* Now what has all this to do with Witt-vector-like constructions? As a set define

$$W(A) = A^{\mathbf{N}} = \{(x_1, x_2, x_3, \cdots)\colon\ x_i \in A\} \tag{9.21}$$

and set up a functorial bijection (of sets) between $W(A)$ and $\Lambda(A)$ by

$$e_A\colon (x_1, x_2, x_3, \cdots) \mapsto \prod_{n=1}^{\infty}(1 - x_n t^n)^{-1} \tag{9.22}$$

and transfer the functorial ring structure on the $\Lambda(A)$ to the $W(A)$ by this bijection. This then defines a unital-commutative-ring-valued functor

$$W\colon \mathbf{CRing} \longrightarrow \mathbf{CRing}$$

(that is isomorphic to Λ, so one might justifiably wonder why one takes the trouble). Let's calculate the ghost components of a Witt vector $(x_1, x_2, x_3, \cdots) \in W(A)$, which by definition means calculating

$$w_n(x) = w_n(x_1, x_2, x_3, \cdots) = s_n(e_A(x)) \tag{9.23}$$

This is easy

$$\begin{aligned} t\frac{d}{dt}\log\Big(\prod_d (1 - x_d t^d)^{-1}\Big) &= t\frac{d}{dt}\sum_d -\Big(\sum_m -m^{-1}(x_d t^d)^m\Big) \\ &= \sum_{d,m} d x_d^m t^{dm} = \sum_{n=1}^{\infty}\sum_{d|n} d x_d^{n/d} \end{aligned} \tag{9.24}$$

and thus the ghost components of the ring valued functor W are given by the (generalized) Witt polynomials

$$w_n(X_1, X_2, \cdots, X_n) = \sum_{d|n} d X_d^{n/d} \tag{9.25}$$

Note that these do indeed generalize the Witt polynomials (5.1) of the previous 8 sections in that, apart from a relabeling the polynomials w_{p^n} of (9.25) are indeed the polynomials w_n of (5.1).

From now on in this whole chapter, that is in all sections that follow, w_n and $w_n(X)$ and $\cdots$ refer to the polynomials as defined by (9.25).

Incidentally, the formula for (second) multiplication of power series, i.e. multiplication in the ring $\Lambda(A)$ of power series written in 'x-coordinates' as in the right hand side of (9.22) can be written down directly as follows

$$\Big(\prod_d (1 - x_d t^d)^{-1}\Big) * \Big(\prod_n (1 - y_n t^n)^{-1}\Big) = \prod_{r,s\geq 1} (1 - x_r^{m/r} y_s^{m/s} t^m)^{-rs/m} \tag{9.27}$$

where on the right hand side m is the least common multiple of r and s. This is a formula that was known to Witt, [419]. Thus there is no absolute need to introduce symmetric functions into the game; it is just very convenient.

As far as I know the first substantial treatment of the big Witt vectors from the symmetric functions point of view is due to Pierre Cartier, [84, 86].

9.28. The a_i of (9.2) and (9.13), and the x_i of (9.22) can be best seen as simply being different ways to coordinatize $\Lambda(A)$. As is usual with two different coordinate systems each has its advantages and diadvantages. One good property of the 'x-coordinates' has already been pointed out. They show that the ring $\Lambda(A)$ is a (far reaching) generalization of the p-adic Witt vectors. Another is that the ghost component formulae, i.e. the Witt polynomials $w_n(X)$ are very simple. For one thing they do not involve mixed terms in the X_i; for another $w_n(X)$ depends only on those X_d for which d is a divisor of n. This latter fact makes it clear that there are many interesting quotient functors of W (which is difficult to see in the Λ formulation). These quotient functors will be described in the next section, 10.

On the other hand there are good well known formuals both ways (from symmetric function theory) that relate the 'a-coordinates' (i.e. the power series coordinates) with their ghost components while inverting the Witt polynomials seems to be a bit of a mess.

Quite a number of formulas from symmetric function theory are important (or at least very useful) for Witt vector theory. In the next subsections there are a few.

9.29. *Very partial symmetric function formularium* (1).[25] Everything takes place in the rings

$$\mathbf{Z}[\xi_i\colon i \in I] \quad \text{respectively} \quad \mathbf{Z}[[\xi_i\colon i \in I]]$$

of polynomials, respectively power series, in a countable infinite collection of commuting indeterminates over the integers. (And also in their analogues $\mathbf{Q}[\xi_i\colon i \in I]$ and $\mathbf{Q}[[\xi_i\colon i \in I]]$ over the rationals.) For some definitional and foundational remarks on what it means to have a polynomial or power series in an infinite number of variables see the appendix to this chapter.

The rings just above are graded by giving all the ξ_i degree 1.

The reader who does not like polynomials and power series in an infinite number of variables can just imagine that things are in terms of a finite number of them, large enough for the business at hand.

9.30. *Partitions and monomial symmetric functions.* A partition $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n)$ is a finite sequence of elements from $\mathbf{N} \cup \{0\}$ in non-increasing order, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_i \geq \cdots$. Partitions with different numbers of trailing zeros are identified. The length of a partition, $\lg(\lambda)$, is the number of its non zero entries; its weight is $\mathrm{wt}(\lambda) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. A different notation for a partition is $(1^{m_1(\lambda)}2^{m_2(\lambda)} \cdots n^{m_n(\lambda)})$ where $m_i(\lambda)$ is the number of entries of it that are equal to i. A partition of weight n is said to be a partition of n and its entries are (also) called the parts of that partition.

For any sequence $\alpha = (a_1, a_2, \cdots, a_m)$ of elements from $\mathbf{N} \cup \{0\}$ let

$$\xi^\alpha = \xi_1^{a_1}\xi_2^{a_2}\cdots\xi_m^{a_m}, \text{ e.g. } \xi^{(0,0,0,2,0,3,0,0)} = \xi_4^2\xi_6^3 \tag{9.31}$$

[25] Not all of the formulas and functions that follow belong to the standard formularium of the symmetric functions.

Given a partition λ (strictly speaking with an infinite string of trailing zeros added), the associated monomial symmetric function is the sum

$$m_\lambda = \sum_\alpha \xi^\alpha \tag{9.32}$$

where the sum is over all *distinct* sequences α (infinite with only a finite number of nonzero entries) that are permutations of λ (with trailing zeros added). Thus, for example,

$$m_{(2,1)} = \sum_{i\neq j} \xi_i^2 \xi_j = \xi_1^2\xi_2 + \xi_2^2\xi_1 + \xi_1^2\xi_3 + \xi_3^2\xi_1 + \xi_2^2\xi_3 + \xi_3^2\xi_2 + \cdots \tag{9.33}$$

$$m_{(1,1)} = \sum_{i<j} \xi_i\xi_j = \xi_1\xi_2 + \xi_1\xi_3 + \xi_2\xi_3 + \cdots \tag{9.34}$$[26]

These are the simplest symmetric functions in that if a symmetric function contains one of the monomials from an m_λ then it also contains all others (with the same coefficient). The monomial symmetric function m_λ is homogeneous of degree $\mathrm{wt}(\lambda) = n$ and they form a basis for the free Abelian group of symmetric functions of weight n.
By definition the complete symmetric functions and the elementary symmetric functions are

$$h_n = \sum_{\mathrm{wt}(\lambda)=n} m_\lambda \quad \text{and} \quad e_n = m_{(1,1,\cdots,1)}\ (n1\text{'s}) \tag{9.35}$$

and for a partition $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_m)$ one writes

$$h_\lambda = h_{\lambda_1}h_{\lambda_2}\cdots h_{\lambda_m} \quad \text{and} \quad e_\lambda = e_{\lambda_1}e_{\lambda_2}\cdots e_{\lambda_m} \tag{9.36}$$[27]

The complete symmetric functions and the elementary ones are related by the Wronski relations

$$\sum_{i+j=n} (-1)^i h_i e_j = 0 \quad \text{for } n \geq 1, \quad \text{where} \quad h_0 = e_0 = 1 \tag{9.37}$$

The h_λ (resp. e_λ) for λ of weight n also form a basis for the symmetric functions of degree n. By definition

$$\mathbf{Symm} = \oplus_n \mathbf{Symm}_n \subset \mathbf{Z}[\xi] \tag{9.38}$$

is the graded ring of symmetric functions in the commuting indeterminates ξ. In this whole chapter it will (usually) be seen as the graded ring in the complete

[26] Note the difference between the two formulae which rests on the word 'distinct' in the sentence just below (9.32).

[27] Note that this notation has a rather different meaning than that for m_λ. It is definitely not true that m_λ is equal to something like $m_{(\lambda_1)}m_{(\lambda_2)}\ldots m_{(\lambda_m)}$.

symmetric functions

$$\mathbf{Symm} = \mathbf{Z}[h_1, h_2, \cdots, h_m, \cdots] = \mathbf{Z}[h] \tag{9.39}$$

For many purposes one could equally well work with the elementary symmetric functions, but the complete ones just work out better. Actually, there are precisely four 'canonical' choices, see subsection 10.18 below on the Liulevicius theorem.

9.40. *Inner product.* One defines a (symmetric positive definite) inner product on **Symm** by declaring the monomial and the complete symmetric functions to form a bi-orthogonal system:

$$< h_\lambda, m_\kappa > = \delta_{\lambda,\kappa} \text{ (Kronecker delta)} \tag{9.41}$$[28]

There is now a remarkable theorem, see [281], (4.6), p. 63:

9.42. *Theorem.* Let $\{u_\lambda\}$ and $\{v_\lambda\}$ be two sets of symmetric functions indexed by all partitions (including the zero partition) that both form a basis for $\mathbf{Symm}_\mathbf{Q} = \mathbf{Symm} \otimes_\mathbf{Z} \mathbf{Q}$ (or **Symm** itself). Then they form a bi-orthonormal system (i.e $< u_\lambda, v_\lambda >= \delta_{\lambda,\kappa}$) if and only if

$$\sum_\lambda u_\lambda(\xi) v_\lambda(\eta) = \prod_{i,j} (1 - \xi_i \eta_j)^{-1} \tag{9.43}$$

Here $\eta = \{\eta_1, \eta_2, \cdots\}$ is a second set of commuting indeterminates that also commute with the ξ's, and $v_\lambda(\eta)$ is the same as $v_\lambda(\xi) = v_\lambda$ but now written as an expression in the η's.

The theorem is especially remarkable to a mathematician with (products of) Witt vectors on his mind[29] and I am far from sure that its consequences in that context have been fully explored and understood.

Theorem 9.42 fits with formula (9.41) for it is indeed the case that ([281], p. 62)

$$\sum_\lambda h_\lambda(\xi) m_\lambda(\eta) = \prod_{i,j} (1 - \xi_i \eta_j)^{-1} \tag{9.44}$$

9.45. *Schur functions.* A symmetric function, i.e. an element of **Symm**, is called positive if when expressed as a sum in the monomial basis $\{m_\lambda\}$ all its coefficient are non negative. All the specific symmetric functions introduced so far, the m_λ, h_λ, e_λ, are positive.

It now turns out that there is a unique positive orthonormal basis of **Symm**. They are called the Schur functions, and are determined by

$$\sum_\lambda s_\lambda(\xi) s_\lambda(\eta) = \prod_{i,j} (1 - \xi_i \eta_j)^{-1} \tag{9.46}$$

and the positivity requirement.

---

[28] This inner product is often called 'Hall inner product'.

[29] See (9.16).

The first few Schur functions are

$$s_{(1)} = m_{(1)} = h_1, \quad s_{(2)} = m_{(2)} + m_{(1,1)} = h_2, \quad s_{(1,1)} = m_{(1,1)} = -h_2 + h_{(1,1)} \tag{9.47}$$

and the ones in weights 3 and 4 are given by the matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix}, \text{ s-m matrix; } \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}, \text{ s-h matrix} \tag{9.48}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 \\ 1 & 3 & 2 & 3 & 1 \end{pmatrix}, \text{ s-m matrix; } \begin{pmatrix} 1 & -1 & 0 & 1 & -1 \\ 0 & 1 & -1 & -1 & 2 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \text{ s-h matrix} \tag{9.49}$$

where the partitions have been ordered lexicographically, (4), (3,1), (2,2), (2,1,1), (1,1,1,1), and the columns give the coefficients of the Schur functions in terms of the monomial symmetric functions (resp. the complete symmetric functions). Thus e.g.

$$\begin{aligned} s_{(3,1)} &= m_{(3,1)} + m_{(2,2)} + 2m_{(2,1,1)} + 3m_{(1,1,1,1)} \\ s_{(2,1,1)} &= h_{(4)} - h_{(3,1)} - h_{(2,2)} + h_{(2,1,1)} \end{aligned} \tag{9.50}$$[30]

There are many ways to define the Schur functions and to write down formulas for them. The standard (and oldest) definition is the one by Jacobi, see [281], p. 40ff. An explicit expression is given by the Jacobi–Trudy formulas

$$s_\lambda = \det(h_i - i + j)_{1 \le i,j \le n} \tag{9.51}$$

Other expressions come from the Schur identity

$$\sum_\lambda s_\lambda = \prod_i (1 - \xi_i)^{-1} \prod_{i<j} (1 - \xi_i \xi_j)^{-1} \tag{9.52}$$

(see [273], 5.4, p. 176). This one has the advantage of showing immediately that Schur symmetric functions are positive (which is far from obvious from (9.51)).

[30] Calculating with the monomial symmetric functions directly is a bit messy (and so is calculating with their explicit expressions). For instance $m_{(3,1)}m_{(1)} = m_{(4,1)} + m_{(3,2)} + 2m_{(3,1,1)}$, $m_{(4,1)}m_{(2)} = m_{(6,1)} + m_{(4,3)} + m_{(4,2,1)}$, $m_{(2,1)}m_{(1)} = 2m_{(2,2)} + m_{(3,1)} + 2m_{(2,1,1)}$. It is fairly easy to see what monomial functions should occur in such a product; things are less clear as regards the coefficients with which they occur. For instance, why is there a 2 in front of $m_{(2,2)}$ in the third formula but not one in front of $m_{(3,2)}$ in the first. Things become a good deal better (in my opinion) when one works in the larger ring **QSymm** $\supset$ **Symm** of quasi-symmetric functions where there is a clear easy to use formula for multiplying quasi-symmetric monomial functions. This is the overlapping shuffle product, see 11.26. See also the appendix to this chapter for some more details.

9.53. *Forgotten symmetric functions.* The complete symmetric functions $\{h_\lambda\}$ form a basis dual to that of the monomial symmetric functions. There also is of course a dual basis to the basis of elementary symmetric functions $\{e_\lambda\}$. These are obtained by replacing the h_i by the e_i in the formulas for the m_λ in the h_i. There seem to be no 'nice' formulas for them, [281], p. 22.

9.54. *Power sum symmetric functions.* The power sum symmetric functions are defined by

$$p_n = m_{(n)} = \sum_i \xi_i^n \tag{9.55}$$

Recall that they are the ghost components of the element

$$H(t) = 1 + h_1 t + h_2 t^2 + h_3 t^3 + \cdots \in \Lambda(\mathbf{Symm}) = 1 + t\mathbf{Symm}[[t]] \tag{9.56}$$

The power sum symmetric functions are related to the complete symmetric functions by the Newton relations

$$nh_n = p_n + h_1 p_{n-1} + h_2 p_{n-2} + \cdots + h_{n-1} p_1 \tag{9.57}$$

which in terms of the generating functions $H(t)$, see (9.56), and

$$P(t) = \sum_{i \geq 1} p_i t^i \quad \text{(so that } p_0 \text{ is set to zero)} \tag{9.58}$$

can be encoded by the 'differental' equation (defining ghost components)

$$P(t) = t \frac{d}{dt} \log H(t) = \frac{tH'(t)}{H(t)} \tag{9.59}$$

From (9.59) one readily obtains a formula for the power sum symmetric functions in terms of the complete symmetric functions, i.e. a universal formula for the ghost components of an element $1 + a_1 t + a_2 t^2 + \cdots$ in terms of its 'a-coordinates'.

$$p_n = \sum_{r_1 + r_2 + \cdots + r_k = n} (-1)^{k+1} r_1 h_{r_1} h_{r_2} \cdots h_{r_k}; \quad r_i \in \mathbf{N} = \{1, 2, \cdots\} \tag{9.60}$$

This formula can readily be inversted by 'solving' (9.59) using formal exponentials.

For each partition λ (with no zero parts) define also

$$p_\lambda = p_{\lambda_1} p_{\lambda_2} \cdots p_{\lambda_n} \tag{9.61}$$

Then, obviously, given the Newton relations, the p_λ form a homogeneous basis over the rationals for the vector space $\mathbf{Symm}_\mathbf{Q}$. But they are not a basis for **Symm** itself. For a partition $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n)$ define the integer z_λ by

$$z_\lambda = 1^{m_1(\lambda)} m_1(\lambda)! 2^{m_2(\lambda)} m_2(\lambda)! \cdots n^{m_n(\lambda)} m_n(\lambda)!$$

where, as before, $m_i(\lambda)$ is the number of parts i in λ. Then ([281], p. 62)

$$\sum_{\lambda} z_{\lambda}^{-1} p_{\lambda}(\xi) p_{\lambda}(\eta) = \prod_{i,j} (1 - \xi_i \eta_j)^{-1} \tag{9.62}$$

showing that suitably normalized p_λ form an orthonormal basis for $\mathbf{Symm}_{\mathbf{Q}}$ (but not of course for **Symm** as these normalized p_λ are not even elements of **Symm**).

9.63. *On the Witt coordinates.* Now consider the transformation from 'a-coordinates' of a 1-power-series, i.e. an element of $\Lambda(A)$, to Witt vector coordinates. This is given by the universal formula, see (9.22)

$$\prod_{d} (1 - x_d t^d)^{-1} = 1 + h_1 t + h_2 t^2 + \cdots \tag{9.64}$$

Assuredly the x_d as defined by (9.64) are symmetric functions and also, obviously, from (9.64) one has

$$x_d = h_d + (\text{homogenous polynomial of weight } d \text{ in the } h_1, \cdots, h_{d-1}) \tag{9.65}$$

and so the

$$x_\lambda = x_{\lambda_1} x_{\lambda_2} x_{\lambda_3} \cdots \tag{9.66}$$

form yet another homogenous basis for **Symm**. It is quite easy to express the h_n in terms of the x_d. Indeed rewriting the left hand side of (9.61) as

$$\prod_{i} (1 + x_i t^i + x_i^2 t^{2i} + \cdots) \tag{9.67}$$

it is immediate that

$$h_n = \sum_{\mathrm{wt}(\lambda)=n} x_\lambda = \sum_{\mathrm{wt}(\lambda)=n} x_1^{m_1(\lambda)} x_2^{m_2(\lambda)} x_3^{m_3(\lambda)} \cdots \tag{9.68}$$

However, finding a formula the other way seems to be rather messy.[31] On the other hand, there is a duality formula. For each symmetric function $f(\xi)$ introduce the shorthand notation

$$f(\xi^i) = f(\xi_1^i, \xi_2^i, \xi_3^i, \cdots) \tag{9.69}$$[32]

Now for any partition $\lambda = (1^{k_1} 2^{k_2} \cdots n^{k_n})$ introduce

$$r_\lambda = h_{k_1}(\xi^1) h_{k_2}(\xi^2) \cdots h_{k_n}(\xi^n)$$

[31] This corresponds to the fact that inverting the Witt vector polynomials appears to be a messy business.

[32] This is the (outer) plethysm with perspect to the power sum symmetric function p_i; see subsections 16.76, 18.35 and 18.37 below.

It is not difficult to check that these also form a basis of **Symm**.[33] There is now the following duality formula

$$<x_\lambda, r_\kappa> = \delta_{\lambda,\kappa} \tag{9.70}$$

i.e. the x_λ form a dual or adjoint basis to the one of the r_λ. This is due to [350], where there is a quite elegant proof using theorem 9.42.

The symmetric functions x_n also have an interesting, even remarkable, positivity property:

9.71. *Theorem.* The symmetric functions $-x_1, x_2, \cdots, x_n, \cdots$ are Schur positive.[34] Here 'Schur positive' means that when expressed as a **Z**-linear combination of the Schur symmetric functions all the coefficients are nonnegative. This is stronger than being positive of course. Theorem 9.71 was conjectured by Christoph Reutenauer, [333, 332], and proved by Thomas Scharf and Jean-Yves Thibon, [350], and also by W Doran. Some more investigations relating to the x_n are reported in loc. cit. Much remains to be done in my opinion, particularly regarding exploiting the positivity result of theorem 9.71.[35]

9.72. *Integrality lemmata. The miracle of the Witt polynomials* (2). For the generalized Witt polynomials $w_n(X)$ there is the same integrality theorem as for the p-adic ones.

9.73. *Theorem.* Let $\varphi(X, Y, Z)$ be a polynomial over the integers in three (or less, or more) commuting indeterminates then there are unique polynomials over the integers $\varphi_n(X_1, X_2, \cdots, X_n; Y_1, Y_2, \cdots, Y_n; Z_1, Z_2, \cdots, Z_n) = \varphi_n(X; Y; Z), n = 1, 2, 3, \cdots$ such that for all $n \geq 1$

$$w_n(\varphi_1(X; Y; Z), \cdots, \varphi_n(X; Y; Z)) = \varphi(w_n(X), w_n(Y), w_n(Z)) \tag{9.74}$$

The proof is (basically) the same as in the p-adic case.

The theory of the functor of the big Witt vectors can now be developed starting from this integrality theorem in just the same way as was done for the p-adic Witt vectors in section 5 above. That is define addition polynomials $\mu_{S,i}(X_1, \cdots, X_i; Y_1, \cdots, Y_i)$ and multiplication polynomials $\mu_{P,i}(X_1, \cdots, X_i; Y_1, \cdots, Y_i)$ by taking the polynomial $\varphi(X, Y)$ in theorem 9.73 respectively equal to $X + Y$ and XY, so that

$$\begin{aligned} w_n(\mu^W_{S,1}(X; Y), \mu^W_{S,n}(X; Y), \cdots, \mu^W_{S,n}(X; Y)) &= w_n(X) + w_n(Y) \\ w_n(\mu^W_{P,1}(X; Y), \mu^W_{P,2}(X; Y), \cdots, \mu^W_{s,n}(X; Y)) &= w_n(X) w_n(Y) \end{aligned} \tag{9.75}$$

[33] This one has a certain amount of special interest in that for any symmetric function f the integrality conditions $< f, r_\lambda > \in \mathbf{Z}$ are equivalent to certain congruences established by Andreas Dress, [118], for testing whether a central function of a symmetric group S_n is a virtual character; see Thomas Scharf and Jean-Yves Thibon, [350].

[34] There is no printing error here; there really is just one minus sign.

[35] Quite generally the systematic use of (partial) orderings and positive properties in Hopf algebra theory (and hence Witt vector theory) has started only fairly recently.

and define an addition and multiplication on Witt vectors $a = (a_1, a_2, a_3, \cdots)$ by

$$(a_1, a_2, a_3, \cdots) +_W (b_1, b_2, b_3, \cdots) = \left(\mu^W_{s,1}(a; b), \mu^W_{s,2}(a; b), \mu^W_{s,3}(a; b), \cdots\right)$$

$$(a_1, a_2, a_3, \cdots) \cdot_W (b_1, b_2, b_3, \cdots) = \left(\mu^W_{P,1}(a; b), \mu^W_{P,2}(a; b), \mu^W_{P,3}(a; b), \cdots\right) \tag{9.76}$$

This then defines the functorial ring $W(A)$ with multiplication and addition given by (9.76) with zero element $(0, 0, 0, \cdots)$ and unit element $(1, 0, 0, 0, \cdots)$.

9.77. Theorem 9.73 is only one of a slew of integrality theorems. One of my favourites is the 'functional equation lemma' which is of great use in formal group theory; particularly in the construction of universal formal groups, see [192], Chapter 1, §2.3.

9.78. *Ingredients for the functional equation integrality lemma.* The ingredients for the functional equation lemma are the following

$$A \subset K, \sigma\colon K \longrightarrow K,\ \mathfrak{p} \subset A,\ p, q, s_1, s_2, s_3, \cdots \tag{9.79}$$

Here A is a subring of a ring K, σ is a ring endomorphism of K, $\mathfrak{a}$ is an ideal in A, p is a prime number and q is a power of p, and the s_i are elements of K. These ingredients are supposed to satisfy the following conditions

$$\sigma(A) \subset A, \sigma(a) \equiv a^q \mod \mathfrak{p} \text{ for all } a \in A,\ p \in \mathfrak{p}, s_i b \in A \text{ for all } b \in \mathfrak{p} \tag{9.80}$$

and also

$$\mathfrak{p}^r b \subset \mathfrak{p} \Rightarrow \mathfrak{p}^r \sigma(b) \subset \mathfrak{p} \text{ for all } r \in \mathbf{N}, b \in K \tag{9.81}$$

a property that is automatically satisfied if the ideal $\mathfrak{p} = (c)$ is principal and $\sigma(c) = uc$ for some unit of A.

Here are three examples of such situations. There are many more.

$$A = \mathbf{Z}, K = \mathbf{Q}, \sigma = \mathrm{id}, q = p, \mathfrak{p} = p\mathbf{Z}, s_i \in p^{-1}\mathbf{Z} \tag{9.82}$$

$$A = \mathbf{Z}_{(p)}, K = \mathbf{Q}, \sigma = \mathrm{id}, q = p, \mathfrak{p} = p\mathbf{Z}_{(p)}, s_i \in p^{-1}\mathbf{Z}_{(p)} \tag{9.83}$$

$$A = \mathbf{Z}[V_1, V_2, \cdots], K = \mathbf{Q}[V_1, V_2, \cdots], \sigma f(V_1, V_2, \cdots) = f(V_1^p, V_2^p, \cdots),$$
$$q = p, \mathfrak{p} = pA, s_i \in p^{-1}A \tag{9.84}$$

9.82. *Constructions for the functional equation lemma.* Now, let $g(X) = b_1 X + b_2 X^2 + \cdots$ be a power series with coefficients in A. Using the ingredients (9.78) construct from it a new power series by the recursion formula (or functional equation)

$$f_g(X) = g(X) + \sum_{i=1}^{\infty} s_i \sigma^i_* f_g(X^{q^i}) \tag{9.85}$$

where $\sigma^i_* f_g(X)$ is the power series obtained from $f_g(X)$ by applying σ^i to the coefficients of $f_g(X)$. Three examples of power series obtained in this way are:

$$X + p^{-1}X^p + p^{-2}X^{p^2} + p^{-3}X^{p^3} + \cdots \tag{9.86}$$

$$\log(1+X) = \sum_{n=1}^{\infty} n^{-1}(-1)^{n+1}X^n \tag{9.87}$$

$$\begin{aligned} &X + p^{-1}V_1X^p + (p^{-2}V_1V_1^p + p^{-1}V_2)X^{p^2} \\ &+ (p^{-3}V_1V_1^pV_1^{p^2} + p^{-2}V_1V_2^p + p^{-2}V_2V_1^{p^2} + p^{-1}V_3)X^{p^3} + \cdots \end{aligned} \tag{9.88}$$

9.89. *Functional equation integrality.*

Lemma. Let $A \subset K, \sigma, \mathfrak{p} \subset A$, $p, q, s_1, s_2, s_3, \cdots$ be as in (9.77) and be such that (9.80) and (9.81) hold. Let $g(X) = b_1X + b_2X^2 + \cdots$ and $\overline{g}(X) = \overline{b}_1X + \overline{b}_2X^2 + \cdots$ be two power series with coefficients in A and let b_1 be invertible. Let this time $f^{-1}(X)$ denote the functional inverse of $f(X)$ (for a power series with zero constant term and nonzero coefficient in degree 1, defined by $f^{-1}(f(X)) = X$. Then

$$\text{The power series } F_g(X,Y) = f_g^{-1}(f_g(X) + f_g(Y)) \text{ has its coefficients in } A \tag{9.90}$$

$$\text{The power series } f_g^{-1}(f_{\overline{g}}(X)) \text{ has its coefficients in } A \tag{9.91}$$

If $h(X)$ is a power series with zero constant term over K, then

$$h(X) \equiv \overline{g}(X) \bmod \mathfrak{p}^r A[[X]] \Leftrightarrow f_g(h(X)) \equiv f_g(\overline{g}(X)) \bmod \mathfrak{p}^r A[[X]] \tag{9.92}$$

Statement (9.90) says that $F_g(X, Y)$ is a (one dimensional) formal group over A with logarithm $f_g(X)$. Example (9.88) is the logarithm of the universal one dimensional p-typical formal group.

There are also more dimensional and infinite dimensional versions and these can be used to construct universal formal groups and also the Witt vectors. For all these statements and the appropriate definitions see [192].

Another application of the functional equation lemma is the following statement, especially useful in the Witt vector context.

9.93. *Lemma.* [192], lemma 17.6.1, page 137; ghost component integrality lemma. Let A be a characteristic zero ring with endomorphisms φ_p for all prime numbers p such that $\varphi_p(a) \equiv a^p$ for all $a \in A$. Then for a given sequence $b_1, b_2, \cdots$ in A there exists another sequence $x_1, x_2, \cdots$ in A such that $w_n(x) = b_n$ for all n (i.e. the sequence $b_1, b_2, \cdots$ lies in the image of the ghost mapping) if and only if

$$\varphi_p(b_n) \equiv b_{np} \bmod (p^{v_p(n)+1}) \tag{9.94}$$

Here, as before, see example 6.10, v_p denotes the p-adic valuation on $\mathbf{Z}$. A vector with components in A which is in the image of $w\colon W(A) \longrightarrow A^{\mathbf{N}}$, $(a_1, a_2, \cdots) \mapsto (w_1(a), w_2(a), \cdots)$ will be called a ghost-Witt vector over A.

As far as I know, this lemma, in this form, first appeared in [86]. Pierre Cartier attributes it to Bernard Dwork and Jean Dieudonné, [107], proposition 1; [132], lemma 1. The formulations there are a bit different.

In e.g. the case $A = \mathbf{Z}$ where one takes all the endomorphisms φ_p to be the identity, the integrality condition (9.94) can be given the following quite elegant formulation. Write down the Mobius inversion transform of the sequence $b_1, b_2, \cdots$, i.e. the sequence

$$c_n = \sum_{d|n} b_{n/d}\mu(d) \tag{9.95}$$

Then the sequence $b_1, b_2, \cdots$ satisfies (9.94) if and only if $c_n \equiv 0 \bmod n$ for all n. This remark seems to be due to Albrecht Dold, [113]. Here the Möbius function $\mu\colon \mathbf{N} \longrightarrow \mathbf{Z}$ is defined by $\mu(1) = 1$, $\mu(n) = (-1)^r$ if n is the product of precisely r different primes and $\mu(n) = 0$ otherwise. Still another formulation of the integrality condition is

$$\sum_{i=1}^{n} b_{(i,n)} \equiv 0 \bmod n \tag{9.96}$$

where (i, n) denotes the greatest common divisor of i and n.

The (integrality aspects of the) theory of Witt vectors can be developed solely on the basis of this lemma 9.93. This is how things are done in [212].

9.97. *Witt vectors over the integers and fixed points and fixed point indices of iterated mappings.* In [113], Dold proves that a sequence of integers $s = (s_1, s_2, s_3, \cdots)$ is the sequence of fixed point indices of the iterates of a continuous mapping if and only if the formal power series

$$\exp\left(-\sum_{i=1}^{\infty} \frac{s_i}{i} t^i\right)$$

has integral coefficients; that is iff s is the ghost vector of a Witt vector over the integers. This is done using the integrality criterium (9.96).

More concretely, call a sequence of nonnegative integers exactly realizable if there is a set X and a map $f\colon X \longrightarrow X$ such that the number of fixed points of the n-th iterate f^n of f is exactly a_n. Then a sequence is exactly realizable if and only if the numbers

$$\sum_{d|n} \mu(d) a_{n/d}$$

are nonnegative and divisible by n, [330]. Similar things happen when studying iterates of mappings of the unit interval into itself, i.e. in the part of dynamical system and chaos theory that deals with such mappings, see [70, 292].

The first extensive treatment that I know of the Witt vector constructions based basically totally on the symmetric function point of view is by Pierre Cartier, [86].

10. The Hopf algebra Symm as the representing algebra for the big Witt vectors

Recall that $\mathbf{Symm} = \mathbf{Z}[h_1, h_2, h_3, \cdots] = \mathbf{Z}[h]$ represents the functor Λ and hence the functor W of the big Witt vectors

$$\Lambda(A) = \{a(t) = 1 + a_1 t + a_2 t^2 + a_3 t^3 + \cdots : a_i \in A\} = \mathbf{CRing}(\mathbf{Z}[h], A) \tag{10.1}$$

The fact that this is an Abelian group valued functor means that there is a comultiplication making **Symm** a coalgebra object in the category **CRing** of commutative unital rings, which in turn means that there is a comultiplication, and a co-unit

$$\mu_s\colon \mathbf{Symm} \longrightarrow \mathbf{Symm} \otimes \mathbf{Symm}, \quad \varepsilon_s\colon \mathbf{Symm} \longrightarrow \mathbf{Z}, \tag{10.2}$$

that are ring morphisms and that make $(\mathbf{Symm}, \mu_s, \varepsilon_s)$[36] a co-associative, cocommutative co-unital coalgebra. See [293] for a lot of material on coalgebras.

Because the addition on the set $\Lambda(A)$ is defined as multiplication of power series and the unit element is the power series 1, the morphisms μ_s, ε_s are given by

$$\mu_s(h_n) = 1 \otimes h_n + \sum_{i=1}^{n-1} h_i \otimes h_{n-i} + h_n \otimes 1, \varepsilon_s(h_n) = 0, \quad n \geq 1 \tag{10.3}$$

(It is often convenient to define $h_0 = 1$ so that the multiplication can be written $\sum_{i+j=n} h_i \otimes h_j$.)

Writing

$$m\colon \mathbf{Symm} \otimes \mathbf{Symm} \longrightarrow \mathbf{Symm}, \ e : \mathbf{Z} \longrightarrow \mathbf{Symm} \tag{10.4}$$

for the multiplication and unit element of **Symm**, in total there is a bialgebra structure $(\mathbf{Symm}, m, e, \mu_s, \varepsilon_s)$. Finally there is an antipode

$$\iota_s : \mathbf{Symm} \longrightarrow \mathbf{Symm} \tag{10.5}$$

The existence of an antipode comes for free in the present case, see below. The antipode is an anti ring morphism (and hence in the present case a ring morphism because of commutativity). The antipode is determined by

$$\iota_s(h_n) = -h_n \tag{10.6}$$

All this makes the total structure $(\mathbf{Symm}, m, e, \mu_s, \varepsilon_s, \iota_s)$ a Hopf algebra over the integers. For the basic theory of Hopf algebras, see [95].

---

[36] The index 'S' here refers to the matter that these are the ring morphisms that give the sum part of the Witt vector functor.

10.7. *Skew Schur functions.* (Very partial symmetric function formularium (2)). Let κ, λ be two partitions and let s_κ, s_λ be the corresponding Schur functions. As the Schur functions form an orthogonal basis, there are coefficients $c^v_{\kappa,\lambda}$ such that

$$s_\kappa s_\lambda = \sum_v c^v_{\kappa,\lambda} s_v \quad \text{or, equivalently, } \langle s_\kappa s_\lambda, s_v\rangle = c^v_{\kappa,\lambda} \tag{10.8}$$

The multiplicity coefficients[37] $c^v_{\kappa,\lambda}$ have a combinatorial interpretation, see [281], Chapter 1, §9, p. 143, and are, hence, nonnegative integers. Define the skew Schur function $s_{\kappa/\lambda}$ by

$$\langle s_{\kappa/\lambda}, s_v\rangle = \langle s_\kappa, s_\lambda s_v\rangle, \text{ or, equivalently, } s_{\kappa/\lambda} = \sum_v c^\kappa_{\lambda,v} s_v \tag{10.9}$$

It turns out that $s_{\kappa/\lambda} = 0$ unless $\lambda \subset \kappa$ which by definition means that $\lambda_i \le \kappa_i$ for all i.[38] There is a determinantal formula for the skew Schur functions as follows

$$s_{\kappa/\lambda} = \det(h_{\kappa_i - \lambda_j - i + j})_{i,j} \tag{10.10}$$

In terms of skew Schur functions the comultiplication of **Symm** can be written

$$\mu_S(s_\kappa) = \sum_\lambda s_{\kappa/\lambda} \otimes s_\lambda \tag{10.11}$$

as follows from the duality formula $\langle xy, z\rangle = \langle x \otimes y, \mu_s(z)\rangle$ in **Symm**.

10.12. *Grading.* **Symm** is a graded Hopf algebra. That is, as an Abelian group it is a direct sum

$$\mathbf{Symm} = \bigoplus_{n=0}^{\infty} \mathbf{Symm}_n \tag{10.13}$$

where the homogenous part of degree (or weight) n is spanned by the monomials in the h_i of weight n where h_i has weight i; i.e. by the monomials h_λ with $\mathrm{wt}(\lambda) = n$, and that moreover the multiplication and comultiplication and unit and co-unit are compatible with the grading, which means

$$m(\mathbf{Symm}_i \otimes \mathbf{Symm}_j) \subset \mathbf{Symm}_{i+j},\ e(\mathbf{Z}) \subset \mathbf{Symm}_0$$

$$\mu_s(\mathbf{Symm}_n) \subset \bigoplus_{i+j=n} \mathbf{Symm}_i \otimes \mathbf{Symm}_j,\ \varepsilon(\mathbf{Symm}_n) = 0 \quad \text{for } n \ge 1$$

$$\iota_s(\mathbf{Symm}_n) \subset \mathbf{Symm}_n$$

Note that $\mathbf{Symm}_0 = \mathbf{Z}$ and from the point of view of this component e and ε identify with the identity on this component.

[37] In the representation theoretic incarnation of **Symm**, see section 18 below, these coefficients turn up as multiplicities of irreducible representations (Littlewood–Richardson rule).

[38] $\kappa \supset \lambda$ is the case if and only if the diagram of κ contains the diagram of λ; whence the notation.

A graded Hopf algebra is connected if its degree 0 component is of rank 1. It is easy to see that for a connected graded bialgebra there is a unique antipode making it a Hopf algebra.

10.14. *Antipodes on connected graded bialgebras.* Indeed, as $H_0 = \mathbf{Z}$ (or whatever ring once is working over), an antipode must be the identity on H_0. Further, if x is primitive, i.e. $\mu_H(x) = 1 \otimes x + x \otimes 1$, the defining requirements for an antipode say that $\iota(x) = -x$. Further, if i is the smallest integer ≥ 1 such that $H_i \neq 0$ all elements of H_i must be primitive by degree considerations (and the counit properties). Finally if the antipode is known on all the H_j with $j < n$ and $x \in H_n$, $\mathrm{id} \otimes \iota$ and $\iota \otimes \mathrm{id}$ are known on all the terms of $\mu_H(x)$ except $1 \otimes x$ and $x \otimes 1$. The defining requirements for an antipode then immediately give a formula for $\iota(x)$ (two formulas really). They must give the same result because anitpodes (if they exist) are unique (by the same argument that is used to show that inverses in groups are unique).

10.15. *Primitives of* **Symm**. For every Hopf algebra it is important to know its primitives. That is the elements x that satisfy

$$\mu(x) = 1 \otimes x + x \otimes 1. \tag{10.16}$$

These form a module over the ring over which the Hopf algebra is defined (in this case $\mathbf{Z}$) and have a Lie algebra structure under the commutator product $[x, y] = xy - yx$. Because of commutativity this Lie product is zero in the present case.

By definition, $p_1 = h_1$ is a primitive. With induction, from the Newton relations (9.57) it follows immediately that all the p_n are primitives. It is not difficult to see that these form a basis over $\mathbf{Z}$ for the module of primitives.[39]

So the comultiplication and counit of **Symm** satisfy

$$\mu_s(p_n) = 1 \otimes p_n + p_n \otimes 1, \varepsilon_s(p_n) = 0 \tag{10.17}$$

and as the power sum symmetric functions form a free polynomial basis (over $\mathbf{Q}$) for the symmetric functions this is a perfectly good description of the Hopf algebra **Symm** (once it is known that it is in fact defined over the integers). Formulas (10.17) of course amount to completely the same thing as saying that at the ghost component level addition is coordinate-wise.

10.18. *Liulevicius theorem,* [269, 268]. There are precisely four graded Hopf algebra automorphisms of **Symm**. They are functorially given by

$$\text{identity}, a(t) \mapsto a(-t), a(t) \mapsto a(t)^{-1}, a(t) \mapsto a(-t)^{-1} \tag{10.19}$$

and they form the Klein 4-group.

---

[39] First observe that if an element is primitive its different homogeneous summands must all be primitives. next observe that a term in the coproduct of a monomial of length n can only cancel against one from another term of the same length; continue by using lexicographic order on exponent sequences. Finally, as $p_n \equiv (-1)^{n+1} h_1^n \mod (h_2, h_3, \cdots)$ (also from the Newton relations with induction), the p_n are not divisible by any natural number other than 1 and as the different p_n are of different degree no sum of them is divisible by any natural number other than 1.

The proof is quite simple and it is surprising that this theorem was discovered so late.[40] Over the rationals of course the situation is quite different and the group of automorphisms is quite large viz a countable sum of copies of the multiplicative group of the rationals.

10.20. Here is the proof. The power sum symmetric functions are primitives for the Hopf algebra structure meaning that

$$\mu_s(p_i) = 1 \otimes p_i + p_i \otimes 1 \tag{10.21}$$

Moreover every homogeneous primitive of weight i is a scalar multiple of p_i. Let φ be a graded automorphism. Automorphisms of Hopf algebras must take primitives into primitives and as a graded automorphism preserves the grading it must be the case that

$$\varphi(p_i) = ap_i$$

with $a \in \{1, -1\}$ because φ is an invertible morphism. The last three automorphisms named in (10.19) take on p_1, p_2 respectively the values

$$-p_1, p_2; -p_1, -p_2; p_1, -p_2$$

and so composing with a suitable one from the four automorphisms (10.19) it remains to analyze the case that

$$\varphi(p_1) = p_1, \varphi(p_2) = p_2$$

Suppose with induction that it has been proved that $\varphi(p_i) = p_i, i < n \geq 3$. Look at the Newton relations (see (9.57))

$$nh_n = p_n + h_1 p_{n-1} + \cdots + h_{n-1} p_1, \tag{10.22}$$

Because the h_i are polynomials in the $p_1, p_2, \cdots, p_i$ (albeit with rational coefficients), $\varphi(h_i) = h_i, i < n$. So applying φ to (10.16) and subtracting the result from (10.16) one finds

$$n(h_n - \varphi(h_n)) = p_n \pm p_n$$

which because $n \geq 3$ and $\varphi(h_n)$ must be an integral polynomial of weight n, and p_n is not divisible by any integer > 1, is only possible if on the right hand side of (10.13) the minus sign applies; i.e. only if $\varphi(p_n) = p_n$.

10.23. *Product comultiplication.* Besides the Abelian group structure on $\Lambda(A)$ and $W(A)$ there is also the product of Witt vectors and power series. These are also functorial and (hence) given by universal polynomials, i.e. by algebra morphisms

$$\mu_p\colon \mathbf{Symm} \longrightarrow \mathbf{Symm} \otimes \mathbf{Symm}, \varepsilon_p : \mathbf{Symm} \longrightarrow \mathbf{Z}$$

[40] This has most probably to do with the unwarranted and regrettable tendency of mathematicians to think primarily of Hopf algebras, etc. as things over a field of characteristic zero.

and these give a second bialgebra structure (**Symm**, m, e, μ_p, ε_p) on **Symm**. Because the functorial unit on the rings $\Lambda(A)$ is given by the power series $1+t+t^2+t^3+\cdots$ the co-unit morphism of rings ε_p is given by

$$\varepsilon_p(h_n) = 1, \quad n = 1, 2, 3, \cdots \tag{10.24}$$

The formula for the morphism governing the multiplication is less easy to describe; certainly explicitly. Various descriptions will be given later in the present section 10. There is no antipode for the second comultiplication of course (otherwise **Symm** would define a field valued functor).

Putting all these structure morphisms together there results the object

$$(\mathbf{Symm}, m, e, \mu_s, \varepsilon_s, \iota_s, \mu_p, \varepsilon_p)$$

which is a coring object in the category **CRing** of commutative unital rings. One axiom that such an object much satisfy is distributivity on both the left and the right of the second comultiplication over the first in the category **CRing** (where the appropriate codiagonal map is the multiplication on a ring). For distributivity on the left this means that the following diagram must be commutative

$$\begin{array}{ccc}
\mathbf{Symm} & \xrightarrow{\mu_s} & \mathbf{Symm} \otimes \mathbf{Symm} \\
\downarrow \mu_p & & \downarrow \mu_p \otimes \mu_p \\
\mathbf{Symm} \otimes \mathbf{Symm} & & \mathbf{Symm} \otimes \mathbf{Symm} \otimes \mathbf{Symm} \otimes \mathbf{Symm} \\
\downarrow \mathrm{id} \otimes \mu_s & & \downarrow \mathrm{id} \otimes \tau \otimes \mathrm{id} \\
\mathbf{Symm} \otimes \mathbf{Symm} \otimes \mathbf{Symm} & \xleftarrow{m \otimes \mathrm{id} \otimes \mathrm{id}} & \mathbf{Symm} \otimes \mathbf{Symm} \otimes \mathbf{Symm} \otimes \mathbf{Symm}
\end{array}$$

where, as usual $\tau : a \otimes b \mapsto b \otimes a$ is the twist morphism. There is a similar diagram for distributivity on the right.

10.25. *Various descriptions of the second comultiplication morphism on* **Symm**. At first sight the easiest description is of course in terms of the power sum symmetric functions as

$$\mu_P(p_n) = p_n \otimes p_n \tag{10.20}$$

but for a variety of reasons this one is of very limited use. It is of course the definition of the multiplication in terms of ghost components.

Much better are the three descriptions that follow from the three expansions of $\prod(1-\xi_i\eta_j)^{-1}$ given in (9.62), (9.44), (9.45). These give the second comultiplication formulas

$$\mu_P(h_n) = \sum_\lambda h_\lambda \otimes m_\lambda \tag{10.21}$$

$$\mu_P(h_n) = \sum_\lambda z_\lambda^{-1} p_\lambda \otimes p_\lambda \tag{10.22}$$

$$\mu_P(h_n) = \sum_\lambda s_\lambda \otimes s_\lambda \tag{10.23}$$

where in all three formulas the sum is over all partitions of n. Explicitly the first few multiplication polynomials are

$$\mu_P(h_1) = h_1 \otimes h_1, \mu_P(h_2) = 2h_2 \otimes h_2 - (h_2 \otimes {h_1}^2 + {h_1}^2 \otimes h_2) + {h_1}^2 \otimes {h_1}^2$$
$$\mu_P(h_3) = 3h_3 \otimes h_3 - 3(h_3 \otimes h_2h_1 + h_2h_1 \otimes h_3) + (h_3 \otimes {h_1}^3 + {h_1}^3 \otimes h_3)$$
$$- (h_2h_1 \otimes {h_1}^3 + {h_1}^3 \otimes h_2h_1) + {h_1}^3 \otimes {h_1}^3 \tag{10.24}$$

10.25. Still other descriptions of the second comultiplication morphism and the (functorial) multiplication on $W(A)$ will be given below, in sections 15 and 11 respectively.

10.26. *Product comultiplication vs inner product.* Theorem 9.42, in particular the expansion formula (9.44), compared with the definition of the multiplication on the Witt vectors determined by

$$(1 - \xi t)^{-1} * (1 - \eta t)^{-1} = (1 - \xi\eta t)^{-1}$$

suggest that there are nontrivial interrelations between the inner product on **Symm** and the second comultiplication (= product comultiplication) on **Symm**. And so there are. For instance cocommutativity of the product comultiplication and symmetry of the inner product are equivalent. At this stage and in this setting this is just an amusing remark. There may be more to it in other contexts.

The inner product is defined by formula (9.41)

$$\langle h_\lambda, m_\mu \rangle = \delta_{\lambda,\mu}$$

Write the monomial symmetric functions as (integer) linear combinations of the h_λ:

$$m_\mu = \sum_\lambda a_\mu^\lambda h_\lambda$$

so that

$$\langle m_\mu, m_{\mu'} \rangle = a_\mu^{\mu'} \tag{10.27}$$

Now (by (9.44))

$$\mu_P(h_n) = \sum_{\mathrm{wt}(\lambda)=n} h_\lambda \otimes m_\lambda = \sum a_\lambda^{\lambda'} h_\lambda \otimes h_{\lambda'} \tag{10.28}$$

By cocommutativity of the product comultiplication, also

$$\mu_P(h_n) = \sum_{\mathrm{wt}(\lambda)=n} m_\lambda \otimes h_\lambda = \sum a_\lambda^{\lambda'} h_{\lambda'} \otimes h_\lambda \tag{10.29}$$

Comparison ot the two expressions (10.28) and (10.29) gives, see (10.27),

$$a_\lambda^{\lambda'} = a_{\lambda'}^\lambda$$

i.e. symmetry of the inner product. The argument also runs the other way.

11. QSymm, the Hopf algebras of quasisymmetric functions and NSymm, the Hopf algebra of noncommutative symmetric functions

When looking at various universality properties of the Witt vectors and **Symm** (which is the topic of the next section) one rapidly stumbles over a (maximally) non commutative version, **NSymm**, and a (maximally) non cocommutative version, **QSymm**. This section is devoted to a brief discussion of these two objects. Somehow a good many things become easier to see and to formulate in these contexts (including certain explicit calculations). As I have said before, e.g. in [200], p. 56; [199], Ch 1, p. 1, once one has found the *right* non commutative version, things frequently become more transparent, easier to understand, and much more elegant.

11.1. *The Hopf algebra of non commutative symmetric functions.* Let $Z = \{Z_1, Z_2, Z_3, \cdots\}$ be a countably infinite set of (noncommuting) indeterminates. As an algebra the Hopf algebra of noncommutative symmetric functions is simply the free associative algebra in these indetermines over the integers

$$\mathbf{NSymm} = Z\langle Z_1, Z_2, Z_3, \cdots\rangle = Z\langle Z\rangle \tag{11.2}$$

The coalgebra structure is given on the generators Z_n by

$$\mu(Z_n) = \sum_{i+j=n} Z_i \otimes Z_j = 1 \otimes Z_n + Z_1 \otimes Z_{n-1} + Z_2 \otimes Z_{n-2} + \cdots + Z_{n-1} \otimes Z_1 + Z_n \otimes 1 \tag{11.3}$$

where the notation $Z_0 = 1$ is used in the expression in the middle. For every word over the natural numbers $\alpha = [a_1, a_2, \cdots, a_m],\ a_i \in \mathbf{N} = \{1, 2, 3, \cdots\}$

$$Z_\alpha = Z_{a_1} Z_{a_2} \cdots Z_{a_m} \tag{11.4}$$

denotes the corresponding (non commutative) monomial. This incudes $Z_{[]} = 1$. These form a basis of the free Abelian group underlying **NSymm**. The co-unit morphism is given by

$$\varepsilon(Z_n) = 0,\ n = 1, 2, 3, \cdots, \varepsilon\big(Z_{[\,]}\big) = 1 \tag{11.5}$$

Give Z_n degree (or weight) n, so that the weight (or degree) of a monomial Z_α is $a_1 + a_2 + \cdots + a_m = \mathrm{wt}(\alpha)$. Then, obviously, **NSymm** is a graded Abelian group

$$\mathbf{NSymm} = \bigoplus_{n\geq 0} \mathbf{NSymm}_n \tag{11.6}$$

with the homogeneous part of degree n spanned by the monomials (11.4) of weight n. Then, obviously, (**Symm**, m, e, μ, ε) is a graded, connected (meaning that $\mathbf{NSymm}_0 = \mathbf{Z}$) bialgebra so that there is (for free, see 10.14) a suitable antipode making **NSymm** a connected, graded, non commutative, cocommutative Hopf algebra.

There is a natural surjective morphism of graded Hopf algebras

$$\mathbf{NSymm} \longrightarrow \mathbf{Symm},\ Z_n \mapsto h_n \tag{11.7}$$

that exhibits **Symm** as the maximal commutative quotient Hopf algebra of **NSymm**. The kernel of (11.7) is the commutator ideal generated by all elements of the form $Z_iZ_j - Z_jZ_i$, $i, j \in \mathbf{N}$.

The first paper to study **NSymm** in depth was probably [162]. It was immediately followed by a slew of other publications [54, 129, 130, 128, 216, 214, 219, 249, 251, 250, 386, 384, 200, 195, 198, 193, 201]. [41]

11.8. *Divided power sequences. Curves*. Given a Hopf algebra H a divided power sequence, or curve, in H is by definition a sequence of elements

$$\gamma = (d_0 = 1, d_1, d_2, \cdots) \tag{11.9}$$

such that

$$\mu_H(d_n) = \sum_{i+j=n} d_i \otimes d_j \tag{11.10}$$

If H is the covariant bialgebra of a formal group F this translates into a mapping of the formal affine line into the formal scheme (variety) underlying F, whence the terminology curve.[42] A curve is often denoted by its generating power series

$$\gamma(t) = 1 + d_1t + d_2t^2 + d_3t^3 + \cdots \tag{11.11}$$

Two such power series can be multiplied and the result is again a curve, i.e a power series in t of which the coefficients satsify (11.10). This defines a functor

$$\mathit{Curve}\colon \mathbf{Hopf} \longrightarrow \mathbf{Group} \tag{11.12}$$

from the category of Hopf algebras to the category of groups. An example of a curve is the universal curve

$$Z(t) = 1 + Z_1t + Z_2t^2 + \cdots \in \mathit{Curve}(\mathbf{NSymm}) \tag{11.13}$$

and it is immediate from the defining property (11.10) and the freeness of **NSymm** as an associative algebra that

11.14. *Theorem* (*universality property of* **NSymm**). For every curve $\gamma(t)$ in a Hopf algebra H there is a unique morphism of Hopf algebras **NSymm** $\longrightarrow H$ that takes the universal curve $Z(t)$ to the given curve $\gamma(t)$.

That is, the functor *Curve* is represented by **NSymm**.[43] The group operation on curves corresponds to the convolution product on **Hopf**(**NSymm**, H). When the Hopf algebra H is commutative this theorem translates into the following.

[41] Many of these papers are about noncommutative symmetric functions over a field of characteristic zero; here, of course, it is **NSymm** over the integers that is really important (here and elsewhere).

[42] There is also another notion in algebra that goes by the name 'divided powers' viz the presence of a sequence of operators (usually on an ideal in a commutative ring) that behave like sending an element x to $(n!)^{-1}x^n$. Hence the term curve, rather than divided power sequence (DPS) is preferred for the notion defined by (11.9), (11.10).

[43] See also [149].

11.15. *Theorem.* (*Second universality property of the Witt vectors*). Let $\gamma(t)$ be a curve in a commutative Hopf algebra H. Then there is a unique morphism of Hopf algebras **Symm** $\longrightarrow H$ that takes the curve

$$h(t) = 1 + h_1 t + h_2 t^2 + \cdots \in \mathit{Curve}(\mathbf{Symm}) \tag{11.16}$$

into the given curve $\gamma(t)$.

In the context of commutative formal groups with H the covariant bialgebra of a commutative formal group this is known as Cartier's first theorem.

11.17. *The Hopf algebra* **QSymm** *of quasi-symmetric functions.* The next sections deal with another generalization of **Symm**, dual to **NSymm** and containing **Symm**.

11.18. *Compositions and partitions.* A composition is a word (of finite length) over the natural numbers. Here a composition will be written

$$\alpha = [a_1, a_2, \cdots, a_m],\ \ a_i \in \mathbf{N} = \{1, 2, 3, \cdots\} \tag{11.19}$$

A composition (11.19) is said to have length m and weight $\mathrm{wt}(\alpha) = a_1 + a_2 + \cdots + a_m$. Associated to a composition α is a partition λ which consists of the a_j re-arranged in nonincreasing order. Or, equivalently, $\lambda = \left(1^{m_1(\alpha)} 2^{m_2(\alpha)} 3^{m_3(\alpha)} \cdots\right)$ where $m_j(\alpha)$ is the number of entries of α equal to j.

11.19. *Monomial quasi-symmetric functions.* Consider again the ring of polynomials, i.e. power series of bounded degree[44] in a countable infinity of commuting indeterminates, i.e. an element of $\mathbf{Z}[\xi_1, \xi_2, \xi_3, \cdots]$. Such a polynomial

$$\sum c_{\alpha, \iota_1, \cdots, i_m} \xi_{i_1}^{a_1} \xi_{i_2}^{a_2} \cdots \xi_{i_m}^{a_m} \tag{11.20}$$

where the sum runs over runs over all compositions $\alpha = [a_1, \cdots, a_m]$ and all index sequences $i_1 < \cdots < i_m$ of the same length as α, is said to be quasi-symmetric if

$$c_{\alpha, \iota_1, \cdots, i_m} = c_{\alpha, j_1, \cdots, j_m} \text{ for all index sequences } i_1 < \cdots < i_m \text{ and } j_1 < \cdots < j_m \tag{11.21}$$

For a given composition $\alpha = [a_1, \cdots, a_m]$ the associated monomial quasi-symmetric function is

$$[a_1, \cdots, a_m] = \sum_{i_1 < i_2 < \cdots < i_m} \xi_{i_1}^{a_1} \xi_{i_2}^{a_2} \cdots \xi_{i_m}^{a_m} \tag{11.22}$$

It is denoted by the same symbol. For example in four variables

$$\begin{aligned}
[1, 2] &= \xi_1\xi_2^2 + \xi_1\xi_3^2 + \xi_1\xi_4^2 + \xi_2\xi_3^2 + \xi_2\xi_4^2 + \xi_3\xi_4^2 \\
[1, 2, 1] &= \xi_1\xi_2^2\xi_3 + \xi_1\xi_2^2\xi_4 + \xi_1\xi_3^2\xi_4 + \xi_2\xi_3^2\xi_4 \\
[3, 2, 1] &= \xi_1^3\xi_2^2\xi_3 + \xi_1^3\xi_2^2\xi_4 + \xi_1^3\xi_3^2\xi_4 + \xi_2^3\xi_3^2\xi_4
\end{aligned} \tag{11.23}$$

[44] See the appendix.

Note that the monomial symmetric function m_λ defined by a partition λ is the sum of all *distinct* monomial quasi-symmetric functions α whose associated partition is λ. For example

$$\begin{aligned} m_{(2,1)} &= [2,1] + [1,2], \ \ m_{(2,1,1)} = [2,1,1] + [1,2,1] + [1,1,2] \\ m_{(3,2,1)} &= [3,2,1] + [3,1,2] + [2,3,1] + [2,1,3] + [1,3,2] + [1,2,3] \end{aligned} \tag{11.24}$$

The monomial quasi-symmetric functions form a free basis of the free Abelian group **QSymm** of all quasi-symmetric functions. This group is graded by assigning to the monomial quasi-symmetric function $\alpha = [a_1, \cdots, a_m]$ the weight $\mathrm{wt}(\alpha)$. Sum and product of quasi-symmetric functions are again quasi-symmetric, so that **QSymm** is a ring.

The ring **Symm** is a subring with the inbedding

$$h_n = \sum_{\mathrm{wt}(\alpha)=n} \alpha \tag{11.25}$$

11.26. *Overlapping shuffle product.* It is useful to have an explicit direct recipe in terms of compositions for the multiplication of monomial quasi-symmetric functions. This is given by the overlapping shuffle product.

Let $\alpha = [a_1, a_2, \cdots, a_m]$ and $\beta = [b_1, b_2, \cdots, b_n]$ be two compositions or words. Take a 'sofar empty' word with $n + m - r$ slots where r is an integer between 0 and $\min\{m, n\}$, $0 \leq r \leq \min\{m, n\}$.

Choose m of the available $n + m - r$ slots and place in it the natural numbers from α in their original order; choose r of the now filled places; together with the remaining $n + m - r - m = n - r$ empty places these form n slots; in these place the entries form β in their original order; finally, for those slots which have two entries, add them. The product of the two words α and β is the sum (with multiplicities) of all words that can be so obtained. So, for instance

$$\begin{aligned} [a,b][c,d] = {} & [a,b,c,d] + [a,c,b,d] + [a,c,d,b] + [c,a,b,d] + [c,a,d,b] \\ & + [c,d,a,b] + [a+c,b,d] + [a+c,d,b] + [c,a+d,b] \\ & + [a,b+c,d] + [a,c,b+d] + [c,a,b+d] + [a+c,b+d] \end{aligned} \tag{11.27}$$

$$[1][1][1] = 6[1,1,1] + 3[1,2] + 3[2,1] + [3] \tag{11.28}$$

It is easy to see that the recipe given above gives precisely the multiplication of (the corresponding basis) quasi-symmetric functions. The shuffles with no overlap of $a_1, \cdots, a_m; b_1, \cdots, b_n$ correspond to the products of the monomials that have no ξ_j in common; the other terms arise when one or more of the ξ's in the monomials making up α and β do coincide. In example (11.27) the first six terms are the *shuffles*; the other terms are '*overlapping shuffles*'.

The term shuffle comes from the familiar rifle shuffle of cardplaying; an overlapping shuffle occurs when one or more cards from each deck don't slide along each other but stick edgewise together; then their values are added.

11.29. *Hopf algebra structure on* **QSymm**. Now introduce on **QSymm** the comultiplication 'cut':

$$\begin{aligned}\mu_{\mathbf{QSymm}}([a_1, a_2, \cdots, a_m]) = 1 \otimes [a_1, a_2, \cdots, a_m] \\ + \sum_{i=1}^{n-1} [a_1, \cdots, a_i] \otimes [a_{i+1}, \cdots, a_m] \\ + [a_1, a_2, \cdots, a_m] \otimes 1\end{aligned} \tag{11.30}$$

and the counit

$$\varepsilon_{QSymm}(\alpha) = \begin{cases} 1 & \text{if } \lg(\alpha) = 0 \\ 0 & \text{if } \lg(\alpha) \geq 1 \end{cases} \tag{11.31}$$

Here recall that the only composition of length zero is the empty composition $1 = [\,]$ which is the unit for the overlapping shuffle multiplication.

Write $m_{\mathbf{QSymm}}$ for the overlapping shuffle multiplication and $e_{\mathbf{QSymm}}$ for the unit morphism determined by $1 \mapsto [\,] = 1$. Then

11.32. *Theorem.* The structure ($\mathbf{QSymm}$, $m_{\mathbf{QSymm}}$, $e_{\mathbf{QSymm}}$, $\mu_{\mathbf{QSymm}}$, $\varepsilon_{\mathbf{QSymm}}$) is a commutative non cocommutative graded connected bialgebra and hence defines a Hopf algebra.

The antipode again comes for free once the the bialgebra statement has been proved because **QSymm** is connected graded. The only thing in proving the bialgebra statement which is not dead easy is the verification that the comultiplication is an algebra morphism for the overlapping shuffle product. Even that can be avoided by using graded duality as described below.

11.33. *Duality between.* **NSymm** *and* **QSymm**. Introduce the nondegenerate graded pairing

$$\langle\,,\rangle\colon \mathbf{NSymm} \times \mathbf{QSymm} \longrightarrow \mathbf{Z},\ \langle Z_\alpha, \beta\rangle = \delta_{\alpha,\beta} \tag{11.34}$$

With respect to this pairing the multiplication (resp. comultiplication) on **NSymm** is dual to the comultiplication (resp. multiplication) on **QSymm**. Similarly the counit (resp. unit) on **NSymm** is dual to the unit (resp. counit) on **QSymm**. These statements amount to the following four formulas.

$$\begin{aligned}\langle m_{\mathbf{NSymm}}(Z_\alpha \otimes Z_\beta), \gamma\rangle &= \langle Z_\alpha \otimes Z_\beta, \mu_{\mathbf{QSymm}}(\gamma)\rangle \\ \langle \mu_{\mathbf{NSymm}}(Z_\alpha), \beta \otimes \gamma\rangle &= \langle Z_\alpha, m_{\mathbf{QSymm}}(\beta \otimes \gamma)\rangle \\ \langle e_{\mathbf{NSymm}}(1), \beta\rangle &= \langle 1, \varepsilon_{\mathbf{QSymm}}(\beta)\rangle \\ \langle \varepsilon_{\mathbf{NSymm}}(\alpha), 1\rangle &= \langle \alpha, e_{\mathbf{QSymm}}(1)\rangle\end{aligned} \tag{11.35}$$

Where $\langle\,,\rangle$ on $\mathbf{Z}$ is the innerproduct $\langle r, s\rangle = rs$. Formulas (11.35) easily imply that if **NSymm** is a bialgebra, so is **QSymm** (and vice-versa). The antipodes on the two

Hopf algebras are (as a consequence) also dual to each other:

$$\langle \iota_{\mathbf{NSymm}}(Z_\alpha), \beta\rangle = \langle Z_\alpha, \iota_{\mathbf{QSymm}}(\beta)\rangle$$

11.36. *Primitives of* **QSymm**. It is an easy exercise to show that the primitives for the first comultiplication, the 'sum comultiplication', i.e. the elements for which

$$\mu_S(x) = 1 \otimes x + x \otimes 1$$

are precisely the linear combinations of the words of length 1.[45]

11.37. *Autoduality of* **Symm**. Under (graded) duality a sub Hopf algebra of a Hopf algebra H corresponds to a quotient Hopf algebra of the graded dual. To avoid notational confusion write temporarily **Symm'** for the algebra of symmetric functions as a subalgebra of **QSymm** and retain the notation $\mathbf{Symm} = \mathbf{Z}[h]$ for the algebra of symmetric functions as a quotient algebra of **NSymm**. The Hopf ideal J in **NSymm** such that $\mathbf{Nsymm}/\boldsymbol{J} = \mathbf{Symm}$ is the commutator ideal spanned by all elements of the form $Z_\alpha(Z_iZ_j - Z_jZ_i)Z_\beta$. It is easy to check that a quasisymmetric function has pairing 0 (under (11.34) with all these elements if and only if it is symmetric. Thus the sub Hopf algebra of **QSymm** corresponding to the quotient **Symm** of **NSymm** is **Symm'**. Also the induced pairing is

$$\langle h_\lambda, m_\kappa\rangle = \delta_{\lambda,\kappa} \tag{11.38}$$

which is the inner product on **Symm** as defined in 9.40.[46]

Moreover as algebras **Symm** and **Symm'** are isomorphic under the isomorphism $h_n \mapsto e_n$. To prove that **Symm** and **Symm'** are isomorphic as Hopf algebras, it remains to check that this isomorphism takes the given comultiplication on **Symm**, $h_n \mapsto \sum_{i+j=n} h_i \otimes h_j$ into the comultiplication induced on **Symm'** by 'cut' the comultiplication on **QSymm**. But $e_n = [1, 1, 1, \cdots, 1]$ as an element of **Symm'** $\subset$ **QSymm** and so 'cut $= \mu_{\mathbf{QSymm}}$' on e_n takes the value

$$\mu_{\mathbf{QSymm}}(e_n) = \sum_{i+j=n} e_i \otimes e_j$$

and things fit perfectly.

11.39. *The second comultiplication on* **QSymm**. There is a second comultiplication (morphism) on **QSymm** that is distributive over the first one on the right but not on the left. Here is the definition. Let $\alpha = [a_1, a_2, \cdots, a_m]$ be a composition. A $(0, \alpha)$-matrix is a matrix whose entries are either zero or one of the a_i, which has no zero columns or zero rows, and in which the entries $a_1, a_2, \cdots, a_m$ occur in their original

[45] The situation for the primitives of **NSymm** is very different. Over the rationals the primitives form the free Lie algebra in countably many generators. Over the integers the description is still more involved and was only recently found; see [198].

[46] This is one main reason for working with the complete symmetric functions rather than with the elementary ones.

order if one orders the entries of a matrix by first going left to right through the first row, then left to right through the second row, etc.

For a matrix M let $column(M)$ be the vector of column sums and $row(M)$ the vector of row sums. For instance

$$M = \begin{pmatrix} 0 & 1 & 0 & 3 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{11.40}$$

is a $(0, [1, 3, 1, 2, 1, 1])$-matrix with $column(M) = [1, 3, 1, 4]$ and $row(M) = [4, 4, 1]$.

The second comultiplication on *QSymm* is now given by

$$\mu_P(\alpha) = \sum_{\substack{M \text{ is a} \\ (0,\alpha)\text{-matrix}}} row(M) \otimes column(M) \tag{11.41}$$

Restricted to $\mathbf{Symm} \subset \mathbf{QSymm}$, this describes the second comultiplication on **Symm** in combinatorial terms. To specify it, it suffices to do this on the elementary symmetric polynomials (because it is a ring morphism). i.e. to use the recipe for compositions of the form $\alpha = [1, 1, 1, \cdots, 1]$. For these compositions, the matrices involved are what are usually called (0, 1)-matrices.

Probably the most famous theorem on (0, 1)-matrices is the Gale-Ryser theorem. Thus it would appear likely that there is some connection between Witt vectors (via symmetric functions) and the Gale Ryser theorem, [150, 345]. And indeed there is (via the additional link of the symmetric functions with the representations of the symmetric groups). It is the Snapper Liebler-Vitale Lam theorem, [196, 254]. This will be described in a little more detail in subsection 18.33 below.

11.42. *Distributivity properties of* μ_P *over* μ_S. The second comultiplication morphism μ_P of (11.39) is distributive on the right over the first comultiplication morphism μ_S but not on the left.

Quite generally let $(H, m, e, \mu, \varepsilon)$ be a bialgebra equipped with a second comultiplication morphism of rings $\mu_P: H \longrightarrow H \otimes H$. For this second comultiplication to be distributive on the right over the first one (in the category of rings) one needs the following diagram (11.43) to commute. (This is precisely what is needed to ensure that the functor represented by H, $A \mapsto \mathbf{CRing}(H, A)$ have as values groups (denoted additively even when they are noncommutative) (as in the case at hand)) equipped with an additional multiplication such that $(a + b)c = ab + bc$.)

$$\begin{array}{ccc} H & \xrightarrow{\mu} & H \otimes H \\ \downarrow \mu_p & & \downarrow \mu_p \otimes \mu_p \\ H \otimes H & & H \otimes H \otimes H \otimes H \\ \downarrow \mu \otimes \mathrm{id} & & \downarrow \mathrm{id} \otimes \mathrm{tw} \otimes \mathrm{id} \\ H \otimes H \otimes H & \xleftarrow{\mathrm{id} \otimes \mathrm{id} \otimes m} & H \otimes H \otimes H \otimes H \end{array} \tag{11.43}$$

Here is a proof that this diagram is commutative in the case at hand. Take a composition α and let M be a $(0, \alpha)$-matrix. Applying $\mu\otimes$ id to $row(M) \otimes column(M)$ is the same thing as cutting the matrix M horizontally to obtain two blocks M_1, M_2 with the same number of columns stacked on top of each other. All terms in the lower left hand corner of diagram (11.43) coming from α thus are of the form $row(M_1) \otimes row(M_2) \otimes column(M)$. Now in $\mu(\alpha)$ (top horizontal arrow in the diagram) consider the term $\alpha_1 \otimes \alpha_2$ for which the length of α_1 is the same as the number of rows of M_1. Let M'_1, M'_2 be the matrices obtaine from M_1, M_2 by removing zero colums. These are respectively a $(0, \alpha_1)$-matrix and a $(0, \alpha_2)$-matrix. The term obtained in the lower right hand corner is now

$$\begin{aligned} &row(M'_1) \otimes row(M'_2) \otimes column(M'_1) \otimes column(M'_2) \\ &\quad = row(M_1) \otimes row(M_2) \otimes column(M'_1) \otimes column(M'_2) \end{aligned} \tag{11.44}$$

Now look at the overlapping shuffle product of two compositions β_1, β_2 in a slightly different way. Intersperse both compositions with zeros in any way to obtain vectors of the same length n and such that the two zero-interspersed compositions when put on top of each other give a $2 \times n$ matrix without zero columns. Take the column sum vector of this $2 \times n$ matrix. Doing this in all possible ways gives the overlapping shuffle product (obviously).

Now, all matrices M which yield given matrices M_1, M_2, are obtained from M'_1, M'_2 by interspersing them with zero columns to have the same number of columns and such that when stacked on top of each other they give a matrix without zero columns.

Combining these two remarks gives that for a given M'_1, M'_2 the possible originating M precisely give the compositions $column(M)$ that arise as terms of the overlapping shuffle product of $column(M'_1)$ and $column(M'_2)$, proving the commutativity of (11.43).

For left distributivity a similar diagram must be commutative. it is obtained from the one at hand by replacing the lower left hand morphism by id $\otimes\, \mu$ and the bottom morphism by $m \otimes$ id $\otimes$ id. It is a trivial matter to check that left distributivity does not hold; it already fails for any $\alpha = [a_1, a_2]$ with $a_1 \neq a_2$.

If one switches 'row' and 'column' in (11.37) there results a second comultiplication morphism that is left distributive but not right distributive.

One easily sees that on a symmetric sum of quasisymmetric monomials μ_P is commutative. Indeed if M is a $(0, \alpha)$ - matrix then $row(M^{\mathrm{tr}}) = column(M)$ and $column(M^{\mathrm{tr}}) = row(M)$ and the transpose M^{tr} is a $(0, \alpha')$-matrix for some permutation α' of α.

Given right distributivity of this second comultiplication on **QSymm** to check that it does indeed give the functorial multiplication of Witt vectors or power series it suffices to check things on power series of the form $1 + at$ which is a triviality.

11.45. *Co-unit for the second comultiplication on* **QSymm**. There is also a counit morphism ε_P for μ_P. It is given by

$$\varepsilon_P([a_1, a_2, \cdots, a_m]) = \begin{cases} 1 & \text{for } m \leq 1 \\ 0 & \text{for } m \geq 2 \end{cases} \tag{11.46}$$

To be a counit for μ_P it must satisfy the condition that the following composition of morphisms is the identify

$$\mathbf{QSymm} \xrightarrow{\mu_P} \mathbf{QSymm} \otimes \mathbf{QSymm} \xrightarrow{\varepsilon_P \otimes \mathrm{id}} \mathbf{QSymm} \otimes \mathbf{QSymm} \xrightarrow{m} \mathbf{QSymm}$$

and also the composition of morphisms obtained from this by switching ε_P and id for the middle arrow.

Let $\alpha = [a_1, a_2, \cdots, a_m]$. As ε_P is zero on anything of length 2 or more the only $(0, \alpha)$-matrices that need to be considered have only one row. There is just one such matrix, viz $(a_1, a_2, \cdots, a_m)$. This gives that the composed morphism in question is indeed the identity.

11.47. *Second multiplication on* **NSymm**. Dually the second comultiplication on **QSymm** gives a second mutliplication morphism (of coalgebras) on **NSymm**. Denoting this one by *, there is the right distributivity formula (which follows from the right distributivity of μ_P over μ_S in **QSymm**).

$$(Z_\alpha Z_\beta) * Z_\gamma = \sum_i \left(Z_\alpha * Z_{\gamma_i'}\right)\left(Z_\beta * Z_{\gamma_i''}\right) \quad \text{where } \mu(Z_\gamma) = \sum_i Z_{\gamma_i'} \otimes Z_{\gamma_i''}$$

This formula is also in [162]. The dual of the recipe for the second comultiplication given in (11.39) is due to Solomon, [373].

11.48. *Unit for the second multiplication on* **NSymm**. The second multiplication, m_P on **NSymm** defines a multiplication on each homogeneous component $\mathbf{NSymm}_n$. The unit for this second multiplication is the element Z_n. So the 'unit' for the second comultiplication on all **NSymm** is the infinite sum

$$1 + Z_1 + Z_2 + Z_3 + \cdots$$

which does not live in **NSymm** but only in a suitable completion.

The same situation holds for **Symm**. Here, for each n the second multiplication, sometimes called inner multiplication, turns each homogenous summand $\mathbf{Symm}_n$ into a commutative ring with unit element h_n, and the sum of all these elements, which does not live in **Symm** itself is the unit element for the second multiplication on all of **Symm**.

So the autoduality of **Symm** is not prefect at this level. The second multiplication and second comultiplication are nicely dual to each other but the second unit only exists in a suitable completion while the second counit is present for **Symm** itself.

This is understandable because the second multiplication and second comultiplication do not respect the grading.

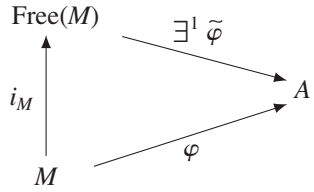
12. Free, cofree and duality properties of Symm

This short section concerns ways of obtaining **Symm** etc. by means of free and cofree constructions and consequences thereof. And open questions concerning these matters. All algebras and coalgebras in this section will come with a unit resp. counit element seen as.morphisms from and to **Z**.

12.1. *Free algebras.* To set the stage consider first the well known case of free algebras. Let M be an Abelian group. A free algebra on M is an algebra Free(M) together with a morphism of Abelian groups $i_M : M \longrightarrow \text{Free}(M)$ such that the following universality property holds

$$\begin{aligned}&\text{For every morphism of Abelian groups } \varphi\colon M \longrightarrow A \text{ to a ring } A\\ &\text{there is a unique morphism of rings } \tilde{\varphi}\colon \text{Free}(M) \longrightarrow A \text{ such that } \tilde{\varphi} i_M = \varphi\end{aligned} \tag{12.2}$$

This defines a functor Free: **AbGroup** $\longrightarrow$ **Ring** that is left adjoint to the (forgetful) functor U: **Ring** $\longrightarrow$ **AbGroup** that forgets about the multiplication. Pictorially, the situation looks as follows.



(12.3)

The symbol group $\exists^1$ in the diagram above means 'there exists a unique'. Such a functor 'Free' exists. It is given by the tensor algebra over M.

12.4. *Tensor algebra and tensor coalgebra.* For an Abelian group M let $T^i M = M^{\otimes i} = M \otimes M \otimes \cdots \otimes M$ (and $T^0 M = \mathbf{Z}$) be the i-fold tensor product. There are natural isomorphisms $\psi_{n,m} : T^n M \otimes T^n M \xrightarrow{\cong} T^{n+m} M$. Now consider

$$\begin{aligned}&TM = \bigoplus_{n=0}^{\infty} T^n M,\\ &i_M\colon M \longrightarrow TM, \quad p_M\colon TM \longrightarrow M,\\ &e_M\colon Z \longrightarrow TM, \quad \varepsilon_M\colon TM \longrightarrow Z\end{aligned} \tag{12.5}$$

where the morphisms i_M, p_M are obtained by identifying M and $T^1 M$ and e_M, ε_M by identifying $\mathbf{Z}$ and $T^0 M$. There is a ring structure on TM given by

$$T^m M \otimes T^n M \xrightarrow{\psi_{m,n}} T^{n+m} M \text{ and } i_M \text{ as unit morphism.} \tag{12.6}$$

This is the tensor algebra and it is the free algebra on M with respect to the (canonical) morphism i_M.

There is also a coalgebra structure on TM given by

$$(\psi_{0,n}^{-1}, \psi_{1,n-1}^{-1}, \psi_{2,n-2}^{-1}, \cdots, \psi_{n-1,1}^{-1}, \psi_{n,0}^{-1})\colon T^n M \longrightarrow \bigoplus_{i=0}^{n} T^i M \otimes T^{n-i} M \tag{12.7}$$

and ε_M as counit morphism

The two structures do not combine to give a bialgebra structure. Far from it.

12.8. *The graded case.* In the present context the important case is when everything in sight is graded and M is a free Abelian group.

So let M be a positively graded Abelian group

$$M = \bigoplus_{i=1}^{\infty} M_i \tag{12.9}$$

Then TM is graded by giving a pure tensor $x_1 \otimes \cdots \otimes x_m$, $x_i \in M_{i_1}$ degree $i_1 + \cdots i_m$. If no grading on M is specified it is treated as graded with all nonzero elements of degree 1. If now M is moreover free with homogeneous basis

$$\{T_u \colon u \in U\}, \quad \text{degree}(T_u) = r_u \tag{12.10}$$

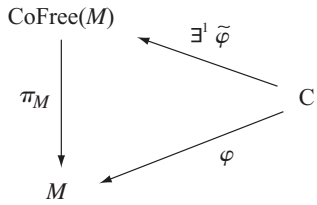
Free(M) is $\mathbf{Z}\langle T_u \colon u \in I\rangle$. For instance if the index set is $\mathbf{N}$ and the degree of T_i is i Free(M) = **NSymm** (as far as the underlying graded algebra is concerned).

There is also a commutative version with $\text{Free}_{\text{comm}}(M)$ being the maximal commutative qoutient of Free(M). So if $M = \bigoplus_{i=1}^{\infty} \mathbf{Z}h_i$, degree $(h_i) = i$, $\text{Free}_{\text{comm}}(M) =$ **Symm**.

12.11. *Cofree coalgebras.* All coalgebras in the following are supposed to come with a co-unit and morphisms of coalgebras are supposed to be compatible with the co-units involved. Given an Abelian group M the cofree coalgebra over M (if it exists) is a coalgebra CoFree(M) (over the integers) together with a morphism of Abelian groups $\pi_M \colon \text{CoFree}(M) \longrightarrow M$ with the following universal property:

$$\begin{aligned} &\text{For every coalgebra } C \text{ together with a morphism of Abelian groups} \\ &\varphi : C \longrightarrow M \text{ there is a unique morphism of coalgebras} \\ &\tilde{\varphi} : C \longrightarrow \text{CoFree}(M) \text{ such that } \pi_M \tilde{\varphi} = \varphi \end{aligned} \tag{12.12}$$

Pictorially this looks as follows:



(12.13)

a picture that is completely dual to the one, see(12.3), for a free algebra over a module. The universality property (12.12), (12.13) says, given existence, that 'CoFree' is right adjoint to the forgetful functor which assigns to a coalgebra the underlying Abelian group. Whether the cofree coalgebra on[47] an Abelian group always exists is unkown. It does do so in the case of free Abelian groups, their duals[48] and certain other cases, see [194].

[47] It pays to be linguistically careful here: free algebra **on** an Abelian group; cofree coalgebra **over** an Abelian group.

[48] These are not free in the infinite rank case.

Let M be (graded) free with (homogeneous) basis $\{T_u\colon u \in I\}$; and let C be a coalgebra together with a morphism of Abelian groups $C \xrightarrow{\varphi} M$. As an Abelian group the tensor coalgebra consists of all noncommutative polynomials in the T_u. This is not always large enough to serve as receiving object for the unique morphism $\tilde{\varphi}$ covering φ as required in (12.12), (12.13). Consider for example the coalgebra $C = \mathbf{Z} \oplus \mathbf{Z}x$, $\mu(x) = x \otimes x$, $\varepsilon(x) = 1$ together with the morphism of Abelian groups that sends $\mathbf{Z}$ to zero and x to one of the basis elements T of M. It is obivous that the n-th component of the covering morphism, $\tilde{\varphi}_n\colon C \longrightarrow T^i M$, must (always) be given by $\tilde{\varphi}_n = \varphi^{\otimes n}\mu_n$ where μ_n is the iterated comultiplication

$$\mu_0 = \varepsilon, \mu_1 = \mathrm{id}, \mu_2 = \mu, \cdots, \mu_{n+1} = (\mu \otimes \mathrm{id} \otimes \cdots \otimes \mathrm{id})\mu_n, \cdots \qquad (12.14)$$

Thus in the example at hand $\tilde{\varphi}_n(x) = T \otimes T \otimes \cdots \otimes T$ (n factors), so that the $\tilde{\varphi}(x)$ 'is' the power series $1 + T + T^2 + \cdots + T^n + \cdots$ which is not in TM. To obtain the free coalgebra on M consider the completion $\hat{T}M$ of all formal noncommutative power series in the indeterminates T_u, $u \in I$. The free coalgebra over M is now something in between TM and $\hat{T}M$ of which the elements can be called for good (but varying) reasons 'rational noncommutative power series', 'representative noncommutative power series', 'Schützenberger-recognizable noncommutative power series', or 'recursive noncommutative power series'. loc. cit.

Things become a good deal easier in the graded case. In this case, i.e. for graded morphisms of graded coalgebras into graded Abelian groups, the tensor coalgebra is the cofree coalgebra. To sees this first observe that a morphism of graded Abelian groups $C \longrightarrow M$ takes the degree 0 component of C to zero. Next for any homogeneous $x \in C$ of degree n, say, all terms in its terated coproduct $\mu_m(x)$ for $m > n$ must contain a factor from C_0. Hence $\tilde{\varphi}_m(x) = \varphi^{\otimes m}\ \mu_m(x) = 0$ for $m > n$ and things are OK.

There is also a cocommutative version where TM is replaced by its maximal cocommutative sub coalgebra of all symmetric tensors.

In the case of graded free Abelian groups the graded free and graded cofree constructions are graded dual in that there is a natural isomorphism

$$\mathrm{CoFree}(M)^{dual} = \mathrm{Free}(M^{dual})$$

For the nongraded case there is also a duality but things are more complicated.[49]

12.15. *Inheritance of structure*. Now let C be a coalgebra and apply the free algebra construction to the augmentation module $I = \mathrm{Ker}(\varepsilon)$. This gives (in any case) an algebra which will still be denoted $\mathrm{Free}(C)$. It inherits a coalgebra structure as follows.

By functoriality the coalgebra morphism gives an algebra morphism

$$\mathrm{Free}(C) \xrightarrow{\mathrm{Free}(\mu)} \mathrm{Free}(C \otimes C) \qquad (12.16)$$

[49] Partly this comes from the fact that while the linear algebraic dual of a coalgebra is immediately an algebra it is not true that the full dual of an algebra is a coalgebra; instead, one must take a submodle of the dual called the zero-dual.

Further there is the natural morphism $i_c \otimes i_c \colon C \otimes C \longrightarrow \mathrm{Free}(C) \otimes \mathrm{Free}(C)$ and by the freeness property this induces a morphism of algebras $\mathrm{Free}(C \otimes C) \longrightarrow \mathrm{Free}(C) \otimes \mathrm{Free}(C)$ which when composed with (12.16) gives the desired coalgebra structure morphism

$$\mathrm{Free}(C) \overset{\mu_{\mathrm{Free}(C)}}{\longrightarrow} \mathrm{Free}(C) \otimes \mathrm{Free}(C) \tag{12.17}$$

By construction it is an algebra morphism and so $\mathrm{Free}(C)$ becomes a bialgebra.

Dually, consider an algebra A and apply the cofree construction to the quotient module $A/\mathbf{Z}$. Denote the result by $\mathrm{CoFree}(A)$. This one inherits an algebra structure in a similar way. Applying CoFree to the multiplication morphism gives a coalgebra morphism

$$\mathrm{CoFree}(A \otimes A) \longrightarrow \mathrm{CoFree}(A) \tag{12.18}$$

On the other hand there is a fairly obvious (canonical) morphism of Abelian groups $\mathrm{CoFree}(A) \otimes \mathrm{CoFree}(A) \longrightarrow (A \otimes A)/\mathbf{Z}$ that by the cofreeness property induces a morphism of coalgebras

$$\mathrm{CoFree}(A) \otimes \mathrm{CoFree}(A) \to \mathrm{CoFree}(A \otimes A) \tag{12.19}$$

which when composed with with (12.18) gives the desired multiplication morphism which is by construction a coalgebra morphism so that the result is again a bialgebra.

12.20. Symm, NSymm *and* **QSymm** *again*. The cofree coalgebra over the integers is

$$\begin{aligned} &\mathrm{CoFree}(\mathbf{Z}) = \mathbf{Z} \oplus \mathbf{Z}Z_1 \oplus \mathbf{Z}Z_2 \otimes \cdots \\ &\mu(Z_n) = \sum_{i=0}^{n} Z_i \otimes Z_{n-i}, \quad \text{where} \quad Z_0 = 1, \varepsilon(Z_n) = \delta_{0,n} \end{aligned} \tag{12.21}$$

It is now easy to see that

$$\mathbf{NSymm} = \mathrm{Free}(\mathrm{CoFree}(\mathbf{Z})), \mathbf{Symm} = \mathrm{Free}_{\mathrm{comm}}(\mathrm{CoFree}(\mathbf{Z})) \tag{12.22}$$

It is far less easy (except via duality) to see in this way that $\mathbf{QSymm} = \mathrm{CoFree}(\mathrm{Free}(\mathbf{Z}))$, but it is interesting to do the exercise as this way the overlapping shuffle product comes about naturally.

It is also not easy to see directly that $\mathbf{Symm} = \mathrm{CoFree}_{\mathrm{cocomm}}(\mathrm{Free}(\mathbf{Z}))$ (except, again, via duality. The duality argument is simple.

$$\begin{aligned} \mathbf{QSymm} = \mathbf{NSymm}^{dual} &= \mathrm{Free}(\mathrm{CoFree}(Z))^{dual} \\ &= \mathrm{CoFree}(\mathrm{CoFree}(\mathbf{Z})^{dual}) = \mathrm{CoFree}(\mathrm{Free}(\mathbf{Z}^{dual})) \\ &= \mathrm{CoFree}(\mathrm{Free}(\mathbf{Z})) \end{aligned}$$

where 'dual' stands for graded dual.

13. Frobenius and Verschiebung and other endomorphisms of Λ and the Witt vectors.

After the rather abstract stuff of the two previous sections, it is a pleasure to return in this one to such concrete things as ghost components of power series and Witt vectors.

There are a vast number of functorial operations on the functorial rings $\Lambda(A) \cong W(A)$ which is no surprise as every ring endomorphism of **Symm** induces such a functorial operation; and, as a free ring on countably many generators there are very many ring endomorphisms. Things become different if one also requires such things as preservation of the Abelian group structure on the rings $W(A)$ or preservation of the full functorial ring structure or one requires other interesting properties.

Probably the easiest to describe are the socalled Verschiebung operations.

13.1. *Verschiebung*. The Verschiebung[50] operators are defined on the $\Lambda(A) = 1 + tA[[t]]$ by

$$\mathbf{V}_n\colon \Lambda(A) \longrightarrow \Lambda(A), \quad a(t) \mapsto a(t^n) \tag{13.2}$$

To see what this means at the ghost component level, apply $t\frac{d}{dt}\log$. Now if

$$\begin{aligned} \frac{d}{dt}\log a(t) &= p_1 + p_2 t + p_3 t^2 + \cdots, \\ t\left(\frac{d}{dt}\log a(t^n)\right) &= t(nt^{n-1}(p_1 + p_2 t^n + p_3 t + \cdots)) \end{aligned}$$

and so, on the ghost component level Verschiebung is given by

$$\mathbf{V}_n p_r = \begin{cases} n p_{r/n} & \text{if } r \text{ is divisible by } n \\ 0 & \text{if } r \text{ is not divisible by } n \end{cases} \tag{13.3}$$

This corresponds to the ring endomorphism of **Symm** (also denoted by $\mathbf{V}_n$)

$$\mathbf{V_n}\colon \mathbf{Symm} \longrightarrow \mathbf{Symm}, \quad h_r \mapsto \begin{cases} h_{r/n} & \text{if } r \text{ is divisible by } n \\ 0 & \text{if } r \text{ is not divisible by } n \end{cases} \tag{13.4}$$

13.5. *Proposition*. The Verschiebung morphisms define additive functorial endomorphisms of the $\Lambda(A)$. That is they are Hopf algebra endomorphisms of **Symm**.

