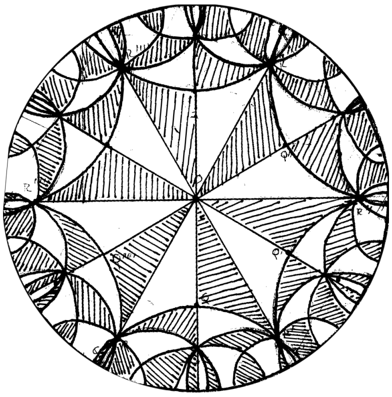
The Word Problem and Combinatorial Methods for Groups and Semigroups

By

Carl-Fredrik Nyberg Brodda



A thesis submitted to the School of Mathematics in partial fulfilment
of the requirements for the degree of Doctor of Philosophy
in the subject of Mathematics.

University of East Anglia
Norwich, United Kingdom

September | 2021

Contents

Acknowledgements

First, I wish to thank my supervisor Bob Gray, for his relentless support, encouragement, patience in listening to many rambling arguments, and very careful reading of my written work. I am very aware of just how lucky I have been with regards to my supervisor. At the same time, I would also like to thank both Laure Daviaud and Tristan Gray, for their warm and kind support (the former), and for preventing me from taking up too much of Bob's time this past month (the latter). I would also like to acknowledge the hospitality shown by the School of Mathematics, its undergraduate students, and the many wonderful researchers and members of staff there, as well as its coffee machine, all of whom I have come to known very well over these three years.

I have made countless new friends during my time at UEA and in Norwich; I could not name them all here, but I would like to especially thank Ben, George, and Mike, for being part of creating such a welcoming atmosphere when I first arrived to the Ph.D. office; all other members of said office for maintaining this atmosphere throughout this time; and finally Monica, for keeping me sane during a global pandemic (no small feat) and giving unparalleled surfing lessons.

Many members of the mathematical community have in various degrees offered interest in my written work and support in my scientific endeavours. This has been a tremendous source of motivation in continuing my research. I would especially like to thank Peter J. Cameron for his continuous assistance, kindness, and encouragement. W. Magnus wrote, the very year I was born, that "mathematicians understand each other no matter where they come from. [...] Nothing is more international than the community of mathematics."[1] I am proud to have become part, and to continue being a part, of this welcoming and accepting international community.

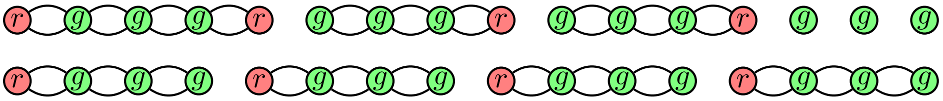
Finally, I wish to thank my family – mamma, pappa, and Stella – for your endless support, patience, and confidence that I would finish what I started. I can only contemplate this with grateful mind and silent wonder; nothing motivates me more than you do.

[1] W. Magnus, "The Significance of Mathematics: The Mathematicians' Share in the General Human Condition", *Amer. Math. Monthly*, 1997, **104**:3 (1997), p. 269.

Preface

This thesis is a contribution to a field of mathematics which, for the past hundred years or so, has been known as semigroup theory. This first sentence already presents the reader with a curious situation: certainly, no mathematician (or otherwise) would claim that the sentence is a mathematical statement; and yet, simultaneously, its meaning could only be understood by a mathematician, for nobody else (sound of mind) would be aware of what a "semigroup" is. In this way we would reach the somewhat puzzling conclusion that the first sentence of this (mathematical) thesis is a non-mathematical sentence, which nevertheless nobody but a mathematician could understand. In light of this, I would like to revise my first sentence.

Before this revision, however, I wish to humanise the abstract topics which are treated within these pages. In a very primitive form, the subject matter of this thesis has been around since the Palaeolithic era. Consider the following situation, based on one considered by Lentin.[1] Let us suppose an ancient bead-maker has many coloured beads. They string together these beads in many patterns, which are then assembled to larger necklaces. They may ask themselves the question: I take three patterns. I repeat the first once, the second twice, and the third three times, and then string them all together into one necklace in that order. Is there some pattern which, when repeated four times, lets me form the same necklace? The answer turns out to be "yes"; I have drawn out one such solution below, where the beads of a given pattern are kept together by small pieces of string. The two rows, each considered as a whole, are identical.



Symbolically, if x, y, z represent the three patterns and x, yy, zzz represent the prescribed repeating of these patterns, then one has a necklace $xyyzzz$. The question above asks: is there some pattern of beads w such that $xyyzzz$ is the same necklace as $wwww$? That is, can we find some patterns x, y, z, w such that $xyyzzz = wwww$? Let r denote a red bead, and g a green bead; let rg denote the pattern with a red bead followed by a green bead, and gr a green followed by a red, etc. The above pictorial solution to our problem then becomes $x = rgggr$, $y = gggr$, $z = g$, and $w = rggg$. The necklace becomes $rgggrgggrgggrggg$. This solution is given symbolically below; the reader is invited to compare it to the pictorial solution.

$$\rlap{$\overbrace{}^{x}\overbrace{}^{y}\overbrace{}^{y}\overbrace{}^{z}\overbrace{}^{z}\overbrace{}^{z}$}\underbrace{rggg}_{w}\underbrace{rggg}_{w}\underbrace{rggg}_{w}\underbrace{rggg}_{w}$$

This bead problem is an example of a question which belongs to the realm of combinatorial semigroup theory. It is a difficult problem in full generality; giving a procedure for solving

[1] A. Lentin, "Équations dans les monoïdes libres", *Math. Sci. Humaines*, 1970, **31**: 5–16.

"bead equations" was only accomplished in the 1980s, in a very long and incredibly intricate paper.[2] Human consciousness is a creative one, and excels at spotting patterns; for this reason there seems to be no doubt that this kind of combinatorial thinking, of overlapping patterns and of noticing possible subdivisions, would have been present in the minds of the bead-makers, if only subconsciously. One finds this proto-combinatorial mindset present also in Antiquity. Building on the ideas of atoms and the indivisible, as developed by Leucippus and Democritus in the 5th century BCE, Lucretius (99–55 BCE) in his *De rerum natura* argued in favour of this view by noting the divisibility of sentences into individual words; words into individual letters; and finally the indivisibility of letters themselves.[3] This thinking, of stringing letters together as one strings beads together, which seems so very natural to us after centuries of being exposed to it, is also combinatorial in nature. Indeed, it is precisely by virtue of this naturality that we could in the first place imagine a bead to be represented by the letter r or g; had we not had this implicit combinatorial mindset, we would when faced with the word $rgggrgggrgggrggg$ be utterly unable to understand the connection with beads, and only read the word as some chant or guttural uttering.