This is obvious from the definition. It can also be checked by the usual ghost component argument using (13.3).

[50] The word means 'shift' and comes for the German. So it should indeed be written with an upper case first letter. In the case of the p-adic Witt vectors it was introduced by Witt.

13.6. *Frobenius operations*. For the Frobenius operations on the $\Lambda(A)$ use again the symmetric function point of view. So again write the universal power series $h(t)$ as

$$h(t) = 1 + h_1 t + h_2 t^2 + \cdots = \prod_i (1 - \xi_i t)^{-1} \tag{13.7}$$

and define

$$\mathbf{f}_n h(t) = \prod_i (1 - \xi_i^n t)^{-1} \tag{13.8}$$

Formally, of course, one observes that the right hand side of (13.8) is symmetric in ξ. So there are universal polynomials $Q_{n,1}(h), Q_{n,2}(h), \cdots \in \mathbf{Symm} = \mathbf{Z}[h_1, h_2, \cdots]$ for the coefficients in the power series (13.8) and now define

$$\mathbf{f}_n a(t) = 1 + Q_{n,1}(a)t + Q_{n,2}(a)t^2 + \cdots \tag{13.9}$$

The sequences of polynomials $Q_n(h) = (Q_{n,1}(h), Q_{n,2}(h), \cdots)$ that define the Frobenius operations are determined by

$$w_r(Q_n(h)) = w_{nr}(h) \tag{13.10}$$

In particular

$$Q_{n,1}(h) = w_n(h) \tag{13.11}$$

and

$$Q_{p,p^r}(h) \text{ only involves the variables } h_1, h_p, \cdots, h_{p^{r+1}} \tag{13.12}$$

The higher $Q_{n,r}(h)$ are difficult to write down explicitely.

The Frobenius Hopf algebra endomorphisms of **Symm** corresponding to the Frobenius operations are (of course) given by $h_r \mapsto Q_{n,r}(h)$.

For later purposes it is useful to have a little explicit knowledge about these polynomials.

13.13. *Lemma*. The polynomials $Q_{n,r}(h)$ determining the Frobenius operations satisfy

$$Q_{p,p^r}(h) \equiv \left(h_{p^r}\right)^p \bmod p, \text{ for } p \text{ a prime number} \tag{13.14}$$

$$Q_{n,r}(h) \equiv n h_{rn} \bmod (h_1, h_2, \cdots, h_{rn-1}) \tag{13.15}$$

Both statements are proved by induction. From (13.11) it follows that (13.15) holds for $r = 1$. Suppose with induction that it holds for all $1 \leq i < r$. By the defining property and this induction assumption $w_r(Q_n) \equiv r Q_{n,r} \bmod (h_1, h_2, \cdots, h_{nr-1})$.

On the other hand $w_{nr}(h) \equiv nr h_{nr} \bmod (h_1, h_2, \cdots, h_{nr-1})$. Comparing these two congruences proves (13.15).

Further (13.14) holds for $r = 0$, again by (13.11). Suppose (13.14) holds for $0 \leq i < r$. It follows that

$$p^i \left(Q_{p,p^i}\right)^{p^{r-i}} \equiv p^i \left(h_{p^{i-1}}\right) p^{r-i+1} \tag{13.16}$$

and consequently the first r terms of

$$w_{p^r}(Q_p) = (Q_{p,1})^{p^r} + p(Q_{p,p})^{p^{r-1}} + \cdots + p^{r-1}(Q_{p,p^{r-1}})^p + p^r Q_{p,p^r}$$

Cancel mod p^r with the first r terms of

$$w_{p^{r+1}}(h) = (h_1)^{p^{r+1}} + p(h_p)^{p^r} + \cdots p^{r-1}(h_{p^{r-1}})^{p^r} + p^r(h_{p^r})^p + p^{r+1}h_{p^{r=1}}$$

leaving

$$p^r Q_{p,p^r} \equiv p^r\left(h_{p^r}\right)^p + p^{r+1}h_{p^{r+1}}$$

proving (13.14).

In the same formal vein let ζ_n be a primitive n-th root of unity. Then always

$$\prod_{j=1}^{n}\left(1 - \zeta_n^j t^{1/n}\right) = (1-t)$$

and so the Frobenius operation can also be written as

$$\mathbf{f}_n a(t) = \prod_{j=1}^{n} a\left(\zeta_n^j t^{1/n}\right) \tag{13.17}$$

The ghost components of $h(t)$ as in (13.7) are the power sums in the ξ and so on the level of ghost components the Frobenius operations are characterized by

$$\mathbf{f}_n p_r = p_{nr} \tag{13.18}$$

The Frobenius operations $\mathbf{f}_p$ for p a prime number have a Frobenius like property:

$$\mathbf{f}_p a(t) \equiv a(t)^{*p} \bmod p \tag{13.19}$$[51]

Here ^{*p} means taking the p-th power in the ring $\Lambda(A)$. The congruence (13.20) takes place in $\Lambda(A)$, so it means that there is a power series $b(t) \in \Lambda(A)$ such that $b(t)^p(\mathbf{f}_p a(t)) = a(t)^{*p}$.

To prove this first observe that for power series of the form $(1-xt)^{-1}$ by the definition of Frobenius and product

$$\mathbf{f}_p(1-xt)^{-1} = (1-x^p t)^{-1} = ((1-xt)^{-1})^{*p} \tag{13.20}$$

Now in any commutative ring, including $\Lambda(A)$, $(y+z)^p \equiv y^p + z^p \bmod p$ and so from (13.20) it follows that (13.19) holds for any finite product power series

$$(1-\xi_1 t)^{-1}\cdots(1-\xi t)^{-1}$$

[51] Let $K/\mathbf{Q}_p$ be an unramified extension of the p-adic numbers with ring of integers A. Then the Galois group $Gal(K/\mathbf{Q}_p)$ is cyclic with a generator σ. Let k be the residue field of K. Then, as extensions of fields $A/\mathbf{Z}_p = W_{p^\infty}(k)/W_{p^\infty}(\mathbf{F}_p)$ and σ is the Frobenius endomorphism of $W_{p^\infty}(k)$. It is characterized by $\sigma(a) \equiv a^p \bmod p$. This has to do with the terminology employed here.

(and in fact under the right circumstances (such as here) also for infinite products).

Alternatively, for an arbitrary series take its Witt vector coordinate expression

$$a(t) = \prod_{d=1}^{\infty} \left(1 - x_d t^d\right)^{-1} \tag{13.21}$$

This stabilizes in that the first n coefficients of $a(t)$ only depend on the first n x-coordinates. As (13.20) already holds for all factors on the right of (13.21) one gets (13.19) in full generality by a simple limit argument.

13.22. *Theorem.* The $\mathbf{f}_n$:**Symm** $\longrightarrow$ **Symm** induce functorial ring endomorphisms on the rings $\Lambda(A)$. They are also the only ring endomorphisms of **Symm** that do so.

13.23. *Corollary.* The only ring automorphism of **Symm** that preserves its coring object structure (as an object in the category of rings) is the identity.

I.e. the object **Symm** is rigid as a coring object in the category **CRing**. Note that in contrast to the Liulevicius theorem 10.8, homogeneity is not needed.

13.24. *Proof of the corollary.* The Frobenius morphisms $\mathbf{f}_n$ satisfy $\mathbf{f}_n p_r = p_{nr}$. So the only one that induces an operation that is always injective is $\mathbf{f}_1 = \mathrm{id}$ (as is seen by looking at the operation on any $\mathbf{Q}$-algebra).[52]

13.25. *Proof of theorem* 13.22. For the first statement of the theorem is suffices to verify this on the ghost components, that is the power sums p_r. On these the sum comultiplication, sum counit, product comultiplication, product counit are respectively given by

$$\begin{aligned} &\mu_s(p_r) = 1 \otimes p_r + p_r \otimes 1, \varepsilon_s(p_r) = 0, \\ &\mu_P(p_r) = p_r \otimes p_r, \quad \varepsilon_P(p_r) = 1 \end{aligned} \tag{13.26}$$

So the characterizing property (13.18) of the Frobenius morphism proves that they preserve the functorial ring structure on the $\Lambda(A)$.

Now let φ be a ring endomorphism of **Symm** that respects the structures (13.26). In particular that means that φ must take primitives of the Hopf algebra **Symm** into primitives. The space of primitives of **Symm** consists of the linear (integer) combinations of the p_r. So the image of p_r under φ is of the form

$$\varphi(p_r) = \sum c_{rj} p_j \tag{13.27}$$

[52] Incidentally, all Frobenius operations except the identity are also not injective on the integers. Indeed let p be a prime number and consider the vector $a = (a_1, a_2, a_3, \cdots) \in \mathbf{Z}^{\mathbf{N}}$ with $a_n = p$ if n is not a multiple of p, and $a_n = 0$ if n is a multiple of p. Then by the ghost-vector criterium in 9.93 (with $A = \mathbf{Z}$ and all the endomorphisms the identity) this vector is a ghost-Witt vector i.e. there is a Witt vector $x \in W(\mathbf{Z})$ such that $w(x) = a$. For this vector $\mathbf{f}_p(x) = 0$.

Now use $\mu_P(\varphi(p_r)) = (\varphi\otimes\varphi)\mu_P(p_r) = \varphi(p_r)\otimes\varphi(p_r)$ and $1 = \varphi(1) = \varphi(\varepsilon_P(p_r)) = \varepsilon_P(\varphi(p_r))$ to find that the coefficients c_{rj} must satisfy

$$c_{rj}c_{rj} = c_{rj}, \quad c_{rj}c_{rj'} = 0 \text{ if } j \neq j', \quad \sum_j c_{rj} = 1$$

and this is only possible if all but one of these coefficients are zero and that last one is equal to 1. Thus there is a mapping $\sigma : \mathbf{N} \longrightarrow \mathbf{N}$ such that

$$\varphi(p_r) = p_{\sigma(r)} \tag{13.28}$$

Now consider the Newton relations (9.57):

$$p_r = rh_r - (p_1h_{r-1} + p_2h_{r-2} + \cdots + p_{r-1}h_1) \tag{13.29}$$

It easily follows with induction that

$$p_r \equiv (-1)^{r+1}h_1^r \bmod(h_2, h_3, \cdots) \tag{13.30}$$

Let $n = \sigma(1)$ and suppose with induction that it has been shown that

$$\varphi(p_u) = p_{nu}, \varphi(h_u) \equiv 0 \bmod(h_2, h_3, \cdots) \quad \text{for } r-1 \geq u \geq 2 \tag{13.31}$$

Now apply φ to the Newton relation (13.29) and use (13.28), (13.30) and the induction hypothesis (13.31) to find that

$$\begin{aligned}(-1)^{\sigma(r)+1}h_1^{\sigma(r)} &\equiv \varphi(p_r)\\ &\equiv r\varphi(h_r) - ((-1)^{(r-1)n+1}h_1^{(r-1)n})(-1)^{n+1}h_1^n) \bmod(h_2, h_3, \cdots)\end{aligned}$$

and so $\sigma(r) = nr$ and $\varphi(h_r) \equiv 0 \mod(h_2, h_3, \cdots)$. It follows that $\varphi = \mathbf{f}_n$.

13.32. Note that the proof does not use any integrality of coefficients statement and so the theorem still holds over the rationals. Indeed it works over any ring and so the theorem is true for the coring object algebras $\mathbf{Symm}_R = \mathbf{Symm} \otimes R$ over any ring R.

13.33. *Frobenius endomorphisms of* **QSymm**. The Frobenius morphisms $\mathbf{f}_n$ on **Symm** extend to Frobenius endomorphisms on **QSymm** $\supset$ **Symm**. The formula is simple

$$\mathbf{f}_n([a_1, a_2, \cdots, a_m]) = [na_1, na_2, \cdots, na_m] \tag{13.34}$$

It is obvious from the original definitions of the first and second comultiplication, and the description of the overlapping shuffle product that these are Hopf algebra endomorphisms that preserve the second comultiplication and second co-unit. They are also the only ring endomorphisms of **QSymm** that do so.

To prove this consider a Hopf algebra endomorphism φ of **QSymm** that preserves the second comultiplication. First it must take primitives into primitives. The primitives are spanned by the compositions of length 1, see 11.36. Further

$$\mu_P([a]) = [a] \otimes [a], \quad \varepsilon_P([a]) = 1$$

and so, exactly as above, there is a mapping σ: $\mathbf{N} \longrightarrow \mathbf{N}$ such that

$$\varphi([a]) = [\sigma(a)] \tag{13.35}$$

Let $\sigma(1) = n$ and let J_2 be the ideal in **QSymm** spanned by all compositions of length 2 or more. This is indeed an ideal. Now also $\varepsilon_P \varphi = \varepsilon_P$ and it follows that $\varphi(J_2) \subset J_2$. Further $[r] \equiv [1]^r \bmod J_2$ and so

$$\varphi([r]) \equiv \varphi([1]^r) = \varphi([1])^r = [n]^r \equiv [nr]$$

and so φ is equal to $\mathbf{f}_n$ on the submodule spanned by the [a]. Now consider a composition of length 2. As φ respects the first comultiplication

$$\begin{aligned} \mu_S(\varphi([a,b])) &= (\varphi \otimes \varphi)(1 \otimes [a,b] + [a] \otimes [b] + [a,b] \otimes 1) \\ &= 1 \otimes \varphi([a,b]) + [na] \otimes [nb] + \varphi([a,b]) \otimes 1 \end{aligned} \tag{13.36}$$

The term $[na] \otimes [nb]$ on the right of (13.36) can only come under μ_S from the composition $[na, nb]$ and if there were any other compositions involved in $\varphi([a, b])$ that would show up after applying μ_S. So $\varphi([a, b]) = [na, nb]$. This argument easily continues.

13.37. *Corollary*. The Frobenius operators on **Symm** are given on the monomial symmetric functions by

$$\mathbf{f}_n m_{(\lambda_1, \lambda_2, \cdots, \lambda_m)} = m_{(n\lambda_1, n\lambda_2, \cdots, n\lambda_m)} \tag{13.38}$$

There are (of course) also other ways of seeing this, but this is a particularly elegant way. Thus on the monomial symmetric functions the Frobenius operators are easy to describe. It is far more difficult to describe them on the h_λ, though of course there are the messy formulas giving the h's in terms of the p's. Dually the Verschiebung operators are easy to describe on the h's and difficult to describe on the monomial symmetric functions.

13.39. *Corollary*. There are no second multiplication preserving Hopf algebra endomorphisms of **NSymm** that descend to a Frobenius endomorphism of **Symm** as a quotient of **NSymm** except the identity.

Indeed such a morphism would involve some degree increasing part. Its dual would therefore involve some degree increasing part and be a second comultiplication preserving Hopf algebra endomorphism of **QSymm**, which is impossible by 13.25 above.

This gives another negative answer, as conjectured, to a question posed in [200, 201].

13.40. The Frobenius morphisms $\mathbf{f}_p$, for p a prime number also satisfy the Frobenius like property

$$[a_1, a_2, \cdots, a_m]^p \equiv [pa_1, pa_2, \cdots, pa_m] \bmod p \tag{13.41}$$

This is more or less immediate from the definition of the overlapping shuffle product. Indeed, the terms of this p-th power are the column sums of all matrices without zero columns of which the rows are obtained from $a_1, a_2, \cdots, a_m$ by interspersing zeros. If all the rows are equal the column sum $[pa_1, pa_2, \cdots, pa_m]$ results. If not all rows

are equal any (non identity) cyclic permutation yields another different matrix of the same type (this uses that p is prime) and so all the other terms in the shuffle power occur with a coefficient divisible by p.

13.42. *'Multiplication by n' operator.* Consider the operator of adding (in $\Lambda(A)$ or $W(A)$) an element to itself n times, i.e. in $\Lambda(A)$ taking the n-th power of a power series. This operator is denoted

$$[n] : \Lambda(A) \longrightarrow \Lambda(A), a(t) \mapsto a(t)^n \tag{13.43}$$

This operator as an endomorphism of **Symm** is given by the composition of maps

$$\mathbf{Symm} \xrightarrow{\mu_n} \mathbf{Symm}^{\otimes n} \xrightarrow{m_n} \mathbf{Symm} \tag{13.44}$$

where μ_n is the n-fold first comultiplication as defined in (12.14) and m_n is the n-fold multiplication. This composed map can be written down for any Hopf algebra H. In general it is not a Hopf algebra endomorphism, nor even an algebra or coalgebra morphism unless H is commutative and cocommutative. Then it is a Hopf algebra endomorphism. In spite of not being a Hopf algebra endomorphism in general these maps have proved to be most useful in various investigations in Hopf algebra theory, see e.g. [323].

13.45. *Homothety operations.* Consider the Hopf algebra $\mathbf{Symm}_R = \mathbf{Symm} \otimes R$ over a ring R. For every R-algebra A and every $u \in R$ consider the operation

$$\langle u \rangle a(t) = a(ut),\ a(t) \in \Lambda(A) \tag{13.46}$$

These are the homothety operators and they clearly define additive functorial Abelian group endomorphisms of the $\Lambda(A)$. The associated Hopf algebra endomorphism of $\mathbf{Symm}_R$ is

$$\langle u \rangle(h_n) = u^n h_n \tag{13.47}$$

As will be seen later the homothety, Verschiebung, and Frobenius endomorphisms together, in a very precise sense, generate all R-Hopf algebra endomorphisms of $\mathbf{Symm}_R$.

There are quite a good many relations among all these operators. They are summed up in the following theorem.

13.48. *Theorem.* There are the following identifies between operations on the functors $\Lambda_R(-)$, $W_R(-)$, respectively the Hopf algebra $\mathbf{Symm}_R$.

$$\langle u \rangle \langle u' \rangle = \langle uu' \rangle \tag{13.49}$$

$$\langle 1 \rangle = \mathbf{f}_1 = \mathbf{V}_1 = \mathrm{id} \tag{13.50}$$

$$\mathbf{V}_m \mathbf{V}_n = \mathbf{V}_{nm} \tag{13.51}$$

$$\mathbf{f}_m \mathbf{f}_n = \mathbf{f}_{mn} \tag{13.52}$$

$$\text{if } \gcd(m,n) = 1, \text{ then } \quad \mathbf{f}_m \mathbf{V}_n = \mathbf{V}_n \mathbf{f}_m \tag{13.53}$$

$$\mathbf{f}_n \mathbf{V}_n = [n] \tag{13.54}$$

$$\mathbf{V}_n(a(t) * \mathbf{f}_n b(t)) = (\mathbf{V}_n a(t)) * b(t), \quad a(t), b(t) \in \Lambda(A) \tag{13.55}$$

$$\langle \mathbf{f}_n x, y \rangle = \langle x, \mathbf{V}_n y \rangle, \quad x, y \in \mathbf{Symm} \tag{13.56}$$

$$\langle u \rangle \mathbf{V}_n = \mathbf{V}_n \langle u^n \rangle \tag{13.57}$$

$$\mathbf{f}_n \langle u \rangle = \langle u^n \rangle \mathbf{f}_n \tag{13.58}$$

$$\langle u \rangle + \langle v \rangle = \sum_{n=1}^{\infty} \mathbf{V}_n \langle r_n(u, v) \rangle \mathbf{f}_n \tag{13.59}$$

where the $r_d(X, Y)$ are the integer coefficient polynomials in two variables determined by

$$(1 - Xt)(1 - Yt) = \prod_{d=1}^{\infty} (1 - r_d(X, Y) t^d) \tag{13.60}$$

These last two, (13.59), (13.60) constitute of course the formula for the addition of Teichmüller representatives in Witt coordinates. Applying $t \frac{d}{dt} \log$ to the two sides of (13.60) it follows that the polynomials $r_j(X, Y)$ are determined by the relation

$$X^n + Y^n = \sum_{d|n} d r_d(X, Y)^{n/d} \tag{13.61}$$

For example

$$r_1(X, Y) = X + Y$$

$$r_2(X, Y) = -XY$$

$$r_3(X, Y) = -(X^2 Y + XY^2)$$

$$r_4(X, Y) = -(X^3 Y + 2X^2 Y^2 + XY^3)$$

$$r_5(X, Y) = -(X^4 Y + 2X^3 Y^2 + 2X^2 Y^3 + XY^4)$$

$$r_6(X, Y) = -(X^5 Y + 3X^4 Y^2 + 4X^3 Y^3 + 3X^2 Y^4 + XY^5)$$

It immediately strikes one's attention that the coefficients in $r_2, r_3 \cdots, r_6$ are all negative. This is true for all $n \geq 2$ as an immediate consequence of the Reutenauer-Scharf-Thibon result 9.71. Another striking fact is that all monomials that possibly can occur do in fact occur with nonzero coefficient. I know of no proof for that but definitely believe it to be true.

13.62. *Caveat.* All these formulas are written from the functorial operations point of view (except (13.56) where the only possible interpretation is in terms of endomorphisms). So, for instance, (13.54) means that for an element $a(t) \in \Lambda(A)$ there

is the equality

$$\mathbf{f}_n(\mathbf{V}_n(a(t))) = [n](a(t)) \tag{13.63}$$

If U_1, U_2, U are functorial operations and u_1, u_2, u are the endomorphism that induce them, then seeing an element $a(t) \in \Lambda(A)$ as a morphism $\mathbf{Symm} \longrightarrow A$ (as must be done to have this correspondence between endomorphisms and operators), the relation beteen the operation U and the endomorphism u is

$$U(a(t)) = a(t) \circ u \tag{13.64}$$

where the small circle denotes composition. It follows that under the correspondence 'functorial operation' $\Leftrightarrow$ 'endomorphism' the order of composition reverses. Indeed

$$U_1(U_2(a(t))) = U_1(a(t) \circ u_2) = (a(t) \circ u_2) \circ u_1 = a(t) \circ (u_2 \circ u_1)$$

13.65. *Proof of theorem* 13.48. Most of these are pretty trivial using the ghost component formalism (see also 5.16). As before for any natural number r and element $a(t) \in \Lambda(A)$ let

$$s_r(a(t)) = \text{coefficient of } t^r \text{ in } t\frac{d}{dt}\log(a(t)) \tag{13.66}$$

Then the characterizing ghost component description of the various functorial operations are

$$s_r(\mathbf{V}_n a(t)) = \begin{cases} n s_{r/n}(a(t)) & \text{if } n \text{ divides } r \\ 0 & \text{if } n \text{ does not divide } r \end{cases} \tag{13.67}$$

$$s_r(\mathbf{f}_n a(t)) = s_{nr}(a(t)) \tag{13.68}$$

$$s_r(\langle u \rangle a(t)) = u^r s_r(a(t)) \tag{13.69}$$

$$s_r([n]a(t)) = n s_r(a(t)) \tag{13.70}$$

So, using this, here is a proof of (13.54):

$$s_r(\mathbf{f}_n \mathbf{V}_n a(t)) = s_{rn}(\mathbf{V}_n a(t)) = n s_r(a(t)) = s_r([n]a(t))$$

using (13.68), (13.67), (13.70) in the order named. And here is how (13.59) is tackled with the ghost component formalism

$$\begin{aligned} s_r(\mathbf{V}_n \langle r_n(u, v) \rangle \mathbf{f}_n a(t) &= \begin{cases} n s_{r/n}(\langle r_n(u, v) \rangle \mathbf{f}_n a(t)) & \text{if } n \text{ divides } r \\ 0 & \text{if } n \text{ does not divide } r \end{cases} \\ &= \begin{cases} n r_n(u, v)^{r/n} s_{r/n}(\mathbf{f}_n a(t)) & \text{if } n \text{ divides } r \\ 0 & \text{if } n \text{ does not divide } r \end{cases} \\ &= \begin{cases} n r_n(u, v)^{r/n} s_r(a(t)) & \text{if } n \text{ divides } r \\ 0 & \text{if } n \text{ does not divide } r \end{cases} \end{aligned}$$

and so

$$s_r\left(\sum_{n=1}^{\infty}{}^{\Lambda}\mathbf{V}_n\,\langle r_n(u,v)\rangle\,\mathbf{f}_n(a(t))\right)=\sum_{n|r}nr_n(u,v)^{r/n}s_r(a(t))=(u^n+v^n)s_r(a(t))$$

$$=s_r\left(\langle u\rangle\,a(t)+_\Lambda\langle v\rangle\,a(t)\right)$$

where of course, as indicated, the sum of elements in $\Lambda(A)$ is taken according to the addition law in the Abelian group $\Lambda(A)$.

Here is the proof of (13.53) as a further illustration of ghost component techniques

$$s_r(\mathbf{f}_m\mathbf{V}_n a(t))=s_{rm}(\mathbf{V}_n a(t))=\begin{cases}s_{rm/n}(a(t)) & \text{if } n \text{ divides } rm\\ 0 & \text{otherwise}\end{cases}$$

and as $(m,n)=1$ the divisibility condition in the formula above is the same as 'n divides r', which is what turns up when calculating $s_r(\mathbf{V}_n\mathbf{f}_m a(t))$.

All the other statements of theorem 13.48 are proved the same way, except of course the duality statement (13.56). For this statement use the orthonormal dual pair of bases $\{h_\lambda\}$, $\{m_\lambda\}$, see (9.40), (9.42). Let $\kappa=(\kappa_1,\kappa_2,\cdots,\kappa_m)$, $\lambda=(\lambda_1,\lambda_2,\cdots,\lambda_n)$ be two partitions. Recall that

$$\mathbf{f}_r m_\lambda=m_{(r\lambda_1,r\lambda_2,\cdots,r\lambda_n)},\ \mathbf{V}_r h_\kappa=\begin{cases}h_{(\kappa_1/r,\kappa_2/r,\cdots,\kappa_m/r)} & \text{if all } \kappa_j \text{ are divisible by } r\\ 0 & \text{otherwise}\end{cases}$$

Thus

$$\langle\mathbf{f}_r m_\lambda,h_\kappa\rangle=\begin{cases}1 & \text{if } \lg(\lambda)=\lg(\kappa) \text{ and } \kappa_i=r\lambda_i \text{ for all } i\\ 0 & \text{otherwise}\end{cases}\tag{13.71}$$

$$\langle m_\lambda,\mathbf{V}_r h_\kappa\rangle=\begin{cases}1 & \text{if } \lg(\lambda)=\lg(\kappa),\ r \text{ divides all } \kappa_i \text{ and } \kappa_i/r=\lambda_i \text{ for all } i\\ 0 & \text{otherwise}\end{cases}\tag{13.72}$$

comparing (13.71) and (13.72) now gives the duality formula (13.56)

13.73. *Convention.* Thus from (13.56) the Frobenius and Verschiebung endomorphisms of $\mathbf{Symm}=\mathbf{Z}[h]$ are (graded) dual to each other. But **Symm** is also isomorphic to its graded dual. That makes it a trifle difficult to agree what should be called Frobenius and what Verschiebung. The convention is that the degree nondecreasing endomorphisms (of these) are always called Frobenius and the degree nonincreasing ones Verschiebung. Both are Hopf algebra endomorphisms.

Then the Frobenius endomorphisms always respect the second comultiplication (and the corresponding co-unit), but not the second multiplication, and thus induce functorial ring endomorphisms of the $W(A)\cong\Lambda(A)=\mathbf{CRing}(\mathbf{Symm},A)$, while the Verschiebung endomorphisms respect the second multiplication (and corresponding unit), but not the second comultiplication, and thus induce ring endomorphisms on the rings of coalgebra morphisms $\mathbf{CoAlg}(C,\mathbf{Symm})$.

13.74. *Ring of Hopf endomorphisms of* **Symm**. Naturally the next step is to try to determine the full ring of Hopf algebra endomorphisms of **Symm**, which is the same as the determination of all additive functorial operations on the functor Λ. Consider again, see 12.11ff, the cofree coalgebra over $\mathbf{Z}$, i.e. the coalgebra

$$\mathrm{CoFree}(\mathbf{Z}) = \bigoplus_{n=0}^{\infty} \mathbf{Z}h_n,\ h_0 = 1,\ \mu(h_n) = \sum_{i+j=n} h_i \otimes h_j,\ \varepsilon(h_n) = 0 \quad \text{for } n \geq 1 \tag{13.75}$$

Now $\mathbf{Symm} = \mathbf{Z}[h]$ is the free algebra on Cofree($\mathbf{Z}$), more precisely it is the free algebra over the augmentation submodule $\bigoplus_{n=1}^{\infty} \mathbf{Z}h_n \subset \mathrm{CoFree}(\mathbf{Z})$ and hence a Hopf algebra endomorphism of **Symm** is the same thing as a morphism of coalgebras from CoFree($\mathbf{Z}$) to **Symm** that takes h_0 to 1.

$$\mathbf{Hopf}(\mathbf{Symm}, \mathbf{Symm}) \cong \mathbf{CoAlg}'(\mathrm{CoFree}(\mathbf{Z}), \mathbf{Symm}) \tag{13.76}$$

where the prime indicates that only morphisms that take h_0 to 1 are permitted. In more pedestrian terms, as $\mathbf{Symm} = \mathbf{Z}[h]$ is free on the generators h_n an algebra endomorphism φ of **Symm** is the same thing as specifying a sequence of polynomials $\varphi(h_n)$ and this sequence yields an endomorphism of Hopf algebras if and only if the sequence $1, \varphi(h_1), \phi(h_2), \varphi(h_3), \cdots$ is a curve (= divided power sequence).

The graded dual of CoFree($\mathbf{Z}$) is $\mathbf{Z}[T]$ the ring of polynomials in one variable T.

Thus, taking graded duals, and using that **Symm** and its graded dual are isomorphic as Hopf algebras (and even as coring objects in the category of rings and as ring objects in the category of coalgebras) one sees that

$$\mathbf{CoAlg}(\mathrm{CoFree}(\mathbf{Z}), \mathbf{Symm}) \cong \mathbf{CRing}(\mathbf{Symm}, \mathbf{Z}[T]) \tag{13.77}$$

and thus, tracing out what the prime in (13.68) means for the right hand side of (13.69),

$$\mathbf{Hopf}(\mathbf{Symm}, \mathbf{Symm}) \cong \mathbf{CRing}'(\mathbf{Symm}, \mathbf{Z}[t]) \tag{13.78}$$

where this time the prime means that only morphisms are allowed that take the degee ≥ 1 part of **Symm** into the degree ≥ 1 part of $\mathbf{Z}[T]$.

But his last object is easy to describe: an element of $\mathbf{CRing}'(\mathbf{Symm}, \mathbf{Z}[t])$ is simply an infinite sequence of polynomials in T with constant terms zero. Things go exactly the same way for $\mathbf{Symm}_A = \mathbf{Symm} \otimes A$.

As the isomorphism between **Symm** and its graded dual takes the h_n into the e_n it is best to see this sequence of polynomials as the images of the e_n.

Thus in this way a Hopf algebra endomorphism φ of $\mathbf{Symm}_A$ is exactly the same thing as an infinite $\times$ infinite matrix with coefficients in A

$$M_\varphi = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots \\ a_{21} & a_{22} & a_{23} & \cdots \\ a_{31} & a_{32} & a_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

such that in each row there are but finitely many coefficients that are nonzero.

It remains to figure out how these matrices are to be added and multiplied (composed) when interpreted as encodings of Hopf endomorphisms of the Hopf algebra **Symm**. That will be the business of section 15 below.

To do something similar for the characteristic p case as well (and the p-adic case) one needs to figure out how the p-adic Witt vectors fit with the power series point of view. That is the subject matter of the next section.

14. Supernatural and other quotients of the big Witt vectors

It is when trying to figure out how the p-adic Witt vectors fit with the power series point of view of sections 9-13 that the Witt vector coordinates have a decided advantage. Recall that in terms of Witt vector coordinates, 9.63, as a set

$$W(A) = \{(x_1, x_2, x_3, \cdots) : x_n \in A\} \tag{14.1}$$

and that two Witt vectors are added and multiplied by means of universal polynomials with integer coordinates

$$\begin{aligned} x +_W y &= (\mu_{S,1}(x, y), \mu_{S,2}(x, y), \mu_{S,3}(x, y), \cdots) \\ x \times_W y &= (\mu_{P,1}(x, y), \mu_{P,2}(x, y), \mu_{P,3}(x, y), \cdots) \end{aligned} \tag{14.2}$$

where $x = (x_1, x_2, x_3, \cdots)$, $y = (y_1, y_2, y_3, \cdots)$, and where the polynomials $\mu_{S,i}$, $\mu_{P,i}$ are recursively given by

$$\begin{aligned} w_n(\mu_{S,1}(X; Y), \mu_{S,2}(X; Y), \mu_{S,3}(X; Y), \cdots) &= w_n(X) + w_n(Y) \\ w_n(\mu_{P,1}(X; Y), \mu_{P,2}(X; Y), \mu_{P,3}(X; Y), \cdots) &= w_n(X)w_n(Y) \end{aligned} \tag{14.3}$$

with

$$w_n(X) = \sum_{d|n} dX_d^{n/d} \tag{14.4}$$

The important fact to notice is now that $w_n(X)$ depends only on the X_d for d a divisor of n, and hence that the n-th addition and multiplication polynomials $\mu_{S,n}$, $\mu_{P,n}$ are polynomials that only involve the X_d and Y_d with d a divisor of n. Thus, for instance, $\mathbf{Z}[X_1, X_p, X_{p^2}, X_{p^3}, \cdots]$, where p is a prime number, is a sub Hopf algebra and sub coring object of $\mathbf{Z}[X] = \mathbf{Z}[X_1, X_2, X_3, \cdots]$, which means that it defines a quotient functor of W, which is manifestly W_{p^∞}, so that the p-adic Witt vectors are a functorial quotient of the big Witt vectors. There are lots more of such quotient functors.

14.5. *Nests*. A nest is a nonempty subset N of the natural numbers $\mathbf{N}$ such that together with any $n \in N$ all the divisors of n are also in N. Some simple examples of nests are

$$\begin{aligned} &\text{for any given fixed } n \in \mathbf{N}, \quad \text{the set } \{1, 2, 3, \cdots, n\} \\ &p^\infty = \{1, p, p^2, p^3, \cdots\} \\ &\text{for any fixed } n \in \mathbf{N}, \text{ the set } \{d : d \text{ divides } n\} \end{aligned} \tag{14.6}$$

Unions and intersections of nests are nests. Every nest contains the natural number 1.

14.7. *Supernatural numbers.* A supernatural number is a formal expression of the form

$$\mathfrak{n} = \prod_p p^{\alpha_p} \quad \alpha_p \in \mathbf{N} \cup \{0, \infty\}$$

where the product is over all prime numbers p. This is exactly the same as specifying for every prime number p an element of the extended natural numbers $\{0\} \cup \mathbf{N} \cup \{\infty\}$. Given a supernatural number $\mathfrak{n}$ there is a nest associated to it (often denoted with the same symbol), viz

$$\mathfrak{n} = N_{\mathfrak{n}} = \{m \in \mathbf{N} \colon v_p(m) \leq \alpha_p \text{ for all prime number} p\}$$

where v_p is the p-adic valuation on the integers. The last two examples in (14.6) are nests coming from a supernatural number. The first example from (14.6) is not of this from (if $n \geq 3$).

14.8. *Supernatural quotients of the big Witt vectors.* Let N be a nest. Let **Symm** $(N) = \mathbf{Z}[X_n : n \in N] \subset$ **Symm**. By the remarks made in the beginning of this section these form sub Hopf algebras of **Symm** and hence define quotient Witt vector functors $W_N(A)$ of the big Witt vectors, called the N-adic Witt vectors. The most important ones are

$W_{p^\infty}(A)$, the p-adic Witt vectors

$W_{p^n}(A)$, the p-adic Witt vectors of length $n+1$

$W_n(A)$, the Witt vectors of length n.

The last named quotient functor is the one defined by the next $\{1, 2, \cdots, n\}$. This notation is not entirely consistent with the other two. However, the quotient Witt vectors defined by the nest $\{d : d|n\}$ are so seldom used (if ever) that this seems justified.

Sometimes the curious appellation "nested Witt vectors" is used for the elements of a $W_N(A)$.

14.9. *Operations on nested Witt vectors.* A first question now is now which of the functorial additive operations $\langle a \rangle$, $\mathbf{V}_r$, $\mathbf{f}_r$ survive to define operations on the various quotients W_N. Here of course these functorial operations are the ones defined by transferring the operations denoted by the same symbols on $\Lambda(A)$. They are therefore characterized by

$$w_r \mathbf{f}_n = w_{rn}, \; w_r \mathbf{V}_n = \begin{cases} w_{r/n} & \text{if } n \text{ divides } r \\ 0 & \text{otherwise} \end{cases}, \quad w_r \langle a \rangle = a^r w_r \qquad (14.10)$$

The answer to the above question of which operations 'descend' to operations on the various W_N is fairly simple.

The homothety operations $\langle a \rangle$ always define a quotient operation.

For the Verschiebung operations the situation is simple. The polynomials defining $\mathbf{V}_n$ are either zero or of the form $X_{r/n}$ for r a multiple of n. So by the nest property they always exist on the quotients W_N. Of course many of them may be zero.

The situation with the Frobenius operators is different and slightly more complicated. Let p be a prime number. The polynomials defining $\mathbf{f}_p$ are degree increasing by a factor of p and because $\mathbf{f}_p$ is characterized by $\mathbf{f}_p(p_n) = p_{pn}$ at the ghost component level, the n-th polynomial defining $\mathbf{f}_p$ involves the indeterminate X_{np}. See also lemma 13.13 above. So for $\mathbf{f}_p$ to descend to a nest quotient in full generality, i.e. to exist for the functor W_N on all of **CRing**, it is necessary and sufficient that the nest contain together with any $n \in N$ also all its p-power multiples $p^i n$. In particular $\mathbf{f}_p$ exists on the p-adic Witt vectors.

However, if one restricts attention to the functor on rings of characteristic p the operation $\mathbf{f}_p$ always exists on the quotients W_N because in that case the Frobenius is given by raising each coordinate of a Witt vector to its p-th power See again lemma 13.13.

14.11. *Möbius function.* The next question is whether there exists lifts, that is an additive functor morphism $W_N \longrightarrow W$ which composed with the quotient map $W \longrightarrow W_N$ gives the identity on W_N. This will involve the Möbius function from number theory and combinatorics. Explicitly this function is defined by

$$\begin{aligned} &\mu(1) = 1 \\ &\mu(n) = (-1)^r \text{ if } n \text{ is the product of } r \text{ different prime numbers} \\ &\mu(n) = 0 \text{ if } n \text{ is divisible by the square of a nontrivial prime number} \end{aligned} \tag{14.12}$$

The function has the characterizing property

$$\sum_{d|n} \mu(d) = 1 \tag{14.13}$$

(which also defines it recursively). Let $\mathbf{N}(p) = \{n \in \mathbf{N}\colon (p, n) = 1\}$ be the set of all natural numbers relatively prime to a given prime number p. It immediately follows from the above that

$$\sum_{d|n, d\in\mathbf{N}(p)} \mu(d) = \begin{cases} 1 & \text{if } n \text{ is a power of } p \\ 0 & \text{otherwise} \end{cases} \tag{14.14}$$

14.15. *Sectioning the projection* $W(A) \longrightarrow W_{p^\infty}(A)$. To set the stage here is the abstract situation. Suppose there is an Abelian group M together with a surjective projection $\pi\colon M \longrightarrow M_p$ to another group M_p. Suppose there is a section, i.e. a morphism of Abelian groups $s\colon M_p \longrightarrow M$ such that $\pi s = id_{M_p}$. Then $\Phi = s\pi$ is an idempotent endomorphism of M with image $s(M_p)$ and kernel $\mathrm{Ker}(\pi)$. Further the projection π induces an isomorphism from $\Phi(M) = s\left(M_p\right)$ to M_p. It is such an idempotent endomorphism Φ that will now be

constructed in the case of the canonical projection

$$\pi_p : W(A) \longrightarrow W_{p^\infty}(A) \tag{14.16}$$

in the case of $\mathbf{Z}_{(p)}$ algebras, i.e. rings A in which all prime numbers except the give prime number p are invertible.

To this end consider the operation

$$\Phi_{p-\mathrm{typ}} = \sum \left[\frac{\mu(n)}{n}\right] \mathbf{V}_n \mathbf{f}_n \tag{14.17}$$

This functorial operation satisfies

$$w_r \Phi_{p-\mathrm{typ}} = \begin{cases} w_r & \text{if } r \text{ is a power of } p \\ 0 & \text{otherwise} \end{cases} \tag{14.18}$$

(and it is characterized by this property, of course). This is an immediate consequence of (14.14) and (14.10). Note that it follows that $\Phi_{p-\mathrm{typ}}$ is idempotent. (First for rings A of characteristic zero and then for all by functoriality).

14.19. *Remark.* $\Phi_{p-\mathrm{typ}}$ commutes with the operations $\langle a\rangle$, $\mathbf{f}_p$, $\mathbf{V}_p$.

14.20. *Comments.* $\Phi_{p-\mathrm{typ}}$ is not easy to write down explicitly, say, in terms of its defining polynomials. It is definitely not something like murdering each coordinate that is not at a p-th power position.

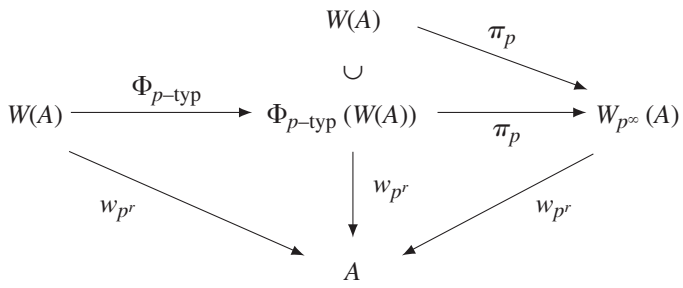
Easy explicit calculations show that it does not preserve the unit element and that it is not multiplicative. This already shows up at the second coordinate for all prime numbers larger than 2 and in the third coordinate for the prime number 2.

14.21. *Proposition.* For each $\mathbf{Z}_{(p)}$-algebra A, a Witt vector $x = (x_1, x_2, x_3, \cdots) \in W(A)$ is in the image of $\Phi_{p-\mathrm{typ}}(A)$: $W(A) \longrightarrow W(A)$ if and only if $\mathbf{f}_{p'} x = 0$ for all prime numbers p' different from p.

First, if $\mathbf{f}_{p'} x = 0$ for all prime numbers different from p, then $\mathbf{f}_n x = 0$ for all $n = \mathbf{N}(p)$, $n \geq 2$ and so $\Phi_{p-\mathrm{typ}} x = [1]x = x$. Second, if A is of characteristic zero and $x \in W(A)$ is in the image of $\Phi_{p-\mathrm{typ}}(A)$ then the characterizing property (14.18) shows that $\mathbf{f}_{p'} x = 0$. Third, suppose that $x \in W(A)$ is in the image $\Phi_{p-\mathrm{typ}}$. Choose any lift $\tilde{x} \in W(\tilde{A})$ of $x \in W(A)$ for some characteristic zero $\mathbf{Z}_{(p)}$-algebra $\tilde{A}$ covering A. Then, by idempotency, $\Phi_{p-\mathrm{typ}}(\tilde{A})\tilde{x}$ is also a lift of x, and so $\mathbf{f}_{p'} \Phi_{p-\mathrm{typ}} \tilde{x} = 0$ implies $\mathbf{f}_{p'} x = 0$.

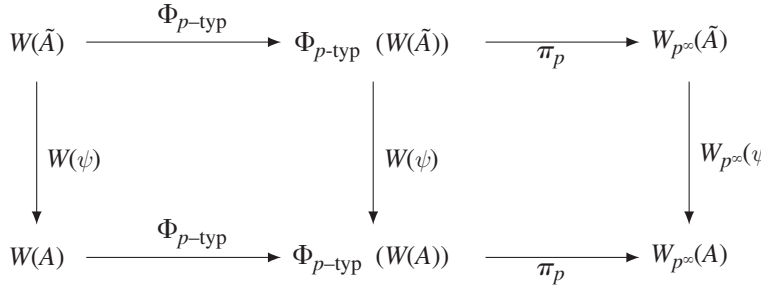
14.21. *Theorem.* For $\mathbf{Z}_{(p)}$-algebras the canonical projection $\pi_p : W(A) \longrightarrow W_{p^\infty}(A)$ induces an isomorphism of Abelian groups $\Phi_{p-\mathrm{typ}}(W(A)) \xrightarrow{\pi_p} W_{p^\infty}(A)$.

Pictorially things are described by the following commutative diagram.



(14.22)

It follows just about immediately that the theorem holds for A of characteristic zero. Now let A



(14.23)

be any $\mathbf{Z}_{(p)}$-algebra and choose a characteristic zero $\mathbf{Z}_{(p)}$-algebra $\tilde{A}$ that covers it with corresponding projection $\psi\colon \tilde{A} \longrightarrow A$ and set $J = \mathrm{Ker}(\psi)$. Consider the commutative diagram (14.23) above.

It follows immediately from the fact that the upper right π_p is an isomorphism that the lower right π_p is surjective. Now let $x \in \Phi_{p-\mathrm{typ}} W(A)$, and take a $y \in W(A)$ that maps to x and take a $\tilde{y} \in W(\tilde{A})$ that lifts it. Then $\tilde{x} = \Phi_{p-\mathrm{typ}}(\tilde{y})$ is a lift of x. Now suppose that $\pi_p(x) = 0$ (lower right hand corner in the diagram). Then $\pi_p(\tilde{x})$ has all its coordinates in J. By the commutativity of diagram (14.22) another, possibly different, lift of x can be obtained by applying $\Phi_{p-\mathrm{typ}}$ to the vector in $W(\tilde{A})$ that has the same coordinates $\pi_p(\tilde{x})$ at the p power spots and zeros elsewhere. Let this vector be $\tilde{x}'$. But the morphism $\Phi_{p-\mathrm{typ}}$ is given by some universal polynomials with coefficients in $\mathbf{Z}_{(p)}$ and so all coordinates of $\tilde{x}'$ are in J, and so $x = W(\psi)\tilde{x}' = 0$. This proves that the lower π_p in diagram (14.23) is also injective.

14.24. The morphism $\Phi_{p-\mathrm{typ}}$ (transferred to the power series case) is an example of 'p-typification.' This is general notion, due to Pierre Cartier, defined for curves in any commutative formal group. The one involved here is the second simplest example, namely the one for the one dimensional multiplicative formal group $\hat{\mathbf{G}}_m$. See [85] and [192], section 15.2.

14.25. *Decomposition.* It follows pretty directly from the preceding that over a $\mathbf{Z}_{(p)}$-algebra each Witt vector can be written uniquely as a sum

$$x = \sum_{n \in \mathbf{N}(p)} \mathbf{V}_n y_n,\ y_n \in \Phi_{p-\mathrm{typ}} W(A) \tag{14.26}$$

of shifted p-typical Witt vectors. And thus for $\mathbf{Z}_{(p)}$-algebras the functor of the big Witt vectors decomposes (as an Abelian group) into a direct product of $\mathbf{N}(p)$ copies of the p-adic Witt vectors.

14.27. *Iterated nested Witt vectors.* Let N and M be two nests. These can be multiplied to form a new nest

$$MN = \{mn : m \in M,\ n \in N\} \subset \mathbf{N} \tag{14.28}$$

There is now, in [28], the theorem (observation) that if $M \cap N = \{1\}$ there is a canonical functorial isomorphism

$$W_{MN}(A) \longrightarrow W_M(W_N(A))$$

which essentially come from the Artin-Hasse exponential $W(-) \longrightarrow W(W(-))$ which is the essential part of the cotriple structure on the functor of the big Witt vectors, see subsections 16.43 and 16.59 below.

The case $M = \{1, p, p^2, p^3, \cdots\}$, $N = \mathbf{N}(p) = (n \in \mathbf{N}\colon (n, p) = 1\}$, so that $MN = \mathbf{N}$ can already be found, more or less, in [336].

This is a good, actually better, substitute for the decomposition (14.26), and it works also for rings that are not $\mathbf{Z}_{(p)}$-algebras

15. Cartier algebra and Dieudonné algebra

Consider again the matter of Hopf algebra endomorphisms of **Symm**, or, equivalently, the matter of determining all functorial additive operations on the Witt vectors $W(A) \cong \Lambda(A)$. More generally consider the Hopf algebra $\mathbf{Symm}_R$ over a ring R and consider the isomorphic functors Λ and W on the category of R-algebras and the additive operations on them, or, equivalently, the Hopf R-algebra endomorphisms of $\mathbf{Symm}_R$. As was already indicated these can be desribed by some infinity $\times$ infinity matrix with only finitely many nonzero entries in each row. See 13.74 above. There are other ways to see such a thing and to describe all Hopf algebra endomorphisms over R.

15.1. $(h-h)$-matrix of a Hopf R-algebra endomorphism of $\mathbf{Symm}_R$. One other way is as follows. Because $\mathbf{Symm}_R$ is free as an algebra over the ring R an algebra endomorphsim φ of it is uniquely specified by a sequence of polynomials, e.g. the images $\varphi(h_n)$ of the free polynomial generators h_n. To be a Hopf algebra endomorphism the sequence of polynomials

$$d_0 = 1, d_1 = \varphi(h_1)\,, d_2 = \varphi(h_2)\,, d_3 = \varphi(h_3), \cdots \tag{15.2}$$

must form a curve (divided power sequence) in $\mathbf{Symm}_R$, which means that

$$\varepsilon_s(d_n) = 0 \quad \text{for } n \geq 1 \quad \text{and} \quad \mu_s(d_n) = \sum_{i+j=n} d_i \otimes d_j \tag{15.3}$$

where μ_s and ε_s are the sum comultiplication and corresponding counit of $\mathbf{Symm}_R$. This means that the d_n have constant terms zero and, I claim, are uniquely specified by the matrix

$$M_{(h-h)}(\varphi) = \left(\left\langle \varphi(h_i), h_j \right\rangle\right)_{i,j} \tag{15.4}$$

I like to call this matrix the $(h-h)$-matrix of the endomorphism φ. There are only finitely many nonzero entries in each row because the $\varphi(h_i)$ are polynomials (and hence bounded in weight).

The proof of the claim is simplicity itself. By the autoduality of $\mathbf{Symm}_R$ one has for example

$$\begin{aligned}\langle \varphi(h_n), h_k h_l \rangle &= \left\langle \sum_{i+j=n} \varphi(h_i) \otimes \varphi(h_j), h_k \otimes h_l \right\rangle \\ &= \sum_{i+j=n} \langle \varphi(h_i), h_k \rangle \left\langle \varphi(h_j), h_l \right\rangle\end{aligned}$$

and thus, with induction, the $(h-h)$-matrix determines the inner products of the polynomials $\varphi(h_n)$ with every monomial in the h's and hence specifies the polynomials themselves completely. One can also of course take any other free polynomial basis for $\mathbf{Symm}_R$, for instance the elementary symmetric functions, which gives the $(h-e)$-matrix of an endomorphism.

It is perhaps interesting to calculate these matrices for some of the more important endomorphisms. For instance the Frobenius and Verschiebung endomorphisms. By definition $\langle h_1, e_1 \rangle = 1$ and $\langle h_n, e_n \rangle = 0$ for $n \geq 2$. It follows with induction using the Newton relations, the fact that $\mathbf{f}_n$ is a ring endomorphism, and duality, that $\langle p_n, e_n \rangle = (-1)^{n+1}$ and hence that $\langle \mathbf{f}_n h_1, e_n \rangle = (-1)^{n+1}$. Using again the Newton relations, ring morphism, and duality and induction one further finds $\langle \mathbf{f}_n h_r, e_{nr} \rangle = 0$. Or use $\langle \mathbf{f}_r h_n, e_{nr} \rangle = \langle h_n, V_r e_{nr} \rangle = \langle h_n, e_n \rangle$ to see this. As all other entries of the matrix must be zero by degree considerations it follows that the $(h-e)$-matrix of the Frobenius Hopf algebra endomorphism $\mathbf{f}_n$ of $\mathbf{Symm}$ has an entry $(-1)^{n+1}$ at spot $(1, n)$ and zeros everywhere else.

Even easier one finds that the $(h-e)$-matrix of the n-th Verschiebung operator has a 1 at spot $(n, 1)$ and zeros everywhere else. This is encouraging and suggests that every matrix of the type specified can arise. And this is indeed the case and can be proved this way. But this is not the easiest or most elegant way to see this.

15.5. *The DE-matrix of a Hopf algebra endomorphism of* $\mathbf{Symm}_R$. Let $R[T]$ be the algebra of polynomials in a single indeterminate over the ring R. The determining example for an additive operation on the functor Λ_R (the functor Λ restricted to the category of R-algebras), and hence for the corresponding Hopf algebra endomorphism of $\mathbf{Symm}_R$, is the element

$$(1 - Tt)^{-1} \in \Lambda(R[T]) \tag{15.6}$$

in the sense that what a functorial operation does to this one example determines it completely. This claim will be examined a bit more closely further on. For the moment the argument is as follows. Let φ be an additive operation and suppose that

$$\varphi((1-Tt)^{-1}) = 1 + a_1(T)t + a_2(T)t^2 + a_3(T)t^3 + \cdots \tag{15.7}$$

where the $a_n(T)$ are polynomials in T with coefficients in R. Because the operation is supposed additive it must take the zero of the Abelian group $\Lambda(R[T])$, which is 1, into itself and so $a_i(0) = 0$ for all i. Then the naive argument goes at follows.

Take any element $b(t) \in \Lambda(A)$ for an R-algebra A. Write it formally as a product

$$b(t) = 1 + b_1 t + b_{21}t^2 + b_3 t^3 + \cdots = \prod_{i=1}^{\infty} (1-\eta_i t)^{-1} \tag{15.8}$$

Then

$$\varphi(b(t)) = \prod_{i=1}^{\infty} (1 + a_1(\eta_i)t + a_2(\eta_i)t^2 + a_3(\eta_i)t^3 + \cdots) \tag{15.9}$$

This is more than a bit shaky. For one thing it is far from clear whether the product on the right hand side exists in a suitable sense. To make sense of things and also for other purposes rewrite the right hand side of 15.7 as a product

$$\varphi((1-Tt)^{-1}) = \prod_{m,n\geq 1} \left(1 - c_{mn}T^n t^m\right)^{-1} \tag{15.10}$$

where for any m there are only finitely many n such that $c_{mn} \neq 0$. This can be done in precisely one way. The finiteness condition comes about precisely because the $a_i(T)$ in (15.7) are polynomials (not power series for instance). The matrix

$$M_{DE}(\varphi) = (c_{mn})_{m,n} \tag{15.11}$$

is the DE-matrix of φ.[53]

15.12. *Proof of the DE principle* (*splitting principle*). Manifestly, by the definition of the Frobenius, Verschiebung and homothety operations the right hand side of (15.10) is the operation

$$\sum_{m,n} \mathbf{V}_m \langle c_{mn}\rangle \mathbf{f}_n \tag{15.13}$$

applied to the determining example power series $(1-Tt)^{-1}$. It remains to show that (15.13) makes unique sense when applied to an arbitrary element $b(t)$. To this end write $b(t)$ in Witt vector coordinates as

$$b(t) = \prod_d (1 - x_d t^d)^{-1} \tag{15.14}$$

[53] DE stands for 'determining example.'

Fix for the moment a power m_0 of t. Because of the finiteness condition there is an n_0 such that $c_{nm} = 0$ for all $m \leq m_0$, $n \geq n_0$. Because of the definition of $\mathbf{V}_m$ the first m_0 coefficients of (15.13) applied to a finite product like (15.14) are determined by the part of (15.13) with $m \leq m_0$. Now an $\mathbf{f}_n$ is given by a series of polynomials of weights $n, 2n, 3n, \cdots$. It follows that $\mathbf{f}_n$ applied to an $(1 - xt^r)^{-1}$ with $r > m_0 n$ gives a result that has the first m_0 coefficients (except the constant term 1) equal to zero. Thus to calculate the first m_0 coefficients of (15.13) applied to (15.14) it suffices to apply it to the finite product

$$\prod_{d \leq m_0 n_0} (1 - x_d t^d)^{-1} = \prod_{i}^{<\infty} (1 - \eta_i t)^{-1} \tag{15.15}$$

where the right hand side of (15.15) is also a finite product. And this is perfectly well defined by additivity and functoriality. This proves the DE principle. This principle, apart form the infinity questions just taken care of, is an algebraic analogue of the splitting principle in algebraic topology as used in, say, topological K-theory.[54]

15.16. *Cartier algebra.* By now it is clear that the ring of all additive operations on the functor is the set of all expressions (15.13) (with the stated finiteness condition on the coefficients involved). This ring, which is also the ring of Hopf R-algebra endomorphisms of the Hopf algebra $\mathbf{Symm}_R$, is called the Cartier algebra on R and denoted $\mathrm{Cart}(R)$.[55]

There are also already a number of calculating rules, as given by the formulas from theorem 13.48 viz formulas (13.49)–(13.54), (13.57)–(13.59). The remaining question is whether these suffice. The most troublesome one (potentially) seems to be adding two expressions like (15.13). So consider a sum

$$\sum_{m,n} \mathbf{V}_m \langle b_{mn} \rangle \mathbf{f}_n + \sum_{m,n} \mathbf{V}_m \langle c_{mn} \rangle \mathbf{f}_n \tag{15.17}$$

Manipulating these directly according to the rules given, which needs especially (13.51), just seems to lead to more and more sums at first sight. More care is needed. To see how things work, calculate what the sum (15.17) perscribes for the first few coefficients, say 2. That means that all terms with a $\mathbf{V}_m$, $m \geq 3$ can be

[54] Suppose one has an additive (in some suitable sense) operation in, say, topological K-theory. To find out what it does to a given bundle over a space take a suitable covering space such that the induced buncle splits into a sum of line bundles. This can always be done. Thus, by functoriality and additivity it suffices to know what the operation does to line bundles. And that in turn is governed (again by functoriality) by what it does to the universal line bundle over infinite dimensional projective space. From this point of view I should have called (15.6) the universal example (overworking the word 'universal').

The 'principle' that under suitable circumstances it suffices to verify properties just for the case of line bundles is (or at least used to be) called the 'verification principle'; see e.g. [27].

[55] When seen as endomorphisms of the Hopf algebra $\mathbf{Symm}_R$ the elements of $\mathrm{Cart}(R)$ need to be written in the form $\sum \mathbf{f}_n \langle c_{mn} \rangle \mathbf{V}_m$. See the 'caveat' 16.62 below. The need for the finiteness condition is perhaps even clearer in this interpretation.

discarded leaving

$$\sum_n \langle b_{1n}\rangle \mathbf{f}_n + \sum_n \mathbf{V}_2 \langle b_{2n}\rangle \mathbf{f}_n + \sum_n \langle c_{1n}\rangle \mathbf{f}_n + \sum_n \mathbf{V}_2 \langle c_{2n}\rangle \mathbf{f}_n \tag{15.18}$$

Now $\langle b_{1n}\rangle + \langle c_{in}\rangle = \sum_{i=1}^{\infty} \mathbf{V}_i \langle r_i(b_{1n}, c_{1n}\rangle \mathbf{f}_i$ and so the $\mathbf{V}_1$ part of the sum (15.17) is given by

$$\sum_n \langle b_{1n} + c_{1n}\rangle \mathbf{f}_n$$

(which is a finite sum). What is left is a sum

$$\mathbf{V}_2 \langle r_2 (b_{1n}, c_{1n})\rangle \mathbf{f}_2 + \sum_n \mathbf{V}_2 \langle b_{2n}\rangle \mathbf{f}_n + \sum_n \mathbf{V}_2 \langle c_{2n}\rangle \mathbf{f}_n$$

and so the $\mathbf{V}_2$ part of (15.17) is given by

$$\mathbf{V}_2 \langle b_{21} + c_{21}\rangle \mathbf{f}_1 + \mathbf{V}_2 \langle (r_2 (b_{12}, c_{12}) + b_{22} + c_{22})\rangle \mathbf{f}_2 + \sum_{n\geq 3} \mathbf{V}_2 \langle (b_{2n} + c_{2n})\rangle \mathbf{f}_n$$

Continuing this way (which is tedious) one sees that the calculating rules given suffice to deal with sums.

Composing operations means dealing with products like

$$\mathbf{V}_m \langle b\rangle \mathbf{f}_n \mathbf{V}_r \langle c\rangle \mathbf{f}_s \tag{15.19}$$

Let d be the greatest common divisor of n and r. Then the product (15.19) is equal to

$$\begin{aligned}\mathbf{V}_m \langle b\rangle \mathbf{f}_n \mathbf{V}_r \langle c\rangle \mathbf{f}_s &= \mathbf{V}_m \langle b\rangle \mathbf{f}_{n/d} \mathbf{f}_d \mathbf{V}_d \mathbf{V}_{r/d} \langle c\rangle \mathbf{f}_s = \mathbf{V}_m \langle b\rangle \mathbf{f}_{n/d} [d] \mathbf{V}_{r/d} \langle c\rangle \mathbf{f}_s \\ &= [d] \mathbf{V}_m \langle b\rangle \mathbf{f}_{n/d} \mathbf{V}_{r/d} \langle c\rangle \mathbf{f}_s = [d] \mathbf{V}_m \langle b\rangle \mathbf{V}_{r/d} \mathbf{f}_{n/d} \langle c\rangle \mathbf{f}_s \\ &= [d] \mathbf{V}_{mr/d} \left\langle b^{r/d} c^{n/d}\right\rangle \mathbf{f}_{sn/d}\end{aligned}$$

So, back to (finite) sums again. The only remaining thing to check is that no infinite sums turn up when multiplying (composing) two things like (15.13) which is immediate.

So the given calculating rules suffice.

15.20. *Witt vectors as endomorphisms.* Quite generally if Ω is a unital-commutative-ring-valued functor on **CRing** to itself its restriction to R-algebras has $\Omega(R)$ as part of its ring of additive endo operations. Indeed an R-algebra structure on a ring A is the same thing as a morphism of rings $R \longrightarrow A$, which on applying Ω yields a morphism of rings $\Omega(R) \longrightarrow \Omega(A)$ which, in turn, for each $x \in \Omega(R)$ an associated additive operation $(x, a) \mapsto xa$.[56]

[56] If the functor takes noncommutative rings as values, or more generally groups with an extra multiplication structure as values, there are two such operations, one on each side and these are additive if and only if resp. left distributivity or right distributivity holds for the extra multiplication structure over the group structure. This applies for instance to the case of the functor defined by the ring of quasi-symmetric functions. **QSymm**, see [200].

So in the present case of the big Witt vectors $W(R)$ must be part of the ring af additive endo operations of $W_R \cong \Lambda_R$. It remains to identify these Witt vectors among the $\sum_{m,n} \mathbf{V}_m \langle c_{mn} \rangle \mathbf{f}_n$.

This is not difficult. Inside the algebras of operations on Λ_R there are the operations of the form

$$\sum_{n=1}^{\infty} \mathbf{V}_n \langle x_n \rangle \mathbf{f}_n, \ x_n \in R \tag{15.21}$$

which take the determining example $(1 - Tt)^{-1}$ into the power series

$$\prod_{n=1}^{\infty} (1 - x_n T^n t^n)^{-1} \tag{15.22}$$

which is a power series with constant term 1 in the variable Tt in Witt coordinate form.

As addition of operations goes pointwise, the power series attached to the sum of two elements of the form (15.21) is the product of the corresponding power series (15.22). It follows that the special operations (15.21) constitute a subgroup of all operations that is (isomorphic to) the additve group of the Witt vectors $W(R)$.

Now consider the composition of an operation (15.21) with an operation of the form $\langle \eta \rangle$ with associated power series $(1 - \eta Tt)^{-1} = \langle \eta \rangle (1 - Tt)^{-1}$. Appling $\langle \eta \rangle$ to (15.22) by definition gives

$$\prod_{n=1}^{\infty} (1 - x_n \eta^n T^n t^n)^{-1}$$

By the definition of the multiplication of Witt vectors, see (9.16) taking a product of a Witt vector $a(tT)$ with one of the form $(1 - \eta Tt)^{-1}$ is the same as substituting ηtT for tT. Thus composition of operations and Witt vector multiplication agree in this case. As should be the case for the operation of $\langle \eta \rangle$ to fit with the algebra structure $\Lambda(R) \longrightarrow \Lambda(A)$. By additivity and distributivity (both of Witt vector multiplication over Witt vector addition and operation composition over operation addition) it folllows that a product of Witt vectors of the form

$$(1 - xt^n T^n)^{-1} * \left(\prod_{n=1}^{\infty} (1 - x_n T^n t^n)^{-1} \right)$$

corresponds exactly to the composition of the corresponding operations. A degree argument now finishes the proof that multiplication of Witt vectors corresponds to the composition of operations of the form (15.21).

15.23. *Theorem.* The Witt vectors $W(R)$ under the correpondence (15.21)–(15.22) form a functorial subring of the rings of operations $\mathrm{End}(\Lambda_R) \cong \mathrm{End}_R(\mathbf{Symm}_R)$.

A somewhat remarkable fact in this case is that in the description 15.16 the subring of operations corresponding to $\Lambda(R)$ comes in Witt vector coordinates.

Still another way to see that the Witt vectors are part of Cart(R) will be briefly discussed further below.

15.24. *Further description of the Cartier algebra.* It is now natural to describe the ring of operations Cart(R) as an overring of $W(R)$, [84]. Let H be the skew polynomial ring over $W(R)$ in commuting indeterminates $\mathbf{f}_p$, p a prime number, with the commutation relation with Witt vectors given by $\mathbf{f}_p x = x^{\mathbf{f}_p}\mathbf{f}_p$ where $x^{\mathbf{f}_p} = x^{(p)}$ denotes the Witt vector arising from applying the p-th Frobenius operation to the Witt vector x. Next take power series in the commuting indeterminates $\mathbf{V}_p$ with commutation relations

$$x\mathbf{V}_p = \mathbf{V}_p x^{\mathbf{f}_p} \quad \text{and} \quad \mathbf{f}_p\mathbf{V}_{p'} = \mathbf{V}_{p'}\mathbf{f}_p \quad \text{for} \quad p \neq p'$$

Let this ring be S. In it take the ideal J generated by the elements

$$\mathbf{f}_p\mathbf{V}_p - p, \quad \mathbf{V}_p x\mathbf{f}_p - x^{\mathbf{V}_p}, \quad x \in W(R), \quad p \text{ a prime number}$$

Then Cart(R) $= S/J$.

15.25. *The ring of additive endo operations of W_{p^∞}.* The situation in the case of the ring valued functor of the p-adic Witt vectors is a bit different. There is no 'determining example' (as far as I can see) like in the case of the big Witt vectors and it is not true that the operations $\mathbf{V} = \mathbf{V}_p$, $\mathbf{f} = \mathbf{f}_p$ (the only surviving Frobenius and Verschiebung operations) together with the homothety operations suffice to generate all of them. In fact for $p \neq 2$ these are not even enough to produce the operation [2] of adding a p-adic Witt vector to itself.

However, as above in 15.20 there are the p-adic Witt vectors $W_{p^\infty}(\mathbf{Z})$ as operations on W_{p^∞} and together with $\mathbf{V}$, $\mathbf{f}$ these do suffice to generate all additive operations. This is the topic of the next few pages. In this case I find it more convenient to work in terms of Hopf algebra endomorphisms of the representing Hopf algebra of the p-adic Witt vectors. This is the Hopf algebra

$$WH^{(p)} = \mathbf{Z}[X_0, X_1, X_2, \cdots] \tag{15.26}$$[57]

with addition and multiplication comultiplications determined by the p-adic Witt polynomials as in section 5 above. For this part of this section 15, temporarily, the notations of section 5 will be used again.

15.27. *Endomorphisms of the infinite dimensional additive group.* To start with consider the infinite dimensional additive group. This is the functor

$$A \mapsto A^{\mathrm{N}}, \text{ represented by the ring } \mathbf{Z}[U] = \mathbf{Z}[U_0, U_1, U_2, \cdots] \tag{15.28}$$

with sum comultiplication and product comultiplication and corresponding counits

$$\mu_s(U_n) = 1 \otimes U_n + U_n \otimes 1, \quad \varepsilon_s(U_n) = 0, \quad \mu_p(U_n) = U_n \otimes U_n, \quad \varepsilon_p(U_n) = 1 \tag{15.29}$$

[57] The "WH" here stands for "Witt-Hopf" (as a sort of mnemonic).

A Hopf algebra endomorphism of this Hopf algebra is given by a series of polynomials with coefficients in the integers

$$G(U) = (G_0(U), G_1(U), G_2(U), \cdots)$$

which are additive. This means that

$$G_n\big(U_1 + U_1', U_2 + U_2', U_3 + U_3', \cdots\big) = G_n(U_1, U_2, U_3, \cdots) + G_n\big(U_1', U_2', U_3', \cdots\big) \quad (15.30)$$

There are over a ring of characteristic zero, such as $\mathbf{Z}$, not many such polynomials. The only ones are the linear ones, i.e. polynomials of the form

$$G_n(U) = \sum_i^{<\infty} a_{n,i} U_i, \quad a_{n,i} \in \mathbf{Z} \quad (15.31)$$

where, as indicated, the sums are finite.

15.32. *Determination of the Hopf algebra endomorphisms of $WH^{(p)}$.* Now let φ be a Hopf algebra endomorphism of $WH^{(p)}$. It is given by a series of polynomials $(\varphi_0, \varphi_1, \varphi_2, \cdots)$. These, when substituted in the polynomials $w_0, w_1, w_2, \cdots$ define an Hopf algebra endomorphism of the infinite additive group and so are linear in the $w_i(X)$. Thus a series of polynomials $(\varphi_0, \varphi_1, \varphi_2, \cdots)$ defines a Hopf algebra endomorphism of $WH^{(p)}$ if and only if

$$w_n(\varphi(X)) = \sum_{i=0}^{\infty} a_{n,i} w_i(X) \quad (15.33)$$

for suitable integers $a_{n,i}$ of which only finitely many are nonzero for each n. Note that the integrality of the coefficients $a_{n,i}$ follows from (15.33). Just look at the coefficients of the powers of X_0 on the left and right hand side of (15.33). It also follows that the right hand side 'comes from' an integral vector over $WH^{(p)}$. This ring is of the type described in the 'ghost component integrality lemma' 9.93. The ring endomorphism φ_p is the one that takes each X_i to X_i^p. Here we are dealing with p-adic Witt vectors so the only denominators involved are powers of the prime number p. So the integrality criterion simplifies to the single sequence for the prime number p. So in the case at hand

$$\sum_{i=0}^{\infty} a_{n,i} w_i(X^p) \equiv \sum_{i=0}^{\infty} a_{n+1,i} w_i(X) \mod p^{n+1} \quad (15.34)$$

is necessary and sufficient for the sequence $\varphi(X)$ as determined by (15.33) to define a Hopf algebra endomorphism of $WH^{(p)}$ and all Hopf algebra endomorphisms of $WH^{(p)}$ arise (uniquely) in this way.

Looking at the coefficients of the powers $X_0^{p^j}$ it follows from (15.34) that the integers $a_{n,i}$ satisfy

$$\begin{aligned} &a_{m,0} \equiv 0 \bmod(p^m), \quad m = 1, 2, 3, \cdots \\ &a_{m,i} \equiv a_{m+1,i+1} \bmod(p^{m+1}), \quad m, i \in \mathbf{N} \cup \{0\} \end{aligned} \tag{15.35}$$

(and these congruences are also sufficient for (15.34) to hold). Now consider the following vectors over the integers

$$\begin{aligned} b_m &= (p^{-m}a_{m,0}, p^{-m}a_{m+1,1}, p^{-m}a_{m+2,2}, \cdots), \quad m = 1, 2, 3, \cdots \\ c_n &= (a_{0,n}, a_{1,n+1}, a_{2,n+2}, a_{3,n+3}, \cdots) \quad n = 0, 1, 2, 3, \cdots \end{aligned} \tag{15.36}$$

By (15.35) these satisfy the criterium of lemma (9.93) for being a p-adic ghost Witt vector. So there are Witt vectors

$$x_m, y_n \in W_{p^\infty}(\mathbf{Z}) \quad \text{such that} \quad w(x_m) = b_m, \quad w(y_n) = c_n \tag{15.37}$$

By definition the endomorphism $\mathbf{f}$ takes $w_r(X)$ into $w_{r+1}(X)$ and the endomorphism $\mathbf{V}$ takes $w_r(X)$ into $pw_{r-1}(X)$ and by the definition of the multiplication of Witt vectors

$$\begin{aligned} w(x \cdot_{W_{p^\infty}} (X_0, X_1, X_2, \cdots) = &(w_0(x)w_0(X), w_1(x)w_1(X), \\ &w_2(x)w_2(X), \cdots) \end{aligned}$$

for any $x \in W_{p^\infty}(Z)$. So the endomorphism

$$\sum_{n=1}^{\infty} \mathbf{f}^n c_n + c_0 + \sum_{m=1}^{\infty} b_m \mathbf{V}^m \tag{15.38}$$

takes $w_r(X)$ into

$$\sum_{n=1}^{\infty} \mathbf{f}^n a_{r,n+r} w_r(X) + a_{r,r} w_r(X) + \sum_{m=1}^{r} \left(p^{-m} a_{r,r-m}\right) p^m w_{r-m}(X) = \sum_{i=1}^{\infty} a_{r,i} w_i(X)$$

exactly as in (15.33). Because a Witt vector $x \in W_{p^\infty}(\mathbf{Z})$ starts with j zeroes, i.e lies in the ideal $\mathbf{V}^j W_{p^\infty}(\mathbf{Z})$, if and only if its associated ghost vector starts with j zeroes (as also follows from the ghost vector criterium of lemma 9.93), the finiteness condition of 15.33) translates into the statement that for each r all but finitely many of the Witt vectors c_n in (15.38) lie in the ideal $\mathbf{V}^r W_{p^\infty}(\mathbf{Z})$. Thus the ring of Hopf algebra endomorphisms of $WH^{(p)}$ consists of all expressions (15.38) with the finiteness condition just stated and with the calculating rules

$$\mathbf{V}\mathbf{f} = p, \quad \mathbf{V}x = x^{(p)}\mathbf{V}, \quad x\mathbf{f} = \mathbf{f}x^{(p)} \tag{15.39}$$

where $x^{(p)}$ denotes the result of applying the operation $\mathbf{f}$ to the p-adic Witt vector x and where $p = [p]$ is the p-fold sum of the unit element in $W_{p^\infty}(\mathbf{Z})$.

15.40. The same technique can be used to determine the ring of endomorphisms of the Hopf algebra **Symm** as the representing Hopf algebra of the big Witt vectors, giving another way to obtain the ring $\mathrm{Cart}_{\mathbf{Z}}$.

15.41. The techniques of (15.32) as a method to determine the endomorphisms of the Hopf algebra $WH_R^{(p)}$ break down when R is not a ring of characteristic zero for two reasons. One is that ghost component calculations as used are then not determining and second because in characteristic $p > 0$ there are other endomorphism of the additive groups than the linear ones; e.g. raising to the p-th power. Still I believe that the same picture holds. One reason is that the picture holds for the $WH_R^{(p)}$ when k is a perfect field of characteristic $p > 0$. This will be described below. However, the proofs are quite different. Still every endomorphism in this case does come from one defined in characteristic zero (after the fact, i.e. after they have been determined in characteristic p). So there still appears to be a bit of work to do.

15.42. *Dieudonné algebras.* Let k be a perfect field of characteristic $p > 0$. The Dieudonné algebra $\mathbf{D}_k$ consists of all expressions

$$\sum_{n=1}^{<\infty} \mathbf{f}^n c_n + c_0 + \sum_{m=1}^{<\infty} b_m \mathbf{V}^m, \quad c_n, b_m \in W_{p^\infty}(k) \tag{15.43}$$

subject to the calculating rules

$$\mathbf{Vf} = \mathbf{fV} = p, \quad \mathbf{V}x = x^{(p)}\mathbf{V}, \quad x\mathbf{f} = \mathbf{f}x^{(p)} \tag{15.44}$$

Recall that in this case the Frobenius operation on $W_{p^\infty}(k)$ is given by

$$x^{(p)} = \left(x_0^p, x_1^p, x_2^p, x_3^p, \cdots\right)$$

and hence is an isomorphism, so that in (15.43) the Frobenius and Verschiebung symbols can be written on the left or the right as convenient. Also recall that (see 13.13) the sequence of polynomials that define the Frobenius endomorphism on $WK_k^{(p)}$ is

$$X_0^p, X_1^p, X_2^p, \cdots$$

so that these endomorphisms in characteristic p take care of the 'raising to the power p' additive operations. One result in the present case is now (see [105], Chapter V, §1, No 3, p. 550):

15.45. *Theorem.* The endomorphim ring (over k) of the sub Hopf algebra $k[X_0, X_1, \cdots X_{n-1}] \subset WH_k^{(p)}$ is isomorphic to the ring $\mathbf{D}_k/\mathbf{D}_k\mathbf{V}^n$.

15.46. *Hirzebruch polynomials.* (Very partial symmetric function formularium (3)). Consider again an additive (functorial) operation on the groups of Witt vectors $\Lambda(A)$.

Such an operation is given uniquely by a series of polynomials

$$1 = K_0, K_1, K_2, \cdots \in \mathbf{Z}[e] = \mathbf{Z}[e_1, e_2, e_3, \cdots] \tag{15.47}$$

such that

$$\mu_s(K_n) = \sum_{i=0}^{n} K_i \otimes K_{n-i}, \text{ where } \quad \mu_s(e_n) = \sum_{i=0}^{n} e_i \otimes e_{n-i} \tag{15.48}$$

is the sum comultiplication on $\mathbf{Symm} = \mathbf{Z}[h] = \mathbf{Z}[e]$.[58] If the K_n are also homogeneous of degree n (where e_i has degree i this is precisely what is defined in [213], Chapter I, §1, p. 9 as a multiplicative sequence.

The operation ψ_K defined by the multiplicative sequence (15.47) is of course

$$\psi_K : a(t) = 1 + a_1 t + a_2 t^2 + \cdots + \mapsto 1 + K_1(a)t + K_2(a)t^2 + \cdots \tag{15.49}$$

A multiplicative sequence is uniquely determined by what in loc. cit. is called its characteristic power series

$$Q(z) = 1 + K_1(1, 0, 0, \cdots)z + K_2(1, 0, 0, \cdots)z^2 + \tag{15.50}$$

which can be seen as the value of the operation on the determining example $1 + zt$ because by the homogeneity condition on the K_n

$$\psi_K(1 + zt) = Q(zt) \tag{15.51}$$

The homogeneity condition is also clearly equivalent to the property that in the DE formalism of (15.5) the DE-matrix of the operation is diagonal and so the operation on $\Lambda(A)$ in question is multiplication by a single Witt vector from $\Lambda(\mathbf{Z})$ (or $\Lambda(\mathbf{Q})$ if rational coefficients are allowed in the K (and in the characteristic power series). This single particular Witt vector is $Q(-t)^{-1}$.

The multiplicative sequence itself is recovered by considering the product

$$\prod_{i=1}^{\infty} Q(\xi_i t)$$

For each n the coefficient of t^n is a homogeneous symmetric polynomial in the ξ and hence a homogeneous polynomial of weight n in the e_i. These are the original multiplicative sequence polynomials by multiplicativity of the sequence or, equivalently, additivity of the operation defined by them.

Inversely, start with any power series $Q(t)$ over the integers or rationals and form the sum and product

$$s_Q(t) = \sum_{i=1}^{\infty} Q(\xi_i t), \quad p_Q(t) = \prod_{i=1}^{\infty} Q(\xi_i t) \tag{15.52}$$

[58] Elementary symmetric functions are used here in order to conform with usage in algebraic topology such as in [213] where these things were first introduced.

where for the sum case it is assumed that $Q(0) = 0$ and for the product case $Q(0) = 1$. The coefficients of t^n in (15.52) are symmetric in the ξ and hence are polynomials in the elementary symmetric functions. These are the n-th additive and multiplicative Hirzebruch polynomials defined by the power series $Q(t)$.

For example the Todd genus operation is defined by the power series

$$Q(z) = \frac{z}{1 - e^{-z}} = 1 + \frac{1}{2}z + \sum_{n=1}^{\infty}(-1)^n \frac{B_n}{(2n)!}z^{2n} \tag{15.53}$$

and the Todd genus of a vector bundle is obtained by applying this operation of the total Chern class $1 + c_1(V)t + c_2(V)t^2 + \cdots + c_n(V)t^n$ of a complex vector bundle of dimension n.

The additive Hirzebruch polynomials serve to define functorial additive operations from $\Lambda(A)$ to A^{N}. An example is the Chern character of a vector bundle defined by applying the operation associated to the power series e^z to the total Chern class of a vector bundle.[59]

15.54. *Open problem.* In subsection 15.5 (on the DE principle), formula 15.7 to be precise, an endomorphism of the Hopf algebra is given by an element of $\Lambda(\mathbf{Z}[T])$, the Witt vectors over the ring of polynomials in one indeterminate. One wonders how the multiplication of these as Witt vectors fits with the other bits of structure (composition and addition of endomorphisms).

16. More operations on the Λ and W functors: λ-rings.

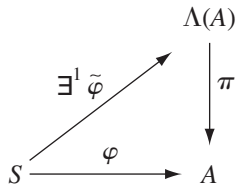
There can be other operations than additive ones on an (Abelian) group valued functor that are important. This happens e.g. in algebraic topology and algebraic geometry with the operations on various functors induced by exterior powers (or, equivalently, symmetric powers) of vectorbundles. There are such functorial operations on the big Witt vectors. Indeed the Witt vectors $\Lambda(A)$ are functorial λ-rings.[60] In this setting the Witt vectors have three more universality properties (that are interrelated).

16.1. *Third universality property of the Witt vectors.* The functor Λ: **CRing** $\longrightarrow$ $\lambda-$ **Ring** is right adjoint to the forgetful functor the other way. This means the following. First there is a canonical projection $\Lambda(A) \xrightarrow{\pi} A$ (which is in fact the functorial morphism s_1 (see (9.8), the first ghost component morphism) and second there is the universal property (that says that $\Lambda(A)$ is the cofree object

[59] Here $Q(0) = 1 \neq 0$ and so a finite factorization $1 + c_1(V)t + \cdots + c_n(V)t^n = \prod_{i=1}^{n}(1 + \beta_i t)$ is used.

[60] I do not know whether the notation 'Λ' for the functor $A \mapsto 1 + tA[[t]]$ has anything to do with the fact that the $1 + tA[[t]]$ are functorial λ-rings. I believe not. However, Λ is also a standard notation for **Symm** the representing object of this functor.

over A) to the effect that for every morphism of rings $\varphi\colon S \longrightarrow A$ from a λ-ring to A there is a unique lift, that is a morphism of λ-rings $\tilde{\varphi}\colon S \longrightarrow \Lambda(A)$ such that $\pi\tilde{\varphi} = \varphi$ as illustrated in the diagram below.[61]



16.2. *Fourth universality property of the Witt vectors.* **Symm**, the representing object of the functor Λ, is the free λ-ring on one generator.

16.3. *Fifth universality property of the Witt vectors.* There is a comonad (= cotriple) structure on Λ and the coalgebras for this comonad are precisely the λ-rings.

16.4. *Definition of λ-rings.* A λ-ring is a unital commutative ring that comes equipped with a an extra collection of operations

$$\lambda^i : A \longrightarrow A, \quad i = 1, 2, 3, \cdots \tag{16.5}$$

which 'behave like exterior powers.' In particular they are not additive for $i \geq 2$. The phrase 'behave like exterior powers' means like the exterior powers of vector bundles over a space or like the exterior powers of representations of a group. Thus the first requirement is that

$$\lambda^n(a+b) = \lambda^n(a) + \sum_{i=1}^{n-1} \lambda^i(a)\lambda^{n-i}(b) + \lambda^n(b) = \sum_{i=0}^{n} \lambda^i(a)\lambda^{n-i}(b) \tag{16.6}$$

where by definition $\lambda^0(a) = 1$ for all $a \in A$. In addition the exterior product operations are supposed to satisfy some properties on products and when iterated in the sense that there are some universal polynomials which give $\lambda^n(xy)$ and $\lambda^m(\lambda^n(x))$ in terms of the $\lambda^i(x)$, $\lambda^j(y)$. These polynomials are specified by defining exterior products on the rings $\Lambda(A)$ and declaring these particular rings with these particular exterior products to be λ-rings.[62] That goes as follows.

---

[61] This property rather nicely takes care of the complaint on page 1 of [263].

[62] Lambda rings were introduced by Alexandre Grothendieck in the course of his investigations into Riemann-Roch type theorems, [48, 178, 179]. In spite of its attractive sounding title the reader is advised to steer clear of [148]. This book contains a great deal of most interesting material; it also contains nasty mistakes (as in the second paragraph of p. 15) and is sloppily written; see also the review by K R Coombes in Math. Rev. (88h:14011).

In the older literature, see e.g. loc. cit. and [27, 239] a lambda ring is a ring with exterior product operations satisfying just the additivity condition (16.6) and the term 'special lambda ring' was used for what here is called a lambda ring.

A morphism of λ-rings is a morphism of rings that respects the exterior product structures, i.e. a unital morphims of rings $\varphi : A \longrightarrow B$ such that

$$\varphi(\lambda^i_A(x)) = \lambda^i_B(\varphi(x)) \tag{16.7}$$

The exterior product structure on the $\Lambda(A)$ is defined by

$$\lambda^m a(t) = \prod_{i_1<i_2<\cdots<i_m} \left(1 - \xi_{i_1}\xi_{i_2}\cdots\xi_{i_m}t\right)^{-1} \text{ when } \quad a(t) = \prod_{i=1}^{\infty}(1-\xi_i t)^{-1} \tag{16.8}$$

By definition a λ-ring is a unital commutative ring A equipped with exterior product operations λ^i such that

$$\begin{aligned} &\lambda^1 = \text{id and the mapping } A \longrightarrow \Lambda(A) \text{ given by} \\ &x \mapsto (1 - \lambda^1(x)t + \lambda^2(x)t^2 - \lambda^3(x)t^3 + \lambda^4(x)t^4 \cdots)^{-1} \end{aligned} \tag{16.9}$$

is a morphism of λ-rings (where $\Lambda(A)$ is given the exterior product structure (16.8)). Writing

$$\lambda_1(x) = 1 + \lambda^1(x)t + \lambda^2(x)t^2 + \lambda^3(x)t^3 + \cdots \tag{16.10}$$

the second part of (16.9) says that

$$x \mapsto \lambda_{-t}(x)^{-1}, \ A \longrightarrow \Lambda(A) \tag{16.11}$$[63]

is a morphism of λ-rings,while the first part says that $s_1\left(\lambda_{-t}^{-1}\right) = \text{id}$.

Of course of make the whole thing consistent it has to be proved that $\Lambda(A)$ with the exterior product as specified by (16.8) is a λ-ring as defined via (16.9). That is it has to be proved that

$$\Lambda(A) \longrightarrow \Lambda(\Lambda(A)) \tag{16.12}$$

is a morphism of λ-rings. This will be done below, mixing in a number of lemmas which will also be useful later.

16.14. *λ-rings vs σ-rings.* Given a ring with exterior product operations λ^n define the corresponding symmetric power operations σ^n by the formula

$$\sum_{i=0}^{n}(-1)^i\lambda^i(x)\sigma^{n-1}(x) = 0, \text{ where } \quad \sigma^0(x) = 1 = \lambda^0(x) \tag{16.15}$$

---

[63] There are a couple minus signs here compared to the definition as given in e.g. [239, 324]. The reason for that is that in the present text (for good reasons) the ring structure on Λ(A) is specified by the 'Hadamard like product' $(1-xt)^{-1} * (1-yt)^{-1} = (1-xyt)^{-1}$ with unit element $(1-t)^{-1}$ instead of $(1+xt) *' (1+yt) = 1+xyt$ with unit element $1+t$ (as in [239, 324]). There are more good reasons for the particular choice made here (Which is also the choice made in the first treatments of this subject, [84, 86]) For instance, as already mentioned, it works out nicer for the autoduality of **Symm**). Still another good reason is that the multiplication $(1-xt)^{-1} * (1-yt)^{-1} = (1-xyt)^{-1}$ is how zeta functions of varieties over a finite field multiply in the sense that $\zeta(X \times Y) = \zeta(X) * \zeta(Y)$, see [263], p. 2.

If the exterior product operations satisfy (16.6) then so do the symmetric product operations. Also, obviously a morphism of rings respects exterior products if and only if it respects symmetric products. Write

$$\sigma_t(x) = 1 + \sigma_1(x)t + \sigma_2(x)t^2 + \sigma_3(x)t^3 + \cdots \tag{16.16}$$

then by definition a ring is a σ-ring if and only if

$$\sigma_t : x \mapsto \sigma_t(x), \quad A \longrightarrow \Lambda(A) \tag{16.17}$$

is a morphism of σ-rings. The relation between the λ^n and σ^m can be succintly written

$$\sigma_t(x) = (\lambda_{-t}(x))^{-1} \tag{16.18}$$

and so σ_t is precisely the mapping that figures in condition (16.11). Thus a σ-ring is exactly the same as a λ-ring. But the formulation of the property is just a tiny bit more elegant in terms of σ-rings (no minus signs).

It is a nice little exercise to show that the σ-ring structure on $\Lambda(A)$ is given more explicitly by

$$\sigma^m\left(\prod_i (1 - \xi_i t)^{-1}\right) = \prod_{i_1 \le i_2 \le \ldots \le i_m} \left(1 - \xi_{i_1}\xi_{i_2}\cdots\xi_{i_m}t\right)^{-1} \tag{16.19}$$

This is in fact the Wronski relations again between elementary and complete symmetric functions (when the complete symmetric functions are written out as the sum of all monomial symmetric functions of the same weight).

16.20. *Adams operations.* Given a ring with exterior product structure operations λ^n the associated Adams operations,[64,65] are defined by

$$\sum_{n=1}^{\infty} \Psi^n(x)t^n = -t\frac{d}{dt}\log(\lambda_{-t}(x)) = t\frac{d}{dt}\log(\sigma_t(x)) \tag{16.21}$$

where the last equality of course follows from (16.18). Here also the formulation is just a bit more elegant in terms of the σ-structure.

---

[64] The operations are named after J Frank Adams who first defined them in algebraic topology in the context of complex vector bundles and topological K-theory, [14].

[65] Consider a Hopf algebra H and let μ_n and m_n be the n-th iterates of its comultiplication and multiplication. Then the composite is the additive map (in general nothing more) $[n]$, the n-fold sum of the identity (under convolution). In some of the literature, e.g. [141, 167, 270, 323] these maps are called Adams operations. That is a bit unfortunate as it does not fit in e.g. the case of **Symm**. For one thing if the Hopf algebra is graded these maps $[n]$ are homogeneous (of degree zero) while the Addams operations are degree increasing. Also it does not fit with the case where the Adams operations came from, the cohomology of the classifying space **BU**. This does not mean that these maps are not important. They are, see loc. cit.

16.22. *Theorem.* The Adams operations on the rings $\Lambda(A)$ with exterior product operations given by (16.8) or, equivalently, with symmetric product operations given by (16.19), are the Frobenius operations.

I like to call this observation the first 'Adams = Frobenius' theorem. The proof is easy. First take a power series of the form $x = (1-at)^{-1}$. Then $\lambda^1(x) = x$, $\lambda^i(x) = 0$ for $i \geq 2$ and hence

$$\lambda_{-u}(x) = 1 - xu, \; -u\frac{d}{du}\log(\lambda_{-u}(x)) = \frac{xu}{1-xu} = xu + x^2u^2 + x^3u^3 + \cdots$$

And so

$$\Psi^n((1-at)^{-1}) = ((1-at)^{-1})^{*n} = (1-a^nt)^{-1} = \mathbf{f}_n((1-at)^{-1})$$

by the definition of product and Frobenius operations on $\Lambda(A)$. Now write an arbitrary element from $\Lambda(A)$ (formally) in the form

$$x = \prod_i (1-\xi_i t)^{-1}, \text{ i.e. } x = x_1 +_\Lambda x_2 +_\Lambda x_3 +_\Lambda \cdots, \quad x_i = (1-\xi_i t)^{-1}$$

in terms of the addition in a $\Lambda(\tilde{A})$. The same calculation as above gives

$$\begin{aligned}\Psi^n(x) &= x_1^n +_\Lambda x_2^n +_\Lambda x_3^n +_\Lambda \cdots = \prod_i \left((1-\xi_i t)^{-1}\right)^{*n} \\ &= \prod_i \left(1-\xi_i^n t\right)^{-1} = \mathbf{f}_n x\end{aligned} \tag{66}$$

16.23. Also observe that by the definitions the Frobenius operations on the $\Lambda(A)$, and hence the Adams operations, commute with the exterior and symmetric power operations.

16.34. *Existence lemma* (*for exterior power* (*and other*) *structures*), [192], lemma 17.6.8, p.138. Let A be a characteristic zero ring with ring endomorphisms $\varphi_n : A \longrightarrow A$ for all $n \in \mathbf{N}$ such that $\varphi_1 = \mathrm{id}$, $\varphi_m\varphi_n = \varphi_{mn}$ and such that $\varphi_p(a) \equiv a^p \bmod(pA)$ for all prime numbers p and $a \in A$. Then there exists a unique mapping $D_A : A \longrightarrow \Lambda(A)$ (resp. $D_A : A \longrightarrow W(A)$) such that $s_n D_A = \varphi_n$ (resp. $w_n D_A = \varphi_n$). Moreover, this mapping is a ring morphism.

This is an immediate consequence of the ghost Witt vector integrality lemma 9.93.

16.35. *Lemma.* Let α: $A \longrightarrow B$ be a ring morphism and let both rings have exterior products λ_A^n, λ_B^n with associated Adams operators Ψ_A^n, Ψ_B^n. Then if $\alpha\Psi_A^n = \Psi_B^n\alpha$ for all n, also $\alpha\lambda_A^n = \lambda_B^n\alpha$.

This comes about because the Adams operations and the exterior products are related by (universal) polynomials with rational coefficients.

---

66 Here, again, the splitting principle is used together with what has been called the 'verification principle,' which says that under suitable circumstances it suffices to verify things for line bundles.

Indeed these polynomials had better be given explicitly as they will also be needed further on. They are as follows

$$\Psi^n = \det \begin{pmatrix} \lambda^1 & 1 & 0 & \cdots & 0 \\ 2\lambda^2 & \lambda^1 & 1 & \ddots & \vdots \\ 3\lambda^3 & \lambda^2 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \lambda^1 & 1 \\ n\lambda^n & \lambda^{n-1} & \cdots & \lambda^2 & \lambda^1 \end{pmatrix} \tag{16.40}$$

$$n!\lambda^n = \det \begin{pmatrix} \Psi^1 & 1 & 0 & \cdots & 0 \\ \Psi^2 & \Psi^1 & 2 & \ddots & \vdots \\ \Psi^3 & \Psi^2 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \Psi^1 & n-1 \\ \Psi^n & \Psi^{n-1} & \cdots & \Psi^2 & \Psi^1 \end{pmatrix} \tag{16.41}$$

Some care must be taken in reading these formulas. E.g. in $\lambda^2 = \frac{1}{2}(\Psi^1)^2 - \frac{1}{2}\Psi^2$, $(\Psi^1)^2(x)$ must not be read as $\Psi^1(\Psi^1(x))$ but as $(\Psi^1(x))^2$.

These determinental formulas are exactly the same as those linking power sums and elementary symmetric functions in symmetric function theory ([281], p. 28), which is as must be because the defining formulas are the same in the two cases.[67]

16.42. *Clarence Wilkerson theorem.* Let A be as above in lemma 16.34. Then there is a unique λ-ring structure on A such that the associated Adams operators are the given ring morphisms φ_n. Moreover the thus defined exterior powers commute with the given morphisms φ_n.

Proof. By the existence lemma 16.34 there is a unique ring morphism $\sigma_t\colon A \longrightarrow \Lambda(A)$ such that $s_n\sigma_t = \varphi_n$ for all n. This defines the exterior powers. Now observe that for all $a \in A$

$$s_m\sigma_t(\varphi_n(a)) = \varphi_m(\varphi_n(a)) = \varphi_{mn}(a)$$

and on the other hand, using the defining property of the Frobenius operation on the Witt vectors

$$s_m\mathbf{f}_n(\sigma_t(a)) = s_{mn}(\sigma_t(a)) = \varphi_{mn}(a)$$

[67] It is because of this that the Adams operations are often called 'power operations'.

Because A is of characteristic zero (so that $s\colon \Lambda(A) \longrightarrow A^{\mathrm{N}}$ is injective) this implies that $\mathbf{f}_n\sigma_t = \sigma_t\varphi_n$. Thus by lemma 16.35 σ_t respects the exterior powers and thus is a morphism of λ-rings as required. □

16.43. *Corollary.* The rings of Witt vectors $\Lambda(A)$ are functorial λ-rings.

Proof. First suppose that A is of characteristic zero. Then so is $\Lambda(A)$. Now let AH be the mapping $\Lambda(A) \longrightarrow \Lambda(\Lambda(A))$ defined by the given exterior product structure on $\Lambda(A)$. At this stage almost nothing is known about what properties it has. However, by definition of the Adams operations these satisfy $s_{n,\Lambda(A)} \circ AH = \Psi^n$. The operations Ψ^n on $\Lambda(A)$ have been shown to be the Frobenius operations (the first Adams = Frobenius theorem) and these Frobenius operations satisfy the conditions of the existence lemma 16.34. So the uniqueness part says that AH is equal to the morphism whose existence is guaranteed by that lemma and that one is a ring morphism by that lemma and respects exterior powers by theorem 16.42. This proves the corollary for characteristic zero rings. For an arbitrary ring take any characteristic zero cover. Everything in sight is functorial and so a little diagram chasing gives the result also in this case. □

16.44. *Remark.* The following diagram commutes for any λ-ring.

$$\begin{array}{ccc} \Lambda(A) & \xrightarrow{\mathbf{f}_n} & \Lambda(A) \\ \uparrow{\scriptstyle \sigma_t} & & \uparrow{\scriptstyle \sigma_t} \\ A & \xrightarrow{\Psi^n} & A \end{array}$$

For instance because $\mathbf{f}_n$ and Ψ^n are given by the same polynomials in the exterior products (see (16.40) and the σ_t are morphisms of λ-rings. In fact more generally if A and B are λ-rings and $A \xrightarrow{\alpha} B$ is a morphism of λ-rings then the morphism commutes with the respective Adams operations.

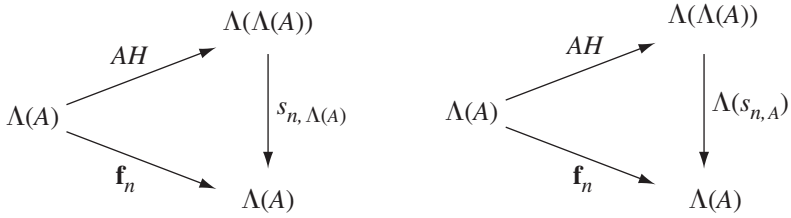
It follows from the commutativity of the above diagram that $\Psi^m\Psi^n = \Psi^{mn}$ and that the Ψ^n respect the exterior powers.

16.45. *Remark.* The notation 'AH' for the functorial morphism of λ-rings $\Lambda(A) \longrightarrow \Lambda(\Lambda(A))$, which could also be properly denoted $\sigma_{t,\Lambda(A)}$, stands for 'Artin-Hasse'. It is in fact a variant of the Artin Hasse exponential[68] in algebraic number

[68] For the definition of the Artin-Hasse exponential in algebraic number theory and some of its uses see e.g. [412] to start with.

theory. More precisely if one takes a finite field k, identifies $\Lambda(k) = W(k)$, takes the quotient $W(k) \longrightarrow W_{p^\infty}$, and sees the latter ring as the unramified mixed characteristic complete discrete valuation ring with residue field k one finds a morphism of rings $W_{p^\infty}(k) \longrightarrow \Lambda(W_{p^\infty}(k))$ which is in fact the classical Artin-Hasse exponential, see [192], sections 17.5, 17.6 and E2.

By its definition the Artin-Hasse exponential AH is such that the left hand diagram below commutes. It follows that the right hand diagram also commutes.



16.46. *Remark.* The simplest example of a (nontrivial) λ-ring is probably the ring of integers with the exterior power structure $\lambda_t(x) = (1 + t)^x$. So in this case λ_t is a kind of exponential, and that is a good way to think about it. In fact there are situations where the presence of an exterior power structure does lead to exponential type isomorphisms between the underlying 'additive' group of a ring and a multiplicative group of units of it, see e.g. [27], [388, 389].

16.47. *The essence of a λ-ring structure.* Let A be a λ-ring. Then it comes with with a morphism of λ-rings $\sigma_t : A \longrightarrow \Lambda(A)$, $\sigma_t(x) = (1 - \lambda^1(x)t + \lambda^2(x)t^2 - \lambda^3(x)t^3 + \cdots)^{-1}$. That means that there are formulas for the exterior product $\lambda^n(x + y)$ of a sum of elements of A because σ_t is additive; there are formulas for the exterior product $\lambda^n(xy)$ of a product of elements of A because σ_t is multiplicative; and there are formulas for the iterations of exterior products $\lambda^m(\lambda^n(x))$ because σ_t is respects exterior products (as a morphism of λ-rings). These formulas are given in terms of the addition, multiplication, and exterior products on the Witt vector ring $\Lambda(A)$. But these formulas are universal;[69] they do not in any way depend on the ring A; they are certain polynomials with integer coefficients determined by certain manipulations with symmetric functions over the integers.

Thus there are formulas for $\lambda^n(x + y)$, $\lambda^n(xy)$, $\lambda^m(\lambda^n(x))$ as polynomials in the $\lambda^i(x)$ and $\lambda^j(y)$ and, which is the real essence of the story, these polynomials have their coefficients in the integers and are the same for all λ-rings A.

These polynomials can be calculated. Either by working directly with the one universal example of the Witt vectors $1 + h_1(\xi)t + h_2(\xi)t^2 + h_3(\xi)t^3 + \cdots$ and $1 + h_1(\eta)t + h_2(\eta)t^2 + h_3(\eta)t^3 + \cdots$ over $\mathbf{Z}[h(\xi); h(\eta)] \subset \mathbf{Z}[\xi; \eta]$, or, which I find easier, by using the determinantal formulas (16.40) and (16.41).

[69] Overworking that unhappy word again; but I know of none other that meets the case.

For example

$$\begin{aligned}
\lambda^3(xy) &= \frac{1}{6}(\Psi^1)^3(xy) - \frac{1}{2}(\Psi^1\Psi^2)(xy) + \frac{1}{3}(\Psi^3)(xy) \\
&= \frac{1}{6}\Psi^1(x)^3\Psi^1(y)^3 - \frac{1}{2}\Psi^1(x)\Psi^1(y)\Psi^2(x)\Psi^2(y) + \frac{1}{3}\Psi^3(x)\Psi^3(y) \\
&= \lambda^1(x)\lambda^2(x)\lambda^1(y)\lambda^2(y) + \lambda^1(x)^3\lambda^3(y) + \lambda^3(x)\lambda^1(y)^3 \\
&\quad - 3\lambda^1(x)\lambda^2(x)\lambda^1(y)^3 - 3\lambda^1(x)^3\lambda^1(y)\lambda^2(y) + 3\lambda^3(x)\lambda^3(y) \\
\lambda^2(\lambda^2(x)) &= \frac{1}{2}(\Psi^1)^2(\lambda^2(x)) - \frac{1}{2}\Psi^2(\lambda^2(x)) \\
&= \frac{1}{2}\lambda^2(x)^2 - \frac{1}{2}\Psi^2(\frac{1}{2}\Psi^1(x)^2 - \frac{1}{2}\Psi^2(x)) \\
&= \frac{1}{2}\lambda^2(x)^2 - \frac{1}{4}\Psi^2(x)^2 + \frac{1}{4}\Psi^4(x) \\
&= \lambda^1(x)\lambda^3(x) - \lambda^4(x)
\end{aligned}$$

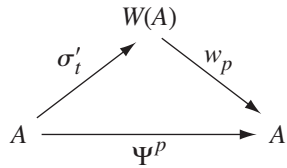
16.48. *Ψ-rings.* A Ψ-ring is a ring together with a family of ring endomorphisms $(\Psi^n)_{n\in\mathbf{N}}$ such that

$$\Psi^1 = \mathrm{id},\ \Psi^m\Psi^n = \Psi^{mn} \tag{16.49}$$

I shall say that A is a 'Ψ-ring with Frobenius morphism like property' if moreover

$$\Psi^p(a) = a^p \bmod p \quad \text{for all prime numbers } p \tag{16.50}$$

Then if A is a λ-ring the associated Adams operations it becomes a Ψ-ring by what was remarked in 16.44 above. It is moreover a Ψ-ring with Frobenius morphism like property. To see this it suffices to remark that that the Adams operations are given in terms of the symmetric powers (resp. the exterior powers) by the same polynomials that relate the power sum symmetric functions to the complete symmetric functions (resp. the elementary symmetric functions). Now $h_p \equiv h_1^p \bmod p$ and the statement follows. Another way to get it is to look at the defining equation for Ψ^p in Witt coordinates:



where σ_t' is σ_t followed by the isomorphism $e_A : \Lambda(A) \longrightarrow W(A)$. Because $w_1\sigma_t' = \mathrm{id}$, $\sigma_t'(a)$ is of the form $(a, x_2, x_3, \cdots)$ and so

$$\Psi^p(a) = w_p(a, x_2, x_3, \cdots) = a^p + px_p \equiv a^p \bmod p$$

16.51. Thus for a characteristic zero ring the statements "A is a λ-ring' and 'A is a Ψ-ring with Frobenius morphism like property' are equivalent. If A is not of characteristic zero this need not be the case.

16.52. Another result in this setting is that if A is a characteristic zero ring with exterior power operations defined on it such that the associated Adams operations turn it into a Ψ-ring then it is a λ-ring.

Note the differences between the two theorems 16.51 and 16.52.

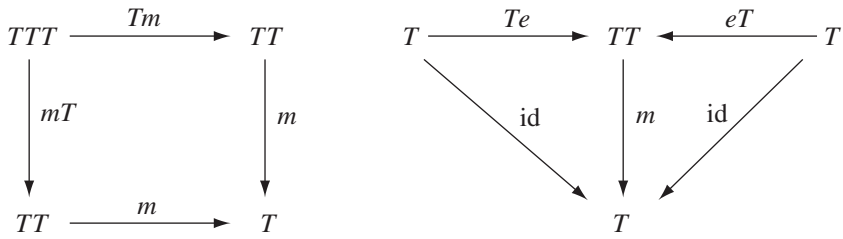
Thus the difference between Ψ-rings and λ-rings is an integrality matter (and such matters can be important). The idea of a Ψ-ring can easily be extended to a noncommutative setting (see below) and dualized. In view of the autoduality of **Symm** there should be some interesting notion of what a 'co-λ-coring' could be. Some sort of co-Ψ-coring with an extra integrality property.

16.53. There are a number of noncommutative rings in various parts of algebra that behave almost like Ψ-rings and λ-rings. Thus it makes sense to develop a theory of noncommutative Ψ-rings just like it was (and is) important to have a theory of noncommutative symmetric functions; the theory of **NSymm** briefly alluded to above in Section 11.

The beginnings of a theory of noncommutative Ψ-rings have been established, see [325] (and [324] for a brief account). And it turns out that this theory relates nicely to the theory of noncommutative symmetric functions, loc. cit.

The next topic is the comonad structure on the functor of the Witt vectors.

16.54. *Monads and comonads.* A monad[70] in a category $\mathcal{C}$ is a triple (T, m, e) consisting of an endofunctor $T\colon \mathcal{C} \longrightarrow \mathcal{C}$, a natural transformation $m\colon TT \longrightarrow T$ (where TT stands for the iterate of T, i.e. $TT(A) = T(T(A))$) and a natural transformation $e\colon \mathrm{id} \longrightarrow T$ such that the following diagrams commute.

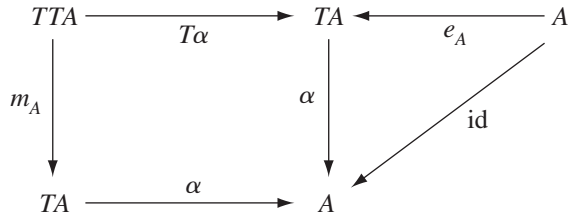


(16.55)

[70] Other names for monad that are used (or have been used) in the literature are 'triad,' 'standard construction,' 'fundamental construction,' 'triple'.

Monads first arose in work of Roger Godement, [171], in connection with the construction of simplicial objects and (standard) resolutions in connection with sheaf cohomology.

Here 'Tm' means 'm' first and then 'T', i.e. at an object C take the morphism $m_C\colon T(C) \longrightarrow TT(C) = T(T(C))$ and then apply the functor T to it to obtain a morphism that is also often denoted $T(m_C)$; on the other hand 'mT' at an object C, i.e. 'T' first and then 'm', is 'm' at the object $T(C)$, often written $m_{T(C)}$.
An algebra (in the category $\mathcal{C}$) for a monad (T, m, e), also called a T-algebra, is an object $A \in \mathcal{C}$ together with a morphism $\alpha\colon TA \longrightarrow A$ such that the following diagram commutes



(16.56)

The opposite notions, i.e. the same notions in the opposite category, are those of a comonad and a coalgebra. Explicitly: a comonad (T, μ, ε) in a category $\mathcal{C}$ is an endo functor T of $\mathcal{C}$ together with a morphism of functors $\mu\colon T \longrightarrow TT$ and a morphism of functors $\varepsilon\colon T \longrightarrow \mathrm{id}$ such that

$$(T\mu)\mu = (\mu T)\mu, \quad (\varepsilon T)\mu = \mathrm{id} = (T\varepsilon)\mu \tag{16.57}$$

And a coalgebra for the comonad (T, μ, ε) is an object in the category $\mathcal{C}$ together with a morphism $\sigma\colon C \longrightarrow TC$ such that

$$\varepsilon_C\sigma = \mathrm{id}, \quad (T(\sigma)\sigma = (\mu_{TC})\sigma \tag{16.58}$$

The equations (16.57) and (16.58) can of course be written out in diagram form, the diagrams involved being the ones one gets by reversing all arrows in (16.55) and (16.56). This seems hardly worth the ink and paper in view of the fact that they will shortly appear explicitly in the case of the particular comonad of the Witt vectors and its coalgebras, which are precisely the λ-rings (or, better, σ-rings).

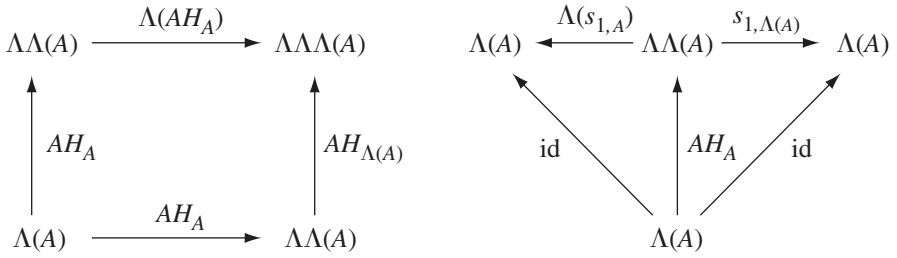
A standard text on monads and comonads is [33]; for a thorough up-to-date account see [284].[71]

16.59. *Comonad structure on the Witt vectors.* The claim is that the Witt vector functor Λ (or W) is a comonad in the category **CRing**, with the comonad morphism given by the Artin-Hasse exponential and the first ghost component morphism and that the coalgebras for this comonad are precisely the λ-rings.

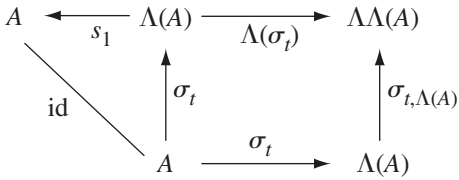
To prove this it must be shown that for each ring A the following diagrams (16.60) are commutative and that an exterior product structure $\sigma_t : A \longrightarrow \Lambda(A)$ gives a

[71] And a, in my opinion, inspiring account.

λ-ring structure if and only if the diagrams (16.61) below are commutative.



(16.60)



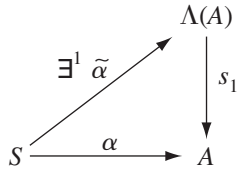
(16.61)

This looks like there is a fair amount of work to do. Especially the triple iterate at the upper right hand corner of the left diagram in (16.60) could seem a bit intimidating. Actually everything has already been proved. Start with the commutativity of the right hand part of diagram (16.61). This is just saying (in a slightly fancy way) that σ_t is a morphism of λ-rings (with the fact that it is a ring morphism coming from the requirement that things take place in the category of rings). The left hand part of (16.61) just says that $\lambda^1 = \mathrm{id}$.

As AH is the morphism determined by the exterior power structure on $\Lambda(A)$ (see the proof of 16.43) the commutativity of the left diagram of (16.60) is just proving that $\Lambda(A)$ is a λ-ring. The remainder of (16.60) is part of the definition of AH (right half) and a consequence of that (see 16.45). The right half of the right diagram of (16.60) is also the first part of the requirement that $\Lambda(A)$ be a λ-ring.

16.62. *Theorem. Cofreeness of the Witt vectors.* The λ-ring $\Lambda(A)$ of Witt vectors over A, together with the first ghost component morphism $s_1 : \Lambda(A) \longrightarrow A$ is the cofree λ-ring over A.

This means that if S is a λ-ring and $\alpha\colon S \longrightarrow A$ is a morphisms of rings (i.e. forget the exterior product structure on S for the moment), then there is a unique morphism of λ-rings $\tilde{\alpha} : S \longrightarrow \Lambda(A)$ that lifts α. This is illustrated in the following diagram.



(16.63)

By now this is simple to prove. First remark that if

$$a(t) = 1 + a_1 t + a_2 t^2 + a_3 t^3 + \cdots = \prod_i (1 - \xi_i t)^{-1} \in \Lambda(A)$$

then by the definition of the λ-ring structure on $\Lambda(A)$ (see (16.8)

$$\lambda^n a(t) \equiv 1 + e_n(\xi)t \equiv 1 + \Big((-1)^{n+1}\left(h_n\left(\xi\right) + P(h_1(\xi), \cdots, h_{n-1}(\xi)\right)\Big)t \mod t^2$$

for some polynomial P with integer coefficients. And so

$$s_1\big(\lambda^n a(t)\big) = (-1)^{n+1} a_n + P\ (a_1, \cdots, a_{n-1}) \tag{16.64}$$

Now let $\alpha : S \longrightarrow A$ be a morphism of rings. Define $\tilde{\alpha} : S \longrightarrow \Lambda(A)$ as the composition $S \xrightarrow{\sigma_t} \Lambda(S) \xrightarrow{\Lambda(\alpha)} \Lambda(A)$. Then this $\tilde{\alpha}$ does the job, and, with induction, it is also the only possibility because it is required to be a morphism of λ-rings and (16.64).[72]

16.65. λ-ring structure on **Symm** itself. There is a very simple and obvious λ-ring structure on the ring $\mathbf{Z}[\xi] = \mathbf{Z}[\xi_1, \xi_2, \xi_3, \cdots]$ of polynomials over the integers, viz the one determined by

$$\lambda^1(\xi_i) = \xi_i, \quad \lambda^j(\xi_i) = 0 \quad \text{if } j \geq 2 \tag{16.66}$$

The easiest way to prove that there is indeed such a λ-ring structure is to consider the ring endomorphisms given by

$$\Psi^n(\xi_i) = \xi_i^n \tag{16.67}$$

These ring endomorphisms satisfy all the requirements of the Wilkerson theorem 16.42 and so there is a λ-ring structure on $\mathbf{Z}[\xi]$ for which the ring endomorphisms (16.67) are the associated Adams operations. The relations between the Adams operations and the exterior product operations are the usual ones linking power sums and elementary symmetric functions. In particular

$$n!\lambda^n = \det \begin{pmatrix} \Psi^1 & 1 & 0 & \cdots & 0 \\ \Psi^2 & \Psi^1 & 2 & \ddots & \vdots \\ \Psi^3 & \Psi^2 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \Psi^1 & n-1 \\ \Psi^n & \Psi^{n-1} & \ldots & \Psi^2 & \Psi^1 \end{pmatrix} \tag{16.68}$$

and it follows immediately that the λ-operations (exterior powers) is casu are indeed those of (16.66).

[72] Again the proof would be more elegant if things were formulated for σ-rings.

Now consider the subring of the symmetric functions $\mathbf{Z}[h] \subset \mathbf{Z}[\xi]$. Obviously this subring is stable under the ring endomorphisms (16.67), so applying the Wilkerson theorem 16.42 again there is a λ-ring structure on $\mathbf{Z}[h]$ for which the morphisms (16.66) are the associated Adams operations which is obviously the restriction of the λ-ring structure (16.65) to the symmetric functions.

16.69. *λ-ring structure on* **QSymm**. In between the symmetric functions and $\mathbf{Z}[\xi]$ there sit the quasi-symmetric functions $\mathbf{Z}[h] = \mathbf{Symm} \subset \mathbf{QSymm} \subset \mathbf{Z}[\xi]$. Obviously, **QSymm** is stable under the ring morphisms (16.67) and so there is a corresponding λ-ring structure on **QSymm**, which is the restriction of the one on $\mathbf{Z}[\xi]$ and which extends the one just discussed on **Symm**. The associated Adams operations are the ring morphisms (16.67) which are the same as the Frobenius endomorphisms of 13.33.

16.70. *Lyndon words.* The proper tails of a word $\alpha = [a_1, a_2, \cdots, a_m]$, $a_i \in \mathbf{N}$, over the natural numbers are the words $[a_i, a_{i+1}, \cdots, a_m]$, $1 < i \leq m$. A word is Lyndon if it is lexicographically smaller than each of its proper tails. For instance [1, 3, 2] is Lyndon and [2, 1, 1] and [1, 1] are not Lyndon. A word α is primitive if the greatest common divisor of its entries is 1.

16.71. *Theorem,* [195, 198] (*Free polynomial generators for* **QSymm** *over the integers*). The elements $\lambda^n\alpha$, $n \in \mathbf{N}$, α primitive word, form a free polynomial basis over the integers of the ring **QSymm**.

This has been a vexing problem since 1972 which was finally solved in 1999, see [193]. **QSymm** is a strong candidate for a theory of noncommutative Witt vectors (as a tool for classifying noncommutative formal groups amoung other things).

16.72. *Discussion.* The many different (?) operations on **Symm**. There are by now some five potentially different unary operations on **Symm** and it is perhaps wise to list them.

(a) The exterior product operations that define the λ-ring structure on **Symm**. These are not additive of course.

(b) The corresponding Adams operations as described by (16.66)

(c) The rings $\Lambda(A)$ are functorial λ-rings. The functor $A \mapsto \Lambda(A)$ is represented by **Symm**: $\Lambda(A) = \mathbf{CRing}(\mathbf{Symm}, A)$. The functorial exterior product operations on $\Lambda(A)$ must therefore come from ring endomorphisms of **Symm**. These endomorphism cannot be Hopf algebra endomorphisms (because otherwise the functorial exterior product operations on $W(A)$ would be additive.

(d) The rings $\Lambda(A)$ have functorial Adams operations. These also must come from ring endomorphisms of **Symm**. Moreover these must be Hopf algebra endomorphisms and even coring object morphisms in the category of rings.

(e) The Frobenius endomorphisms as defined in subsection 13.6. These are the ones that induce the functorial Frobenius operarions on the $\Lambda(A)$.

The first Adams = Frobenius theorem, 16.22, says that the morphisms (d) and (e) are the same. By the remarks already made the only one which could conceivable also be the same is (b). And this is indeed the case

16.73. *Theorem.* (Second Adams = Frobenius theorem). The Adams operations Ψ^n coming from the λ-ring structure on **Symm** are the same as the endomorphisms $\mathbf{f}_n$ of **Symm** that induce the Frobenius endomorphisms of the functor $\Lambda(A)$ of the big Witt vectors.

PROOF. This follows immediately from the fact that $\mathbf{f}_n(p_m) = p_{nm}$ compared with (16.66). □

There is of course still more structure on **Symm**. **Symm** being autodual there are also all the duals of (a)–(e). It is not evident what all these dual operations are.

16.74. *Theorem.* Universal λ-ring on one generator. The ring of symmetric polynomials with the λ-ring structure defined above is the universal λ-ring on one generator.[73]

This means the following. For each λ-ring A and element $a \in A$ there is a unique morphism of λ-rings $\varphi : \mathbf{Symm} \longrightarrow A$ such that $\varphi(e_1) = a$.

PROOF. As is easily verified from e.g. (16.58) the λ-ring structure on **Symm** satisfies $\lambda^n(e_1) = e_n$. It follows that the only ring morphism that could possibly work is defined by φ: $e_n \mapsto \lambda^n(a)$. And there is such a ring morphism because **Symm** is free on the e_n. That this is actually a morphism of λ-rings requires a bit more work, as follows. Let x be an element of **Symm**, i.e. a polynomial in the elementary symmetric functions

$$x = P_x\ (e_1, e_2, \cdots) \equiv P_x\ \left(\lambda^1(e_1), \lambda^2(e_1), \cdots\right)$$ □

Now consider $\lambda^n(x)$. Because composition of lambda operations, a lambda operation applied to a product, and a lambda operation applied to a sum, are given by 'universal polynomials', that means the same polynomials for any λ-ring, see 16.47, there is a universal polynomial Q_{n,P_x} (with coefficients in the integers)[74] such that for any λ-ring and any element a in it

$$\lambda^n(P_x(\lambda^1(a), \lambda^2(a), \cdots)) = Q_{n,P_x}(\lambda^1(a), \lambda^2(a), \cdots)$$

Also φ is a ring homomorphism and so commutes with polynomials

$$\varphi(Q(x_1, x_2, \cdots)) = Q(\varphi(x_1), \varphi(x_2), \cdots)$$

[73] There is a far reaching generalization. As will be discussed below **Symm** is isomorphic to RS the direct sum of the representation rings of the symmetric groups. For a fixed finite group G let $G \times_{wr} S_n$ be the wreath product of G and S_n. The direct sum of the representation rings $R(G \times_{wr} S_n)$ is the free lambda ring on the irreducible representations of G as generators, [280].

[74] This polynomial is in fact the plethysm $e_n \circ P_x$, see subsection 16.76 below.

It follows that φ commutes with the lambda operations, so that it is a morphism of λ-rings.

16.75. *Comonadability vs cofreeness vs representability vs freeness.* The various notions that passed under review in the previous few subsections are far from being unrelated. The following matters were discussed.

(a) The functor $\Lambda : \mathbf{CRing} \longrightarrow \lambda - \mathbf{Ring}$ takes a ring A into the cofree λ-ring over A.
(b) The functor Λ comes with a comonad structure and the coalgebras of this comad are precisely the λ-rings.
(c) The functor Λ is representable. The representing ring is a λ-ring and is the free λ-ring on one generator.

The relation (a)–(b) is a well known part of monad and comonad theory (from the early days of this theory). Every monad or comonad comes from an adjoint pair of functors, see [284, 33], and of course cofreeness of $\Lambda(A) \longrightarrow A$ translates into the observation that Λ is right adjoint to the functor λ-**Ring** $\longrightarrow$ **CRing** that forgets about the exterior product structure.

Here is how the relation (b)–(c) goes.[75] Let $\mathcal{C}$ be a category and let (T, μ, ε) be a comonad in $\mathcal{C}$. Now let $(Z, z \in T(Z))$ represent the functor T. That is, there is a functorial bijection $\mathcal{C}(Z, A) \longrightarrow T(A)$, $f \mapsto T(f)(z)$. The comonad structure gives in particular a morphism $\sigma\colon Z \longrightarrow TZ$, viz the image of id_Z under $\mu_Z : T(Z) = \mathcal{C}(Z, Z) \longrightarrow T(T(Z)) = \mathcal{C}(Z, T(Z))$. This defines a 'coalgebra for T' structure on Z. Now let (A, σ) be a coalgebra for the comonad T and let a be an element of A. Consider the element $\sigma(a) \in T(A) = \mathcal{C}(Z, A)$. This gives a unique morphism of T-coalgebras that takes z into a. There are of course a number of things to verify both at this categorical level and to check that these categorical considerations fit with the explicit constructions carried out in the previous subsections. This is straightforward.

16.76. *Plethysm.* (Very partial symmetric function formularium (4)). Given an element $a \in A$ in a λ-ring A and a polynomial $f \in \mathbf{Symm}$ define

$$\alpha_a(f) = \beta_f(a) = f\Big(\lambda^1 a, \lambda^2 a, \cdots\Big) \tag{16.77}$$

where in the last part of (16.77) f is seen as a polynomial in the elementary symmetric functions; i.e. in the expression for f as a polynomial in the elementary symmetric functions these elementary symmetric functions are replaced by the exterior powers of the element a. If f is seen as a polynomial in the complete symmetric functions the same result is obtained by replacing these with the symmetric powers of the element a. This can lead to confusion so it is better to have a description that does not depend

[75] I have not found this relation in the monad literature (which does not mean it is not there somewhere). There should be a general theorem to this effect also in a more general context than 'categories of sets with structure.'

on what free polynomial basis is used for **Symm**. That goes as follows. Give $a \in A$ and $f \in \mathbf{Symm}$

$$\begin{aligned}&\text{Let } \alpha_a : \mathbf{Symm} \longrightarrow A \text{ be the unique morphism of } \lambda\text{-rings}\\ &\text{such that } \phi_a(h_1) = \phi_a|e_1| = a. \text{ Then } \alpha_a(f) = \beta_f(a)\end{aligned} \tag{16.78}$$

This looks at $a \in A$ as something that defines (and is defined by) a morphism of λ-rings from the free λ-ring on one generator to A.

Another way to look at (16.77) is as the definition of a functorial operation on λ-rings defined by the polynomial f. This is the most used way to look at (16.77), (16.78).

In particular **Symm** itself is a λ-ring, and so, taking a $g \in \mathbf{Symm}$ for a in (16.77), (16.78) there results a new kind of composition of polynomials

$$f \circ g = \alpha_g(f) = \beta_f(g) = f\big(\lambda^1 g,\ \lambda^2 g, \cdots\big) \tag{16.79}$$

This composition law on **Symm** is called plethysm. More precisely it is called 'outer plethysm.' There is also something called 'inner plethysm' which has to do with the λ-ring structures on the homogeneous summands $\mathbf{Symm}^{(n)}$ of **Symm**.[76] In the representation theoretic incarnation of **Symm** in terms of representations of the general linear groups, outer plethysm corresponds to composition of representations.

In terms of the λ-ring interpretation mentioned above it also corresponds to composition: the functorial operation on λ-rings defined by $f \circ g$ on λ-rings is the composition of the functorial operations defined by f and g:

$$\beta_{f\circ g}(a) = \beta_f\big(\beta_g(a)\big) \tag{16.80}$$

Plethysm is associative

$$(f \circ g) \circ h = f \circ (g \circ h) \tag{16.81}$$

This is a special case of (16.80).

16.80. *Calculation of plethysms.* Write $g \in \mathbf{Symm} \subset \mathbf{Z}$ as a sum of monomials in the ξ $g = \sum_\tau c_\tau \xi^\tau$ and write (formally)

$$\prod (1 + y_i t) = \prod_\tau \big(1 + \xi^\tau t\big)^{c_\tau} \tag{16.81}$$

then

$$f \circ g = f(y_1, y_2, \cdots) \tag{16.82}$$

where f is seen as a symmetric polynominal in the ξ.

---

[76] These homogeneous summands are rings under the second multiplication on **Symm**. In the representation theoretic incarnation of **Symm** they correspond with the ring of (virtual) representations $R(S_n)$ of the symmetric groups S_n. The exterior powers there are the exterior powers of representations.

Thus for instance $f \circ p_n = f\big(\xi_1^n, \xi_2^n, \cdots\big)$, a plethysm that was used in 9.93 above when discussing the Witt symmetric functions which give the Witt coordinates of a power series.

To see that (16.82)–(16.83) fits with the definition, first note that for the λ-ring structure on $\mathbf{Z}[\xi]$, $\lambda^1(\xi_i) = \xi_i$, $\lambda^n(\xi_i) = 0$ for $n \geq 2$. Thus $\sigma_t(\xi_i) = (1 - \xi_i t)^{-1}$ and because σ_t is a morphism of rings

$$\sigma_t(g) = \prod_\tau \big(1 - \xi^\tau t\big)^{-c_\tau} \tag{16.83}$$

and so

$$\lambda_t(g) = \sigma_{-t}(g)^{-1} = \prod_\tau \big(1 + \xi^\tau t\big)^{c_\tau} = \prod (1 + y_i t)$$

so that the elementary symmetric functions in the y are the $\lambda^n g$.

16.84. *Distributivity of the functorial plethysm operations.* The plethysm operations β_f on λ-rings A are not additive. But they are distributive over addition and multiplication in a Hopf algebra way as follows. Let

$$\mu_S(f) = \sum_i f'_{S,i} \otimes f''_{S,i}, \quad \mu_P(f) = \sum_i f'_{P,i} \otimes f''_{P,i} \tag{16.85}$$

Then

$$\beta_f(a+b) = \sum_i f'_{S,i}(a) f''_{S,i}, (b), \quad \beta_f(ab) = \sum_i f'_{P,i}(a) f''_{P,i}(b) \tag{16.86}$$

This is seen as follows given $a, b \in A$, they define Witt vectors $\sigma_t(a), \sigma_t(b) \in \Lambda(A)$ which, in turn, are morphisms of rings $\mathbf{Symm} \longrightarrow A$. Because σ_t is a morphism of rings the Witt vectors of $a+b$ and ab are the Witt vector sum and Witt vector product of the Witt vectors of a and b. As morphisms of rings $\mathbf{Symm} \xrightarrow{\varphi_a, \varphi_b} A$. Witt vectors are summed and multiplied as follows. The sum Witt vector corresponds to the composite

$$\mathbf{Symm} \xrightarrow{\mu_S} \mathbf{Symm} \otimes \mathbf{Symm} \xrightarrow{\varphi_a \otimes \varphi_b} A \otimes A \xrightarrow{m_A} A$$

and the product Witt vector corresponds to the composite

$$\mathbf{Symm} \xrightarrow{\mu_P} \mathbf{Symm} \otimes \mathbf{Symm} \xrightarrow{\varphi_a \otimes \varphi_b} A \otimes A \xrightarrow{m_A} A$$

Formula (16.86) follows.

For those polynomials for which there are nice formulas for the sum comultiplication morphism μ_S or for the product comultiplication μ_p (16.85) and (16.86) can give useful formulas for calculating plethysm operations. For instance, using 10.7

$$s_\kappa \circ (f + g) = \sum_{\lambda \subset \kappa} (s_\lambda \circ f)\,(s_{\kappa/\lambda} \circ g) \tag{16.87}$$

It also follows (again) from (16.85), (16.86) that the plethysm operations β_{p_n} on λ-rings defined by the power sum symmetric functions p_n are both additive and

multiplicative; they are of course the Adams operations (explaining once more why these are often called "power operations.'

Rather few of the plethysm operations defined by a polynomial are additive: only those that come from a sum of power sum symmetric functions. On the other hand there are all the additive operations generated by the Frobenius, Verschiebung and homothety operations (the Cartier ring). It is unknown whether all these together with the plethysm operations generate all operations on the functor Λ.

There is quite a bit of literature on plethysm, mostly in a representation theoretic context and much is on how to calculate it in special cases. A sampling is [8, 66, 82, 282, 313, 394, 395], [312, 306, 43, 42, 429].

17. Necklace rings

There is (for suitable rings) a third coordinatization of the unital power series over a ring A, i.e. the elements of $\Lambda(A) = 1 + tA[[t]]$, besides the power series coordinates and Witt vector coordinates considered so far. These coordinates go by the name necklace coordinates. The three systems of coordinates are related by

$$1 + a_1 t + a_2 t^2 + a_2 t^2 + \cdots = \prod_{i=1}^{\infty} \left(1 - x_i t^i\right)^{-1} = \prod_{i=1}^{\infty} \left(1 - t^i\right)^{-c_i} \qquad (17.1)$$

Unlike in the case of Witt vector coordinates it is not always possible to find necklace coordinates. It can certainly be done for the case that A is a $\mathbf{Q}$-algebra, but there is but little interest in that as in that case

$$W(A) \cong \Lambda(A) \cong Gh(A) = A^{\mathbf{N}}$$

But there are quite a few rings A that are not $\mathbf{Q}$-algebras for which such a representation in necklace coordinates can always be found with the $c_i \in A$; in particular the rather important case $A = \mathbf{Z}$. This will be discussed a bit further below. On the other hand this can not be done for the case that A is a ring of polynomials, in particular not for $A = \mathbf{Symm}$ which is the context in which the universal example of a one power series lives; that is the power series

$$h(t) = 1 + h_1 t + h_2 t^2 + h_3 t^3 + \cdots \qquad (17.2)$$

from which all others are obtained.

Why I like to call the coordinates on the right hand side of (17.1) by the name necklace coordinates will become apparent in a minute.

17.3. *Necklace polynomials.* Consider a totally ordered alphabet of α letters, say $\{1, 2, \cdots, \alpha\}$. A word over this alphabet is primitive (also called aperiodic) if it is not a concatenation power of a strictly smaller word. A word is Lyndon if it is strictly smaller in the lexicographic order than any of its (non-identity) cyclic permutations. So in particular it is primitive. For the equivalence of this definition of Lyndon word with the one used in 16.71 see e.g. [273], Section 4.4.

A necklace (also called circular word) is an equivalence class under cyclic permutations of words. A necklace is primitive if the words in the equivalence class are primitive. Thus the Lyndon words can be regarded as a systematic choice of representatives of primitive necklaces.

The formula for the number of primitive necklaces of length n on α letters, i.e. with α different colors of beads, has been known since Colonel Moreau, [298], in 1872. It is

$$M(\alpha;\, n) = n^{-1} \sum_{d|n} \mu(d)\alpha^{n/d} \tag{17.4}$$

These expressions, seen as polynomials in an indeterminate α are known as the necklace polynomials. Note that though these polynomials are integer valued for every integer argument they do not have integer coefficients. For instance for a prime number p

$$M(\alpha;\, p) = p^{-1}(\alpha^p - \alpha) \tag{17.5}$$

The same expression (17.4) turns up in other contexts. For instance in the theory of free Lie algebras. Consider the free Lie algebra generated by α symbols. Give each symbol weight one. The free Lie algebra (over the integers or any field) is then graded and the graded part of weight n has rank $M(\alpha, n)$. This can be seen for instance by the so called Hall set construction of a basis (as an Abelian group) for the free Lie algebra, using the set of Lyndon words as a Hall set. See Chapter 5 in [334] for details. In this context of free Lie algebras (17.4) is known as the Witt formula, [418].

In loc. cit., p. 153, Witt writes: "Es ist merkwürdig, dass diese Rangformel übereinstimmt mit der bekannten Gausschen Formel für die Anzahl der Primpolynome $x^n + a_1 x^{n-1} + \cdots + a_n$ im Galoisfeld von q Elementen." Later, Solomon Golomb, [174], found indeed a bijection between primitive necklaces and irreducible polynomials. This correspondence is not yet entirely satisfactory in that it depends on the choice of a primitive element for the Galois field. See also [335] for a discussion of this correspondence and other occurences of expression (17.4).

Consider the power series in Witt coordinates

$$\sum_{n\geq 0} \beta_n t^n = \prod_{n\geq 1} \left(1 - M(\alpha, n)t^n\right)^{-1} \tag{17.6}$$

Then β_n is the rank of the homogeneous component of degree n of the free associative algebra over $\mathbf{Z}$ in α symbols. This results from the Witt formula by using the Poincaré-Birkhoff-Witt theorem combined with the Milnor-Moore type result that says that the free associative algebra is the universal enveloping algebra of the free Lie algebra on the same set of symbols. For some further manifestations of the necklace polynomials and related expressions in such varied fields of inquiry as the Feynman identity, the elliptic modular function, multiple zeta values, (symbolic) dynamics and fixed points, formal groups, see also [70, 101, 113, 218, 234, 261, 260, 292, 299, 316, 317].

17.7. *Necklace polynomial formulas.* Here are two interesting formulas from [291].

$$M(\alpha\beta; n) = \sum_{[i,j]=n} (i,\, j)M(\alpha;\, i)M(\beta;\, j) \tag{17.8}$$

where (i, j) denotes the greatest common divisor of i and j, and $[i, j]$ denotes their least common multiple.

$$M(\beta^r; n) = \sum \frac{j}{n} M(\beta; j) \tag{17.9}$$

where the sum ranges over all j such that $[j, r] = nr$ (which implies that j is a multiple of n).

17.10. *Cyclotomic identity.* The so called cyclotomic identity is

$$\frac{1}{(1-\alpha t)} = \prod_{n=1}^{\infty} \left(\frac{1}{1-t^n} \right)^{M(\alpha;n)} \tag{17.11}$$

(It is a tradition to write it precisely this way.) All three identities (17.8), (17.8), (17.11) are identities about polynomials. As such it suffices to prove them for integer values of the variables α, β which is (usually) done by combinatorial means. Cf e.g. [290, 291].

There is a very elegant 'symmetric' generalization of the cyclotomic identity due to Volker Strehl, [375]

$$\prod_{n=1}^{\infty} \left(\frac{1}{1-\alpha t^n} \right)^{M(\beta;n)} = \prod_{n=1}^{\infty} \left(\frac{1}{1-\beta t^n} \right)^{M(\alpha;n)} \tag{17.12}$$[77]

There are some other generalizations which will be described later in a section devoted to generalized Witt vectors.

17.13. *Motivational remarks regarding the necklace algebra functor.* Now consider power series expressed by means of necklace coordinates as in (the right hand side of) 17.1. Multiplying two such expression means adding their necklace coordinates. The multiplication defined on the Witt vector ring is determined by

$$\frac{1}{(1-\alpha t)} * \frac{1}{(1-\beta t)} = \frac{1}{(1-\alpha\beta t)} \tag{17.14}$$

Taking into account distributivity, the cyclotomic identity (17.11) and formula (17.9) this dictates the following definition of the necklace algebra (and its ghost component morphisms).

17.15. *Definition of the necklace algebra functor.* As an Abelian group the necklace ring over a ring A is the infinite product $A^{\mathbf{N}}$ of all sequences $(a_1, a_2, a_3, \ldots)$, $a_i \in A$ with component wise addition. Two such sequences are multiplied according to the rule

$$\begin{aligned} (a_1, a_2, a_3, \cdots) * (b_1, b_2, b_3, \cdots) &= (c_1, c_2, c_3, \cdots) \\ c_n &= \sum_{[i,j]=n} (i, j) a_i b_j \end{aligned} \tag{17.16}$$

[77] This has the flavour of a 'reciprocity formula' and one wonders if it relates to some other 'reciprocity results' in mathematics.

This is clearly functorial and this functor will be denoted $Nr : \mathbf{CRing} \longrightarrow \mathbf{CRing}$. There is a functorial ghost components ring morphism $u : Nr(A) \longrightarrow Gh(A) = A^{\mathbf{N}}$ to the direct product of $\mathbf{N}$ copies of the ring A defined by

$$u(a) = (u_1(a), u_2(a), u_3(a), \cdots), \quad u_n(a) = \sum_{d|n} da_d \tag{17.17}$$

Given a power series for which necklace coordinates exist, so that (17.1) holds, one then has for these coordinates

$$w(x) = s(a) = u(c) \tag{17.18}$$

in the ghost ring $Gh(A) = A^{\mathbf{N}}$ (with component-wise addition and multiplication).

Necklace rings (in the sense defined above[78] are a special kind of convolution rings as defined in [403, 402, 405]. For some first investigations in the algebraic theory of necklace rings see [404, 322].

17.19. *Binomial rings.* Obviously, from (17.1), necklace coordinates are hardly compatible with the presence of torsion.

Further as

$$(1 - t^n)^x = 1 - xt^n + \frac{x(x-1)}{2!}t^{2n} - \frac{x(x-1)(x-2)}{3!}t^{3n} + \cdots \tag{17.20}$$

it seems necessary for necklace coordinates to exist over a ring A to require that together with an $x \in A$ it also contains the binomial coefficients $(n!)^{-1}x(x-1)\cdots(x-n+1)$. This leads to the idea of a binomial ring.

A (commutative unital) ring A is said to be binomial if it is torsion free (as a $\mathbf{Z}$-module) and if together with any x in it it also contains the binomial coefficients

$$\binom{x}{n} = \frac{x(x-1)(x-2)\cdots(x-n+1)}{n!} \tag{17.21}$$

These rings have made their appearance before in 1958, in the fundamental work of Philip Hall on the theory of nilpotent groups, [186], often referred to as the 'Edmonton notes."

Obviously any $\mathbf{Q}$-algebra is binomial. But there are many rings that are binomial that are not algebras over the rationals. First and foremost the ring of integers Z and certain rings of functions with values in the integers (under pointwise addition and multiplication), notably polynomials.

17.22. *Free binomial rings.* Let $X = \{X_i : i \in I\}$ be a set (of indeterminates). By definition $IVal[X]$ is the ring of all polynomials with rational coefficients that take integer values on integers; i.e. integer valued polynomials. These are also sometimes called numerical polynomials.

[78] There is another kind of object in this more or less same corner of algebra that has 'necklace' in its name. Viz necklace Lie algebras, necklace Hopf algebra, necklace Lie coalgebra, see [55, 151, 170, 352]. These have little to do with the necklace approach to Witt vectors.

Examples are the binomial coefficients polynomials

$$\binom{X_i}{n} = \frac{X_i(X_i-1)(X_i-2)\cdots(X_i-n+1)}{n!} \tag{17.22}$$

and it is a theorem that the monomials in these form a basis of $IVal[X]$ over the integers (as an Abelian group), see [45], §45, p. 240ff.

These are indeed the free binomial rings in the technical sense that the functor $IVal : \mathbf{Set} \longrightarrow \mathbf{BinRing}$ (where **BinRing** stands for the sub category of binomial rings in **CRing**) is left adjoint to the forgetful functor the other way, see [134], p. 168, proposition 2:

$$\mathbf{BinRing}(IVal[X], A) \cong \mathbf{Set}(X, A) \tag{17.23}$$

A monograph on integer valued polynomials is [80].[79]

Other examples of binomial rings are the p-adic integers, the profinite completion $\hat{\mathbf{Z}}$ of the integers and localizations of the integers, and a special kind of λ-rings.

17.24. *Binomial rings vs λ-rings.* Let A be a torsion free ring such that $x^p \equiv x \mod p$ for all prime numbers p. Then, taking $\varphi_n = \mathrm{id}$ for all n the conditions of the Wilkerson theorem 16.42 are satisfied and so there is a λ-ring structure on A for which all the Adams operations are the identity. For this λ-ring structure

$$\lambda^n(x) = \binom{x}{n}, \quad \lambda_t(x) = (1+t)^x, \quad \sigma_t(x) = (1-t)^{-x} \tag{17.25}$$

and thus A is a binomial ring. To see (17.25) simply calculate the determinant of the matrix $M_x = (m_{i,j})$, $m_{i,j} = x$ for $i \geq j$, $m_{i,i+1} = i$, $m_{i,j} = 0$ for $j \geq i+2$, see (16.41). Actually one can do a little better and prove that a λ-ring for which all the Adams operations are equal to the identity is automatically torsion free, see [134], section 5.

Inversely let A be a binomial ring then taking the (tentative) exterior product operations to be the binomial coefficients one gets a ring with exterior products with associated Adams operations that are all the identity and hence a λ-ring structure.

17.26. *Frobenius and Verschiebung on necklace rings.* From the above it is clear that for binomial rings necklace coordinates always exist and that the necklace ring over a binomial ring is isomorphic (functorially) to the ring of Witt vectors over that ring.

Using this isomorphism one can of course transfer the Frobenius and Verschiebung operations to necklace rings. Actually these are always defined. Here are the explicit formulas.

$$\text{For } c = (c_1, c_2, c_3, \cdots) \in Nr(A), \quad \mathbf{f}_r c = (b_1, b_2, b_3, \cdots)$$
$$b_n = \sum_{[j,r]=nr} n^{-1} j c_j \quad \mathbf{V}_r c = (\underbrace{0, 0, \cdots, 0, c_1}_{r}, \underbrace{0, 0, \cdots, 0, c_2}_{r}, \cdots) \tag{17.27}$$

[79] As far as I know this is also the only monograph on the topic.

and one easily proves the usual formulas; either directly or via the ghost components or via …, such as

$$\mathbf{V}_r\mathbf{V}_s = \mathbf{V}_{rs},\ \mathbf{f}_r\mathbf{f}_s = \mathbf{f}_{rs},\ \mathbf{f}_r\mathbf{V}_r = [r],\ \mathbf{f}_r\mathbf{V}_s = \mathbf{V}_s\mathbf{f}_r \text{ if } (r,s) = 1 \tag{17.28}$$

17.29. *Witt vectors vs necklaces in general.* Let A be a torsion free ring.[80] Consider the mapping

$$\psi_A : Nr(A) \longrightarrow \Lambda(A),\ (c_1, c_2, c_3, \cdots) \mapsto \prod_{n=1}^{\infty}(1-t^n)^{-c_n} \tag{17.30}$$

This is by the very definition of addition and multiplication on $Nr(A)$ a morphism of rings and it is compatible with the Frobenius and Verschiebung morphims on the two sides.

When A is a binomial ring it is an isomorphism. When A is not binomial this is not the case. However, it is an isomorphism for $A_{\mathbf{Q}} = A \otimes \mathbf{Q}$ and it is (of course) dead easy to describe the subring of $Nr(A_{\mathbf{Q}})$ which corresponds to the subring $\Lambda(A) \subset \Lambda(A_{\mathbf{Q}})$.

For each $\alpha \in A$ let $M(\alpha)$ be the necklace vector

$$M(\alpha) = (M(\alpha;1), M(\alpha;2), M(\alpha;3), \cdots) \in Nr(A_{\mathbf{Q}}) \tag{17.31}$$

Then by the cyclotomic identity and the definition of the Verschiebung operations

$$\psi_{A_{\mathbf{Q}}}(M(\alpha)) = (1-\alpha t)^{-1}, \quad \psi_{A_{\mathbf{Q}}}(\mathbf{V}_r M(\alpha)) = (1-\alpha t^r)^{-1} \tag{17.32}$$

and it follows that $\psi_{A_{\mathbf{Q}}}$ induces an isomorphism

$$\left\{ \sum_{r=1}^{\infty} \mathbf{V}_r M(\alpha_r) : \alpha_i \in A \right\} \xrightarrow{\psi_{A_{\mathbf{Q}}}} \Lambda(A) \subset \Lambda(A_{\mathbf{Q}}) \tag{17.33}$$

By the left half of (17.32) it is clear that the necklace vectors $M(\alpha)$ are the analogs of Teichmüller representatives.

For a binomial ring, like the integers, it follows from (17.31)–(17.33) that each necklace vector can be written uniquely as an infinite sum

$$\sum_{r=1}^{\infty} \mathbf{V}_r M(\alpha_r),\ M(\alpha_r) = (M(\alpha_r;1), M(\alpha_r;2), M(\alpha_r;3), \cdots) \in Nr(A) \tag{17.34}$$

17.35. *Modified necklace rings.* The description (17.33) of the ring of Witt vectors in terms of necklace vectors is not very satisfactory or elegant. One can in fact do a

[80] As far as I know no investigations have been carried out on the necklace ring over A when A is not torsion free, for instance when A is a finite field. Yet it is in that setting that something interesting may happen that is different from what happens in the case of the Witt vectors. Also the analogue of the p-typification endomorphism defined by the Frobenius and Verschiebung operations has not been investigated for necklace rings.

good deal better using modified necklace polynomials, [316]. In loc. cit. Young-Tak Oh defines modified necklace polynomials over a λ-ring A which explicitly involve the Adams operations as follows

$$M(r; n) = n^{-1} \sum_{d|n} \mu(d) \Psi^d(r^{n/d}) \tag{17.36}$$

He also introduces much related[81] multivariable versions of the necklace polynomials

$$M(X; n) = n^{-1} \sum_{d|n} \mu(d)\, p_d(X) \tag{17.37}$$

where the p_d's are the power sums in the X's. These polynomials satisfy very similar properties to (17.7) and (17.8) and lead to a definition of necklace rings for λ-rings which are isomorphic to the rings of Witt vectors.

It would be very nice if there were combinatorial interpretations of these modified necklace polynomials.

17.38. *Adjoints of the inclusion* **BinRing** $\subset$ **CRing**. Now consider the inclusion of the binomial rings into the category of rings. This inclusion functor has both a left adjoint (free objects), Bin^U, and a right adjoint (cofree objects), Bin_U, characterized and defined by the functorial properties

$$\mathbf{BinRing}(Bin^U(A), B) \cong \mathbf{CRing}(A, B), \mathbf{BinRing}(B, Bin_U(A)) \cong \mathbf{CRing}(B, A)$$

for binomial rings B and rings A.

The construction of $Bin^U(A)$ is much like $IVal[X]$ compared to $\mathbf{Z}[X]$. For $Bin_U(A)$ take $\Lambda(A)$ and take the subring of all elements on which all the Frobenius operations are the identity. For more details and proofs see [134].

17.39. The functors *BinW*. The sum comultiplication μ_s and the product comultiplication μ_p on $\mathbf{Z}[h]$ that define the ring valued functor of the big Witt vectors, extend uniquely to $IVal[h_1, h_2, h_3, \cdots]$ and $IVal[h_1, h_2, h_3, \cdots, h_n]$ to define coring objects and hence ring-valued sub functors of the Witt vectors and the truncated Witt vectors. These would seem to merit some investigation. Even the simplest one $IVal[h_1]$ already has thought provoking properties, see [45], p. 241ff.

17.40. *Carryless Witt vectors.* Like the real numbers in decimal notation the Witt vectors, when added or multiplied, involve 'carry-overs.' The necklace approach to them can be used to describe a carry-less version, [291]. The price, however, is high, too high in my opinion, in that Witt vectors in this approach are seen as equivalence classes rather than single objects.

[81] This again involves the splitting principle technique/philosophy.

17.41. *To ponder and muse about.* A nice theory of necklace rings as in [316] works precisely for λ-rings, i.e. the very objects which are coalgebras for the main functor involved, the one of the Witt vectors. There is something circular about this; almost incestuous, of much the same flavour as one often meets in various parts of category theory. To me this is something that needs to be pondered and mulled over a bit.

17.42. *Apology.* Within the published literature on necklace algebras one finds the notion of what are called aperiodic rings. The formulas involved are, in my opinion, essentially empty given the necklace formulas and, better, Witt vector formulas. So in this chapter 1 will not discuss these aperiodic rings.

17.43. A selection of references on necklaces and necklace algebras is [21, 22, 55, 70, 115, 120, 144, 145, 174, 260, 261, 291, 294, 303, 314, 315, 316, 322, 329, 399, 401, 404].

18. Symm vs $\bigoplus_n R(S_n)$

That symmetric functions and representations of the symmetric groups representations of the general linear groups have much to do with one another has been known since the early days of the previous century (Alfred Young, Issai Schur, Georg F Frobenius, ...). The realization that these things become more elegant and better understandable from the Hopf algebra point of view is of a more recent date, [269, 268, 239, 425]. Here, mostly, the latter approach will be outlined. For more details see [199], Chapter 4. First, however, here is a streamlined version of what might be called classical Schur-Frobenius theory as presented in [281], pages 112–114.

18.1. *The ring $R(S)$.* Let S_n be the group of permutations on n letters, usually taken to be $\{1, 2, \ldots, n\}$. For each n let $R(S_n)$ be the free Abelian (Grothendieck) group spanned by the irreducible representations of the symmetric group S_n; or, what is the same, the Grothendieck group of isomorphism classes of (complex) representations of S_n, or, what is again the same, the free Abelian group with a basis the irreducible characters of S_n. An element from $R(S_n)$ that comes from an actual representation will be called real (as opposed to virtual (not as opposed to complex)). These are the ones that are nonnegative integral sums of irreducible representation. Other elements of $R(S_n)$ are sometimes referred to as virtual representations.

As an Abelian group $R(S)$ is the direct sum of all these groups of representations:

$$R(S) = \bigoplus_{n=0}^{\infty} R(S_n) \tag{18.2}$$

where, by decree, $R(S_0) = \mathbf{Z}$. A product is defined on $R(S)$ as follows. Take a representation ρ of S_p and a representation σ of S_q. Taking the (outer) tensor product gives a representation $\rho \otimes \sigma$ of $S_p \times S_q$ on the tensor product of the representation spaces of ρ and σ. Now consider $S_p \times S_q$ as the subgroup of $S_n = S_{p+q}$ of those permutations that take the set of the first p elements into itself and that take the set

of the last q elements into itself. Now induce the representation $\rho \otimes \sigma$ up to S_n. Further take $1 \in \mathbf{Z} = R(S_0)$ to be the unit element. This defines an associative and commutative multiplication on $R(S)$ that makes it into a graded ring. As a rule this outer product of two irreducible representations is not irreducible and determining the multiplicities of the irreducible representations that occur in it is and always has been a major part of the representation theory of the symmetric groups (Littlewood-Richardson rule[82]).

18.3. *Scalar product.*[83] If f, g are functions on a finite group G their scalar product is defined by

$$\langle f, g\rangle_G = \frac{1}{\#G}\sum_{x\in G} f(x)g(x^{-1}) \tag{18.4}$$

This scalar product is used to define a scalar product on all of $R(S)$ by

$$\langle f, g\rangle = \sum_{n\geq 0}\langle f_n, g_n\rangle_{S_n} \tag{18.5}$$

where $f = \sum f_n$, $g = \sum g_n \in R(S)$ (and $\langle 1, 1\rangle = 1$ for $1 \in R_0(S) = \mathbf{Z}$).

18.6. *Characteristic map.* Let $w \in S_n$ be a permutation on n letters. It decomposes as a product of disjoint cycles and the lengths of these cycle define a partition $\lambda(w)$ of n called the cycles type of w. Define a mapping

$$\psi : S_n \longrightarrow \mathbf{Symm}_n, \quad w \mapsto p_{\lambda(w)} \tag{18.7}$$

where $p_{\lambda(w)}$ is the power sum monomial, see (9.61), defined by the cycle type of w. The next step is the definition of a morphism of Abelian groups called the characteristic map

$$\mathrm{ch} : R(S) \longrightarrow \mathbf{Symm}_{\mathbf{C}} = \mathbf{Symm} \otimes_{\mathbf{Z}} \mathbf{C} \tag{18.8}$$

as follows. If f is a (virtual) character of S_n

$$\begin{aligned}\mathrm{ch}(f) = \langle f, \psi\rangle_{S_n} &= \frac{1}{n!}\sum_{w\in S_n} f(w)\psi(w^{-1}) \\ &= \frac{1}{n!}\sum_{w\in S_n} f(w)\psi(w) = \sum_{\mathrm{wt}(\lambda)=n} z_\lambda^{-1} f_\lambda p_\lambda \end{aligned} \tag{18.9}$$[84]

where f_λ is the value of f on the cycle class λ and z_λ is the number from (9.62) that gives the scalar product of the power sum monomials with themselves and where it

[82] [230], theorem 2.8.13, p. 93.

[83] This inner product is sometimes called 'Hall inner product.'

[84] Where for the second expression in (18.9) definition (18.4) is extended a bit in that g has its values in **Symm**.

is used that the cycle class of $w \in S_n$ is the same as that of w^{-1}. (Recall that f as a character has the same value on each element of a conjugacy class). It follows that, using also $\langle p_\lambda, p_\mu \rangle = z_\lambda \delta_{\lambda,\mu}$,

$$\langle \mathrm{ch}(f), \mathrm{ch}(g) \rangle = \sum_{\mathrm{wt}(\lambda)=n} z_\lambda^{-1} f_\lambda g_\lambda = \langle f, g \rangle_{S_n} \tag{18.10}$$

and hence that ch is an isometry. The basic theorem is now that

18.11. *Theorem.* The characteristic map is (induces) an isometric isomorphism from $R(S)$ onto **Symm**.

There are a certain number of things to prove. But this is not a text on representation theory, let alone a text on the representation theory of the symmetric groups, that vast and fascinating subject; so I will restrict myself to a brief sketch. First, ch is multiplicative. This is handled by Frobenius reciprocity (which will also turn up later in the Hopf algebra approach).

Next one shows that the identity character η_n, i.e. the character of the trivial representation of S_n corresponds under ch to the complete symmetric function h_n. (The sign character corresponds to the elementary symmetric function e_n.)

Further for each partition λ of n *define*

$$\chi_\lambda = \det\bigl(\eta_{\lambda_i - i + j}\bigr)_{1 \le i,\, j \le n} \in R(S_n) \tag{18.12}$$

These are (possibly virtual) characters of S_n. Their images under ch are the Schur functions and so $\langle \chi_\kappa, \chi_\lambda \rangle = \delta_{\kappa,\lambda}$ and as ch is an isometry they are up to sign irreducible characters. As the number of conjugacy classes of S_n is equal to the number of partitions of n they must form a basis for $R(S_n)$ and so ch indeed induces an isomorphism of $R(S_n)$ onto **Symm**.

It remains to verify that the χ_λ are in fact real characters (as opposed to virtual) which is done by checking their value on the trivial permutations.

18.13. *Enter Hopf algebras.* The topologist knows **Symm** as $H^*(\mathbf{BU}; \mathbf{Z})$, the cohomology of the classifying space **BU** of the unitary group. And he remembers that this is a Hopf algebra. So wouldn't it be nice if $R(S)$ were a Hopf algebra too and if ch were an isomorphism of Hopf algebras. This is indeed the case. The first to notice that $R(S)$ is a Hopf algebra would appear to have been Burroughs, [79]. It was, however, Arunas Liulevicius, [269, 268], who first systematically exploited this point of view. To quote a bit more from the second of his two papers: "The aim of this paper is to present a ridiculously simple proof of a theorem on representations rings of the symmetric groups which concisely presents theorems of Frobenius, Atiyah, and Knutson" (and Schur in my view of things).

Whether this approach is really simpler than the very streamlined presentation of Macdonald outlined above is debatable. A Hopf algebra is a heavy structure. But it certainly brings in new and more functorial and universal points of view, which are, I believe, important.

18.14. *The Hopf algebra structure on $R(S)$.* For a composition $\alpha = [a_1, a_2, \cdots, a_m]$ of weight n the corresponding Young subgroup is $S_\alpha = S_{a_1} \times S_{a_2} \times \cdots \times S_{a_m}$. It consists of all permutations that map all the sets of letters $\{1, \cdots a_1\}, \{a_1+1, \cdots, a_1+a_2\}, \cdots \{a_1+\cdots a_{m-1}+1, \cdots, a_1+\cdots+a_m = n\}$ to themselves. Define a coproduct structure by

$$\mu_S(\rho) = 1 \otimes \rho + \sum_{i=1}^{n-1} \mathrm{Res}^{S_n}_{S_i \times S_{n-i}}(\rho) + \rho \otimes 1 \in \bigoplus_{i=0}^{n} R(S_i) \otimes R(S_{n-i}) \tag{18.15}$$

and a counit (augmentation)

$$\varepsilon_S : RS \longrightarrow \mathbf{Z},\ \varepsilon = \begin{cases} \mathrm{id} & \text{on } R(S_0) = \mathbf{Z} \\ 0 & \text{on } R(S_n),\ n \geq 1 \end{cases} \tag{18.16}$$

The basic fact is now that together with the ring structure already defined and used this makes $R(S)$ a Hopf algebra. The harder part of this is to prove the compatibility of the multiplication with the comultiplication. This is taken care of by the Mackey double coset theorem which describes what happens when a representation is first induced up from a subgroup and then restricted to another subgroup, see [199], Chapt. 4 for details. Actually, it is the projection formula which does the job (together with a description of double cosets of Young subgroups of the type $S_i \times S_{n-i} \subset S_n$. This 'induction restriction projection formula' is the following:[85]

18.17. *Theorem.* Projection formula, Frobenius axiom. Let G be a finite group with subgroup H, and let V be an H-module and W a G-module, then

$$\mathrm{Ind}_H^G \left(V \otimes \mathrm{Res}_H^G(W) \right) = \mathrm{Ind}_H^G(V) \otimes W \tag{18.18}$$

18.19. *Hopf algebras continue their insidious work.* For an account of how to use the Hopf algebraic structure so far described in the representation theory of the symmetric groups, see the two papers of Liulevicius already quoted. Here I will now continue to describe some more Hopf algebraic structure, culminating in the Zelevinsky structure theorem. This involves further fairly heavy machinery, and certainly does not give the easiest way to get at the representation theory of the symmetric groups. But this way is, in my view, quite important.

As to the various bits of structure on $R(S)$ the situation is now as follows:

(i) $R(S)$ is a connected graded $\mathbf{Z}$ module.
(ii) There is a preferred basis consisting of 1 and the irreducible representations with a corresponding inner product for which this basis is orthonormal.
(iii) Multiplication is positive (because taking the outer tensor product of true (real) representations and than inducing up gives a true representation).

[85] Caveat: this is **not** the same instance of a projection formula as occurred i theorem 13.48.

(iv) Comultiplication is positive (because restricting a true representation yields true representations.

(v) Comultiplication is multiplication preserving (or, equivalently, the multiplication is comultiplication preserving). This is an immediate consequence of the Mackey double coset theorem combined with the description of double cosets of Young subgroups.

(vi) Comultiplication and multiplication are dual to each other with respect to the inner product. This follows from Frobenius reciprocity in the form that induction is adjoint to restriction (both on the left and on the right) combined with the fact that the scalar product of two characters on a group counts the number of irreducible representations that they have in common.

(vii) The counit morphism is a morphism of algebras.

(viii) There is just one element of the preferred basis that is primitive. This is the unique irreducible representation of S_1. Indeed for all $n \geq 2$ a real representation of S_n must restrict to some real (as opposed to virtual) representation of $S_1 \times S_{n-1}$.

18.20. *PSH algebras.* The acronym 'PSH' stands for 'positive selfadjoint Hopf.' This implies sort of that there is also a positive definite inner product and that we are working over something like the integers or the reals, where positive makes sense. What is not mentioned is that these Hopf algebras are also supposed to be connected and graded. As will be seen the assumptions 'positive, selfadjoint, graded' are all three very strong; so strong that these algebras can actually be classified. Indeed they are all products of one example[86] and that example is the Hopf algebra of the symmetric functions. The notion is due to Zelevinsky, [425], and the classification theorem is his.

The precise definition is as follows. A PSH algebra is a connected graded Hopf algebra over the integers, so that

$$H = \bigoplus_n H_n, \quad H_0 = \mathbf{Z}, \quad rk(H_n) < \infty \tag{18.21}$$

which is free as an Abelian group and which comes with a given, 'preferred' homogeneous basis $\{\omega_i : i \in I\} = \mathcal{B}$. Define an inner product $\langle\ ,\ \rangle$ on H by declaring this basis to be orthonormal. Then the further requirements are:

$$\text{Selfadjointness: } \langle xy, z\rangle = \langle x \otimes y, \mu(z)\rangle,\ \langle \varepsilon(x), a\rangle = \langle x, e(a)\rangle, \quad x, y, z \in H, a \in \mathbf{Z} \tag{18.22}$$

$$\text{Positivity: let } \omega_i\omega_j = \sum_r a^r_{i,j}\omega_r, \quad \mu(\omega_r) = \sum_{i,j} b^{i,j}_r \omega_i \otimes \omega_j, \quad \text{then } a^r_{i,j}, b^{i,j}_r \geq 0 \tag{18.23}$$

[86] More precisely they are all products of that one example, but the factors are possibly degree shifted.

The Zelevinsky classification theorem now says that a PSH algebra with just one primitive among the preferred basis elements is isomorphic to the Hopf algebra of the symmetric functions (as a Hopf algebra). The proof proceeds by the inductive construction in any PSH algebra with just one primitive preferred basis element p of a series of elements that behave just like the h_n, e_n, p_n of symmetric function theory. The key observation here is that the powers of p must involve all preferred basis elements. In the representations theoretic case $R(S)$ this corresponds to the fact that by definition of the outer multiplication p^n is the regular representation of S_n. Here p is the identity representation of S_1. For the details of the proof see [425], [199] Chapter 4.[87]

18.24. *Bernstein morphism.*[88] The final step of the proof involves a construction that I think of particular interest, that is not yet well understood, and that deserves more study.

Let H be any graded commutative and associative Hopf algebra. Let

$$\mu_n(x) = \sum_i x_{i,1} \otimes x_{i,2} \otimes \cdots \otimes x_{i,n} \tag{18.25}$$

be the n-fold comultiplication written as a sum of tensor products of homogeneous components. Now define

$$\beta_n : H \longrightarrow H[\xi_1, \xi_2, \ldots] = H \otimes \mathbf{Z}[\xi_1, \xi_2, \ldots]$$

by

$$\beta_n(x) = \sum_i x_{i,1} x_{i,2} \cdots x_{i,n} \xi_1^{\deg(x_{i,1})} \xi_2^{\deg(x_{i,2})} \cdots \xi_n^{\deg(x_{i,n})} \tag{18.26}$$

Because H is coassociative and cocommutative this is symmetric in the variables $\xi_1, \cdots, \xi_n$ so that there is an induced algebra morphism[89]

$$\beta_n : H \longrightarrow H \otimes \mathbf{Z}[h_1, \ldots, h_n]$$

and because H is graded this stabilizes in n giving the Bernstein morphism[90]

$$\beta : H \longrightarrow H \otimes \mathbf{Symm} \tag{18.27}$$

In the case that H is **Symm** this turns out to be the product comultiplication morphism!. The final step of the proof of the Zelevinsky classification theorem is now to

---

87 The proof as written down in loc. cit. is just my own attempt to write down Zelevinsky's proof in a way that I could understand it.

88 The construction is due to Joseph N Bernstein.

89 The fact that this is a morphism of algebras uses commutativity; otherwise multiplication is not an algebra morphism.

90 For a general graded commutative Hopf algebra the Bernstein morphism defines a coaction of **Symm** on H and then by duality also an action of **Symm** on the graded dual of H.

compose the Bernstein morphism with a morphism $H \longrightarrow \mathbf{Z}$ that takes the inductively constructed analogues of the h_n all to one. In the case $H = \mathbf{Symm}$ this is the counit morphism of the product comultiplication morphism μ_P.[91]

As promised in 11.39 above, there now follow a few remarks on (0, 1)-matrices and their links with Witt vectors and representation theory. This requires some preparation.

18.28. *Majorization ordering.* Let $\alpha = (a_1, a_2, \cdots, a_n)$ and $\beta = (b_1, b_2, \cdots, b_n)$ be two vectors of non-negative real numbers of the same l_1-norm, i.e. $a_1 + a_2 + \cdots + a_n = b_1 + b_2 + \cdots + b_n$. Denote by $\overline{\alpha} = (\overline{a}_1, \overline{a}_2, \cdots, \overline{a}_n)$ a reordering (rearrangement, permutation) of α such that $\overline{a}_1 \geq \overline{a}_2 \geq \cdots \geq a_n$.

The majorization ordering is now defined by

$$\alpha \geq_{maj} \beta \Leftrightarrow \sum_{i=1}^{r} \overline{a}_i \geq \sum_{i=1}^{r} \overline{b}_i, \quad r = 1, 2, \cdots, n \tag{18.29}$$

This ordering occurs is many parts of mathematics under many different names: majority ordering, dominance ordering, natural ordering, specialization ordering, Snapper ordering, Ehresmann ordering, mixing ordering. Parts of mathematics where it plays an important role include: families of algebraic geometric vectorbundles, families of representations, families of nilpotent matrices, Grassmann manifolds, control theory, representation theory, thermodynamics, convex function theory (Schur convex functions[92]), doubly stochastic matrices, (0, 1) matrices, inequality theory (Muirhead inequalities, a far reaching generalization of the geometric mean-arithmetic mean inequality), representation theory, ... See [3]. Many of these uses of the majorization ordering are strongly related, see [202].

18.30. *Conjugate partition.* Let $\alpha = [a_1, a_2, \cdots, a_m]$ be a partition of n. Then the conjugate partition $\alpha^{conj} = \left[a'_1, a'_2, \cdots, a'_{m'}\right]$ is defined by

$$a'_i = \#\{j : a_j \geq i\} \tag{18.31}$$

So, for instance, $[4, 4, 3, 2, 1, 1, 1]^{conj} = [7, 4, 3, 2]$. (If the partition is displayed as a diagram (either in the French or Anglo-Saxon manner), the conjugate partition looks at columns instead of rows.)

There are (at least) three[93] applications of the majorization ordering in the representation theory of the symmetric groups. In addition there is the Gale-Ryser theorem which is of immediate relvance here as it deals with the existence of (0, 1)-matrices

[91] Note that without the characteristic map isomorphism or the Zelevinsky theorem itself, it is not yet clear that **Symm** is PSH (but $R(S)$ is). The trouble is positivity; specifically the fact that the product of two Schur symmetric functions is a nonnegative linear combination of Schur functions. This seems a fact that is not so easy to establish directly (without going to representation theory) (and, hence, is a bit of a blemish on symmetric function theory).

[92] Same Schur; totally different topic.

[93] Another (related one), due to Kraft and de Concini, is too far from the present topic to discuss.

and thus has things to say about the second comultiplication on **Symm** (and hence the multiplication of Witt vectors.

18.32. *Gale-Ryser theorem.* Let α and β be two partitions then there is a $(0,1)-$matrix with row sum α and column sum β if and only if $\alpha^{conj} \geq_{maj} \beta$.

Of course a similar theorem holds for compositions instead of partitions; this amounts to taking permutations of columns and permutations of rows.

For example take the $(0,1)-$matrix

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

This one has as row sum the composition $[2,3,2]$ with associated partition $[3,2,2]$; and the column sum is the composition $[1,2,1,1,2]$ with associated partition $[2,2,1,1,1]$. Also $[3,2,2]^{conj} = [3,3,1]$. And, indeed $[3,3,1] \geq_{maj} [2,2,1,1,1]$.

Of the three theorems in the representation theory of the symmetric groups that involve the majorization ordering the one closest to the present concerns (Witt vectors) is the Snapper Liebler-Vitale Lam theorem.

18.33. *Snapper Liebler-Vitale Lam theorem.* Let α and β be partitions and let S_α be the Young subgroup defined by α. Then ρ_β, the irreducible representation corresponding to β, occurs in $Ind_{S_\alpha}^{S_n}(I)$ if and only if $\alpha \leq_{maj} \beta$.

Here I stands for the trivial representation of S_α. For a proof see [203] and/or the references therein.

For completeness sake and possible future applications, below there is the Ruch-Schönhofer theorem. Under the isomorphism between **Symm** and $R(S)$ the trivial representation of S_n corresponds to the complete symmetric function h_n and the sign representation A_{S_n} corresponds to the elementary symmetric function e_n. So this theorem could certainly be relevant.

18.34. *Ruch-Schönhofer theorem.* The representations $\mathrm{Ind}_{S_\alpha}^{S_n}\left(I_{S_\alpha}\right)$ and $\mathrm{Ind}_{S_\beta}^{S_n}\left(A_{S_\beta}\right)$ have an irreducible representation in common (which is the same as saying that their inner product is nonzero), if and only if $\alpha \leq_{maj} \beta$.

18.35. *Outer plethysm and inner plethysm.* In section 16.67 the composition operator 'outer plethysm' on **Symm** made its appearance. But **Symm** is isomorphic to $R(S)$ as Hopf algebras (with extra product multiplication and extra coproduct multiplication with counit). Thus there is an outer plethysm operation on $R(S)$. The problem is to describe this operation in representation theoretic terms. This has been called the 'outer plethysm problem.'

The other way: each of the summands of $R(S) = \bigoplus_{n=0}^{\infty} R(S_n)$ is a $\lambda-$ring in its own right, giving rise to plethym operations called inner plethysm, and the 'inner plethysm problem' is to describe these in symmetric function terms when transferred to **Symm**.

18.36. *Outer plethysm problem.* Outer plethysm gives in particular a composition of I_{S_n}, the trivial representation of S_n with any $\sigma \in R(S_k)$. The result is denoted $h_n(\alpha)$ in [239], p. 135. The trivial (or identity) representation of S_n corresponds to $h_n \in \mathbf{Symm}$, hence the notation.

In loc. cit., p. 135 Donald Ivar Knutson guesses[94] the following representation theoretic description of $h_n(\alpha)$. Given a representation α of S_k on V first construct the induced representation of the wreath product $S_n[S_k]$ on $V^{\otimes n}$, then induce up the representation obtained to S_{nk} using the natural inclusion $S_n[S_k] \subset S_{nk}$.

This 'guess' has since turned out to be correct, see [221, 313, 394].

18.37. *Inner plethysm problem.* The $\lambda-$ring structure on $R(S_n)$ gives of course rise to Frobenius operators which via the isomorphism give rise to (degree preserving) Adams-Frobenius type operators on the homogeneous component of weight (degree) n of **Symm**. Using the inner product there are corresponding adjoint operators. These are described nicely in section 3 of [349].

Besides the five references just quoted here is a selection of further references on plethysm: [8, 43, 42, 59, 66, 82, 267, 265, 266, 282, 306, 312, 313, 341, 351, 385, 394, 395, 429].

19. Burnside rings

Let G be a finite group. The Burnside ring $B(G)$ is one of the fundamental (representation like) rings attached to G. It is the Grothendieck ring of finite G-sets with sum and product inducted by disjoint union and Cartesian product respectively. Or, equivalently, the ring of permutation representations.

So as a group $B(G)$ is a finitely generated Abelian group with as (canonical) basis the transitive G-sets (orbits). According to some the Burnside ring was introduced by Andreas Dress in [117], where he proved a somewhat unexpected kind of result, viz. that a finite group is solvable if and only if the spectrum of its Burnside ring is Zariski connected, i.e. if and only if 0 and 1 are the only idempotents in $B(G)$.

Others attribute the introduction of the concept to Louis Solomon [372].

By now there is a substantial literature on Burnside rings: the ZMath database lists at the moment[95] 305 publications with "Burnside ring" in title or abstract. For a fair selection see the bibliography of [61].

"[The Burnside ring] is in many ways the universal object to consider when looking at the category of G-sets. It can be viewed an analogue of the ring $\mathbf{Z}$ of integers for this category."

"The ring $B(G)$ is also functorial with respect to G and subgroups of G and this leads to the Mackey functor or Green functor point of view. . . . The Burnside Mackey functor is a typical example of projective Mackey functor. It is also a universal object in the category of Green functors."

[94] His word.

[95] 19 April 2008.

These are two quotes from the introduction of [61]. So if the Burnside ring of an arbitrary finite group is already as nice as all that, what about the Burnside ring of a really nice group like the integers. That one must then be super-nice. And it is. It turns out to be the ring of Witt vectors of the integers, $W(\mathbf{Z})$, a discovery due to Andreas Dress and Christian Siebeneicher, [115, 120].[96]

19.1. *G-sets.* Let G be a group. A G-set is a set X together with an action (on the left) of G on it; that is a mapping $G \times X \longrightarrow X$, $(g, x) \mapsto gx$ such that $g(h(x)) = (gh)(x)$, $g, h \in G$, $x \in X$ and $e(x) = x$ where e is the identity element of G. A morphism of G-sets is a mapping $f : X \longrightarrow Y$ such that $f(gx) = g(f(x))$. This defines the category G-**Set** of G-sets.

For $x \in X$ the subgroup

$$G_x = \{g \in G : gx = x\} \tag{19.2}$$

is the stabilizer of $x \in X$ and

$$Gx = \{gx : g \in G\} \tag{19.3}$$

is the orbit of x (or through x). If H is a subgroup multiplication on the left induces a G-set structure on the set of left cosets G/H which has a single orbit (a transitive G-set). The orbit Gx of an element is isomorphic as a G-set to G/G_x.

A G-set X gives rise to a permutation representation $\rho(X)$ over any ring of which the underlying module is the free module with basis X and with the action of G given by the the permutations of basis elements determined by the G-set structure of X.

19.4. *Induction and restriction for G-sets.* Let H be a subgroup of a group G. Then if X is a G-set retricting the action to H gives an H-set, defining a functor Res^G_H: $G - \mathbf{Set} \longrightarrow H - \mathbf{Set}$.

As is to be expected there is also a functor the other way called induction. Let Y be an H-set. Consider the set $G \times Y$ with the equivalence relation determined by

$$(g, y) \sim (gh^{-1}, hy),\ g \in G, h \in H, y \in Y \tag{19.5}$$

The set of equivalence classes is suggestively denoted $G \times_H Y$. (Note that $(g, y) \mapsto \left(gh^{-1}, hy\right)$ defines an action of H on $G \times Y$ and that the equivalence classes are the orbits for this action.) Multiplication on the left $\left(g', (g, y)\right) \mapsto (g'g, y)$ induces an action of G on $G \times_H Y$ giving a G-set denoted $\mathrm{Ind}^G_H(Y)$. It is obvious how to make this into a functor Ind^G_H: $H - \mathbf{Set} \longrightarrow G - \mathbf{Set}$. Induction is left adjoint to restriction (but not right adjoint in general). There is also a product formula (projection formula, Frobenius identity)

$$X \times \mathrm{Ind}^G_H(Y) \cong \mathrm{Ind}^G_H\left(\mathrm{Res}^G_H(X) \times Y\right) \tag{19.6}$$

For more details on this and some more related material see e.g. [409] p. 811, [61] pp. 744–745.

[96] This also of course, given the 'universality remark' just quoted, provides a seventh universality property of the Witt vectros. Working out what this one really means is still an open matter.

Restriction and induction are compatible with the notions of the same name for representations; i.e. they are compatible with the mapping ρ that assigns to a G-set its associated permutation representation.

19.7. *Almost finite G-sets.* Given a group G its profinite completion $\hat{G}$ is its completion for the topology of normal subgroups of finite index. Or, equivalently it is the projective limit

$$\hat{G} = \varprojlim G/N, \quad N \text{ runs over the normal subgroups of finite index} \tag{19.8}$$

A G-set X for a group G is almost finite if for each $x \in X$ the stabilizer subgroup G_x is a subgroup of finite index (so that all orbits are finite) and such that the subsets of invariants

$$X^U = \{x \in X : G_x \subseteq U\} \tag{19.9}$$

are finite for every subgroup of finite index. (So that there are only finitely many orbits of isomorphism type G/U for each such U.)

It does not matter whether one works in this with the group G or its profinite completion. In the case of a profinite group one works with G-spaces. These are Hausdorff spaces with a continuous action[97] of G. See [115], p. 5. Most papers that deal with the present subject (Burnside rings and Witt-Burnside rings) take the profinite point of view.

For a subgroup of finite index $H \subset G$ restriction takes almost finite G-sets into almost finite H-sets and induction takes almost finite H-sets into almost finite G-sets.

Finite disjoint unions and finite Cartesian products of almost finite G-sets are almost finite.

19.10. *Burnside theorem.* [78] Chapter XII theorem 1. Let G be a finite group and X and Y finite G-sets. Then the following are equivalent

(i) The G-sets X and Y are isomorphic

(ii) For any subgroup H of G the sets of invariants X^H and Y^H have the same cardinality.

This still holds for almost finite G-sets where one only need consider subgroups of finite index.

19.11. *Burnside ring.* The (completed) Burnside ring $\hat{B}(G)$ of a group is now defined as the Grothendieck ring of the category of almost finite G-sets.

It can also be defined as the projective limit of the usual Burnside rings (of finite groups) $B(G/N)$ where N ranges over all normal subgroups (resp. closed normal subgroups) of finite index.[98]

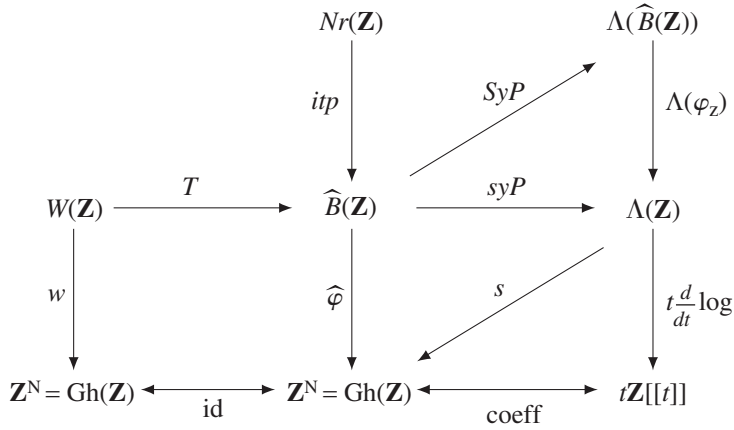
[97] Here G is given the topology defined by the normal subgroups of finite index (which define the profinite structure) and the set X is given the discrete topology.

[98] This explains the notation $\hat{B}$ and the terminology 'completed.'

19.12. *Almost finite cyclic sets.* In the case of the group of integers all this specializes as follows. A cyclic set is simply a set with a left action of the group of integers on it. That is it is a set with a specified bijection. A cyclic set is almost finite if every orbit is of finite length and if for every $n \in \mathbf{N}$ there are only finitely many orbits of length n. Obviously (finite) disjoint unions and (finite) Cartesian products of almost finite cyclic sets are again almost finite and so there is the Grothendieck ring $\hat{B}(\mathbf{Z})$ of almost finite cyclic sets.