Let us consider a slightly more abstract problem. Let us say our bead-maker could, at any point, replace any two adjacent red beads by exactly three green beads anywhere inside any necklace they have made. Conversely, any three adjacent green beads could be replaced with two red beads. Symbolically, we impose that the equality $rr = ggg$ hold. Given two necklaces, one could ask: which necklaces can be transformed into one another? For example, the necklace $grrg$ could be transformed into the necklace $ggggg$, by replacing the interior red beads, and this latter necklace could be transformed into the necklace $ggrr$, by replacing the right-most run of green beads. Thus, $grrg$ and $ggrr$ are, in this sense, equivalent necklaces, even though there is no *single* transformation that turns one into the other. It thus seems like there is something non-trivial in telling necklaces apart. Is there a procedure by which, given two necklaces, one can always tell if the necklaces are the same, or not? For example, are the necklaces $rrgg$ and $ggrr$ the same? It may come as a surprise to the reader that the general problem of, given an equality $u = v$ of two patterns of red and green beads (e.g. $rgrrg = rgg$ or $gg = ggr$), finding a procedure for telling necklaces apart with respect to this form of equality, remains *even today* an open problem! Indeed, I recently wrote an extensive survey whose subject matter is this very problem, and the century-long history of attempts to solve it.[4]

In modern terminology, the first bead problem posed above is asking for solutions to the Diophantine equation $xy^2z^3 = w^4$ over a free monoid; the second is asking for a solution to the word problem in the monoid $\mathrm{Mon}\langle r, g \mid r^2 = g^3 \rangle$ and, more generally, for a solution to the word problem in one-relation monoids $\mathrm{Mon}\langle r, g \mid u = v \rangle$. These are fundamental questions about *free monoids*. To be clear, our bead-maker, through no fault of their own, would not have been asking these more abstract questions about free monoids, nor would they have been aware that these objects are what they were subconsciously studying. For this reason, and somewhat regrettably, it would therefore not be terribly accurate to say that the first semigroup theorists worked alongside mammoth hunters. Instead, the theory would lay dormant, awaiting an appropriate level of mathematical maturity to be attained by humanity before it could be treated. To put it mildly: this would take some time. In the words of Lentin: "*adieu, monoïdes*

[2] G. S. Makanin, "The problem of the solvability of equations in a free semigroup" (in Russian), *Mat. Sbornik* 1977, **103**(2): 147–236.

[3] Lucretius, *De rerum natura*, I, verses 823–826; II, verses 688–694; III, verses 1013–1018.

[4] C.-F. Nyberg-Brodda, "The word problem for one-relation monoids: a survey", *Semigroup Forum*, 2021, **103**:2, 297–355.

libres, adieu pour des siècles!"[5] It would not be until the early 20th century that a more general and abstract framework would develop, allowing for proper investigations of free monoids and quotients of the same. This framework is the realm of *combinatorial semigroup theory*.

Palaeolithic considerations completed, I am now able to present a revised, non-antinomic, and meaningful first sentence: this thesis represents a contribution to combinatorial semigroup theory. Even if one disregards our bead-maker as a stretch, this combinatorial semigroup theory is older than the "ordinary". By "ordinary", I here mean that point of view which studies semigroups from the point of view of their properties as abstract algebraic objects, rather than from that which regards them as combinatorial objects. In other words, ordinary (or perhaps "axiomatic") semigroup theory is the area which studies (by means that I shall not describe) the properties of collections of objects, and the means by which these objects can be composed with one another to form another object of the collection. Thus, adding integers is a semigroup-theoretic operation; as is composing functions; as is composing beads to form necklaces. The advantage of studying such general and ubiquitous operations from an abstract point of view is the general applicability of any results; the drawback is that the innumerable disparate natures of the objects in question means that one can rarely prove anything at all for the class of all semigroups. Thus, to approach semigroups from this internal, axiomatic point of view, one must always begin with a restriction to some interesting class. There is nothing particular about semigroup theory in this regard; quoting M.-P. Schützenberger, "it is not the general notion of a group that is interesting; it is always a very precise class of groups that we study. Lie groups are very special objects; they are admirable due to their applications, but their properties teach me nothing about Coxeter groups, and vice versa."[6] However, the approach via combinatorial semigroup theory is not such a restriction. It represents a different perspective on semigroup theory. One might bear in mind the words of Roger Lyndon, who describes combinatorial group theory as "just a state of mind"[7]. From the point of view lent to us by this state of mind, certain semigroups which by pure means of axiomatisation would seem entirely unnatural (for example, one-relation semigroups) become natural; while, conversely, certain semigroups which are easy to axiomatise (for example, von Neumann regular semigroups) seem rather difficult to approach.

It is, for this reason, not an intention of mine (or an opinion of mine) to state that the approach to the subject of semigroup theory in this thesis, or indeed of combinatorial semigroup theory in general, is superior to the ordinary. Rather, I wish to convey that once this combinatorial perspective has been attained, the intricate world of semigroup theory opens up to all areas of mathematics, and researchers from vastly different backgrounds should find something they will (at the very least) recognise within. That is not to say that anyone doing so shall find themselves one morning woken up, like Grigor Samsa, as a monstrous semigroup theorist. Indeed, it grants the benefit of attacking a given problem with a multitude of methods form a multitude of branches, each drawn from vast and disparate fields; one may study the overlaps of words or their language-theoretic properties; describe rewriting systems and their (co)homological properties; describe and study algebraic structures graphically; investigate sequences of elementary transformations as topological homotopy relations or decision-theoretically; simulate semigroups via finite automata; and

[5] A. Lentin, *ibid.*, p. 6.

[6] A. Connes; A. Lichnerowicz; M.-P. Schützenberger, *Triangle of thoughts*, (Providence, RI: Amer. Math. Soc., 2001), p. 30.

[7] R. Lyndon, "Problems in combinatorial group theory", *Ann. of Math. Stud.* **111**, 1987, 3–33, p. 3.

each such branch (and countless others), serving to understand some class of combinatorial semigroups, itself branches indefinitely. What a remarkable world of endless discovery!

I have deliberately not given an explicit definition of combinatorial semigroup theory in this preface. If one were to attempt such a definition, one ought to bear in mind that the first appearance of the term combinatorial *group* theory was as the title of the (now famous) 1966 book by Magnus, Karrass & Solitar.[8] The very first sentence of the book by Chandler & Magnus on the history of that subject begins: "combinatorial group theory may be characterised as the theory of groups which are given by generators and relations".[9] Perhaps, then, one could add the prefix *semi-* to both occurrences of the word *group* in that sentence, and thereby have obtained a characterisation of combinatorial semigroup theory? Again quoting Lyndon, "this hardly does justice to the goals or methods of the subject". It seems better to leave the area implicitly defined; there is not yet any book published with the title *Combinatorial Semigroup Theory.* On the other hand, there are many connections with the group-theoretic perspective in this thesis. The "father of semigroup theory", A. K. Sushkevič, wrote in his own Ph. D. thesis (published 99 years ago!) that a problem about semigroups – or indeed monoids – would be considered solved if it could be reduced to a problem about groups.[10]. This principle will be applied consistently also in this thesis; for example, when reducing problems about special monoids to their group of units in Chapter 3.

One part of the title of my thesis remains to be explained. The importance of the word problem to combinatorial semigroup theory cannot be overstated. Indeed, the first paper dealing whatsoever with combinatorial semigroup theory studies only this problem, in a form that is remarkably similar to the modern formulation of the same; this is the 1914 paper by the Norwegian mathematician A. Thue.[11] This paper was the true genesis of combinatorial semigroup theory, to borrow a turn of phrase from H. Wussing.[12] It was published three years after the appearance of a much more famous paper by M. Dehn[13], which introduced the word problem for groups, but unlike Dehn's highly topological considerations the phrasing by Thue is entirely combinatorial. Thue's paper laid the groundwork for the theory of *rewriting systems* (also called *Thue systems*) in an attempt to solve this problem. In 1942, M. Newman, one of the many early British pioneers of semigroup theory who worked at Bletchley Park during the war, published an article which properly set up the combinatorial foundations of the theory of such systems, including a proof of a key lemma which today bears his name.[14] Newman does not refer to Thue's paper. It seems instead that the first to recognise the importance of that paper was E. Post, when he in 1947 gave one of that magnificent year's proofs of the general *unsolvability* of the word problem for semigroups; Thue's name appears already in the title of the paper.[15] Many have worked on these problems, and other related problems since; far too many to recount here. We mention only that another famous member

[8] W. Magnus; A. Karrass; D. Solitar, *Combinatorial group theory. Presentations of groups in terms of generators and relations*, (New York-London-Sydney: Interscience Publishers 1966).