If $B(\mathbf{Z})$ denotes the Grothendieck ring of finite cyclic sets, $\hat{B}(Z)$ is the completion of $B(\mathbf{Z})$ under a suitable natural topology on $B(\mathbf{Z})$, see [120], p. 3 and below.

19.13. *A remarkable commutative diagram. The $W(\mathbf{Z})$, $\hat{B}(\mathbf{Z})$, $\Lambda(\mathbf{Z})$, $Nr(\mathbf{Z})$ isomorphisms.* The ring $\hat{B}(\mathbf{Z})$ fits in the following commutative diagram.



(19.14)

All the horizontal arrows in this diagram are isomorphisms (of rings with operators) and so is the morphism denoted '*itp*' (which stands for 'interpretation'). The morphisms w, $\hat{\varphi}$, s, $t\frac{d}{dt}\log$ are ghost component morphisms and injective. Finally, $\varphi_{\mathbf{Z}}$ and $\Lambda(\varphi_{\mathbf{Z}})$ are surjections and SyP is an injection. Those morphisms which have not already occurred in earlier sections, such as w, s and the logarithmic derivative, will be elucidated below of course.

One of the more remarkable aspects of this diagram is the fact that the composite $syP \circ T$ precisely embodies the very nasty coordinate change from Witt vector coordinates to power series coordinates encoded by the power series identity

$$\prod_n \frac{1}{(1 - x_n t^{n)}} = 1 + a_1 t + a_2 t^2 + \cdots \tag{19.15}$$

I consider this a main contribution from [120]. (The necklace coordinates also fit in as suggested by the diagram.)

19.16. *The ghost component morphism* $\hat{\varphi}: \hat{B}(\mathbf{Z}) \longrightarrow \mathrm{Gh}(\mathbf{Z}) = \mathbf{Z}^{\mathbf{N}}$. Given an almost finite cyclic set X and an element $n \in \mathbf{N}$, consider the subgroup of finite index $n\mathbf{Z} \subseteq \mathbf{Z}$ and define

$$\varphi_{n\mathbf{Z}}(X) = \#\{x \in X : x \text{ is invariant under } n\,\mathbf{Z}\} \tag{19.17}$$

This induces ring morphisms $\hat{B}(\mathbf{Z}) \longrightarrow \mathbf{Z}$ and these combine to define the ghost component ring morphism $\hat{\varphi}: \hat{B} \longrightarrow \mathrm{Gh}(\mathbf{Z})$. The morphism $\hat{\varphi}$ is injective because of the (extension to almost finite sets of the) Burnside theorem 19.10.

Take the restrictions of the $\varphi_{n\mathbf{Z}}$ to $B(\mathbf{Z})$ and give $B(\mathbf{Z})$ the coarsest topology for which all these restrictions are continuous (with $\mathbf{Z}$ discrete). Then $\hat{B}(\mathbf{Z})$ is the completion of $B(\mathbf{Z})$ for this topology.

The morphism $\hat{\varphi}$ is injective; it is not surjective; an element $y \in \mathrm{Gh}(\mathbf{Z})$ is in its image if and only if

$$\sum_{d|n} \mu(d^{-1}n)\, y(d) \equiv 0 \mod n \tag{19.18}$$

for all $n \in \mathbf{N}$. The same condition turned up in connection with the necklace ring in section 17 above ([113, 118, 126], [391] pp. 11–12).

19.19. *The isomorphism itp*: $Nr(\mathbf{Z}) \longrightarrow \hat{B}(\mathbf{Z})$. The only transitive almost finite cyclic sets are the coset spaces $C_n = \mathbf{Z}/n\mathbf{Z}$. So every element of $\hat{B}(\mathbf{Z})$ is a (possibly infinite) sum

$$\sum_n b_n C_n,\ b_n \in \mathbf{Z} \tag{19.20}$$

As a group $Nr(\mathbf{Z}) = \mathbf{Z}^{\mathbf{N}}$ with coordinate wise addition. So assigning to an element $\beta = (b_1,\ b_2,\ b_3, \ldots) \in \mathbf{Z}^{\mathbf{N}}$ the formal difference of almost finite cyclic sets

$$\sum_{b_n>0} b_n C_n \ - \sum_{b_n<0} (-b_n) C_n$$

defines an isomorphism '*itp*' of Abelian groups $Nr(\mathbf{Z}) \longrightarrow \hat{B}(\mathbf{Z})$. It remains to see how *itp* behaves with respect to multiplication.

Now observe[99] that the product of two transitive cyclic sets C_r and C_s decomposes as the sum of (r, s) copies of $C_{[r,s]}$. Here (r, s) is the greatest common divisor of r and s and $[r, s]$ is their least common multiple. Given the definition of the multiplication on the necklace ring, see 17.15, it follows that *itp* is an isomorphism of rings.

19.21. *The isomorphism* $T: W(\mathbf{Z}) \longrightarrow \hat{B}(\mathbf{Z})$. This isomorphism of rings is denoted τ in [120] and called "Teichmüller" there. This may be (indeed, is) appropriate in some sense but is also potentially confusing in view of the Teichmüller representatives mapping $\tau: A \longrightarrow W(A)$ in Witt vector theory.

[99] It appears, see [120] p. 5, that it was this observation that lead to the investigations of Andreas Dress and Christian Siebeneicher that culminated in [115, 120, 126].

The definition of the isomorphism T involves induction. Let X be a $\mathbf{Z}$-set. Via the isomorphism $\mathbf{Z} \longrightarrow n\mathbf{Z} \subset \mathbf{Z}, r \mapsto nr$ it can be seen as an $n\mathbf{Z}$-set. Now induce[100] this one up to $\mathbf{Z}$. The result is denoted $\mathrm{ind}_n(X)$. In concreto this works out as follows. Consider the product $\mathbf{Z} \times X$ and the equivalence relation on it defined by $(z, x) \sim (z - nu, ux), u \in \mathbf{Z}$. Then $\mathrm{ind}_n(X)$ is the set of equivalence classes of this equivalence relation with the action induced by the left action of $\mathbf{Z}$ on itself, $(z', (z, x)) \mapsto (z' + z, x)$.

It readily follows that

$$\mathrm{ind}_n(C_r) = C_{rn}$$

and hence that

$$\mathrm{ind}_n\left(\sum_r b_r C_r\right) = \sum_r b'_r C_r \tag{19.22}$$

where

$$b'_r = \begin{cases} b_{r/n} & \text{if } n \text{ divides } r \\ 0 & \text{otherwise} \end{cases} \tag{19.23}$$

and hence that

$$\varphi_{r\mathbf{Z}}(\mathrm{ind}_n(X)) = \begin{cases} n\varphi_{r/n}(X) & \text{if } n \text{ divides } r \\ 0 & \text{otherwise} \end{cases} \tag{19.24}$$

So, at the ghost component level ind_n behaves just like Verschiebung in the case of the big Witt vectors.

Now let $q \in \mathbf{N} \cup \{0\}$. With $q^{(\mathbf{Z})}$ denote the set of maps $\mathbf{Z} \longrightarrow \{1, 2, \cdots, q\}$ that factor through some set of cosets $\mathbf{Z} / n\mathbf{Z}$. That is the continous maps $\mathbf{Z} \longrightarrow \{1, 2, \cdots, q\}$ where $\{1, 2, \cdots, q\}$ has the discrete topology and $\mathbf{Z}$ the topology of subgroups of finite index. The action of $\mathbf{Z}$ is given by

$$(zf)(z') = f(z' - z) \tag{19.25}$$

Note that $0^{(\mathbf{Z})} = \emptyset$. It is fairly immediate that

$$(qq')^{(\mathbf{Z})} = q^{(\mathbf{Z})} \times q'^{(\mathbf{Z})}, \ \varphi_{n\mathbf{Z}}(q^{(\mathbf{Z})}) = q^n \tag{19.26}$$

$$\hat{\varphi}(q^{(\mathbf{Z})}) = (1, q, q^2, q^3, \ldots) = \mathrm{coeff}\,(1 - q)^{-1} \tag{19.27}$$

Finally define

$$T : W(\mathbf{Z}) \longrightarrow \hat{B}(\mathbf{Z}), (x_1, x_2, x_3, \cdots) \mapsto \sum_{n=1}^{\infty} \mathrm{ind}_n\left(x_n^{(\mathbf{Z})}\right) \tag{19.28}$$

[100] See 19.4 above.

Then it follows from what has been noted just above that

$$\varphi_n \mathbf{Z}(x_1, x_2, x_3, \cdots) = \sum_{d|n} d x_d^{n/d} \tag{19.29}$$

and that

$$\prod_{n=1}^{\infty} (1 - x_n t^n)^{-1} = \prod_{n=1}^{\infty} (1 - t^n)^{-b_n} \Leftrightarrow T(x_1, x_2, x_3, \cdots) = \sum_{i=1}^{\infty} b_i C_i \tag{19.30}$$

Formula (19.29) shows that the lower left square in the diagram (19.14) is commutative. Using the characterizations of the image of the ghost components morphisms w and $\hat{\varphi}$, see (9.95) and (19.20) above, and their injectivity, it follows that T is an isomorphism.

Further (19.30) says that the composite morphism $(itp)^{-1} \circ T$ exactly embodies the coordinate transformation between Witt vector coordinates and necklace coordinates encoded by the power series identity

$$\prod_{n=1}^{\infty} (1 - x_n t^n)^{-1} = \prod_{n=1}^{\infty} (1 - t^n)^{-b_n}$$

19.31. *Symmetric powers of* G-sets. Given a G-set X its n-th symmetric power is obtained by first taking the n-fold Cartesian product X^n with diagonal action

$$g(x_1, x_2, \cdots, x_n) = (gx_1, gx_2, \cdots, gx_n) \tag{19.32}$$

The symmetric group S_n acts on this by permuting coordinates

$$\sigma(x_1, x_2, \cdots, x_n) = (x_{\sigma^{-1}(1)}, x_{\sigma^{-1}(2)}, \cdots, x_{\sigma^{-1}(n)})$$

and this action commutes with the diagonal action of G. So there is an induced action of G on the set of orbits $S^n X$. This is the n-th symmetric power of the G-set X. Alternatively, and better, an element from $S^n X$ can be described as a multiset of size n, i.e. a function $f : X \longrightarrow \mathbf{N} \cup \{0\}$ such that $\sum_{x \in X} f(x) = n$. In this picture the action is $(gf)(x) = f(g^{-1}x)$. The symmetric powers of almost finite G-sets are almost finite.

It follows rather immediately from the construction that

$$S^n (X \amalg Y) = \coprod_{i+j=n} S^i X \times S^j Y \tag{19.33}$$

and it follows that the map

$$X \mapsto \sum_{n=1}^{\infty} (S^n X) t^n \tag{19.34}$$

induces a morphism of Abelian groups

$$SyP\colon \hat{B}(\mathbf{Z}) \longrightarrow \Lambda(\hat{B}(\mathbf{Z})) \tag{19.35}$$

Now compose this with $\Lambda(\varphi_{\mathbf{Z}}) : \Lambda(\hat{B}(\mathbf{Z})) \longrightarrow \Lambda(\mathbf{Z})$ to obtain a morphism of Abelian groups

$$syP : \hat{B}(\mathbf{Z}) \longrightarrow \Lambda(\mathbf{Z}) \tag{19.36}$$

(and the commutative upper triangle of diagram (19.14)). Thus the coefficient of t^n in $syP(X)$ is the number of invariant elements in the n-th symmetric power $S^n X$.

It now turns out that in fact syP is an isomorphism of rings and even an isomorphism of $\lambda-$rings. The key to that is the easy observation that

$$\varphi_{\mathbf{Z}}(S^n C_r) = \begin{cases} 1 & \text{if } r \text{ divides } n \\ 0 & \text{otherwise} \end{cases} \tag{19.37}$$

(Indeed, a function $f : \{0, 1, \cdots, r-1\} \longrightarrow \mathbf{N} \cup \{0\}$ is invariant under 'left shift modulo r of its argument' if and only if all its values are equal.)

Formula (19.37) serves to prove that the lower right square of diagram (19.14) is commutative (as well as the two triangles there, that syP is an isomorphism of rings, and that the composed morphism $syP \circ itp$ embodies the coordinate change encoded by the power series identity

$$\prod_{n=1}^{\infty}(1 - t^n)^{-b_n} = 1 + a_1 t + a_2 t^2 + a_3 t^3 + \tag{19.38}$$

When combined with (19.30) this shows that the composite morphism $syP \circ T : W(\mathbf{Z}) \longrightarrow \Lambda(\mathbf{Z})$ precisely gives the nasty coordinate change formulas between Witt vector coordinates and power series coordinates encoded by the power series identity

$$\prod_{n=1}^{\infty}(1 - x_n t^n)^{-1} = 1 + a_1 t + a_2 t^2 + a_3 t^3 + \cdots$$

which I feel is a major insight from [120].

19.39. *Frobenius and Verschiebung on* $W(\mathbf{Z})$, $\hat{B}(\mathbf{Z})$, $\Lambda(\mathbf{Z})$. On all the rings in diagram (19.14) there are Frobenius and Verschiebung operators. All have been defined before except the ones on $\hat{B}(\mathbf{Z})$. On this Burnside ring 'Verschiebung' is 'ind' and Frobenius is restriction (followed by the identification coming from the obvious (canonical) isomorphism of Abelian groups $\mathbf{Z} \longrightarrow n\mathbf{Z}$.

It now turns out (and is verified with no great difficulty) that all morphism indicated in the diagram are compatible with the Frobenius operators and all except SyP are compatible with Verschiebung. (SyP is most definitely not compatible with Verschiebung.)

As a matter of fact, identifying $\hat{B}(\mathbf{Z})$ with $\Lambda(\mathbf{Z})$ via syP turns SyP into the morphism of λ-rings $\sigma_t : \Lambda(\mathbf{Z}) \longrightarrow \Lambda(\Lambda(\mathbf{Z}))$ of λ-ring theory,[101] see section 16 above.

[101] This is one more indication that it is better to work with symmetric powers than with exterior powers.

19.40. This concludes the treatment here of the Burnside ring of the integers (or its profinite completion $\hat{\mathbf{Z}}$). I have pretty much followed [120] apart from some interspersed remarks. I hope and believe that the outline above is sufficient that the reader can fill in all details. But if needed they can be found in loc. cit.

However, there is more to the Burnside ring story in connection with Witt vectors. For every profinite group there is a Witt vector like functor **CRing** $\longrightarrow$ **CRing** called the Witt Burnside functor. Below there is the main theorem from [126] about them. First some definitions and notation.

19.41. Let G be a profinite group. That is a projective limit $G = \varprojlim G/N$ of finite quotient group. Give G the topology defined by the collection of normal subgroups N of finite index. A G-space is a Hausdorff spcae on which G acts continuously from the left. Such a G-space is almost finite if it is discrete and if for any open subgroup $U \subset G$ the number of invariants under U is finite. Set

$$\varphi^U(X) = \#\{x \in X\colon ux = x \text{ for all } u \in U\} \tag{19.42}$$

For any open subgroup U there is the transitive almost finite (in fact finite) G-space of left cosets G/U. Denote with $osg(G)$ the set of all open subgroups of G and with $cosg(G)$ the quotient of conjugacy classes of open subgroups.

Finally let $\hat{B}(G)$ be the "completed Burnside ring" of G, that is the Grothendieck ring of (isomorphism classes) of almost finite G-spaces with addition induced by disjoint union and multiplication induced by the Cartesian product of G-spaces.

19.43. *Existence theorem of the Witt-Burnside functors* [126]. Let G be a profinite group. There exists a unique covariant functor W_G from the category of unital commutative rings to itself such that as a set $W_G(A)$ is the set $A^{cosg(G)}$ of all functions from $cosg(G)$ to A. That is all functions from $osg(G)$ to A that are constant on conjugacy classes of subgroups, and with $W_G(h)\colon W_G(A) \longrightarrow W_G(B)$ given by composition $\alpha \mapsto \alpha \circ h$, $\alpha \in W_G(A)$, such that for all open subgroups the map

$$\psi_U^A : W_G(A) \longrightarrow A \tag{19.44}$$

defined by

$$\sum_{U \subset_{scjg} V \subset G} \varphi^U(G/V)\alpha(V)^{(V:U)} \tag{19.45}$$

is a natural transformation of functors from W_G to the identity. Here the sum (9.45) is over all V such that U is subconjugate to U (denoted $U \subset_{scjg} V$), which means that there is a $g \in G$ such that U is a subgroup of gVg^{-1}, the 'index' $(V{:}U) = (G{:}U)/(G{:}V)$, and in the sum (9.45) exactly one summand is taken for each conjugacy class of subgroups V with $U \subset_{scjg} V$.

Moreover, $W_G(\mathbf{Z}) = \hat{B}(G)$ and $W_{\mathbf{z}} = W$ the functor of the big Witt vectors.

There are also Frobenius and Verschiebung like functorial endomorphisms coming respectively from restriction and induction. The Frobenius morphism are ring

morphisms, the Verschiebung morphism are morphisms of Abelian groups. These morphisms have the usual kinds of properties.

Since the appearance of the foundational papers [126, 120] a number of other papers have appeared on the topic of Witt-Burnside functors, giving refinements, further developments, applications and interrelations, simplifications and complications; see [73, 116, 122, 135, 156, 175, 318, 319, 315].

19.46. *β-rings.* It seems clear from [370] that there is no good way to define a λ-ring structure on Burnside rings, see also [158]. There are (at least) two different choices giving pre-λ-rings but neither is guaranteed to yield a λ-ring. Of the two the symmetric power construction seems to work best.

Instead one needs what are called β-operations[102] (power operations), first introduced in [56], and β-rings. There is a β-operation for each conjugacy class of subgroups of the symmetric groups S_n and they are constructed by means of symmetric powers of G-sets. Every λ-ring is a β-ring (but not vice versa).

There is some vagueness about what is precisely the right definition of a β-ring, see, e.g. the second paragraph on page 2 of [183] and the second comment at the end of §3 of that preprint.

However, it seems clear that the free β-ring on one generator must be the ring

$$B(S) = \bigoplus_{n=0}^{\infty} B(S_n) \tag{19.47}$$

This is the direct sum of the Burnside rings of the symmetric groups (with $B(S_0) = \mathbf{Z}$ by decree) equipped with an outer product defined completely analogously as in the case of $R(S)$, see section 18 above. I.e.

$$XY = \mathrm{Ind}_{S_i \times S_j}^{S_{i+j}}(X \times Y) \tag{19.48}$$

There is also a coproduct making it a Hopf algebra and the underlying ring is again a ring of polynomials in countably many indeterminates over the integers.

The natural morphism

$$\rho : B(S) \longrightarrow R(S) \tag{19.49}$$

that assigns to an S_n-set the corresponding permutation representation is a surjective (but not injective) morphism of Hopf algebras and also a morphism of λ-rings, see [183] p. 11.

A selection of papers on β-rings is [56, 158, 183, 300, 301, 344, 397, 396, 433].

Coda.

Some people write about the Witt polynomials as mysterious polynomials that come out of nowhere.[103] To me they are so elegant and natural that they simply cry out for

[102] I have tried to track down where the appellation 'beta' comes from unsuccessfully. But it seems likely that it is some sort of philosophical mix between the 'B' from Burnside and the 'λ' from λ-ring.

[103] Also, see section 3 above, they definitely do not come out nowhere.

deep study. This is my main reason for presenting things here as I have done above, even though, as has been shown, they are not really needed.

Should a reader inadvertently get really interested in the Witt vector ring functors he/she is recommended to work through the 57 exercises on the subject in [63], pp. 26 worth, just for the formulations only, mostly contributed, I have been told, by Pierre Cartier. He/she can then continue with the exercises on Cohen rings in same volume which also involve a fair amount of Witt vector stuff.

Appendix. The algebra of symmetric functions in infinitely many indeterminates

In several places in the sections above there occur expressions like

$$\prod_{i=1}^{\infty}\left(1+\xi_i t+\xi_i^2 t^2+\xi_i^3 t^3+\cdots\right) \tag{A.1}$$

and statements that the coefficients of each power in t in (A.1) are symmetric functions in the infinity of commuting variables $\xi_1, \xi_2, \xi_3, \cdots$ and that they are, hence, polynomials in the elementary (or complete) symmetric functions in the infinity of commuting indeterminates $\xi_1, \xi_2, \xi_3, \ldots$. The few pages in this appendix are meant for those who feel (rightly) a bit nervous about statements like this even though the meaning seems intuitively clear.[104]

A.2. *Power series in infinitely many variables.* Let I be any set and let $\{\xi_i : i \in I\}$ be a corresponding set of commuting indeterminates. An exponent sequence for I is a map $e : I \longrightarrow \mathbf{N} \cup \{0\}$ of finite support. I.e. there are only finitely may $i \in I$ for which $e(i) \neq 0$. Let $E(I)$ be the set of all exponent sequences. For each $e \in E(I)$ introduce a symbol ξ^e. Giving I a total order, one can think of ξ^e as

$$\xi^e = \prod_{e(i)\neq 0} \xi_i^{e(i)} \tag{A.3}$$

where the product (monomial) on the right hand side is written down in the order specified by an ordering of I.[105] Two monomials, i.e. symbols, ξ^e are multiplied by the rule $\xi^e\xi^{e'} = \xi^{e+e'}$ where $e+e'$ is the point-wise sum $(e+e')(i) = e(i)+e'(i)$. The ring of formal power series over a base ring k is now defined as

$$k\langle\langle \xi_i : i \in I\rangle\rangle = \left\{ \sum_{e\in E(I)} a_e \xi^e \colon a_e \in k \right\} \tag{A.4}$$

[104] To my mind the matters touched upon here illustrate well issues in foundational mathematics (intuitionism) having to do with realized infinities vs potential infinites.

[105] But this is not needed.

Two such expressions as on the right of (A.4) are multiplied by the rule

$$\left(\sum_{e\in E(I)} a_e\xi^e\right)\left(\sum_{e\in E(I)} b_e\xi^e\right)=\left(\sum_{e\in E(I)} c_e\xi^e\right), c_e=\sum_{e'+e''=e} a_{e'}b_{e''} \qquad \text{(A.5)}^{106}$$

This product rule is well defined because of the finite support condition on exponent sequences.

A product like (A.1) is now by definition interpreted as the formal power series one obtains by multiplying any finite number of factors from it. That is

$$\begin{aligned}&\prod_{i=1}^{\infty}\left(1+\sum_{\mathrm{wt}(e_i)\geq 1} a_{i,e_i}\xi^{e_i}t^{\mathrm{wt}(e_i)}\right)=1+\sum_{\mathrm{wt}(e)\geq 1} c_e\xi^e t^{\mathrm{wt}(e)},\\ &\mathrm{wt}(e)=\sum_i e(i),\quad c_e=\sum_{\substack{e_1+\cdots+e_r=e\\ \mathrm{wt}(e_i)\geq 1}} a_{i_1,e_1}\cdots a_{i_r,e_r}\end{aligned} \qquad \text{(A.6)}$$

and similar products with the 'counting variable' t left out.

A.7. *Symmetric and quasisymmetric power series.* Probably every one (who is likely to get his fingers on this chapter) knows, or can easily guess at, what is a symmetric polynomial (over the integers, or any other base rink k).

Say, the polynomial is in n (commuting) variables $\xi_1, \xi_2, \cdots, \xi_n$. Then the polynomial $f=f(\xi_1, \xi_2, \cdots, \xi_n)$ is symmetric if for every permutation σ of $\{1, 2, \ldots, n\}$

$$f(\xi_{\sigma(1)}, \xi_{\sigma(2)}, \cdots, \xi_{\sigma(n)})=f(\xi_1, \xi_2, \cdots, \xi_n) \qquad \text{(A.8)}$$

And, in that case, the main theorem of symmetric functions says that: the polynomial f is a polynomial p_f in the complete symmetric functions $h_1, h_2, \cdots, h_n$ or, equivalently, in the elementary symmetric functions, and "the polynomial p_f is independent of the number of variables involved provided there are enough of them" (meaning more than or equal to the degree of f). The last phrase needs explaining. Also this strongly suggests that the best way to work with symmetric polynomials is to take an infinity of variables.

A little bit more notation is useful. Let S_I be the group (under composition) of all bijections $\sigma\colon I \longrightarrow I$ such that there are only finitely many $i \in I$ with $\sigma(i) \neq i$. For $S_{\mathbf{N}}$ the notation S_∞ is usually used. For an exponent sequence e let e^σ be the exponent sequence $e^\sigma(i)=e(\sigma^{-1}(i))$.[107]

---

[106] All this just formalizes what everyone knows intuitively. One can also do these things for noncommuting indeterminates, and things are (curiously enough?) actually easier in that case. For totally ordered index sets one can go much further and make sense of infinite ordered products and sums by using injective limit ideas to make sense of infinite ordered sums and in finite products of elements of the base field and the integers by treating them as sequences with two sequences equal if they eventually agree.

[107] The exponent '-1' is there to ensure that $e^{\sigma\tau}=(e^\sigma)^\tau$; not that that is important in the present context.

A power series $\sum_{e\in E(I)} a_e\xi^e$ is now symmetric if and only if $a_e = a_{e^\sigma}$ for all $\sigma \in S_I$. A polynomial in an infinite set of commutating indeterminates is a finite sum $\sum_{e\in E(I)} a_e\xi^e$ i.e. a power series with all but finitely many of the coefficients a_e unequal to zero. A power series $\sum_{e\in E(I)} a_e\xi^e$ is of bounded degree if and only if there is a natural number n such that $a_e = 0$ for all e with $\mathrm{wt}(e) > n$. The complete symmetric functions in the ξ are by definition

$$h_n = \sum_{\mathrm{wt}(e)=n} \xi^e \tag{A.9}$$

The main theorem for symmetric power series in an infinity of indeterminates now takes the following form.

A.10. *Theorem.* Every symmetric power series can be uniquely written as a power series in the complete symmetric functions. A bounded degree symmetric power series is a polynomial in the complete symmetric functions.

There is of course a similar theorem in terms of the elementary symmetric functions.

The algebra **Symm** is the algebra of polynomials in the countable set of complete symmetric functions h_n for the case of the index set $I = \mathbf{N}$.

A.11. *Projective limit description.* For the finicky (or pernickety) the following projective limit construction is perhaps more congenial. For each n consider the algebra morphism

$$\pi_{n+1,n}\colon k[\xi_1, \cdots, \xi_n, \xi_{n+1}] \longrightarrow k[\xi_1, \cdots, \xi_n],$$
$$\xi_i \mapsto \xi_i \text{ for } i \in \{1, \cdots, n\}, \xi_{n+1} \mapsto 0$$

These are graded algebra morphisms

Let $\mathbf{Symm}_k^{(n)}$ be the subalgebra of symmetric polynomials $k[x_1, \cdots, x_n]$. This gives a graded projective system of graded algebras

$$\pi_{n+1,n}\colon \mathbf{Symm}_k^{(n+1)} \longrightarrow \mathbf{Symm}_k^{(n)}$$

and **Symm** is the graded projective limit of this system. So, for instance, the symmetric power series $\sum_{e\in E(\mathbf{N})} \xi^e$ is not in **Symm**. In this picture theorem A.10 is obtained by the usual theorem for symmetric functions in a finite number of variables and using that for each given degree the coefficients involved in that degree stabilize as $n \to \infty$.

References

[1] *Brauer group*, in M. Hazewinkel, ed., *Encyclopaedia of mathematics, volume 1*, Kluwer Acad. Publ., 1988, 479–480. http://eom.springer.de.

[2] *Central simple algebra*, in M. Hazewinkel, ed., *Encyclopaedia of mathematics, volume 2*, Kluwer Acad. Publ., 1988, 87. http://eom.springer.de.

[3] *Majorization ordering*, in M. Hazewinkel, ed., *Encyclopaedia of Mathematics, volume 6*, Kluwer Acad. Publ., 1990, 74–76. http://eom.springer.de.
[4] *Witt vector, Wikipedia*, 2007. http://en.wikipedia.org/wiki/Witt_vector.
[5] *Witt vector*, in M. Hazewinkel, ed., *Encyclopaedia of mathematics, volume 9*, Kluwer Acad. Publ., 1993, 505–507. http://eom.springer.de/W/w098100.htm.
[6] *Witt vectors*, *PlanetMath*, 2005. http://planetmath.org/encyclopaedia/Witt Vectors.html.
[7] *Witt vectors*, in S. Iyanaga and Y. Kawada, eds., *Encyclopaedic dictionary of mathematics*, MIT, 1960, 1690–1691.
[8] Abotteen, Essam A, *Inner plethysm in the representation ring of the general linear group*, Missouri J. of the math. Sci. **2** (1990), 115–131.
[9] Abotteen, Essam A, *On the structure of the Hopf representations ring of the symmetric groups*, Missouri J. of the math. Sci. **4** (1992), 20–29.
[10] Abrashkin, V. A., *Explicit formulae for the Hilbert symbol of a formal group over the Witt vectors*, Izvestiya Mathematics **61** (1997), 463–515.
[11] Abrashkin, V.A., *Galois moduli of period p-group schemes over a ring of Witt vectors*, Mathematics of the USSR-Izvestiya **51** (1987), 1–46.
[12] Abrashkin, V.A., *The image of the Galois group for some crystalline representations*, Izvestiya Mathematics **63** (1999), 1–36.
[13] Abrashkin, V.A., *Modular-Representations of the Galois group of a local field, and a generalization of the Shafarevich conjecture*, Mathematics of the USSR-Izvestiya **53** (1989), 469–518.
[14] Adams, J. Frank, *Vector fields on spheres*, Ann. of Math. **75** (1962), 603–632. See also the $J(X)$ series of papers: On the groups $J(X)$ I-IV, *Differential and combinatorial topology symposium in honor of Marston Morse*, Princeton Univ. Press, 1965, 121–143; Topology **2** (1963), 181–195; Topology **3** (1965), 193–222; Topology **5** (1966), 21–71; Topology **7** (1968), 331 (correction).
[15] Agboola, A. and D. Burns, *Grothendieck groups of bundles on varieties over finite fields*, K-Theory **23** (2001), 251–303.
[16] Albert, Adrian A, *Cyclic fields of degree p^n over F of characteristic p*, Bull. Amer. math. Soc. **40** (1934), 625–631.
[17] Almkvist, Gert, *Endomorphisms of finitely generated projective modules over a commutative ring*, Arkiv för Matematik **11** (1973), 263–301.
[18] Almkvist, Gert, *The Grothendieck ring of the category of endomorphisms*, J. of Algebra **28** (1974), 375–388.
[19] Almkvist, Gert, *Integrity of ghosts*, preprint, 2006.
[20] Almkvist, Gert, *K-theory of endomorphisms*, J. of Algebra **55** (1978), 308–340. Erratum: J. of Algebra **68** (1981), 520–521.
[21] Alon, Noga, *Splitting necklaces*, Advances in Mathematics **63** (1987), 247–253.
[22] Alon, Noga, and Douglas B West, *The Borsuk-Ulam theorem and bisection of necklaces*, Proc. Amer. math. soc. **98** (1986), 623–628.
[23] Artin, Emil and O Schreier, *Eine Kenzeichnung der reell abgeschlossenen Körper*, Abh. math. Sem. Hamburg **5** (1927), 225–231.
[24] Artin, Michael and Barry Mazur, *Formal groups arising from algebraic varieties*, Ann. sci. de l'École norm. sup. **10** (1977), 87–131.
[25] Ash, A., *Non-vanishing of alternants*, Linear Algebra and its Applications **411** (2005), 348–355.
[26] Atiyah, Michael Francis and Graeme B Segal, *Exponential isomorphisms for λ-rings*, Quartely J. Math. Oxford **22** (1971), 371–378.
[27] Atiyah, Michael Francis and D O Tall, *Group representations, λ-rings and the J-homomorphism*, Topology **8** (1969), 253–297.
[28] Auer, R., *A functorial property of nested Witt vectors*, Journal of Algebra **252** (2002), 293–299.
[29] Badra, A., *Deformations of finite p-group schemes to a formal scheme*, Comptes Rendus de l' Academie des Sciences, Serie I-Mathematique **325** (1997), 177–181.
[30] Baker, Andrew and Birgit Richter, *Quasisymmetric functions from a topological point of view*, Math. Dept., Univ. of Glasgow, Scotland; Math. Dept. Univ. of Hamburg, Germany, 2006.

[31] Ballico, E., *Applications of the Lifting from characteristic p of some rational n-fold*, Journal of pure and applied Algebra **103** (1995), 285–286.
[32] Balmer, Paul, *Witt groups*, in E.M. Friedlander and et al., eds., *Handbook of K-theory. Volume 2*, Springer, 2005, 539–279.
[33] Barr, Michael and Charles Wells, *Toposes, triples, and theories*, Springer, 1985.
[34] Barsotti, Iacopo, *On Witt vectors and periodic group varieties*, Illinois Journal of Mathematics **2** (1958), 99–110.
[35] Bélair, L., *Difference equations in Witt vectors*, Comptes Rendus de l' Academie des Sciences, Série I-Mathématique **340** (2005), 99–102.
[36] Bélair, L., *Integral valued difference rational functions over Witt vectors*, Comptes Rendus de l' Academie des Sciences, Série I-Mathématique **339** (2004), 83–86.
[37] Bélair, L. and A. Macintyre, *The Frobenius automorphism of Witt vectors*, Comptes Rendus de l' Academie des Sciences, Série I-Mathématique **331** (2000), 1–4.
[38] Bélair, L., A. Macintyre and T. Scanlon, *Model theory of the Frobenius on the Witt vectors*, American Journal of Mathematics **129** (2007), 665–721.
[39] Bélair, Luc, *Approximation for Frobenius algebraic equations in Witt vectors*, Université de Quebec, Montréal, Quebec, Canada, 2006.
[40] Bélair, Luc, *Vecteurs de Witt*, Publ. math. Univ. de Paris VII **33** (1990), 141–149.
[41] Berger, L., *Limits of crystalline representations*, Compositio mathematica **140** (2004), 1473–1498.
[42] Bergeron, François, *A combinatorial outlook on symmetric functions*, J. of combinatorial Theory, series A **50** (1989), 226–234.
[43] Bergeron, François, *Une combinatoire du pléthysme*, J. of combinatorial Theory, series A **46** (1987), 291–305.
[44] Bergman, G.M., *Ring schemes*: *The Witt scheme (with appendix)*, in D. Mumford, ed., *Lectures on curves on an algebraic surface*, Princeton Univ. Press, 1966, 171–192.
[45] Bergman, George, M. and Adam, O. Hausknecht, *Cogroups and co-rings in categories of associative rings*, American math. Soc., 1996.
[46] Berstein, Israel, *On cogropus in the category of graded algebras*, Trans. Amer. math. Soc. **115** (1965), 257–269.
[47] Berthelot, P., S. Bloch and H. Esnault, *On Witt vector cohomology for singular varieties*, Compositio mathematica **143** (2007), 363–392.
[48] Berthelot, Pierre, *Généralités sur les lambda-anneaux*, in P. Berthelot, A. Grothendieck and L. Illusie, eds, *Séminaire de géométrie algébrique du Bois Marie 1966/67 (SGA 6)*, Springer, 1971, 279–364.
[49] Blache, R.G., *A Stickelberger theorem for p-adic Gauss sums*, Acta arithmetica **118** (2005), 11–26.
[50] Bleher, F.M., *Universal deformation rings and dihedral 2-groups*, 2007. arXiv:0705.0834 [math.RT].
[51] Bleher, F.M., *Universal deformation rings and dihedral defect groups*, 2007. arXiv:math.RT/0607571.
[52] Bleher, F.M., *Universal deformation rings and Klein four defect groups*, Transactions of the American mathematical Society **354** (2002), 3893–3906.
[53] Bleher, F.M. and T. Chinburg, *Deformations with respect to an algebraic group*, Illinois Journal of Mathematics **47** (2003), 899–919.
[54] Blessnohl, Dieter and Manfred Schocker, *Noncommutative character theory of the symmetric group*, Imperial College Press, 2005.
[55] Bocklandt, Ralf and Lieven Le Bruyn, *Necklace Lie algebras and noncommutative symplectic geometry*, Math. Zeitschrift **240** (2002), 141–167.
[56] Boorman, Evelyn Hutterer, *S-operations in representation theory*, Transactions of the American mathematical Society **205** (1975), 127–149.
[57] Borger, James, *The basic geometry of Witt vectors*, Dept. Math., Australian national University, Canberra, Australia, 2007.

[58] Borger, James and Bart de Smit, *Galois theory and integral models of lambda rings*, Math. Sci. Inst., Australian national Univ., Canberra, 2007.

[59] Borger, James and Ben Wieland, *Plethystic algebra*, Advances in Mathematics **194** (2005), 246–283. arXiv:math.AC/0407227.

[60] Borwein, J. and S. Lou, *Asymptotics of a sequence of Witt vectors*, Journal of Approximation Theory **69** (1992), 326–337.

[61] Bouc, Serge, *Burnside rings*, in M. Hazewinkel, ed., *Handbook of Algebra. Volume 2*, North Holland (imprint of Elsevier), 2000, 739–804.

[62] Bourbaki, Nicholas, *Algèbre commutative. Chapitres 5 et 6*, Masson, 1964. Chapitre 6: Valuations.

[63] Bourbaki, Nicholas, *Algèbre commutative. Chapitres 8 et 9*, Masson, 1983. Chapitre IX: Anneaux locaux Noethériens complets, §1: Vecteurs de Witt.

[64] Bourbaki, Nicholas, *Algèbre. Chapitre 4: Polynômes et fractions rationelles; Chapitre 5: Corps commutatifs*, Hermann, 1959.

[65] Bousfield, A K, *On lambda-rings and the K-theory of infinite loop spaces*, K-theory **10** (1996), 1–30.

[66] Brennan, Joseph P, *The restriction of the outer plethysm to a Young Subgroup*, Comm. in Algebra **21** (1993), 1029–1036.

[67] Breuil, C., *Construction of p-adic semi-stable representations*, Annales scientifiques de l'École normale supérieure **31** (1998), 281–327.

[68] Breuil, C., *Group schemes and filtered modules*, Comptes Rendus de l' Academie des Sciences, Série I-Mathématique **328** (1999), 93–97.

[69] Breuil, C., *Log-syntomic topology, log-crystalline cohomology and Cech cohomology*, Bulletin de la Soc. mathématique de France **124** (1996), 587–647.

[70] Brucks, K.M, *MSS sequences, colorings of necklaces, and periodic points of* $f(z) = z^2 - 2$, Adv. applied Math. **8** (1987), 434–445.

[71] Brun, M., *Witt vectors and equivariant ring spectra*, 2004. arXiv:math.AT/0411567.

[72] Brun, M., *Witt vectors and equivariant ring spectra applied to cobordism*, Proceedings of the London mathematical Society **94** (2007), 351–385.

[73] Brun, M., *Witt vectors and Tambara functors*, Advances in Mathematics **193** (2005), 233–256. Preprint version: arXiv:math.AC/0304495.

[74] Brun, Morton, *Introduction to Witt vectors*, preprint Fachbereich Mathematik/Informatik, Univ. Osnabrück, 2006.

[75] Bryden, J. and K. Varadarajan, *Witt vectors which are rational functions*, Communications in Algebra **21** (1993), 4263–4270.

[76] Buch, A., J.F. Thomsen, N. Lauritzen and V. Mehta, *Frobenius morphisms module p(2)*, Comptes Rendus de l' Academie des Sciences, Série I-Mathématique **322** (1996), 69–72.

[77] Buchheim, C. and H. Frommer, *Reduction of tori split over tamely ramified extensions*, Comptes Rendus de l' Academie des Sciences, Série I-Mathématique **338** (2004), 219–221.

[78] Burnside, William S, *Theory of groups of finite order*, Cambridge Univ. Press, 1911. Dover reprint: 1955.

[79] Burroughs, J, *Operations in Grothendieck rings and the symmetric group*, Canadian Journal of Mathematics **26** (1974), 543–550.

[80] Cahen, P.-J. and J.-L. Chabert, *Integer valued polynomials*, Amer. math. Soc., 1997.

[81] Calderbank, A.R., P. Hanlon and S. Sundaram, *Representations of the symmetrical group in deformations of the free Lie algebra*, Transactions of the American mathematical Society **341** (1994), 315–333.

[82] Carré, Christophe and Jean-Yves Thibon, *Plethysm and vertex operators*, Adv. in Math. **13** (1992), 390–403.

[83] Cartier, Pierre, *Groupes de Lubin-Tate généralisés*, Inventiones mathematicae **35** (1976), 273–284.

[84] Cariter, Pierre, *Groupes formels associés aux vecteurs de Witt généralisés*, Compt. Rend. Acad. Sci. Paris, Ser. A-B **265** (1967), A49–A52.

[85] Cartier, Pierre, *Modules associés à un groupe formel commutatif. Courbes typiques*, Compt. Rend. Acad. Sci. Paris, Ser. A-B **265** (1967), A129–A132.
[86] Cartier, Pierre, *Séminaire sur les groupes formels à l'IHES*, 1972. Unpublished lecture notes.
[87] Chevalley, Claude, *La théorie du symbole de restes normiques*, J. reine und angew. Math. (Crelle) **169** (1933), 140–157.
[88] Christol, Gilles, *Opération de Cartier et vecteurs de Witt*, J. of combinatorial Theory, series A **104** (1971), 217–263.
[89] Clauwens, F.J.B.J, *K-theory, lambda rings, and formal groups*, Compositio mathematica **65** (1988), 223–240.
[90] Clauwens, F.J-B.J, *Commuting polynomials and lambda ring structures on* $\mathbf{Z}[x]$, J. pure and applied Algebra **95** (1994), 261–269.
[91] Clauwens, F.J-B.J, *The K-theory of lambda-rings. Part I: Construction of the logarithmic invariant*, Compositio mathematica **61** (1987), 295–328.
[92] Clauwnes, F.J-B.J, *The K-theory of lambda-rings. Part II: Invertibility of the logarithmic map*, Compositio mathematica **92** (1994), 205–225.
[93] Cohen, I.S., *Commutative rings with restricted minimum condition*, Duke math. J. **17** (1950), 27–42.
[94] Cohen, I.S., *On the structure and ideal theory of complete local rings*, Transactions of the American mathematical Society **59** (1946), 54–106.
[95] Cohen, Miriam, Shlomo Gelaki and Sara Westreich, *Hopf algebras*, in M. Hazewinkel, ed., *Handbook of Algebra. Volume 4*, Elsevier, 2006, 173–239.
[96] Cohn, P.M., *Algebra. Volume 3*, Wiley, 1991.
[97] Colmez, Pierre, *Le corps des périodes p-adiques*, Comp. Rendus Acad. Sci. Paris, Série I **310** (1990), 321–324.
[98] Connes, Alain, *Cohomologie cyclique et foncteurs* Ext^n, CR Acad. Sci. Paris, Sér. I Math **296** (1983), 953–958.
[99] Contou-Carrère, Carlos, *Jacobienne locale, groupe de bivecteurs de Witt universel, et symbole modéré*, Compt. Rend. Acad. Sci. Paris, Sér. I (1994), 743–746.
[100] Cronheim, Arno, *Dual numbers, Witt vectors, and Hjelmslev planes*, Geometria dedicata **7** (1978), 287–302.
[101] Da Costa, G.A.T.F and J. Variane Jr, *Feynman identity: a special case revisited*, Letters in mathematical Physics **73** (2005), 221–235.
[102] Dauns, John, *A concrete approach to division rings*, Heldermann Verlag, 1982.
[103] de Cataldo, M.A.A., *Vanishing via lifting to second Witt vectors and a proof of an isotriviality result*, Journal of Algebra **219** (1999), 255–265.
[104] de Graaf, W.A., *Constructing homomorphisms between Verma modules*, Journal of Lie Theory **15** (2005), 415–428.
[105] Demazure, Michel and Pierre Gabriel, *Groupes algébriques*, North Holland, 1970. Chap. V, \$1–3.
[106] Demchenko, O., *Covariant Honda theory*, Tohoku mathematical Journal **57** (2005), 303–319.
[107] Dieudonné, Jean, *On the Artin-Hasse exponential series*, Proc. Amer. math. Soc. **8** (1957), 210–214.
[108] Dobrev, V.K. *Singular vectors of quantum group representations for straight Lie-algebra roots*, Letters in Mathematical Physics **22** (1991), 251–266.
[109] Dobrev, V.K. *Singular vectors of representations of quantum groups*, Journal of Physics A-mathematical and general **25** (1992), 149–160.
[110] Dobrev, V.K. and M. El Falaki, *Quantum group* $U_q(A(l))$ *singular vectors in Poincaré-Birkhoff-Witt basis*, Letters in Mathematical Physics **49** (1999), 47–57.
[111] Dobrev, V.K. and M. El Falaki, *Quantum group* $U_q(D\text{-}l)$ *singular vectors in the Poincare-Birkhoff-Witt basis*, Journal of Physics A-mathematical and general **33** (2000), 6321–6332.
[112] Dobrev, V.K. and P. Truini, *Irregular* $U_q(sl(3))$ *representations at roots of unity via Gel'fand-(Weyl)-Zetlin basis*, Journal of mathematical Physics **38** (1997), 2631–2651.

[113] Dold, Albrecht, *Fixed point indices of iterated maps*, Inventiones mathematicae **74** (1983), 419–435.

[114] Dress, A. and T. Muller, *Decomposable functors and the exponential principle*, Advances in Mathematics **129** (1997), 188–221.

[115] Dress, A.W.M. and C. Siebeneicher, *The Burnside Ring of infinite groups, Witt vectors, necklace algebras, lambda-rings and all that*, Dept. Math., University of Bielefeld, Germany, 1986. Preprint version of 120.

[116] Dress, A.W.M. and C. Siebeneicher, *A multinomial Identity for Witt vectors*, Advances in Mathematics **80** (1990), 250–260.

[117] Dress, Andreas W.M., *A characterisation of solvable groups*, Math. Zeitschrift **110** (1969), 213–217.

[118] Dress, Andreas W.M., *Congruence relations characterizing the representation ring of the symmetric group*, J. of Algebra **101** (1986), 350–364.

[119] Dress, Andreas W.M., *Contributions to the theory of induced representations*, in H. Bass, ed., *Algebraic K-theory II*, Springer, 1973, 183–242.

[120] Dress, Andreas W.M., and Christian Siebeneicher, *The Burnside ring of the infinite cyclic group and its relations to the necklace algebra, λ-rings, and the universal ring of Witt vectors*, Adv. in Math. **78** (1989), 1–41.

[121] Dress, Andreas W.M., and Christian Siebeneicher, *On the number of solutions of certain linear diophantine equations*, Hokkaido math J. **19** (1990), 385–401.

[122] Dress, Andreas W.M., and Ernesto Vallejo, *A simple proof for a result by Kratzer and Thévenaz concerning the embedding of the Burnside ring into the ghost ring*, J. of Algebra **154** (1993), 356–359.

[123] Dress, Andreas W.M., *Notes on the theory of representations of finite groups and related topics with two appendices, A: The Witt ring as a Mackey functor, B: A relation between Burnside- and Witt-rings*, Dept. Math., University of Bielefeld, Germany, 1971.

[124] Dress, Andreas W.M., and Christian Siebeneicher, *On the integrality of the Witt polynomials, Séminaire Lotharingien de Combinatoire*, 1992.

[125] Dress, Andreas W.M., Christian Siebeneicher and Tomoyuki Yoshida, *An application of Burnside rings in elementary finite group theory*, Advances in Mathematics **91** (1992), 27–44.

[126] Dress, Andreas W.M., and Christian Siebeneicher, *The Burnside ring of profinite groups and the Witt vector construction*, Adv. in Math. **70** (1988), 87–132.

[127] Dubrovin, B.A., *Generalized Witt groups (Russian)*, Mat. Zametki **13** (1973), 419–426.

[128] Duchamp, G., D. Krob, B. Leclerc and J.-Y. Thibon, *Fonctions quasi-symmétriques, fonctions symmétriques noncommutatives, et algèbres de Hecke à $q = 0$*, CR Acad. Sci. Paris, Série I- Math. **322** (1996), 107–112.

[129] Duchamp, Gérard, Florent Hivert and Jean-Yves Thibon, *Noncommutative symmetric functions VI: free quasisymmetric functions and related algebras*, Int. J. Algebra Comput. **12** (2002), 671–717.

[130] Duchamp, Gérard, Alexander Klyachko, Daniel Krob and Jean–Yves Thibon, *Noncommutative symmetric functions III: deformations of Cauchy and convolution algebras*, Discrete Math. Theor. Comput. Sci. **1** (1996), 159–216.

[131] Dumas, François, *Hautes dérivations et anneaux de séries non commutatifs en charactérisituque nulle*, Compt. Rend. Acad. Sci. PAris, Sér. I Math. **303** (1986), 383–385.

[132] Dwork, Bernard, *Norm residue symbol in local number fields*, Abh. math. Sem. Hamburg **22** (1958), 180–190.

[133] Dwyer, W.G., M.J. Hopkins and D.M. Kan, *The homotopy theory of cyclic sets*, Trans. Amer. math. Soc. **291** (1985), 281–289.

[134] Elliott, J., *Binomial rings, integer-valued polynomials, and lambda-rings*, Journal of pure and applied Algebra **207** (2006), 165–185.

[135] Elliott, J., *Constructing Witt-Burnside rings*, Advances in Mathematics **203** (2006), 319–363.

[136] Endler, O., *Valuation theory*, Springer, 1972.

[137] Faith, Carl., *Algebra II. Ring theory*, Springer, 1976.

[138] Faith, Carl., *Projective ideals in Cohen rings*, Archiv Math. **26** (1975), 588–594.
[140] Faltin, F, N. Metropolis, B. Ross and G.-C. Rota, *The real numbers as a wreath product*, Adv. in Math. **16** (1975), 278–304.
[141] Feigin, Boris .I and Boris .I. Tsygan, *Additive K-theory*, in Y.I. Manin, ed., *K-theory: arithmetic and geometry*, Springer, 1987, 97–209.
[142] Ferrero, Miguel, Antonio Paques and Andrzej Solecki, *On $\mathbf{Z}_p$-extensions of commutative rings*, J. Pure and applied Algebra **72** (1991), 5–22.
[143] Finotti, L.R. A., *Minimal degree liftings in characteristic 2*, Journal of pure and applied Algebra **207** (2006), 631–673.
[144] Fredricksen, Harold and Irving J. Kessler, *An algorithm for generating necklaces of beads in two colors*, Discrete Mathematics **61** (1986), 181–188.
[145] Fredrickesn, Harold and James Maiorana, *Necklaces of beads in k colors and k-ary de Bruijn sequences*, Discrete Mathematics **23** (1978), 207–210.
[146] Fresse, Benoit, *Cogroups in algebras over an operad are free algebras*, Comm. math. Helv. **73** (1998), 637–676.
[147] Fripertinger, Harald, *Endliche Gruppenaktionen auf Funktionenmengen. Das Lemma von Burnside, Repräsentanenkonstruktionen, Anwendungen in der Musiktheorie*, Bayreuther Mathematische Schriften **45** (1993), 19–132.
[148] Fulton, William and Serge Lang, *Riemann-Roch algebra*, Springer, 1985.
[149] Gabriel, Peter, *Universelle Eigenschaften der Wittsche Vektoren und der Einseinheitenalgebra einer Potenzreihenalgebra in einer Veränderlichen*, Jahresbericht Deutsche Mathematiker Vereinigung **72** (1970), 116–121.
[150] Gale, D., *A theorem on flows in networks*, Pacific Journal of Mathematics **7** (1957), 1073–1083.
[151] Gan, Wee Liang and Travis Schedler, *The necklace Lie coalgebra and renormalization algebras*, 2007. arXiv:math-ph/0702055v3.
[152] Garsia, A.M., and Nolan Wallach, *Qsym over Sym is free*, J. of combinatorial Theory, Series A **104** (2003), 217–263.
[153] Garsia, A.M., and Nolan Wallach, *r-Qsym is free over Sym*, J. of combinatorial Theory, Series A **114** (2006), 704–732.
[154] Gaudier, Henri, *Algèbres de group du groupe additif*, Bull. Soc. math. de France **117** (1989), 233–245. http://emis.kaist.ac.kr/journals/SLC/opapers/s21.html.
[155] Gaudier, Henri, *Multiplication des matrices et vecteurs de Witt*, Séminaire Lotharingien de Combinatoire, IRMS Strassbourg **21** (1989). http://emis.kaist.ac.kr/journals/SLC/opapers/s21.html.
[156] Gaudier, Henri, *Relêvement des coefficients binômiaux dans les vecteurs de Witt*, Séminaire Lotharingien de Combinatoire, IRMA Strassbourg **18** (1988), 93–108. http://www.mat.univie.ac.at/∼slc/s18gaudier.html.
[157] Gavarini, F., *A PBW basis for Lusztig's form of untwisted affine quantum groups*, Communications in Algebra **27** (1999), 903–918.
[158] Gay, C.D., G.C. Morris and I. Morris, *Computing Adams operations on the Burnside ring of a finite group*, J. reine und angew. Math. (Crelle) **341** (1983), 87–97.
[159] Gebhard, David D., *Noncommutative symmetric functions and the chromatic polynomial*, Dept Math., Michigan State Uni., East Lansing, Michigan, USA, 1994.
[160] Geisser, T., *On K-3 of Witt vectors of length two over finite fields*, K-Theory **12** (1997), 193–226.
[161] Geissinger, Ladnor, *Hopf algebras of symmetric functions and class functions,* in D. Foata, ed., *Combinatoier et réprésentation du groupe symmétrique*, Springer, 1977, 168–181.
[162] Gelfand, Israel M., Daniel Krob, Alain Lascoux, Bernard Leclerc, Vladimir S Retakh and Jean- Yves-Thibon, *Noncommutative symmetric functions*, Adv. in Math. **112** (1995), 218–348.
[163] Gerstenhaber, Murray, *On the deformation theory of rings and algebras*, Ann. of Math. **79** (1964), 59–103.
[164] Gerstenhaber, Murray, *On the deformation theory of rings and algebras: II*, Ann. of Math. **84** (1966), 1–19.

[165] Gerstenhaber, Murray, *On the deformation theory of rings and algebras: III*, Ann. of Math. **88** (1968), 1–34.
[166] Gerstenhaber, Murray, *On the deformation theory of rings and algebras: IV*, Ann. of Math. **99** (1974), 257–276.
[167] Gerstenhaber, Murray and S. Don Shack, *The shuffle bialgebra and the cohomology of commutative algebras*, J. pure and appl. Algebra **70** (1991), 263–272.
[168] Gerstenhaber, Murray and Clarence Wilterson, *On the deformation theory of rings and algebras V: Deformation of differential graded algebras*, in J. McCleary, ed., *Higher homotopy structures in topology and mathematical physics*, Amer. math. Soc., 1999, 89–102.
[169] Gingold, H. and A. Knopfmacher, *Analytic properties of power product expansions*, Canadian Journal of Mathematics-Journal Canadien de Mathématiques **47** (1995), 1219–1239.
[170] Ginzburg, Victor and Travis Schedler, *Moyal quantization and stable homology of necklace Lie algebras*, 2006. arXiv:math/0605704v2 [math. QA].
[171] Godement, Roger, *Topologie algébrique et théorie des faisceaux*, Hermann, 1964.
[172] Goerss, P., J. Lannes and F. Morel, *Vecteurs de Witt non commutatifs et représentabilité de l'homologie modulo p*, Inventiones mathematicae **108** (1992), 163–227.
[173] Goerss, Paul, J. Lannes and F. Morel, *Hopf algebras*, *Witt vectors*, *and Brown-Gitler spectra*, in M.C. Tangora, ed., *Algebraic topology*, *Oaxtepec 1991*, Amer. math. Soc., 1993, 217–263.
[174] Golomb, Solomon W., *Irreducible polynomials*, *synchronization codes*, *primitive necklaces*, *and the cyclotomic algebra*, in R.C. Bose and T. A. Dowling, eds., *Combinatorial mathematics and its applications*, Univ of North Carolina Press, 1967, 358–370.
[175] Graham, John J., *Generalized Witt vectors*, Advances in Mathematics **99** (1993), 248–263. Preprint version: "A new approach to the Witt-Burnside ring", Dept. pure Math., Univ. of Sydney, Australia, Research report 90–5, March 1990.
[176] Green, B., *Realizing deformations of curves using Lubin-Tate formal groups*, Israel Journal of Mathematics **139** (2004), 139–148.
[177] Greenberg, Marvin J., *Unit Witt vectors*, Proc. Amer. math. Soc. **13** (1962), 72–73.
[178] Grothendieck, Alexander, *La théorie des classes de Chern*, Bull. Soc. math. de France **86** (1958), 137–154.
[179] Grothendieck, Alexandre, *Classes de faisceaux et théorème de Riemann-Roch*, in P. Berthelot, A. Grothendieck and L. Illusie, eds., *Séminaire de géométrie algébrique du Bois Marie 1966/67* (*SGA 6*), Springer, 1971, 20–77.
[180] Gubeladze, J., *Higher K-theory of toric varieties*, K-Theory **28** (2003), 285–327.
[181] Gubeladze, J., *Toric varieties with huge Grothendieck group*, Advances in Mathematics **186** (2004), 117–124.
[182] Guerrini, Luca, *Formal and analytic deformations of the Witt algebra*, Lett. in math. Physics **46** (1998), 121–129.
[183] Guillot, Pierre, *Adams operations in cohomotopy*, 2008. arXiv:math/0612327v1 [math.AT]
[184] Guillot, Pierre, *The representation ring of a simply connected Lie group as a lambda-ring*, Comm. Algebra **35** (2007), 875–883.
[185] Haboush, W.J., *Infinite dimensional algebraic geometry; Algebraic structures on p-adic groups and their homogeneous spaces*, Tohoku mathematical Journal **57** (2005), 65–117.
[186] Hall, Philip, *Nilpotent groups*. *Notes of lectures given at the Canadian mathematical congress*, *summer seminar*, *Univ. of Alberta*, *August 1957*, Dept. Math., Queen Mary College, London, UK, 1969. These notes are often referred to as the "Edmonton notes".
[187] Haraguchi, Yuki, *On noncommutative extensions of G_a by G_m over an F_p-algebra*, Tsukuba J. of Math. **29** (2005), 405–435.
[188] Harder, G., *Wittvektoren*, Jahresbericht Deutsche Mathematiker Vereinigung **99** (1997), 18–48. A translation into English of this paper appears in "Ina Kersten (ed.), E. Witt: Gesammelte Abhandlungen, Springer, 1998, 165–194", see [238].
[189] Hartshorne, Robin, *Algebraic geometry*, Springer, 1977.
[190] Hasse, Helmut, *Theorie der relativ zyklischen algebraische Funktionenkörper*, *insbesondere bei endlichem Konstantenkörper*, J reine angew. Math. (= Crelle) **172** (1934), 37–54.

[191] Hasse, Helmut and Friedrich Karl Schmidt, *Die Struktur diskret bewerter Körper*, J reine angew. Math. (= Crelle) **170** (1934), 4–63.
[192] Hazewinkel, M., *Formal groups and applications*, Acad. Press, 1978. Free electronic copy: 'Darenet/Cream of Science' <www.darenet.nl/en/page/language.view/keur.page>.
[193] Hazewinkel, Michiel, *The algebra of quasi-symmetric functions is free over the integers*, Advances in Mathematics **164** (2001), 283–300.
[194] Hazewinkel, Michiel, *Cofree coalgebras and multivariable recursiveness*, J. of pure and applied Algebra **183** (2003), 61–103.
[195] Hazewinkel, Michiel, *Explicit polynomial generators for the ring of quasisymmetric functions over the integers*, 2002. arXiv:math/0410365.
[196] Hazewinkel, Michiel, *On the Snapper Liebler-Vitale Lam theorem on permutation representations of the symmetric froup*, J. pure and applied Algebra **23** (1982), 29–32.
[197] Hazewinkel, Michiel, *Operations in the K-theory of endomorphisms*, J. of Algebra **84** (1983), 285–305. Free electronic copyg: 'Darenet/Cream of Science' <www.darenet.nl/en/page/language.view/keur.page>.
[198] Hazewinkel, Michiel, *The primitives of the Hopf algebra of noncommutative symmetric functions over the integers*, CWI, 2001. arXiv:math/0410365.
[199] Hazewinkel, Michiel, *Six chapters on Hopf algebras*. Preliminary version of the first six (of seven) chapters on Hopf algebras for: "Michiel Hazewinkel, Nadiya Gubareni, Volodymyr Kirichenko, *Algebras*, *rings*, *and modules*. *Volume 3*", Springer, to appear., 2008.
[200] Hazewinkel, Michiel, *Symmetric functions*, *noncommutative symmetric functions and quasisymmetric functions*, Acta appl. Math. **75** (2003), 55–83. Free electronic copy: 'Darenet/Cream of Science' <www.darenet.nl/en/page/language.view/keur.page>. Also arXiv:math/0410468.
[201] Hazewinkel, Michiel, *Symmetric functions*, *noncommutative symmetric functions and quasisymmetric functions II,* Acta appl. Math. **85** (2005), 319–340. Free electronic copy: 'Darenet/Cream of Science' <www.darenet.nl/en/page/language.view/keur.page>. Also arXiv:math/0410470.
[202] Hazewinkel, Michiel and Clyde F. Martin, *Representations of the symmetric groups*, *the specialization order*, *systems*, *and Grassmann manifolds*, Enseignement math. **29** (1983), 53–87.
[203] Hazewinkel, Michiel and Ton Vorst, *On the Snapper Liebler-Vitale Lam theorem on permutation representations of the symmetric groups*, J. pure and applied Algebra **23** (1982), 29–32.
[204] Hellegouarch, Yves and François Recher, *Canonical t-modules*, J. of Algebra **187** (1997), 323–272.
[205] Hellegouarch, Yves and François Recher, *Relèvement de modules de Drinfeld en charactéristique zéro*, Comp. Rendus Math. Acad. Sci., Soc. royale de Canada **15** (1993), 167–172.
[206] Hesselholt, L., *Correction to Witt vectors of non-commutative rings and topological cyclic homology*, [Hesselholt, 1997 #108], Acta Math. **95** (2005), 155–160.
[207] Hesselholt, L., *Witt vectors of non-commutative rings and topological cyclic homology*, Acta Mathematica **178** (1997), 109–141.
[208] Hesselholt, L., Witt vectors of non-commutative rings and topological cyclic homology (vol 178, pg 109, 1997), Acta mathematica 195 (2005), 55–60.
[209] Hesselholt, L. and I. Madsen, *On the K-theory of finite algebras over Witt vectors of perfect fields*, Topology **36** (1997), 29–101.
[210] Hesselholt, Lars, *The absolute de Rham-Witt complex*, MIT, Cambridge, Massachusetts, USA, 2005.
[211] Hesselholt, Lars, *Galois cohomology of Witt vectors of algebraic integers*, Math. Proc. of the Cambridge phil. Soc. **137** (2004), 55157.
[212] Hesselholt, Lars, *Lecture notes on Witt vectors*, MIT, Cambridge, Massachusetts, USA, 2005.
[213] Hirzebruch, Friedrich, *Topological methods in algebraic geometry*, Springer, 1966. Original edition: Neue topologische Methoden in der algebraische Geometrie, Springer, 1962.

[214] Hivert, Florent, *Hecke algebras, difference operators and quasi-symmetric functions*, Adv. Math. **155** (2000), 181–238.
[215] Hivert, Florent, Alain Lascoux and Jean-Yves Thibon, *Noncommutative symmetric functions and quasi symmetric functions with two and more parameters*, Inst. Gaspard Monge, Univ. Marne-la-Vallée, 2001.
[216] Hivert, Florent, Alain Lascoux and Jean-Yves Thibon, Noncommutative symmetric functions and quasisymmetric functions with two and more parameters, Inst. Gaspard Monge, Univ. Marne-la-Vallée, 2001.
[217] Hivert, Florent, Jean-Christophe Novelli and Jean-Yves Thibon, *Une analogue du monoïde plaxique pour les arbres binaires de recherche*, 2002.
[218] Hoffman, Michael E., *The algebra of multiple harmonic series*, J. of Algebra **194** (1997), 477–495.
[219] Hoffman, Michael E., *Quasi-shuffle products*, J. algebraic Combinatorics **11** (2000), 49–68.
[220] Hoffman, P., *Exponential maps and λ-rings*, J. pure and appl. Algebra **27** (1983), 131–162.
[221] Hoffman, Peter, *τ-rings and wreath product representations*, Springer, 1979.
[222] Holland, Martin P., *K-theory of endomorphism rings and of rings of invariants*, J. of Algebra **191** (1997), 668–685.
[223] Hyodo, Osamu, *A cohomological construction of Swan representations over the Witt ring I, II*, Proc. Japan Acad. Ser. A **64** (1988), 300–303, 350–351.
[224] Hyodo, Osamu, *On the de Rham-Witt complex attached to a semi-stable family*, Compositio mathematica **78** (1991), 241–260.
[225] Illusie, Luc, *Complex de de Rham-Witt et cohomologie cristalline*, Ann. sci. École norm. sup. **12** (1979), 501–661.
[226] Ischebeck, Friedrich and Volker Kokot, *Modules of Witt vectors*, in R. Tandon, ed., *Algebra and number theory*, Hindustan Book Agency, 2005.
[227] Itzykson, C., *Geometry of Differential-Equations and Projective-Representations of the Witt Algebra*, International Journal of Modern Physics B **6** (1992), 1969–2003.
[228] Jacobi, Carl Gustav, *De functiones alternantibus*, J. reine und angew. Math. (Crelle) **22** (1841), 360–371. Also in: Werke **3**, 439–452.
[229] Jacobson, Nathan, *Abstract algebra. Volume 3: theory of fields and Galois theory*, van Nostrand, 1964. pp. 124–132, 234–236.
[230] James, Gordon and Adalbert Kerber, *The representation theory of the symmetric group*, Addison-Wesley, 1981.
[231] Jean, Sandrine, *Witt vectors over local fields of positive characteristic and conjugacy classes of p-torsion elements in the Nottingham group*, copy of overhead transparancies, 2007.
[232] Kambayashi, T., *On the Witt vectors over nonperfect rings*, J. Math. Kyoto Univ. **8** (1968), 279–283.
[233] Kanesaka, Kiyomi and Koji Sekiguchi, *Representation of Witt vectors by formal power series and its applications*, Tokyo J. of Math. **2** (1979), 249–270.
[234] Kang, Seok-Jin and Myung-Hwan Kim, *Free Lie algebras, generalized Witt formula, and the denominator identity*, J. of Algebra **183** (1996), 560–594.
[235] Kaplansky, Irving, *Topics in commutative ring theory*, Dept. of Mathematics, University of Chicago, 1974.
[236] Kashiwabara, Takuji and W. Stephen Wilson, *The Morava K-theory and Brown-Peterson cohomology of spaces related to BP*, J. Math. Kyoto Univ. **41** (2001), 43–95.
[237] Kedlaya, K. S., *Power series and p-adic algebraic closures*, Journal of Number Theory **89** (2001), 324–339.
[238] Kersten, Ina, ed., *Ernst Witt. Collected papers. Gesammelte Abhandlungen*, Springer, 1998.
[239] Knutson, Donald, *Lambda rings and the representation theory of the symmetric groups*, Springer, 1973.
[240] Koch, A., *Endomorphisms of monogenic Hopf algebras,* Communications in Algebra **35** (2007), 747–758.

[241] Koch, A., *The Hopf algebra of a uniserial group*, Pacific Journal of Mathematics **215** (2004), 347–356.
[242] Koch, A., *Monogenic bialgebras over finite fields and rings of Witt vectors*, Journal of pure and applied Algebra **163** (2001), 193–207.
[243] Koch, A., *Monogenic Hopf algebras and local Galois module theory*, Journal of Algebra **264** (2003), 408–419.
[244] Koch, A., *Witt subgroups and cyclic Dieudonné modules killed by p*, Rocky Mountain J. of Mathematics **31** (2001), 1023–1038.
[245] Koch, Alan, *Lifting Witt subgroups to characteristic zero*, New York J. Math. **4** (1998), 127–136. Journal is electronic only.
[246] Kokot, Volker, *Moduln von Wittvektoren*, Münster, 2000.
[247] Kozma, Ilan, *Witt vectors and complex cobordism*, Topology **13** (1974), 389–394.
[248] Kraft, Hanspeter, *Kommutative algebraische Gruppen und Ringe*, Springer, 1975.
[249] Krob, D., B. Leclerc and J.-Y. Thibon, *Noncommutative symmetric functions II: transformations of alphabeths*, Int. J. of Algebra and Computation **7** (1997), 181–264.
[250] Krob, D., and J.-Y. Thibon, *Noncommutative symmetric functions V: a degenerate version of* $U_q(gl_N)$, Int. J. of Algebra and Computation **9** (1997), 405–430.
[251] Krob, Daniel and Jean-Yves Thibon, *Noncommutative symmetric functions IV: quantum linear groups and Hecke algebras at* $q=0$, J. of algebraic Combinatorics **6** (1997), 339–376.
[252] Kung, Joseph P.S., ed., *Gian-Carlo Rota on combinatorics*, Birkhäuser, 1995.
[253] Labelle, J., and Y.N. Yeh, *The relation between Burnside rings and combinatorial species*, J. of combinatorial Theory, series A **50** (1989), 269–284.
[254] Lam, T.Y., *Young diagrams*, *Schur functions*, *the Gale-Ryser theorem and a conjecture of Snapper*, J. pure and applied Algebra **10** (1977), 81–94.
[255] Lang, Serge, *Algebra*, Addison-Wesley, 1965. Chapter VIII, exercises 21–26 (pp 233ff); Chapter XII, exercises 8–10 (pp. 313ff).
[256] Langer, Andreas and Thomas Zink, *De Rham-Witt cohomology for a proper and smooth morphism*, (2003).
[257] Lasalle, Michel, *Une q-specialisation pour les fonctions symétriques monomiales*, Adv. Math. **162** (2001), 217–242.
[258] Lazard, Michel, *Bemerkungen zur Theorie der bewerteten Körper und Ringe*, Math. Nachr. **12** (1954), 67–73.
[260] Lenart, Cristian., *Formal group-theoretic generalizations of the necklace algebra, including a q-deformation*, Journal of Algebra **199** (1998), 703–732.
[261] Lenart, Cristian, *Necklace algebras and Witt vectors associated with formal group laws. Extended abstract*, Dept. Math., University of Manchester, UK.
[262] Lenart, Cristian, *Symmetric functions, formal group laws, and Lazard's theorem*, Journal of Algebra **134** (1998), 219–239.
[263] Lenstra, Hendrik, *Construction of the ring of Witt vectors*, Univ. of California at Berkeley, California, USA, 2002. Notes by John Voigt.
[264] Li, Changchun, *The forms of the Witt group schemes*, J. of Algebra **186** (1996), 182–206.
[265] Littlewood, D.E., *The inner plethysm of S-functions*, Canadian Journal of Mathematics **10** (1958), 1–16.
[266] Littlewood, D.E., *Products and plethysms of characters with orthogonal*, *symplectic and symmetric groups*, Canadian Journal of Mathematics **10** (1958), 17–32.
[267] Littlewood, D.E., *The theory of group characters*, Oxford University Press, 1950.
[268] Liulevicius, Arunas, *Arrows. Symmetries and representation rings*, J. pure and applied Algebra **19** (1980), 259–273.
[269] Liulevicius, Arunas, *Representation rings of the symmetric groups - a Hopf algebra approach*, Mat. Inst., Aarhus Univ., Denmark, 1976.
[270] Loday, Jean-Louis, *Cyclic homology*, Springer, 1992.
[271] Lorenz, Falko, *Einführung in die Algebra II*, Spektrum akademischer Verlag, 1997.

[272] Lorenz, Falko and Peter Roquette, *The theorem of Grunwald-Wang in the setting of valuation theory*, in F.-V. Kuhlmann, S. Kuhlmann and M. Marshall, eds., *Valuation theory and its applications. Volume II*, Amer. math. Soc., 2002, 175–212. Free electronic version: <www.rzuser.uni-heidelberg.de/∼ci3>.

[273] Lothaire, M., *Algebraic combinatorics on words*, Cambridge Univ. Press, 2002.

[274] Lubin, Jonathan and John Tate, *Formal moduli for one-parameter formal Lie groups*, Bull. Soc. math. de France **94** (1966), 49–59.

[275] Lück, Wolfgang, *The Burnside ring and equivariant stable cohomotopy for infinite groups*, Fachbereich Math., Univ. Münster, Germany, 2006. arXiv:AT/0504051;

[276] Luque, Jean-Gabriel and Jean-Yves Thibon, *Noncommutative symmetric functions associated with a code*, *Lazard elimination*, *and Witt vectors*, Discrete mathematics and theoretical Computer Science **9** (2007), 59–72.

[277] Mac Lane, Saunders, *Subfields and automorphism groups of p-adic fields*, Ann. of Math. **40** (1939), 423–442.

[278] Mac Lane, Saunders and Friedrich Karl Schmidt, *The generation of inseparable fields*, Proc. national Acad. Sci. USA **27** (1941), 583–587.

[279] Macdonald, I.G., *A new class of symmetric functions*, *Publ. IRMA Strassbourg, Actes 20-e Séminaire Lotharingien*, IRMA, 1988, 131–171. First edition 1979.

[280] Macdonald, I.G., *Polynomial functors and wreath products*, J. pure and applied Algebra **18** (1980), 173–204.

[281] Macdonald, I.G., *Symmetric functions and Hall polynomials*, Oxford Univ. Press, 1995. First edition 1979.

[282] Malvenuto, Claudia and Christophe Reutenauer, *Plethysm and conjugation of quasi-symmetric functions*, Discrete Math. **193** (1998), 225–233.

[283] Mammone, P and A Merkurjev, *On the corestriction of p^n-symbol*, Israel J. of Mathematics **76** (1991), 73–80.

[284] Manes, Ernie, *Monads of sets*, in M. Hazewinkel, ed., *Handbook of algebra*. *Volume 3*, North Holland (imprint of Elsevier), 2003, 67–153.

[285] Matignon, M., *Abelian p-groups of type* $(p, \ldots, p)$ *and p-adic open disks*, Manuscripta Mathematica **99** (1999), 93–109.

[286] McNinch, G.J., *Faithful representations of SL_2 over truncated Witt vectors*, Journal of Algebra **265** (2003), 606–618.

[287] Mendez, M. A., *The umbral calculus of symmetric functions*, Advances in Mathematics **124** (1996), 207–271.

[288] Mendez, Miguel, *Multisets and the combinatorics of symmetric functions*, Adv. in Math. **102** (1993), 95–125.

[289] Messing, William and Thomas Zink, *De Jong's theorem on homomorphisms of p-divisible groups*, (2002).

[290] Metropolis, Nick and Gian-Carlo Rota, *The cyclotomic identity, Combinatorics and algebra*, 1984, 19–27. Also: [252], 390–396.

[291] Metropolis, Nick and Gian-Carlo Rota, *Witt vectors and the algebra of necklaces*, Adv. Math **50** (1983), 95–125. Also: [252], 359–389.

[292] Metropolis, Nick, M.L. Stein and P.R. Stein, *On finite limit sets for transformations on the unit interval*, J. combinatorial Theory **15** (1973), 25–44.

[293] Michaelis, Walter, *Coassociative coalgebras*, in M. Hazewinkel, ed., *Handbook of algebra*. *Volume 3*, North Holland (imprint of Elsevier), 2003, 587–788.

[294] Miller, Robert L., *Necklaces, symmetries and self-reciprocal polynomials*, Discrete Mathematics **22** (1978), 25–33.

[295] Milne, J.S, *Étale cohomology*, Princeton Univ. Press, 1980.

[296] Milnor, John and Dale Husemoller, *Symmetric bilinear forms*, Springer, 1973.

[297] Molev, Alexander, *Noncommutative symmetric functions and Laplace operators for classical Lie algebras*, Letters in math. Physics **35** (1995), 135–143.

[298] Moreau, C., *Sur les permutations ciculaires distincts*, Nouv. Ann. Math. **11** (1872), 309–314.

[299] Moree, Pieter, *The formal series Witt transform*, Max Planck Inst., Bonn, Germany, 2005.
[300] Morris, I., and C.B. Wensley, *Adams operations and lambda-operations in beta-rings*, Discrete Mathematics **50** (1984), 253–270.
[301] Morris, I., and C.D. Wensley, *Lambda-operations in beta-rings*, Proc. Cambridge phil. Soc. **121** (1997), 247–267.
[302] Morris, Robert A., *Derivations of Witt vectors with application to K_2 of truncated polynomial rings and Laurent series*, J. pure and applied Algebra **18** (1980), 91–96.
[303] Mramor, Neza, *Splitting necklaces*, Obz. mat Fiz **40** (1993), 75–80.
[304] Nagata, Masayoshi, *Local rings*, Wiley, 1962.
[305] Nakamura, Tetsuo, *On torsion points of formal groups over a ring of Witt vectors*, Math. Zeitschrift **193** (1986), 397–404.
[306] Nava Z., Oscar A., *On the combinatorics of plethysm*, J. of combinatorial Theory, series A **46** (1987), 212–251.
[307] Newman, Kenneth, *Constructing sequences of divided powers*, Proc. Amer. math. Soc. **31** (1972), 32–38.
[308] Newman, Kenneth, *A realization of the additive Witt group*, Proc. Amer. math. Soc. **76** (1979), 39–42.
[309] Newman, Kenneth, *Tensor products of Witt Hopf algebras*, Comm. in Algebra **6** (1976), 761–773.
[310] Nitsuma, Yasuhiro, *On the extensions of W_n by $G^{(m)}$ over a $\mathbf{Z}_{(p)}$-algebra*, Tsukuba J. of Math. **29** (2005), 437–470.
[311] Niziol, W., *Duality in the cohomology of crystalline local systems*, Compositio mathematica **109** (1997), 67–97.
[312] Ochoa, Gustavo, *Outer plethysm, Burnside rings and beta-rings*, J. pure and appl. Algebra **55** (1988), 173–195.
[313] Ochoa, S., *A complete description of the outer plethysm in $R(S)$*, Rev. Acad. cienciae Zaragoza **42** (1987), 119–122.
[314] Oh, Young-Tak, *Classification of the ring of Witt vectors and the necklace ring associated with the formal group law X+Y-qXY*, Journal of Algebra **310** (2007), 325–350.
[315] Oh, Young-Tak, *Generalized Burnside-Grothendieck ring functor and aperiodic ring functor associated with profinite groups*, Journal of Algebra **291** (2005), 607–648.
[316] Oh, Young-Tak, *Necklace rings and logarithmic functions*, Advances in Mathematics **205** (2006), 434–486.
[317] Oh, Young-Tak, *Nested Witt vectors and their q-deformation*, Journal of Algebra **309** (2007), 683–710.
[318] Oh, Young-Tak, *R-analogue of the Burnside ring of profinite groups and free Lie algebras*, Advances in Mathematics **190** (2005), 1–46. Corrigendum: ibid. **192** (2005), 226–227.
[319] Oh, Young-Tak, *q-deformation of Witt-Burnside rings*, 2006. arXiv:math/0411353v3 [math.RA]. Version 2/3 of this preprint differs substantially from version 1.
[320] Pajitnov, A. V., *Closed orbits of gradient flows and logarithms of non-Abelian Witt vectors*, K-Theory **21** (2000), 301–324.
[321] Pajitnov, A. V. and A. A. Ranicki, *The Whitehead group of the Novikov rings*, K-Theory **21** (2000), 325–365.
[322] Parmenter, M. M. and E. Spiegel, *Algebraic properties of necklace rings*, Communications in Algebra **26** (1998), 1625–1632.
[323] Patras, F., *La décomposition en poids des algèbres de Hopf*, Ann. Inst. Fourier **43** (1993), 1067–1087.
[324] Patras, Frédéric, *Lambda rings*, in M. Hazewinkel, ed., *Handbook of algebra. Volume 3,* North Holland (imprint of Elsevier), 2005, 961–986.
[325] Patras, Frédéric, *Lambda-anneaux non commutatifs*, Comm. Algebra **23** (1995), 2067–2078.
[326] Pink, Richard, *Finite group schemes*, ETH Zürich, Switzerland, 2005.
[327] Plyushchay, Mikhail, S., *Deformed Heisenberg algebra with reflection*, 1997. arXiv:hep-th/901091v1.

[328] Pop, Horia, *Noncommutative p-adic rings and Witt vectors with coefficients in a separable algebra*, J. pure and applied Algebra **48** (1987), 271–279.
[329] Prasad, Lakshman and S.S. Iyengar, *An asymptotic equality for the number of necklaces in a shuffle exchange network*, Theoretical Comp. Sci. **102** (1992), 355–365.
[330] Puri, Y. and T. Ward, *Arithmetic and growth of periodic orbits*, J. of integer Sequences **4** (2001), Article 01.2.1.
[331] Ravenel, Douglas, C., *Complex cobordism and stable homotopy groups of spheres*, Academic Press, 1986. Appendix 2.
[332] Reutenauer, C., *On symmetric functions related to Witt vectors and the free Lie algebra,* Advances in Mathematics **110** (1995), 234–246.
[333] Reutenauer, C., *Sur des fonctions symmétriques reliées aux vecteurs de Witt*, Comptes Rendus de l' Academie Des Sciences, Série I Mathématique **312** (1991), 487–490.
[334] Reutenauer, Christophe, *Free Lie algebras*, Oxford University Press, 1993.
[335] Reutenauer, Christophe, *Mots circulaires et polynômes irréductibles*, Ann. Sci. Math. Québec **12** (1988), 275–285.
[336] Roberts, Leslie G., *The ring of Witt vectors*, in A. V. Geramita, ed., *Curves seminar at Queen's*. Volume XI. Queen's papers pure appl. Math. 105, Queen's University, Kingston, Ontario, Canada, 1997, 3–366.
[337] Ronco, M., *Free Lie algebra and lambda-ring structure*, Bulletin of the Australian Mathematical Society **50** (1994), 373–382.
[338] Roquette, Peter, *Class field theory in characteristic p, its origin and development*, in K. Miyake, ed., *Class field theory - its centenary and prospects*, Math Soc. of Japan, 2001, 549–631. Free electronic version (of 23 July 2003): <www.rzuser.uni-heidelberg.de/∼ci3>.
[339] Roquette, Peter, *History of valuation theory. Part I*, in F.-V. Kuhlmann, S. Kuhlmann and M. Marshall, eds., *Valuation theory and its applications. Volume I,* Amer. math. Soc., 2002, 291–356. Free electronic version: <www.rzuser.uni-heidelberg.de/∼i3>.
[340] Rosas, Mercedes H. and Gian-Carlo Rota, *A combinatorial overview of the Hopf algebra of macMahon symmetric functions*, Annals of Combinatories **6** (2002), 195–207.
[341] Rota, Gian-Carlo and Joel A Stein, *Plethystic Hopf algebras*, Proc. national Acad. Sci. USA **91** (1994), 13057–13061.
[342] Roux, Bernard, *Hautes sigma-dérivations et anneaux de valuation discrète non commutatifs en charactéristique nulle*, Compt. Rend. Acad. Sci. Paris, Sér. I Math. **303** (1986), 943–946.
[343] Russell, Peter, *Continuous derivations of the ring of Witt vectors*, J. pure and applied Algebra **6** (1975), 259–263.
[344] Rymer, N. W., *Power operations on the Burnside ring*, J. London math. Soc. **15** (1977), 75–80.
[345] Ryser, H. J., *Combinatorial properties of matrices of zeros and ones*, Canadian Journal of Mathematics **9** (1957), 371–377.
[346] Saibi, M., *Upper bounds of modulo p(n) trigonometric sums*, Compositio mathematica **116** (1999), 311–319.
[347] Saïdi, Mohammed, *On the degeneration of étale* $\mathbf{Z}/p\mathbf{Z}$ *and* $\mathbf{Z}/p^2\mathbf{Z}$*-torsors in equal characteristic* $p > 0$, Hiroshima math. J. **37** (2007), 315–341.
[348] Scanlon, T., *Local Andre-Oort conjecture for the universal abelian variety*, Inventiones mathematicae **163** (2006), 191–211.
[349] Scharf, Thomas and Jean-Yves Thibon, *A Hopf algebra approach to inner plethysm*, Adv. in Math. **104** (1994), 30–58.
[350] Scharf, Thomas and Jean-Yves Thibon, *On Witt vectors and symmetric functions*, Algebra Colloquium **3** (1996), 231–238.
[351] Scharf, Thomas and Jean-Yves Thibon, *Über die Adamsoperatoren des inneren Plethysmus*, Sitz. ber. Math. -Nat. wiss. KL., Akad. Gem.nütz. Wiss. Erfurt **4** (1992), 125–138.
[352] Schedler, Travis, *A Hopf algebra quantizing a necklace Lie algebra canonically associated to a quiver*, 2004. arXiv:math/0406200v2 [math.QA].
[353] Schmalz, Bernd, *Verwendung von Untergruppenleitern zur Bestimmung von Doppelnebenklassen*, Bayreuther Mathematische Schriften **31** (1990), 109–143.

[354] Schmid, Hermann Ludwig, *Zur Arithmetik der Zyklischen p-Körper*, J. reine und angew. Math. (Crelle) **176** (1937), 161–167.
[355] Schmid, Hermann Ludwig, *Zyklische algebraische Funktionenkörper vom Grade p^n über endlichen Konstantenkörper der Charakteristik p*, J. reine und angew. Math. (Crelle) **175** (1936), 108–123.
[356] Schmid, Hermann Ludwig and Ernst Witt, *Unverzweigte abelsche Körper vom Exponenten p^n über einem algebraischen Funktionenkörper der Charakteristik p*, J. reine und angew. Math. (Crelle) **176** (1937), 168–173.
[357] Schmidt, Friedrich Karl, *Die Theorie der Klassenkörper über einem Körper algebraischer Funktionen in einer Unbestimmten und mit endlichem Koeffizientenbereich*, Sitz.-Ber. phys. med. Soz. **62** (1931), 267–284.
[358] Schoeller, Colette, *Groupes affines, commutatifs, unipotents sur un corps non parfait*, Bull. de la Soc. math. de France **100** (1972), 241–300.
[359] Schröer, S., *Some Calabi-Yau threefolds with obstructed deformations over the Witt vectors*, Compositio mathematica **140** (2004), 1579–1592.
[360] Schröer, Stefan, *The T^1-lifting theorem in positive characteristic*, J. algebraic Geometry **12** (2003), 699–714.
[361] Schur, Issai, *Über eine Klasse von Matrizen die such eine gegebenen Matrix zuordnen lassen. Dissertation, Berlin*, 1901. Gesammelte Abhandlungen **4**, 1–72.
[362] Sekiguchi, Koji, *The Lubin-Tate theory for formal power series fields with finite coefficient fields*, J. of Number Theory **18** (1984), 360–370.
[363] Sekiguchi, T. and N. Suwa, *A note on extensions of algebraic and formal groups III*, Tôhoku mathematical Journal **49** (1997), 241–257.
[364] Sekiguchi, Tsutom and Noriyuki Suwa, *A note on extensions of algebraic and formal groups, V*, Japan J. of Math. **29** (2003), 221–284.
[365] Sekiguchi, Tsutomu, *On the deformations of Witt groups to tori*, *II*, J. of Algebra **138** (1991), 273–297.
[366] Sekiguchi, Tsutomu and Noriyuki Suwa, *Théories de Kummer-Artin-Schreier-Witt*, Comp. Rendus Acad. Sci. Paris, Série I Math. **319** (1994), 105–110.
[367] Serre, J.-P., *Corps locaux*, Hermann, 1968. Chap. II, \$6.
[368] Shay, P. Brian, *Bipolynomial Hopf algebras H*(**BSU**; **Z**) et al.*, J. pure and applied Algebra **9** (1977), 163–168.
[369] Shay, P. Brian, *Erratum: Representations for p-typical curves*, J. of Algebra **51** (1978), 326–334. This article replaces the earlier one in J. of Algebra **45**.
[370] Siebeneicher, Christian, *Lambda-ring Strukturen auf dem Burnsidering der Permetations-darstellungen einer endlichen Gruppe*, Math. Zeitschrift **146** (1976), 223–238.
[371] Silverman, J. H., N. P. Smart and F. Vercauteren, *An algebraic approach to NTRU (q=2(n)) via Witt vectors and overdetermined systems of nonlinear equations, in Security in Communication Networks*, Springer-Verlag Berlin, Berlin, 2005, 278–293.
[372] Solomon, Louis, *The Burnside algebra of a finite group*, J. of combinatorial Theory **2** (1967), 603–615.
[373] Solomon, Louis, *A Mackey formula in the group ring of a Coxeter group*, J. of Algebra **41** (1976), 255–268.
[374] Stienstra, Jan, Marius van der Put and Bert van der Marel, *On p-adic monodromy*, Math. Zeitschrift **208** (1991), 309–325.
[375] Strehl, V., *Cycle counting for isomorphism types of endofunctions*, Bayreuther math. Schriften **40** (1992), 153–167.
[376] Sullivan, John Brendan, *Products of Wit groups*, Illinois Journal of Mathematics **19** (1975), 27–32.
[377] Suwa, Noriyuki, *Hodge-Witt cohomology of complete intersections*, J. math. Soc. of Japan **45** (1993), 295–300.
[378] Takeuchi, Mitsuhiro, *On the structure of commutative affine group schemes over a nonperfect filed*, Manuscripta mathematica **16** (1975), 101–136.

[379] Tall, D. O. and G. C. Wraith, *Representable functors and operations on rings*, Proc. London math. Society **20** (1970), 619–643.
[380] Teichmüller, Oswald, *Diskret bewertete perfekte Körper mit unvollkommenem Restklassenkörper*, J. reine und angew. Math. (Crelle) **176** (1937), 141–156.
[381] Teichmüller, Oswald, *p-Algebren*, Deutsche Math. **1** (1936), 362–388.
[382] Teichmüller, Oswald, *Über die Struktur diskret bewerteter perfekter Körper*, Nachr. Ges. Wiss. Göttingen N.F. **1** (1936), 151–161.
[383] Terakawa, H., *On the Kawamata-Viehweg vanishing theorem for a surface in positive characteristic*, Archiv der Mathematik **71** (1998), 370–375.
[384] Thibon, J.-Y and B.-C.-V. Ung, *Quantum quasisymmetric functions and Hecke algebras*, J. Phys. A: math. gen. **29** (1996), 7337–7348.
[385] Thibon, Jean-Yves, *The inner plethysm of symmetric functions and some of its applications*, Bayreuther math. Schriften **40** (1992), 177–201.
[386] Thibon, Jean-Yves, *Lectures on noncommutative symmetric functions*, in J. R. Stembridge, J.-Y. Thibon and M. A. A. v. Leeuwen, eds., *Interaction of combinatorics and representation theory*, Math. Soc. of Japan, 2001, 39–94.
[387] Thomas, Laura, *Ramification groups in Artin-Schreier-Witt extensions*, J. des théorie des nombres de Bordeaux **17** (2005), 689–720.
[388] tom Dieck, Tammo, *The Artin-Hasse logarithm for lambda-rings*, in G. Carlsson, R. L. Cohen, H. R. Miller and D. C. Ravenel, eds., *Algebraic topology. Proceedings of an international conference, Arcata, California*, *1986*, Springer, 1989, 409–415.
[389] tom Dieck, Tammo, *Der Artin-Hasse-Logarithmus für lambda-Ringe*, Math. Gottingensis, Schriftenr. Sonderforschungsbereich Geom. Anal. **47** (1986).
[390] tom Dieck, Tammo, *Transformation groups*, Walter de Gruyter, 1987.
[391] tom Dieck, Tammo, *Tranformation groups and representation theory*, Springer, 1979.
[392] tom Dieck, Tammo and Ted Petrie, *Geometric modules over the Burnside ring*, Inventiones mathematicae **47** (1978), 273–287.
[393] Tyc, A, *Witt groups of commutative formal groups*, Bull. de l'Acad. Pol. des Sci. Sér. des Sci. Math., Nat. et Phys. **28** (1975), 1233–1340.
[394] Uehara, Hiroshi, Essam Abotteen and Man-Wai Lee, *Outer plethysms and lambda-rings*, Arch. Math. **46** (1986), 216–224.
[395] Uehara, Hiroshi and Robert A DiVall, *Hopf algebra of class functions and inner plethysms*, Hiroshima math. J. **12** (1982), 225–244.
[396] Vallejo, Ernesto, *The free beta-ring on one generator*, J. pure and applied Algebra **56** (1993), 95–108.
[397] Vallejo, Ernesto, *Polynomial operations from Burnside rings to representation functors*, J. pure and appl. Algebra **65** (1990), 163–190.
[398] van den Dries, L., *On the elementary theory of rings of Witt vectors with a multiplicative set of representatives for the residue held*, Manuscripta Mathematica **98** (1999), 133–137.
[399] Varadarajan, K., *Aperiodic rings, necklace rings and Witt vectors - II*, Proc. Indian Acad. Sci. **99** (1989), 7–15.
[400] Varadarajan, K., *Verschiebung and Frobenius operators*, Proc Indian Acad Sci (Math Sci) **100** (1990), 37–43. Erratum: ibid, p. 303.
[401] Varadarajan, K. and K Wehrhahn, *Aperiodic rings, necklace rings and Witt vectors*, Adv. Math. **81** (1990), 1–29.
[402] Veldsman, S., *Chain conditions on convolution rings*, Communications in Algebra **35** (2007), 371–388.
[403] Veldsman, S., *Convolution rings*, Algebra Colloquium **13** (2006), 211–238.
[404] Veldsman, S., *Necklace rings and their radicals*, Mathematica Pannonica **12** (2001), 269–289.
[405] Veldsman, S., *The radical theory of convolution rings*, Bull. Acad. de Stiinte a Republicii Moldava. Mathematica **1** (2004), 98–115.
[406] Vidal, Robert, *Anneaux de valuation discrète complets non commutatifs*, Transactions of the American mathematical Society **267** (1981), 65–81.

[407] Voloch, J. F. and J. L. Walker, *Homogeneous weights and exponential sums*, Finite Fields and their Applications **9** (2003), 310–321.
[408] Vostokov, S. V. and A. N. Gurevich, *A relationship between the Hilbert and the Witt symbols*, J of mathematical Sciences **89** (1998), 1108–1112.
[409] Webb, Peter, *A guide to Mackey functors*, in M. Hazewinkel, ed., *Handbook of Algebra. Volume 2*, North Holland (imprint of Elsevier), 2000, 805–836.
[410] Weinstein, Felix V., *Filtering bases: a tool to compute cohomologies of abstract subalgebras of the Witt algebra*, in D. Fuchs, ed., *Unconventional Lie algebras*, Amer. math. Soc., 1993, 155–216.
[411] Weston, T., *Unobstructed modular deformation problems*, American Journal of Mathematics **126** (2004), 1237–1252.
[412] Whaples, G., *Generalized local class field theory III: second form of the existence theorem. Structure of analytic groups*, Duke math. J. (1954), 575–581.
[413] Widiger, A, *Nichtprime Cohen-Ringe*, Period. math. Hungar. **12** (1981), 141–160.
[414] Wilkerson, Clarence, *Lambda rings, binomial domains, and vector bundles over* $\boldsymbol{CP}^{\infty}$, Comm. Algebra **10** (1982), pp. 311–328.
[415] Witt, Ernst, *Der Existenzsatz für abelsche Funktionenkörper*, J reine angew. Math. (= Crelle) **173** (1935), 43–51. Also in [238], pp. 69–77.
[416] Witt, Ernst, *Konstruktion von galoisschen Körpern der Charakteristik p zu vorgegebenen Gruppe der Ordnung* p^n, J reine angew. Math. (= Crelle) **174** (1936), 237–245. Also in [238], pp. 120–128.
[417] Witt, Ernst, *p-Algebren und Pfaffsche Formen*, Abh. math. Sem. Hamburg **22** (1958), 308–315. Also in [238], pp. 133–141.
[418] Witt, Ernst, *Treue Darstellung Liescher Ringe*, J. reine und angew. Math. (Crelle) **177** (1937), 152–160.
[419] Witt, Ernst, *Vektorkalkül und Endomorphismen der Einspotenzreihengruppe*, in I. Kersten, ed., *Ernst Witt*, Springer, 1969, 157–164. Originally unpublished. Notes for a colloquium Lecture in Hamburg, June 1964; mailed by Witt to Peter Gabriel, Sept. 1969.
[420] Witt, Ernst, *Zyklische Körper und Algebren der Charakteristik p von Grad* p^n. *Struktur diskret bewerteter Körper mit vollkommenem Restklassenköper der Charakteristik p*, J reine angew. Math. **176** (1937), 126–140. Also in [238], pp. 142–156.
[421] Yau, Donald, *Cohomology of lambda-rings*, J. of Algebra **284** (2005), 37–51.
[422] Yau, Donald, *Deformation theory of modules*, Comm. Algebra **33** (2005), 2351–2359.
[423] Yau, Donald, *Moduli space of filtered lambda-ring structures over a filtered ring*, 2003.
[424] Yoshida, Tomoyuki, *The generalized Burnside ring of a finite group*, Hokkaido math. J. **19** (1990), 509–574.
[425] Zelevinsky, A V, *Representations of the finite classical groups*, Springer, 1980.
[426] Zhang, Hechun, *Witt-like algebras and their q-analogues*, Linear and multilinear Algebra **42** (1997), 221–231.

Additional references.

[427] Anderson, Greg W., *t-Motives*, Duke math. J. **53** (1986), 457–502.
[428] Azymaya, Goro, *On maximally central algebras*, Nagoya math. J. **2** (1951), 119–150.
[429] Baker, T. H., *Symmetric function products and plethysms and the boson-fermion correspondence*, J. Physics A: math. gen. **28** (1995), 589–606.
[430] Baldassari, Francesco, *Una funzione intera p-adica a valori interi*, Ann. della Scuola norm. sup. di Pisa, Classe di Scienze (4) **2** (1975), 321–331.
[431] Berthelot, Pierre, *Théorie de Dieudonné sur un anneau de valuation parfait*, Ann. scient. de l'École norm. sup. (4) **13** (1980), 225–268.
[432] Berthelot, Pierre, *Cohomologie rigide et théorie des D-modules*, in F. Baldassarri, S. Bosch, B. Dwork, eds., *p-adic analysis*, Springer, 1990, 80–124.
[433] Blass, Andreas, *Natural endomorphisms of Burnside* rings, Trans. Amer. math. Soc. **253** (1979), 121–137.

[434] Bleher, Frouke and Ted Chinburg, *Operations on ring structures preserved by normalized automorphisms of group rings*, J. of Algebra **215** (1999), 531–542.
[435] Candilera, Maurizio and Valentino Cristante, *Witt realization of p-adic Barsotti-Tate groups*, in Valentino Cristante and William Messing, eds., *Barsotti symposium in algebraic geometry*, Acad. Press, 1994, 65–123.
[436] Candilera, Maurizio and Valentino Cristane, *Periods and duality of p-adic Barsotti-Tate groups*, Ann. della Scuola norm. sup. di Pisa, Classe di Scienze (4) **22** (1995), 545–593.
[437] Candilera, Maurizio, *Vettori di Witt ed anelli di Fontaine*, Dip. Mat. Univ. Padova, preprint.
[438] Candilera, Maurizio, *A note about p-adic logarithms*, Dip. Mat. Univ. Padova, preprint.
[439] Cline, Edward, Brian Parshall and Leonard Scott, *Cohomology, hyperalgebras, and representations*, J. of Algebra **63** (1980), 98–123.
[440] Cristante, Valentino, *Theta functions and Barsotti-Tate groups*, Ann. della Scuola norm. sup. di Pisa, Classe di Scienze (4) **7** (1980), 181–215.
[441] Cristante, Valentino, *Witt realization of p-adic Barsotti-Tate groups; some applications*, in Valentino Cristante and William Messing, eds., *Barsotti symposium in algebraic geometry*, Acad. Press, 1994, 205–216.
[442] Devinatz, Ethan, S. and Michael J. Hopkins, *The action of the Morava stabilizer group on the Lubin-Tate Moduli space of lifts*, Amer. J. of Math. **117** (1995), 669–710.
[443] Diarra, Bertin, *Point de vue séries formelles des nombres p-adiques*, preprint, Lab. de Math. pures, Univ. Blaise Pascal, Clairmont-Ferrand, France.
[444] Drinfel'd, V. G., *Elliptic modules*, Math. USSR Sbornik **23** (1974), 561–592.
[445] Ekedahl, Torsten, *Duality for the de Rham-Witt complex*, preprint, Dep. Math., Chalmers Univ. Techn. 1982.
[446] Fontaine, Jean-Marc, *Le corps des périodes p-adiques*, in Jean-Marc Fontaine, ed., Périodes p-adiques, Soc. math. de France, 1994, 59–111. Avec un appendice par Pierre Colmez.
[447] Fontaine, Jean-Marc, *Cohomologie de de Rham, cohomologie cristalline et réprésentations p-adiques*, in *Algebraic geometry Tokyo-Kyoto*, Springer 1983, 86–108.
[448] Fontaine, Jean-Marc, *Sur certains types de réprésentations p-adiques du groupe de Galois d'un corps local; construction d'un anneau de Barsotti-Tate*, Annals of Math. **115** (1982), 520–577.
[449] Greenberg, Marvin J., *Algebraic rings*, Trans. Amer. math. Soc. **111** (1964), 472–481.
[450] Greenberg, Marvin J., *Schemata over local rings*, Ann. of Math. (2) **73** (1961), 624–648.
[451] Hesselholt, Lars and Ib Madsen, *on the de Rham-Witt comples in mixed characteristic*, preprint, Univ. of Åarhus, 2002.
[452] Hesselholt, Lars and Ib Madsen, *Cyclic polytopes and the K-theory of truncated polynomial algebras*, Inventiones math. **130** (1997), 73–97.
[453] Izhboldin, O. T., *On the torsion subgroup of Milnor's K-groups*, Soviet Math. Dokl. **35** (1987), 493–495.
[454] Kirkman, E, C. Procesi and L. Small, *A q-analog for the Vitasoro algebra*, Comm. in Algebra **22** (1994), 3755–3774.
[455] Kurihara, Masato, *Abelian extensions of an absolutely unramified local field with general residue field*, Inventiones math. **93** (1988), 451–480.
[456] Langer, Andreas and Thomas Zink, *De Rham - Witt cohomology and displays*, Documenta math. **12** (2007), 147–191.
[457] Lazarev, A., *Deformations of formal groups and stable homotopy theory*, Topology **36** (1987), 1317–1331.
[458] Moore, John C. and Larry Smith, *Hopf algebras and multiplicative fibrations I*, Amer. J. of Math. **90** (1968), 752–780.
[459] Moore, John C. and Larry Smith, *Hopf algebras and multiplicative fibrations II*, Amer. J. of Math. **90** (1968), 1113–1150.
[460] Oort, Frans, *Lifting algebraic curves*, *Abelian varieties, and their endomorphisms to characteristic zero*, in Spencer J. Bloch, ed., *Algebraic geometry Bowdoin 1985*, Amer. math. Soc., 1987, 165–195.

[461] Oort, Frans and David Mumford, *Deformations and liftings of finite, commutative group schemes*, Inventiones math. **5** (1968), 317–334.
[462] Pajitnov, A. V., *Closed orbits of gradient flows and logarithms of non-Abelian Witt vectors*, 2000. arXiv:math.DG/9908010.
[463] Schneider, Peter, *Arithmetic of formal groups and applications I: universal norm subgroups*, Inventiones math. **87** (1987), 587–602.
[464] Schoeller, Colette, *F-H-algèbres sur un corps*, C. R. Acad. Sci. Paris Série A **265** (1967), 655–658.
[465] Schoeller, Colette, *Étude de la catégorie des algèbres de Hopf commutatives connexes sur un corps*, Manuscripta math. **3** (1970), 133–155.
[466] Sekiguchi, T., Frans Oort and N. Suwa, *On the deformation of Artin - Schreier to Kummer*, Ann. Sci. de l'École norm. sup. (4) **22** (1989), 345–375.
[467] Serre, Jean-Pierre, *Exemples de variétés projectives en charactéristique p non relevables en charactéristique zéro*, Proc. national Acad. Sci. USA **47** (1961), 108–109.
[468] Yau, Donald, *Unstable K-cohomology algebra is filtered λ-ring*, Int. J. of Math. and math. Sci.**10** (2003), 593–405.
[469] Yau, Donald, *On λ-rings and topological realization,* Int. J. of Math. and math. Sci.**13** (2006), 1–21.
[470] Zhang, Qifan, *Polynomial functions and permutation polynomials over some finite commutative rings*, J. Number Theory **105** (2004), 192–202.
[471] Zhang, Qifan, *Witt rings and permutation polynomials*, Algebra Colloquium **12** (2005), 161–169.

Crystal Graphs and the Combinatorics of Young Tableaux

Jae-Hoon Kwon

Department of Mathematics, University of Seoul, Seoul 130-743, Korea
E-mail: jhkwon@uos.ac.kr

Contents

HANDBOOK OF ALGEBRA, VOL. 6
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(08)00208-8

1. Introduction

The quantum group $U_q(\mathfrak{g})$ of a symmetrizable Kac–Moody algebra $\mathfrak{g}$ [17] is a q-analogue of the universal enveloping algebra $U(\mathfrak{g})$, which was introduced independently by Drinfel'd [7] and Jimbo [15], and it has been one of the most important objects in representation theory in connection with many areas of mathematics and mathematical physics.

In [21, 22], Kashiwara introduced a *crystal base* of an integrable representation of $U_q(\mathfrak{g})$. A crystal base is usually understood as a basis at $q = 0$, or a crystallized basis because the parameter q in the quantum group $U_q(\mathfrak{g})$ corresponds to the absolute temperature in certain exactly solvable lattice models in statistical mechanics. Remarkably, to a crystal base, we can associate a unique colored oriented graph called a *crystal graph*, which reflects the combinatorial structure of a given integrable representation. Therefore, with crystal base theory, many problems in representation theory are reduced to problems in combinatorics, and also certain combinatorial results naturally obtain representation theoretical meanings (interpretations).

On the other hand, the notion of *Young tableaux* together with its combinatorics plays an essential role in the representation theory of the general linear Lie algebra $\mathfrak{gl}_n$ or the symmetric group S_k. There are many beautiful algorithms by which we understand the combinatorics of Young tableaux, based on the Schensted bumping rule, the Schützenberger sliding rule, and the Knuth equivalence. Among them, the *Robinson–Schensted–Knuth correspondence* and the *Littlewood–Richardson rule* are the most important ones closely related to representation theory (see, for example, [10,41,43]).

The purpose of this survey article is to give an exposition on the relations between the classical combinatorics of Young tableaux and the crystal graphs of integrable $U_q(\mathfrak{gl}_n)$-modules, as an introduction to crystal base theory for those who are not familiar with this area (see also [5] that deals with similar topics, but is written from a different point of view). This article is organized as follows. In Section 2, we recall the definition of the quantum group $U_q(\mathfrak{gl}_n)$, its integrable representations, and their crystal bases. In Section 3, we review the Kashiwara and Nakashima description of the crystal graphs of integrable highest weight $U_q(\mathfrak{gl}_n)$-modules [26]. In Section 4, we explain the Robinson–Schensted correspondence and the Littlewood–Richardson rule in the context of crystal graphs. Our explanation is based on the crucial fact that the Knuth equivalence on words is a special case of the crystal equivalence. In Section 5, we discuss the Robinson–Schensted–Knuth correspondence in a similar vein, and give another proof of it using the notion of crystals. We also give some applications to combinatorics and representation theory.

2. Quantum group and crystal base

2.1. *Quantum group*

For $n \geq 2$, let $P_n = \bigoplus_{1 \leq i \leq n} \mathbf{Z}\epsilon_i$ be the free abelian group with a basis $\{\epsilon_1, \ldots, \epsilon_n\}$[1], called the *weight lattice*. There is a natural symmetric bilinear form $(\ ,\ )$ on P_n given

[1] $\epsilon_1, \ldots, \epsilon_n$ are formal symbols.

by $(\epsilon_i, \epsilon_j) = \delta_{ij}$ for $1 \le i, j \le n$. Let $\langle\ ,\ \rangle$ denote the canonical pairing on $P_n^\vee \times P_n$, where $P_n^\vee = \mathrm{Hom}(P_n, \mathbf{Z})$ is the *dual weight lattice*. Set $I_n = \{1, \dots, n-1\}$. Then, for $i \in I_n$ the *simple root* α_i is given by $\alpha_i = \epsilon_i - \epsilon_{i+1}$, and the *simple coroot* h_i is defined to be the unique element in $P_n^\vee$ such that $\langle h_i, \lambda\rangle = (\alpha_i, \lambda)$ for all $\lambda \in P_n$. Note that $\big(\langle h_j, \alpha_i\rangle\big)_{i,j\in I_n} = \big((\alpha_i, \alpha_j)\big)_{i,j\in I_n}$ is the Cartan matrix of A_{n-1} type (see [17]).

Let $\mathfrak{gl}_n$ be the general linear Lie algebra over a field K of characteristic 0 and let q be an indeterminate. We define $U_q(\mathfrak{gl}_n)$ to be the associative algebra over $K(q)$ with 1 generated by q^h, e_i, and f_i ($h \in P_n^\vee$ and $i \in I_n$) subject to the following relations;

$$
\begin{aligned}
&q^0 = 1, \quad q^{h+h} = q^h q^{h'},\\
&q^h e_i q^{-h} = q^{\langle h,\alpha_i\rangle} e_i, \quad q^h f_i q^{-h} = q^{-\langle h,\alpha_i\rangle} f_i,\\
&e_i f_j - f_j e_i = \delta_{ij}\frac{q^{h_i} - q^{-h_i}}{q - q^{-1}},\\
&e_i e_j = e_j e_i, \quad f_i f_j = f_j f_i, \quad \text{if } |i-j| > 1,\\
&e_j e_i^2 - (q + q^{-1}) e_i e_j e_i + e_i^2 e_j = 0,\\
&f_j f_i^2 - (q + q^{-1}) f_i f_j f_i + f_i^2 f_j = 0, \quad \text{if } |i-j| = 1,
\end{aligned}
$$

for $h, h' \in P_n^\vee$ and $i, j \in I_n$. We call $U_q(\mathfrak{gl}_n)$ the *quantized universal enveloping algebra of* $\mathfrak{gl}_n$ (or simply, the *quantum group of* $\mathfrak{gl}_n$).

Also, we define $U_q(\mathfrak{sl}_n)$ to be the subalgebra of $U_q(\mathfrak{gl}_n)$ generated by e_i, f_i and $q^{\pm h_i}$ for $i \in I_n$. Note that for each $i \in I_n$, $U_q(\mathfrak{gl}_n)_i$, the subalgebra of $U_q(\mathfrak{gl}_n)$ generated by e_i, f_i, $q^{\pm h_i}$ is isomorphic to $U_q(\mathfrak{sl}_2)$.

A left $U_q(\mathfrak{gl}_n)$-module M is called *integrable* if e_i and f_i act locally nilpotently on M. Let $\mathcal{O}_{\mathrm{int}}$ be the category of integrable $U_q(\mathfrak{gl}_n)$-modules M satisfying the following conditions;

(1) M has a weight space decomposition, that is, $M = \bigoplus_{\lambda\in P_n} M_\lambda$, where

$$M_\lambda = \{\, u \in M \mid q^h u = q^{\langle h,\lambda\rangle} u \text{ for all } h \in P_n^\vee \,\},$$

and $\dim_{K(q)} M_\lambda < \infty$ for $\lambda \in P_n$ (λ is called a *weight of* M if $M_\lambda \neq \{0\}$).

(2) The weights of M are dominated by a finite number of elements in P_n, that is, there exist $\lambda_1, \dots, \lambda_r \in P_n$ such that every weight of M is contained in $\lambda_k - \sum_{i\in I_n} \mathbf{Z}_{\ge 0}\alpha_i$ for some $1 \le k \le r$.

Note that any M in $\mathcal{O}_{\mathrm{int}}$ is a direct sum of finite dimensional representations of $U_q(\mathfrak{gl}_n)_i \simeq U_q(\mathfrak{sl}_2)$ for each $i \in I_n$.

The representation theory of $U_q(\mathfrak{gl}_n)$ is almost parallel to that of its Lie algebra $\mathfrak{gl}_n$. Set $P_n^+ = \{\, \lambda \in P_n \mid \langle h_i, \lambda\rangle \ge 0 \text{ for } i \in I_n \,\}$. An element in P_n^+ is called a *dominant integral weight*. We may identify P_n with $\mathbf{Z}^n$, and hence P_n^+ with

$$\mathbf{Z}_+^n = \{\, \lambda = (\lambda_1, \dots, \lambda_n) \in \mathbf{Z}^n \mid \lambda_i \ge \lambda_{i+1} \text{ for } 1 \le i \le n-1 \,\}.$$

There is a partial ordering $\le$ on P_n such that $\mu \le \lambda$ if and only if $\lambda - \mu \in \sum_{i\in I_n} \mathbf{Z}_{\ge 0}\alpha_i$ for $\lambda, \mu \in P_n$.

For $\lambda \in P_n^+$, let $V(\lambda)$ be the $U_q(\mathfrak{gl}_n)$-module generated by v_λ[2] subject to the following relations;

$$q^h v_\lambda = q^{\langle h,\lambda\rangle} v_\lambda, \quad e_i v_\lambda = 0, \quad f_i^{1+\langle h_i,\lambda\rangle} v_\lambda = 0,$$

for $h \in P_n^\vee$ and $i \in I_n$. Then $V(\lambda)$ is an irreducible $U_q(\mathfrak{gl}_n)$-module, and any $U_q(\mathfrak{gl}_n)$-module M in $\mathcal{O}_{\text{int}}$ is completely reducible with irreducible summands that are isomorphic to $V(\lambda)$ for some $\lambda \in P_n^+$, that is, M is isomorphic to $\bigoplus_{\lambda\in P_n^+} V(\lambda)^{\oplus m_\lambda}$ for some $m_\lambda \in \mathbf{Z}_{\geq 0}$ (see [36]). We call $V(\lambda)$ the *irreducible highest weight module with highest weight* λ, and v_λ the *highest weight vector*, since for each weight μ of $V(\lambda)$, we have $\mu \leq \lambda$.

The quantum group $U_q(\mathfrak{gl}_n)$ has a Hopf algebra structure, where the comultiplication $\Delta : U_q(\mathfrak{gl}_n) \to U_q(\mathfrak{gl}_n) \otimes U_q(\mathfrak{gl}_n)$ is given by

$$\begin{aligned}
\Delta q^h &= q^h \otimes q^h, \\
\Delta e_i &= e_i \otimes q^{-h_i} + 1 \otimes e_i, \\
\Delta f_i &= f_i \otimes 1 + q^{h_i} \otimes f_i,
\end{aligned}$$

for $h \in P_n^\vee$ and $i \in I_n$. So, we can define a $U_q(\mathfrak{gl}_n)$-module structure on the tensor product of two $U_q(\mathfrak{gl}_n)$-modules. In particular, the category $\mathcal{O}_{\text{int}}$ is closed under tensor product.

Remark 2.1. For more detailed expositions on quantum groups of symmetrizable Kac–Moody algebras and their representations, see the books [3, 12, 14, 36].

2.2. *Crystal base*

Let us briefly review the crystal base theory developed by Kashiwara (see also [24] for a general review on crystal bases). For $M \in \mathcal{O}_{\text{int}}$ and $i \in I_n$, M is a direct sum of finite dimensional irreducible representations of $U_q(\mathfrak{gl}_n)_i$. Then each $u \in M_\lambda$ ($\lambda \in P_n$) can be written uniquely as

$$u = \sum_{k\geq 0, -\langle h_i,\lambda\rangle} f_i^{(k)} u_k,$$

with $e_i u_k = 0$ for all $k \geq 0$, where

$$[a] = \frac{q^a - q^{-a}}{q - q^{-1}} \ (a \geq 1), \quad [k]! = \prod_{a=1}^{k} [a], \quad f_i^{(k)} = \frac{1}{[k]!} f_i^k.$$

Then we define

$$\tilde{e}_i u = \sum_{k\geq 1} f_i^{(k-1)} u_k, \quad \tilde{f}_i u = \sum_{k\geq 0} f_i^{(k+1)} u_k.$$

[2] More precisely, $V(\lambda) = U_q(\mathfrak{gl}_n)/\left(\sum_i U_q(\mathfrak{gl}_n)e_i + \sum_h U_q(\mathfrak{gl}_n)(q^h - q^{\langle h,\lambda\rangle}) + \sum_i U_q(\mathfrak{gl}_n) f_i^{1+\langle h_i,\lambda\rangle}\right)$, and v_λ is the image of $1 \in U_q(\mathfrak{gl}_n)$ in $V(\lambda)$.

The operators $\tilde{e}_i, \tilde{f}_i : M \to M$ are called *Kashiwara operators*.

Let A be the subring of $K(q)$ that consists of rational functions regular at $q = 0$. A free A-module L is called a *crystal lattice of* M if

(1) L is an A-lattice of M, that is, $M = K(q) \otimes_A L$,
(2) $L = \bigoplus_{\lambda \in P_n} L_\lambda$, where $L_\lambda = M_\lambda \cap L$,
(3) $\tilde{e}_i L \subset L$ and $\tilde{f}_i L \subset L$ for $i \in I_n$.

Note that the Kashiwara operators also induce K-linear endomorphisms on L/qL. Now, a *crystal base of* M is a pair (L, B) satisfying the following conditions;

(1) L is a crystal lattice of M,
(2) B is a K-basis of L/qL,
(3) $B = \bigsqcup_{\lambda \in P_n} B_\lambda$, where $B_\lambda = (L_\lambda/qL_\lambda) \cap B$,
(4) $\tilde{e}_i B \subset B \cup \{0\}$ and $\tilde{f}_i B \subset B \cup \{0\}$ for $i \in I_n$.
(5) for $b, b' \in B$ and $i \in I_n$, $\tilde{f}_i b = b'$ if and only if $\tilde{e}_i b' = b$.

It is not difficult to see that $\dim_{K(q)} M_\lambda = |B_\lambda|$ for $\lambda \in P_n$. Moreover, to a crystal base (L, B) of M, we can associate an I_n-colored oriented graph, where the set of vertices is given by B and the colored arrows are determined by

$$b \xrightarrow{i} b' \quad \text{if and only if} \quad b' = \tilde{f}_i b \qquad (i \in I_n).$$

We call B together with the I_n-colored oriented graph structure the *crystal graph of* M.

THEOREM 2.2 (Existence of a crystal base [21, 22]). *For* $\lambda \in P_n^+$, *let*

$$L(\lambda) = \sum_{r \geq 0,\, i_1, \ldots, i_r \in I_n} A \tilde{f}_{i_1} \ldots \tilde{f}_{i_r} v_\lambda,$$

$$B(\lambda) = \{ \tilde{f}_{i_1} \ldots \tilde{f}_{i_r} v_\lambda \pmod{qL(\lambda)} \mid r \geq 0,\ i_1, \ldots i_r \in I_n \} \backslash \{0\}.$$

Then $(L(\lambda), B(\lambda))$ *is a crystal base of* $V(\lambda)$.

It follows directly from the theorem aforementioned that any $M \in \mathcal{O}_{\text{int}}$ has a crystal base. That is, if M is isomorphic to $\bigoplus_{\lambda \in P_n^+} V(\lambda)^{\oplus m_\lambda}$ for some $m_\lambda \in \mathbf{Z}_{\geq 0}$, then a crystal base of M can be given by $(\bigoplus_{\lambda \in P_n^+} L(\lambda)^{\oplus m_\lambda}, \bigsqcup_{\lambda \in P_n^+} B(\lambda)^{\oplus m_\lambda})$.

THEOREM 2.3 (Uniqueness of a crystal base [21, 22]). *For* $M \in \mathcal{O}_{\text{int}}$, *let* (L, B) *be a crystal base of* M. *Then there exists an isomorphism of* $U_q(\mathfrak{gl}_n)$*-modules*

$$\psi : M \longrightarrow \bigoplus_{\lambda \in P_n^+} V(\lambda)^{\oplus m_\lambda}$$

for some $m_\lambda \in \mathbf{Z}_{\geq 0}$ $(\lambda \in P_n^+)$ *such that* ψ *restricts to an* A*-linear isomorphism from* L *to* $\bigoplus_{\lambda \in P_n^+} L(\lambda)^{\oplus m_\lambda}$, *and the induced* K*-linear map* $\overline{\psi}$ *gives a bijection from* B *to* $\bigsqcup_{\lambda \in P_n^+} B(\lambda)^{\oplus m_\lambda}$.

Note that the crystal graph $B(\lambda)$ is connected, and any element in $B(\lambda)$ is generated by the highest weight element $\overline{v_\lambda} = v_\lambda + qL(\lambda) \in L(\lambda)/qL(\lambda)$, so that for any $b \in B(\lambda)$, we have $b = \tilde{f}_{i_1} \dots \tilde{f}_{i_r} \overline{v_\lambda}$ for some $i_1, \dots, i_r \in I_n$. Also, the highest weight element in $B(\lambda)$ can be characterized by the fact that $\tilde{e}_i \overline{v}_\lambda = 0$ for all $i \in I_n$.

By the uniqueness theorem, there exists a unique I_n-colored oriented graph B associated with each $M \in \mathcal{O}_{\text{int}}$, where each connected component in B is isomorphic to $B(\lambda)$ for some $\lambda \in P_n^+$ as an I_n-colored oriented graph.

EXAMPLE 2.4. Let us consider crystal bases of integrable $U_q(\mathfrak{sl}_2)$-modules. Suppose that $U_q(\mathfrak{sl}_2)$ is generated by e, f, and $q^{\pm h}$. Each dominant integral weight is determined by its value on h, which is a nonnegative integer. For $m \geq 0$, let $V(m)$ be the irreducible highest weight module with highest weight m, and v_m the highest weight vector. Then $\dim_{K(q)} V(m) = m + 1$ with a basis $\{ v_m, fv_m, \dots, f^m v_m \}$. If we put

$$L(m) = \bigoplus_{0 \leq k \leq m} A f^{(k)} v_m, \quad B(m) = \{ \overline{f^{(k)} v_m} \mid 0 \leq k \leq m \},$$

then $(L(m), B(m))$ is a crystal base of $V(m)$. The associated crystal graph of $V(m)$ is given by

$$\overline{v_m} \longrightarrow \overline{fv_m} \longrightarrow \cdots \longrightarrow \overline{f^{(m)} v_m}.$$

Note that the weight of $\overline{f^{(k)} v_m}$ is $m - 2k$ for $0 \leq k \leq m$.

For $M \in \mathcal{O}_{\text{int}}$, let (L, B) be a crystal base of M. For each $i \in I_n$, (L, B) is also a crystal base of M as a representation of $U_q(\mathfrak{gl}_n)_i \simeq U_q(\mathfrak{sl}_2)$. Hence, the connected component of each b in B, as the crystal graph of a $U_q(\mathfrak{sl}_2)$-module, is equal to the crystal graph of $V(m)$ for some $m \geq 0$, called the *i-string of b*. Set

$$\varepsilon_i(b) = \max\{ k \in \mathbf{Z}_{\geq 0} : \tilde{e}_i^k b \neq 0 \},$$
$$\varphi_i(b) = \max\{ k \in \mathbf{Z}_{\geq 0} : \tilde{f}_i^k b \neq 0 \},$$

which denote the number of i-arrows coming into b and going out of b in the i-string of b, respectively[3]. Then we have

$$\varphi_i(b) - \varepsilon_i(b) = \langle h_i, \text{wt}(b) \rangle,$$

where $\text{wt}(b)$ denotes the weight of b.

One of the most important features of a crystal base is that it has a nice behavior under tensor product, which can be stated as follows;

THEOREM 2.5 (Tensor product of crystal bases [21,22]). *For $M_i \in \mathcal{O}_{\text{int}}$, let (L_i, B_i) be the crystal base of M_i $(i = 1, 2)$. Then $(L_1 \otimes L_2, B_1 \otimes B_2)$ is a crystal base of*

[3] Note that ε_i is a function on B, and ϵ_i is a basis element of P_n.

$M_1 \otimes M_2$. *Moreover, for* $b_1 \otimes b_2 \in B_1 \otimes B_2$ *and* $i \in I_n$, *we have*

$$\tilde{e}_i(b_1 \otimes b_2) = \begin{cases} \tilde{e}_i b_1 \otimes b_2, & \text{if } \varphi_i(b_1) \geq \varepsilon_i(b_2), \\ b_1 \otimes \tilde{e}_i b_2, & \text{if } \varphi_i(b_1) < \varepsilon_i(b_2), \end{cases}$$

$$\tilde{f}_i(b_1 \otimes b_2) = \begin{cases} \tilde{f}_i b_1 \otimes b_2, & \text{if } \varphi_i(b_1) > \varepsilon_i(b_2), \\ b_1 \otimes \tilde{f}_i b_2, & \text{if } \varphi_i(b_1) \leq \varepsilon_i(b_2). \end{cases}$$

The formulas mentioned earlier for the actions of $\tilde{e}_i$ and $\tilde{f}_i$ are often called the *tensor product rule*, and this rule plays a crucial role when we construct various crystal graphs (see the next section). The tensor product rule also implies that under the same hypothesis,

$$\varepsilon_i(b_1 \otimes b_2) = \max(\varepsilon_i(b_1), \varepsilon_i(b_2) - \langle h_i, \mathrm{wt}(b_1)\rangle),$$
$$\varphi_i(b_1 \otimes b_2) = \max(\varphi_i(b_1) + \langle h_i, \mathrm{wt}(b_2)\rangle, \varphi_i(b_2)).$$

Example 2.6. Let us illustrate the tensor product rule for the crystal bases of $U_q(\mathfrak{sl}_2)$-modules. We keep the notations of the previous example. Let $B(3) = \{\, b_k = \overline{f^{(k)} v_3} \mid k = 0, 1, 2, 3 \,\}$ and $B(2) = \{\, b'_k = \overline{f^{(k)} v_2} \mid k = 0, 1, 2 \,\}$ be the crystal graphs of $V(3)$ and $V(2)$, respectively. Then the crystal graph $B(3) \otimes B(2)$ is given as follows;

$$\begin{array}{ccccccc} b_0 \otimes b'_0 & \longrightarrow & b_1 \otimes b'_0 & \longrightarrow & b_2 \otimes b'_0 & \longrightarrow & b_3 \otimes b'_0 \\ & & & & & & \downarrow \\ b_0 \otimes b'_1 & \longrightarrow & b_1 \otimes b'_1 & \longrightarrow & b_2 \otimes b'_1 & & b_3 \otimes b'_1 \\ & & & & \downarrow & & \downarrow \\ b_0 \otimes b'_2 & \longrightarrow & b_1 \otimes b'_2 & & b_2 \otimes b'_2 & & b_3 \otimes b'_2 \end{array}$$

This implies that $B(3) \otimes B(2) = B(5) \sqcup B(3) \sqcup B(1)$, and hence $V(3) \otimes V(2) \simeq V(5) \oplus V(3) \oplus V(1)$.

Remark 2.7. The results on crystal bases presented here were proved in case of the quantum group $U_q(\mathfrak{g})$ of a symmetrizable Kac–Moody algebra $\mathfrak{g}$ by Kashiwara [22]. Also, it is an important problem to give an explicit realization of the crystal graph of a $U_q(\mathfrak{g})$-module. There have been many papers on this problem: for classical Lie algebras, there are *tableaux realizations* by Kashiwara and Nakashima [26]. For affine Lie algebras, there are *path realizations* using perfect crystals [20]. For a symmetrizable Kac–Moody algebra $\mathfrak{g}$, Littelmann realized the crystal graph of an irreducible integrable highest weight module in terms of certain piecewise linear paths in the weight lattice, often called the *Littelmann path model* [34]. Recently, Nakajima introduced the *monomial realization*, where the crystal graphs are given in terms of certain monomials in commuting variables [25, 39].

3. Crystals and Young tableaux

3.1. *Crystal*

Motivated by the colored oriented graphs induced from crystal graphs of integrable $U_q(\mathfrak{gl}_n)$-modules, we may define the notion of (abstract) *crystals*.

DEFINITION 3.1. ([23, 24]) Let P_n be the weight lattice, $P_n^\vee$ the dual weight lattice, $\{\alpha_i \mid i \in I_n\}$ the set of simple roots, and $\{h_i \mid i \in I_n\}$ the set of simple coroots given in 2.1. A $\mathfrak{gl}_n$-*crystal* is a set B together with the maps

$$\begin{aligned} &\mathrm{wt} : B \to P_n, \\ &\varepsilon_i, \varphi_i : B \to \mathbf{Z} \cup \{-\infty\}, \\ &\mathrm{e}_i, \mathrm{f}_i : B \to B \cup \{\mathbf{0}\}, \end{aligned}$$

for $i \in I_n$, satisfying the following conditions;

(1) for $i \in I_n$ and $b \in B$, we have $\varphi_i(b) = \langle h_i, \mathrm{wt}(b)\rangle + \varepsilon_i(b)$,
(2) if $\mathrm{e}_i b \in B$ for $i \in I_n$ and $b \in B$, then

$$\varepsilon_i(\mathrm{e}_i b) = \varepsilon_i(b) - 1, \quad \varphi_i(\mathrm{e}_i b) = \varphi_i(b) + 1, \quad \mathrm{wt}(\mathrm{e}_i b) = \mathrm{wt}(b) + \alpha_i,$$

(3) if $\mathrm{f}_i b \in B$ for $i \in I_n$ and $b \in B$, then

$$\varepsilon_i(\mathrm{f}_i b) = \varepsilon_i(b) + 1, \quad \varphi_i(\mathrm{f}_i b) = \varphi_i(b) - 1, \quad \mathrm{wt}(\mathrm{f}_i b) = \mathrm{wt}(b) - \alpha_i,$$

(4) $\mathrm{f}_i b = b'$ if and only if $b = \mathrm{e}_i b'$ for all $i \in I_n$, $b, b' \in B$,
(5) If $\varphi_i(b) = -\infty$, then $\mathrm{e}_i b = \mathrm{f}_i b = \mathbf{0}$,

where $\mathbf{0}$ is a formal symbol and $-\infty$ is the smallest element in $\mathbf{Z} \cup \{-\infty\}$ and is taken to satisfy $-\infty + n = -\infty$ for all $n \in \mathbf{Z}$.

REMARK 3.2. (1) A $\mathfrak{gl}_n$-crystal B is an I_n-colored oriented graph, where

$$b \xrightarrow{i} b' \quad \text{if and only if} \quad b' = f_i b \qquad (i \in I_n).$$

We also call e_i and f_i $(i \in I_n)$ *Kashiwara operators*.

(2) For $\lambda \in P_n^+$, $B(\lambda)$ is a $\mathfrak{gl}_n$-crystal with $\mathrm{e}_i = \widetilde{e}_i$ and $\mathrm{f}_i = \widetilde{f}_i$ for $i \in I_n$ (see 2.2). In fact, the $\mathfrak{gl}_n$-crystals, which we deal with in this chapter, will always satisfy the following additional conditions;

$$\varepsilon_i(b) = \max\{k \in \mathbf{Z}_{\geq 0} : \mathrm{e}_i^k b \neq 0\}, \quad \varphi_i(b) = \max\{k \in \mathbf{Z}_{\geq 0} : \mathrm{f}_i^k b \neq 0\},$$

for $b \in B$ and $i \in I_n$. Such crystals are called *seminormal* [24]. There are $\mathfrak{gl}_n$-crystals that are not isomorphic to a $B(\lambda)$, but play important roles in the construction of various crystal graphs, for example the crystal graphs of the Verma modules and the negative part of a quantum group and so on (see [23]).

Let B_1 and B_2 be $\mathfrak{gl}_n$-crystals. Let $B_1 \otimes B_2$ be $\{ b_1 \otimes b_2 \mid b_i \in B_i \ (i = 1, 2) \}$ as a set. Define

$$\begin{aligned}
\mathrm{wt}(b_1 \otimes b_2) &= \mathrm{wt}(b_1) + \mathrm{wt}(b_2), \\
\varepsilon_i(b_1 \otimes b_2) &= \max(\varepsilon_i(b_1), \varepsilon_i(b_2) - \langle h_i, \mathrm{wt}(b_1) \rangle), \\
\varphi_i(b_1 \otimes b_2) &= \max(\varphi_i(b_1) + \langle h_i, \mathrm{wt}(b_2) \rangle, \varphi_i(b_2)), \\
\mathrm{e}_i(b_1 \otimes b_2) &= \begin{cases} \mathrm{e}_i b_1 \otimes b_2, & \text{if } \varphi_i(b_1) \geq \varepsilon_i(b_2), \\ b_1 \otimes \mathrm{e}_i b_2, & \text{if } \varphi_i(b_1) < \varepsilon_i(b_2), \end{cases} \\
\mathrm{f}_i(b_1 \otimes b_2) &= \begin{cases} \mathrm{f}_i b_1 \otimes b_2, & \text{if } \varphi_i(b_1) > \varepsilon_i(b_2), \\ b_1 \otimes \mathrm{f}_i b_2, & \text{if } \varphi_i(b_1) \leq \varepsilon_i(b_2), \end{cases}
\end{aligned}$$

where we assume that $\mathbf{0} \otimes b_2 = b_1 \otimes \mathbf{0} = \mathbf{0}$. Then, it is straightforward to check that $B_1 \otimes B_2$ is a $\mathfrak{gl}_n$-crystal with respect to the wt, ε_i, φ_i, e_i, f_i $(i \in I_n)$ on $B_1 \otimes B_2$, and it is called the *tensor product of* B_1 and B_2.

Finally, let us fix some terminology. Let B be a $\mathfrak{gl}_n$-crystal. A subset $B' \subset B$ is called a *subcrystal of* B if B' itself is a $\mathfrak{gl}_n$-crystal with respect to the wt, ε_i, φ_i, e_i, f_i $(i \in I_n)$ of B. Let B_1 and B_2 be two $\mathfrak{gl}_n$-crystals. The *direct sum* $B_1 \oplus B_2$ is the disjoint union of B_1 and B_2. An *isomorphism* $\psi : B_1 \to B_2$ of $\mathfrak{gl}_n$-crystals is an isomorphism of I_n-colored oriented graphs which preserves wt, ε_i, and φ_i $(i \in I_n)$. We say that B_1 *is isomorphic to* B_2, and write $B_1 \simeq B_2$. For $b_i \in B_i$ $(i = 1, 2)$, let $C(b_i)$ denote the connected component of b_i as an I_n-colored oriented graph. We say that b_1 *is* ($\mathfrak{gl}_n$-) *crystal equivalent to* b_2 if there is an isomorphism of $\mathfrak{gl}_n$-crystals $C(b_1) \to C(b_2)$ sending b_1 to b_2, and write $b_1 \simeq_{\mathfrak{gl}_n} b_2$.

3.2. *Crystal of words*

For $n \geq 2$, let

$$\mathbf{B}_n = \{ 1 < 2 < \cdots < n \}$$

be a linearly ordered set. Then $\mathbf{B}_n$ can be made into a $\mathfrak{gl}_n$-crystal whose associated I_n-colored oriented graph is given by

$$1 \xrightarrow{1} 2 \xrightarrow{2} \cdots \xrightarrow{n-2} n-1 \xrightarrow{n-1} n,$$

where $\mathrm{wt}(b) = \epsilon_b$. Note that $\mathbf{B}_n$ is the crystal graph of the natural representation $K(q)^{\oplus n}$ [26].

Let $\mathcal{W}_n$ be the set of all finite words with letters in $\mathbf{B}_n$. That is,

$$\mathcal{W}_n = \left(\bigoplus_{r \geq 1} \mathbf{B}_n^{\otimes r} \right) \oplus \{\emptyset\},$$

where $\emptyset$ denotes the empty word. Then $\mathcal{W}_n$ is a $\mathfrak{gl}_n$-crystal, where $\{\emptyset\}$ forms a trivial crystal, that is, $\mathrm{wt}(\emptyset) = 0, \mathrm{e}_i\emptyset = \mathrm{f}_i\emptyset = \mathbf{0}$, and $\varepsilon_i(\emptyset) = \varphi_i(\emptyset) = 0$ for all $i \in I_n$. In fact, $\mathcal{W}_n$ is the crystal graph of the tensor algebra generated by the natural representation. For simplicity, let us write a nonempty word $w_1 \otimes \cdots \otimes w_r \in \mathbf{B}_n^{\otimes r}$ as $w_1 \cdots w_r$.

Following the tensor product rule of the crystals, we can describe the Kashiwara operators $\mathrm{e}_i, \mathrm{f}_i : \mathcal{W}_n \to \mathcal{W}_n \cup \{\mathbf{0}\}$ $(i \in I_n)$ induced from those on $\mathbf{B}_n$ explicitly;

(1) Suppose that a nonempty word $w = w_1 \cdots w_r$ is given. To each letter w_k, we assign

$$\epsilon^{(i)}(w_k) = \begin{cases} +, & \text{if } w_k = i, \\ -, & \text{if } w_k = i+1, \\ \cdot, & \text{otherwise,} \end{cases}$$

and let $\epsilon^{(i)}(w) = (\epsilon^{(i)}(w_1), \ldots, \epsilon^{(i)}(w_r))$.

(2) We replace a pair $(\epsilon^{(i)}(w_s), \epsilon^{(i)}(w_{s'})) = (+, -)$ such that $s < s'$ and $\epsilon^{(i)}(w_t) = \cdot$ for $s < t < s'$ with $(\cdot, \cdot)$ in $\epsilon^{(i)}(w)$, and repeat this process as long as possible until we get a sequence with no $+$ placed to the left of $-$. This is called the *i-signature of w*.

(3) We call the right-most $-$ in the i-signature of w the *i-good $-$ sign*, and define $\mathrm{e}_i w$ to be the word obtained by applying e_i to $i+1$ corresponding to the i-good $-$ sign. If there is no i-good $-$ sign, then we define $\mathrm{e}_i w = \mathbf{0}$.

(4) We call the left-most $+$ in the i-signature of w the *i-good $+$ sign*, and define $\mathrm{f}_i w$ to be the word obtained by applying f_i to i corresponding to the i-good $+$ sign. If there is no i-good $+$ sign, then we define $\mathrm{f}_i w = \mathbf{0}$.

EXAMPLE 3.3. Suppose that

$$w = 1\,2\,4\,2\,1\,1\,2\,3\,2\,1\,1\,3 \in \mathbf{B}_4^{\otimes 12}.$$

Then

$$\begin{aligned} \epsilon^{(1)}(w) &= (+, -, \cdot, -, +, +, -, \cdot, -, +, +, \cdot) \\ &= (\not+, \not-, \cdot, -, \not+, \not+, \not-, \cdot, \not-, +, +, \cdot) \\ &= (\cdot, \cdot, \cdot, \ominus, \cdot, \cdot, \cdot, \cdot, \cdot, \oplus, +, \cdot) \end{aligned}$$

where $\oplus$ and $\ominus$ denote the i-good signs. We have

$$\mathrm{e}_1 w = 1\,2\,4\,\mathbf{1}\,1\,1\,2\,3\,2\,1\,1\,3, \quad \mathrm{f}_1 w = 1\,2\,4\,2\,1\,1\,2\,3\,2\,\mathbf{2}\,1\,3.$$

3.3. *Crystal of a Young tableau*

A *partition* is a nonincreasing sequence of nonnegative integers $\lambda = (\lambda_k)_{k \geq 1}$ such that all but a finite number of its terms are zero. Each λ_k is called a *part of* λ,

and the number of nonzero parts is called the *length of* λ and denoted by $\ell(\lambda)$. We also write $\lambda = (1^{m_1}, 2^{m_2}, \ldots)$, where m_i is the number of occurrences of i in λ, and $|\lambda| = \sum_{k\geq 1} \lambda_k = \sum_{i\geq 1} i m_i$. Recall that a partition $\lambda = (\lambda_k)_{k\geq 1}$ is identified with a *Young diagram*, which is a collection of nodes (or boxes) in left-justified rows with λ_k nodes in the k-th row numbered from top to bottom. We denote by $\mathscr{P}$ the set of all partitions, and set $\mathscr{P}_n = \{\, \lambda \in \mathscr{P} \mid \ell(\lambda) \leq n \,\}$ for $n \geq 1$.

A *Young tableau* T is obtained by filling a Young diagram λ with the entries from $\mathbf{N}$. We say that T is *semistandard* if the entries in each row (resp. column) are weakly (resp. strictly) increasing from left to right (resp. from top to bottom). If all the entries in T are distinct, then T is called *standard*. We say that λ is the *shape of* T, and write $\mathrm{sh}(T) = \lambda$.

For $\lambda \in \mathscr{P}_n$, let $\mathbf{B}_n(\lambda)$ be the set of all semistandard (Young) tableaux of shape λ with entries in $\mathbf{B}_n = \{1, \ldots, n\}$. For each $T \in \mathbf{B}_n(\lambda)$, let $w(T)$ (or $w_{\mathrm{col}}(T)$) be the word obtained by reading the entries of T column by column from right to left, and in each column from top to bottom. Note that a semistandard tableau can be uniquely recovered from the associated word. Hence the map $T \mapsto w(T)$ gives a natural embedding of $\mathbf{B}_n(\lambda)$ into $\mathcal{W}_n$, and we may view $\mathbf{B}_n(\lambda)$ as a subset of $\mathcal{W}_n$.

EXAMPLE 3.4.

$$\mathbf{B}_5((3,2)) \ni \begin{array}{ll} 1 & 2 \; 2 \\ 4 & 5 \end{array} \underset{\text{reading}}{\overset{\text{Column}}{\longmapsto}} \quad 2\,2\,5\,1\,4 \;\in \mathcal{W}_5$$

For $T \in \mathbf{B}_n(\lambda)$, the weight of T is given by $\mathrm{wt}(T) = \sum_{b\in\mathbf{B}_n} \mu_b \epsilon_b \in P_n$, where μ_b is the number of occurrences of b in T. Indeed, $\mathbf{B}_n(\lambda)$ together with $\mathbf{0}$ is stable under $\mathrm{e}_i, \mathrm{f}_i$ $(i \in I_n)$, and hence $\mathbf{B}_n(\lambda)$ is a subcrystal of $\mathcal{W}_n$.

THEOREM 3.5 ([26]). *For $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathscr{P}_n$, $\mathbf{B}_n(\lambda)$ is a $\mathfrak{gl}_n$-crystal, which is isomorphic to the crystal graph of the irreducible highest weight module with highest weight $\sum_{b\in\mathbf{B}_n} \lambda_b \epsilon_b \in P_n^+$.*

REMARK 3.6. (1) For $\lambda \in \mathscr{P}_n$, let H_n^λ be the semistandard tableaux in $\mathbf{B}_n(\lambda)$ such that each k-th row is filled with k $(1 \leq k \leq n)$. Then $\mathrm{wt}(H_n^\lambda) \geq \mathrm{wt}(T)$ for all $T \in \mathbf{B}_n(\lambda)$, that is, H_n^λ is the highest weight vector of weight $\lambda_1\epsilon_1 + \cdots + \lambda_n\epsilon_n$[4]. Hence, each $T \in \mathbf{B}_n(\lambda)$ can be obtained from H_n^λ by applying a finite number of f_i.

(2) The way of reading the entries of a tableau by which we give a crystal structure on $\mathbf{B}_n(\lambda)$ is not unique. In fact, we may also define $\mathrm{e}_i, \mathrm{f}_i$ on $\mathbf{B}_n(\lambda)$ with respect to various readings (see, for example, [2]).

[4] We may identify a partition $\lambda = (\lambda_1, \ldots, \lambda_n)$ with a dominant integral weight $\lambda_1\epsilon_1 + \cdots + \lambda_n\epsilon_n$.

EXAMPLE 3.7. Here is the $\mathfrak{gl}_3$-crystal $\mathbf{B}_3((2,1))$, where the action of the f_i on semistandard tableaux is given by that on the associated words in $\mathcal{W}_3$.

$$\begin{array}{ccccc}
\begin{array}{ll}1 & 1\\ 2 & \end{array} & \xrightarrow{2} & \begin{array}{ll}1 & 1\\ 3 & \end{array} & \xrightarrow{1} & \begin{array}{ll}1 & 2\\ 3 & \end{array}\\
\downarrow 1 & & & & \downarrow 1\\
\begin{array}{ll}1 & 2\\ 2 & \end{array} & \xrightarrow{2} & \begin{array}{ll}1 & 3\\ 2 & \end{array} & & \begin{array}{ll}2 & 2\\ 3 & \end{array}\\
 & & \downarrow 2 & & \downarrow 2\\
 & & \begin{array}{ll}1 & 3\\ 3 & \end{array} & \xrightarrow{1} & \begin{array}{ll}2 & 3\\ 3 & \end{array}
\end{array}$$

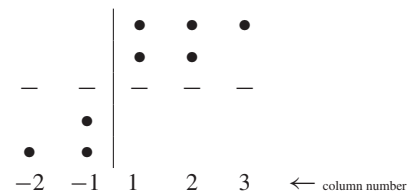
In general, we can also define a $\mathfrak{gl}_n$-crystal for semistandard tableaux of skew shape. For $\lambda, \mu \in \mathscr{P}$ such that $\lambda \supset \mu$ (that is, $\lambda_k \geq \mu_k$ for all k), let λ/μ be the diagram given by removing the nodes of μ from λ. Let $\mathbf{B}_n(\lambda/\mu)$ be the set of all semistandard tableaux of shape λ/μ with entries in $\mathbf{B}_n$. Just as mentioned earlier, we can check that $\mathbf{B}_n(\lambda/\mu)$ is a $\mathfrak{gl}_n$-crystal. The crystal $\mathbf{B}_n(\lambda/\mu)$ is not necessarily connected. We will discuss its explicit decomposition into its connected components in Section 4.

3.4. *Crystal of a rational Young tableau*

Let us describe the crystal graph of the irreducible highest weight $U_q(\mathfrak{gl}_n)$-module with arbitrary highest weight $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbf{Z}_+^n = P_n^+$.

An element $\lambda = (\lambda_1, \ldots, \lambda_n)$ in $\mathbf{Z}_+^n$ is called a *generalized partition* of length n, and it may be identified with a *generalized Young diagram* in the following way. First, we fix a vertical line. Then for each λ_k, we place $|\lambda_k|$ nodes (or boxes) in the k-th row in a left-justified (resp. right-justified) way with respect to this vertical line if $\lambda_k \geq 0$ (resp. $\lambda_k \leq 0$). For example,

$\lambda = (3, 2, 0, -1, -2) \quad \longleftrightarrow$



where the horizontal line denotes an empty row. We enumerate the columns of a diagram as mentioned earlier.

Let $\mathbf{B}_n^\vee = \{-n < -n+1 < \cdots < -1\}$ be the *dual crystal* of $\mathbf{B}_n$ whose associated graph is given by

$$-n \xrightarrow{n-1} -n+1 \xrightarrow{n-2} \cdots \xrightarrow{1} -1,$$

where $\mathrm{wt}(-b) = -\mathrm{wt}(b) = -\epsilon_b$ for $b \in \mathbf{B}_n$ (cf [24]).

Let T be a tableau obtained by filling a generalized Young diagram $\lambda = (\lambda_1, \ldots, \lambda_n)$ of length n with the entries in $\mathbf{B}_n \sqcup \mathbf{B}_n^\vee$. We call T a *rational semistandard tableau of shape* λ if

(1) the entries in the columns indexed by positive (resp. negative) numbers are from $\mathbf{B}_n$ (resp. $\mathbf{B}_n^\vee$),
(2) the entries in each row (resp. column) are weakly (resp. strictly) increasing from left to right (resp. from top to bottom),
(3) if $b_1 < \cdots < b_s$ (resp. $-b'_1 < \cdots < -b'_t$) are the entries in the 1-st (resp. the -1-st) column ($s + t \leq n$), then $b''_i \leq b_i$ for $1 \leq i \leq s$, where $\{b''_1 < \cdots < b''_{n-t}\} = \mathbf{B}_n \backslash \{b'_1, \cdots, b'_t\}$.

We denote by $\mathbf{B}_n(\lambda)$ the set of all rational semistandard tableaux of shape λ.

REMARK 3.8. The notion of rational semistandard tableaux was introduced by Stembridge [42] to describe the characters of rational representations of $\mathfrak{gl}_n$ over the complex numbers.

EXAMPLE 3.9. For $\lambda = (3, 2, 0, -1, -2)$, we have

$$\begin{array}{cc|ccc} & & 1 & 2 & 2 \\ & & 3 & 4 & \\ - & - & - & - & - \\ & -5 & & & \\ -4 & -3 & & & \end{array} \in \mathbf{B}_5(\lambda).$$

Now, embed $\mathbf{B}_n(\lambda)$ into $(\mathbf{B}_n \oplus \mathbf{B}_n^\vee)^{\otimes N}$ (where $N = \sum_{k \geq 1} |\lambda_k|$) by column reading of tableaux, and apply e_i, f_i to $\mathbf{B}_n(\lambda)$ ($i \in I_n$). Then $\mathbf{B}_n(\lambda)$ together with $\mathbf{0}$ is stable under e_i, f_i, and becomes a $\mathfrak{gl}_n$-crystal, which is in fact isomorphic to the crystal graph of the irreducible highest weight $U_q(\mathfrak{gl}_n)$-module with highest weight λ (which is $\sum_{b \in \mathbf{B}_n} \lambda_b \epsilon_b$).

Let us explain the relation between the crystals of rational semistandard tableaux and ordinary semistandard tableaux. Let T be a rational semistandard tableau in $\mathbf{B}_n((0^{n-t}, -1^t))$ $(0 \leq t \leq n)$ with the entries $-b_1 < \cdots < -b_t$ in the -1-st column. We define $\sigma(T)$ be the tableau in $\mathbf{B}_n((1^{n-t}, 0^t))$ with the entries $b'_1 < \cdots < b'_{n-t}$ in the 1-st column, where $\{b'_1 < \cdots < b'_{n-t}\} = \mathbf{B}_n \setminus \{b_1 < \cdots < b_t\}$. If $t = n$, then we define $\sigma(T)$ to be the empty tableau.

In general, for $\lambda \in \mathbf{Z}_+^n$ and $T \in \mathbf{B}_n(\lambda)$, we define $\sigma(T)$ to be the tableau obtained by applying σ to the -1-st column of T. Note that the shape of $\sigma(T)$ is $\lambda + (1^n)$. For example, when $n = 5$, we have

$$\sigma\left(\begin{array}{cc|ccc} & & 1 & 2 & 2 \\ & & 3 & 4 & \\ - & - & - & - & - \\ & -5 & & & \\ -4 & -3 & & & \end{array}\right) = \begin{array}{c|cccc} \mathbf{1} & 1 & 2 & 2 \\ \mathbf{2} & 3 & 4 & \\ \mathbf{4} & & & \\ - & - & - & - \\ -4 & & & \end{array}.$$

LEMMA 3.10. *For $\lambda \in \mathbf{Z}_+^n$, the map*

$$\sigma : \mathbf{B}_n(\lambda) \cup \{\mathbf{0}\} \to \mathbf{B}_n(\lambda + (1^n)) \cup \{\mathbf{0}\},$$

($\sigma(\mathbf{0}) = \mathbf{0}$) is a bijection which commutes with $\mathrm{e}_i, \mathrm{f}_i$ *for* $i \in I_n$, *where* $\mathrm{wt}(\sigma(T)) = \mathrm{wt}(T) + (\epsilon_1 + \cdots + \epsilon_n)$ *for* $T \in \mathbf{B}_n(\lambda)$. *In particular,* $\mathbf{B}_n(\lambda)$ *and* $\mathbf{B}_n(\lambda + (1^n))$ *gives the same* I_n*-colored oriented graph.*

In other words, the crystal $\mathbf{B}_n(\lambda)$ for $\lambda \in \mathbf{Z}_+^n$ is essentially the same as the crystal of ordinary semistandard tableaux $\mathbf{B}_n(\lambda + (d^n))$ for $\lambda + (d^n) \in \mathscr{P}_n$. In particular, it is connected with the unique highest weight element, say H_n^λ.

EXAMPLE 3.11. For $\lambda = (3, 2, 0, -1, -2)$,

$$H_n^\lambda = \begin{array}{ccc|ccc} & & & 1 & 1 & 1 \\ & & & 2 & 2 & \\ - & - & - & - & - & - \\ & & -5 & & & \\ & -5 & -4 & & & \end{array}$$

is the unique highest weight element in $\mathbf{B}_5(\lambda)$. Note $H_n^\lambda = \sigma^{-2}\left(H_n^{\lambda+(2^5)}\right)$.

4. Crystal equivalence

In classical representation theory, the decompositions of finite-dimensional $\mathfrak{gl}_n$-modules are explained using the characters of the representations, which are symmetric polynomials in n variables. In particular, there are well-known combinatorial algorithms, called the *Robinson–Schensted correspondence* and the *Littlewood–Richardson rule*, which provide in a bijective way the character identities corresponding to the decomposition of the tensor power of the natural representation and the tensor product of two irreducible highest weight representations, respectively (for a background on the combinatorics of Young tableaux, which will be needed in this section, see [10, 37, 41]).

In this section, we will see that these algorithms obtain more concrete representation theoretical meanings in the language of crystals, that is, they yield *morphisms of crystals*, and hence explain the decompositions of the corresponding crystals directly.

4.1. *Knuth equivalence*

For a given word $w \in \mathcal{W}_n$ and three consecutive letters xyz in w, consider the transformations

$$w = \ldots xyz \ldots \longmapsto w' = \ldots xzy \ldots, \quad \text{if } y \le x < z,$$
$$w = \ldots xyz \ldots \longmapsto w' = \ldots yxz \ldots, \quad \text{if } y < x \le z,$$

which are called the *elementary Knuth transformations*. Then, we say that two words w and w' in $\mathcal{W}_n$ are *Knuth equivalent* if they can be transformed into each other by a sequence of elementary Knuth transformations and their inverses, and write $w \simeq_{\mathrm{K}} w'$.

In general, for two semistandard tableau T and T' with entries in $\{1, \ldots, n\}$, we define $T \simeq_{\mathrm{K}} T'$ if $w(T) \simeq_{\mathrm{K}} w(T')$.

REMARK 4.1. Note that the following two semistandard tableaux (of skew shapes) are Knuth equivalent.

$$\begin{array}{ll} y & x \\ z & \end{array} \simeq_{\mathrm{K}} \begin{cases} \begin{array}{ll} & x \\ y & z \end{array}, & \text{if } y \le x < z, \\ \begin{array}{ll} & y \\ z & x \end{array}, & \text{if } y < x \le z \end{cases}$$

where the tableau on the right-hand side can be obtained by *sliding* the two entries of the tableau in the left-hand side in such a way that it still remains semistandard, and also vice versa.

Next, let us recall the *Schensted column bumping algorithm* for semistandard tableaux: For $\lambda \in \mathscr{P}_n$ and $T \in \mathbf{B}_n(\lambda)$, we define $(T \leftarrow b)$ $(b \in \mathbf{B}_n)$ to be the tableau obtained from T by applying the following procedure;

(1) Let b' be the smallest entry in the first (or the left-most) column that is greater than or equal to b.
(2) Replace b' by b (b' is *bumped out* of the first column). If there is no such b', put b at the bottom of the first column and stop the procedure.
(3) Repeat (1) and (2) on the next column with b'.

Note that $(T \leftarrow b) \in \mathbf{B}_n(\mu)$ for some $\mu \in \mathscr{P}_n$, where μ is obtained by adding a node to λ. Now, for a given word $w = w_1 \cdots w_r \in \mathcal{W}_n$, define

$$P(w) = ((\cdots((w_1 \leftarrow w_2) \leftarrow w_3) \cdots) \leftarrow w_r).$$

$P(w)$ is called the *P-tableau of* w or the *insertion tableau of* w. Because the Knuth equivalence of words with three letters can be understood in terms of *sliding* of entries in skew diagrams (see Remark 4.1), it is not difficult to see that $w \simeq_{\mathrm{K}} P(w)$ (or $w \simeq_{\mathrm{K}} w(P(w))$).

The following is the key fact, which connects the theory of crystals and the classical combinatorics of Young tableaux;

THEOREM 4.2 ([35]). *For $w, w' \in \mathcal{W}_n$, we have*

$$w \simeq_{\mathrm{K}} w' \Longleftrightarrow w \simeq_{\mathfrak{gl}_n} w'.$$

PROOF. ($\Rightarrow$) By definition of the tensor product of crystals, we may assume that $w = xyz$ and w is transformed to w' by an elementary Knuth transformation. Identifying w and w' with the column words of the corresponding semistandard tableaux of skew shapes respectively (see Remark 4.1), it is straightforward to check that the elementary Knuth transformation naturally extends to an isomorphism of $\mathfrak{gl}_n$-crystals between the connected components of two words w and w'.

($\Leftarrow$) Since $w \simeq_{\mathrm{K}} P(w)$ and $w' \simeq_{\mathrm{K}} P(w')$, we have $P(w) \simeq_{\mathfrak{gl}_n} P(w')$. Suppose that $\mathrm{sh}(P(w)) = \lambda$ and $\mathrm{sh}(P(w')) = \mu$. There exists a sequence of indices $i_1, \cdots, i_r \in I_n$ and nonnegative integers $a_k (1 \le k \le r)$ such that

$$\mathrm{e}_{i_1}^{a_1} \dots \mathrm{e}_{i_r}^{a_r} P(w) = H_n^{\lambda}.$$

Since $P(w) \simeq_{\mathfrak{gl}_n} P(w')$, we must have

$$\mathrm{e}_{i_1}^{a_1} \dots \mathrm{e}_{i_r}^{a_r} P(w') = H_n^{\mu}.$$

Hence, we get $\mathrm{wt}(H_n^{\lambda}) = \mathrm{wt}(H_n^{\mu})$, which implies $\lambda = \mu$ and $H_n^{\lambda} = H_n^{\mu}$. Since $P(w) = \mathrm{f}_{i_r}^{a_r} \cdots \mathrm{f}_{i_1}^{a_1} H_n^{\lambda}$, we conclude that $P(w) = P(w')$. In particular, we have $w \simeq_{\mathrm{K}} w'$ □

COROLLARY 4.3. *Each connected component in $\mathcal{W}_n$ is isomorphic to a $\mathbf{B}_n(\lambda)$ for some $\lambda \in \mathscr{P}_n$.*

COROLLARY 4.4 (cf [30]). *Let T and T' be two semistandard tableaux. If $T \simeq_{\mathrm{K}} T'$, then $T = T'$. That is, for a given word $w \in \mathcal{W}_n$, there exists a unique semistandard tableau T which is Knuth equivalent to w.*

4.2. *Robinson–Schensted correspondence*

Let us consider the decomposition of $\mathbf{B}_n^{\otimes r}$ $(r \ge 2)$. For $w = w_1 \dots w_r \in \mathcal{W}_n$, suppose that $P(w)$ is of shape λ. Then we define $Q(w)$ to be the standard tableau of shape λ such that if w_i is inserted into $(((\cdots(w_1 \leftarrow w_2) \cdots) \leftarrow w_{i-1})$ to create a node in λ, then we fill the node with i. We call $Q(w)$ the *Q-tableau of w* or the *recording tableau*.

EXAMPLE 4.5. For example, if $w = 31224 \in \mathcal{W}_4$, then we have

$$P(w) = \begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 2 \\ \cline{1-1} 4 \\ \cline{1-1} \end{array}, \quad Q(w) = \begin{array}{|c|c|c|} \hline 1 & 2 & 4 \\ \hline 3 \\ \cline{1-1} 5 \\ \cline{1-1} \end{array}.$$

Then the *Robinson–Schensted correspondence* is that the map $w \mapsto (P(w), Q(w))$ for $w \in \mathcal{W}_n$ gives a one-to-one correspondence between the two sets, $\mathbf{B}_n^{\otimes r}$ and $\bigsqcup_{\lambda} \mathbf{B}_n(\lambda) \times ST_r(\lambda)$, where the union is taken over all partitions $\lambda \in \mathscr{P}_n$ such that $|\lambda| = r$, and $ST_r(\lambda)$ denotes the set of all standard tableaux of shape λ with entries in $\{1, \dots, r\}$.

Because each connected component in $\mathbf{B}_n^{\otimes r}$ is isomorphic to a $\mathbf{B}_n(\lambda)$ for some $\lambda \in \mathscr{P}_n$ with $|\lambda| = r$ and $\mathbf{B}_n(\lambda)$ is connected, the multiplicity of $\mathbf{B}_n(\lambda)$ in $\mathbf{B}_n^{\otimes r}$, say m_{λ}, is equal to the number of the words in $\mathbf{B}_n^{\otimes r}$ such that $P(w) = H_n^{\lambda}$, that is, $m_{\lambda} = |ST_r(\lambda)|$ by the Robinson–Schensted correspondence.

Moreover, we can check that for $w \in \mathbf{B}_n^{\otimes r}$ and $i \in I_n$,

$$Q(w) = Q(\mathrm{f}_i w), \quad \text{if } \mathrm{f}_i w \neq \mathbf{0},$$

(equivalently, $Q(w) = Q(\mathrm{e}_i w)$, if $\mathrm{e}_i w \neq \mathbf{0}$). In other words, the Q-tableaux are invariant under the Kashiwara operators, and it turns out that the Q-tableau of w represents the connected component of w in $\mathbf{B}_n^{\otimes r}$. Summarizing the aforementioned arguments, we have the following crystal version of the Robinson–Schensted correspondence;

THEOREM 4.6 ([31]). *The map $w \mapsto (P(w), Q(w))$ gives the following isomorphism of $\mathfrak{gl}_n$-crystals;*

$$\mathbf{B}_n^{\otimes r} \simeq \bigoplus_{\substack{\lambda \in \mathscr{P}_n, |\lambda| = r \\ Q \in ST_r(\lambda)}} \mathbf{B}_n(Q),$$

where $\mathbf{B}_n(Q) = \mathbf{B}_n(\lambda) \times \{ Q \}$ for $\lambda \in \mathscr{P}_n$ with $|\lambda| = r$ and $Q \in ST_r(\lambda)$ is a $\mathfrak{gl}_n$-crystal isomorphic to $\mathbf{B}_n(\lambda)$.

A word $w = w_1 \dots w_r \in \mathcal{W}_n$ is called a *lattice permutation* if for $1 \leq k \leq r$ and $1 \leq i \leq n-1$, the number of occurrences of i in $w_1 \dots w_k$ is no less than that of $i+1$. Then it follows from the crystal structure on $\mathcal{W}_n$ (see 3.2) that $\mathrm{e}_i w = \mathbf{0}$ for all $i \in I_n$ (that is, $P(w) = H_n^\lambda$ for some $\lambda \in \mathscr{P}_n$) if and only if w is a lattice permutation.

REMARK 4.7. A connection between the Robinson–Schensted correspondence and the representations of $U_q(\mathfrak{gl}_n)$ was first observed by Date et al. [6].

4.3. *Littlewood–Richardson rule*

Next, let us discuss the decomposition of $\mathbf{B}_n(\mu) \otimes \mathbf{B}_n(\nu)$ for $\mu, \nu \in \mathscr{P}_n$. For $T \in \mathbf{B}_n(\mu)$ and $T' \in \mathbf{B}_n(\nu)$, let $w(T') = w_1 w_2 \dots w_r$. We define

$$P(T \otimes T') = ((((T \leftarrow w_1) \leftarrow w_2) \cdots) \leftarrow w_r)$$

(usually denoted as $(T \leftarrow T')$). Clearly, we have $T \otimes T' \simeq_{\mathfrak{gl}_n} P(T \otimes T')$.

Next, define $Q(T \otimes T')$ to be the semistandard tableau of shape λ/μ ($\lambda = \mathrm{sh}(P(T \otimes T'))$ such that if w_i is in the k-th row of T' and inserted into $(((\cdots(T \leftarrow w_1) \leftarrow w_2) \cdots) \leftarrow w_{i-1})$ to create a node in λ/μ, then we fill the node with k. We call $Q(T \otimes T')$ the *recording tableau of* $(T \leftarrow T')$.

EXAMPLE 4.8. Suppose that we have

$$T = \begin{array}{ccc} 1 & 2 & 2 \\ 3 & & \end{array} \quad \text{and} \quad T' = \begin{array}{cc} 2 & 2 \\ 3 & 4 \\ 4 & \end{array}.$$

Then we have

$$P(T \otimes T') = \begin{array}{cccc} 1 & 2 & 2 & 2 \\ 2 & 3 & & \\ 3 & 4 & & \\ 4 & & & \end{array} \quad \text{and} \quad Q(T \otimes T') = \begin{array}{cccc} \bullet & \bullet & \bullet & 1 \\ \bullet & 1 & & \\ 2 & 2 & & \\ 3 & & & \end{array} .$$

Recall that given λ, μ, and ν in $\mathscr{P}$ such that $\mu \subset \lambda$ and $|\lambda| = |\mu| + |\nu|$, a semistandard tableau T of shape λ/μ is called a *Littlewood–Richardson tableau of shape* λ/μ *with content* ν if

(1) the number of occurrences of k in T is equal to ν_k for $k \geq 1$,
(2) $w(T)$ is a lattice permutation.

We denote by $\mathrm{LR}^{\lambda}_{\mu\,\nu}$ the set of all Littlewood–Richardson tableaux of shape λ/μ with content ν.

Then, the map $T \otimes T' \mapsto (P(T \otimes T'), Q(T \otimes T'))$ gives a bijection between the two sets $\mathbf{B}_n(\mu) \otimes \mathbf{B}_n(\nu)$ and $\bigsqcup_{\lambda \in \mathscr{P}_n} \mathbf{B}_n(\lambda) \times \mathrm{LR}^{\lambda}_{\mu\,\nu}$ [44]. Because the connected components are completely determined by $T \otimes T'$ such that $P(T \otimes T') = H^{\lambda}_n$, the multiplicity of $\mathbf{B}_n(\lambda)$ in $\mathbf{B}_n(\mu) \otimes \mathbf{B}_n(\nu)$ is equal to $|\mathrm{LR}^{\lambda}_{\mu\,\nu}|$ called the Littlewood–Richardson coefficient.

Moreover, as in the case of the Robinson–Schensted correspondence, we have for $T \otimes T' \in \mathbf{B}_n(\mu) \otimes \mathbf{B}_n(\nu)$ and $i \in I_n$,

$$Q(T \otimes T') = Q(\mathrm{f}_i(T \otimes T')), \quad \text{if } \mathrm{f}_i(T \otimes T') \neq \mathbf{0}.$$

Hence, the Q-tableau is invariant under the Kashiwara operators, and it represents the connected component of $T \otimes T'$ in $\mathbf{B}_n(\mu) \otimes \mathbf{B}_n(\nu)$. Summarizing, we also have the crystal version of the Littlewood–Richardson rule;

THEOREM 4.9 (cf [40]). *The map* $T \otimes T' \mapsto (P(T \otimes T'), Q(T \otimes T'))$ *gives the following isomorphism of* $\mathfrak{gl}_n$*-crystals;*

$$\mathbf{B}_n(\mu) \otimes \mathbf{B}_n(\nu) \simeq \bigoplus_{\substack{\lambda \in \mathscr{P}_n \\ Q \in \mathrm{LR}^{\lambda}_{\mu\nu}}} \mathbf{B}_n(Q),$$

where $\mathbf{B}_n(Q) = \mathbf{B}_n(\lambda) \times \{ Q \}$ *for* $\lambda \in \mathscr{P}_n$ *and* $Q \in LR^{\lambda}_{\mu\,\nu}$ *is a* $\mathfrak{gl}_n$*-crystal isomorphic to* $\mathbf{B}_n(\lambda)$.

For a skew Young diagram λ/μ, let us consider $\mathbf{B}_n(\lambda/\mu)$. Then the decomposition of $\mathbf{B}_n(\lambda/\mu)$ into its connected components reduces to finding all $T \in \mathbf{B}_n(\lambda/\mu)$ equivalent to H^{ν}_n, where ν is the content (or weight) of T. For $T \in \mathbf{B}_n(\lambda/\mu)$, it follows that

$$\begin{aligned} T \simeq_{\mathfrak{gl}_n} H^{\nu}_n &\iff \mathrm{e}_i T = \mathbf{0} \ \text{ for all } i \in I_n \\ &\iff w(T) \text{ is a lattice permutation} \\ &\iff T \in \mathrm{LR}^{\lambda}_{\mu\nu}. \end{aligned}$$

Therefore, we obtain the skew Littlewood–Richardson rule;

$$\mathbf{B}_n(\lambda/\mu) \simeq \bigoplus_{\substack{\nu \in \mathscr{P}_n \\ T \in \mathrm{LR}_{\mu\nu}^{\lambda}}} \mathbf{B}_n(T),$$

where $\mathbf{B}_n(T)$ is the connected component of T in $\mathbf{B}_n(\lambda/\mu)$, which is isomorphic to $\mathbf{B}_n(\nu)$.

REMARK 4.10. (1) There are analogues of the material in this section for classical Lie algebras of B, C, D type, using the Kashiwara and Nakashima description [1,32,33]. See also [2,19] for a generalization to Lie superalgebras.

(2) The combinatorics of Young tableaux presented here can be formalized using the notion of the *plactic algebra* introduced by Lascoux and Schützenberger [30], which is a noncommutative ring of the Knuth equivalence classes of the words in $\mathcal{W}_n$ with the multiplication given by juxtaposition (or tensor product) of words. In [35], using his path models, Littelmann gave a nice generalization of the plactic algebra for complex semisimple Lie algebras, which consists of the crystal equivalence classes of certain piecewise linear paths in the weight lattice, where the multiplication is given by the concatenation of paths.

5. Bicrystals

Generalizing the Robinson–Schensted correspondence, Knuth introduced an algorithm that establishes a bijection between $m \times n$ matrices of nonnegative integers and pairs of semistandard tableaux of the same shape [27] (now called the *Robinson–Schensted–Knuth correspondence*). The associated character identity explains the decomposition of the symmetric algebra generated by the tensor product of the natural representations of $\mathfrak{gl}_m$ and $\mathfrak{gl}_n$ into irreducible representations of the Lie algebra $\mathfrak{gl}_m \oplus \mathfrak{gl}_n$ over the complex numbers [13]. In this section, we explain the Knuth result in terms of crystals (see [4,29]), and discuss its applications to combinatorics and representation theory.

5.1. *Robinson–Schensted–Knuth correspondence*

First, let us consider a crystal of biwords. Let m, n be nonnegative integers not both zero. Set

$$\begin{aligned}\Omega_{m,n} = \{\, (\mathbf{i}, \mathbf{j}) \in \mathcal{W}_m \times \mathcal{W}_n \mid \\ (1)\ \mathbf{i} = i_1 \dots i_r \text{ and } \mathbf{j} = j_1 \dots j_r \quad \text{for some } r \geq 0, \\ (2)\ (i_1, j_1) \leq \cdots \leq (i_r, j_r)\},\end{aligned}$$

where the ordering is a rather weird kind of lexicographic type ordering given by

$$(i, j) < (k, l) \iff (j < l) \text{ or } (j = l \text{ and } i > k)$$

for (i, j) and $(k, l) \in \mathbf{B}_m \times \mathbf{B}_n$. Now, for $i \in I_m$ and $(\mathbf{i}, \mathbf{j}) \in \Omega_{m,n}$, we define

$$\mathrm{e}_i(\mathbf{i}, \mathbf{j}) = (\mathrm{e}_i\mathbf{i}, \mathbf{j}), \quad \mathrm{f}_i(\mathbf{i}, \mathbf{j}) = (\mathrm{f}_i\mathbf{i}, \mathbf{j}),$$

where we assume that $x_i(\mathbf{i}, \mathbf{j}) = \mathbf{0}$ if $x_i\mathbf{i} = \mathbf{0}$ $(x = \mathrm{e}, \mathrm{f})$. We set $\mathrm{wt}(\mathbf{i}, \mathbf{j}) = \mathrm{wt}(\mathbf{i}) \in P_m$, $\varepsilon_i(\mathbf{i}, \mathbf{j}) = \varepsilon_i(\mathbf{i})$ and $\varphi_i(\mathbf{i}, \mathbf{j}) = \varphi_i(\mathbf{i})$ $(i \in I_m)$. Then the set of biwords $\Omega_{m,n}$ together with wt, e_i, f_i, ε_i, φ_i $(i \in I_m)$ becomes a $\mathfrak{gl}_m$-crystal. In fact, it is isomorphic to

$$\bigoplus_{\ell_1,\ldots,\ell_n \in \mathbf{Z}_{\geq 0}} \mathbf{B}_m(\ell_1) \otimes \cdots \otimes \mathbf{B}_m(\ell_n),$$

where $\mathbf{B}_m(\ell)$ is a $\mathfrak{gl}_m$-crystal with highest weight $\ell\epsilon_1$ for $\ell \geq 0$ (note that $\mathbf{B}_m(\ell_j)$ corresponds to subwords $i_s i_{s+1} \ldots i_t$ of $\mathbf{i}$ in $(\mathbf{i}, \mathbf{j}) \in \Omega_{m,n}$ such that $j_s = j_{s+1} = \ldots = j_t = j$).

Next, set

$$\Omega^*_{m,n} = \{\, (\mathbf{k}, \mathbf{l}) \in \mathcal{W}_m \times \mathcal{W}_n \mid (\mathbf{l}, \mathbf{k}) \in \Omega_{n,m} \,\}.$$

Similarly, for $j \in I_n$ and $(\mathbf{k}, \mathbf{l}) \in \Omega^*_{m,n}$, we define

$$\mathrm{e}^*_j(\mathbf{k}, \mathbf{l}) = (\mathbf{k}, \mathrm{e}_j\mathbf{l}), \quad \mathrm{f}^*_j(\mathbf{k}, \mathbf{l}) = (\mathbf{k}, \mathrm{f}_j\mathbf{l}),$$

and set $\mathrm{wt}^*(\mathbf{k}, \mathbf{l}) = \mathrm{wt}(\mathbf{l}) \in P_n$, $\varepsilon^*_j(\mathbf{k}, \mathbf{l}) = \varepsilon_j(\mathbf{l})$, and $\varphi^*_j(\mathbf{k}, \mathbf{l}) = \varphi_j(\mathbf{l})$ $(j \in I_n)$. Then also the set of biwords $\Omega^*_{m,n}$ together with wt^*, e^*_j, f^*_j, ε^*_j, φ^*_j $(j \in I_n)$ is a $\mathfrak{gl}_n$-crystal, which is isomorphic to

$$\bigoplus_{(\ell_1,\ldots,\ell_m) \in \mathbf{Z}^m_{\geq 0}} \mathbf{B}_n(\ell_1) \otimes \cdots \otimes \mathbf{B}_n(\ell_m).$$

Let $\mathbf{M}_{m\times n}$ be the set of all $m \times n$ matrices with nonnegative integral entries. We assume that the row (resp. column) numbers are indexed by $\mathbf{B}_m$ (resp. $\mathbf{B}_n$). For $(\mathbf{i}, \mathbf{j}) \in \Omega_{m,n}$, define $A(\mathbf{i}, \mathbf{j}) = (a_{ij})$ to be the matrix in $\mathbf{M}_{m\times n}$, where a_{ij} is the number of k' such that $(i_k, j_k) = (i, j)$ for $i \in \mathbf{B}_m$ and $j \in \mathbf{B}_n$. For $(\mathbf{k}, \mathbf{l}) \in \Omega^*_{m,n}$, we define $A(\mathbf{k}, \mathbf{l}) \in \mathbf{M}_{m\times n}$ in the same way. Then the map $(\mathbf{i}, \mathbf{j}) \mapsto A(\mathbf{i}, \mathbf{j})$ (resp. $(\mathbf{k}, \mathbf{l}) \mapsto A(\mathbf{k}, \mathbf{l})$) gives a bijection from $\Omega_{m,n}$ (resp. $\Omega^*_{m,n}$) to $\mathbf{M}_{m\times n}$, where the pair of empty words $(\emptyset, \emptyset)$ corresponds to the zero matrix. With these identifications, $\mathbf{M}_{m\times n}$ becomes a $\mathfrak{gl}_m$ and $\mathfrak{gl}_n$-crystal, simultaneously.

EXAMPLE 5.1. Suppose that

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 2 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix} \in \mathbf{M}_{4\times 4}.$$

Then $A = A(\mathbf{i}, \mathbf{j}) = A(\mathbf{k}, \mathbf{l})$ for $(\mathbf{i}, \mathbf{j}) \in \Omega_{4,4}$ and $(\mathbf{k}, \mathbf{l}) \in \Omega^*_{4,4}$, where

$$\mathbf{i} = 3\,1\,2\,2\,4\,4\,1\,3\,1, \quad \mathbf{k} = 1\,1\,1\,2\,2\,3\,3\,4\,4,$$
$$\mathbf{j} = 1\,1\,2\,2\,3\,3\,3\,4\,4, \quad \mathbf{l} = 4\,3\,1\,2\,2\,4\,1\,3\,3.$$

DEFINITION 5.2. Suppose that B is a $\mathfrak{gl}_m$-crystal and also a $\mathfrak{gl}_n$-crystal under $\mathrm{e}_i, \mathrm{f}_i$ $(i \in I_m)$ and $\mathrm{e}_j^*, \mathrm{f}_j^*$ $(j \in I_n)$, respectively. We call B a $(\mathfrak{gl}_m, \mathfrak{gl}_n)$-*bicrystal* if $\mathrm{e}_i, \mathrm{f}_i$ commute with $\mathrm{e}_j^*, \mathrm{f}_j^*$ $(i \in I_m, j \in I_n)$, where we understand the Kashiwara operators as the associated maps from $B \cup \{\mathbf{0}\}$ to itself (that is, $x_i\mathbf{0} = x_j^*\mathbf{0} = \mathbf{0}$, for $x = \mathrm{e}, \mathrm{f}$).

For a $(\mathfrak{gl}_m, \mathfrak{gl}_n)$-*bicrystal* B, if we put $\mathbf{I}_{m,n} = I_m \sqcup I_n^*$ $(I_n^* = \{\, k^* \mid 1 \le k \le n-1 \,\})$, then B can be viewed as an $\mathbf{I}_{m,n}$-colored oriented graph, where

$$b \xrightarrow{j^*} b' \quad \text{if and only if} \quad b' = \mathrm{f}_j^* b \qquad (j \in I_n).$$

EXAMPLE 5.3. For $\lambda \in \mathscr{P}_m$ and $\mu \in \mathscr{P}_n$, $\mathbf{B}_m(\lambda) \times \mathbf{B}_n(\mu)$ is a $(\mathfrak{gl}_m, \mathfrak{gl}_n)$-bicrystal where $x_i(b_1, b_2) = (x_i b_1, b_2), x_j^*(b_1, b_2) = (b_1, x_j^* b_2)$ for $(b_1, b_2) \in \mathbf{B}_m(\lambda) \times \mathbf{B}_n(\mu)$ and $x = \mathrm{e}, \mathrm{f}$.

Now, we can also check that $\mathbf{M}_{m\times n}$ is a $(\mathfrak{gl}_m, \mathfrak{gl}_n)$-bicrystal. For $A \in \mathbf{M}_{m\times n}$, suppose that $A = A(\mathbf{i}, \mathbf{j}) = A(\mathbf{k}, \mathbf{l})$ for $(\mathbf{i}, \mathbf{j}) \in \Omega_{m,n}$ and $(\mathbf{k}, \mathbf{l}) \in \Omega_{m,n}^*$. Let us define

$$P(A) = P(\mathbf{i}), \quad Q(A) = P(\mathbf{l}).$$

Note that $P(A) \simeq_{\mathfrak{gl}_m} A$ and $Q(A) \simeq_{\mathfrak{gl}_n} A$ by definition of the crystal structures on $\mathbf{M}_{m\times n}$. Since $\mathbf{M}_{m\times n}$ is a $(\mathfrak{gl}_m, \mathfrak{gl}_n)$-bicrystal, it follows that

$$P(\mathrm{f}_j^* A) = P(A), \quad \text{if } \mathrm{f}_j^* A \neq \mathbf{0} \text{ for } j \in I_n,$$
$$Q(\mathrm{f}_i A) = Q(A), \quad \text{if } \mathrm{f}_i A \neq \mathbf{0} \text{ for } i \in I_m.$$

In other words, $\mathrm{f}_j^* A \simeq_{\mathfrak{gl}_m} A$ and $\mathrm{f}_i A \simeq_{\mathfrak{gl}_n} A$.

Let C be a connected component in $\mathbf{M}_{m\times n}$ as an $\mathbf{I}_{m,n}$-colored oriented graph. For $A \in C$, suppose that $P(A) \in \mathbf{B}_m(\lambda)$ and $Q(A) \in \mathbf{B}_n(\mu)$ for some $\lambda \in \mathscr{P}_m$ and $\mu \in \mathscr{P}_n$. Then the map $A \mapsto (P(A), Q(A))$ extends to an isomorphism of $(\mathfrak{gl}_m, \mathfrak{gl}_n)$-bicrystals;

$$C \longrightarrow \mathbf{B}_m(\lambda) \times \mathbf{B}_n(\mu)$$

(for detailed verifications, see [4, 28]).