[9] B. Chandler; W. Magnus, *The history of combinatorial group theory* (New York: Springer, 1982), p. 3.

[10] A. K. Sushkevič, *The theory of operations as the general theory of groups*, Ph. D. thesis, (Voronezh State University, 1922). 80 pp. The quote is taken from §38.

[11] A. Thue, "Problem über Veränderungen von Zeichenreihen nach gegebenen Regeln", *Christiana [Oslo] Videnskaps-selskabs Skrifter, I. Math. naturv. Klasse*, 1914, **10**.

[12] H. Wussing, *Die Genesis des abstrakten Gruppenbegriffes. Ein Beitrag zur Entstehungsgeschichte der abstrakten Gruppentheorie* (Berlin: VEB, 1969).

[13] M. Dehn, "Über unendliche diskontinuierliche Gruppen", *Math. Ann.* 1911, **71**(1): 116–144.

[14] M. H. A. Newman, "On theories with a combinatorial definition of 'equivalence'", *Ann. of Math. (2)*, 1942, **43**: 223–243.

[15] E. Post, "Recursive unsolvability of a problem of Thue", *J. Symbolic Logic*, 1947, **12**: 1–11.

of Bletchley Park, Alan Turing, proved in 1950 that the word problem is unsolvable for cancellative semigroups.[16] This would be his last paper on pure mathematics. To end this informal discussion on the word problem, I would like to quote P. Hall, writing in 1958: "in spite of, or perhaps because of, their relatively concrete and particular character, [word problems] appear, to me at least, to offer an amiable alternative to the ever popular pursuit of abstractions."[17] I have certainly found this to be the case.

Above all, I believe combinatorial semigroup theory is a very human subject. Certainly, the methods, results, and proofs of this thesis will not appear particularly human (at times, they may even appear inhumane) to one not familiar with mathematics, or indeed the subject. However, I believe that the fundamental questions of combinatorial semigroup theory could be explained – as our favourite bead-maker has demonstrated with some of the problems – in a short time to anyone interested. The simple human curiosity which posed such questions in the first place is never far away.

Throughout this thesis, and especially in Chapter 1, there is an abundance of historical references and connections. To be clear, this is not a thesis on the history of mathematics. However, just as history is the accumulated experience of mankind, the history of mathematics is the accumulated experience of mathematicians. It would be insufferably arrogant not to take this experience into account. Mathematics, like philosophy, is virtually inseparable from its history.[18] For this reason, I have also made virtually no attempt to perform such a separation. I will only make two further comments on the presence of history of mathematics within this thesis. First, W. Magnus, one of the great pioneers of combinatorial group theory, was in 1934 given the task of writing an article on *general group theory* and its history. He consulted Emmy Noether, who gave him the following rather laconic piece of advice: "First write down what you know. Then check the literature and expand".[19] In writing this thesis, I have attempted to follow this advice. Finally, there are, without a doubt, many anachronisms and minor inaccuracies as to the idea history of the areas of mathematics presented within this thesis. But, quoting J. L. Borges, "reality is partial to symmetries and slight anachronisms".[20]

In many places, including this preface, the reader may already have noted that the presence of footnotes and additional embellishments somewhat detracts from a perfectly streamlined presentation of the material. There are two reasons for this. The first is related to the fact that achieving a truly streamlined presentation is a very difficult task; and, as J. Jaynes puts it, "poems are rafts clutched at by men drowning in inadequate minds".[21] I hope that such words may aid in forgiving the addition of such "poems" which constitute this – rather (but hopefully not all too) often – extraneous material. The second, more important, is that this is how the material was first presented to me, as I read it, finding new papers; surprising links across decades and centuries, countries and languages, hinting at a depth that I could not achieve by a linear narrative. I can only hope that my excitement comes across in my writing!

C.-F. Nyberg-Brodda
September 30, 2021

[16] A. Turing, "The word problem in semi-groups with cancellation", *Ann. of Math.*, 1950, **52**: 491–505.

[17] P. Hall, "Some word-problems", *J. London Math. Soc.*, 1958, **33**: 482–496, p. 496.

[18] Quoting H. Edwards, "Read the masters!", p. 108; in L. Steen, *Mathematics tomorrow*, (Springer, 1981).

[19] Chandler and Magnus, *ibid.*, p. 75.

[20] J. L. Borges, *Fictions; The South.* p. 148.

[21] J. Jaynes, *The origin of consciousness in the breakdown of the bicameral mind*, (Houghton Mifflin, Boston: 1976). Quote from p. 256.

Abstract

The subject matter of this thesis is combinatorial semigroup theory. It includes material, in no particular order, from combinatorial and geometric group theory, formal language theory, theoretical computer science, the history of mathematics, formal logic, model theory, graph theory, and decidability theory.

In Chapter 1, we will give an overview of the mathematical background required to state the results of the remaining chapters. The only originality therein lies in the exposition of *special* monoids presented in §1.3, which unifies the approaches by several authors.

In Chapter 2, we introduce some general algebraic and language-theoretic constructions which will be useful in subsequent chapters. As a corollary of these general methods, we recover and generalise a recent result by Brough, Cain & Pfeiffer that the class of monoids with context-free word problem is closed under taking free products.

In Chapter 3, we study language-theoretic and algebraic properties of special monoids, and completely classify this theory in terms of the group of units. As a result, we generalise the Muller-Schupp theorem to special monoids, and answer a question posed by Zhang in 1992.

In Chapter 4, we give a similar treatment to weakly compressible monoids, and characterise their language-theoretic properties. As a corollary, we deduce many new results for one-relation monoids, including solving the rational subset membership problem for many such monoids. We also prove, among many other results, that it is decidable whether a one-relation monoid containing a non-trivial idempotent has context-free word problem.

In Chapter 5, we study context-free graphs, and connect the algebraic theory of special monoids with the geometric behaviour of their Cayley graphs. This generalises the geometric aspects of the Muller-Schupp theorem for groups to special monoids. We study the growth rate of special monoids, and prove that a special monoid of intermediate growth is a group.

Chapter 1

Background and Introduction

Synopsis

In this discursive and definition-heavy chapter we shall give the necessary mathematical background for the reader to be able to approach and understand the results and proofs of the following chapters, as well as appreciate the broader picture into which they fit. None of the results presented in this chapter should be considered as original. In §1.1 we first present the elements of semigroup theory, and how it relates to groups and monoids. As part of this, we present an introduction to some aspects of free monoids and free groups, including combinatorics on words and the theory of presentations. We present certain decision problems which will be of interest in this thesis, and certain structural properties of monoids and groups which aid in solving such problems. In §1.2 we present the elements of formal language theory and rewriting systems. This includes an overview of studying the word problem of monoids as a language-theoretic object. In §1.3, we give an overview of the classical theory of special monoids, and provide a unification of the various approaches that have been made to the subject. In §1.4 we give an overview of some aspects of graph theory. This includes the notion of context-free graphs and Cayley graphs of monoids and groups. We give an overview of the connection between formal logic and graphs, and end the section with a brief incursion into geometric group and semigroup theory. In §1.5, we give a case study of one of the most fundamental objects of combinatorial group theory – one-relator groups. Finally, in §1.6 we present a table containing referential material on various properties and decision problems for different classes of groups.