Hence, to describe the decomposition of $\mathbf{M}_{m\times n}$ as a $(\mathfrak{gl}_m, \mathfrak{gl}_n)$-bicrystal, it suffices to find all the highest weight elements A in $\mathbf{M}_{m\times n}$, that is, those A for which $(P(A), Q(A)) = \left(H_m^\lambda, H_n^\mu\right)$ for some $\lambda \in \mathscr{P}_m$ and $\mu \in \mathscr{P}_n$, or equivalently, $\mathrm{e}_i A = \mathrm{e}_j^* A = \mathbf{0}$ for all $i \in I_m$ and $j \in I_n$. It is not difficult to verify that $A = (a_{ij})$ is of highest weight if and only if (1) $a_{ij} = 0$ unless $i = j$, and (2) the diagonal entries, say $\lambda_i = a_{ii}$ for $1 \le i \le \min\{m, n\}$, are weakly decreasing, and hence they form a partition $\lambda = (\lambda_i)_{i\ge 1} \in \mathscr{P}_m \cap \mathscr{P}_n$. For example, if $m \le n$, then

$$A = \begin{pmatrix} \lambda_1 & & & & & \\ & \lambda_2 & & & & \\ & & \ddots & & & \\ & & & \lambda_{m-1} & & \\ & & & & \lambda_m & \end{pmatrix}.$$

Then, the connected component of A is isomorphic to $\mathbf{B}_m(\lambda) \times \mathbf{B}_n(\lambda)$. Therefore, we obtain the crystal version of the Robinson–Schensted–Knuth correspondence.

THEOREM 5.4 ([4]). *The map $A \mapsto (P(A), Q(A))$ gives the following isomorphism of $(\mathfrak{gl}_m, \mathfrak{gl}_n)$-bicrystals;*

$$\mathbf{M}_{m\times n} \simeq \bigoplus_{\lambda\in\mathscr{P}_m\cap\mathscr{P}_n} \mathbf{B}_m(\lambda) \times \mathbf{B}_n(\lambda).$$

REMARK 5.5. The map $A \mapsto (P(A), Q(A))$ above is slightly different from Knuth's original description. But if we give $\mathbf{M}_{m\times n}$ another bicrystal structure by replacing $\mathbf{B}_m$ and $\mathbf{B}_n$ (the index sets of rows and columns in $\mathbf{M}_{m\times n}$) with their duals, then we recover the original Knuth map with a little modification, which amounts to a morphism of crystals. (See [10, 28] for other variations of the Knuth correspondence and bicrystal structures on $\mathbf{M}_{m\times n}$.)

As an application to combinatorics, let us consider the crystal of symmetric matrices with respect to a diagonal action of the Kashiwara operators. For $n \geq 2$, set

$$\mathfrak{M}_n = \{\, A \in \mathbf{M}_{n\times n} \mid A = A^t \,\}.$$

For $i \in I_n$ and $A \in \mathfrak{M}_n$, we define

$$\mathbf{e}_i A = \mathrm{e}_i \mathrm{e}_i^* A = \mathrm{e}_i^* \mathrm{e}_i A, \quad \mathbf{f}_i A = \mathrm{f}_i \mathrm{f}_i^* A = \mathrm{f}_i^* \mathrm{f}_i A.$$

Then we have

$$(\mathbf{e}_i A)^t = (\mathrm{e}_i \mathrm{e}_i^* A)^t = \mathrm{e}_i^* \mathrm{e}_i A^t = \mathbf{e}_i A \in \mathfrak{M}_n \cup \{\mathbf{0}\},$$
$$(\mathbf{f}_i A)^t = (\mathrm{f}_i \mathrm{f}_i^* A)^t = \mathrm{f}_i^* \mathrm{f}_i A^t = \mathbf{f}_i A \in \mathfrak{M}_n \cup \{\mathbf{0}\}.$$

Hence, $\mathbf{e}_i, \mathbf{f}_i : \mathfrak{M}_n \to \mathfrak{M}_n \cup \{\mathbf{0}\}$ are well-defined operators. Note that $P(A) = Q(A)$, and

$$P(\mathbf{e}_i A) = \mathrm{e}_i P(A), \quad P(\mathbf{f}_i A) = \mathrm{f}_i P(A).$$

Now, if we put $\mathrm{wt}(A) = \mathrm{wt}(P(A)) = \mathrm{wt}(Q(A)) \in P_n$, $\varepsilon_i(A) = \max\{\, k : \mathbf{e}_i^k A \neq \mathbf{0} \,\}$, and $\varphi_i(A) = \max\{\, k : \mathbf{f}_i^k A \neq \mathbf{0} \,\}$ for $i \in I_n$ and $A \in \mathfrak{M}_n$, then $\mathfrak{M}_n$ is a $\mathfrak{gl}_n$-crystal. Because the highest weight elements in $\mathfrak{M}_n$ are exactly those in $\mathbf{M}_{n\times n}$, we have

PROPOSITION 5.6 (cf [27]). *$\mathfrak{M}_n$ is a $\mathfrak{gl}_n$-crystal, which decomposes as follows;*

$$\mathfrak{M}_n \simeq \bigoplus_{\lambda\in\mathscr{P}_n} \mathbf{B}_n(\lambda).$$

For $A = (a_{ij}) \in \mathfrak{M}_n$, let

$$\mathfrak{o}(A) = |\,\{\, i \in \mathbf{B}_n \mid a_{ii} \equiv 1 \pmod 2 \,\}\,|\,.$$

With the help of the crystal structure on $\mathfrak{M}_n$, we may easily deduce the following fact, which comes from [27].

PROPOSITION 5.7. *Let* $\mathfrak{M}_n(k) = \{ A \in \mathfrak{M}_n \mid \mathfrak{o}(A) = k \}$ *for* $k \geq 0$. *Then* $\mathfrak{M}_n(k)$ *is a subcrystal of* $\mathfrak{M}_n$, *and decomposes as follows;*

$$\mathfrak{M}_n(k) \simeq \bigoplus_{\substack{\lambda \in \mathscr{P}_n \\ o(\lambda)=k}} \mathbf{B}_n(\lambda),$$

where $o(\lambda)$ *is the number of odd parts in* λ.

PROOF. Note that for $A \in \mathfrak{M}_n$ and $i \in I_n$, if $\mathbf{f}_i A \neq \mathbf{0}$, then

$$\mathrm{tr}(\mathbf{f}_i A) = \mathrm{tr}(A) \ \text{ or } \mathrm{tr}(A) \pm 2.$$

This implies that $\mathfrak{M}_n(k)$ together with $\mathbf{0}$ is stable under $\mathbf{e}_i$ and $\mathbf{f}_i$, and hence it is a subcrystal of $\mathfrak{M}_n$. Moreover, if A is a highest weight element in $\mathfrak{M}_n(k)$, that is, $P(A) = H_n^\lambda$ for some $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathscr{P}_n$, then A is a diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$, and $\mathfrak{o}(A) = k$ is equal to the number of odd parts in λ. □

In particular, we have

$$\mathfrak{M}_n(0) \simeq \bigoplus_{\substack{\lambda \in \mathscr{P}_n \\ \lambda : \text{even}}} \mathbf{B}_n(\lambda),$$

where a partition λ is called *even* if each part of λ is an even integer.

REMARK 5.8. A relation between the diagonal entries of a symmetric matrix and the shape of the corresponding tableau was first observed by Knuth [27]. The present exposition is a variation. (See [28] for a general statement including the result in [27]).

5.2. *Dual Robinson–Schensted–Knuth correspondence*

Let us now discuss the dual Robinson–Schensted–Knuth correspondence. Let $\mathbf{M}'_{m\times n}$ be the set of all $m \times n$ matrices with entries 0 or 1. We also assume that the row (resp. column) numbers are indexed by $\mathbf{B}_m$ (resp. $\mathbf{B}_n$).

As in the case of $\mathbf{M}_{m\times n}$, we identify a matrix in $\mathbf{M}'_{m\times n}$ with a biword reading the row and column indices of nonzero entries in the matrix with respect to a certain linear ordering. First, we set

$$\begin{aligned}\Gamma_{m,n} = \{ (\mathbf{i}, \mathbf{j}) \in \mathcal{W}_m \times \mathcal{W}_n \mid \\ (1)\ \mathbf{i} = i_1 \ldots i_r \text{ and } \mathbf{j} = j_1 \ldots j_r \text{ for some } r \geq 0, \\ (2)\ (i_1, j_1) \prec \cdots \prec (i_r, j_r) \},\end{aligned}$$

where for (i, j) and $(k, l) \in \mathbf{B}_m \times \mathbf{B}_n$, the linear ordering $\prec$ (right to left lexicographic ordering) is given by

$$(i, j) \prec (k, l) \quad \Longleftrightarrow \quad (j < l) \ \text{ or } (j = l \ \text{ and } \ i < k).$$

We define $\mathrm{e}_i, \mathrm{f}_i : \Gamma_{m,n} \to \Gamma_{m,n} \cup \{\mathbf{0}\}$ $(i \in I_m)$ by

$$\mathrm{e}_i(\mathbf{i}, \mathbf{j}) = (\mathrm{e}_i\mathbf{i}, \mathbf{j}), \quad \mathrm{f}_i(\mathbf{i}, \mathbf{j}) = (\mathrm{f}_i\mathbf{i}, \mathbf{j}),$$

for $(\mathbf{i}, \mathbf{j}) \in \Gamma_{m,n}$. Set $\mathrm{wt}(\mathbf{i}, \mathbf{j}) = \mathrm{wt}(\mathbf{i}) \in P_m$, $\varepsilon_i(\mathbf{i}, \mathbf{j}) = \varepsilon_i(\mathbf{i})$, and $\varphi_i(\mathbf{i}, \mathbf{j}) = \varphi_i(\mathbf{i})$ $(i \in I_m)$. Then $\Gamma_{m,n}$ is a $\mathfrak{gl}_m$-crystal, which is isomorphic to

$$\bigoplus_{0 \le \ell_1, \dots, \ell_n \le m} \mathbf{B}_m(1^{\ell_1}) \otimes \cdots \otimes \mathbf{B}_m(1^{\ell_n}),$$

where $\mathbf{B}_m(1^\ell)$ is a $\mathfrak{gl}_m$-crystal with highest weight $\epsilon_1 + \cdots + \epsilon_\ell$ for $0 \le \ell \le m$.

Next, we set

$$\begin{aligned}\Gamma^*_{m,n} = \{\, &(\mathbf{k}, \mathbf{l}) \in \mathcal{W}_m \times \mathcal{W}_n \mid \\ &(1)\ \mathbf{k} = k_1 \cdots k_r \text{ and } \mathbf{l} = l_1 \cdots l_r \text{ for some } r \ge 0, \\ &(2)\ (k_1, l_1) \prec' \cdots \prec' (k_r, l_r)\,\},\end{aligned}$$

where for (i, j) and $(k, l) \in \mathbf{B}_m \times \mathbf{B}_n$, the linear ordering $\prec'$ is given by

$$(i, j) \prec' (k, l) \iff (i > k) \text{ or } (i = k \text{ and } j < l).$$

Similarly, we define $\mathrm{e}^*_j, \mathrm{f}^*_j : \Gamma^*_{m,n} \to \Gamma^*_{m,n} \cup \{\mathbf{0}\}$ $(j \in I_n)$ by

$$\mathrm{e}^*_j(\mathbf{k}, \mathbf{l}) = (\mathbf{k}, \mathrm{e}_j\mathbf{l}), \quad \mathrm{f}^*_j(\mathbf{k}, \mathbf{l}) = (\mathbf{k}, \mathrm{f}_j\mathbf{l}),$$

for $(\mathbf{k}, \mathbf{l}) \in \Gamma^*_{m,n}$. Set $\mathrm{wt}^*(\mathbf{k}, \mathbf{l}) = \mathrm{wt}(\mathbf{l}) \in P_n$, $\varepsilon^*_j(\mathbf{k}, \mathbf{l}) = \varepsilon_j(\mathbf{l})$, and $\varphi^*_j(\mathbf{k}, \mathbf{l}) = \varphi_j(\mathbf{l})$ $(j \in I_n)$. Then $\Gamma^*_{m,n}$ is a $\mathfrak{gl}_n$-crystal isomorphic to

$$\bigoplus_{0 \le \ell_1, \dots, \ell_m \le n} \mathbf{B}_n(1^{\ell_1}) \otimes \cdots \otimes \mathbf{B}_n(1^{\ell_m}).$$

For $(\mathbf{i}, \mathbf{j}) \in \Gamma_{m,n}$ (resp. $(\mathbf{k}, \mathbf{l}) \in \Gamma^*_{m,n}$), we define $A(\mathbf{i}, \mathbf{j}) \in \mathbf{M}'_{m\times n}$ (resp. $A(\mathbf{k}, \mathbf{l}) \in \mathbf{M}'_{m\times n}$) in the same way as in 5.1, which yields a bijection between $\Gamma_{m,n}$ (resp. $\Gamma^*_{m,n}$) and $\mathbf{M}'_{m\times n}$. Under these maps, $\mathbf{M}'_{m\times n}$ becomes a $(\mathfrak{gl}_m, \mathfrak{gl}_n)$-bicrystal. For $A \in \mathbf{M}'_{m\times n}$, if we define $P(A) = P(\mathbf{i})$ and $Q(A) = P(\mathbf{l})$, where $A = A(\mathbf{i}, \mathbf{j}) = A(\mathbf{k}, \mathbf{l})$ for $(\mathbf{i}, \mathbf{j}) \in \Gamma_{m,n}$ and $(\mathbf{k}, \mathbf{l}) \in \Gamma^*_{m,n}$, then the map $A \mapsto (P(A), Q(A))$ gives an isomorphism of $(\mathfrak{gl}_m, \mathfrak{gl}_n)$-bicrystals from each connected component in $\mathbf{M}'_{m\times n}$ to $\mathbf{B}_m(\lambda) \times \mathbf{B}_n(\mu)$ for some $\lambda \in \mathscr{P}_m$ and $\mu \in \mathscr{P}_n$.

Now, we can check that $A = (a_{ij}) \in \mathbf{M}'_{m\times n}$ is a highest weight element if and only if it satisfies the following condition;

$$a_{ij} = 0 \Longrightarrow a_{i+1\,j} = a_{i\,j+1} = 0,$$

whenever $i + 1 \in \mathbf{B}_m$ and $j + 1 \in \mathbf{B}_n$.

Example 5.9. For $m = 4$ and $n = 8$,

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & & \\ 1 & 1 & 1 & 1 & & & & \\ 1 & & & & & & & \end{pmatrix}$$

is a highest weight element, and $(P(A), Q(A)) = (H_m^{\lambda}, H_n^{\lambda'})$, where $\lambda = (8, 6, 4, 1)$ and λ' is the conjugate of λ (cf [37]).

In general, for a highest weight element $A = (a_{ij}) \in \mathbf{M}'_{m\times n}$, put $\lambda = (\lambda_1, \ldots, \lambda_m)$ with $\lambda_i = \sum_{j=1}^{n} a_{ij}$ for $1 \le i \le m$. Then λ is a partition and $(P(A), Q(A)) = (H_m^{\lambda}, H_n^{\lambda'})$. Summarizing the aforementioned arguments, we have

THEOREM 5.10 ([4]). *The map $A \mapsto (P(A), Q(A))$ gives the following isomorphism of $(\mathfrak{gl}_m, \mathfrak{gl}_n)$-bicrystals;*

$$\mathbf{M}'_{m\times n} \simeq \bigoplus_{\substack{\lambda\in\mathscr{P}_m \\ \lambda'\in\mathscr{P}_n}} \mathbf{B}_m(\lambda) \times \mathbf{B}_n(\lambda').$$

REMARK 5.11. The $\mathfrak{gl}_n$-crystal structure on $\mathbf{M}'_{m\times n}$ is not obtained from its $\mathfrak{gl}_m$-crystal structure just by taking the transpose of a matrix, as in case of $\mathbf{M}_{m\times n}$. The reasons for this can be found in [28].

5.3. *Decomposition of Fock space representations*

Let $\mathfrak{gl}_\infty$ be the Lie algebra of $\mathbf{Z}\times\mathbf{Z}$-matrices over K with only finite number of nonzero entries. As an application of the dual Robinson–Schensted–Knuth correspondence, let us give an explicit decomposition of the n-fold tensor power of the Fock space representation of $\mathfrak{gl}_\infty$ or $U_q(\mathfrak{gl}_\infty)$.

The quantum group $U_q(\mathfrak{gl}_\infty)$ is defined in a similar way as $U_q(\mathfrak{gl}_n)$, where the weight lattice is $P = \bigoplus_{i\in\mathbf{Z}} \mathbf{Z}\epsilon_i$, the simple root is $\alpha_i = \epsilon_i - \epsilon_{i+1}$, the simple coroot h_i satisfies $\langle h_i, \lambda\rangle = (\alpha_i, \lambda)$ for all $\lambda \in P$ ($i \in \mathbf{Z}$), and so on. Let us denote by I the index set for the simple roots.

To consider integrable highest weight $U_q(\mathfrak{gl}_\infty)$-modules, we will use the extended weight lattice $\widehat{P}$ instead of P, where

$$\widehat{P} = \mathbf{Z}\Lambda_0 \oplus \bigoplus_{i\in\mathbf{Z}} \mathbf{Z}\epsilon_i,$$

and $\Lambda_0 \in \mathrm{Hom}\left(\bigoplus_{i\in I} \mathbf{Z}h_i, \mathbf{Z}\right)$ is given by $\langle h_i, \Lambda_0\rangle = \delta_{0i}$ for $i \in I$. For $c \in \mathbf{Z}$, let Λ_c be the fundamental weight, that is, $\Lambda_c(h_i) = \delta_{ic}$ for $i \in I$.

Then all the results in Section 2 hold for $U_q(\mathfrak{gl}_\infty)$, and we can thus define the notion of (abstract) $\mathfrak{gl}_\infty$-crystals in the same way. Now, consider the following $\mathfrak{gl}_\infty$-crystal $\mathbf{B}$;

$$\cdots \xrightarrow{-2} -1 \xrightarrow{-1} 0 \xrightarrow{0} 1 \xrightarrow{1} 2 \xrightarrow{2} \cdots,$$

where $\mathrm{wt}(b) = \epsilon_b$ for $b \in \mathbf{Z}$. Note that $\mathbf{B}$ is the crystal graph of the natural representation of $U_q(\mathfrak{gl}_\infty)$. Unlike $\mathfrak{gl}_n$-crystals, a connected component in $\mathbf{B}^{\otimes r}$ does not contain a highest weight element. So, we use the q-deformed Fermionic Fock space (or semi-infinite wedge space) representation $\mathscr{F}$ of $U_q(\mathfrak{gl}_\infty)$ [11] and its crystal graph (cf [38]) to describe the crystal graphs of integrable highest weight $U_q(\mathfrak{gl}_\infty)$-modules.

Let $\mathbf{F}$ be the set of *semi-infinite* words $w = \cdots w_{-3}w_{-2}w_{-1}w_0$ with letters in $\mathbf{B}$ such that

(1) $w_{i-1} < w_i$ for $i \le 0$,
(2) there exists an integer $c \in \mathbf{Z}$ such that $w_i = i + c$ for $i \ll 0$,

where c is called a *charge of* w. For example, the charge of $w = \cdots -5-4-1\,2\,4 \in \mathbf{F}$ is -1. For $w \in \mathbf{F}$, we define

$$\mathrm{wt}(w) = \Lambda_0 + \sum_{i\in\mathbf{B}} m_i\epsilon_i \in \widehat{P},$$

where $m_i = \big|\{\, k \mid w_k = i \,\}\big| - \delta_{-i,|i|}$. Because $m_i = 0$ for almost all $i \in \mathbf{B}$, $\mathrm{wt}(w)$ is well-defined. For each $i \in I$, we define the Kashiwara operators $\mathrm{e}_i, \mathrm{f}_i : \mathbf{F} \longrightarrow \mathbf{F} \cup \{\mathbf{0}\}$ as in the case of $\mathcal{W}_n$ (see 3.2). Since for $w = \cdots w_{-3}w_{-2}w_{-1}w_0 \in \mathbf{F}$ and $i \in I$, $\epsilon^{(i)}(w) = (\cdots, \epsilon^{(i)}(w_{-3}), \epsilon^{(i)}(w_{-2}), \epsilon^{(i)}(w_{-1}), \epsilon^{(i)}(w_0))$ has at most one $+$ or $-$ (we read the signs from left to right), e_i and f_i ($i \in I$) are well-defined. Set $\varepsilon_i(w) = \max\{\, k : \mathrm{e}_i^k w \neq \mathbf{0} \,\}$ and $\varphi_i(w) = \max\{\, k : \mathrm{f}_i^k w \neq \mathbf{0} \,\}$. Then $\mathbf{F}$ becomes a $\mathfrak{gl}_\infty$-crystal, and it is isomorphic to the crystal graph of the Fock space representation $\mathscr{F}$ (cf [38]).

Remark 5.12. Note that a semi-infinite word w in $\mathbf{F}$ of charge c can be identified with a partition $\lambda = (\lambda_k)_{k\ge1}$ such that $\lambda_k = w_{1-k} - c + k - 1$ for $k \ge 1$. For example,

$$w = \cdots -5-4-1\,1\,4 \quad\longleftrightarrow\quad \lambda = \begin{array}{l}\bullet\ \bullet\ \bullet\ \bullet\ \bullet\\ \bullet\ \bullet\ \bullet\\ \bullet\ \bullet\end{array}\ .$$

Let $\mathbf{B}(c)$ be the set of words in $\mathbf{F}$ of charge $c \in \mathbf{Z}$. Then $\mathbf{B}(c)$ is a subcrystal of $\mathbf{F}$ isomorphic to the crystal graph of the irreducible highest weight module with highest weight Λ_c (cf [38]). Hence, we have

$$\mathbf{F} = \bigoplus_{c\in\mathbf{Z}} \mathbf{B}(c),$$

where $\mathbf{B}(c)$ is a connected component with the highest weight element $H^c = \cdots\ c-3\ c-2\ c-1\ c$.

Let $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbf{Z}_+^n$ be a generalized partition of length n. We call an n-tuple of semi-infinite words $\mathbf{w} = (w^{(1)}, \cdots, w^{(n)})$ a *semi-infinite semistandard tableau of shape* λ if

(1) $w^{(i)} = \cdots w_{-2}^{(i)}w_{-1}^{(i)}w_0^{(i)} \in \mathbf{B}(\lambda_{n-i+1})$ for $1 \le i \le n$,
(2) $w_k^{(i)} \ge w_{k+d_i}^{(i+1)}$ for $1 \le i < n$ and $k \le 0$, where $d_i = \lambda_{n-i+1} - \lambda_{n-i}$.

We denote by $\mathbf{B}(\lambda)$ the set of all semi-infinite semistandard tableaux of shape λ. We may identify each $\mathbf{w} \in \mathbf{B}(\lambda)$ with a semi-infinite tableau with infinitely many rows and n columns, where each row of $\mathbf{w}$ reads (from left to right) as follows;

$$w_{k+d_1+\cdots+d_{n-1}}^{(n)} \cdots w_{k+d_1+d_2}^{(3)} w_{k+d_1}^{(2)} w_k^{(1)} \qquad (k \le 0).$$

EXAMPLE 5.13. Let $\lambda = (3, 1, -2, -3)$, and let $\mathbf{w} = (w^{(1)}, w^{(2)}, w^{(3)}, w^{(4)})$ be given by

$w^{(4)}$	$w^{(3)}$	$w^{(2)}$	$w^{(1)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
-4	-4	-4	-4
-3	-3	-2	-1
-2	-2	0	
-1	-1		
0	1		
2	3		
3			
5			

.

Then $\mathbf{w}$ is a semi-infinite semistandard tableau of shape λ.

For $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbf{Z}_+^n$ and $\mathbf{w} = (w^{(1)}, \ldots, w^{(n)}) \in \mathbf{B}(\lambda)$, we may assume that $\mathbf{w} = w^{(1)} \otimes \cdots \otimes w^{(n)} \in \mathbf{F}^{\otimes n}$, and consider $\mathbf{B}(\lambda)$ as a subset of $\mathbf{F}^{\otimes n}$. Then we can check that $\mathbf{B}(\lambda)$ together with $\mathbf{0}$ is stable under e_i and f_i ($i \in I$). Hence, $\mathbf{B}(\lambda)$ is a $\mathfrak{gl}_\infty$-crystal. Moreover, it is connected with the unique highest weight element $H^\lambda = H^{\lambda_n} \otimes \cdots \otimes H^{\lambda_1}$.

THEOREM 5.14. *For $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbf{Z}_+^n$, $\mathbf{B}(\lambda)$ is a $\mathfrak{gl}_\infty$-crystal, which is isomorphic to the crystal graph of the irreducible highest weight module with highest weight $\Lambda_{\lambda_1} + \cdots + \Lambda_{\lambda_n}$.*

REMARK 5.15. Since each word in $\mathbf{F}$ can be identified with a partition, a semi-infinite semistandard tableau in $\mathbf{B}(\lambda)$ can be identified with an n-tuple of partitions where the semistandard condition can be expressed in terms of certain inclusions among the associated partitions. This identification naturally gives rise to a realization of the crystal graph of an irreducible highest weight representation of an affine quantum group (cf [8, 16, 38]). Here we will use the tableaux realization of $\mathbf{B}(\lambda)$ to apply the combinatorics of Young tableaux given in the previous sections.

Our goal is to give an explicit decomposition of $\mathbf{F}^{\otimes n}$ for $n \geq 2$. We start with another description of $\mathbf{F}^{\otimes n}$ in terms of matrices of nonnegative integers. Set

$$\begin{aligned}\mathbf{M}'_{\infty\times n} = \{A = (a_{ij})_{i\in\mathbf{B},\, j\in\mathbf{B}_n} \mid \\ (1)\ a_{ij} \in \{0, 1\}, \\ (2)\ a_{ij} = 1 \text{ for all } i \ll 0, \text{ and } a_{ij} = 0 \text{ for all } i \gg 0\}.\end{aligned}$$

For $\mathbf{w} = w^{(1)} \otimes \cdots \otimes w^{(n)} \in \mathbf{F}^{\otimes n}$, set $A(\mathbf{w}) = (a_{ij}) \in \mathbf{M}'_{\infty\times n}$, where a_{ij} is the number of occurrences of i in $w^{(j)}$ ($1 \leq j \leq n$). The map $\mathbf{w} \mapsto A(\mathbf{w})$ is a bijection from $\mathbf{F}^{\otimes n}$ to $\mathbf{M}'_{\infty\times n}$, where each $w^{(j)}$ corresponds to the j-th column of $A(\mathbf{w})$ for $1 \leq j \leq n$.

EXAMPLE 5.16.

$$\mathbf{F}^{\otimes 3} \ni \mathbf{w} = \begin{array}{ccc} \vdots & \vdots & \vdots \\ -4 & -4 & -4 \\ -3 & -3 & -1 \\ -2 & -1 & 0 \\ 2 & 3 & 1 \\ 3 & & 2 \\ & & 3 \end{array} \longrightarrow A(\mathbf{w}) = \left(\begin{array}{ccc} \vdots & \vdots & \vdots \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \end{array}\right) \in \mathbf{M}'_{\infty\times 3},$$

where the horizontal line is between the 0-th row and the 1-st row of the matrix.

Now define a bicrystal structure on $\mathbf{M}'_{\infty\times n}$, which is naturally induced from the one on $\mathbf{M}'_{m\times n}$ in 5.2. For $m > 0$, we define

$$\mathrm{Res}_m : \mathbf{M}'_{\infty\times n} \longrightarrow \mathbf{M}'_{2m+1\times n}$$

by $\mathrm{Res}_m A = (a_{ij})_{-m\le i\le m,\, j\in I_n}$ for $A = (a_{ij}) \in \mathbf{M}'_{\infty\times n}$, where we assume that instead of $\mathbf{B}_{2m+1}$ and I_{2m+1}, the row index set for $\mathbf{M}'_{2m+1\times n}$ and the index set for the simple roots of $\mathfrak{gl}_{2m+1}$ are given by $\{-m, \dots, -1, 0, 1, \dots, m\}$ and $\{-m+1, \dots, 0, \dots, m-1\}$, respectively. For $A \in \mathbf{M}'_{\infty\times n}$ and $i \in I$, we define $\mathrm{e}_i A$ and $\mathrm{f}_i A$ to be the unique elements in $\mathbf{M}'_{\infty\times n} \cup \{\mathbf{0}\}$ satisfying

$$\mathrm{Res}_m(\mathrm{e}_i A) = \mathrm{e}_i(\mathrm{Res}_m A), \quad \mathrm{Res}_m(\mathrm{f}_i A) = \mathrm{f}_i(\mathrm{Res}_m A),$$

for all sufficiently large $m > |i|$, where $\mathrm{Res}_m \mathbf{0} = \mathbf{0}$. We set

$$\mathrm{wt}(A) = n\Lambda_0 + \sum_{i\in\mathbf{B}} m_i\epsilon_i \in \widehat{P}, \quad \text{where } m_i = \sum_{j\in\mathbf{B}_n} (a_{ij} - \delta_{-i,|i|}),$$

$$\varepsilon_i(A) = \max\{k : \mathrm{e}_i^k A \neq \mathbf{0}\}, \quad \varphi_i(A) = \max\{k : \mathrm{f}_i^k A \neq \mathbf{0}\}.$$

Then $\mathbf{M}'_{\infty\times n}$ is a $\mathfrak{gl}_\infty$-crystal, and the map $\mathbf{w} \mapsto A(\mathbf{w})$ is an isomorphism of $\mathfrak{gl}_\infty$-crystals from $\mathbf{F}^{\otimes n}$ to $\mathbf{M}'_{\infty\times n}$ since each column in $A \in \mathbf{M}'_{\infty\times n}$ can be identified with a word in $\mathbf{F}$.

Next, let us define a $\mathfrak{gl}_n$-crystal structure on $\mathbf{M}'_{\infty\times n}$. Since $\mathbf{M}'_{\infty\times n}$ can be identified with $\mathbf{F}^{\otimes n}$, this will induce a $\mathfrak{gl}_n$-crystal structure on $\mathbf{F}^{\otimes n}$. For $A \in \mathbf{M}'_{\infty\times n}$ and $j \in I_n$, define $\mathrm{e}_j^* A$ and $\mathrm{f}_j^* A$ to be the unique elements in $\mathbf{M}'_{\infty\times n} \cup \{\mathbf{0}\}$ satisfying

$$\mathrm{Res}_m(\mathrm{e}_j^* A) = \mathrm{e}_j^*(\mathrm{Res}_m A), \quad \mathrm{Res}_m(\mathrm{f}_j^* A) = \mathrm{f}_j^*(\mathrm{Res}_m A),$$

for all sufficiently large $m > 0$. Set

$$\mathrm{wt}^*(A) = \sum_{j \in \mathbf{B}_n} m_j \epsilon_j \in P_n, \quad \text{where } m_j = \sum_{i \in \mathbf{B}} (a_{ij} - \delta_{-i,|i|}),$$

$$\varepsilon_j^*(A) = \max\{k : (\mathrm{e}_j^*)^k A \neq \mathbf{0}\}, \quad \varphi_j^*(A) = \max\{k : (\mathrm{f}_j^*)^k A \neq \mathbf{0}\}.$$

Then $\mathbf{M}'_{\infty \times n}$ is a $\mathfrak{gl}_n$-crystal. Since the crystal structures on $\mathbf{M}'_{\infty \times n}$ are induced from those on $\mathbf{M}'_{2m+1 \times n}$, it is clear that $\mathbf{M}'_{\infty \times n}$ is a $(\mathfrak{gl}_\infty, \mathfrak{gl}_n)$-bicrystal.

Suppose that $A = (a_{ij}) \in \mathbf{M}'_{\infty \times n}$ is given and $A = A(\mathbf{w})$ for $\mathbf{w} = w^{(1)} \otimes \cdots \otimes w^{(n)} \in \mathbf{F}^{\otimes n}$. Identifying a semi-infinite word in $\mathbf{F}$ with a semi-infinite semistandard tableau of a single column, we can define an algorithm of inserting a semi-infinite word in $\mathbf{F}$ into another one in a similar way as in 4.3 so that we obtain a semi-infinite semistandard tableau with two columns. Now we define the P-tableau of A to be

$$P(A) = (((\cdots (w^{(1)} \leftarrow w^{(2)}) \cdots) \leftarrow w^{(n)}),$$

which is a semi-infinite semistandard tableau with n columns[5]. Then $P(A)$ is $\mathfrak{gl}_\infty$-crystal equivalent to A.

Example 5.17. Let A be the matrix given in Example 5.16. Then

$$P(A) = \begin{array}{c} \vdots \\ -4 \\ -3 \\ -2 \\ 2 \\ 3 \end{array} \leftarrow \begin{array}{c} \vdots \\ -4 \\ -3 \\ -1 \\ 3 \end{array} \leftarrow \begin{array}{c} \vdots \\ -4 \\ -1 \\ 0 \\ 1 \\ 2 \\ 3 \end{array} = \begin{array}{ccc} & \vdots & \\ -4 & -4 & -4 \\ -3 & -3 & 2 \\ -2 & -1 & 3 \\ -1 & 3 & \\ 0 & & \\ 1 & & \\ 2 & & \\ 3 & & \end{array}.$$

Next, define the Q-tableau of A as follows. Choose a sufficiently large m such that $a_{ij} = 1$ for all $i \leq -m$ and $a_{ij} = 0$ for all $i \geq m$, and let Q_m be the unique semistandard tableau which is $\mathfrak{gl}_n$-crystal equivalent to $\mathrm{Res}_m A$, that is, $Q_m = Q(\mathrm{Res}_m A)$. Then the connected components of A and Q_m are isomorphic as I_n-colored oriented graphs, but $\mathrm{wt}^*(A) = \mathrm{wt}(Q_m) - (m+1)(\epsilon_1 + \cdots + \epsilon_n)$. So, if we define the Q-tableau of A to be

$$Q(A) = \sigma^{-m-1}(Q_m)$$

(see 3.4), then $Q(A)$ is a rational semistandard tableau, which is $\mathfrak{gl}_n$-crystal equivalent to A. Note that $Q(A)$ is independent of the choice of m.

[5] More precisely, $P(A)$ is the unique semi-infinite semistandard tableau such that $\mathrm{Res}_m(P(A)) = P(\mathrm{Res}_m(A))$ for all $m \gg 0$, where $P(A)$ is identified with its matrix form in $\mathbf{M}'_{\infty \times n}$.

EXAMPLE 5.18. Let A be as in Example 5.16. Then

$$\mathrm{Res}_3 A = \left(\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{array}\right) \in \mathbf{M}'_{7\times 3},$$

and

$$Q_3 = Q(\mathrm{Res}_3 A) = \begin{array}{ccccccccc} 1 & 1 & 1 & 1 & 3 & 3 & 3 \\ 2 & 2 & 2 & & & & \\ 3 & 3 & & & & & \end{array}.$$

Therefore,

$$Q(A) = \sigma^{-4}(Q_3) = \left|\begin{array}{ccc} & & \\ & -3 & \\ -3 & -2 & \end{array}\right| \begin{array}{ccc} 3 & 3 & 3 \\ & & \end{array}.$$

It can now be concluded that each connected component in $\mathbf{M}'_{\infty\times n}$ as a $(\mathfrak{gl}_\infty, \mathfrak{gl}_n)$-bicrystal is isomorphic to $\mathbf{B}(\lambda) \times \mathbf{B}_n(\mu)$ for some λ and $\mu \in \mathbf{Z}_+^n$. Moreover, $A = (a_{ij}) \in \mathbf{M}'_{\infty\times n}$ is of highest weight if and only if

$$a_{ij} = 0 \Longrightarrow a_{i+1\,j} = 0, \quad \text{and} \quad a_{i\,j+1} = 0 \ (j < n)$$

(see 5.2). For a highest weight element $A \in \mathbf{M}'_{\infty\times n}$, if λ_j is set to be

$$\lambda_j = \sum_{i\in\mathbf{B}} (a_{ij} - \delta_{-i,|i|}) \quad (j \in \mathbf{B}_n),$$

then $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbf{Z}_+^n$ and $(P(A), Q(A)) = (H^\lambda, H_n^\lambda)$.

Summarizing, we obtain the decomposition of $\mathbf{F}^{\otimes n}$ and hence $\mathscr{F}^{\otimes n}$, which can be viewed as a semi-infinite analogue of the Robinson–Schensted–Knuth correspondence.

THEOREM 5.19. *The map* $\mathbf{w} \mapsto (P(A(\mathbf{w})), Q(A(\mathbf{w})))$ *gives the following isomorphism of* $(\mathfrak{gl}_\infty, \mathfrak{gl}_n)$*-bicrystals;*

$$\mathbf{F}^{\otimes n} \simeq \bigoplus_{\lambda\in\mathbf{Z}_+^n} \mathbf{B}(\lambda) \times \mathbf{B}_n(\lambda).$$

REMARK 5.20. The decomposition of the corresponding representation of Lie algebras was first obtained by Frenkel [44] (see also [18]).

References

[1] T.H. Baker, An insertion scheme for C_n crystals, Prog. Math. **191** (2000), 1–48.

[2] G. Benkart, S.-J. Kang, M. Kashiwara, Crystal bases for the quantum superalgebra $U_q(\mathfrak{gl}(m,n))$, J. Am. Math. Soc. **13** (2000), 295–331.

[3] V. Chari, A. Pressley, *A Guide to Quantum Groups*, Cambridge University Press, Cambridge, 1994.

[4] V.I. Danilov, G.A. Koshevoy, Bi-crystals and crystal $(GL(V), GL(W))$ duality, RIMS Preprint no. 1485, (2004).

[5] V.I. Danilov, G.A. Koshevoy, Arrays and the combinatorics of Young tableaux, Russian Math. Surv. **60** (2005), 269–334.

[6] E. Date, M. Jimbo, T. Miwa, Representations of $U_q(\mathfrak{gl}(n,\mathbb{C}))$ at $q=0$ and the Robinson-Schensted correspondence, in: L. Brink, D. Friedan, A.M. Polyakov eds., *Physics and Mathematics of Strings, Memorial Volume of V. Knizhnik*, World Scientific, 1990, 185–211.

[7] V.G. Drinfel'd, Hopf algebras and the quantum Yang-Baxter equations, Soviet Math. Dokl. **32** (1985), 254–258.

[8] O. Foda, B. Leclerc, M. Okado, J.-Y. Thibon, T.A. Welsh, Branching functions of $A^{(1)}_{n-1}$ and Jantzen-Seitz problem for Ariki-Koike algebras, Adv. Math. **141** (1999), 322–365.

[9] I. Frenkel, Representations of affine Lie algebras, Hecke modular forms and Korteweg-de Vries type equations, Lect. Notes Math. **933** (1982), 71–110.

[10] W. Fulton, *Young tableaux, London Mathematical Society Student Texts*, 35, Cambridge University Press, Cambridge, 1997.

[11] T. Hayashi, q-analogues of Clifford and Weyl algebras—Spinor and oscillator representations of quantum enveloping algebras, Comm. Math. Phys. **127** (1990), 129–144.

[12] J. Hong, S.-J. Kang, Quantum groups and crystal bases, Graduate Stud. Math. **42**, Amer. Math. Soc., 2002.

[13] R. Howe, Remarks on classical invariant theory, Trans. Am. Math. Soc. **313** (1989), 539–570.

[14] J.C. Janzten, Lectures on quantum groups, Graduate Stud. Math. **6**, Amer. Math. Soc., 1996.

[15] M. Jimbo, A q-difference analogue of $U(\mathfrak{g})$ and the Yang-Baxter equations, Lett. Math. Phys. **10** (1985), 63–69.

[16] M. Jimbo, K. Misra, T. Miwa, M. Okado, Combinatorics of representations of $U_q(\widehat{\mathfrak{sl}(n)})$ at $q=0$, Comm. Math. Phys. **136** (1991), 543–566.

[17] V.G. Kac, *Infinite Dimensional Lie Algebras*, 3rd ed., Cambridge University Press, Cambridge, 1990.

[18] V.G. Kac, A. Radul, Representation theory of the vertex algebra $W_{1+\infty}$, Transform. Groups **1** (1996), 41–70.

[19] S.-J. Kang, J.-H. Kwon, Tensor product of crystal bases for $U_q(\mathfrak{gl}(m,n))$-modules, Comm. Math. Phys. **224** (2001), 705–732.

[20] S.J. Kang, M. Kashiwara, K. Misra, T. Miwa, T. Nakashima, A. Nakayashiki, Perfect crystal of quantum affine Lie algebra, Duke Math. J. **68** (1992), 499–607.

[21] M. Kashiwara, Crystalizing the q-analogue of universal enveloping algebras, Comm. Math. Phys. **133** (1990), 249–260.

[22] M. Kashiwara, On crystal bases of the q-analogue of universal enveloping algebras, Duke Math. J. **63** (1991), 465–516.

[23] M. Kashiwara, Crystal bases and Littelmann's refined Demazure character formula, Duke Math. J. **71** (1993), 839–858.

[24] M. Kashiwara, On crystal bases, Representations of groups, CMS Conference Proceedings, American Mathematics Society, Providence, RI, 1995, pp. 155–197.

[25] M. Kashiwara, Realizations of crystals, Contemp. Math. **325** (2003), 133–139.

[26] M. Kashiwara, T. Nakashima, Crystal graphs for representations of the q-analogue of classical Lie algebras, J. Algebra **165** (1994), 295–345.

[27] D. Knuth, Permutations, matrices, and the generalized Young tableaux, Pacific J. Math. **34** (1970), 709–727.

[28] J.-H. Kwon, Crystal graphs for Lie superalgebras and Cauchy decompositions, J. Algebraic Combin. **25** (2007), 57–100.
[29] A. Lascoux, Double crystal graphs, Prog. Math. **210** (2003), 95–114.
[30] A. Lascoux, M.P. Schützenberger, Le monoïde plaxique in non-commutative structures in algebra and geometric combinatorics, Quad. "Ricerca Sci." **109** (1981), 129–156.
[31] B. Leclerc, J.-Y. Thibon, The Robinson-Schensted correspondence, crystal bases, and the quantum straightening at $q = 0$, Electron. J. Combin. **3** (1996), no. 2, Research Paper 11, approx. 24 pp. (electronic).
[32] C. Lecouvey, Schensted-type correspondence, plactic monoid, and jeu de taquin for type C_n, J. Algebra **247** (2002), 295–331.
[33] C. Lecouvey, Schensted-type correspondences and plactic monoids for types B_n and D_n, J. Algebraic Combin. **18** (2003), 99–133.
[34] P. Littelmann, Paths and root operators in representation theory, Ann. Math. **142** (1995), 499–525.
[35] P. Littelmann, A plactic algebra for semisimple Lie algebras, Adv. Math. **124** (1996), 312–331.
[36] G. Lusztig, *Introduction to Quantum Groups*, Progress in Math. **110**, Birkhäuser, 1993.
[37] I.G. Macdonald, *Symmetric functuins and Hall polynomials*, 2nd ed., Oxford University Press, New York, 1995.
[38] K. Misra, T. Miwa, Crystal base for the basic representation of $U_q(\widehat{\mathfrak{sl}}(n))$, Comm. Math. Phys. **134** (1990), 79–88.
[39] H. Nakajima, t-analogs and q-characters of quantum affine algebras of type A_n and D_n, Contemp. Math. **325** (2003), 141–160.
[40] T. Nakashima, Crystal base and a generalization of Littlewood-Richardson rule for the clasical Lie algebras, Comm. Math. Phys. **154** (1993), 215–243.
[41] B. Sagan, *The Symmetric Group–Represenations, Combinatorial Algorithms, and Symmetric Functions*, Wadworth and Brooks, Cole, 1991.
[42] J.R. Stembridge, Rational tableaux and the tensor algebra of $\mathfrak{gl}_n$, J. Combin. Theory Ser. A **46** (1987), 79–120.
[43] S. Sundaram, Tableaux in the representation theory, in: D. Stanton, Ed., *Invariant Theory and Tableaux (Minneapolis, MN, 1988)*, IMA Vol. Math. Appl., Vol. **19**, Springer-Verlag, New York, 1990, 191–225.
[44] G.P. Thomas, On Schensted's construction and the multiplication of Schur functions, Adv. Math. **30** (1978), 8–32.

Quivers and Representations*

Xueqing Chen and Ki–Bong Nam

Department of Mathematical and Computer Sciences,
University of Wisconsin–Whitewater, Whitewater, WI53190, USA
E-mail: chenx@uww.edu; namk@uww.edu

Tomáš Pospíchal

Department of Mathematics and Statistics,
University of Windsor, Windsor, ON N9B 3P4, Canada
E-mail: tpospich@uwindsor.ca

Contents

* Dedicated to Professor Hazewinkel on the occasion of his 65th birthday.

HANDBOOK OF ALGEBRA, VOL. 6
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(08)00209-X

1. Introduction

Why quivers? In this introductory section, we provide some motivation for the use of the language of quivers in mathematics. Quivers arise in many areas of mathematics, including representation theory, algebraic geometry, and differential geometry, Kac–Moody algebras, and quantum groups.

A **quiver** is just a directed graph, and a **representation** of a quiver associates a vector space to each vertex and a linear map to each arrow. The notions of quiver and its representations were introduced by P. Gabriel in the early 1970s of the past century. The introduction of quivers marked the starting point of the modern representation theory of finite-dimensional associative algebras. A number of remarkable connections to other mathematical fields have been discovered, in particular to Hall algebras, quantum groups, elliptic Lie algebras, and more recently cluster algebras.

In the following short overview of these topics, we highlight some of the main contributors.

Representations of finite-dimensional algebras: (P. Gabriel, M. Auslander, I. Reiten, V. Dlab, C. M. Ringel, V. Kac et al.).

The representation theory of finite-dimensional algebras can be formulated using the language of quivers. A basic finite-dimensional hereditary algebra over an algebraically closed field k is isomorphic to a path algebra $k\mathcal{Q}$ associated to a quiver $\mathcal{Q}$. If it is basic but not hereditary, it is isomorphic to a certain quotient algebra of a path algebra $k\mathcal{Q}$, where the quiver $\mathcal{Q}$ depends only on homological properties of the simple modules of the algebra.

In effect, the study of representations of a finite-dimensional algebra over an algebraically closed field is reduced to the study of a quotient of a path algebra of a certain quiver. If the base field is not algebraically closed, one may use "species" (or valued graphs) introduced by V. Dlab and C. M. Ringel in [1] instead of quivers to realize finite-dimensional algebras. Methods based on quivers and their representations have been used in the past 30 years extensively to describe the structure of length categories (Abelian categories where every object has a finite composition series), which arise very frequently in algebra. These techniques provide much better understanding of the indecomposable objects and sometimes allow a fairly complete description of the category.

Ringel–Hall algebras: (C. M. Ringel, J. A. Green).

Hall algebras were introduced by E. Steinitz and P. Hall in the context of counting subgroups of finite Abelian groups. In a broad sense, a Hall algebra is a tool allowing one to code a category. C. M. Ringel realized that Hall algebras provide a link between the representation theory of quivers on one side and Lie algebras and quantum groups on the other. This discovery had a profound influence on representation theory.

The **Ringel–Hall algebra** is an algebra, which has a basis labeled by isomorphism classes of modules of a finite-dimensional hereditary algebra (or, more specially, representations of a quiver). The "structure constants" describing the multiplicative structure (relative to this basis) reflect the structure of the underlying modules. The Ringel–Hall algebra encodes a large amount of information about the modules of finite-dimensional hereditary algebra. Amazingly, the generic composition algebra of the Ringel–Hall algebra is isomorphic to the positive part of a quantum group. Via the Ringel–Hall algebra approach, J. Xiao and L. Peng [2] provided a realization of all symmetrizable Kac–Moody algebras, and Y. Lin and L. Peng [3] obtained a realization of some elliptic Lie algebras.

Canonical bases of quantum groups: (G. Lusztig).

A **canonical basis** of the quantum group $\mathbf{U}_v^+(\mathfrak{g})$ was first introduced by G. Lusztig in [4] via an elementary algebraic definition when $\mathfrak{g}$ is a simple finite-dimensional Lie algebra. Later, he examined this basis in the framework of Ringel–Hall algebras. This canonical basis provides a natural setting for classical combinatorial results in the representation theory of simple Lie algebras. The original algebraic definition does not work for quantum groups of arbitrary Kac–Moody algebras. The difficulty in constructing the basis arises from the need to define suitable analogs of imaginary root vectors. By using topological methods [5], G. Lusztig constructed this canonical basis of the quantum group of a symmetric Kac–Moody algebra. It was shown that the "global crystal basis" introduced by M. Kashiwara and this canonical basis coincide. This canonical basis has very nice properties, which make it a powerful tool for the study of representation theory of quantum groups, Hecke algebras, quantized Schur algebra, etc. It is particularly suited for the investigation of the structure of integrable modules of quantum groups and their tensor products.

Quiver variety: (G. Lusztig and H. Nakajima).

By introducing **quiver variety**, G. Lusztig provided a geometric realization of the positive part of the universal enveloping algebra of a simply laced Kac–Moody algebra $\mathfrak{g}$ using constructible functions. A geometric realization of the positive part of a quantum group using perverse sheaf permitted the construction of a canonical basis, which plays a major role in representation theory. The quiver varieties introduced by H. Nakajima yield a geometric realization of the highest weight representations of simply laced Kac–Moody algebras.

Category $\mathcal{O}$ and stratified algebras: (E. Cline, B. Parshall, and L. Scott).

Quasi-hereditary algebras were introduced by E. Cline, B. Parshall, and L. Scott in their efforts to understand Kazhdan–Lusztig theory and developed

further by V. Dlab and C. M. Ringel. Quasi-hereditary algebras and their generalizations may be viewed as generalizations of quiver algebras more appropriate for understanding the relationship between standard modules and irreducible modules in the category $\mathcal{O}$, perverse sheaves, and categories of representations of algebraic groups in positive characteristic.

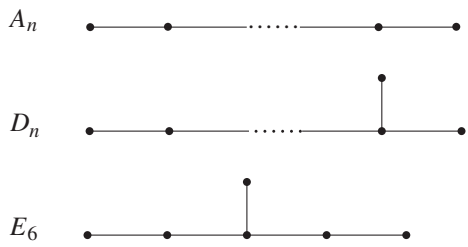
Some of the most interesting generalizations of quasi-hereditary algebras, called **stratified algebras**, have been defined in independent (and slightly different) ways by I. Ágoston, V. Dlab and E. Lukács, and E. Cline, B. Parshall, and L. Scott. In each case, the stratification involved occurs categorically at the derived category level and internally in terms of particular ideals of the algebra.

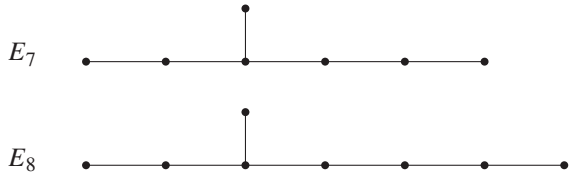
Weighted projective lines and canonical algebras: (W. Geigle and H. Lenzing).

W. Geigle and H. Lenzing in [6] introduced a new class of curves called **weighted projective lines**. These curves are essentially lines in a weighted projective space and behave like smooth projective curves with respect to coherent sheaves and vector bundles. In particular, the category of coherent sheaf coh $\mathfrak{C}$ over a weighted projective line $\mathfrak{C}$ has Serre duality, almost-split sequences, and each coherent sheaf splits into a direct sum of a vector bundle and a torsion sheaf.

There is a bijective correspondence between weighted projective lines and **canonical algebras** as studied by C. M. Ringel in [7]. By means of a tilting sheaf, one has an equivalence $D^b(\mathrm{coh}\,\mathfrak{C}) \simeq D^b(\mathrm{mod}-\Lambda)$ of the derived categories of coh $\mathfrak{C}$ and the category mod $-\Lambda$ of finite-dimensional Λ–modules, where Λ is the canonical algebra associated to $\mathfrak{C}$. In the case where the virtual genus of $\mathfrak{C}$ is one, a brief account of the classification of indecomposable bundles on C can be given. The equivalence $D^b(\mathrm{coh}\,\mathfrak{C}) \simeq D^b(\mathrm{mod}-\Lambda)$ establishes a link between M. F. Atiyah's classification of vector bundles on elliptic curves and C. M. Ringel's classification of modules over canonical algebras of tubular type.

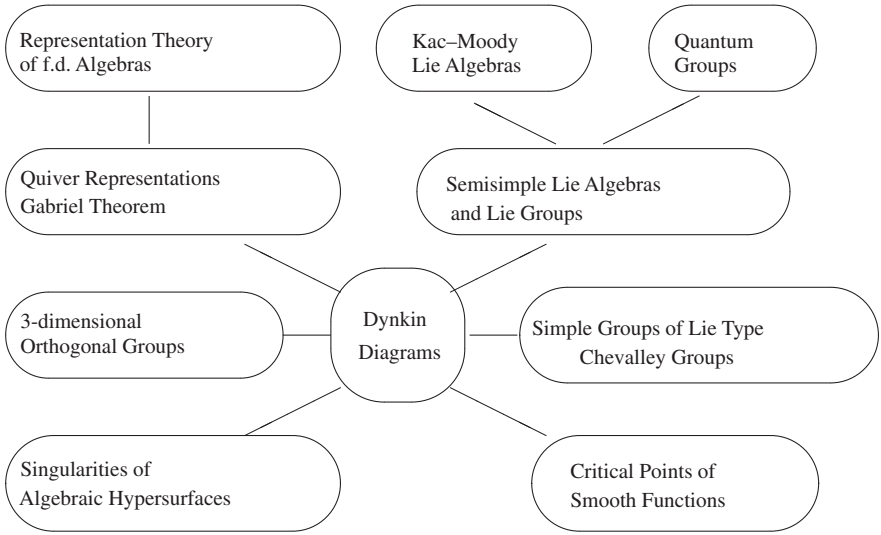
Dynkin diagrams of type A, D, and E:





The **Dynkin diagrams** A_n, D_n, E_6, E_7, and E_8 have appeared in many different parts of mathematics.

The A–D–E classifications: In accordance with the ubiquity of *A–D–E* classifications [8, 9], the following graph shows some of the main occurrences of Dynkin diagrams in mathematics.



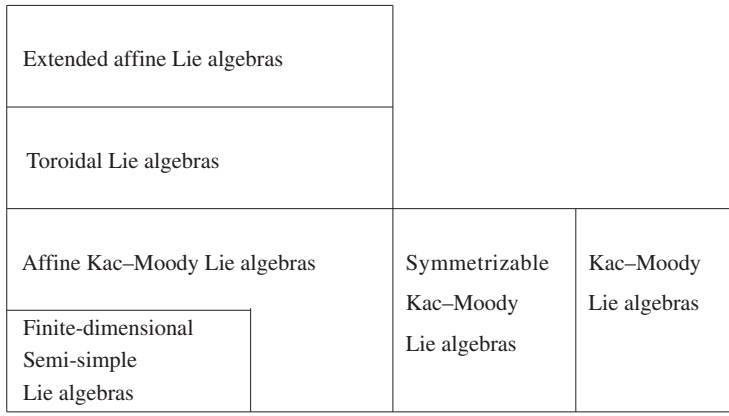
As V.I. Arnold said in [8]:

> The problem is to find a common origin of all *A–D–E* classification theorems and to substitute a priori proofs to a posteriori verifications of the parallelism of the classifications.

Are there direct relationship among some fields of mathematics where the same diagrams occur? Can results already developed in one field be applied to get new result and new information in another field?

In the following sections, you will see why these diagrams have come up in the representation theory of finite-dimensional algebras and related topics and how the representation theory of finite-dimensional algebras is connected with other fields of mathematics.

Classes of Lie algebras: The class of semisimple Lie algebras has been extended to several more general classes. We list some of those related to the topic of this chapter without mentioning details.

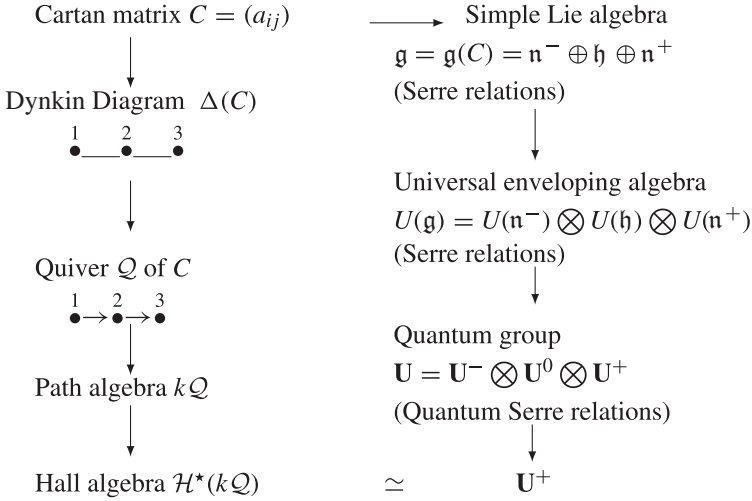


The finite-dimensional algebras can be classified with respect to their representation types. Every algebra is either of finite type, tame type, or wild type. In this chapter, the close connection between the classification of finite-dimensional algebras and the classification of Kac–Moody algebras will be given.

The main problems of the representation theory of finite-dimensional algebras are:

- to study the structure of the given algebra;
- to describe all indecomposable modules and their homological properties;
- to classify the given algebra.

The following graph (c.f. [10]) illustrates the relations between the representation theory of quivers and Lie theory, and also provides a guideline for the following sections.



In Section 2, we recall the basic terminology of Kac–Moody algebras and quantum groups, which is needed in the following sections. We introduce the language of quivers and their representations in Section 3 and describe the relation between quivers and finite-dimensional algebras. Section 4 provides the first connection between the representation theory of quivers and Lie theory; Gabriel theorem is explained. We give the definition of the Ringel–Hall algebra and discuss the amazing isomorphism between the generic composition algebra and the positive part of the corresponding quantum group in Section 5. In Section 6, by using the Ringel–Hall algebra approach, we construct root vectors, the Poincaré–Birkhoff–Witt (PBW) basis, and the canonical basis of a quantum group of finite or affine type. Section 7 provides a realization of all symmetrizable Kac–Moody algebras and some elliptic Lie algebras of type $D_4^{(1,1)}$, $E_6^{(1,1)}$, $E_7^{(1,1)}$, and $E_8^{(1,1)}$ via the root category of hereditary algebras and tubular algebras of the same type, respectively.

Throughout this chapter, the composition of any two maps $X \xrightarrow{f} Y \xrightarrow{g} Z$ is denoted by gf and the set of non-negative integers is denoted by $\mathbf{N}$.

2. Kac–Moody algebras and quantum groups

We will use the Cartan datum to unify a variety of set-ups in different mathematical fields.

Definition 2.0.1 (G. Lusztig). A **Cartan datum** is a pair $\Delta = (I, (\ ,\ ))$ where

- I is a finite set;
- $(\ ,\ )$ is a symmetric bilinear form on $\mathbf{Z}^I$.

It is assumed that

- $(i, i) \in \{2, 4, 6, \dots\}$ for any $i \in I$;
- $2\dfrac{(i, j)}{(i, i)} \in \{0, -1, -2, \dots\}$ for any $i \neq j$ in I.

A Cartan datum Δ is **symmetric** Cartan datum if $(i, i) = 2$ for all $i \in I$; we call Δ **simply laced** if it is symmetric and $(i, j) \in \{0, -1\}$ for all $i \neq j$. A Cartan datum Δ is said to be **irreducible** Cartan datum if I is nonempty and for any $i \neq j$ in I, there exists a sequence $i = i_1, i_2, \dots, i_m = j$ in I such that $(i_p, i_{p+1}) < 0$ for $p = 1, 2, \dots, m - 1$.

A matrix C is called **symmetrizable** matrix if there exists an invertible diagonal matrix D such that DC is symmetric. Note that any symmetric matrix is symmetrizable.

Every Cartan datum determines a symmetrizable (generalized) Cartan matrix in the following way: write $a_{ij} = 2\dfrac{(i, j)}{(i, i)}$, and let $C = (a_{ij})_{i,j\in I}$. In the following until the end of this section, assume $I = \{1, \dots, n\}$.

DEFINITION 2.0.2. A **realization** of a symmetrizable (generalized) Cartan matrix $C = (a_{ij})_{1\le i,j\le n}$ is a triple $(\mathfrak{h}, \Pi, \Pi^\vee)$, where $\mathfrak{h}$ is a complex vector space, $\Pi = \{\alpha_1, \alpha_2, \ldots, \alpha_n\} \subset \mathfrak{h}^*$, and $\Pi^\vee = \{h_1, h_2, \ldots, h_n\} \subset \mathfrak{h}$ such that

(1) both Π and $\Pi^\vee$ are linearly independent;
(2) $\dim \mathfrak{h} = 2n - \operatorname{rank}(A)$;
(3) $\langle h_i, \alpha_j \rangle = a_{ij}$ for all $i, j \in I$,

where $\langle\ ,\ \rangle : \mathfrak{h} \times \mathfrak{h}^* \longrightarrow \mathbf{C}$, $\langle h, \alpha\rangle = \alpha(h)$.

DEFINITION 2.0.3. Let $C = (a_{ij})_{1\le i,j\le n}$ be a symmetrizable (generalized) Cartan matrix. The associated **Kac–Moody algebra** $\mathfrak{g} = \mathfrak{g}(C)$ is defined as the complex Lie algebra with generators e_i, f_i $(i = 1, \ldots, n)$ and $\mathfrak{h}$ and relations:

$$
\begin{aligned}
&[h, h'] = 0, \qquad \forall\ h, h' \in \mathfrak{h};\\
&[e_i, f_j] = \delta_{ij} h_i;\\
&[h, e_i] = \alpha_i(h) e_i, \qquad \forall\ h \in \mathfrak{h};\\
&[h, f_i] = -\alpha_i(h) f_i, \qquad \forall\ h \in \mathfrak{h};\\
&(ad\ e_i)^{1-a_{ij}} e_j = 0, \qquad \text{if } i \neq j;\\
&(ad\ f_i)^{1-a_{ij}} f_j = 0, \qquad \text{if } i \neq j,
\end{aligned}
$$

where $\operatorname{ad} e_i = [e_i, -]$. The elements e_i and f_j are called **Chevalley generators**.

Let $\mathfrak{n}^+$, respectively $\mathfrak{n}^-$, be the Lie subalgebra of $\mathfrak{g}$ generated by $\{e_i | 1 \le i \le n\}$, respectively $\{f_i | 1 \le i \le n\}$. The Kac–Moody algebra $\mathfrak{g}$ has triangular decomposition: $\mathfrak{g} = \mathfrak{n}^- \oplus \mathfrak{h} \oplus \mathfrak{n}^+$.

Denote by $\mathfrak{g}' = \mathfrak{g}'(C)$ the subalgebra of $\mathfrak{g}$ generated by all Chevalley generators e_i and f_j.

THEOREM 2.0.4 [11]. *Let $\mathfrak{h}' \subset \mathfrak{h}$ be the span of $h_1, \ldots, h_n$. Then*

(1) $\mathfrak{g}' = \mathfrak{n}^- \oplus \mathfrak{h}' \oplus \mathfrak{n}^+$;
(2) $\mathfrak{g}' = [\mathfrak{g}, \mathfrak{g}]$.

REMARK 2.0.5. The algebra $\mathfrak{g}' = \mathfrak{g}'(C)$ is the derived algebra of a Kac–Moody algebra $\mathfrak{g}$. The algebra $\mathfrak{g}'$ can be defined as the complex Lie algebra with $3n$ generators e_i, f_i, h_i $(1 \le i \le n)$ and the following defining relations:

$$
\begin{aligned}
&[h_i, h_j] = 0, \qquad \forall\ i, j;\\
&[e_i, f_j] = \delta_{ij} h_i;\\
&[h_i, e_j] = a_{ij} e_j;\\
&[h_i, f_j] = -a_{ij} f_j;\\
&(ad\ e_i)^{1-a_{ij}} e_j = 0, \qquad \text{if } i \neq j;\\
&(ad\ f_i)^{1-a_{ij}} f_j = 0, \qquad \text{if } i \neq j.
\end{aligned}
$$

Note that the Kac–Moody algebra $\mathfrak{g}$ and its derived algebra $\mathfrak{g}'$ share a common structure and representation theory.

Any element $h \in \mathfrak{h}$ gives rise to a decomposition of $\mathfrak{g}$ into ad_h-eigenspaces, and since $\mathfrak{h}$ is commutative, there is a simultaneous such decomposition

$$\mathfrak{g} = \oplus_{\alpha \in \mathfrak{h}^*} \mathfrak{g}_\alpha,$$

where

$$\mathfrak{g}_\alpha = \{x \in \mathfrak{g} \mid [h, x] = \alpha(h)x, \ \forall h \in \mathfrak{h}\}.$$

A nonzero element $\alpha \in \mathfrak{h}^*$ such that $\mathfrak{g}_\alpha \neq \{0\}$ is called a **root** of $\mathfrak{g}$, and the set of roots

$$\Phi = \{\alpha \in \mathfrak{h}^* \mid \mathfrak{g}_\alpha \neq \{0\}\}$$

is called the **root system** of $\mathfrak{g}$ (relative to $\mathfrak{h}$). The number

$$\operatorname{mult} \alpha := \dim \mathfrak{g}_\alpha$$

is called the multiplicity of α. Note that $\Pi = \{\alpha_1, \ldots, \alpha_n\} \subset \Phi$. Every root $\beta \in \Phi$ can be written as

$$\beta = \sum_{i=1}^{n} l_i \alpha_i$$

with integer coefficients l_i, all non-negative or all nonpositive. The root in Π is then called **simple root**. If all $l_i \geq 0$ (respectively, all $l_i \leq 0$), we call β **positive root** (respectively **negative root**). The collections of positive and negative roots will usually be denoted by Φ^+ and Φ^-. The set Φ comes with a canonical polarization $\Phi = \Phi^+ \sqcup \Phi^-$ and $\Phi^+ = -\Phi^-$.

For each $i = 1, \ldots, n$, define the **simple reflection** s_i of $\mathfrak{h}^*$ by the formula

$$s_i(\iota) = \iota - \langle h_i, \iota \rangle \alpha_i, \qquad \iota \in \mathfrak{h}^*.$$

Note that $s_i \in GL(\mathfrak{h}^*)$. The subgroup $W = W(C)$ of $GL(\mathfrak{h}^*)$ generated by all simple reflections is called the **Weyl group** of $\mathfrak{g}$.

A root α belonging to $W \cdot \Pi$ is called **real root** and obviously satisfies $\dim \mathfrak{g}_\alpha = 1$. A root that is not real is called **imaginary root** and may have higher multiplicity. The imaginary roots are W–invariant, which means $W \cdot \Phi_{\text{img}} = \Phi_{\text{img}}$. Let Φ^+_{real} be the set of all positive real roots and Φ^+_{img} be the set of all positive imaginary roots. We know that every positive root is either real or imaginary, thus we have $\Phi^+ = \Phi^+_{\text{real}} \sqcup \Phi^+_{\text{img}}$.

For simplicity, we only consider case of an irreducible and simply laced Cartan datum $\Delta = (I, (\ ,\ ))$ in the following sections, except Section 7.

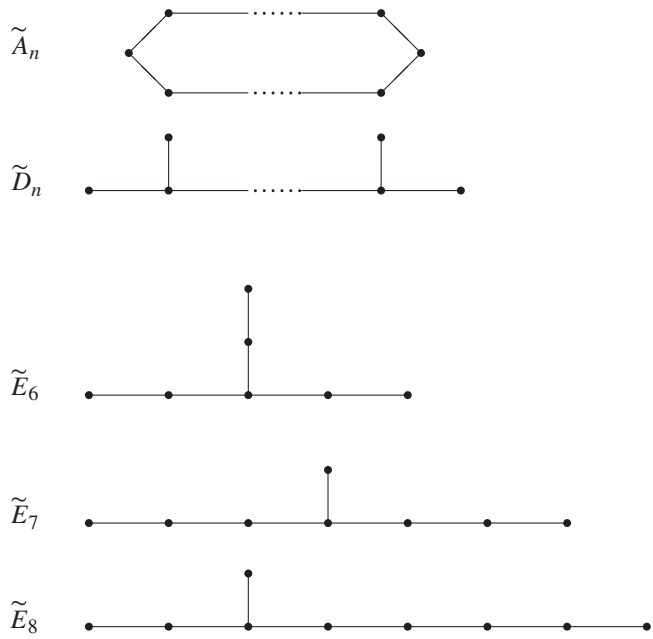
If Δ is symmetric and irreducible, then the corresponding Cartan matrix is symmetric and indecomposable. We can associate a graph of a symmetric Cartan matrix. Its vertices are labeled by $1, 2, \ldots, n$ ($n = |I|$). If $i \neq j$, the vertices i, j are jointed by $a_{ij}a_{ji}$ edges. Recall that a Cartan matrix is called **indecomposable** if its associated graph is connected.

THEOREM 2.0.6. *Let* $\mathfrak{g}(C)$ *be a Kac–Moody algebra with* C *symmetric and indecomposable. Then,*

- $\mathfrak{g}(C)$ *is a simple Lie algebra* $\iff$ *the graph of* C *is one of Dynkin diagrams* A_n, D_n, E_6, E_7, *and* E_8.
- $\mathfrak{g}(C)$ *is affine Kac–Moody algebra* $\iff$ *the graph of* C *is one of Euclidean diagrams* $\widetilde{A}_n$, $\widetilde{D}_n$, $\widetilde{E}_6$, $\widetilde{E}_7$, *and* $\widetilde{E}_8$.

The Dynkin diagrams of type A, D, and E have been displayed in the previous section.

The **Euclidean diagrams** (called by some authors "extended Dynkin diagrams") of type $\widetilde{A}$, $\widetilde{D}$, and $\widetilde{E}$ are as follows:



Quantum group of type $\Delta = (I, (\ ,\ ))$. Quantum groups were initially developed in the context of the inverse scattering methods and used to solve integrable quantum systems. V. Drinfel'd and M. Jimbo showed independently in 1985 that quantum groups are Hopf algebras that are often deformations of universal enveloping algebras of Lie algebras. Since then, quantum groups have brought a new level of understanding to a wide variety of areas of physical and mathematical research. In mathematics, they have been used to develop invariants of knots; other applications have been found in category theory and also in the representation theory of algebraic groups in prime characteristic. They are very influential in modern developments of ring theory and provide an inspiration to the new field of quantum geometry.

The Cartan datum considered in this chapter is nothing but a symmetrizable generalized Cartan matrix together with its W-invariant scalar product $(,)$ defined on $\mathbf{Z}^I$, where W is the Weyl group and $\mathbf{Z}^I$ is the lattice generated by the root system Φ.

Let $\mathfrak{g} = \mathfrak{g}(C)$ be a symmetrizable Kac–Moody Lie Algebra of type $\Delta = (I, (,))$, where $\Delta = (I, (,))$ is a Cartan datum. Let $\mathbf{Q}(v)$ be the field of rational function in the variable v.

DEFINITION 2.0.7 (Drinfel'd–Jimbo). The **quantum group** (or "quantized enveloping algebra") $\mathbf{U} = \mathbf{U}_v(\mathfrak{g})$ of type Δ is defined as a $\mathbf{Q}(v)$-algebra with generators E_i, F_i, K_i, and K_{-i}, $i \in I$ subject to the relations:

$$\begin{aligned} K_i K_{-i} &= 1 = K_{-i} K_i, \\ K_i K_j &= K_j K_i, \\ K_i E_j &= v^{(i,j)} E_j K_i, \\ K_i F_j &= v^{-(i,j)} F_j K_i, \\ E_i F_j - F_j E_i &= \delta_{ij} \frac{K_i - K_{-i}}{v_i - v_i^{-1}} \end{aligned}$$

for any $i, j \in I$, where $v_i = v^{\frac{(i,i)}{2}}$, and

$$\sum_{t=0}^{1-a_{ij}} (-1)^t \begin{bmatrix} 1-a_{ij} \\ t \end{bmatrix}_{\varepsilon_i} E_i^t E_j E_i^{1-a_{ij}-t} = 0$$

$$\sum_{t=0}^{1-a_{ij}} (-1)^t \begin{bmatrix} 1-a_{ij} \\ t \end{bmatrix}_{\varepsilon_i} F_i^t F_j F_i^{1-a_{ij}-t} = 0$$

for any $i \neq j$ in I, where $a_{ij} = 2\frac{(i,j)}{(i,i)}$, $\varepsilon_i = \frac{(i,i)}{2}$; for $i \in I$, we use the notation

$$[s]_{\varepsilon_i} = \frac{(v^{\varepsilon_i})^s - (v^{\varepsilon_i})^{-s}}{(v^{\varepsilon_i}) - (v^{\varepsilon_i})^{-1}} = (v^{\varepsilon_i})^{s-1} + (v^{\varepsilon_i})^{s-3} + \cdots + (v^{\varepsilon_i})^{-s+1},$$

$$[s]_{\varepsilon_i}! = \prod_{r=1}^{s} [r]_{\varepsilon_i}, \text{ and}$$

$$\begin{bmatrix} s \\ r \end{bmatrix}_{\varepsilon_i} = \frac{[s]_{\varepsilon_i}!}{[r]_{\varepsilon_i}![s-r]_{\varepsilon_i}!}$$

here, r and s are non-negative integers, and $r \leq s$. For convenience, we set $[0]_{\varepsilon_i}! = 1$. In the case when $\varepsilon_i = 1$, write $[s] = [s]_1$, $[s]! = [s]_1!$, and $\begin{bmatrix} s \\ r \end{bmatrix} = \begin{bmatrix} s \\ r \end{bmatrix}_1$.

Let $\mathcal{A} = \mathbf{Z}[v, v^{-1}]$. Define the **divided powers subalgebra** ${}_{\mathcal{A}}\mathbf{U}$ to be the $\mathcal{A}$-subalgebra of $\mathbf{U}_v(\mathfrak{g})$ generated by the elements of the set

$$\{K_{\pm i}, E_i^{(r)} = \frac{E_i^r}{[r]_{\varepsilon_i}!}, F_i^{(r)} = \frac{F_i^r}{[r]_{\varepsilon_i}!} | i \in I, r \geq 0\}.$$

We define ${}_{\mathcal{A}}\mathbf{U}^{\pm}$ in an analogous manner.

Next, we define the bar involution $\overline{(\,)}$ of $\mathbf{U}_v(\mathfrak{g})$ by

$$\overline{E_i} = E_i, \overline{F_i} = F_i, \overline{v} = v^{-1}.$$

The involution preserves both $\mathbf{U}^{\pm}$ and ${}_{\mathcal{A}}\mathbf{U}^{\pm}$.

3. Quivers and their representations

We will work over a base field k. Unless mentioned otherwise, k is assumed to be algebraically closed.

3.1. *Quivers and their representations*

Definition 3.1.1. As mentioned before, a quiver is a directed graph. Formally, a finite **quiver**

$$\mathcal{Q} = (\mathcal{Q}_0, \mathcal{Q}_1, s, t : \mathcal{Q}_1 \longrightarrow \mathcal{Q}_0)$$

is given by

- $\mathcal{Q}_0 =$ a finite set of vertices $= \{1, 2, \ldots, n\}$;
- $\mathcal{Q}_1 =$ a finite set of arrows $= \{\rho : s(\rho) \xrightarrow{\rho} t(\rho)\}$,

where the two maps s and t assign to each arrow $\rho \in \mathcal{Q}_1$ its source $s(\rho) \in \mathcal{Q}_0$ and its target $t(\rho) \in \mathcal{Q}_0$, respectively.

Note that multiple arrows between the same pair of vertices are allowed, as well as **loops** (arrows which start and end at the same vertex).

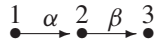
The quiver $\mathcal{Q}$ is said to be **connected** quiver if the underlying graph (which is obtained from $\mathcal{Q}$ by forgetting about the orientation of the arrows) is a connected graph. We only consider connected quiver in the following.

A nontrivial **path** σ in a quiver $\mathcal{Q}$ is a sequence $\sigma = \sigma_1\sigma_2 \ldots \sigma_m$ of arrows, which satisfies $t(\sigma_i) = s(\sigma_{i+1})$ for all $1 \leq i \leq m-1$. Often, we shall denote it by a diagram like the following:

$$\sigma = \bullet \xrightarrow{\sigma_1} \bullet \xrightarrow{\sigma_2} \bullet \cdots \xrightarrow{\sigma_m} \bullet.$$

This path starts at $s(\sigma_1)$ and terminates at $t(\sigma_m)$, m is the length of a path σ. We denote by $s(\sigma)$ and $t(\sigma)$ the starting and terminating vertex of the path σ. For each vertex i,

we denote by e_i the trivial path of length zero, which starts and terminates at i. A nontrivial path σ is called an **oriented cycle** if $s(\sigma) = t(\sigma)$.

EXAMPLE 3.1.2. The quiver $\mathcal{Q}$:  has three vertices and two arrows.

REMARK 3.1.3. It is trivial to see that every simply laced Dynkin diagram can be made into a quiver by assigning an orientation to each edge. Note that there are lots of ways of orienting the edges of a Dynkin diagram to make a quiver. For example, the following two quivers

are both obtained from the same Dynkin diagram of type A_3.

Given a quiver $\mathcal{Q}$, we now define $\mathrm{Rep}(\mathcal{Q})$, the **category of representations of a quiver** $\mathcal{Q}$. We will restrict our discussion to finite-dimensional representations of a quiver.

The ingredients of $\mathrm{Rep}(\mathcal{Q})$ are given as follows:

- objects: a (finite dimensional) **representation** V of a quiver $\mathcal{Q}$ is given by a finite-dimensional k-vector space V_i for each $i \in \mathcal{Q}_0$ and a linear map $V_\rho : V_{s(\rho)} \to V_{t(\rho)}$ for each $\rho \in \mathcal{Q}_1$;
- morphisms: $\theta : V \to V'$ is given by linear maps $\theta_i : V_i \to V'_i$ for each $i \in \mathcal{Q}_0$ satisfying

$$\begin{array}{ccc} V_{s(\rho)} & \xrightarrow{V_\rho} & V_{t(\rho)} \\ {\scriptstyle \theta_{s(\rho)}}\downarrow & & \downarrow{\scriptstyle \theta_{t(\rho)}} \\ V'_{s(\rho)} & \xrightarrow{V'_\rho} & V'_{t(\rho)} \end{array} \qquad V'_\rho \theta_{s(\rho)} = \theta_{t(\rho)} V_\rho;$$

- composition: $V \xrightarrow{\theta} V' \xrightarrow{\phi} V''$ is given by $(\phi\theta)_i = \phi_i \theta_i$;
- the **direct sum** $V' \oplus V''$ of two representations of a quiver $\mathcal{Q}$ is given by

$$(V' \oplus V'')_i = V'_i \oplus V''_i, \qquad i \in \mathcal{Q}_0,$$

$$(V' \oplus V'')_\rho : V'_{s(\rho)} \oplus V''_{s(\rho)} \xrightarrow{\begin{pmatrix} V'_\rho & 0 \\ 0 & V''_\rho \end{pmatrix}} V'_{t(\rho)} \oplus V''_{t(\rho)}, \ \rho \in \mathcal{Q}_1;$$

- V is a trivial (or zero) representation if $V_i = 0$ for each $i \in \mathcal{Q}_0$ and $V_\rho = 0$ for each $\rho \in \mathcal{Q}_1$;
- V is **indecomposable** representation of a quiver $\mathcal{Q}$ if $V \not\cong V' \oplus V''$, where V' and V'' are nontrivials;

- the **dimension vector** of a representation of a quiver $\mathcal{Q}$ is denoted by

$$\underline{\dim}\, V = (\dim_k V_i)_{i \in \mathcal{Q}_0};$$

- for each $i \in \mathcal{Q}_0$, the **simple** representation S_i of a quiver $\mathcal{Q}$ is given by

$$S_{i_j} = \begin{cases} k & \text{when } j = i \\ 0 & \text{otherwise} \end{cases} \qquad \text{and } S_{i_\rho} = 0 \text{ for all } \rho \in \mathcal{Q}_1.$$

It is easy to see that the simple representation has a standard vector as dimension vector, namely

$$\alpha_i = \underline{\dim}\, S_i = (0, 0, \ldots, \overset{i}{1}, \ldots, 0, 0).$$

EXAMPLE 3.1.4. Consider the quiver $\mathcal{Q}$: $\overset{1}{\bullet} \longrightarrow \overset{2}{\bullet} \longrightarrow \overset{3}{\bullet}$

A representation V_1 is given by $0 \xrightarrow{[0]} k \xrightarrow{[1]} k$ and another representation V_2 is given by $k \xrightarrow{[1]} k \xrightarrow{[1]} k$. Both are finite dimensional and indecomposable. A morphism $V_1 \to V_2$ is given by

$$\begin{array}{ccccc} 0 & \xrightarrow{[0]} & k & \xrightarrow{[1]} & k \\ {\scriptstyle [0]}\downarrow & & {\scriptstyle [1]}\downarrow & & \downarrow{\scriptstyle [1]} \\ k & \xrightarrow{[1]} & k & \xrightarrow{[1]} & k \end{array} .$$

It is easy to compute that the following representation $k^2 \xrightarrow{[0,1]} k \xrightarrow{[1]} k$ is decomposable and

$$k^2 \xrightarrow{[0,1]} k \xrightarrow{[1]} k \quad \simeq \quad k \xrightarrow{[1]} k \xrightarrow{[1]} k \quad \oplus \quad k \xrightarrow{[0]} 0 \xrightarrow{[0]} 0.$$

Moreover, we can see that

$$\underline{\dim}\left(k^2 \xrightarrow{[0,1]} k \xrightarrow{[1]} k\right) = (2, 1, 1) = 2\alpha_1 + \alpha_2 + \alpha_3.$$

THEOREM 3.1.5 (Krull–Remak–Schmidt Theorem). *Every finite-dimensional representation V of a quiver $\mathcal{Q}$ is isomorphic to a direct sum of indecomposable representations. This decomposition is unique up to isomorphism and permutation of factors.*

Thus, the problem of classifying representations of a quiver requires us to classify all indecomposable representations and then to find what morphisms between those indecomposables can exist.

We summarize the main problems of the representation theory of quivers as follows (cf. [12]):

- to develop methods for constructing indecomposable representations;
- to look for suitable invariants to identify indecomposable representations;

- to show that a given list of indecomposable representations is complete;
- to find connections with other areas of mathematics.

3.2. *Representations of $\mathcal{Q}$ and problems in linear algebra*

A whole range of problems of linear algebra can be formulated in a uniform way in terms of representations of quivers.

EXAMPLE 3.2.1 (Equivalence of matrices). Consider the quiver $\mathcal{Q}: \overset{2}{\bullet} \xrightarrow{\alpha} \overset{1}{\bullet}$

In this case, a finite-dimensional representation is given by a linear map $f: V_2 \to V_1$, the map itself may be given by a matrix A. The following diagram illustrates the situation when this representation and another one (given by a matrix B) are equivalent (i.e. isomorphic in the category of representations). The isomorphism is realized by two invertible matrices P and Q:

$$\begin{array}{ccc} k^m & \xrightarrow{A} & k^n \\ {\scriptstyle P}\downarrow & & \downarrow{\scriptstyle Q} \\ k^m & \xrightarrow{B} & k^n \end{array} \qquad A = Q^{-1}BP.$$

The classification of equivalence classes of linear maps amounts to classifying the $n \times m$ matrices up to "row and column equivalence." As is well known from elementary linear algebra, the equivalence classes of matrices are completely determined by their size and rank. Every representation of dimension vector (m, n) is a direct sum of three types of indecomposable representations: $S_1: 0 \xrightarrow{[0]} k$, $S_2: k \xrightarrow{[0]} 0$, and $S_{2,1}: k \xrightarrow{[1]} k$. The $n \times m$ matrix has rank r when the summand $S_{2,1}$ occurs r times in its decomposition. The classification of representations of this quiver is trivial since we only have three indecomposable representations up to isomorphism.

EXAMPLE 3.2.2 (Similarity of matrices). Consider the quiver $\mathcal{Q}:$ 

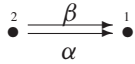
Giving a finite-dimensional representation simply means giving an endomorphism of a finite-dimensional vector space $f: V_1 \to V_1$. The following diagram illustrates the situation where two representations are isomorphic:

$$\begin{array}{ccc} k^n & \xrightarrow{A} & k^n \\ {\scriptstyle P}\downarrow & & \downarrow{\scriptstyle P} \\ k^n & \xrightarrow{B} & k^n \end{array} \qquad A = P^{-1}BP.$$

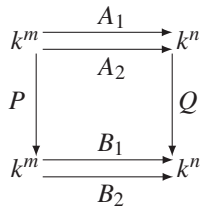
The classification of isomorphism classes is equivalent to the classification of matrices up to conjugation. This problem is answered by the well-known **Jordan classification**. In terms of indecomposable representations, this means that

for a dimension vector (n), there is a one parameter family of indecomposable endomorphisms

$$J_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix}_{n\times n}, \quad \lambda \in k.$$

EXAMPLE 3.2.3 (Equivalence of matrix pairs). Consider the **Kronecker quiver** $\mathcal{Q}$: 

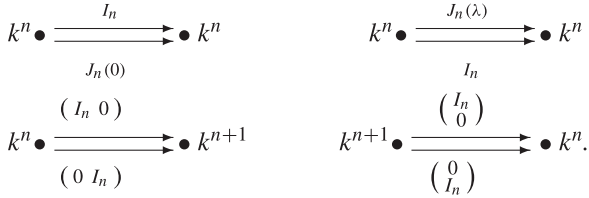
A representation means a pair of linear maps: $f_\alpha : V_2 \to V_1$ and $f_\beta : V_2 \to V_1$. Two representations are isomorphic if the following holds:



$$A_1 = Q^{-1}B_1P,$$
$$A_2 = Q^{-1}B_2P.$$

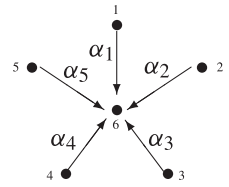
The problem of classifying such pairs of matrices was considered by K. Weierstrass, but his solution was incomplete. The problem was completely solved by L. Kronecker in 1890.

Here, we list isomorphism classes of all indecomposable representations:



This time there are infinitely many indecomposables, but we can classify them. An effective classification of the representations of this quiver is possible.

EXAMPLE 3.2.4 (The five subspace problem). Consider the quiver:

$\mathcal{Q}$: 

A representation of this star quiver is given by

$$V = \{V_1, V_2, V_3, V_4, V_5, V_6; f_{\alpha_i} : V_i \to V_6, 1 \le i \le 5\}.$$

If all maps are injective, we can have such a representation as a subspace configuration. There exists a two-parameter family of pairwise nonisomorphic indecomposable representations,

$$V_{\lambda,\mu} = \left\{k, k, k, k, k, k^2; \binom{1}{0}, \binom{0}{1}, \binom{1}{1}, \binom{1}{\lambda}, \binom{1}{\mu}\right\}.$$

Unfortunately, one cannot effectively find a complete list of all indecomposable representations. The complete classification of representations of this quiver seems impossible.

3.3. *Representation types of quivers*

Based on the above discussion, we can distinguish quivers with respect to their representation types.

The **representation type** of a quiver $\mathcal{Q}$ is one of the following three:

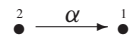
finite type: the number of isomorphism classes of indecomposable representations is finite;

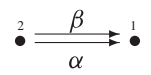
tame type: the number of isomorphism classes of indecomposable representations is infinite, but a complete list of indecomposables can be described explicitly. In each dimension, there exist only finitely many one-parameter families of pairwise nonisomorphic indecomposables;

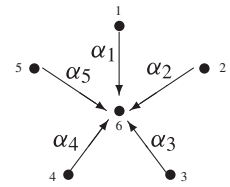
wild type: the quiver is neither of tame nor of finite type. As it turns out, in this case one cannot provide a "complete list" of indecomposable representations.

We do not provide the precise definitions of tame and wild, as those are rather technical. It is well-known that every quiver is either of finite, tame, or wild type.

REMARK 3.3.1. The representation types of the quivers discussed in the examples are as follows:

Example 3.2.1 is of finite type: 

Example 3.2.3 is of tame type: 

Example 3.2.4 is of wild type: 

The basic result in representation theory of quivers is the following theorem classifying the quivers with respect to different representation types.

THEOREM 3.3.2 [1, 13]. *Let $\mathcal{Q}$ be a quiver with no oriented cycles, and let Γ be the underlining graph of $\mathcal{Q}$. We get this graph when we forget the orientation of arrows in $\mathcal{Q}$. Then,*

- *$\mathcal{Q}$ is of finite type $\iff$ Γ is Dynkin.*
- *$\mathcal{Q}$ is of tame type $\iff$ Γ is Euclidean.*
- *All other quivers are of wild type.*

3.4. *The path algebra of a quiver*

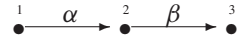
Starting with a quiver, one can construct an algebra that has some remarkable properties.

DEFINITION 3.4.1. Let $\mathcal{Q}$ be a quiver. The **path algebra** $k\mathcal{Q}$ is the k-vector space with

- basis: {all paths in $\mathcal{Q}$};
- multiplication given by:

$$\sigma\tau = \begin{cases} \text{obvious composition} & \text{if } t(\sigma) = s(\tau), \\ 0 & \text{otherwise.} \end{cases}$$

This product is extended to the entire space $k\mathcal{Q}$ using distributivity and linearity. In this way, one obtains an associative algebra $k\mathcal{Q}$. The "zero-length" path starting and ending at the vertex i is denoted by e_i. Note that products are read from left to right (unlike composition of morphisms).

EXAMPLE 3.4.2. Consider the quiver $\mathcal{Q}$: 

The path algebra $k\mathcal{Q}$ has basis $\{e_1, e_2, e_3, \alpha, \beta, \alpha\beta\}$. A couple of examples of products are: $\beta\alpha = \beta\beta = \alpha\beta e_1 = e_3\alpha\beta = 0$; $\alpha\beta e_3 = \alpha\beta$.

EXAMPLE 3.4.3. If $\mathcal{Q}$ consists of one vertex and one loop, that is, $\mathcal{Q}$: 

, then $k\mathcal{Q} \simeq k[T]$. If $\mathcal{Q}$ has one vertex and n loops, then $k\mathcal{Q}$ is the free associative algebra on n letters.

The basic properties of the path algebra $k\mathcal{Q}$ are dictated by the quiver $\mathcal{Q}$.

PROPOSITION 3.4.4. *Set $\Lambda = k\mathcal{Q}$. Note the following facts (cf. [14]) :*

(1) *the set $\{e_i \mid i \in \mathcal{Q}_0\}$ is a complete set of primitive orthogonal idempotents of Λ;*
(2) *Λ has an identity given by $\sum_{i=1}^{n} e_i$;*
(3) *the spaces $e_i\Lambda$, Λe_j, and $e_i\Lambda e_j$ have as bases the paths starting at i, terminating at j and starting at i and terminating at j, respectively;*
(4) *distinct e_i's are not equivalent, that is, $e_i\Lambda \not\simeq e_j\Lambda$ for $i \neq j$;*

(5) $\Lambda_\Lambda = \oplus_{i=1}^n e_i\Lambda$, *and each* $e_i\Lambda$ *is an indecomposable projective right* Λ*-module*;
(6) Λ *is finite dimensional* $\iff$ $\mathcal{Q}$ *has no oriented cycles*;
(7) *if* $\mathcal{Q}$ *has no oriented cycles, then* Λ *is a hereditary algebra.*

Recall that an algebra (or more generally, a ring) is called **hereditary** if every submodule of a projective module is projective.

The following theorem tells us that we can identify finite-dimensional representations of a quiver $\mathcal{Q}$ with finitely generated (finite dimensional) right $k\mathcal{Q}$-modules.

THEOREM 3.4.5. *We have the following equivalence*:

$$\mathrm{Rep}(\mathcal{Q}) \simeq \mathrm{mod}\, -k\mathcal{Q}.$$

PROOF. We only provide the construction. If M is a $k\mathcal{Q}$-module, define a representation V with $V_i = Me_i$ and $V_\rho(x) = x\rho = x\rho e_{t(\rho)} \in V_{t(\rho)}$.

If V is a representation, define a module M as follows. Let $M = \bigoplus_{i\in\mathcal{Q}_0} V_i$. Let $V_i \xrightarrow{\varepsilon_i} M \xrightarrow{\pi_i} V_i$ be the canonical maps. Then,

$$xe_i = \varepsilon_i\pi_i(x),\ \ x\sigma_1\sigma_2\ldots\sigma_m = \varepsilon_{t(\sigma_m)}V_{\sigma_m}\ldots V_{\sigma_2}V_{\sigma_1}\pi_{s(\sigma_1)}(x).$$

□

A module M over $k\mathcal{Q}$ can be identified with a representation of $\mathcal{Q}$, which is given by a set of k-vector spaces indexed by $\mathcal{Q}_0$ along with a set of linear maps indexed by $\mathcal{Q}_1$. Thus, $M = (M_i, f_\alpha)_{i\in\mathcal{Q}_0,\alpha\in\mathcal{Q}_1}$, where $f_\alpha \in \mathrm{Hom}_k(M_i, M_j)$ if α is an arrow from i to j. For instance, let S_i be the simple representation corresponding to a vertex i, then it can be identified with the module, still denoted by $S_i = (S_{i_j}, f_\alpha)$ where $S_{i_j} = k$, if $j = i$, $S_{i_j=0}$, if $j \neq i$, $f_\alpha = 0$ for all $\alpha \in \mathcal{Q}_1$. Then, S_i is a simple module and, in the case that $k\mathcal{Q}$ is finite dimensional, any simple module is of this form.

REMARK 3.4.6. Dealing with representations and morphisms of representations of a quiver $\mathcal{Q}$ amounts to the same thing as dealing with modules and homomorphisms of modules over the path algebra $k\mathcal{Q}$. It is very easy to compute with representations since only vector spaces and matrices are involved.

In the following, we will not distinguish between representations of a quiver and modules of corresponding path algebra.

Given a finite-dimensional k-algebra A, consider it as a right module over itself and decompose it into indecomposables $\Lambda = \oplus_{i=1}^n P_i$, which are projective as is easily seen. The algebra A is called **basic** if P_i is not isomorphic to P_j for $i \neq j$. Note that every finite-dimensional algebra is Morita equivalent to a basic finite-dimensional algebra. The next two theorems elucidate the relationship of finite-dimensional associative algebras and path algebras of quivers.

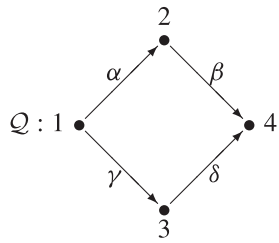
THEOREM 3.4.7. *A basic finite-dimensional hereditary algebra over an algebraically closed field* k *is isomorphic to the path algebra* $k\mathcal{Q}$ *for some quiver* $\mathcal{Q}$.

Since there are finite-dimensional k-algebras that are not hereditary, not every basic finite-dimensional k-algebra is isomorphic to the path algebra of a quiver. The following theorem is one of the cornerstones of the theory.

THEOREM 3.4.8. *A basic finite-dimensional algebra over an algebraically closed field k is isomorphic to a certain quotient algebra of the path algebra $k\mathcal{Q}$ for some quiver $\mathcal{Q}$.*

REMARK 3.4.9. A quotient algebra of a path algebra can be described by admissible relations. An admissible relation on $\mathcal{Q}$ over a field k is a k-linear combination of paths of length ≥ 2. Without loss of generality, we can assume that all paths involved in one relation have the same starting vertex and the same terminating vertex. Given a set $\mathcal{S}$ of admissible relations, consider the $k\mathcal{Q}$-ideal $\mathcal{I}$ generated by $\mathcal{S}$, and call such an **admissible ideal**. The quotient algebra $k\mathcal{Q}/\mathcal{I}$ can be finite dimensional even if $k\mathcal{Q}$ is infinite dimensional. Just as before, a module over $k\mathcal{Q}/\mathcal{I}$ can be identified with a representation of $\mathcal{Q}$, where the set of maps is to satisfy the additional conditions that the linear combinations of homomorphisms corresponding to relations in $\mathcal{S}$ are zero. Let us reiterate that it is well known that any basic algebra over algebraically closed field k is isomorphic to $k\mathcal{Q}/\mathcal{I}$ for some quiver $\mathcal{Q}$ and some admissible ideal $\mathcal{I}$. Denote by $\mathrm{Rep}(\mathcal{Q}, \mathcal{I})$ the full subcategory of $\mathrm{Rep}(\mathcal{Q})$ consisting of the representations of $\mathcal{Q}$ bounded by $\mathcal{I}$ (see [7]). Then we have $\mathrm{Rep}(\mathcal{Q}, \mathcal{I}) \backsimeq \mathrm{mod}\, -k\mathcal{Q}/I$.

EXAMPLE 3.4.10 (Quiver with zero relations). Consider the quiver



.

Choose the ideal $\mathcal{I} =< \alpha\beta - \gamma\delta >$ generated by one admissible relation. The quotient algebra $k\mathcal{Q}/\mathcal{I}$ has a basis $\{\overline{e_1}, \overline{e_2}, \overline{e_3}, \overline{e_4}, \overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\delta}, \overline{\alpha\beta} = \overline{\gamma\delta}\}$.

It is well known that any finite-dimensional algebra is Morita equivalent to precisely one basic finite-dimensional algebra. Thus, most problems from the representation theory of finite-dimensional algebras can be understood as problems about representations of quivers.

REMARK 3.4.11. In case k is not algebraically closed field, one may have to resort to the use of **species** (also known as a "modulated quiver") instead of a simple quiver. The interested readers may consult [1].

The representation type of an algebra is defined in the same way as it was done for quivers. Alas, the rigorous definitions of tame and wild type make it far from obvious

that each algebra is precisely one of the three types. Fortunately, this highly nontrivial fact is true.

THEOREM 3.4.12 [15]. *Let A be a finite-dimensional k-algebra. Then A is of finite, tame, or wild representation type.*

3.5. *Auslander–Reiten theory*

Let A be a finite-dimensional k-algebra. We would like to record information on the category $\operatorname{mod} -A$ of right A–modules in terms of a directed graph. It seems most appropriate that vertices should represent some modules and arrows should represent some homomorphisms. Since $\operatorname{mod} -A$ is a Krull–Remak–Schmidt category, we only allow one point for each isomorphism class of indecomposable modules. As for the arrows, they should stand for "irreducible homomorphism," that is, those with no "nontrivial" factorization.

DEFINITION 3.5.1. A homomorphism $f : X \to Y$ in $\operatorname{mod} -A$ is **irreducible** morphism provided

(a) f is neither a split monomorphism nor a split epimorphism, and
(b) if $f = f_1 f_2$, either f_1 is a split epimorphism or f_2 is a split monomorphism, where $f_2 : X \to Z$ and $f_1 : Z \to Y$.

DEFINITION 3.5.2. A homomorphism $f : X \to Y$ in $\operatorname{mod} -A$ is **right almost split** morphism provided

(a) it is not a split epimorphism, and
(b) any morphism $Z \to Y$, which is not a split epimorphism factors through f.

The definition of a **left almost split** morphism is dual.

DEFINITION 3.5.3. An exact sequence $0 \longrightarrow X \xrightarrow{f} Y \xrightarrow{g} Z \longrightarrow 0$ in $\operatorname{mod} -A$ is called an **almost split sequence** (also called **Auslander–Reiten sequence**) if f is left almost split and g is right almost split.

Almost split sequences can be characterized in terms of irreducible morphisms.

THEOREM 3.5.4 [16]. *An exact sequence $0 \longrightarrow X \xrightarrow{f} Y \xrightarrow{g} Z \longrightarrow 0$ is an almost split sequence if and only if f and g are both irreducible.*

We can construct almost split sequences in $\operatorname{mod} -A$ through a functor. First, consider the A-dual functor

$$(-)^t = \operatorname{Hom}_A(-, A) : \operatorname{mod} -A \to \operatorname{mod} -A^{\mathrm{op}},$$

where mod $-A^{\mathrm{op}}$ is the category of all finite-dimensional left A-modules. Start by approximating each right A-module M by projective modules. Let

$$P_1 \xrightarrow{p_1} P_0 \xrightarrow{p_0} M \to 0$$

be a minimal projective presentation of M, that is, an exact sequence such that $p_0 : P_0 \longrightarrow M$ and $p_1 : P_1 \longrightarrow \operatorname{Ker} p_0$ are projective covers. Applying the left exact and contravariant functor $(-)^t$, we obtain an exact sequence of left A-modules

$$0 \longrightarrow M^t \xrightarrow{p_0^t} P_0^t \xrightarrow{p_1^t} P_1^t \longrightarrow \operatorname{Coker} p_1^t \longrightarrow 0.$$

Denote Coker p_1^t by $\operatorname{Tr}(M)$ and call it the **transpose** of M. The transpose functor transforms right A-module into left A-module. Thus, if we wish to define an endofunctor of mod $-A$, we need to compose it with another duality between right and left A-modules, namely the standard duality

$$D = \operatorname{Hom}_k(-, k).$$

The **Auslander–Reiten translations** are defined to be the compositions of D and Tr, namely, set $\tau = D\operatorname{Tr}$ and $\tau^{-1} = \operatorname{Tr} D$.

The existence theorem of almost split sequences is described as follows.

THEOREM 3.5.5 [16]. *Let A be a finite-dimensional k-algebra.*

(a) *For any indecomposable nonprojective A-module M, there exists an almost split sequence* $0 \longrightarrow \tau M \longrightarrow E \longrightarrow M \longrightarrow 0$;
(b) *For any indecomposable noninjective A-module N, there exists an almost split sequence* $0 \longrightarrow N \longrightarrow F \longrightarrow \tau^{-1} N \longrightarrow 0$.

In the following, a device for studying all left and right almost split morphisms simultaneously is introduced.

DEFINITION 3.5.6. The **Auslander–Reiten quiver** $\Gamma(A)$ of mod $-A$ is constructed as follows:

- the set of vertices Γ_0 consists of the isomorphism classes $[M]$ of finite-dimensional indecomposable modules;
- let $[M]$, $[N]$ be the vertices in $\Gamma(A)$ corresponding to the indecomposable modules M, N in mod $-A$. The arrows $[M] \to [N]$ are in bijective correspondence with the vectors of a basis of the k-vector space $\operatorname{Irr}(M, N)$ (all irreducible morphisms from M to N).

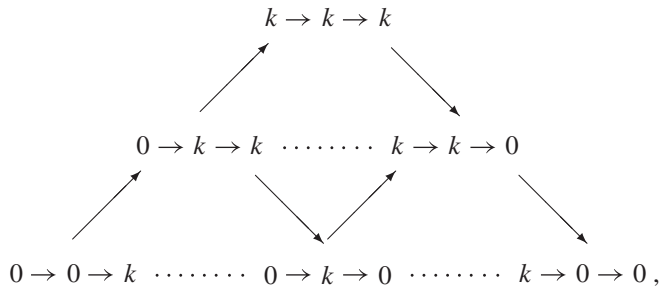
The Auslander–Reiten quiver of A encodes a lot of information about the homological properties of mod $-A$.

Now, $\Gamma(A)$ can be decomposed into components in a straightforward way: two objects X and Y lie in the same component if and only if there exist a natural number n, indecomposable modules X_i, for $1 \leq i \leq n$, and for each i, either an irreducible

morphism $f_i : X_i \to X_{i+1}$ or an irreducible morphism $g_i : X_{i+1} \to X_i$ with $X_1 = X$ and $X_n = Y$.

THEOREM 3.5.7. *Let $\mathcal{Q}$ be a quiver without oriented cycles. Then, $k\mathcal{Q}$ is of finite representation type if and only if $\Gamma(k\mathcal{Q})$ is a finite connected quiver* (*i.e. has only a finite connected component*).

EXAMPLE 3.5.8. Consider the quiver $\mathcal{Q}$: $\overset{1}{\bullet} \longrightarrow \overset{2}{\bullet} \longrightarrow \overset{3}{\bullet}$. The Auslander–Reiten quiver $\Gamma(k\mathcal{Q})$ looks as follows:



where the dotted lines indicate the Auslander–Reiten translation. The Auslander–Reiten quiver encodes much information about the representations and morphisms between them.

REMARK 3.5.9. Given a quiver $\mathcal{Q}$ without oriented cycles, let $\Lambda = k\mathcal{Q}$ be the path algebra. If it is of finite type, then one can construct all the indecomposables systematically and completely by following the "flow" of the Auslander–Reiten quiver $\Gamma(\Lambda)$. Start with indecomposable projectives $e_i\Lambda$ (or, alternatively, the injectives $\mathrm{Hom}_k(\Lambda e_i, k)$), for all $i \in \mathcal{Q}_0$, which correspond to the vertices of the quiver $\mathcal{Q}$, and then apply the inverse of the Auslander–Reiten translation functor τ^{-1} (or τ) repeatedly to get all the indecomposables.

3.6. *Quadratic forms of quivers and algebras*

DEFINITION 3.6.1. Let A be a basic finite-dimensional k-algebra (associative, with unit) and let $S_1, S_2, \ldots, S_n$ be a complete set of isomorphism classes of simple right A-modules. Given any A-module M, the **dimension vector** of M is the vector

$$\underline{\dim}\, M = \sum_{i=1}^{n} [M : S_i]\alpha_i,$$

where $[M : S_i]$ is the **Jordan–Hölder multiplicity** of S_i in M and α_i is the standard coordinate vector (or, equivalently, the dimension vector of the simple module S_i).

REMARK 3.6.2. Let Q be a quiver without oriented cycles. Let M be an $k\mathcal{Q}$-module. The dimension vector of M is the vector

$$\underline{\dim}\, M = (\dim_k Me_i)_{i\in\mathcal{Q}_0} = \sum_{i\in\mathcal{Q}_0} [M : S_i]\alpha_i.$$

DEFINITION 3.6.3. Let A be a k-algebra. The **Grothendieck group** of A is the Abelian group $G(A) = \mathfrak{F}/\mathfrak{F}'$, where $\mathfrak{F}$ is the free Abelian group having as basis the set of the isomorphism classes $[M]$ of modules M in mod $-A$ and $\mathfrak{F}'$ is the subgroup of $\mathfrak{F}$ generated by all elements $[M] - [L] - [N]$ corresponding to an exact sequence

$$0 \longrightarrow L \longrightarrow M \longrightarrow N \longrightarrow 0.$$

THEOREM 3.6.4. *Let A be a basic finite-dimensional k-algebra and let $S_1, S_2, \ldots, S_n$ be a complete set of isomorphism classes of simple right A-modules. Then there exists an isomorphism* $\underline{\dim} : G(A) \longrightarrow \mathbf{Z}^n$ *such that* $\underline{\dim}([M]) = \underline{\dim}\, M$.

Note that $G(A)$ can be identified with $\mathbf{Z}^n$.

PROPOSITION 3.6.5. *Let A be a basic k-algebra of finite global dimension. Let n be the rank of its Grothendieck group. The* ***Euler form*** *of A is the $\mathbf{Z}$-bilinear form defined on $\mathbf{Z}^n$ given by*

$$\langle \underline{\dim}\, M, \underline{\dim}\, N\rangle = \sum_{j=0}^{\infty} (-1)^j \dim_k \operatorname{Ext}_A^j(M, N),$$

where M and N are A-modules. The symmetric Euler form is given by

$$(\underline{\dim}\, M, \underline{\dim}\, N) = \langle \underline{\dim}\, M, \underline{\dim}\, N\rangle + \langle \underline{\dim}\, N, \underline{\dim}\, M\rangle.$$

DEFINITION 3.6.6. The **Euler quadratic form** of an algebra A is the form $\mathfrak{q}_A : \mathbf{Z}^n \longrightarrow \mathbf{Z}$ given by

$$\mathfrak{q}_A(\underline{\dim}\, M) = \langle \underline{\dim}\, M, \underline{\dim}\, M\rangle = \sum_{j=0}^{\infty} (-1)^j \dim_k \operatorname{Ext}_A^j(M, M),$$

where M is an A-module.

P. Gabriel classified the quivers of finite type in terms of Dynkin diagrams, and shortly thereafter I.N. Bernstein, I.M. Gelfand, and V.A. Ponomarev gave a new and elegant proof of Gabriel's theorem based on the connection between Dynkin diagrams and the quadratic form of quiver being positive definite.

Let $\mathcal{Q}$ be a quiver without oriented cycles. Let $n = |\mathcal{Q}_0|$.

DEFINITION 3.6.7. The **Euler form** (also known as the Ringel form) of a quiver $\mathcal{Q}$ is the bilinear form on $\mathbf{Z}^n$ given by

$$\langle x, y\rangle = \sum_{i\in\mathcal{Q}_0} x_i y_i - \sum_{\rho\in\mathcal{Q}_1} x_{s(\rho)} y_{t(\rho)}.$$

DEFINITION 3.6.8. The **symmetric Euler form** (or the **Cartan form**) of the quiver $\mathcal{Q}$ is given by

$$(x, y) = \langle x, y\rangle + \langle y, x\rangle.$$

Note that the symmetric Euler form is independent of the orientation of $\mathcal{Q}$.

DEFINITION 3.6.9. The **Tits form** of a quiver $\mathcal{Q}$ is defined by

$$\mathfrak{q}_{\mathcal{Q}}(x) = \langle x, x\rangle = \frac{1}{2}(x, x) = \sum_{i\in\mathcal{Q}_0} x_i^2 - \sum_{\rho\in\mathcal{Q}_1} x_{s(\rho)} x_{t(\rho)}.$$

LEMMA 3.6.10. *Let Q be a quiver without oriented cycles. Then, the Euler quadratic form $\mathfrak{q}_A$ of the path algebra $A = k\mathcal{Q}$ and the Tits form of the quiver $\mathcal{Q}$ coincide. Moreover,*

$$\mathfrak{q}_A(x) = \sum_{i\in\mathcal{Q}_0} x_i^2 - \sum_{i,j\in\mathcal{Q}_0} n_{ij} x_i x_j,$$

where $n_{ij} = \dim_k \operatorname{Ext}^1_A(S_i, S_j)$ = the number of arrows between i and j.

LEMMA 3.6.11. *Let Q be a quiver without oriented cycles. The symmetric Euler form $(,)$ on $\mathbf{Z}^n$ of the quiver $\mathcal{Q}$ (or $k\mathcal{Q}$) can be given by*

$$(\alpha_i, \alpha_j) = \begin{cases} -n_{ij} & \text{if } i \neq j; \\ 2 & \text{if } i = j, \end{cases}$$

where α_i is the ith standard coordinate vector (or the dimension vector of simple $k\mathcal{Q}$-module S_i). Define $C = (a_{ij})_{1\leq i,j\leq n}$, where $a_{ij} = (\alpha_i, \alpha_j)$. Then, C is a symmetric Cartan matrix.

Indeed, it is quite convenient to describe a finite quiver $\mathcal{Q}$ with no oriented cycles and vertex set $\mathcal{Q}_0 = \{1, 2, \ldots, n\}$ in terms of a Cartan matrix $C = (a_{ij})$. By the above, $a_{ii} = 2$ for $1 \leq i \leq n$, and $-a_{ij}$ $(= -a_{ji})$ is the number of arrows between i and j. The matrix C will then be called the Cartan matrix of a quiver $\mathcal{Q}$.

EXAMPLE 3.6.12. Consider the quiver $\mathcal{Q}$: $\overset{1}{\bullet}\longrightarrow\overset{2}{\bullet}\longrightarrow\overset{3}{\bullet}$.

The associated Cartan matrix of $\mathcal{Q}$ is

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

If $\mathcal{Q}$ does not contain oriented cycles, then $\Lambda = k\mathcal{Q}$ is hereditary. The Euler form is given by

$$\langle \underline{\dim}\, M, \underline{\dim}\, N\rangle = \dim_k \operatorname{Hom}_\Lambda(M, N) - \dim_k \operatorname{Ext}^1_\Lambda(M, N)$$

for any Λ–modules M and N. Since Λ is hereditary, $\langle \underline{\dim}\, M, \underline{\dim}\, N\rangle$ depends only on the dimension vectors $\underline{\dim}\, M$ and $\underline{\dim}\, N$.

COROLLARY 3.6.13. *Let $\mathcal{Q}$ be a quiver with no oriented cycles. The index set I $(= \mathcal{Q}_0)$ and the symmetric Euler form $(,)$ of $\mathcal{Q}$ give a realization of a Cartan datum Δ.*

Using the Tits form of a quiver, the set of roots Φ can be characterized as follows.

DEFINITION 3.6.14. Define $\Phi = \{\alpha \in \mathbf{Z}^n \mid \alpha \neq 0, \mathfrak{q}_\mathcal{Q}(\alpha) \leq 1\}$, the set of **roots**. A root α is **real** if $\mathfrak{q}_\mathcal{Q}(\alpha) = 1$ and **imaginary** if $\mathfrak{q}_\mathcal{Q}(\alpha) \leq 0$.

4. Dimension vectors and positive roots

In this section, let k be any field. We will see a close connection between the representation theory of quivers and Lie algebras.

4.1. *Gabriel theorem*

Let $\mathfrak{g}(C)$ be the Kac–Moody algebra associated to the Cartan matrix of a quiver $\mathcal{Q}$. Sometimes, we will say that the Kac–Moody algebra $\mathfrak{g}(C)$ and the path algebra $k\mathcal{Q}$ share the same Cartan datum (or the related diagram). Ignoring the orientation of $\mathcal{Q}$, the underlying graph Γ is the Dynkin graph of $\mathfrak{g}(C)$. Based on the root space decomposition of $\mathfrak{g}(C)$, we can obtain the root system Φ and a set of simple roots $\{\alpha_i \mid i \in I\}$ indexed by the vertices of Dynkin graph, which coincides with the root system defined in 3.6.14.

The following result is one of the founding stones of the representation theory of algebras.

THEOREM 4.1.1 [1,13]. *Let $\mathcal{Q}$ be a quiver without oriented cycles and $\Lambda = k\mathcal{Q}$ be the path algebra of $\mathcal{Q}$. Let $\mathcal{P}_{Ind}$ be the set of isomorphism classes of all indecomposable Λ-modules.*

If Λ and $\mathfrak{g}(C)$ share the same Dynkin (or Euclidean) diagram, then there exists a surjective map from $\mathcal{P}_{Ind}$ to Φ^+. The map is given by

$$[M] \longmapsto \sum_{i\in\mathcal{Q}_0} (\dim_k Me_i)\alpha_i.$$

Moreover, if Q is of finite type, then the above map is bijective.

REMARK 4.1.2.

- If we identify $\alpha_i = \underline{\dim}\, S_i = (0, 0, \ldots, \overset{i}{1}, \ldots, 0, 0)$, then

$$\underline{\dim}\, M = \sum_{i \in \mathcal{Q}_0} (\dim_k Me_i)\alpha_i.$$

  Thus, the dimension vector of the indecomposable Λ-module is nothing but a (corresponding) positive root;
- If there is an indecomposable representation of dimension vector α, then α is a root;
- If $\alpha \in \Phi^+_{real}$, then there is a unique indecomposable of dimension vector α (up to isomorphism);
- If $\alpha \in \Phi^+_{img}$, then there are many indecomposables of dimension vector α (up to isomorphism).

In the following, we will not distinguish between the dimension vector of an indecomposable Λ-module and the corresponding positive root of $\mathfrak{g}(C)$.

REMARK 4.1.3. Gabriel's result was generalized to arbitrary quivers in [17] by V. Kac who proved that indecomposable representations occur in dimension vectors that are the roots of Kac–Moody Lie algebra associated to a given graph. In particular, these dimension vectors do not depend on the orientation of the arrows in $\mathcal{Q}$.

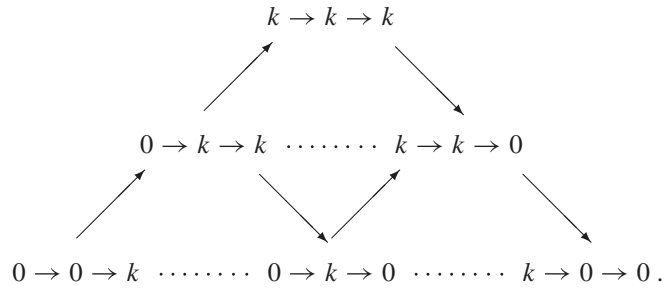
Given a quiver with loops, the symmetric bilinear form is defined by

$$(\alpha_i, \alpha_j) = \begin{cases} -n_{ij} & \text{if } i \neq j, \\ 2 - \#\{\text{loops at i}\} & \text{if } i = j. \end{cases}$$

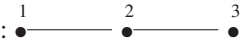
One can define roots for any graph, and more generally for "species" (see [1]). Using species, all possible Dynkin diagrams can arise.

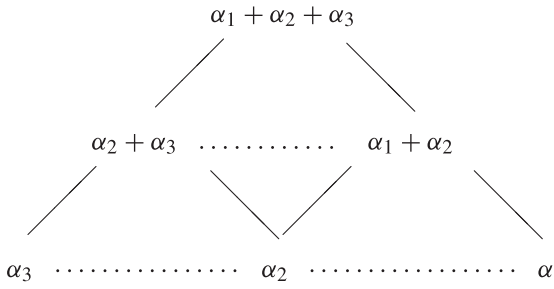
EXAMPLE 4.1.4. Consider the quiver $\mathcal{Q}$: $\overset{1}{\bullet} \longrightarrow \overset{2}{\bullet} \longrightarrow \overset{3}{\bullet}$.

The Auslander–Reiten quiver of the path algebra $k\mathcal{Q}$ looks as follows:



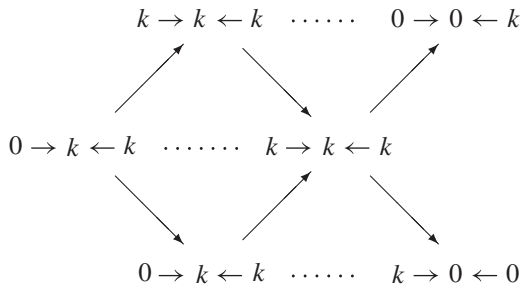
According to Gabriel theorem, all positive roots of a simple Lie algebra of type A_3 can be listed in a manner that is compatible with the above Auslander–Reiten quiver.

The positive roots of the Lie algebra of type A_3:  are



EXAMPLE 4.1.5. Consider the quiver $\mathcal{Q}$:  .

The Auslander–Reiten quiver of the path algebra $k\mathcal{Q}$ looks as follows:



This Auslander–Reiten quiver has a different shape as from previous one even though the underlying graph of the above two quivers are the same. But at the dimension vector level, they are the same. All dimension vectors are the positive roots of Lie algebra of type A_3.

4.2. *BGP reflection functors*

Let $\mathcal{Q}$ be a quiver without oriented cycles and $t \in \mathcal{Q}_0$. Let $\sigma_t\mathcal{Q}$ be the quiver obtained from $\mathcal{Q}$ by changing the orientation of all arrows that have t as a starting or terminating vertex.

Assume t is a sink for $\mathcal{Q}$. We now define the **Bernstein–Gelfand–Ponomarev reflection** functor:

$$\sigma_t^+ : \mathrm{Rep}(\mathcal{Q}) \to \mathrm{Rep}(\sigma_t\mathcal{Q}).$$

Let $V = (V_i, V_\rho)_{i\in\mathcal{Q}_0, \rho\in\mathcal{Q}_1}$ be a representation of $\mathcal{Q}$. The representation

$$\sigma_t^+ V = (V_i', V_\rho')_{i\in(\sigma_i\mathcal{Q})_0, \rho\in(\sigma_t\mathcal{Q})_1}$$

of $\sigma_t\mathcal{Q}$ is defined as follows:

- $V'_i = V_i$ for $i \neq t$, V'_t is the kernel of
$$\bigoplus_{\alpha:s(\alpha)\to t} V_{s(\alpha)} \to V_t,$$
where the sum is taken over all arrows in $\mathcal{Q}$ with terminating vertex t;
- $V'_\rho = V_\rho$ for all arrows $\rho : i \to j$ in $\mathcal{Q}_1$ with $j \neq t$. If $\rho : i \to t$ is an arrow in $\mathcal{Q}$, then $\rho' : t \to i$ is an arrow in $\sigma_t\mathcal{Q}$, and $V'_{\rho'}$ is the composition of the inclusion of V'_t into $\bigoplus_{\alpha:s(\alpha)\to t} V_{s(\alpha)}$ with the projection onto the direct summand V_i.

Dually, if t is a source for $\mathcal{Q}$, we can define another BGP-reflection functor

$$\sigma_t^- : \mathrm{Rep}(\mathcal{Q}) \to \mathrm{Rep}(\sigma_t\mathcal{Q}).$$

These functors are the analog of simple reflections in Lie algebra theory and will play a similar role.

PROPOSITION 4.2.1. *The functors σ_t^+ and σ_t^- induce quasi-inverse equivalences between the full subcategory of* $\mathrm{Rep}(\mathcal{Q})$ *of the representations having no direct summand isomorphic to S_t and the full subcategory of* $\mathrm{Rep}(\sigma_t\mathcal{Q})$ *of the representations having no direct summand isomorphic to S_t.*

DEFINITION 4.2.2. A sequence $i_m, \ldots, i_1$ is called a **sink sequence** for $\mathcal{Q}$ if i_m is a sink for $\mathcal{Q}$, and for any $1 \leq t < m$, the vertex i_t is a sink for the quiver $\sigma_{i_{t+1}} \ldots \sigma_m\mathcal{Q}$.

DEFINITION 4.2.3. Let $\mathcal{Q}$ be a quiver with no oriented cycles and let $i_n, \ldots, i_1$ be a sink sequence that uses each vertex of $\mathcal{Q}$ exactly once. We can then define the **Coxeter functors** $C^+ = \sigma_{i_1}^+ \ldots \sigma_{i_n}^+$ and $C^- = \sigma_{i_n}^- \ldots \sigma_{i_1}^-$. Both are endofunctors of $\mathrm{Rep}(\mathcal{Q})$.

REMARK 4.2.4. Given a Cartan matrix C, one has the following correspondence between Lie theory and the representation theory of finite-dimensional algebras:

Symmetric Kac–Moody algebra	$\leftrightarrow$	Hereditary path algebra
Symmetric Cartan matrix	$\leftrightarrow$	Symmetric Euler form
Root	$\leftrightarrow$	Dimension vector
Root system	$\leftrightarrow$	Auslander–Reiten quiver
Reflection s_i	$\leftrightarrow$	BGP reflection functor σ_i

5. Ringel–Hall algebras

In this section, we let k be the finite field $\mathbf{F}_{p^r}$. Let $\mathcal{Q}$ be a quiver without oriented cycles, and let $\Lambda = k\mathcal{Q}$ be the corresponding path algebra that is a finite-dimensional hereditary algebra over the finite field k. Recall that the index set $I = \mathcal{Q}_0$ and the symmetric Euler form of $\mathcal{Q}$ together give a realization of a Cartan datum Δ.

5.1. *Ringel–Hall algebras*

Let $\mathcal{P}$ be the set of isomorphism classes of finite-dimensional Λ-modules and $I \subset \mathcal{P}$ be the set of isomorphism classes of simple Λ-modules. We choose a representative $V_\alpha \in \alpha$ for any $\alpha \in \mathcal{P}$.

For $\alpha, \beta \in \mathcal{P}$, we write

$$\langle \alpha, \beta \rangle = \langle \underline{\dim}\, V_\alpha, \underline{\dim}\, V_\beta \rangle,$$

and that defines the Euler form $\langle -, - \rangle$ on $\mathbf{Z}^I$. The symmetric Euler form $(-, -)$ is given by $(\alpha, \beta) = \langle \alpha, \beta \rangle + \langle \beta, \alpha \rangle$ on $\mathbf{Z}^I$.

Let R be a commutative integral domain containing $\mathbf{Q}(v)$, where $v^2 = q$, $q = |k|$, and $\mathbf{Q}(v)$ is the rational function field of v.

Definition 5.1.1 (C. M. Ringel). The **Ringel–Hall algebra** $\mathcal{H}(\Lambda)$ is the free R-module with

- basis: $\{u_\alpha \mid \alpha \in \mathcal{P}\} = \{\text{isomorphism classes of f. d. } \Lambda - \text{modules}\}$;
- multiplication given by

$$u_\alpha u_\beta = v^{\langle \alpha, \beta \rangle} \sum_{\lambda \in \mathcal{P}} g^\lambda_{\alpha\beta} u_\lambda$$

for all $\alpha, \beta \in \mathcal{P}$, where $g^\lambda_{\alpha\beta}$ is the number of submodules V' of V_λ such that V_λ / V' and V' lie in the isomorphism classes α and β respectively, that is, $g^\lambda_{\alpha\beta} = \#\{V' \subseteq V_\lambda \mid V' \backsimeq V_\beta,\ V/V' \backsimeq V_\alpha\}$.

It is easy to verify that $\mathcal{H}(\Lambda)$ is an associative $\mathbf{N}^I$-graded R-algebra with the identity element u_0. The grading $\mathcal{H}(\Lambda) = \oplus_{\mathfrak{r} \in \mathbf{N}^I} \mathcal{H}_\mathfrak{r}$ is defined as follows: for each $\mathfrak{r} \in \mathbf{N}^I$, $\mathcal{H}_\mathfrak{r}$ is the R-span of the set $\{u_\lambda \mid \lambda \in \mathcal{P}, \underline{\dim}\, V_\lambda = \mathfrak{r}\}$.

We denote by $\mathcal{C}(\Lambda)$ the R-subalgebra of $\mathcal{H}(\Lambda)$, which is generated by $u_i, i \in I$. $\mathcal{C}(\Lambda)$ is called the **composition algebra** of Λ.

Lemma 5.1.2. *If Λ is of finite type, then $\mathcal{H}(\Lambda) = \mathcal{C}(\Lambda)$.*

Remark 5.1.3. If Λ is of finite type, whenever we change the ground field to be another finite field, the filtration number $g^\lambda_{\alpha\beta}$ can be evaluated by a polynomial at $|k|$, that is, $g^\lambda_{\alpha\beta} = \phi^\lambda_{\alpha\beta}(|k|)$ for some polynomial $\phi^\lambda_{\alpha\beta}$. The polynomial $\phi^\lambda_{\alpha\beta}$ occurring in this way is called a **Hall polynomial**.

C.M. Ringel [18, 19] has proved that the elements $u_i, i \in I$ satisfy the quantum Serre relations

$$\sum_{t=0}^{1-a_{ij}} (-1)^t \begin{bmatrix} 1 - a_{ij} \\ t \end{bmatrix} u_i^t u_j u_i^{1-a_{ij}-t} = 0,$$

for any $i \neq j$ in I, where $a_{ij} = \dfrac{2(i,j)}{(i,i)} = (i,j) = (\underline{\dim}\, S_i, \underline{\dim}\, S_j)$, and

$$[s] = \frac{v^s - v^{-s}}{v - v^{-1}} = v^{s-1} + v^{s-3} + \cdots + v^{-s+1}$$

$$[s]! = \prod_{r=1}^{s} [r]\,, \quad \begin{bmatrix} s \\ r \end{bmatrix} = \frac{[s]!}{[r]![s-r]!}$$

where r, s are non-negative integers, and $r \leq s$.

5.2. *Quantum groups and Ringel–Hall algebras*

Let $\bar{k}$ be the algebraic closure of k. For any $n \in \mathbf{N} \setminus \{0\}$, let $F(n)$ be a subfield of $\bar{k}$ such that $[F(n) : k] = n$. If we let $\Lambda(n) = \Lambda \otimes_k F(n)$, then $\Lambda(n)$ is a finite-dimensional hereditary $F(n)$-algebra corresponding to the same Cartan datum as that of Λ. We also have the Hall algebra $\mathcal{H}_n = \mathcal{H}_n(\Lambda(n))$ of $\Lambda(n)$. Define

$$\mathcal{H} = \prod_{n>0} \mathcal{H}_n.$$

Without creating confusion, we still use letter v to represent $v = (v_n)_n \in \mathcal{H}$, where $v_n = \sqrt{|F(n)|}$. Obviously, v lies in the center of $\mathcal{H}$ and is transcendental over the rational field $\mathbf{Q}$. Let $u_i = (u_i(n))_n \in \mathcal{H}$ satisfy that $u_i(n)$ is the element of $\mathcal{H}(\Lambda(n))$ corresponding to $S_i(n)$, where $S_i(n)$ is the simple $\Lambda(n)$-module that lies in the class i.

Definition 5.2.1. The **generic composition algebra** $\mathcal{C}^\star(\Lambda)$ of the Cartan datum Δ is defined to be the subring of $\mathcal{H}$ generated by $\mathbf{Q}$ and the elements v, v^{-1}, and u_i $(i \in I)$. The integral generic composition algebra ${}_{\mathcal{A}}\mathcal{C}^\star(\Lambda)$ is the $\mathcal{A}$–subalgebra of $\mathcal{C}^\star(\Lambda)$ generated by $u_i^{(m)} = \dfrac{u_i^m}{[m]!}$ $(i \in I)$, where $\mathcal{A} = \mathbf{Z}[v, v^{-1}]$.

Let $\mathbf{U}_v^+(\mathbf{g})$ be the positive part of the Drinfel'd–Jimbo quantum group corresponding to the Cartan datum Δ. A fundamental theorem of Green and Ringel establishes a wonderful connection between generic composition algebras and quantum groups.

Theorem 5.2.2 [18, 20]. *If $\mathfrak{g}$ and Λ have the same Cartan datum, then there is a $\mathbf{Q}(v)$-algebra isomorphism*

$$\mathcal{C}^\star(\Lambda) \simeq \mathbf{U}_v^+(\mathfrak{g}),$$

which maps $u_i \longrightarrow E_i$ for all $i \in I$. Even more, ${}_{\mathcal{A}}\mathcal{C}^\star(\Lambda)$ is isomorphic to ${}_{\mathcal{A}}\mathbf{U}_v^+(\mathbf{g})$ by sending $u_i^{(m)}$ to $E_i^{(m)}$ for all $i \in I$.

Thanks to this isomorphism, our discussion is applicable directly to generic composition algebras and quantum groups, even though we have started with our considerations with a composition algebra over finite field.

REMARK 5.2.3. If we extend $\mathbf{U}_v^+(\mathbf{g})$ to the Borel part of $\mathbf{U}_v(\mathbf{g})$ and $\mathcal{C}^\star(\Lambda)$ to a certain extension $\mathcal{C}(\Lambda, \mathbf{T})$, then η can be extended canonically to be a bijection of Hopf algebras (see [20–22]).

REMARK 5.2.4. If Λ is of finite type, the existence of Hall polynomials makes it possible to define the generic composition algebra by

$$u_\alpha u_\beta = v^{\langle\alpha,\beta\rangle} \sum_{\lambda\in\mathcal{P}} \phi_{\alpha\beta}^{\lambda}(v^2) u_\lambda,$$

where v is an indeterminate.

So far, all the Kac–Moody algebras considered have been symmetric. For nonsymmetric cases, one should use "species" (introduced by V. Dlab and C. M. Ringel in [1]) instead of quivers to construct an algebra that realizes the positive part of the corresponding quantum group.

6. PBW basis and canonical basis

Let $\mathfrak{g} = \mathfrak{g}(C)$ be the Kac–Moody algebra of type C, let $\mathbf{U} = \mathbf{U}_v(\mathfrak{g})$ be the corresponding quantum group, and let $\mathbf{U}^+ = \mathbf{U}_v^+(\mathbf{g})$ be its positive part. For simplicity, we only consider the case when the Cartan matrix C is symmetric.

Let $\mathcal{Q}$ be a quiver without oriented cycles whose Cartan matrix is C, and let $\Lambda = k\mathcal{Q}$ be the hereditary path algebra of $\mathcal{Q}$. We will not consider the quivers $\widetilde{A}_n$ with cyclic orientation in this chapter. Nevertheless, the subcategory of $\mathrm{Rep}(\widetilde{A}_n)$ consisting of all nilpotent representations of the cyclic quiver $\widetilde{A}_n$ plays a very important role in the representation theory of algebras.

Because of the canonical isomorphism between $\mathbf{U}^+$ and the generic composition algebra $\mathcal{C}^\star(\Lambda)$ of Λ (where $\mathfrak{g}$ and Λ have the same diagram), we are going to be able to use representation theory of algebras to analyze the structure of the corresponding Kac–Moody algebras and related quantum groups.

The so-called canonical basis of $\mathbf{U}^+$ was first introduced by G. Lusztig in [4] via an elementary algebraic definition in the case when $\mathfrak{g}$ is simple finite-dimensional Lie algebra. Subsequently, G. Lusztig studied canonical bases in the framework of Ringel–Hall algebras. The canonical basis was characterized by being integral, being invariant under the bar involution, and spanning a certain $\mathbf{Z}[v^{-1}]$-lattice $\mathbf{L}$ with a specific image in the quotient $\mathcal{L}/v^{-1}\mathcal{L}$, where $\mathcal{L}$ is generated by any basis of PBW type.

This algebraic definition does not work for quantum group of arbitrary Kac–Moody algebras. The difficulty in constructing a basis for $\mathcal{L}$ arises from the need to define suitable analogs of imaginary root vectors. By using topological methods, G. Lusztig [5] constructed the canonical basis of a quantum group of symmetric Kac–Moody algebra. It was shown that the "global crystal basis" introduced by M. Kashiwara and the canonical basis coincide. The canonical basis has very nice properties, but there are also still many unanswered questions.

There are different approaches to the construction of the canonical basis: an algebraic approach [4, 23–25], a geometric approach [5], and a quiver approach [26, 27]. We will not comment on the geometric approach here.

To construct the canonical basis, we start with a PBW basis, which involves root vectors.

6.1. *Root vectors in quantum groups*

According to G. Lusztig [28], there exists an action of the braid group corresponding to Δ on $\mathbf{U}$. Applying the standard generators $\mathcal{T}''_{i,1}$, $\mathcal{T}''^{-}_{i,1}$, $i \in I$ of the braid group to the generators of $\mathbf{U}$ in an admissible order, we obtain a family of linearly independent elements in $\mathbf{U}^+$. Since these elements degenerate into a basis of $\oplus_{\alpha\in\Phi^+_{\mathrm{real}}}\mathfrak{g}_\alpha$ by the specialization $v \to 1$, we call these elements the **real root vectors** of $\mathbf{U}^+$. If Δ is of finite type, they provide a complete set of root vectors.

Note that by using results of C. M. Ringel [27] and J. Xiao [22], the braid group action can be realized through BGP-reflection functors. The operations $\mathcal{T}''_{i,1}$ are given by σ_i on the Ringel–Hall algebra according to the canonical isomorphism $\mathcal{C}^\star(\Lambda) \backsimeq \mathbf{U}^+$.

For affine Kac–Moody algebra $\mathfrak{g}$, there exist imaginary roots. Several authors have introduced imaginary root vectors for $\mathbf{U}$ (see [23, 29–31]). Those imaginary root vectors cannot be obtained by Lusztig's operations.

Since the Auslander–Reiten quiver is a convenient device for visualizing the module category of a finite-dimensional algebra, we may ask what kind of information about root vectors in $\mathbf{U}^+$ can be read off from the Auslander–Reiten quiver of Λ. C. M. Ringel [27] gives an answer, providing an interpretation for the root vectors obtained by Lusztig's braid group action in $\mathcal{C}^\star(\Lambda)$.

6.2. *PBW basis and canonical basis for $\mathcal{C}^\star(\Lambda)$ of finite type*

We start with a path algebra $\Lambda = k\mathcal{Q}$, where k is a finite field and Q is of finite type. Let $\mathfrak{g}$ be a finite-dimensional simple Lie algebra of type A, D, or E, and let Φ^+ be the set of all positive roots, W its Weyl group, and $\mathbf{U}$ the corresponding quantum group.

Since Λ is of finite type, there are finitely many isomorphism classes of indecomposable Λ modules. The isomorphism classes of all indecomposables are in one-to-one correspondence with the positive roots Φ^+ of the corresponding simple Lie algebra. Every indecomposable Λ–module can be uniquely determined by a positive root $\alpha \in \Phi^+$. We will denote this indecomposable Λ-module by V_α if $\underline{\dim}\, V_\alpha = \alpha$. It is well known that $\mathrm{mod}-\Lambda$ (identified with $\mathrm{Rep}(\mathcal{Q})$) is representation-directed, (see [1, 32]); this means that there exists a total ordering of the positive roots, say $\beta_1, \beta_2, \ldots, \beta_N$, where N is the cardinality of Φ^+ such that $\mathrm{Hom}(V_{\beta_i}, V_{\beta_j}) \neq 0$ implies that $i \leq j$. Such an ordering is called $\mathcal{Q}$-admissible. Consider this $\mathcal{Q}$-admissible ordering of positive roots. There exists a sequence $i_1, \ldots, i_N$ of vertices of $\mathcal{Q}$, with the following properties:

(1) the sequence $i_N, \ldots, i_1$ is a sink sequence for $\mathcal{Q}$;
(2) we have $\beta_l = s_{i_1} s_{i_2} \ldots s_{i_{l-1}}(\alpha_{i_l})$, where α_{i_l} is a simple root.

There exists a reduced expression of the longest element

$$w_0 = s_{i_1} \dots s_{i_{N-1}} s_{i_N}$$

in W such that $i_N, i_{N-1}, \dots, i_1$ is a sink sequence for $\mathcal{Q}$ derived from an $\mathcal{Q}$-admissible ordering of positive roots. The element w_0 gives a total ordering on Φ^+, which is the same as the $\mathcal{Q}$-admissible ordering: $\beta_1 = \alpha_{i_1} < \beta_2 = s_{i_1}(\alpha_{i_2}) < \cdots < \beta_N = s_{i_1} s_{i_2} \cdots s_{i_{N-1}}(\alpha_{i_N})$. G. Lusztig defined the root vectors in $\mathbf{U}$ by

$$E_{\beta_l} := \mathcal{T}''_{i_1,1} \mathcal{T}''_{i_2,1} \dots \mathcal{T}''_{i_{l-1},1}(E_{\alpha_{i_l}}).$$

THEOREM 6.2.1 [27]. *In* $\mathcal{C}^\star(\Lambda)$,

$$E_{\beta_l} = v^\theta u_{\beta_l},$$

where $\theta = -\dim_k(V_{\beta_l}) + \dim_k \operatorname{End}_\Lambda V_{\beta_l}$.

Let $\mathbf{N}^{\Phi^+}$ be the set of multiplicity functions from $\Phi^+ = \{\beta_1, \beta_2, \dots, \beta_N\}$ to $\mathbf{N}$. Given $\mathbf{c} \in \mathbf{N}^{\Phi^+}$, we define

$$E^{\mathbf{c}} = E_{\beta_1}^{(\mathbf{c}(\beta_1))} E_{\beta_2}^{(\mathbf{c}(\beta_2))} \dots E_{\beta_N}^{(\mathbf{c}(\beta_N))},$$

where $E_{\beta_i}^{(r)} = \dfrac{E_{\beta_i}^r}{[r]!}$ for any $r \in \mathbf{N}$.

THEOREM 6.2.2 [4, 27]. $\{E^{\mathbf{c}} \mid \mathbf{c} \in \mathbf{N}^{\Phi^+}\}$ *is a PBW* $\mathcal{A}$*-basis of* ${}_{\mathcal{A}}\mathbf{U}^+$.

This **PBW basis** depends on the choice of a reduced expression of the longest element w_0.

THEOREM 6.2.3 [4, 24]. *For each given* $\mathbf{c} \in \mathbf{N}^{\Phi^+}$, *there exists a unique* $b(\mathbf{c}) \in_{\mathcal{A}} \mathbf{U}^+$ *such that*

(a) $\overline{b(\mathbf{c})} = b(\mathbf{c})$;
(b) $b(\mathbf{c}) = E^{\mathbf{c}} + \sum_{\mathbf{c} < \mathbf{c}'} a_{\mathbf{c},\mathbf{c}'} E^{\mathbf{c}'}$, *where* $a_{\mathbf{c},\mathbf{c}'} \in v^{-1}\mathbf{Z}[v^{-1}]$ *and* $<$ *is the lexicographic order on* $\mathbf{N}^{\Phi^+}$.

The basis $\{b(\mathbf{c}) \mid \mathbf{c} \in \mathbf{N}^{\Phi^+}\}$ *is the* ***canonical basis*** *introduced by G. Lusztig. This basis is* independent *of the choice of the reduced expression of the longest element* w_0.

A geometric realization of the canonical basis can be found in [4], where the theory of perverse sheaves is involved.

6.3. *Root vectors in* $\mathcal{C}^\star(\Lambda)$ *of tame type*

We start with a path algebra $\Lambda = k\mathcal{Q}$ where k is a finite field and Q is of tame type. Let $\mathfrak{g}$ be a symmetric affine Kac–Moody algebra of type $\widetilde{A}$, $\widetilde{D}$ or $\widetilde{E}$, let W be its Weyl group and $\mathbf{U}$ be the corresponding quantum group.

To construct a PBW basis and a canonical basis of $\mathbf{U}$, some new problems arise, which do not appear in the finite case (cf. [30]):

(1) there do exist imaginary roots $m\delta$ ($m \in \mathbf{Z}\backslash\{0\}$), which are fixed by the action of Weyl group, so we cannot apply the action of braid group to the simple root vectors to obtain the imaginary root vectors of a quantum group;
(2) there does no longer exist a longest element of the Weyl group. In the finite case, the longest element w_0 gives a way to totally order all positive roots. We need to find another way to order all roots;
(3) the real roots are divided into two disjoint orbits by the action of the Weyl group.

Let $I = \{0, 1, 2, \ldots, n\}$ be the index set of all simple roots of $\mathfrak{g}$. Define a doubly infinite sequence

$$\ldots i_{-1}, i_0, i_1, i_2 \ldots$$

by setting $i_l = i_{l(\mathrm{mod}\ N)}$, where N is the length of some special element in W. For a detailed description of this special element, we refer to [23, 24]. Note that $s_{i_m} \ldots s_{i_t}$ is reduced for any $m < t$.

Set

$$\beta_l = \begin{cases} s_{i_0} s_{i_{-1}} \ldots s_{i_{l+1}}(\alpha_{i_l}) & \text{if } l \leq 0, \\ s_{i_1} s_{i_2} \ldots s_{i_{l-1}}(\alpha_{i_l}) & \text{if } l > 0. \end{cases}$$

LEMMA 6.3.1 [23, 24]. $\{\beta_l \mid l \in \mathbf{Z}\} = \Phi^+_{\mathrm{real}}$.

The imaginary roots cannot be obtained in the same manner (through the action of Weyl group on simple roots).

Let δ be the smallest imaginary root. Then the imaginary roots of $\mathfrak{g}$ are the elements $l\delta$ for $l \in \mathbf{Z}$, $l \neq 0$ and mult $(l\delta) = n$. Let

$$\Phi^+_{\mathrm{img}} = \{(l\delta, t) \mid l > 0, 1 \leq t \leq n\}$$

be the set of all imaginary roots counted with multiplicity.

Define a total order on Φ^+ by setting

$$\beta_0 < \beta_{-1} < \cdots < (\delta, 1) < \cdots < (\delta, n) < (2\delta, 1) < \cdots (2\delta, n) < \cdots < \beta_2 < \beta_1.$$

We now define real root vectors for each positive real root as follows:

$$E_{\beta_l} = \begin{cases} \mathcal{T}''^{-1}_{i_0,1} \mathcal{T}''^{-1}_{i_{-1},1} \ldots \mathcal{T}''^{-1}_{i_{l+1},1}(E_{\alpha_{i_l}}) & \text{if } l \leq 0, \\ \mathcal{T}''_{i_1,1} \mathcal{T}''_{i_2,1} \ldots \mathcal{T}''_{i_{l-1},1}(E_{\alpha_{i_l}}) & \text{if } l > 0. \end{cases}$$

Having defined real root vectors, we continue to define the imaginary root vectors.

For $l > 0$, $i \in I\backslash\{0\}$, set

$$\widetilde{\psi}_{l\delta,i} = E_{l\delta-\alpha_i} E_{\alpha_i} - v^{-2} E_{\alpha_i} E_{l\delta-\alpha_i}.$$

Following [29], we introduce "integral" **imaginary root vectors** $E_{l\delta,i}$ ($l \geq 0, 1 \leq i \leq n$) in $\mathbf{U}^+$ by $E_{0,i} = 1$ and

$$E_{l\delta,i} = \frac{1}{[l]} \sum_{t=1}^{l} v^{t-l} \widetilde{\psi}_{t\delta,i} E_{(l-t)\delta,i}.$$

Let $\mathcal{Q}$ be a quiver with no oriented cycles and $|\mathcal{Q}_0| = n+1$. Let $i_{n+1}, i_n, \ldots, i_1$ be a sink sequence that uses each vertex of $\mathcal{Q}$ exactly once. We can define the Coxeter functors: $C^+ = \sigma^+_{i_1} \ldots \sigma^+_{i_{n+1}}$ and $C^- = \sigma^-_{i_{n+1}} \ldots \sigma^-_{i_1}$. Both are endofunctors of mod $-\Lambda$ (or Rep($\mathcal{Q}$)). According to Theorem 4.1.1, if $\alpha \in \Phi^+_{\text{real}}$, then there is a unique indecomposable Λ–module of dimension vector α (up to isomorphism) and if $\alpha \in \Phi^+_{\text{img}}$, then there are many indecomposable Λ–modules of dimension vector α (up to isomorphism). We can use the Coxeter functors $C^\pm$ to partition all indecomposable Λ-modules.

Definition 6.3.2. Given an indecomposable Λ-module V, we call it

- **preprojective** indecomposable module if $(C^+)^r V = 0$ for $r \gg 0$;
- **preinjective** indecomposable module if $(C^-)^r V = 0$ for $r \gg 0$;
- **regular** indecomposable module if it is neither preprojective nor preinjective. Moreover, $(C^\pm)^r V \simeq V$ for some $r > 0$.

All preprojective and preinjective modules may be constructed in a uniform way through the action of BGP reflection functors.

For $1 \leq l \leq n+1$, we can obtain all indecomposable projective Λ-modules by

$$\mathbf{P}_{i_l} = \sigma^-_{i_{n+1}} \sigma^-_{i_n} \sigma^-_{i_{l+1}} (S_{i_l}),$$

where S_{i_l} is the simple module of the path algebra of $\sigma_{i_l} \sigma_{i_{l-1}} \ldots \sigma_{i_1} \mathcal{Q}$.

Dually, all indecomposable injective Λ-modules are given by

$$\mathbf{I}_{i_l} = \sigma^+_{i_1} \sigma^+_{i_2} \sigma^+_{i_{l-1}} (S_{i_l}),$$

where S_{i_l} is the simple module of the path algebra of $\sigma_{i_l} \sigma_{i_{l+1}} \ldots \sigma_{i_{n+1}} \mathcal{Q}$.

If V is an indecomposable preprojective module, then

$$V \simeq (C^-)^t (\mathbf{P}_{i_l}) = \left(\sigma^-_{i_{n+1}} \ldots \sigma^-_{i_1} \right)^t \left(\sigma^-_{i_{n+1}} \sigma^-_{i_n} \sigma^-_{i_{l+1}} (S_{i_l}) \right)$$

for some $t \geq 0$. Even more, $\underline{\dim}\, V = (s_{i_{n+1}} \ldots s_{i_1})^r (s_{i_{n+1}} s_{i_n} s_{i_{l+1}} (\alpha_{i_l}))$.

If V is an indecomposable preinjective module, then

$$V \simeq (C^+)^t (\mathbf{I}_{i_l}) = \left(\sigma^+_{i_1} \ldots \sigma^+_{i_{n+1}} \right)^t \left(\sigma^+_{i_1} \sigma^+_{i_2} \sigma^+_{i_{l-1}} (S_{i_l}) \right)$$

for some $t \geq 0$. Also $\underline{\dim}\, V = (s_{i_1} \ldots s_{i_{n+1}})^r (s_{i_1} s_{i_2} s_{i_{l-1}} (\alpha_{i_l}))$.

Using a sink sequence $i_{n+1}, i_n, \ldots, i_1$ of all vertices of $\mathcal{Q}$, we can define a doubly infinite sequence

$$\ldots, i_{-1}, i_0, i_1, i_2, \ldots$$

by setting $i_l = i_{l(\text{mod } n+1)}$ for $l \in \mathbf{Z}$.

Set

$$\hat{\beta}_l = \begin{cases} s_{i_0} s_{i_{-1}} \dots s_{i_{l+1}}(\alpha_{i_l}) & \text{if } l \leq 0, \\ s_{i_1} s_{i_2} \dots s_{i_{l-1}}(\alpha_{i_l}) & \text{if } l > 0. \end{cases}$$

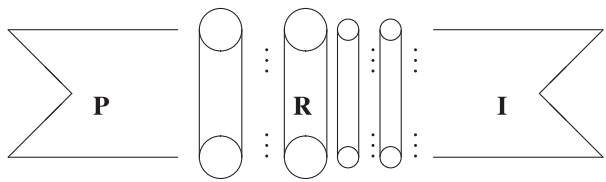
We have already shown that $\hat{\beta}_l$ for $l \leq 0$ is the dimension vector of an indecomposable preprojective module and $\hat{\beta}_l$ for $l > 0$ is the dimension vector of an indecomposable preinjective module. Using the actions of BGP reflections on simple modules, we can only get all the indecomposable preprojective and preinjective modules. According to the Auslander–Reiten quiver of Λ, the set Φ^+ can be partitioned as follows. The roots

$$\{\hat{\beta}_l \mid l \in \mathbf{Z}\}$$

are called **irregular roots**. The dimension vector of any indecomposable regular module is called a **regular root**. Hence, we have

$$\Phi^+ = \{\text{irregular roots}\} \sqcup \{\text{regular roots}\}.$$

The Auslander–Reiten quiver of Λ looks as follows:

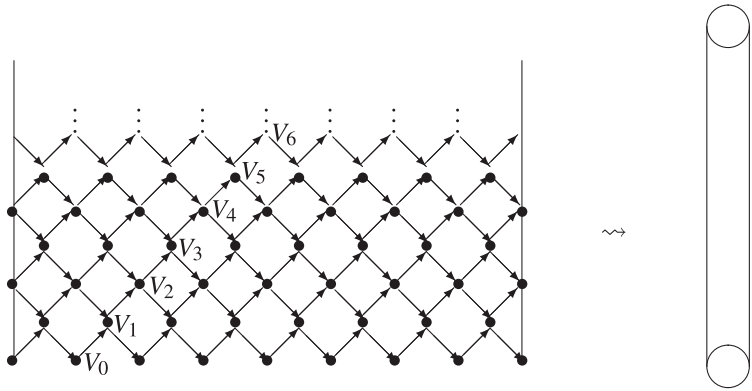


P : Preprojective component

R : Regular components

I : Preinjective component

There exists a unique preprojective (or preinjective) component that consists of all indecomposable preprojective (or preinjective) modules. There are lots of regular components containing all indecomposable regular modules. We have the following geometric picture of a regular component of the AR–quiver:



Here there is the possibility that $(C^+)^n V_0 \backsimeq V_0$ for some $n > 0$. Then $(C^+)^n V_i \backsimeq V_i$ for all i and we can identify all points on those two vertical lines to obtain what is called a stable tube. If $n = 1$, the tube is called homogeneous stable tube.

Note that only positive real root uniquely determines an indecomposable Λ-module (up to isomorphism).

LEMMA 6.3.3 [27]. $\{\hat{\beta}_l \mid l \in \mathbf{Z}\} \subseteq \Phi^+_{\text{real}}$ *and* {*Regular roots*} $= \Phi^+_{\text{img}} \sqcup$ {*Regular real roots*}.

We can order all irregular roots as

$$\hat{\beta}_0 < \hat{\beta}_{-1} < \hat{\beta}_{-2} < \cdots < \hat{\beta}_3 < \hat{\beta}_2 < \hat{\beta}_1.$$

Let $V_{\hat{\beta}_l}$ for $l \in \mathbf{Z}$ be the indecomposable preprojective or preinjective module with $\underline{\dim}\, V_{\hat{\beta}_l} = \hat{\beta}_l$. The above order of positive roots is compatible with the structure of the preprojective and preinjective components of the Auslander–Reiten quiver of Λ, which means $\operatorname{Hom}_\Lambda(V_{\hat{\beta}_j}, V_{\hat{\beta}_t}) \neq 0$ implies $\hat{\beta}_j < \hat{\beta}_t$.

Similarly as before, we can define real root vectors for each irregular real root:

$$E_{\hat{\beta}_l} = \begin{cases} \mathcal{T}''_{i_0,1}{}^{-1} \mathcal{T}''_{i_{-1},1}{}^{-1} \cdots \mathcal{T}''_{i_{l+1},1}{}^{-1}(E_{\alpha_{i_l}}) & \text{if } l \leq 0, \\ \mathcal{T}''_{i_1,1} \mathcal{T}''_{i_2,1} \cdots \mathcal{T}''_{i_{l-1},1}(E_{\alpha_{i_l}}) & \text{if } l > 0. \end{cases}$$

Recall that $\mathcal{T}''_{i,1}$ can be realized by σ_i under the identification $\mathcal{C}^\star(\Lambda) \backsimeq \mathbf{U}^+$.

THEOREM 6.3.4 [27]. *Let $V_{\hat{\beta}_l}$ be an indecomposable preprojective or preinjective Λ-module with $\underline{\dim}\, V_{\hat{\beta}_l} = \hat{\beta}_l$ for $l \in \mathbf{Z}$, then in $\mathcal{C}^\star(\Lambda)$ (or $\mathbf{U}^+$),*

$$E_{\hat{\beta}_l} = v^\theta u_{\hat{\beta}_l},$$

where $\theta = -\dim_k(V_{\hat{\beta}_l}) + \dim_k \operatorname{End}_\Lambda V_{\hat{\beta}_l}$.

PROPOSITION 6.3.5 [33, 34]. *There is an algorithm to express $E_{\hat{\beta}_l}$ as linear combination of simple root vectors.*

In this way, we only construct irregular real root vectors. Now the natural next question is how to construct imaginary and regular real root vectors in Ringel–Hall algebras. For the quiver of type $\widetilde{A_1}$—the Kronecker quiver, we answer the question in next subsection. The question still remains open for the general case.

6.4. *Kronecker quiver*

In this section, we only consider the Kronecker quiver $\mathcal{Q} : 2 \bullet \rightrightarrows \bullet 1$.

Let Φ^+ be the positive root system of Lie algebra $\widehat{sl_2}$. Let k be a finite field. The path algebra $\mathcal{K} = k\mathcal{Q}$ is the "smallest" tame hereditary algebra. We will provide a

realization of all imaginary root vectors in the generic composition algebra $\mathcal{C}^\star(\mathcal{K})$. Moreover, an integral PBW basis and the canonical basis of this algebra are obtained. The corresponding quantum group is $\mathbf{U}_v(\widehat{sl_2})$.

The Auslander–Reiten quiver of $\mathcal{K}$ consists of one preprojective component, one preinjective component (which have the same structure as those components of $k\mathcal{Q}$ for k being algebraically closed field), and a family of homogeneous tubes of regular modules parameterized by the set of all monic irreducible polynomials over k (which is different from that of $k\mathcal{Q}$ when k is an algebraically closed field).

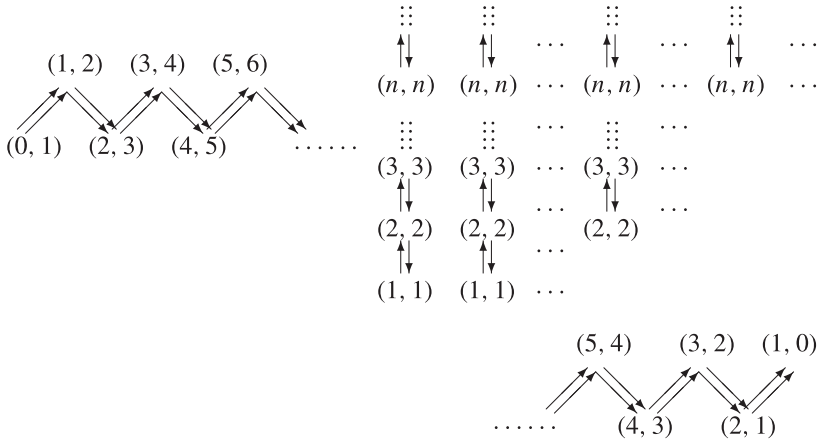
Let $\alpha_1 = \underline{\dim}\, S_1$ and $\alpha_2 = \underline{\dim}\, S_2$. Put $\delta = \alpha_1 + \alpha_2$. The indecomposable preprojective and preinjective modules have the dimension vectors $n\delta + \alpha_1$ and $n\delta + \alpha_2$ with $n \in \mathbf{N}$, respectively. Moreover, each indecomposable preprojective and preinjective module is uniquely determined by its dimension vector. The indecomposable regular modules have the dimension vectors $n\delta$ with $n \in \mathbf{N} \setminus \{0\}$. There is no regular real root. Every regular root is an imaginary root.

Put $(r, t) = r\alpha_1 + t\alpha_2$. We have

$$\Phi^+ = \{(l+1, l), (m, m), (n, n+1) \mid l \geq 0, m \geq 1, n \geq 0\},$$

which is the set of dimension vector of all indecomposable representations of the Kronecker quiver.

The Auslander–Reiten quiver of $\mathcal{K}$ looks as follows:



In $\mathbf{U}^+$, we have for $i \neq j \in \{1, 2\}$

$$E_{n\delta+\alpha_1} = \underbrace{\mathcal{T}''_{1,1}{}^{-1}\mathcal{T}''_{2,1}{}^{-1}\mathcal{T}''_{1,1}{}^{-1}\dots\mathcal{T}''_{i,1}{}^{-1}}_{n \text{ times}}(E_j),$$

$$E_{n\delta+\alpha_2} = \underbrace{\mathcal{T}''_{2,1}\mathcal{T}''_{1,1}\mathcal{T}''_{2,1}\dots\mathcal{T}''_{i,1}}_{n \text{ times}}(E_j),$$

where $\mathcal{T}''_{1,1}$ and $\mathcal{T}''_{2,1}$ are the braid group operations defined in [28].

According to C. M. Ringel [27], we can define the following root vectors in $\mathcal{C}^\star(\mathcal{K})$, which correspond to all positive real roots:

$$E_1 = E_{\alpha_1} = u_1, \qquad E_2 = E_{\alpha_2} = u_2,$$

$$E_{n\delta+\alpha_1} = v^{-\dim_k V_{n\delta+\alpha_1}+\dim_k \mathrm{End}_\Lambda(V_{n\delta+\alpha_1})} u_{n\delta+\alpha_1} = v^{-2n} u_{n\delta+\alpha_1},$$

$$E_{n\delta+\alpha_2} = v^{-\dim_k V_{n\delta+\alpha_2}+\dim_k \mathrm{End}_\Lambda(V_{n\delta+\alpha_2})} u_{n\delta+\alpha_2} = v^{-2n} u_{n\delta+\alpha_2}.$$

for any $n \in \mathbf{N}$.

Following I. Damiani [30] (or [23]), there are the imaginary root vectors

$$\widetilde{\psi}_{n\delta} = E_{n\delta-\alpha_1} E_1 - v^{-2} E_1 E_{n\delta-\alpha_1}, \quad \text{for } n \in \mathbf{N}\backslash\{0\},$$

and integral imaginary root vectors

$$E_{0\delta} = 1 \text{ and } E_{n\delta} = \frac{1}{[n]} \sum_{i=1}^{n} v^{i-n} \widetilde{\psi}_{i\delta} E_{(n-i)\delta}.$$

THEOREM 6.4.1 [35, 36]. *In $\mathcal{C}^\star(\mathcal{K})$, we have*

$$E_{n\delta} = v^{-2n} \sum_{\substack{\alpha \in \mathcal{P},\, V_\alpha \text{ regular},\\ \underline{\dim} V_\alpha = n\delta}} u_\alpha.$$

The total ordering on Φ^+ (adapted for the structure of the Auslander–Reiten quiver) is as follows:

$$\alpha_1 < \delta + \alpha_1 < \cdots < (n-1)\delta + \alpha_1 < n\delta + \alpha_1 < \cdots$$
$$\cdots < \delta < 2\delta < \cdots < (n-1)\delta < n\delta < \cdots$$
$$< n\delta + \alpha_2 < (n-1)\delta + \alpha_2 < \cdots < \delta + \alpha_2 < \alpha_2.$$

Let $\mathbf{N}^{\Phi^+}$ be the set of finitely supported multiplicity functions from Φ^+ to $\mathbf{N}$. Given $\mathbf{c} \in \mathbf{N}^{\Phi^+}$, if

$$\{\alpha \in \Phi^+ \mid \mathbf{c}(\alpha) \neq 0\} = \{\beta_1 < \beta_2 < \ldots \beta_n\},$$

define $E^{\mathbf{c}}$ be the monomial formed by multiplying the divided powers of real root vectors and imaginary root vectors, that is,

$$E^{\mathbf{c}} = E_{\beta_1}^{(\mathbf{c}(\beta_1))} E_{\beta_2}^{(\mathbf{c}(\beta_2))} \ldots E_{\beta_n}^{(\mathbf{c}(\beta_n))},$$

where $E_{\beta_m}^{(\mathbf{c}(\beta_m))} = \dfrac{E_{\beta_1}^{\mathbf{c}(\beta_1)}}{[m]!}$ if β_m is real, and $E_{\beta_m}^{(\mathbf{c}(\beta_m))} = E_{\beta_m}^{\mathbf{c}(\beta_m)}$ if β_m is imaginary.

Then, there is the following theorem:

THEOREM 6.4.2 [35, 37]. *The set $\{E^{\mathbf{c}} \mid \mathbf{c} \in \mathbf{N}^{\Phi^+}\}$ is a PBW $\mathcal{A}$–basis of ${}_{\mathcal{A}}\mathcal{C}^\star(\mathcal{K})$ (or ${}_{\mathcal{A}}\mathbf{U}_v^+(\widehat{sl_2})$).*

For $\mathbf{d} = (d_1, d_2) \in \mathbf{N}^2$, write $E(\mathbf{d}) = E_2^{(d_2)} E_1^{(d_1)}$. Similarly, for $\mathbf{c} \in \mathbf{N}^{\Phi^+}$, define

$$E(\mathbf{c}) = E(\mathbf{c}(\beta_1))E(\mathbf{c}(\beta_2)) \dots E(\mathbf{c}(\beta_m)),$$

which is a monomial on the divided powers of the Chevalley generators E_1 and E_2. Therefore, $E(\mathbf{c}) \in \mathcal{C}^\star(\mathcal{K})$.

We will now describe a triangular relation between the PBW basis and the above monomial basis.

Let $\varphi : \mathbf{N}^{\Phi^+} \to \mathbf{N}^2$ be the map given by

$$\varphi(\mathbf{c}) = \sum_{\alpha \in \Phi^+} \mathbf{c}(\alpha)\alpha.$$

Then for any $\mathbf{d} \in \mathbf{N}^2$, $\varphi^{-1}(\mathbf{d})$ gives the size of all isomorphism classes of modules, which have the dimension vector $\mathbf{d}$ and is a finite set.

PROPOSITION 6.4.3 [26]. *For any* $\mathbf{c} \in \mathbf{N}^{\Phi^+}$,

$$E(\mathbf{c}) = \sum_{\mathbf{c}' \in \varphi^{-1}(\varphi(\mathbf{c}))} h_{\mathbf{c}'}^{\mathbf{c}} E^{\mathbf{c}'}$$

such that

(1) $h_{\mathbf{c}'}^{\mathbf{c}} \in \mathbf{Z}[v, v^{-1}]$;

(2) $h_{\mathbf{c}}^{\mathbf{c}} = 1$;

(3) *if* $h_{\mathbf{c}'}^{\mathbf{c}} \neq 0$ *and* $\mathbf{c}' \neq \mathbf{c}$, *then* $\mathbf{c}' < \mathbf{c}$, *where* $<$ *is the geometric order related to the corresponding modules*;

(4) $E(\mathbf{c})$ *is bar-invariant, that is,* $\overline{E(\mathbf{c})} = E(\mathbf{c})$.

PROPOSITION 6.4.4 [26]. *For any* $\mathbf{c} \in \mathbf{N}^{\Phi^+}$, *set*

$$\overline{E^{\mathbf{c}}} = \sum_{\mathbf{c}' \in \mathbf{N}^{\Phi^+}} \omega_{\mathbf{c}'}^{\mathbf{c}} E^{\mathbf{c}'}.$$

Then,

(1) $\omega_{\mathbf{c}'}^{\mathbf{c}} \in \mathbf{Z}[v, v^{-1}]$;

(2) $\omega_{\mathbf{c}}^{\mathbf{c}} = 1$;

(3) *if* $\omega_{\mathbf{c}'}^{\mathbf{c}} \neq 0$ *and* $\mathbf{c}' \neq \mathbf{c}$, *then* $\mathbf{c}' < \mathbf{c}$.

Consider the bar involution in Section 2, we have $\overline{\overline{E^{\mathbf{c}}}} = E^{\mathbf{c}}$ for any $\mathbf{c} \in \mathbf{N}^{\Phi^+}$. Hence,

$$\overline{\overline{E^{\mathbf{c}}}} = \overline{\sum_{\mathbf{c}'} \omega_{\mathbf{c}'}^{\mathbf{c}} E^{\mathbf{c}'}} = \sum_{\mathbf{c}', \mathbf{c}''} \overline{\omega_{\mathbf{c}'}^{\mathbf{c}}} \omega_{\mathbf{c}''}^{\mathbf{c}'} E^{\mathbf{c}''}.$$

Since $\{E^{\mathbf{c}} \mid \mathbf{c} \in \mathbf{N}^{\Phi^+}\}$ is an $\mathcal{A}$–basis of $\mathcal{C}^\star(\mathcal{K})$, we have

$$\sum_{\mathbf{c}'} \overline{\omega_{\mathbf{c}'}^{\mathbf{c}}} \omega_{\mathbf{c}''}^{\mathbf{c}'} = \delta_{\mathbf{c}\mathbf{c}''},$$

and therefore, one can uniquely solve the system of equations

$$\zeta_{\mathbf{c}'}^{\mathbf{c}} = \sum_{\mathbf{c}' \le \mathbf{c}'' \le \mathbf{c}} \omega_{\mathbf{c}'}^{\mathbf{c}''} \overline{\zeta_{\mathbf{c}''}^{\mathbf{c}}}$$

with unknowns $\zeta_{\mathbf{c}'}^{\mathbf{c}} \in \mathbf{Z}[v^{-1}]$ such that

(1) $\zeta_{\mathbf{c}}^{\mathbf{c}} = 1$;

(2) $\zeta_{\mathbf{c}'}^{\mathbf{c}} \in v^{-1}\mathbf{Z}[v^{-1}]$ for all $\mathbf{c}' < \mathbf{c}$.

For any $\mathbf{c} \in \varphi^{-1}(\mathbf{d})$ and $\mathbf{d} \in \mathbf{N^2}$, set

$$\varepsilon^{\mathbf{c}} = \sum_{\mathbf{c}' \in \varphi^{-1}(\mathbf{d})} \zeta_{\mathbf{c}'}^{\mathbf{c}} E^{\mathbf{c}'}.$$

THEOREM 6.4.5 [26]. *The set* $\{\varepsilon^{\mathbf{c}} \mid \mathbf{c} \in \varphi^{-1}(\mathbf{d}), \mathbf{d} \in \mathbf{N^2}\}$ *provides the canonical basis of* ${}_{\mathcal{A}}\mathcal{C}^\star(\mathcal{K})$.

The importance of the Kronecker quiver lies in the existence of a full exact embedding from the category of regular modules of $\mathcal{K}$ to the category of regular modules of any tame hereditary algebra with underlining quiver $\widetilde{A}_n$, $\widetilde{D}_n$, $\widetilde{E}_6$, $\widetilde{E}_7$, or $\widetilde{E}_8$.

THEOREM 6.4.6 [1]. *There exists a fully faithful exact functor*

$$\mathcal{F} : \mathrm{mod} - \mathcal{K} \longmapsto \mathrm{mod} - k\mathcal{Q},$$

which preserves regular modules, where $\mathcal{Q} = \widetilde{A}_n$ (*excluding the case of a cyclic orientation*), $\widetilde{D}_n$, $\widetilde{E}_6$, $\widetilde{E}_7$, *or* $\widetilde{E}_8$. *Even more, this functor gives rise to an injection of Ringel–Hall algebras* (*which will still be denoted by* $\mathcal{F}$).

More recently, P. Baumann and C. Kassel [38] described the Ringel–Hall algebra of the category of coherent sheaves on the projective line and recovered M. Kapranov's isomorphism between a certain subalgebra of this Ringel–Hall algebra and the "positive part" of $\mathbf{U}_v(\widehat{sl_2})$. In combination with the isomorphism between the generic composition algebra $\mathcal{C}^\star(\mathcal{K})$ and $\mathbf{U}_v^+(\widehat{sl_2})$, the real root vectors $E_{n\delta+\alpha_1}$ and the imaginary root vectors $E_{n\delta}$ are related to locally free coherent sheaves and the torsion sheaves, respectively.

6.5. *PBW basis and canonical basis for* $\mathcal{C}^\star(\Lambda)$ *of tame type*

We use the same assumptions presented in Subsection 6.3. There are different approaches to the construction of the canonical basis of $\mathbf{U}^+$. G. Lusztig [5]

parameterized the canonical basis elements by using the representation theory of $\mathcal{Q}$ and algebraic geometry. J. Beck, V. Chari, and A. Pressley [23] provided an algebraic characterization of the canonical basis. J. Beck and H. Nakajima [24] obtained a purely algebraic construction of the PBW basis. Z. Lin, J. Xiao, and G. Zhang [26] used the Ringel–Hall algebra approach to construct the PBW basis and the canonical basis of $\mathcal{C}^\star(\Lambda)$.

Let $\Phi^+_{\text{prep}} = \{\hat{\beta}_l \mid l \le 0\}$ and $\Phi^+_{\text{prei}} = \{\hat{\beta}_l \mid l > 0\}$. Let $\mathbf{N}^{\Phi^+_{\text{prep}}}$ be the set of finitely supported multiplicity functions from Φ^+_{prep} to $\mathbf{N}$. Similarly, we define $\mathbf{N}^{\Phi^+_{\text{prei}}}$.

Given $\mathbf{c}_+ \in \mathbf{N}^{\Phi^+_{\text{prep}}}$, if

$$\{\alpha \in \Phi^+_{\text{prep}} \mid \mathbf{c}_+(\alpha) \neq 0\} = \{\hat{\beta}'_1 < \hat{\beta}'_2 < \cdots < \hat{\beta}'_n\},$$

define $E^{\mathbf{c}_+}$ to be the monomial formed by

$$E^{\mathbf{c}_+} = E_{\hat{\beta}'_1}^{(\mathbf{c}(\hat{\beta}'_1))} E_{\hat{\beta}'_2}^{(\mathbf{c}(\hat{\beta}'_2))} \cdots E_{\hat{\beta}'_n}^{(\mathbf{c}(\hat{\beta}'_n))},$$

where $E_{\hat{\beta}'_m}^{(\mathbf{c}(\hat{\beta}'_m))} = \dfrac{E^{\mathbf{c}(\hat{\beta}'_m)}_{\hat{\beta}'_m}}{[m]!}$. Similarly, define $E^{\mathbf{c}_-}$ for $\mathbf{c}_- \in \mathbf{N}^{\Phi^+_{\text{prei}}}$.

We now define a set $\mathcal{M} = \{(\mathbf{c}_+, (\pi_1, \pi_2, \ldots, \pi_s), w, \mathbf{c}_-)\}$ by the following rule:

- $(\pi_1, \pi_2, \ldots, \pi_s)$ is a collection of partitions and each π_t is related to a non-homogeneous tube;
- w is a partition $(w_1, w_2, \ldots, w_t)$ for some t.

Then for each $\mathbf{c} \in \mathcal{M}$, define

$$E^{\mathbf{c}} = E^{\mathbf{c}_+} E_{\pi_1} E_{\pi_2} \cdots E_{\pi_s} \mathcal{F}(E_{w_1\delta}) \mathcal{F}(E_{w_2\delta}) \cdots \mathcal{F}(E_{w_t\delta}) E^{\mathbf{c}_+},$$

where $E_{\pi_1} E_{\pi_2} \cdots E_{\pi_s}$ is the product of some elements in $\mathcal{C}^\star(\Lambda)$, which are related to all nonhomogeneous tubes, and F is the functor mentioned in 6.4.6. We omit the detailed discussion; interested readers are referred to [26].

THEOREM 6.5.1 [26]. *The set $\{E^{\mathbf{c}} \mid \mathbf{c} \in \mathcal{M}\}$ is an $\mathcal{A}$-basis of ${}_{\mathcal{A}}\mathcal{C}^\star(\Lambda)$. The set $\{\varepsilon^{\mathbf{c}} \mid \mathbf{c} \in \mathcal{M}\}$ provides the canonical basis of ${}_{\mathcal{A}}\mathcal{C}^\star(\Lambda)$, where $\varepsilon^{\mathbf{c}}$ is defined in a similar manner as was done for the Kronecker quiver.*

7. Root categories, Kac–Moody algebras and elliptic Lie algebras

7.1. *A construction of the Kac–Moody algebra $\mathfrak{g}'(C)$.*

Let Λ be a finite-dimensional hereditary algebra over a field k (or, more specifically, $\Lambda = k\mathcal{Q}$ for some quiver $\mathcal{Q}$ without oriented cycles). Let $\mathfrak{g}(C)$ be a symmetrizable Kac–Moody algebra associated to a symmetrizable Cartan matrix C, and let $\mathfrak{g}'(C)$ be its derived algebra, that is, $\mathfrak{g}'(C) = [\mathfrak{g}(C), \mathfrak{g}(C)]$.

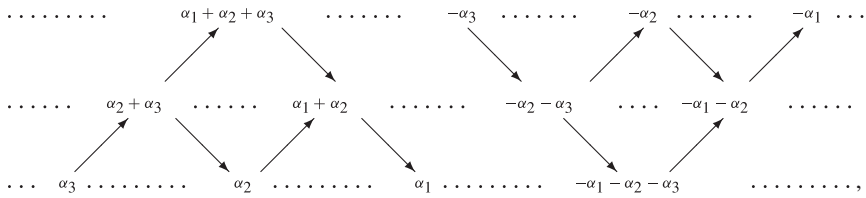
According to Theorem 4.1.1, if Λ and $\mathfrak{g}(C)$ (or $\mathfrak{g}'(C)$) share the same Cartan datum, then the map $\underline{\dim}$ induces a surjection between the isomorphism classes of all

indecomposable Λ–modules and the positive roots of $\mathfrak{g}(C)$ (or $\mathfrak{g}'(C)$). Moreover, if Λ is of finite type, the map $\underline{\dim}$ is bijective. It is natural to ask whether it is possible to recover the Lie algebra structure of $\mathfrak{g}'(C)$ in terms of Λ–modules. For Λ of finite type, this was successfully done by C. M. Ringel [39,40] via the Ringel–Hall algebra construction. Let $K(\mathrm{mod}-\Lambda)$ be the free Abelian group with basis indexed by the set of isomorphism classes of indecomposable Λ–modules. C.M. Ringel showed that $K(\mathrm{mod}-\Lambda)$ is the Chevalley $\mathbf{Z}$-form of $\mathfrak{n}^+$ and hence $K(\mathrm{mod}-\Lambda)\bigotimes_{\mathbf{Z}}\mathbf{C}$ and $\mathfrak{n}^+$ are isomorphic Lie algebras, where $\mathfrak{g}'(C)=\mathfrak{n}^-\bigoplus\mathfrak{h}'\bigoplus\mathfrak{n}^+$ is a triangular decomposition of $\mathfrak{g}'(C)$.

One of the important contributions to the representation theory of algebras is the study of derived categories by D. Happel [41,42]. The **derived category** $\mathcal{D}^b(\Lambda)$ is a **triangulated category** with the translation $\mathcal{T}$ obtained from the category of bounded complexes over Λ by localizing with respect to the set of all **quasi-isomorphisms**. The interested readers are referred to B. Keller's paper [43] in an earlier volume of this handbook for a detailed description of derived categories and their uses.

There is a full embedding of $\mathrm{mod}-\Lambda$ into $\mathcal{D}^b(\Lambda)$ as a full subcategory, which sends each Λ-module X into the **stalk complex** $X^\bullet=(X^i,d^i)$ with $X^0=X$, $X^i=0$ for $i\neq 0$, and $d^i=0$ for all $i\in\mathbf{Z}$. We keep using the same symbol $\mathrm{mod}-\Lambda$ for this subcategory. The dimension vectors of all indecomposable complexes in $\mathrm{mod}-\Lambda$ are all positive roots and the dimension vectors of all indecomposable complexes in $\mathcal{T}(\mathrm{mod}-\Lambda)$ are all negative roots.

EXAMPLE 7.1.1. Consider the quiver $\mathcal{Q}:\ \overset{1}{\bullet}\longrightarrow\overset{2}{\bullet}\longrightarrow\overset{3}{\bullet}$. The categorical structure of the derived category $\mathcal{D}^b(\Lambda)$ can be visualized as follows:



where each indecomposable object is represented by its dimension vector.

To recover the whole Lie algebra $\mathfrak{g}'(C)$ (not only $\mathfrak{n}^+$) using modules of algebras, we should consider the **orbit category** $\mathcal{R}=\mathcal{D}^b(\Lambda)/\mathcal{T}^2$ of $\mathcal{D}^b(\Lambda)$ under the automorphism $\mathcal{T}^2$. The translation $\mathcal{T}$ of a derived category $\mathcal{D}^b(\Lambda)$ induces an automorphism of $\mathcal{R}$ of order 2, which is again denoted by $\mathcal{T}$. $\mathcal{R}$ inherits a triangulated category structure from $\mathcal{D}^b(\Lambda)$ with $\mathcal{T}$ as the translation. Note that $\mathrm{mod}-\Lambda$ can be embedded into $\mathcal{R}$ as a full subcategory. The root category $\mathcal{R}$ is divided by $\mathrm{mod}-\Lambda$ into two parts, $\mathrm{mod}-\Lambda$ and $\mathcal{T}(\mathrm{mod}-\Lambda)$.

D. Happel [41] pointed out that if Λ is of finite type, then the isomorphism classes of indecomposable objects in $\mathcal{R}$ correspond bijectively to all roots of $\mathfrak{g}'(C)$ (not only positive roots). For this reason, $\mathcal{R}$ is called the **root category**. It was soon found that

the category $\mathcal{R} = \mathcal{D}^b(\Lambda)/\mathcal{T}^2$ provided a solid framework, which is compatible with the root system of $\mathfrak{g}'(C)$ even when Λ is not of finite type.

In [11], V. Kac asked a fundamental question about the realization of nonaffine Lie algebras. This question was answered by L. Peng and J. Xiao in [2]. Not only that, the authors actually gave a realization of all symmetrizable Kac–Moody algebras in a global way via the root categories $\mathcal{R} = \mathcal{D}^b(\Lambda)/\mathcal{T}^2$. In the remainder of this section, we will summarize their work.

Assume now that the field k is a finite field and $q = |k|$, the cardinality of k.

Given a root category $\mathcal{R}$ of Λ, we consider the Grothendieck group $G(\mathcal{R})$ as usual, that is, $G(\mathcal{R})$ is the quotient of a free Abelian group with basis $\{[M] \mid M \in \mathcal{R}\}$ indexed by the isomorphism classes of all objects in $\mathcal{R}$, subject to the relations $[Y]-[X]-[Z]$ arising from the triangles of form $X \to Y \to Z \to \mathcal{T}X$ in $\mathcal{R}$. For any $M \in \mathcal{R}$, we denote by $\underline{\dim}\, M$ the canonical image of M in $G(\mathcal{R})$, called the **dimension vector** of M.

Given $X, Y, L \in \mathcal{R}$, put

$$W(X, Y, L) = \{(f, g, h) \in \operatorname{Hom}(X, L) \times \operatorname{Hom}(L, Y) \times \operatorname{Hom}(Y, \mathcal{T}X) \mid X \xrightarrow{f} L \xrightarrow{g} Y \xrightarrow{h} \mathcal{T}X \text{ is a triangle}\}.$$

One can define an action of $\operatorname{Aut}(X) \times \operatorname{Aut}(Y)$ on $W(X, Y, L)$ as follows:

$$(a, c) \circ (f, g, h) = (fa, c^{-1}g, (\mathcal{T}a)^{-1}hc)$$

for any $(a, c) \in \operatorname{Aut}(X) \times \operatorname{Aut}(Y)$, $(f, g, h) \in W(X, Y, L)$. Denote the orbit space by

$$V(X, Y, L) = W(X, Y, L)/\operatorname{Aut}(X) \times \operatorname{Aut}(Y).$$

$V(X, Y, L)$ is a finite set and

$$F^L_{Y,X} = |V(X, Y, L)|$$

denotes the cardinality of $V(X, Y, L)$.

Let $\operatorname{ind}\mathcal{R}$ denote the set of representatives of isomorphism classes of all indecomposable objects in $\mathcal{R}$. For any indecomposable object $X \in \mathcal{R}$, let $d(X) = \dim_k(\operatorname{End} X/\operatorname{rad}\operatorname{End} X)$.

For any $M \in \mathcal{R}$, write h_M for the dimension vector $\underline{\dim}\, M$ of M. Then the Grothendieck group $G(\mathcal{R})$ is generated by $\{h_M \mid M \in \mathcal{R}\}$ and $h_L = h_M + h_N$ if there exists a triangle of form $M \to L \to N \to \mathcal{T}M$. Thus, $h_M = -h_{\mathcal{T}M}$. Consider the rational extension $\mathbf{Q} \bigotimes_{\mathbf{Z}} G(\mathcal{R})$. Denote by $\mathbf{h}$ the subgroup of $\mathbf{Q} \bigotimes_{\mathbf{Z}} G(\mathcal{R})$ generated by $\frac{h_M}{d(M)}$, $M \in \operatorname{ind}\mathcal{R}$. Note that the Grothendieck group $G(\mathcal{R})$ is a subgroup of $\mathbf{h}$.