Common notation

Throughout this thesis, some notation and definitions will be fixed. The usual notation of set theory will be adopted. In particular $\in$ denotes set membership; $\subset$ indicates proper inclusion; and $\subseteq$ inclusion. The empty set is denoted $\varnothing$. The set of natural numbers $\{0, 1, \dots, \}$ will be denoted $\mathbb{N}$ or ω. The set of integers, rational numbers, and real numbers will respectively be denoted by $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$. The first infinite cardinal will be denoted by $\aleph_0$. A countable set is one of cardinality $\leq \aleph_0$. When considering a function $f \colon A \times B \to C$, then for every $a \in A, b \in B$, and $c \in C$ we will write $f(a, b) = c$ rather than $f((a, b)) = c$. If $R \subseteq X \times X$ is an equivalence relation, then for $x \in X$ we let $[x]_R$ denote the equivalence class of x, i.e. the set of elements $\{y \in X \mid xRy\}$. A function $f \colon A \to B$ is formally a subset of $A \times B$. We will assume the reader is familiar with the notions of recursive and recursively enumerable sets and functions.

We shall later define groups; once we have done so, we shall assume that all standard notions from group theory become immediately familiar to the reader (not many will be necessary). If G is a group with a finite index subgroup H such that H has a property $\mathcal{P}$ of groups (for example, being free), then we say that G is *virtually* $\mathcal{P}$ (for example, being virtually free). That is, a group G being virtually free is synonymous to G being free-by-finite, i.e. there existing a short exact sequence

$$1 \to H \to G \to K \to 1$$

where H is a free group, and K is a finite group. The commutator $[g, h]$ of two elements g, h of a group G denotes the element $ghg^{-1}h^{-1}$. Let P be a property of groups preserved under isomorphism. A group G is said to be *residually* P if for every non-trivial $g \in G$ there exists a homomorphism $\phi_g \colon G \to H$ to a group H with property P such that $\phi_g(g) \neq 1$. Analogously, for a monoid (semigroup) property P we define a monoid (semigroup) M to be *residually* P if for any pair $m, n \in M$ with $m \neq n$ there exists a homomorphism $\phi_{m,n} \colon M \to H$ to a monoid (semigroup) H with property P such that $\phi_{m,n}(m) \neq \phi_{m,n}(n)$. We say that a group G is *linear* if it is a subgroup of $\mathrm{GL}_n(\mathbb{F})$ for some $n \geq 1$ and a field $\mathbb{F}$.

Finally, the end of a proof is signified by $\square$, and the end of an example is signified by $\triangle$.

1.1 Semigroups, monoids, groups

The material herein is standard; we shall base the exposition given on Berstel & Perrin [54], and include material from Ljapin [283], Clifford & Preston [112, 113], and Howie [222, 224].

A *semigroup* $(S, \cdot)$ is a set S together with a binary operation $\cdot \colon S \times S \to S$, called the *multiplication* in S, such that $\cdot$ is associative, i.e. $x \cdot (y \cdot z) = (x \cdot y) \cdot z$. When speaking of a semigroup $(S, \cdot)$, we shall often (almost exclusively) refer only to S, the underlying set, when speaking of the semigroup, keeping the binary operation implicit. We shall generally write $x \cdot y$ simply as xy, and a simple induction yields that expressions such as $xyzw$ have a definite meaning, and we need never use parentheses.[1] A *monoid* is a semigroup M in which there exists an element $1 \in M$, such that for all $x \in M$ we have $x \cdot 1 = 1 \cdot x = x$. Such an element is called an *identity element* of M. An identity element is always unique.

If M is a monoid with identity 1, then $m \in M$ is called *left invertible* (or a *left unit*) if there exists some $m' \in M$ with $m'm = 1$. Symmetrically, we say that $m \in M$ is *right invertible* (or a *right unit*) if there exists some $m' \in M$ with $mm' = 1$. In these cases, m' is called a *left*

[1] This is what P. Hall calls “the paradox of the pointlessness of punctuation” [199, p. 485].

(resp. *right*) *inverse* for m. Note that left and right inverses are not, in general, unique. We say that m is *invertible* if it has both a left inverse m' and a right inverse m'', in which case an easy exercise shows that $m' = m''$. A *group* is a monoid M with identity 1 such that every $x \in M$ is invertible. The (necessarily) unique inverse of x is then denoted x^{-1}. Note that $(x^{-1})^{-1} = x$.

Let M be a monoid. If $X, Y \subseteq M$, then by $X \cdot Y$, or simply XY, we mean the set

$$XY = \{xy \mid x \in X, y \in Y\}.$$

This is an associative operation on subsets of M, so e.g. XYZ has a definite meaning for all $X, Y, Z \subseteq M$. If $x \in M$, then for brevity we shall write xY resp. Yx rather than $\{x\}Y$ resp. $Y\{x\}$. If $X \subseteq M$, then we define $X^0 = \{1\}$ and recursively set $X^n = XX^{n-1}$. If $X \subseteq M$ and $X^2 \subseteq X$, then we say that X is closed under multiplication, and we say that X is a *subsemigroup* of M. If additionally $1 \in X$, then we say that X is a *submonoid* of M, and we denote this by $X \leq M$. Note that a subsemigroup X of a monoid M can itself be a monoid without being a submonoid of M; that is, X might have an identity without this element being the identity of M. Such subsemigroups will *not* be called submonoids. The set of invertible elements of M forms a submonoid of M, and is called the *group of units* of M, denoted $U(M)$. If $e \in M$ is such that $e^2 = e$, then e is an *idempotent* of M. The set of idempotents of M is denoted $E(M)$. Note that $1 \in E(M)$. For each idempotent $e \in M$, the set

$$eMe = \{eme \mid m \in M\}$$

is a subsemigroup of M, and is a monoid with e as its identity element; it is not hard to see that it is the largest monoid contained in M with this property. The subsemigroup eMe is called the *localisation* of M at e. The group of units $U(eMe)$ of eMe is a group whose identity element is e. The *maximal subgroups* of M is the set $\{U(eMe) \mid e \in E(M)\}$. In particular $U(M)$ is a maximal subgroup of M.

We will call a *congruence* ϱ on a semigroup, monoid, or group M any equivalence relation $\varrho \subseteq M \times M$ on M such that, for all $m, m' \in M$ and all $x \in M$, we have

$$m \varrho m' \implies (mx)\varrho(m'x) \quad \text{and} \quad (xm)\varrho(xm').$$

We define *semigroup, monoid*, and *group homomorphisms* in the standard way; in particular, if $\varphi \colon M \to N$ is a monoid homomorphism for monoids M, N with identity elements 1_M resp. 1_N, then we require $\varphi(1_M) = 1_N$. For a monoid homomorphism $\varphi \colon M \to N$, the equivalence relation $\sim_\varphi \subseteq M \times M$ defined by $x \sim_\varphi y$ if and only if $\varphi(x) = \varphi(y)$ is easily seen to be a congruence. This congruence is called the *kernel* (sometimes *nuclear congruence*) of φ, and is denoted $\ker \varphi$. Conversely, if ϱ is a congruence on M, then the set M/ϱ of equivalence classes of ϱ has a natural monoid structure inherited from M, and the map $M \to M/\varrho$ sending an element to its equivalence class is a monoid homomorphism with kernel ϱ.