Now define a symmetric bilinear function $(-|-)$ on $\mathbf{h} \times \mathbf{h}$ (and by extension also on $(\mathbf{Q} \bigotimes_{\mathbf{Z}} \mathbf{h}) \times (\mathbf{Q} \bigotimes_{\mathbf{Z}} \mathbf{h})$) by

$$(h_X|h_Y) = \dim_k \operatorname{Hom}(X, Y) - \dim_k \operatorname{Hom}(X, \mathcal{T}Y) + \dim_k \operatorname{Hom}(Y, X) - \dim_k \operatorname{Hom}(Y, \mathcal{T}X)$$

for any $X, Y \in \mathcal{R}$. On $G(\mathcal{R})$, this is just the symmetric Euler form of $\mathcal{R}$.

Let $\mathbf{n}$ be the free Abelian group with a basis $\{u_X | X \in \text{ind}\,\mathcal{R}\}$, which is bijective to the set of isomorphism classes of all indecomposable objects in $\mathcal{R}$. Let

$$\mathbf{g} = \mathbf{h} \oplus \mathbf{n}$$

be a direct sum of $\mathbf{Z}$-modules. Consider the quotient group

$$\mathbf{g}_{(q-1)} = \mathbf{g}/(q-1)\mathbf{g}.$$

Mildly abusing notation, denote by u_M, h_M also the residues in the quotient group corresponding to $M \in \mathcal{R}$.

THEOREM 7.1.2 [2]. $\mathbf{g}_{(q-1)}$ *is a Lie algebra over* $\mathbf{Z}/(q-1)$ *with its bilinear operation* $[-,-]$ *defined as follows:*

(1) *for any two indecomposable objects* $X, Y \in \mathcal{R}$,

$$[u_X, u_Y] = \begin{cases} \displaystyle\sum_{L \in \text{ind}\,\mathcal{R}} \left(F^L_{XY} - F^L_{YX}\right) & \text{if } Y \not\simeq \mathcal{T}X, \\ \dfrac{h_X}{d(X)} & \text{if } Y \simeq \mathcal{T}X; \end{cases}$$

(2) *for any objects* $X, Y \in \mathcal{R}$ *with* Y *indecomposable,*

$$[h_X, u_Y] = -(h_X|h_Y)u_Y \text{ and } [u_Y, h_X] = -[h_X, u_Y];$$

(3) $[h, h'] = 0$, $\forall\ h, h' \in \mathbf{h}$.

REMARK 7.1.3. Indeed, the positive part of the above Lie algebra can be obtained through the (untwisted) Ringel–Hall algebra.

Recall the definition of **untwisted Ringel–Hall algebra** $\overline{\mathcal{H}}(\Lambda)$. It is a free Abelian group with basis $\{u_{[M]} \mid M \in \text{mod}-\Lambda\}$ indexed by the isomorphism classes of all finite-dimensional Λ–modules. Write simply u_M for $u_{[M]}$. The multiplication is defined by

$$u_X \star u_Y = \sum_{[M], M \in \text{mod}-\Lambda} g^M_{XY} u_M.$$

This is an associative ring with unit u_0. Let ind Λ be the set of representatives of isomorphism classes of all indecomposable modules in mod $-\Lambda$, and let $K(\text{mod}-\Lambda)$ be the subgroup of $\overline{\mathcal{H}}(\Lambda)$ generated (freely) by $\{u_X \mid X \in \text{ind}\,\Lambda\}$. Then the proof of Proposition 5 in [39] implies that over the residue ring $\mathbf{Z}/(q-1)$, $K(\text{mod}-\Lambda)$ is a Lie subalgebra of $\overline{\mathcal{H}}(\Lambda)$, that is, the quotient group $K(\text{mod}-\Lambda)/(q-1)K(\text{mod}-\Lambda)$ is a Lie subalgebra of the quotient ring $\overline{\mathcal{H}}(\Lambda)/(q-1)\overline{\mathcal{H}}(\Lambda)$, where the Lie bracket is the commutator $[u_X, u_Y] = u_X \star u_Y - u_Y \star u_X$.

The root category $\mathcal{R}$ has two parts, mod $-\Lambda$ and $\mathcal{T}(\text{mod}-\Lambda)$. Hence, $\mathbf{n} = K(\text{mod}-\Lambda) \oplus K(\mathcal{T}(\text{mod}-\Lambda))$, where $K(\mathcal{T}(\text{mod}-\Lambda))$ is the free Abelian group with

the basis $\{u_{\mathcal{T}(X)} \mid X \in \operatorname{ind}\Lambda\}$ and

$$\mathbf{g} = K(\operatorname{mod}-\Lambda) \oplus \mathbf{h} \oplus K(\mathcal{T}(\operatorname{mod}-\Lambda)).$$

Hence, the Lie algebra $K(\operatorname{mod}-\Lambda)/(q-1)K(\operatorname{mod}-\Lambda)$ is a subalgebra of $\mathfrak{g}_{(q-1)}$ and

$$\begin{aligned}\mathbf{g}_{(q-1)} = {} & K(\operatorname{mod}-\Lambda)/(q-1)K(\operatorname{mod}-\Lambda) \oplus \mathbf{h}/(q-1)\mathbf{h} \\ & \oplus K(\mathcal{T}(\operatorname{mod}-\Lambda))/(q-1)K(\mathcal{T}(\operatorname{mod}-\Lambda))\end{aligned}$$

is a triangular decomposition of $\mathbf{g}_{(q-1)}$.

Let E be a field extension of k. Then, $\Lambda \otimes_k E$ is a hereditary E-algebra, and for $M \in \operatorname{mod}-\Lambda$, the module $M \otimes_k E$ has a canonical $\Lambda \otimes_k E$-module structure. Recall that E is called **conservative extension** for $X \in \operatorname{ind}\Lambda$ if $(\operatorname{End}_\Lambda / \operatorname{rad}\operatorname{End}_\Lambda X) \otimes_k E$ is a field again so that $X \otimes_k E$ is an indecomposable $\Lambda \otimes_k E$-module. An indecomposable Λ–module X is called **exceptional Λ-module** if $\operatorname{Ext}^1_\Lambda(X, X) = 0$. Note that all simple Λ-modules are exceptional. It follows from C.M. Ringel [44] that for an exceptional module, the endomorphism algebra is isomorphic to the endomorphism algebra of some simple Λ-module S.

Let $\bar{k}$ be the algebraic closure of k and set

$$\begin{aligned}\Omega = \{E \ : \ & k \subseteq E \subseteq \bar{k} \text{ is a finite field extension and conservative for} \\ & \text{all exceptional } \Lambda\text{-modules}\} \\ = \{E \ : \ & k \subseteq E \subseteq \bar{k} \text{ is a finite field extension and conservative for} \\ & \text{all exceptional objects in } \mathcal{R}\}.\end{aligned}$$

The set Ω is an infinite set.

Let $\mathcal{R}^E$ be the root category of $\Lambda \otimes_k E$. Then, as before, we have $\mathbf{h}^E$, $\mathbf{n}^E$, and $\mathbf{g}^E = \mathbf{h}^E \oplus \mathbf{n}^E$, and $\mathbf{g}^E_{(|E|-1)} = \mathbf{g}^E/(|E|-1)\mathbf{g}^E$ which is a Lie algebra over $\mathbf{Z}/(|E|-1)$. Consider the direct product

$$\prod_{E\in\Omega} \mathbf{g}^E_{(|E|-1)}$$

of Lie algebras and let $\mathcal{LE}(\mathcal{R})$ be the Lie subalgebra of $\prod_{E\in\Omega} \mathfrak{g}^E_{(|E|-1)}$ generated by the set $\{u_X = (u_{X\otimes_k E})_{E\in\Omega} \mid X \text{ is an exceptional object in } \mathcal{R}\}$.

THEOREM 7.1.4 [2]. *If Λ and $\mathfrak{g}'(C)$ share the same Cartan datum, then there is a Lie algebra isomorphism*

$$\Psi : \mathbf{C} \otimes_{\mathbf{Z}} \mathcal{LE}(\mathcal{R}) \longrightarrow_{\mathbf{C}} \mathfrak{g}'(C)$$

given by

$$\begin{aligned} -\frac{h_{S_i}}{d(S_i)} &\mapsto h_i, \quad 1 \le i \le n, \\ u_{S_i} &\mapsto e_i, \quad 1 \le i \le n, \\ u_{\mathcal{T}(S_i)} &\mapsto -f_i, \quad 1 \le i \le n, \end{aligned}$$

where $\{S_i | 1 \le i \le n\}$ is the set of all nonisomorphic simple Λ–modules.

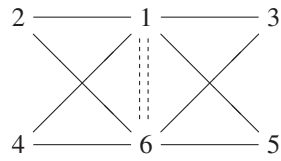
7.2. *A construction of the elliptic Lie algebra* $\mathbf{g}(X)$

For the real root system determined by a Cartan datum, one can construct a Lie algebra, namely, the Kac–Moody algebra, that has a nice decomposition of root spaces. The same question can be raised for a generalized root system, especially for an elliptic root system. There have been numerous attempts along these lines, leading to intersection matrix Lie algebras, vertex algebras, toroidal algebras, extended affine Lie algebras, and so on.

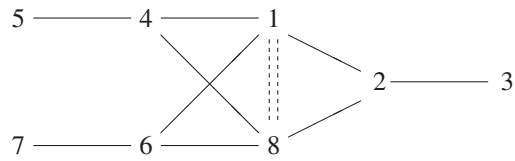
New advances have been made by K. Saito and D. Yoshii [45], who described simply laced elliptic Lie algebras from three different points of view. The first one is from the vertex algebra attached to an elliptic root lattice. In this structure, the root spaces are of an explicit form. The second one is a presentation given by Chevalley generators and generalized Serre relations attached to the elliptic Dynkin diagram. The third one is an amalgamation of an affine Kac–Moody algebra and a Heisenberg algebra. As mentioned in [45], simply laced elliptic Lie algebras are in fact isomorphic to the corresponding 2-toroidal algebras (2-extended affine Lie algebras). More recently, Y. Lin and L. Peng [3] constructed elliptic Lie algebras of type $D_4^{(1,1)}$, $E_6^{(1,1)}$, $E_7^{(1,1)}$, or $E_8^{(1,1)}$ via the root category of the tubular algebras of the same type.

In this subsection, we only deal with the following elliptic Dynkin diagrams:

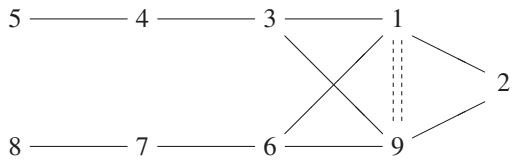
$D_4^{(1,1)}$:



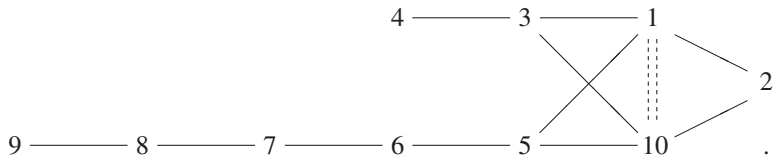
$E_6^{(1,1)}$:



$E_7^{(1,1)}$:



$E_8^{(1,1)}$:



.

Consider n-dimensional $\mathbf{Q}$-space U, where n is the number of vertices of an elliptic Dynkin diagram of type $D_4^{(1,1)}$, $E_6^{(1,1)}$, $E_7^{(1,1)}$, or $E_8^{(1,1)}$. Let $\Pi = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be the canonical basis of U. The symmetric bilinear form $(-,-) : U \times U \to \mathbf{Q}$ attached to an elliptic Dynkin diagram of type $D_4^{(1,1)}$, $E_6^{(1,1)}$, $E_7^{(1,1)}$, or $E_8^{(1,1)}$ is defined by

$$(\alpha_i, \alpha_j) = \begin{cases} 2 & \text{if } i = j; \\ -1 & \text{if there is a real edge between } i \text{ and } j; \\ 2 & \text{if there is a double dotted edge between } i \text{ and } j; \\ 0 & \text{otherwise.} \end{cases}$$

Then the matrix attached to the symmetric bilinear form $(-,-)$, called an **elliptic Cartan matrix**, is positive semidefinite with corank 2. Note that it is not a generalized Cartan matrix in the sense of Kac–Moody algebras [11] because it contains positive entries $a_{1n} = a_{n1} = 2$ in the off-diagonal part.

DEFINITION 7.2.1 (K. Saito and D. Yoshii). The **elliptic Lie algebra** of a given elliptic Cartan matrix is defined by Chevalley generators and generalized Serre relations as follows:

Generators: $\{\alpha_i \mid 1 \le i \le n\}$ and $\{e_{\pm i} \mid 1 \le i \le n\}$.
Relations:

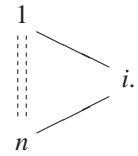
0.
$$[\alpha_i, \alpha_j] = 0, \ \ 1 \le i, j \le n;$$

I.
$$[e_i, e_{-i}] = \alpha_i, \ \ 1 \le i \le n;$$

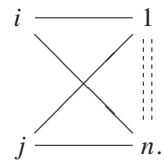
II.1. $[\alpha_i, e_j] = (\alpha_i, \alpha_j) e_j$, for $i, j = \pm 1, \pm 2, \ldots, \pm n$, where $\alpha_{-i} = -\alpha_i$;

II.2. $(\operatorname{ad} e_i)^{\max\{1, 1-(\alpha_i, \alpha_j)\}} e_j = 0$, for $i, j = \pm 1, \pm 2, \ldots, \pm n$;

III. $[[e_i, e_1], e_n] = 0$
$[[e_{-i}, e_{-1}], e_{-n}] = 0$
for



IV. $[[[e_1, e_n], e_i], e_j] = 0$
$[[[e_{-1}, e_{-n}], e_{-i}], e_{-j}] = 0$
$[[[e_{-1}, e_n], e_i], e_j] = 0$
$[[[e_1, e_{-n}], e_{-i}], e_{-j}] = 0$
for

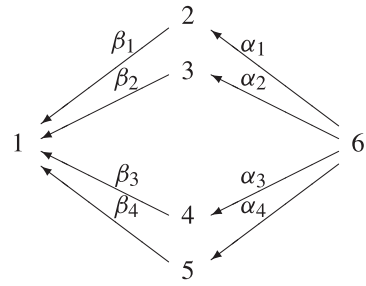


For a generalized Cartan matrix, the relations of type 0–II are defining relations for a Kac–Moody algebra. But, for an elliptic Cartan matrix, one has to consider additional relations as given by III–IV above.

Tubular algebras form a certain special class of associative algebras of global dimension two, which can be determined by quivers with relations. Their module categories and derived categories are of wide interest (see [6,7]), for example, W. Geigle and H. Lenzing [6] found that the derived category of a tubular algebra is isomorphic to the derived category of coherent sheaves on a weighted projective line. (An algebra of the form $B = \mathrm{End}_A(T_A)$, where A is connected, hereditary, representation–infinite, and T_A is a preprojective tilting module, is said to be a **concealed algebra**. In case A is, in addition, tame, then B is said to be a **tame concealed algebra**). By definition, a **tubular algebra** is a **tubular extension** of a tame concealed algebra of extension type $\mathbf{T}(r_1, r_2, \ldots, r_t)$, where $(r_1, r_2, \ldots, r_t)$ is equal to $(2, 2, 2, 2)$, $(3, 3, 3)$, $(4, 4, 2)$, or $(6, 3, 2)$. We refer to [7] for a very technical definition of tubular extension. For a tubular algebra Λ of type $\mathbf{T}(r_1, r_2, \ldots, r_t)$, its derived category has a nice structure and consists of infinitely many tubular families, where each tubular family is determined essentially by a tame concealed algebra and contains infinitely many homogenous tubes and t nonhomogenous tubes of rank $r_1, r_2, \ldots, r_t$, respectively.

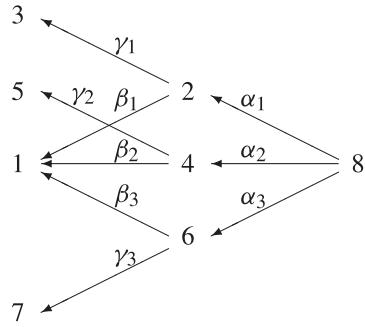
Y. Lin and L. Peng considered the following tubular algebras given by quivers with relations, that is, $\Lambda = k\mathcal{Q}/I$:

$\mathbb{T}(2,2,2,2)$:

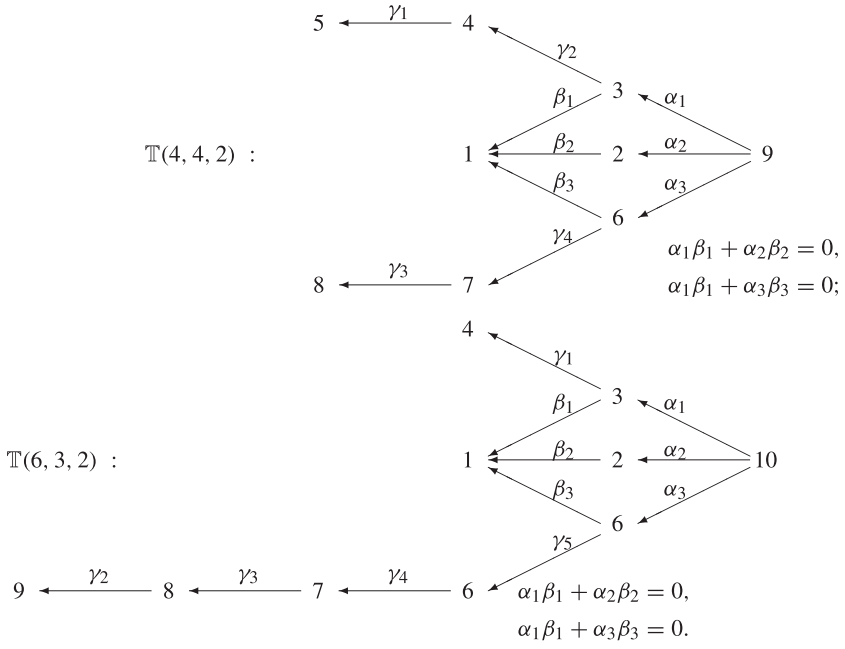


$$\alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3 = 0,$$
$$\alpha_1\beta_1 + \lambda\alpha_2\beta_2 + \alpha_4\beta_4 = 0,$$
$$\lambda \in k \setminus \{0, 1\};$$

$\mathbb{T}(3,3,3)$:



$$\alpha_1\beta_1 + \alpha_2\beta_2 = 0,$$
$$\alpha_1\beta_1 + \alpha_3\beta_3 = 0;$$



For each of the four quivers above, if we forget the orientation and add a dotted edge between the vertices 1 and n (here, $n = 6, 8, 9, 10$), then the diagram obtained coincides with an elliptic Dynkin diagram. We extend the usual terminology to such tubular algebras, calling them of type $D_4^{(1,1)}$, $E_6^{(1,1)}$, $E_7^{(1,1)}$, and $E_8^{(1,1)}$.

Using the explicit structure of the derived categories of tubular algebras given by D. Happel and C.M. Ringel [46], Y. Lin and L. Peng proved the following theorem:

THEOREM 7.2.2 [3]. *Let Λ be the tubular algebra of type $X_l^{(1,1)}$ with the isomorphism classes of simple modules S_i, $1 \le i \le n$, where $X_l = D_4^{(1,1)}$, $E_6^{(1,1)}$, $E_7^{(1,1)}$, or $E_8^{(1,1)}$. Let $\mathbf{g}$ be the elliptic Lie algebra of type $X_l^{(1,1)}$, with generators α_i and $e_{\pm i}$, $1 \le i \le n$, and $\mathcal{L}(\mathcal{R})$ the Ringel–Hall Lie algebra over $\mathbf{Q}$ of the root category $\mathcal{R}$ of Λ. Then, there is a Lie algebra isomorphism*

$$\Psi : \mathcal{L}(\mathcal{R}) \longrightarrow \mathbf{g}$$

given by

$$\begin{aligned} h_{S_i} &\mapsto \alpha_i, \qquad 1 \le i \le n, \\ u_{S_i} &\mapsto e_i, \qquad 1 \le i \le n, \\ u_{\mathcal{T}(S_i)} &\mapsto -e_{-i}, \quad 1 \le i \le n. \end{aligned}$$

8. Guide to the literature

A wealth of material concerning the representation theory of finite-dimensional algebras (or quivers) can be found in the books [7, 16, 47] and online notes [14]. For the unexplained notation concerning the theory of Lie algebra, we refer to the book [11], for those concerning quantum groups to the book [28], and for derived categories to the book [42] and the paper [43]. The material related to Ringel–Hall algebras is discussed in the articles [7, 18–21, 27, 39, 40]. For Section 4, we recommend [1, 13]. Finally, the articles [2] and [3] contain a detailed exposition of the topics discussed in Section 7.

Acknowledgements

X. Chen and T. Pospíchal would like to thank Professor Vlastimil Dlab for his help and support. X. Chen is grateful to Professor Jie Xiao for directing him to do research in the field of Ringel–Hall algebras. We thank Professor H. Nakajima for sending us his paper [48], which helped us to organize the material in Section 6.

References

[1] V. Dlab, C.M. Ringel, Indecomposable representations of graphs and algebras, Mem. Amer. Math. Soc. **6** (1976), vol. **173**, Providence, RI.

[2] L. Peng, J. Xiao, Triangulated categories and Kac–Moody algebras, Invent. Math. **140** (2000), 563–603.

[3] Y. Lin, L. Peng, Elliptic Lie algebras and tubular algebras, Adv. Math. **196** (2005), 487–530.

[4] G. Lusztig, Canonical bases arising from quantized enveloping algebras, J. Amer. Math. Soc. **3** (1990), 447–498.

[5] G. Lusztig, Quivers, perverse sheaves, and quantized enveloping algebras, J. Amer. Math. Soc. **4** (1991), 365–421.

[6] W. Geigle, H. Lenzing, A class of weighted projective curves arising in representation theory of finite dimensional algebras, in: G.-M. Greuel, G. Trautmann, eds., *Singularities, Representations of Algebras, and Vector Bundles* (*Lambrecht, 1985*), Lecture Notes in Math., vol. **1273**, 265–297, Springer–Verlag, Berlin, 1987.

[7] C.M. Ringel, Tame algebras and integral quadratic forms, Lecture Notes in Math., vol. **1099**, Springer–Verlag, Berlin–New York, 1984.

[8] V.I. Arnold, The *A-D-E* classifications, in: *Mathematical Developments Arising from Hilbert Problems*, Proc. Sympos. Pure Math., vol. **28**, American Mathematical Society, Providence, RI, 1976, pp. 46.

[9] M. Hazewinkel, W. Hesselink, D. Siersma, F. Veldkamp, The ubiquity of Coxeter-Dynkin diagrams (an introduction to the *A-D-E* problem), Nieuw Arch. Wisk. **25** (3) (1977), 257–307.

[10] J.A. Green, Hall algebras and quantum groups, Textos de Matemática. Série B [Texts in Mathematics. Series B], 4. Universidade de Coimbra, Departamento de Matemática, Coimbra, 1994, pp. vi+23.

[11] V. Kac, Infinite dimensional Lie algebras, 3rd ed., Cambridge University Press, Cambridge, UK, 1990.

[12] C.M. Ringel, Representation theory of finite-dimensional algebras, in: *Representations of Algebras* (*Durham, 1985*), 7–79, London Mathematical Society Lecture Note Series, vol. **116**, Edited by P. J. Webb. Cambridge University Press, Cambridge, UK, 1986.

[13] P. Gabriel, Unzerlegbare Darstellungen I, Manuscripta Math. **6** (1972), 71–103.
[14] W. Crawley-Boevey, Lectures on representations of quivers, lectures given in 1992 at Oxford Univ., available at http://www.amsta.leeds.ac.uk/∼pmtwc/quivlecs.pdf
[15] Yu.A. Drozd, Tame and wild matrix problems, (Russian) Representations and quadratic forms (Russian), Akad. Nauk Ukrain. SSR Inst. Mat. Kiev **154** (1979), 39–74.
[16] M. Auslander, I. Reiten, S. Smalø, *Representation Theory of Artin Algebras*, Cambridge Studies in Advanced Math., vol. **36**, Cambridge University Press, Cambridge, 1995.
[17] V. Kac, Infinite root systems, representations of graphs and invariant theory, Invent. Math. **56** (1980), 57–92.
[18] C.M. Ringel, Hall algebras and quantum groups, Invent. Math. **101** (1990), 583–591.
[19] C.M. Ringel, Hall algebras revisited, in: *Quantum Deformations of Algebras and their Representations* (Ramat-Gan, 1991/1992; Rehovot, 1991/1992), 171–176, Israel Math. Conf. Proc., **7**, Bar-Ilan Univ., Ramat Gan, 1993.
[20] J.A. Green, Hall algebras, hereditary algebras and quantum groups, Invent. Math. **120** (1995), 361–377.
[21] C.M. Ringel, Green's theorem on Hall algebras, in *Representation Theory of Algebras and Related Topics*; CMS Conf. Proc., vol. **19**, (1996), 185–245, Amer. Math. Soc., Providence, RI.
[22] J. Xiao, Drinfeld double and Ringel–Green theory of Hall algebras, J. Algebra **190** (1997), 100–144.
[23] J. Beck, V. Chari, A. Pressley, An algebraic characterization of the affine canonical Basis, Duke. Math. J. **99** (1999), 455–487.
[24] J. Beck, H. Nakajima, Crystal bases and two-sided cells of quantum affine algebras, Duke Math. J. **123** (2004), 335–402.
[25] M. Kashiwara, On crystal bases of the q-analogue of universal enveloping algebras, Duke Math. J. **63** (1991), 465–516.
[26] Z. Lin, J. Xiao, G. Zhang, Representations of tame quivers and affine canonical bases, preprint, arXiv:0706.1444.
[27] C.M. Ringel, PBW–bases of quantum groups, J. Reine Angew. Math. **470** (1996), 51–88.
[28] G. Lusztig, Introduction to quantum groups, *Progress in Math*., vol. **110**, Birkhäuser, Boston, 1993.
[29] V. Chari, A. Pressley, Quantum affine algebras at roots of unity, Representation Theory **1** (1997), 280–328.
[30] I. Damiani, A basis of type Poincaré–Birkhoff–Witt for the quantum algebra $\widehat{sl}(2)$, J. Algebra **161** (1993), 291–310.
[31] F. Gavarini, A PBW basis for Lusztig's form of untwisted affine quantum groups, Comm. Algebra **27** (1999), 903–918.
[32] I.N. Bernstein, I.M. Gelfand, V.A. Ponomarev, Coxeter functors and Gabriels theorem, Uspekhi Math. Nauk **28** (1973), 19–33.
[33] X. Chen, J. Xiao, Root vectors arising from Auslander-Reiten Quivers, J. Algebra **222** (1999), 328–356.
[34] X. Chen, J. Xiao, Exceptional sequence in Hall algebra and quantum group, Compositio Math. **117** (1999), 161–187.
[35] X. Chen, Root vectors of the composition algebra of the Kronecker algebra, Algebra Discrete Math. **1** (2004), 37–56.
[36] K. McGerty, The Kronecker quiver and bases of quantum affine sl_2, Adv. Math. **197** (2005), 411–429.
[37] P. Zhang, PBW–bases of the composition algebra of the Kronecker algebra, J. Reine Angew. Math. **527** (2000), 97–116.
[38] P. Baumann, C. Kassel, The Hall algebra of the category of coherent sheaves on the projective line, J. Reine Angew. Math. **533** (2001), 207–233.
[39] C.M. Ringel, Hall algebras, in *Topics in Algebra*, Part 1 (Warsaw, 1988), 433–447, Banach Center Publ., 26, Part 1, PWN, Warsaw, 1990.
[40] C.M. Ringel, Hall polynomials for the representation-finite hereditary algebras, Adv. Math. **84** (1990), 137–178.

[41] D. Happel, On the derived category of a finite-dimensional algebra, Comment. Math. Helv. **62** (1987), 339–389.

[42] D. Happel, Triangulated categories in the representation theory of finite-dimensional algebras, London Mathematical Society Lecture Note Series, vol. **119**, Cambridge University Press, Cambridge, UK, 1988.

[43] B. Keller, Derived categories and their uses, Handbook of algebra, vol. **1**, 671–701, North-Holland, Amsterdam, 1996.

[44] C.M. Ringel, The braid group action on the set of exceptional sequences of a hereditary Artin algebra, in: *Abelian Group Theory and Related Topics* (Oberwolfach, 1993), 339–352, Contemp. Math., **171**, Amer. Math. Soc., Providence, RI, 1994.

[45] K. Saito, D. Yoshii, Extended affine root systems IV (Simple laced elliptic Lie algebras), Publ. RIMS. Kyoto Univ. **36** (2000), 385–421.

[46] D. Happel, C.M. Ringel, The derived category of a tubular algebra, Lecture Notes in Math., vol. **1273**, 156–180, Springer–Verlag, Berlin–Heidelberg–New York, 1986.

[47] I. Assem, D. Simson, A. Skowroński, *Elements of the representation theory of associative algebras*, vol. **1**: *Techniques of representation theory*. London Mathematical Society Student Texts, vol. **65**. Cambridge University Press, Cambridge, 2006.

[48] H. Nakajima, Crystal, Canonical and PBW bases of quantum affine algebras, Algebraic groups and homogeneous spaces, 389–421, Tata Inst. Fund. Res. Stud. Math., Tata Inst. Fund. Res., Mumbai, 2007.

Canonical Decompositions and Invariants for Data Analysis

Marlos Viana

University of Illinois at Chicago Eye Center,
Chicago, IL 60612, USA
E-mail: viana@uic.edu

Contents

HANDBOOK OF ALGEBRA, VOL. 6
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(08)00210-6

1. Introduction

Riley et al. [3] observed that

> If a physical system is such that after application of a particular symmetry transformation the final system is indistinguishable from the original system, then its behavior, and hence the functions that describe its behavior, must have the corresponding property of invariance when subject to the same transformations.

The study of these transformations is a study of the symmetries of the system. More generally, as Bacry [1] shows, the study of the symmetries of a physical system often suggests the study of the symmetries of certain physical laws and theories, and frequently, leads to symmetry-related principles in the form of conservation laws. In this chapter, a data analytic interpretation of these principles is proposed.

To illustrate, consider the group $G = \{1, v, h, o\}$ of vertical (v) and horizontal (h) line reflections, point reflection (o), and the identity (1) under composition of functions. Its multiplication table

$*$	1	v	h	o
1	1	v	h	o
v	v	1	o	h
h	h	o	1	v
o	o	h	v	1

(1.1)

shows that G is Abelian. Of course, these are the automorphisms of a rectangle. Examples of data indexed by these transformations appear, for example, in [5] and [6, Chapter 1]. Indicating by $\{x_1, x_v, x_h, x_o\}$ the data indexed by these transformations, it then follows that the four scalar quantities,

$$\begin{aligned} &\mathcal{I}_1 = x_1 + x_o + x_v + x_h, \quad \mathcal{I}_v = x_1 + x_v - x_o - x_h, \\ &\mathcal{I}_h = x_1 + x_h - x_o - x_v, \quad \text{and} \quad \mathcal{I}_o = x_1 + x_o - x_v - x_h, \end{aligned} \tag{1.2}$$

are the invariants of interest in the data space under the regular group action. These data summaries vary only up to companion points in the same invariant subspace. For example, the invariant $x_1 + x_o - x_v - x_h$ is such that

$$x_{t*1} + x_{t*o} - x_{t*v} - x_{t*h} = \pm(x_1 + x_o - x_v - x_h) \quad \text{for all} \quad t \in G.$$

Ideally, then, it is of interest to interpret and describe both the underlying group actions and the resulting invariants.

2. A general principle of interpretability

As a general principle, group actions can be interpreted as degrees of arbitrariness, or irrelevance, that are chosen upon the labels and need to be resolved. For example,

it can be shown that the sum and $n-1$ sets of pairwise comparisons in a sample of size n are the invariants that result when the assignment of observations to the labels $\{1, 2, \ldots, n\}$ is (determined to be) arbitrary in such a way that its degree of arbitrariness amounts to shuffling the labels in all possible $n!$ ways. These invariants, in dimensions of 1 and $n-1$, respectively, appear in the practice of data analysis as the sample mean and its standard deviation. If, however, the data are labeled by n equally spaced points along a circle and the degree of arbitrariness is now restricted to shuffling the labels according to the n cyclic permutations, then the resulting invariants are the n linear combinations of the original data, with as coefficients the n sets of Fourier coefficients $\cos(2\pi jk/n) + i\sin(2\pi jk/n)$, $j, k = 0, \ldots, n-1$, that is, the n finite Fourier transforms of the observed data.

Applying this principle to the example introduced in Section 1, the invariants (1.2) can be characterized as those data that should be retained when the arbitrariness of where is left (right) and where is up (down) is resolved. This degree of arbitrariness can be identified by studying the group action, defined by the multiplication table in (1.1), each line of which relabels what is up and down and what is left–right. It can be referred to as *planar orientation* so that then the pairwise comparisons in (1.2) are the invariants of planar orientation. For example, $x_1 + x_o - x_v - x_h$ gives a one-dimensional subspace regardless of the planar orientation. Equivalently, it compares the responses to point symmetry against the combined responses of vertical and horizontal responses, regardless of the planar orientation. In the present case, all four invariants are in dimension of one, adding to the dimension (4) of the original vector space for the data indexed by the elements of G.

A simpler example is given by the invariants $x_\ell \pm x_r$ associated with bilateral, say, left (ℓ) and right (r) measurements, determined by the identity transformation and the transposition of left and right measurements. Here, left–right is the arbitrariness resolved by $x_\ell \pm x_r$.

When an invariant is derived from a regular group action (the group acting on itself by multiplication on the right or on the left), it is called a *regular invariant*. In the following sections, the data analytic methods of symmetry studies will be applied to determine and interpret the regular invariants associated with experimental data indexed by the permutations of three and four objects.

3. Symmetry studies

Methods of representation theory can be systematically applied in the analysis of data indexed by a set of labels where a group action can be defined and often justified by an experimental context. While group representations provide the connection between the labels and the data space, the canonical decomposition theorem, reducing the identity matrix in the data space to a sum of algebraically orthogonal projections, provides the conditions with which the statistical theory of quadratic forms can be formulated to study the resulting decomposition of the sum of squares or analysis of variance. The study of these projections also leads to useful summaries of the data in

the form of canonical invariants, which can be used as descriptive tools or interpreted for further statistical inference.

These basic elements constitute the core of all symmetry studies as introduced by the author in [6], where a detailed introduction to a number of data analytical applications is described. Specifically, every symmetry study includes the identification of the following:

(1) A set V of labels with v elements;
(2) The observed data x defined in a linear subspace $\mathcal{V} \subset \mathbb{R}^v$ and indexed by those labels (the structured data);
(3) A rule or group action, according to which the symmetry transformations in a group G are applied to V;
(4) The classes and multiplicities of the resulting elementary orbits, subsets of V, on which G acts transitively;
(5) The resulting linear representations of these actions in the corresponding data vector subspaces;
(6) The canonical projection matrices $\mathcal{P}_\chi$ indexed by the irreducible characters χ of G;
(7) The canonical invariants $\mathcal{P}_\chi x$ on the data x, and their interpretations;
(8) A statistical analysis of the canonical invariants and, if applicable, their analysis of variance $x'x = \sum_\chi x'\mathcal{P}_\chi x$ based on the canonical decomposition $I = \sum_\chi \mathcal{P}_\chi$ of the identity operator in the data subspaces, as a sum of algebraically orthogonal projections.

When necessary, these canonical projections may be grouped together to describe $\mathcal{V}$. Specifically, if there are q_λ orbits of type λ, each with o_λ elements, classifying the $v = \sum_\lambda q_\lambda o_\lambda$ points in $\mathcal{V}$, then for each class of orbits with $v_\lambda = q_\lambda o_\lambda$ points, the identity matrix in $\mathcal{V}$ reduces according to $I_{v_\lambda} = \mathbb{P}_1^\lambda + \mathbb{P}_2^\lambda + \ldots$, where $\mathbb{P}_j^\lambda$ is the canonical projection constructed with q_λ copies of the canonical projection $\mathcal{P}_j^\lambda$ in the vector space associated with the elementary orbits of type λ and indexed by the irreducible representations $j = 1, 2, \ldots$ of the underlying group G. The reduction in the original space $\mathcal{V}$ is then $I_v = \mathrm{Diag}(\ldots, I_{v_\lambda}, \ldots)$, with as many components as the different types of orbits identified by the action of G on V.

To illustrate, consider the set $V = \{uu, yy, uy, yu\}$ of binary sequences in length of two or, equivalently, the set of all mappings s from $\{1, 2\}$ into $\{u, y\}$ so that $\varphi_1(\tau, s) = s\tau^{-1}$,

$$\varphi_1 : \{1, 2\} \xrightarrow{\tau^{-1}} \{1, 2\} \xrightarrow{s} \{u, y\}, s \in V,\ \tau \in S_2 = \{1, (12)\}, \tag{3.1}$$

and $\varphi_2(\sigma, s) = \sigma s$,

$$\varphi_2 : \{1, 2\} \xrightarrow{s} \{u, y\} \xrightarrow{\sigma} \{u, y\},\ s \in V,\ \sigma \in S_2 = \{1, (uy)\}, \tag{3.2}$$

are actions of S_2 on V. The action φ_1 classifies the sequences up to permutations of the positions $\{1, 2\}$ of the letters, whereas φ_2 classifies the sequences up to permutations of the symbols $\{u, y\}$. The action φ_1 generates three orbits $\{uu\}$, $\{yy\}$, and $\{uy, yu\}$ and φ_2 generates two orbits $\{uu, yy\}$ and $\{uy, yu\}$.

In the position symmetry case, the representation is isomorphic to

$$\xi_1 = \mathrm{Diag}\left(1, 1, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right),\ \xi_t = \mathrm{Diag}\left(1, 1, \mathrm{Diag}\left(1, 1, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right)\right),\ t \equiv (12),$$

taking values in $GL(\mathbb{R}^4)$. Equivalently, writing $\mathcal{V} = \mathbb{R}^4$, it is observed that the ξ determines the decomposition

$$\mathcal{V} = \mathcal{W}_1 \oplus \mathcal{W}_2 \oplus \mathcal{W}_3,$$

where the stable subspaces $\mathcal{W}_1$ and $\mathcal{W}_2$ reduce isomorphically as a subspace $\mathcal{U}_1$ associated with the unit representation, and $\mathcal{W}_3$ reduces as the sum of an isomorphic copy of $\mathcal{U}_1$ and a subspace $\mathcal{U}_{\mathrm{Sgn}}$ associated with the sign (Sgn) representation. Therefore,

$$\mathcal{V} = \underbrace{\mathcal{U}_1 \oplus \mathcal{U}_1 \oplus \mathcal{U}_1}_{\mathcal{V}_1} \oplus \mathcal{U}_{\mathrm{Sgn}} = \mathcal{V}_1 \oplus \mathcal{V}_{\mathrm{Sgn}}, \tag{3.3}$$

showing a decomposition of $\mathcal{V}$ into a direct sum of the irreducible representations of S_2 in which the isomorphic copies are collected together. This is the *canonical decomposition* of $\mathcal{V}$. Next, to construct the projections $\mathcal{P}_1$ and $\mathcal{P}_{\mathrm{Sgn}}$ of $\mathcal{V}$ on the irreducible subspaces $\mathcal{V}_1$ and $\mathcal{V}_{\mathrm{Sgn}}$, define

$$\mathcal{P}_\beta = n_\beta \sum_{\tau \in G} \overline{\chi}_\beta(\tau) \xi_\tau / g,$$

where χ_β is the irreducible character of the irreducible representation β of G with $|G| = g$ elements and n_β its dimension. In the present case ($G = S_2$), $\beta \in \{1, \mathrm{Sgn}\}$ with corresponding characters $\chi_1 \equiv 1$ and $\chi_{\mathrm{Sgn}} = \mathrm{Sgn}_\tau$ so that $\mathcal{P}_1 = (\xi_1 + \xi_t)/2$ and $\mathcal{P}_{\mathrm{Sgn}} = (\xi_1 - \xi_t)/2$. When the projections are evaluated relative to a basis of $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_{\mathrm{Sgn}}$ on which

$$\xi_\tau = \mathrm{Diag}(I_3 \otimes 1_\tau, \mathrm{Sgn}_\tau),$$

where I_n indicates the $n \times n$ identity matrix, the results are $\mathrm{Diag}(1, 1, 1, 0)$ and $\mathcal{P}_{\mathrm{Sgn}} = \mathrm{Diag}(0, 0, 0, 1)$.

It then follows that (1) $\mathcal{P}_1^2 = \mathcal{P}_1$, $\mathcal{P}_{\mathrm{Sgn}}^2 = \mathcal{P}_{\mathrm{Sgn}}$; (2) $\mathcal{P}_1 \mathcal{P}_{\mathrm{Sgn}} = \mathcal{P}_{\mathrm{Sgn}} \mathcal{P}_1 = 0$, and (3) $I_4 = \mathcal{P}_1 + \mathcal{P}_{\mathrm{Sgn}}$ so that $\mathcal{P}_\beta$ is a projection on a subspace isomorphic to $\mathcal{V}_\beta$. These properties remain valid if $\mathcal{P}_\beta$ is evaluated relative to any representation $M \xi_\tau M^{-1}$ equivalent to ξ. In this case, $\mathcal{P}_\beta$ transforms as $M \mathcal{P}_\beta M^{-1}$ and (1), (2), and (3) remain unchanged. For example, relative to the basis for $\mathcal{V}$ indexed by $\{uu, yy, uy, yu\}$,

$$\mathcal{P}_1 = \mathrm{Diag}(1, 1, \mathcal{A}_2); \quad \mathcal{P}_{\mathrm{Sgn}} = \mathrm{Diag}(0, 0, \mathcal{Q}_2),$$

where $\mathcal{A}_n$ indicates the $n \times n$ matrix with all entries equal to $1/n$ and $\mathcal{Q}_n = I_n - \mathcal{A}_n$.

Similarly, in the letter-symmetry case, ξ is given by

$$\xi_1 = \mathrm{Diag}(I_2, I_2), \quad \xi_t = \mathrm{Diag}\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right),$$

so that now $\mathcal{V} = \mathcal{W}_1 \oplus \mathcal{W}_2$ and in each one of these two stable subspaces ξ reduces isomorphically as the sum of the unit and the sign representations. Collecting the

isomorphic copies of the corresponding irreducible subspaces $\mathcal{U}_1$ and $\mathcal{U}_{\mathrm{Sgn}}$, it follows that

$$\mathcal{V} = \underbrace{\mathcal{U}_1 \oplus \mathcal{U}_1}_{\mathcal{V}_1} \oplus \underbrace{\mathcal{U}_{\mathrm{Sgn}} \oplus \mathcal{U}_{\mathrm{Sgn}}}_{\mathcal{V}_{\mathrm{Sgn}}},$$

which is the canonical decomposition of $\mathcal{V}$, with corresponding projections $\mathcal{P}_1 = (\xi_1 + \xi_t)/2$ and $\mathcal{P}_{\mathrm{Sgn}} = (\xi_1 - \xi_t)/2$. When these matrices are evaluated relative to a basis of $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_{\mathrm{Sgn}}$ on which

$$\xi_\tau = \mathrm{Diag}(I_2 \otimes 1_\tau, I_2 \otimes \mathrm{Sgn}_\tau),$$

the result is $\mathcal{P}_1 = \mathrm{Diag}(1, 1, 0, 0)$ and $\mathcal{P}_{\mathrm{Sgn}} = \mathrm{Diag}(0, 0, 1, 1)$.

It then follows that $\mathcal{P}_\beta$ is a projection on a subspace isomorphic to $\mathcal{V}_\beta$ and, regardless of the chosen basis for $\mathcal{V}$, $\mathcal{P}_1^2 = \mathcal{P}_1$, $\mathcal{P}_{\mathrm{Sgn}}^2 = \mathcal{P}_{\mathrm{Sgn}}$; $\mathcal{P}_1\mathcal{P}_{\mathrm{Sgn}} = \mathcal{P}_{\mathrm{Sgn}}\mathcal{P}_1 = 0$ and $I_4 = \mathcal{P}_1 + \mathcal{P}_{\mathrm{Sgn}}$. In particular,

$$\mathcal{P}_1 = \mathrm{Diag}(\mathcal{A}_2, \mathcal{A}_2), \quad \mathcal{P}_{\mathrm{Sgn}} = \mathrm{Diag}(\mathcal{Q}_2, \mathcal{Q}_2).$$

The arguments illustrated in the above example lead to a proof of following result [4, 6]:

THEOREM 3.1 (Canonical decomposition). *Let ρ be a linear representation of G into $GL(\mathcal{V})$, and let $\rho_1, \ldots, \rho_h$ be the distinct nonisomorphic irreducible representations of G with g elements, with corresponding characters $\chi_1, \ldots, \chi_h$ and dimensions $n_1, \ldots, n_h$. Then,*

$$\mathcal{P}_i = n_i \sum_{\tau \in G} \overline{\chi}_i(\tau)\rho_\tau / g$$

is a projection of $\mathcal{V}$ onto a subspace $\mathcal{V}_i$, sum of m_i isomorphic copies of the irreducible subspaces associated with ρ_i, $i = 1, \ldots, h$. Moreover, $\mathcal{P}_i\mathcal{P}_j = 0$, for $i \neq j$, $\mathcal{P}_i^2 = \mathcal{P}_i$ and $\sum_i \mathcal{P}_i = I_v$, where $v = \dim \mathcal{V} = \sum_{i=1}^h m_i n_i$.

The Fourier inverse formula In several experiments, however, the data are naturally indexed by G in itself so that $V = G$. In this case, there is a one-to-one correspondence

$$\widehat{x}(\beta) = \sum_{\tau \in G} x_\tau \beta_\tau \quad \longleftrightarrow \quad x_\tau = \sum_{\beta \in \widehat{G}} n_\beta \mathrm{tr}\,[\beta_{\tau^{-1}} \widehat{x}(\beta)]$$

between the experimental data $\{x_\tau, \tau \in G\}$ and the Fourier transforms $\widehat{x}(\beta)$ over the irreducible representations β of G, in dimension of n_β, where in the above expression, $\widehat{G}$ indicates the set of all irreducible representations of G. Observe, in particular, that the (regular) invariants in (1.2) are the Fourier transforms of the data indexed by G. More generally, it is given as follows.

PROPOSITION 3.1. *The rows of the Fourier transform $\widehat{x}(\beta)$ matrix reduce as β under the regular action.*

PROOF. Assume, without loss of generality, that β is in dimension of 2; select the first row of $\widehat{x}(\beta)$ and write β_{ij}^{τ} to indicate the ij entry of β_{τ}. The objective here is to show that G acting on the components

$$u = \sum_{\tau \in G} \tau \beta_{11}^{\tau}, \quad v = \sum_{\tau \in G} \tau \beta_{12}^{\tau}$$

of the first row of $\widehat{x}(\beta)$ as points in the group algebra of G according to the regular action $(u\sigma^{-1}, v\sigma^{-1})$ gives β_{σ} so that the representation coincides with (one of the two copies of) β, that is,

$$u\sigma^{-1} = u\beta_{11}^{\sigma} + v\beta_{21}^{\sigma}, \quad v\sigma^{-1} = u\beta_{12}^{\sigma} + v\beta_{22}^{\sigma}.$$

Indeed,

$$u\beta_{11}^{\sigma} + v\beta_{21}^{\sigma} = \sum_{\tau} \tau \left[\beta_{11}^{\tau}\beta_{11}^{\sigma} + \beta_{12}^{\tau}\beta_{21}^{\sigma}\right] = \sum_{\tau} \tau \beta_{11}^{\tau\sigma} = \sum_{\gamma} (\gamma \beta_{11}^{\gamma})\sigma^{-1} = u\sigma^{-1}$$

and similarly $u\beta_{12}^{\sigma} + v\beta_{22}^{\sigma} = v\sigma^{-1}$, proving the proposed equality and concluding the proof. □

Replacing τ by x_{τ} (the scalar data indexed by G) so that the rows of the Fourier transform $\widehat{x}(\beta)$ should give a basis for an invariant subspace tagged by β. If $\beta \in \widehat{G}$ is in dimension of m, then there are m such bases, corresponding to the m copies with which β appears in the regular representation of G. Repeating the same argument for the remaining rows of $\widehat{x}(\beta)$, it is seen that, in matrix form, Proposition 3.1 can be expressed as

$$\widehat{x}(\beta)\beta_{\sigma^{-1}} = \sum_{\tau} x_{\tau\sigma}\beta_{\tau}. \tag{3.4}$$

In particular, the rows (or columns if the regular action is from the right) form a set of bases for two isomorphic copies of β. In the next two sections, these results will be applied to the study of the regular invariants of S_3 and S_4.

4. The regular invariants of S_3

Consider the observed frequencies, shown in Table (4.1), with which the DNA words in the permutation orbit

$$\mathcal{O}_{\text{act}} = \{\text{act, tac, cta, cat, tca, atc}\} = \{s\tau^{-1}; \tau \in S_3\}$$

of the mapping $s = \text{act}$ appear in nine subsequent 900-bp-long regions along the BRU isolate of the human immunodeficiency virus type I. The entire 9229-bp-long single-strain DNA sequence is available from the National Center for Biotechnology Information[1] data base using the accession number K02013.

[1] http://www.ncbi.nlm.nih.gov/.

It is assumed in this section that the elements of S_3 are written in the order of $\{1, (123), (132), (12), (13), (23)\}$ so that $\mathcal{O}_{\text{act}}$ is obtained by evaluating $s\tau^{-1}$ in that order, where $s(1) = a$, $s(2) = c$, and $s(3) = t$. In what follows, moreover, the components of vectors such as

$$x' = (\text{act}, \text{tac}, \text{cta}, \text{cat}, \text{tca}, \text{atc})$$

indicate in short notation the data and at the same time the corresponding permutations in S_3. Other common examples occur in ranking preferences studies (e.g. [2]) or scalar data in psychometric studies, where scalar responses to the different sequences of presentation of a stimuli are recorded.

	1	2	3	4	5	6	7	8	9
act	8	16	16	7	17	11	12	6	14
tac	7	17	13	15	9	11	18	5	17
cta	15	8	14	9	14	15	8	5	16
cat	14	15	16	14	21	17	15	10	8
tca	11	18	10	17	11	16	14	9	13
atc	7	15	9	13	11	11	11	12	10
total	62	89	78	75	83	81	78	47	78

(4.1)

Having identified the labels and the structured data, it is remarked here again that their notation will often be identified so that, for example, act $\equiv x_{\text{act}}$ and act + tac is to be interpreted as $x_{\text{act}} + x_{\text{tac}}$. In particular, the case in which the permutations in S_3 act on the labels (which are in one-to-one correspondence with S_3) will be studied. The regular action will be defined by its multiplication table

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	3	1	6	4	5
3	3	1	2	5	6	4
4	4	5	6	1	2	3
5	5	6	4	3	1	2
6	6	4	5	2	3	1

, (4.2)

where the permutations 1, (123), (132), (12), (13), and (23) are identified with 1, 2, 3, 4, 5, and 6, respectively. The resulting representation is the regular representation of S_3, indicated here by ϕ. For example, when $\tau = (12)$, ϕ_τ is the permutation matrix defined by rows 1 and 4 of Table (4.2). Because this action is transitive, Step (4) in the summary presented in Section 3 leads to a simple single-orbit symmetry study.

4.1. *The canonical projections*

Indicate the irreducible representations of S_3 by 1, the symmetric or trivial representation; by Sgn, the alternating (or sign) representation (both in dimension of 1); and

by β in dimension of two. Their table of characters is given by

τ	χ_1	χ_{Sgn}	χ_β
1	1	1	2
(123)	1	1	−1
(132)	1	1	−1
(12)	1	−1	0
(13)	1	−1	0
(23)	1	−1	0

,

from which the evaluation of the canonical projections

$$\mathcal{P}_\chi = n_\chi \sum_{\tau \in G} \overline{\chi}(\tau)\phi_\tau/6$$

gives $\mathcal{P}_1$ the 6×6 matrix with entries equal to $1/6$,

$$\mathcal{P}_{\text{Sgn}} = 1/6 \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 \end{bmatrix},$$

$$\mathcal{P}_\beta = 1/3 \begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & -1 & -1 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & -1 & 2 \end{bmatrix}.$$

4.2. *Interpreting the regular invariants*

Borrowing language from physics [1], the permutations $\{1, (123), (132)\}$ will be referred to as rotations and $\{(12), (13), (23)\}$ as reversals. The terminology is a consequence of the action of S_3 on the ordered edges of a regular triangle, which identifies the three-fold rotations seen from each side of the triangle's plane. Applying the interpretability principles introduced in Section 2, it is observed that

(1) $\mathcal{P}_1 x$ gives the symmetric invariant act + tac + cta + cat + tca + atc;
(2) $\mathcal{P}_{\text{Sgn}} x$ gives the alternating invariant act + tac + cta − cat − tca − atc, comparing the total effect of rotations with the total effect of reversals, and
(3) $\mathcal{P}_\beta x$ evaluates as an invariant of within-rotation variability and within-reversal variability, giving an invariant subspace in dimension of four.

The degree of arbitrariness induced by the regular action can be interpreted as the arbitrariness intrinsic to the notion of rotations and reversals, which in a sense is equivalent to a notion of planar orientation. To see this, observe that although there are $3! = 6$ distinct assignments of the symbols $\{\mathrm{a}, \mathrm{c}, \mathrm{t}\}$ to the labeled (as $1, 2, 3$, say) vertices of a regular triangle, there are only two distinct assignments to those vertices when their labels are removed (by shuffling): namely to the two distinct classes

$$\{(\mathrm{a}, \mathrm{c}), (\mathrm{c}, \mathrm{t}), (\mathrm{t}, \mathrm{a})\} \text{ and } \{(\mathrm{c}, \mathrm{a}), (\mathrm{a}, \mathrm{t}), (\mathrm{t}, \mathrm{c})\}$$

of oriented edges. These two classes are the rotations and the reversals. It is simple to verify that the action of the permutations act, tac, cta, cat, tca, and atc on these edges replicates precisely the regular action of S_3. With this language, it can be said that (1)–(3) above are *the canonical invariants of rotations and reversals*. It is remarked however that, more generally, (1)–(3) are the regular invariants of S_3.

4.3. *A set of bases for the regular invariants*

It is simple to verify that

$$\beta : \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}, \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix},$$

where the matrices follow the order of S_3 described above, gives an irreducible representation of S_3. The resulting Fourier transform $\widehat{x}(\beta) = \sum_\tau x_\tau \beta_\tau$ of the structured data x at β is then

$$\widehat{x}(\beta) = \begin{bmatrix} \mathrm{act} - \mathrm{cta} - \mathrm{tca} + \mathrm{atc} & \mathrm{tac} - \mathrm{cta} + \mathrm{cat} - \mathrm{tca} \\ -\mathrm{tac} + \mathrm{cta} + \mathrm{cat} - \mathrm{atc} & \mathrm{act} - \mathrm{tac} + \mathrm{tca} - \mathrm{atc} \end{bmatrix} \simeq \beta \oplus \beta.$$

By Proposition 3.1, it identifies the two bases

$$\mathcal{I}_\beta = \{\mathrm{act} - \mathrm{cta} - \mathrm{tca} + \mathrm{atc}, \quad \mathrm{tac} - \mathrm{cta} + \mathrm{cat} - \mathrm{tca}\}$$

and

$$\mathcal{I}'_\beta = \{-\mathrm{tac} + \mathrm{cta} + \mathrm{cat} - \mathrm{atc}, \quad \mathrm{act} - \mathrm{tac} + \mathrm{tca} - \mathrm{atc}\}$$

from rows 1 and 2 of $\widehat{x}(\beta)$, respectively. The two (isomorphic) subspaces spanned by these two bases reduce as β under the regular action of S_3, that is, (3.4) is obtained.

The alternating invariant Clearly, a trivial application of Proposition 3.1 with the alternating character gives the invariant $\mathcal{I}_{\mathrm{Sgn}} = \mathrm{act} + \mathrm{tac} + \mathrm{cta} - \mathrm{cat} - \mathrm{tca} - \mathrm{atc}$ comparing rotations and reversals.

The vectors $\{\mathcal{I}_1, \mathcal{I}_{\mathrm{Sgn}}, \mathcal{I}_\beta, \mathcal{I}'_\beta\}$ define the (rows of the) change of basis matrix

$$B = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 0 & 0 & -1 & 1 \\ 0 & 1 & -1 & 1 & -1 & 0 \\ 0 & -1 & 1 & 1 & 0 & -1 \\ 1 & -1 & 0 & 0 & 1 & -1 \end{bmatrix},$$

relative to which

$$B\mathcal{P}_1 B^{-1} = \mathrm{Diag}(1,0,0,0,0,0),\ B\mathcal{P}_{\mathrm{Sgn}} B^{-1} = \mathrm{Diag}(0,1,0,0,0,0),\ \text{and}$$
$$B\mathcal{P}_\beta B^{-1} = \mathrm{Diag}(0,0,1,1,1,1).$$

Similarly, the regular action of S_3 extended to the components of $\{\mathcal{I}_1, \mathcal{I}_{\mathrm{Sgn}}, \mathcal{I}_\beta, \mathcal{I}'_\beta\}$ gives the representation

$$\tilde{\phi}_\tau = \mathrm{Diag}(1, \mathrm{Sgn}_\tau, \mathcal{I}_2 \otimes \beta_\tau),$$

relative to which the canonical projections are, respectively, $B\mathcal{P}_1 B^{-1}$, $B\mathcal{P}_{\mathrm{Sgn}} B^{-1}$, and $B\mathcal{P}_\beta B^{-1}$.

Interpreting voting preferences In the context of voting preferences, in which the data are the frequencies with which voters select among all permutations of the rankings of three candidates or objects, it is observed that

$$\mathcal{I}_\beta = \{\underline{\mathrm{a}}\mathrm{ct} - \mathrm{ct}\underline{\mathrm{a}} - \mathrm{tc}\underline{\mathrm{a}} + \underline{\mathrm{a}}\mathrm{tc},\ \ \mathrm{t}\underline{\mathrm{a}}\mathrm{c} - \mathrm{ct}\underline{\mathrm{a}} + \mathrm{c}\underline{\mathrm{a}}\mathrm{t} - \mathrm{tc}\underline{\mathrm{a}}\} \equiv \{v_1, v_2\}$$

can be characterized (as highlighted by $\underline{\mathrm{a}}$) as a basis for the space jointly describing "more radical" (v_1, first and third rank) and "less radical" (v_2, second and third rank) comparisons. The same interpretation, not surprisingly, applies to $\mathcal{I}'_\beta$, with the symbols *c* and *a* interchanged. From Proposition 3.1, these two position-comparisons depend on the symbols only up to linear superpositions of the components of $\mathcal{I}_\beta$. For example, the difference of these two components compares first and second ranks.

Interpreting planar rotations and reversals Following from Section 4.3, the regular invariants $\mathcal{I}_{\mathrm{Sgn}}$ and $\mathcal{I}_\beta$ (or $\mathcal{I}'_\beta$) can be interpreted as

(1) the comparison between rotations and reversals, in dimension 1;
(2) the two-component within-rotation, within-reversal variability, in dimension 2 (each).

More specifically, indicating by $\hookrightarrow$ and $\hookleftarrow$ the two opposing (chiral) orientations of planar rotation, it is observed that

$$v_1 = \underbrace{\underline{\mathrm{a}}\mathrm{ct} - \mathrm{cta}}_{\hookleftarrow \text{two-fold}} + \underbrace{\underline{\mathrm{a}}\mathrm{tc} - \mathrm{tca}}_{\hookrightarrow \text{one-fold}},\ \ \text{and}\ \ v_2 = \underbrace{\underline{\mathrm{t}}\mathrm{ac} - \mathrm{cta}}_{\hookleftarrow \text{one-fold}} + \underbrace{\underline{\mathrm{c}}\mathrm{at} - \mathrm{tca}}_{\hookrightarrow \text{two-fold}}.$$

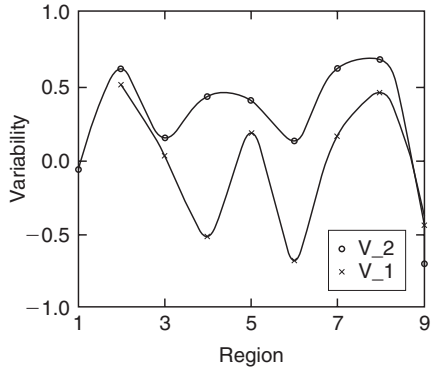


Fig. 1. The components of the regular invariant $\mathcal{I}_\beta$ (chiral variability) along the BRU isolate.

Each component is a within-rotation, within-reversal comparison. However, they are distinguished by the way with which direction and phase interact. They describe a right-hand two-fold followed by a left-hand one-fold screw motion, and a left-hand one-fold followed by a right-hand two-fold screw. It is suggested here that this property be interpreted as chiral variation.

4.4. *A numerical example*

Figure 1 displays the two components

$$v_1 = \text{act} - \text{cta} - \text{tca} + \text{atc} \quad \text{and} \quad v_2 = \text{tac} - \text{cta} + \text{cat} - \text{tca}$$

of the regular invariant $\mathcal{I}_\beta$ for the DNA (log-transformed) frequency data from Table (4.1) along the nine subsequent regions of the genome, suggesting the presence of a dominant direction of higher (chiral) variability.

5. The regular invariants of S_4

The complete symmetry study of data indexed by S_4 is the study of five vector (invariant) subspaces,

$$\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_{\text{Sgn}} \oplus 2\mathcal{V}_\eta \oplus 3\mathcal{V}_\beta \oplus 3\mathcal{V}_\gamma,$$

reducing[1] the original data space (see also [2]). Each of the three copies of the subspaces $\mathcal{V}_\beta$ and $\mathcal{V}_\gamma$ are of dimension three, whereas the two copies of $\mathcal{V}_\eta$ are of dimension two. There are, in addition, the alternating ($\mathcal{V}_{\text{Sgn}}$) and the trivial ($\mathcal{V}_1$) subspaces,

[1] If G has g elements and its r irreducible representations occur with multiplicities d_i then $g = d_1^2 + d_2^2 + \cdots + d_r^2$. E.g. [5].

both of dimension 1. In what follows, a set of bases $\{\mathcal{I}_\beta, \mathcal{I}'_\beta, \mathcal{I}''_\beta\}$, $\{\mathcal{I}_\gamma, \mathcal{I}'_\gamma, \mathcal{I}''_\gamma\}$, and $\{\mathcal{I}_\eta, \mathcal{I}'_\eta\}$ for $3\mathcal{V}_\beta$, $3\mathcal{V}_\gamma$, and $2\mathcal{V}_\eta$, respectively, will be described.

It is assumed, from now on, that the elements of S_4 are listed sequentially following the rows of Matrix (5.1)

$$S_4: \quad \begin{array}{|cccccc|} \hline 1 & (12) & (13) & (14) & (23) & (24) \\ (34) & (123) & (132) & (234) & (243) & (134) \\ (143) & (124) & (142) & (1234) & (13)(24) & (1432) \\ (1324) & (12)(34) & (1423) & (1243) & (14)(23) & (1342) \\ \hline \end{array} \tag{5.1}$$

so that when S_4 acts (via $s\tau^{-1}$) on the positions $\{1, 2, 3, 4\}$ of the symbols in $s = \text{agct}$, the resulting complete set of labels (also indicating the data x indexed by those labels) is given by Matrix (5.2).

$$x: \quad \begin{array}{|cccccc|} \hline \text{agct} & \text{gact} & \text{cgat} & \text{tgca} & \text{acgt} & \text{atcg} \\ \text{agtc} & \text{cagt} & \text{gcat} & \text{atgc} & \text{actg} & \text{tgac} \\ \text{cgta} & \text{tacg} & \text{gtca} & \text{tagc} & \text{ctag} & \text{gcta} \\ \text{tcag} & \text{gatc} & \text{ctga} & \text{catg} & \text{tcga} & \text{gtac} \\ \hline \end{array}\,. \tag{5.2}$$

5.1. *The regular invariants in dimension of* 2

Following Viana [6, Section 3.9], a two-dimensional irreducible representation of S_4, corresponding to the partition $\lambda = 22$, can be obtained from the transitive action of S_4 on the classes

$$\boxed{\begin{array}{c|c} 1 & 2 \\ \hline 3 & 4 \end{array}}, \quad \boxed{\begin{array}{c|c} 1 & 3 \\ \hline 2 & 4 \end{array}}, \quad \boxed{\begin{array}{c|c} 1 & 4 \\ \hline 2 & 3 \end{array}}, \quad \boxed{\begin{array}{c|c} 2 & 3 \\ \hline 1 & 4 \end{array}}, \quad \boxed{\begin{array}{c|c} 2 & 4 \\ \hline 1 & 3 \end{array}}, \quad \boxed{\begin{array}{c|c} 3 & 4 \\ \hline 1 & 2 \end{array}}$$

of equivalent Young diagrams, shown in Table (5.3), where the classes are labeled as $1, 2, \ldots, 6$, respectively. The action gives a permutation representation, indicated here by ξ, with which the projections

$$\mathcal{P}_i = \sum_{\tau \in G_i} \text{Sgn}_\tau \xi_\tau$$

in dimension of one based on the column stabilizers

$$G_1 = \{1, (12), (34), (12)(34)\}, \quad G_2 = \{1, (13), (24), (13)(24)\}, \quad \text{and}$$
$$G_3 = \{1, (14), (23), (14)(23)\}$$

identify the bases

$$u = \boxed{\begin{array}{c|c} 1 & 3 \\ \hline 2 & 4 \end{array}} - \boxed{\begin{array}{c|c} 1 & 4 \\ \hline 2 & 3 \end{array}} - \boxed{\begin{array}{c|c} 2 & 3 \\ \hline 1 & 4 \end{array}} + \boxed{\begin{array}{c|c} 2 & 4 \\ \hline 1 & 3 \end{array}} \equiv 2 - 3 - 4 + 5$$

$$v = \boxed{\begin{array}{c|c} 1 & 2 \\ \hline 3 & 4 \end{array}} - \boxed{\begin{array}{c|c} 1 & 4 \\ \hline 2 & 3 \end{array}} - \boxed{\begin{array}{c|c} 2 & 3 \\ \hline 1 & 4 \end{array}} + \boxed{\begin{array}{c|c} 3 & 4 \\ \hline 1 & 2 \end{array}} \equiv 1 - 3 - 4 + 6.$$

The action of S_4 on $\{u, v\}$, shown in the two rightmost columns of Table (5.3), gives the representation (η) of interest. The resulting matrix representation and its table of characters are shown in (5.4) and (5.5), respectively, corresponding to the elements of S_4 shown on Table (5.1).

τ	$\frac{1\mid 2}{3\mid 4}$	$\frac{1\mid 3}{2\mid 4}$	$\frac{1\mid 4}{2\mid 3}$	$\frac{2\mid 3}{1\mid 4}$	$\frac{2\mid 4}{1\mid 3}$	$\frac{3\mid 4}{1\mid 2}$	u	v
1	1	2	3	4	5	6	u	v
(12)	1	4	5	2	3	6	$-u$	$-u+v$
(13)	4	2	6	1	5	3	$u-v$	$-v$
(14)	5	6	3	4	1	2	v	u
(23)	2	1	3	4	6	5	v	u
(24)	3	2	1	6	5	4	$u-v$	$-v$
(34)	1	3	2	5	4	6	$-u$	$-u+v$
(123)	4	5	1	6	2	3	$-u+v$	$-u$
(234)	2	3	1	6	4	5	$-v$	$u-v$
(134)	4	6	2	5	1	3	$-u+v$	$-u$
(124)	5	4	6	1	3	2	$-v$	$u-v$
(1234)	4	5	1	6	2	3	$u-v$	$-v$
(1324)	6	4	2	5	3	1	$-u$	$-u+v$
(1243)	5	1	4	3	6	2	v	u

(5.3)

η :

1	0	−1	0	1	−1	0	1	0	1	1	−1
0	1	−1	1	0	−1	1	0	1	0	0	−1
−1	0	−1	1	0	−1	0	−1	−1	1	−1	1
−1	1	−1	0	1	−1	1	−1	−1	0	−1	0
0	−1	0	−1	−1	1	1	−1	1	0	1	−1
1	−1	1	−1	−1	0	0	−1	0	1	0	−1
−1	0	1	0	−1	0	0	1	1	0	0	1
−1	1	0	1	−1	1	1	0	0	1	1	0

(5.4)

Bases for the regular invariants in dimension of two Applying Proposition 3.1 to the irreducible representation η, with the labels in (5.1), gives (from the rows of the Fourier transform of x at η) the bases $\{v_1, v_2\}$ and $\{w_1, w_2\}$ shown, in matrix form, in (5.6)–(5.9).

$$\chi_\eta : \begin{array}{|cccccc|} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 0 & 2 & 0 \\ 0 & 2 & 0 & 0 & 2 & 0 \end{array}, \tag{5.5}$$

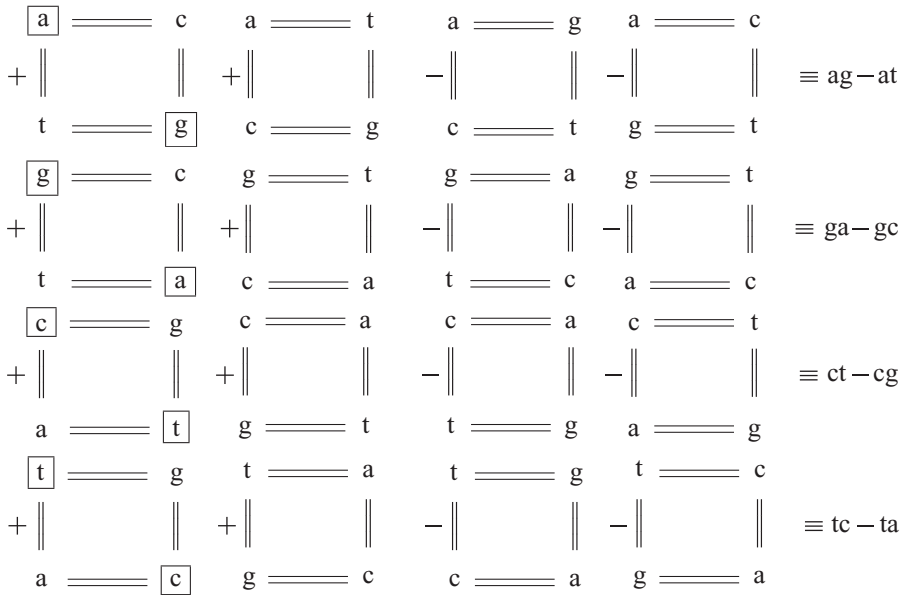
$$v_1 = \begin{bmatrix} \text{agct} & -\text{gact} & \text{cgat} & 0 & 0 & \text{atcg} \\ -\text{agtc} & -\text{cagt} & 0 & 0 & -\text{actg} & -\text{tgac} \\ 0 & 0 & -\text{gtca} & \text{tagc} & \text{ctag} & \text{gcta} \\ -\text{tcag} & \text{gatc} & -\text{ctga} & 0 & \text{tcga} & 0 \end{bmatrix}, \tag{5.6}$$

$$v_2 = \begin{bmatrix} 0 & 0 & -\text{cgat} & \text{tgca} & \text{acgt} & -\text{atcg} \\ 0 & \text{cagt} & -\text{gcat} & -\text{atgc} & \text{actg} & \text{tgac} \\ -\text{cgta} & -\text{tacg} & \text{gtca} & -\text{tagc} & 0 & -\text{gcta} \\ 0 & 0 & 0 & \text{catg} & 0 & \text{gtac} \end{bmatrix}, \tag{5.7}$$

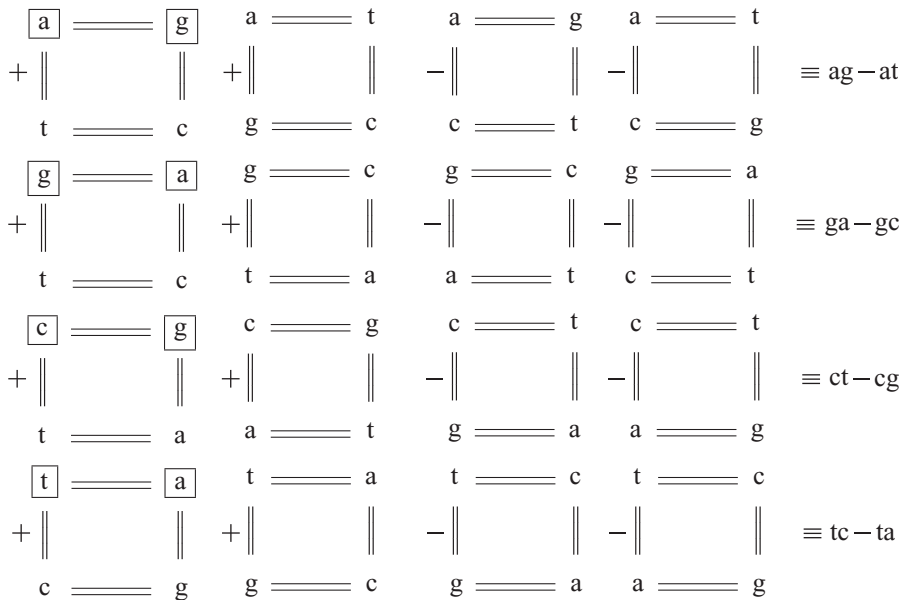
$$w_1 = \begin{bmatrix} 0 & -\text{gact} & 0 & \text{tgca} & \text{acgt} & 0 \\ -\text{agtc} & -\text{cagt} & \text{gcat} & \text{atgc} & -\text{actg} & -\text{tgac} \\ \text{cgta} & \text{tacg} & -\text{gtca} & 0 & 0 & 0 \\ -\text{tcag} & 0 & -\text{ctga} & \text{catg} & 0 & \text{gtac} \end{bmatrix}, \tag{5.8}$$

$$w_2 = \begin{bmatrix} \text{agct} & \text{gact} & -\text{cgat} & 0 & 0 & -\text{atcg} \\ \text{agtc} & 0 & -\text{gcat} & -\text{atgc} & 0 & 0 \\ -\text{cgta} & -\text{tacg} & 0 & -\text{tagc} & \text{ctag} & -\text{gcta} \\ \text{tcag} & \text{gatc} & \text{ctga} & 0 & \text{tcga} & 0 \end{bmatrix}. \tag{5.9}$$

Interpretation Write $uv = \underline{u}m\underline{v}n + \underline{u}n\underline{v}m$, where u, v, m, and n are distinct symbols in $\{\text{a}, \text{g}, \text{c}, \text{t}\}$ so that, for example, $\text{ac} = \text{agct} + \text{atcg}$. With that notation in mind, the following diagram illustrates the 16 components of w_1 (Note that uv sums over nonadjacent positions in the square):



Similarly, redefining $uv = \underline{u}\underline{v}mn + \underline{u}\underline{v}nm$, summing now over two adjacent positions in the square, it follows that w_2 is given by



Similar interpretation is obtained for the basis $\{v_1, v_2\}$.

5.2. *The regular invariants in dimension of* 3

Table (5.10) shows one of the two irreducible representation of S_4 of dimension three [6] indexed in correspondence with the elements of S_4 in Table (5.1). Applying Proposition 3.1 to β with the labels in (5.1), the resulting bases (rows of the Fourier transform of x at β) are shown in (5.11), (5.12), and (5.13), indicated respectively by $\mathcal{I}_\beta$, $\mathcal{I}'_\beta$, and $\mathcal{I}''_\beta$,

$\beta =$

1	0	0	−1	0	0	0	−1	0	0	0	−1	1	1	0	1	0	1
0	1	0	1	1	0	−1	0	0	0	1	0	0	−1	0	0	1	−1
0	0	1	1	0	1	0	0	1	−1	0	0	0	−1	1	0	0	−1
1	0	0	0	1	0	−1	−1	0	1	1	0	1	0	1	0	−1	0
0	0	1	−1	−1	0	1	0	0	0	−1	1	0	0	−1	0	0	1
0	1	0	0	−1	1	1	0	1	0	−1	0	0	1	−1	−1	0	0
0	0	−1	0	0	1	−1	0	−1	0	1	0	0	−1	1	−1	0	−1
−1	0	0	0	1	−1	1	1	0	0	−1	1	−1	0	−1	1	0	0
0	1	0	−1	0	−1	1	0	0	−1	−1	0	0	0	−1	1	1	0
0	−1	1	−1	0	0	0	1	−1	0	0	1	0	1	−1	−1	−1	0
0	0	−1	1	0	1	−1	−1	0	−1	0	−1	0	−1	0	1	0	1
−1	0	−1	1	1	0	0	−1	0	0	1	−1	−1	−1	0	1	0	0

. (5.10)

$$\begin{bmatrix} \text{agct} - \text{gact} + \text{acgt} + \text{atcg} + \text{agtc} - \text{gcat} + \text{actg} + \text{atgc} - \text{gtca} - \text{gcta} - \text{gatc} - \text{gtac} \\ -\text{cgat} + \text{acgt} - \text{gcat} + \text{cagt} + \text{atgc} - \text{tgac} - \text{ctag} + \text{tagc} + \text{ctga} - \text{tcag} - \text{gtac} + \text{tcga} \\ -\text{tgca} + \text{atcg} + \text{actg} - \text{cgta} - \text{gtca} + \text{tacg} - \text{gcta} + \text{ctag} - \text{ctga} + \text{tcag} - \text{tcga} + \text{catg} \end{bmatrix}, \quad (5.11)$$

$$\begin{bmatrix} \text{gact} - \text{cgat} + \text{gcat} - \text{cagt} - \text{cgta} + \text{gtca} + \text{gcta} - \text{ctag} - \text{ctga} + \text{gatc} + \text{gtac} - \text{catg} \\ \text{agct} + \text{gact} + \text{tgca} - \text{acgt} + \text{atcg} - \text{cagt} - \text{atgc} + \text{gtca} + \text{tacg} - \text{tagc} - \text{ctga} - \text{tcga} \\ -\text{atcg} + \text{agtc} - \text{actg} + \text{atgc} + \text{tgac} - \text{tacg} - \text{ctag} + \text{tagc} + \text{gatc} - \text{tcag} + \text{gtac} - \text{catg} \end{bmatrix}, \quad (5.12)$$

$$\begin{bmatrix} \text{gact} - \text{tgca} + \text{gcat} - \text{tgac} + \text{gtca} - \text{tacg} + \text{gcta} - \text{tagc} + \text{gatc} - \text{tcag} + \text{gtac} - \text{tcga} \\ -\text{acgt} + \text{agtc} - \text{cagt} + \text{actg} - \text{atgc} + \text{cgta} + \text{gcta} - \text{tagc} - \text{ctga} + \text{gatc} - \text{tcga} + \text{catg} \\ \text{agct} + \text{gact} + \text{cgat} + \text{acgt} - \text{atcg} + \text{gcat} + \text{cagt} - \text{actg} - \text{tacg} - \text{ctag} - \text{tcag} - \text{catg} \end{bmatrix}. \quad (5.13)$$

each one spanning one copy of $\mathcal{V}_\beta$. Inspection of these bases shows that

(1) $\mathcal{I}_\beta$ describes "g versus a" at positions 1, 3, 4;
(2) $\mathcal{I}'_\beta$ describes "g versus c" at positions 1, 3, 4;
(3) $\mathcal{I}''_\beta$ describes "g versus t" at positions 1, 3, 4.

The position 2, not present above, is in fact g's position in the generating sequence "agct." In each subspace, the three coordinates correspond to positions 1, 3, 4.

The reduction in the second invariant space in dimension three The companion (distinct) irreducible representation of S_4 is simply $\gamma_\tau = \text{Sgn}_\tau \beta_\tau$. It is obtained by multiplying the matrices β_τ in (5.10) by the parity Sgn_τ of the corresponding permutation. Consequently, the bases for the invariant subspaces of interest are the rows of $F_\gamma = \sum_\tau x_\tau \text{Sgn}_\tau \beta_\tau$,

$$\begin{bmatrix} \text{agct} + \text{gact} - \text{acgt} - \text{atcg} - \text{agtc} - \text{gcat} + \text{atgc} + \text{actg} - \text{gtca} + \text{gcta} - \text{gatc} + \text{gtac} \\ \text{cgat} - \text{acgt} + \text{cagt} - \text{gcat} + \text{atgc} - \text{tgac} - \text{tagc} - \text{ctag} + \text{tcag} - \text{ctga} + \text{tcga} + \text{gtac} \\ \text{tgca} - \text{atcg} + \text{actg} - \text{cgta} + \text{tacg} - \text{gtca} + \text{ctag} + \text{gcta} - \text{tcag} + \text{ctga} - \text{catg} - \text{tcga} \end{bmatrix}, \quad (5.14)$$

$$\begin{bmatrix} -\text{gact} + \text{cgat} - \text{cagt} + \text{gcat} - \text{cgta} + \text{gtca} - \text{ctag} - \text{gcta} + \text{gatc} + \text{ctga} + \text{catg} - \text{gtac} \\ \text{agct} - \text{gact} - \text{tgca} + \text{acgt} - \text{atcg} - \text{cagt} - \text{atgc} + \text{tacg} + \text{gtca} + \text{tagc} + \text{ctga} - \text{tcga} \\ \text{atcg} - \text{agtc} + \text{atgc} - \text{actg} + \text{tgac} - \text{tacg} - \text{tagc} - \text{ctag} + \text{tcag} + \text{gatc} + \text{catg} - \text{gtac} \end{bmatrix}, \quad (5.15)$$

$$\begin{bmatrix} -\text{gact} + \text{tgca} + \text{gcat} - \text{tgac} - \text{tacg} + \text{gtca} + \text{tagc} - \text{gcta} + \text{tcag} + \text{gatc} - \text{tcga} - \text{gtac} \\ \text{acgt} - \text{agtc} - \text{cagt} - \text{atgc} + \text{actg} + \text{cgta} + \text{tagc} - \text{gcta} + \text{gatc} + \text{ctga} - \text{catg} - \text{tcga} \\ \text{agct} - \text{gact} - \text{cgat} - \text{acgt} + \text{atcg} + \text{cagt} + \text{gcat} - \text{actg} - \text{tacg} - \text{ctag} + \text{tcag} + \text{catg} \end{bmatrix}. \quad (5.16)$$

indicated, respectively, by $\mathcal{I}_r, \mathcal{I}'_r, \mathcal{I}''_r$.

To outline the interpretation of these subspaces, the following notation is introduced: indicate by $\pm u_k$ the sum of (the data indexed by) the permutations with the symbol u in the kth position that differ only by cyclically moving the remaining symbols in one direction (+) and in the opposite direction (−). Then, for example, the first component $\mathcal{I}_1^\gamma = \text{agct}+\text{gact}-\text{acgt}-\text{atcg}-\text{agtc}-\text{gcat}+\text{atgc}+\text{actg}-\text{gtca}+\text{gcta}-\text{gatc}+\text{gtac}$ of $\mathcal{I}_\gamma$ can be represented as $\pm \text{a}_1 \pm \text{g}_1$. It then follows that $\mathcal{I}_1^\gamma = \pm \text{a}_1 \pm \text{g}_1$, $\mathcal{I}_2^\gamma = \pm \text{a}_3 \pm \text{g}_3$, $\mathcal{I}_3^\gamma = \pm \text{a}_4 \pm \text{g}_4$. Similarly, $\mathcal{I}_1^{\gamma'} = \pm \text{c}_1 \pm \text{g}_1, \mathcal{I}_2^{\gamma'} = \pm \text{c}_3 \pm \text{g}_3$, $\mathcal{I}_3^{\gamma'} = \pm \text{c}_4 \pm \text{g}_4, \mathcal{I}_1^{\gamma''} = \pm \text{t}_1 \pm \text{g}_1, \mathcal{I}_2^{\gamma''} = \pm \text{t}_3 \pm \text{g}_3$, and $\mathcal{I}_3^{\gamma''} = \pm \text{t}_4 \pm \text{g}_4$.

Interpretation Keeping in mind that in the generating permutation the symbol g occupies position number 3, it is observed that each subspace compares g against the cyclic (eventually chiral) changes of the remaining symbols in the remaining positions, one subspace for each one of the three positions 1, 2, or 4 to which g is moved to.

The joint reduction of the data space The vectors $\{\mathcal{F}_1, \mathcal{I}_{\mathrm{Sgn}}, \mathcal{I}_\eta, \mathcal{I}'_\eta, \mathcal{I}_\beta, \mathcal{I}'_\beta, \mathcal{I}''_\beta, \mathcal{I}_\gamma, \mathcal{I}'_\gamma, \mathcal{I}''_\gamma\}$ define the (rows of the) matrix B shown in Appendix A, relative to which $B\mathcal{P}_1 B^{-1} = \mathrm{Diag}(1, 0, \ldots, 0)$, $B\mathcal{P}_{\mathrm{Sgn}} B^{-1} = \mathrm{Diag}(0, 1, 0, \ldots, 0)$, $B\mathcal{P}_\eta B^{-1} = \mathrm{Diag}(0, 0, I_4, 0, \ldots, 0)$, $B\mathcal{P}_\beta B^{-1} = \mathrm{Diag}(0, \ldots, 0, I_9, 0, \ldots, 0)$, and $B\mathcal{P}_\gamma B^{-1} = \mathrm{Diag}(0, \ldots, 0, I_9)$.

The regular action of S_4 extended to the components of the joint basis gives a representation isomorphic to

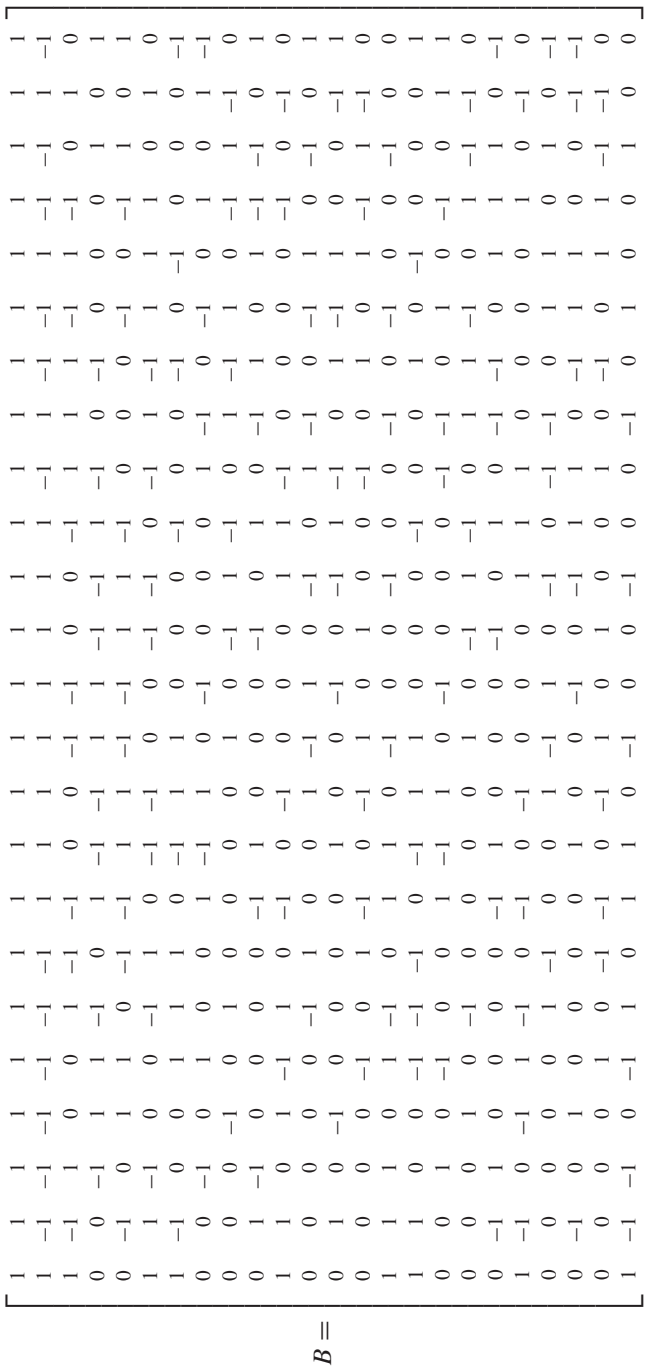
$$\mathrm{Diag}(1, \mathrm{Sgn}_\tau, \eta, \eta_\tau, \beta_\tau, \beta_\tau, \beta_\tau, \gamma_\tau, \gamma_\tau, \gamma_\tau).$$

6. Comments and summary

This chapter developed the regular invariants for S_3 and S_4 and obtained their interpretations within the context of data indexed by these two groups. These regular invariants are to be understood in analogy to the usual invariants (leading to the sample mean and variance) that are obtained under the action of the full symmetric group shuffling the sample labels $\{1, 2, \ldots, n\}$. The regular action shuffles the experimental labels (permutations) by multiplication and introduces a degree of arbitrariness. The interpretation of the regular invariants, in data analytic applications, should be better understood so that the corresponding parametric hypotheses can be formulated and scientific explanations suggested.

It is also observed that symmetry studies of regular (hence transitive) actions are simpler to interpret. That is, in the regular case, Step (4) described in Section 3 always gives a single orbit. Otherwise, when orbits appear, as illustrated by (3.3), the canonical projections collect together pieces from different orbits and label the resulting classes by the irreducible representations, in one-to-one correspondence with the canonical projections. This "crossing out into nonisomorphic subspaces" clearly complicates any eventual interpretations. This is why the representations might be restricted to the underlying (transitive) orbits, for data analytic purposes.

Appendix A



References

[1] K.F. Riley, M.P. Hobson, S.J. Bence, *Mathematical Methods for Physics and Engineering*, 2nd ed., Cambridge University Press, New York, NY, 2002.
[2] H. Bacry, *Leçons sur la theorie des groupes et les symétries des particules elémentaires*, Dunot, Paris, France, 1967.
[3] M. Viana, Symmetry studies and decompositions of entropy, Entropy **8**, no. 2, (2006) 88–109.
[4] M. Viana, *Symmetry Studies- An Introduction to the Analysis of Structured Data in Applications*, Cambridge University Press, New York, NY, 2008.
[5] J-P. Serre, *Linear Representations of Finite Groups*, Springer-Verlag, New York, 1977.
[6] P. Diaconis, A generalization of spectral analysis with application to ranked data, Ann. Stat. **17** (1989) 949–979.

Index

M

Q