Let M be a monoid and $X \subseteq M$. The *submonoid generated by* X is denoted $\langle X \rangle$, and is defined as the smallest submonoid of M containing X. Equivalently, it is the set of all finite (including empty) products $x_1 x_2 \cdots x_n$, where $x_i \in X$ for every $1 \leq i \leq n$. We will sometimes write X^* instead of $\langle X \rangle$. In particular, $\varnothing^* = \langle \varnothing \rangle = \{1\}$. If $\langle X \rangle = M$, then we say that X *generates* M, or alternatively that it is a *generating set* for M. If there exists some $X \subseteq M$ such that $|X| < \aleph_0$ and $\langle X \rangle = M$, then we say that M is *finitely generated* (by X). If no such X exists, then we say that M is *non-finitely generated.* We extend all above notions to subsemigroups, disallowing empty products. That is, the subsemigroup generated by $\varnothing$ is not a well-defined object. For groups, we do need some further comments. If M is a monoid and $X \subseteq M$ consists entirely of invertible elements, then the *subgroup generated by* X is the smallest subgroup of M containing X. We shall usually denote the subgroup generated by X as $\langle X \rangle_{\text{gp}}$.

Note that, if X^{-1} denotes the set of inverses of elements from X, we have $\langle X \rangle_{\text{gp}} = \langle X \cup X^{-1} \rangle$.

We extend the usage of *finitely generated subgroup* or indeed *finitely generated groups* to this context; however, note that a subgroup of a monoid is finitely generated (as a subgroup) if and only if it is finitely generated as a submonoid. A group is also finitely generated as a group if and only if it is finitely generated as a monoid. Thus, we may without ambiguity speak of *finitely generated groups*. The benefit of defining finite generation as above is that a finitely generated group means essentially the same to us as it does to an ordinary harmless group theorist.[2]

A *left ideal* of a semigroup (or monoid) S is a non-empty subset $I \subseteq S$ such that $SI \subseteq S$. Symmetrically, a *right ideal* is a subset $I \subseteq S$ such that $IS \subseteq S$. A *two-sided ideal*, or simply *ideal*, is one which is a left and right ideal. The ideal structure of semigroups and monoids is rich.[3] While, for the most part, we are not concerned with this structure, there are some fundamental equivalence relations, known as *Green's relations*, which are based on this ideal structure. These relations will be primarily used for notational convenience. One could instead, if so inclined, directly phrase these relations in terms of divisibility conditions on the elements. Given a monoid M, we define four equivalence relations $\mathscr{R}, \mathscr{L}, \mathscr{J}$, and $\mathscr{H}$, as follows:

$$\begin{aligned} m \mathscr{R} m' &\iff mM = m'M, \\ m \mathscr{L} m' &\iff Mm = Mm', \\ m \mathscr{J} m' &\iff MmM = Mm'M, \\ m \mathscr{H} m' &\iff m \mathscr{R} m' \quad \text{and} \quad m \mathscr{L} m'. \end{aligned}$$

These relations were first introduced by J. Green in 1951 [174], and form the cornerstones of large swathes of semigroup theory. Note that for semigroups one must (of course) append an identity to the semigroup in order for the above relations to be equivalence relations. Now for any monoid, we have $\mathscr{R}, \mathscr{L} \subseteq \mathscr{J}$, and it is not hard to show (see e.g. [54, Proposition 5.1]) that $\mathscr{R} \circ \mathscr{L} = \mathscr{L} \circ \mathscr{R}$, where $\circ$ denotes composition of binary relations. This relation $\mathscr{R} \circ \mathscr{L}$ is usually denoted $\mathscr{D}$. We shall not need many non-trivial properties of Green's relations.

Various restricted classes of monoids are of interest in this thesis (and beyond). A monoid M is said to be *left cancellative* if for all $x, y, z \in M$, we have $xy = xz$ implies $y = z$. We symmetrically define *right cancellative*, and say that M is *cancellative* if it is both left and right cancellative. All groups are clearly cancellative. A monoid M is said to be *regular* if for every $x \in M$ there exists some $y \in M$ such that $xyx = x$ and $yxy = y$. Such a y is called an *inverse* (or *pseudo-inverse*) for x. Note that it might not be the case that e.g. $xy = 1$. If a regular monoid has unique inverses, then M is said to be an *inverse* monoid. Overloading the notation $^{-1}$, we shall in inverse monoids denote the unique inverse of an element x by x^{-1}, and note that this terminology satisfies such identities as $(xy)^{-1} = y^{-1}x^{-1}$. However, this inverse need not be a group inverse. Context will always make it clear which type of inverse is discussed.

We shall now describe in more detail some aspects of the theory of a certain class of monoids, called *free* monoids. Studying free monoids from a combinatorial perspective is known as *combinatorics on words*, for reasons that will soon become apparent.

[2] There are counterintuitive results for finitely generated semigroups, however, which indicate the benefits of considering monoids rather than semigroups. The *direct product* $M \times N$ of two monoids M, N is defined as the monoid with underlying set the Cartesian product of M and N, and the natural associated multiplication. The direct product of two finitely generated groups is clearly finitely generated, and similarly one can show that a direct product of two finitely generated monoids is finitely generated. However, the direct product of two finitely generated semigroups need not be finitely generated [420].

[3] Note that any ring is a monoid with respect to multiplication; hence the study of monoid ideals contains the study of ring ideals. However, there exist (easy examples of) monoids which cannot be "extended" to become the multiplicative structure of a ring [223].

1.1.1 Combinatorics on words

Let A be a finite set. Then A is an *alphabet*. A *word* w over A is any finite sequence of elements $(a_1, a_2, \dots, a_n)$, where $a_i \in A$ for all $1 \leq i \leq n$. The empty sequence is denoted ε or 1, depending on the context, and is called the *empty word*. A sequence of length one is called a *letter*. A word $(a_1, a_2, \dots, a_n)$ will for brevity be written $a_1 a_2 \cdots a_n$. Thus we have $A \subseteq A^*$. The *free semigroup* A^+ on A consists of all non-empty words over A, together with the operation of word concatenation, i.e. the result of writing one word followed by the other. This is clearly an associative operation. The *free monoid* A^* on A consists of all words over A, together with the same operation of word concatenation. The free monoid is also, at times, denoted as $A^{<\omega}$. As sets, $A^* = A^+ \cup \{\varepsilon\}$. Equality in the free monoid or semigroup is denoted $\equiv$, and is called *graphical equality* of words.[4] We shall presently, in the theory of presentations, see why this notation is useful. The *length* $|w|$ of a word $w \equiv a_1 a_2 \cdots a_n$ is the number of letters n in w. Note that $|\varepsilon| = 0$, and it is the unique word with this property. To simplify notation, words are sometimes written using exponents: for $w \in A^*$, define $w^0 := \varepsilon$, $w^1 := w$, and $w^{n+1} := ww^n$ for $n \geq 1$. The *reverse* w^{rev} of a word $w \equiv a_1 a_2 \cdots a_n$ is simply the word $a_n a_{n-1} \cdots a_1$. Note that $\varepsilon^{\text{rev}} = \varepsilon$, and that reversal is an anti-homomorphism, i.e. $(uv)^{\text{rev}} = v^{\text{rev}} u^{\text{rev}}$ for all words $u, v \in A^*$. For a set $X \subseteq A^*$, we set $X^{\text{rev}} = \{x^{\text{rev}} \mid x \in X\}$. For a fixed total order $<_A$ of the finite set A, we define the *short-lex* order $<_s$ on the free monoid A^* as follows: for distinct $u, v \in A^*$, if $|u| < |v|$, then $u <_s v$. If $|u| = |v|$, then if $u \equiv a_1 a_2 \cdots a_k$ and $v \equiv a'_1 a'_2 \cdots a'_k$ where $a_i \in A$ for $1 \leq i \leq k$, with $k = |u| = |v|$, then let j be the leftmost index $1 \leq j \leq k$ such that a_i and a'_i differ. Then $u <_s v$ if and only if $a_i <_A a'_i$. For example, if $A = \{a, b\}$ with $a <_A b$, then $aba <_s abbab$, and $aaa <_s aba$. The shortlex order should be familiar to anyone who can compare the size of numbers in the decimal system.

The free monoid is a more natural object than the free semigroup, especially when regarded as a combinatorial object. One key reason for this is simple, and underpins much of combinatorics: much like the empty union $\bigcup \varnothing$ is $\varnothing$; like the empty sum of integers is defined to be 0; and like the empty product of integers is defined to be 1; we have that the empty product of words is defined to be ε. Hence in a free semigroup not every product is well-defined (!). Doing combinatorial semigroup theory with free semigroups is thus not entirely unlike doing combinatorics without 0. For this reason, free monoids will play a far more prominent rôle in this thesis than their semigroup counterpart.

The *free group* on A requires some more work to define. The free group on $\varnothing$ is the free monoid on $\varnothing$, i.e. the trivial group. If $A \neq \varnothing$, then let A^{-1} be a set of symbols such that (i) A^{-1} is in bijective correspondence with A via a map $^{-1}$; and (ii) $A \cap A^{-1} = \varnothing$. A word over $A \cup A^{-1}$ is called *freely reduced* (or simply *reduced*) if it does not contain a subword of the form xx^{-1} or $x^{-1}x$ for some $x \in A$. The elements of the free group on A are the set of reduced words. The *free reduction* of a word $w \in (A \cup A^{-1})^*$ is an operation which is defined recursively as follows: if w is reduced, then the free reduction of w is simply w itself. If w contains some subword of the form xx^{-1} or $x^{-1}x$ for some $x \in A$, i.e. $w \equiv w_0 x x^{-1} w_1$ or $w \equiv w_0 x^{-1} x w_1$ for some $w_0, w_1 \in (A \cup A^{-1})^*$, then the free reduction of w is defined as the free reduction of the word $w_0 w_1$. It is not hard to show that the free reduction of a word is uniquely defined (regardless of the order in which the reductions are carried out) using elementary results from rewriting systems. For example, the free reduction of the word $xyyx^{-1}xy^{-1}xx^{-1}$ is xy. The operation of the free group is *reducing concatenation*, which first concatenates the two freely

[4] Especially in older Soviet literature, graphical equality is usually denoted $\overline{\overline{}}$.

reduced words in question, and then freely reduces the resulting word. We say that a word is *cyclically reduced* if it is reduced and is not of the form xwx^{-1} or $x^{-1}wx$ for any letter $x \in A$. We shall denote by $\sigma_a(w)$ the *exponent sum* of the letter $a \in A$ in the word w (over A or over $A \cup A^{-1}$). This is defined recursively as follows: if $|w| = 0$, then $\sigma_a(w) := 0$. If $w \equiv a^{\pm 1}w'$, then $\sigma_a(w) := \sigma_a(w') \pm 1$. If $w \equiv b^{\pm 1}w'$ for some other letter b, then $\sigma_a(w) := \sigma_a(w')$. For example, $\sigma_a(aba^{-1}ab^{-1}) = 1$, and $\sigma_b(aba^{-1}ab^{-1}) = 0$.

A word $w \in A^*$ is a *prefix* of $w' \in A^*$ if there exists a word $u \in A^*$ such that $w' \equiv wu$, and it is a *proper prefix* if u is non-empty. The empty word is a proper prefix of every word except itself. We define *suffix* and *proper suffix* entirely symmetrically. The word $w \in A^*$ is called a *subword* (or *factor*) of $w' \in A^*$ if there exist $u, v \in A^*$ such that $w' \equiv uwv$, and it is called a *proper subword* if u or v is non-empty. For example, *ghm* is a subword of *arghmgog*, but *agmo* is not. Clearly every prefix and every suffix of a word is also a subword of that word. We say that a word is *self-overlap free* (sometimes also called a *bifix-free word* or *hypersimple word*) if none of its non-empty proper prefixes is also a suffix. For example, the word $ababb$ is self-overlap free, but the word $xyzabcxyz$ is not.

Let A be an alphabet. A subset $X \subseteq A^*$ is called a *code* over A if for all $n, m \geq 1$ and $x_1, \dots, x_n, x'_1, \dots, x'_m \in X$, we have that the condition

$$x_1 x_2 \cdots x_n \equiv x'_1 x'_2 \cdots x'_m$$

implies that $n = m$ and

$$x_i \equiv x'_i \quad \text{for} \quad i = 1, \dots, n.$$

In other words, a code is a subset of a free monoid which freely generates a free submonoid. In particular, a code never contains the empty word 1. Any subset of a code is clearly a code, and the empty set is a code. In fact, it is not hard to show that if $X \subseteq A^*$ with $X \not\subseteq \{1\}$, then $\langle X \rangle$ is a free monoid if and only if X is a code.[5]

For any alphabet A, the set $X = A^p$ for any $p \geq 1$ is a code, called the *uniform code* of words of length p. As in [54, Example 1.3], if $A = \{a, b\}$ and $X = \{aa, baa, ba\}$, then X is a code. It is surprisingly tricky to algorithmically check whether a finite set of words is a code; however, it is decidable, and an explicit algorithm can be found in [54, I.3].[6]

We say that a code $X \subseteq A^*$ is a *prefix code* if no word in X is a proper suffix of another element in X. Similarly, a code $X \subseteq A^*$ is a *suffix code* if no word in X is a proper prefix of another element in X. For example, if $\alpha \in A^+$ is a self-overlap free word, then it is not hard to show that $\alpha(A^* \setminus A^*\alpha A^*)$ is a suffix code (see e.g. [264, Lemma 3.4]). Clearly X is a prefix code if and only if X^{rev} is a suffix code. We say that a code X is a *biprefix code* if it is a prefix and a suffix code. The sets $X = \{ab, aabb, aaabbb\}$ and $Y = \{abba\}$ are both biprefix codes, while $Y = \{abba, bba\}$ is not, as *bba* is a suffix of $abba$. We shall primarily be interested in biprefix codes in this thesis, but the notion of prefix and suffix codes will be notationally useful when discussing compression in Chapter 4. Note that a uniform code is clearly always a biprefix code, and if X is a biprefix code, then X^p is again a biprefix code for any $p \geq 1$. We say that a code X is an *infix code* if no word from X is a proper factor of another word in X. We say that

[5] Any subgroup of a free group is itself a free group. This is a famous theorem from 1921 due to Nielsen [376] (in the finitely generated case) and 1927 due to Schreier [426] (in general). The corresponding statement is not true for submonoids of free monoids. That is, there are submonoids of free monoids which are not free (!). Thus there are submonoids of a free monoid which cannot be generated by a code.

[6] One can extend the notion of codes from subsets of free monoids to subsets of general monoids. There are, however, many easy examples even of matrix monoids in which the problem of deciding whether a finite subset is a code is undecidable, cf. §1.1.4 and [259, 103, 55].

a code X is *overlap-free* if no pair of words from X have a non-trivial overlap, i.e. no proper non-empty prefix of any word is also a proper non-empty suffix of a word, and vice versa. Thus, for example, neither $\{ab, bba\}$ nor $\{abba\}$ are overlap-free, but $\{ab, aabb\}$ is overlap-free.

1.1.2 Presentations

For all their importance in the theory of codes, free monoids and semigroups will play an even more central rôle in this thesis due to the theory of *presentations*. A *monoid presentation* is an ordered pair (A, R), where A is a set and $R \subseteq A^* \times A^*$. The set A is called the *generators* of the presentation, and the set R is called the set of *defining relations* of the presentation. A *semigroup presentation* is an ordered pair (A, R), where A is a set and $R \subseteq A^+ \times A^+$, and otherwise the terminology is identical. A *group presentation* is an ordered pair (A, R), where A is a set in bijective correspondence with a set A^{-1} such that $A \cap A^{-1} = \varnothing$, and $R \subseteq (A \cup A^{-1})^* \times (A \cup A^{-1})^*$. We will generally denote a monoid presentation as $\text{Mon}\langle A \mid R\rangle$, a semigroup presentation as $\text{Sgp}\langle A \mid R\rangle$, and a group presentation as $\text{Gp}\langle A \mid R\rangle$. We will denote by $\langle A \mid R\rangle$ an arbitrary presentation, if no distinction is needed as to whether it is a semigroup, monoid, or group presentation.

Presentations have a rich and long history, which we cannot go into at any depth here; we do, however, mention briefly the fact that W. R. Hamilton [201] in a brief note in 1856 gave the first ever presentation of a group (the alternating group A_5) in his study of his remarkable and fascinating *icosian calculus*, though he did not at all phrase it in the language of e.g. free groups.[7] Instead, this was done by von Dyck [145] in 1882, who formalised the theory of presentations (of groups), and gave presentations for the finite cyclic groups C_n and the dihedral groups D_n for all $n \geq 1$, as well as presentations for A_4, S_4, and A_5. Writing down presentations for the alternating groups A_n and the symmetric groups S_n for arbitrary $n \geq 1$ would not be done until 1896, by E. H. Moore [360].[8]

Returning to definitions, if $\langle A \mid R\rangle$ is a presentation (of any type), and if (u, v) is a defining relation of the presentation, then we will generally write that $u = v$ is a defining relation of (A, R). We may also e.g. use the following two pieces of notation interchangeably:

$$\langle A \mid u_i = v_i \ (i \in I)\rangle \quad \leftrightarrow \quad \langle A \mid \{(u_i, v_i) \mid i \in I\}\rangle,$$

when I is some indexing set for the elements of R. Let $\langle A \mid R\rangle$ be a presentation. If A and R are finite, then we say that the presentation is *finite*. If $|A| \leq \aleph_0$ and R is a recursively enumerable set, then we say that $\langle A \mid R\rangle$ is a *recursive presentation*. Before continuing with the forthcoming connection between semigroup/monoid/group presentations and semigroups/monoids/groups, we emphasise the general fact that this thesis will almost exclusively concern itself with finite presentations. We shall never consider cases when $|A| > \aleph_0$.[9]

The key part of a presentation is the natural algebraic structure which one can associate to it. Any presentation $\langle A \mid R\rangle$ induces an equivalence relation ϱ_R on the corresponding free object

[7] I thank J. Stillwell for bringing this fact to my attention.

[8] I thank J. East for bringing this fact to my attention.

[9] There are many good (?) reasons to be somewhat frightened of infinitely presented monoids and groups, and of uncountable monoids and groups in general. For example, there exists, by Shelah [442], a group of cardinality $\aleph_1$ in which every proper subgroup is countable. The existence of such groups was predicted by Kurosh [267]. Furthermore, assuming the continuum hypothesis, the group constructed by Shelah does not admit any non-trivial topology as a topological group. The unnerved reader may be assured that this thesis will not stray into questions regarding questions such as the continuum hypothesis, and will only deal with the safety of the harmless countable.

on A. First, let $\sim_R$ denote the relation

$$u \sim_R v \qquad \iff \qquad u \equiv u_0 r u_1, v \equiv u_0 s u_1 \quad \text{and} \quad (r, s) \in R \text{ or } (s, r) \in R.$$

If $u \sim_R v$, then we say that u and v are related by a single *elementary transformation*. Let ϱ_R denote the symmetric, transitive closure of the reflexive, transitive closure of $\sim_R$. Then ϱ_R is clearly a congruence on the free object on A, and is called the *congruence induced by* R. Thus, given a presentation $\langle A \mid R \rangle$ of any type, if F_A denotes the corresponding free object on A then the quotient F_A/ϱ_R has the structure of the same type as the presentation. We then say that F_A/ϱ_R is the object of that type *defined by the presentation*. For example, the free commutative monoid $\mathbb{N} \times \mathbb{N}$ is isomorphic to the monoid defined by the presentation $\text{Mon}\langle a, b \mid ab = ba \rangle$, and the free abelian group $\mathbb{Z} \times \mathbb{Z}$ is isomorphic to the group defined by the presentation $\text{Gp}\langle a, b \mid ab = ba \rangle$. If u, v are two elements of F_A and $M = F_A/\varrho_R$, then we will write $u =_M v$ (or $u = v$ in M) if $u \varrho_R v$. We will say that u *represents* the element $m \in M$ if $m = [u]_{\varrho_R}$. The homomorphism associated with the quotient F_A/ϱ_R will be denoted by $\pi_R \colon F_A \to F_A/\varrho_R$, and is called the *canonical homomorphism* associated with the presentation $\langle A \mid R \rangle$. For example, a word u represents the identity element 1 if and only if $\pi_R(u) = 1$ if and only if $u \varrho_R \varepsilon$ if and only if $u =_M 1$.

Remark 1.1.1. Although monoids and semigroups are very similar, sometimes the contrast between semigroup and monoid presentations becomes quite important. For example, the monoid defined by the presentation $\text{Mon}\langle a \mid a^2 = a \rangle$ is certainly not a group, as an easy induction proves that no non-empty word represents the identity element; indeed, this monoid has exactly two elements $\{1, a\}$, with multiplication defined by setting $1 \cdot 1 = 1$ and $1 \cdot a = a \cdot 1 = a \cdot a = a$. In particular, a is not invertible. However, the semigroup defined by the presentation $\text{Sgp}\langle a \mid a^2 = a \rangle$ has only one element a with multiplication defined by $a \cdot a = a$, and is a group with identity a.

Because of the above remark, semigroup presentations will not be used in this thesis; instead, monoid and group presentations will hold centre stage. Further to this, we shall generally not hold the distinction between a presentation and the algebraic object it defines in any high regard. For example, we may speak of "the monoid $\text{Mon}\langle A \mid R \rangle$" or ask "when does the group $G = \text{Gp}\langle A \mid R \rangle$ admit a monoid presentation with the same number of defining relations?". We shall extend the usage of terms for presentations to the object defined by them; for example, a monoid (or group) will be said to be *finitely presented* if it is isomorphic to the monoid (group) defined by a finite monoid (group) presentation.

As before, we shall almost exclusively deal with finitely presented monoids and groups throughout this thesis. There are strong connections between monoid presentations and group presentations. For example, a finitely presented group $\text{Gp}\langle A \mid R \rangle$ is also finitely presented as a monoid, by adding new generators A^{-1} and considering the monoid presentation

$$\text{Mon}\langle A \cup A^{-1} \mid R \cup \{aa^{-1} = 1, a^{-1}a = 1 \mid a \in A\} \rangle$$

which clearly defines a group isomorphic to $\text{Gp}\langle A \mid R \rangle$. Note that if $M = \text{Mon}\langle A \mid R \rangle$ defines a group, then $M = \text{Gp}\langle A \mid R \rangle$. In particular, if a group is finitely presented as a monoid, then it is also finitely presented as a group.

There are many surprising results connecting group and monoids presentations. We cannot mention them all, but we end by mentioning a short result, attributed by R. Book to C. Wrathall [68]: if $\text{Mon}\langle A \mid R \rangle$ is a free monoid, then $\text{Gp}\langle A \mid R \rangle$ is a free group.

1.1.3 Free products

Let $S_1 = \text{Sgp}\langle A_1 \mid R_1 \rangle$ and $S_2 = \text{Sgp}\langle A_2 \mid R_2 \rangle$ be two semigroups, where A_1 and A_2 are disjoint alphabets. We define the *semigroup free product* $S_1 * S_2$ of S_1 and S_2 to be the semigroup with the presentation $\text{Sgp}\langle A_1 \cup A_2 \mid R_1 \cup R_2 \rangle$. Semigroup free products may at first glance appear quite unnatural to the combinatorial group theorist. For example, the semigroup free product of two trivial semigroups (where the trivial semigroup is the unique semigroup $\text{Sgp}\langle e \mid e^2 = e \rangle$ with one element) is infinite, and has non-trivial properties, see e.g. [453]. Furthermore, note that even if S_1 and S_2 are themselves monoids, their respective identity elements will not be identity elements of $S_1 * S_2$. The following is easy and standard to prove by induction on the number of applications of defining relations, giving a "normal form lemma" for semigroup free products.

Lemma 1.1.1. *Let S_1, S_2 be two semigroups, generated by disjoint sets A_1, A_2, respectively. Let $\Pi = S_1 * S_2$ denote their semigroup free product. Then for $u, v \in (A_1 \cup A_2)^+$ we have $u =_\Pi v$ if and only if there exist unique $u_0, v_0, u_1, v_1, \dots, u_n, v_n$ such that*

(1) $u \equiv u_0 u_1 \cdots u_n$ and $v \equiv v_0 v_1 \cdots v_n$;

(2) $u_i, v_i \in A^+_{X(i)}$ and $u_i =_{S_{X(i)}} v_i$ for all $i \geq 0$,

where $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$.

Let $M_1 = \text{Mon}\langle A_1 \mid R_1 \rangle$ and $M_2 = \text{Mon}\langle A_2 \mid R_2 \rangle$ be two monoids, where A_1 and A_2 are disjoint alphabets. We define the *monoid free product* $M_1 * M_2$ of M_1 and M_2 to be the monoid with the presentation $\text{Mon}\langle A_1 \cup A_2 \mid R_1 \cup R_2 \rangle$. Thus the monoid free product identifies the identity elements of the factors; that is, $1_{M_1 * M_2} = 1_{M_1} = 1_{M_2}$, and we denote this identity element simply as 1. In particular, the free product of two trivial monoids is itself trivial. Let $u \equiv u_0 u_1 \cdots u_n$ be a word in $(A_1 \cup A_2)^*$ such that $u_i \in A^*_{X(i)}$ for all $0 \leq i \leq n$, where $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$. Clearly every word in $(A_1 \cup A_2)^*$ can be written in this form. We say that u is *reduced* if it is empty, or else $u_i \neq_{M_{X(i)}} 1$ for all $0 \leq i \leq n$. In particular, a word $u \equiv u_0 u_1 \cdots u_n$ has at least one factor u_i empty if and only if $u \equiv \varepsilon$.

Proposition 1.1.2. *Let M_1, M_2 be two monoids generated by disjoint sets A_1, A_2, respectively, and let $M_1 * M_2$ denote their monoid free product. Then for every word $u \in (A_1 \cup A_2)^*$ there exists $u' \in (A_1 \cup A_2)^*$ such that $u' =_{M_1 * M_2} u$, and u' is reduced.*

Proof. To prove this proposition, we will assume temporarily that the reader is familiar with the language and terminology of rewriting systems, see §1.2.2. Suppose u is non-empty, for otherwise there is nothing to show. Write $u \equiv u_0 u_1 \cdots u_n$ uniquely, where the u_i are from alternating factors as above. Let $I_{M_i} = \{(w \to 1) \mid w \in \text{IP}^{M_i}_{A_i}, |w| > 0\}$ for $i = 1, 2$. The rewriting system $I = I_{M_1} \cup I_{M_2}$, a subset of $(A_1 \cup A_2)^* \times (A_1 \cup A_2)^*$, is terminating (though certainly not confluent in general) and it is clear that any word that is I-irreducible is reduced in the above sense. Furthermore, I is $(M_1 * M_2)$-equivariant, as any word equal to 1 in a factor of a monoid free product is also equal to 1 in the product. Thus u' can be taken as any irreducible descendant of u under I, and by $M_1 * M_2$-equivariance we have $u' =_{M_1 * M_2} u$. □

We say that any u' as in the statement of Proposition 1.1.2 is a *reduced form* of u. Certainly, this is not unique, in the sense that there may be many different words representing the same word. However, there is one important form of uniqueness for reduced forms. Given a reduced word w', we can uniquely factor it as $w'_0w'_1 \cdots w'_n$, where the w'_i come alternatingly from A_1^* and A_2^*, and none of the w'_i (with $w'_i \in A_j^*$, say) are such that $w'_i =_{M_j} 1$ unless $w' \equiv \varepsilon$, in which case all $w' \equiv w'_0 \equiv \varepsilon$. In particular, none of the w'_i are empty unless all of them are. We call the factorisation $w'_0w'_1 \cdots w'_n$ the *normal form* of the reduced word w'. This normal form of a fixed reduced form is unique. Furthermore, reduced forms are unique up to equality of the representatives chosen for the alternating factors in their normal forms, in the following sense.

Lemma 1.1.3. *Let M_1, M_2 be two finitely generated monoids, generated by finite disjoint sets A_1, A_2, respectively. Let $\Pi = M_1 * M_2$ denote their monoid free product. Let $u, v \in \{A_1, A_2\}^*$. Let u' resp. v' be reduced forms of u resp. v, with normal forms*

$$u' \equiv u_0u_1 \cdots u_m \qquad \text{resp. } v' \equiv v_0v_1 \cdots v_n.$$

Then we have $u =_\Pi v$ if and only if

(1) $n = m$, and

(2) $u_i, v_i \in A^*_{X(i)}$ and $u_i =_{M_{X(i)}} v_i$ for all $0 \leq i \leq n$,

where $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$.

This lemma can be proved using e.g. van der Waerden's trick, see [483]. From the above lemma, one immediately deduces that the word problem in a monoid free product is decidable if and only if the word problem is decidable in each of the factors. The language-theoretic aspects of the word problem of a (semigroup or monoid) free product has not been extensively studied, however; we shall revisit this theme in Chapter 2, and prove some rather general theorems. We remark, finally, that we have adopted the excellent means of parametrisation in a free product as above, i.e. using a function $X(j)$, from Brough, Cain & Pfeiffer [80]. We shall use this notion in multiple places throughout this thesis, notably in Chapter 2.

1.1.4 Decision problems

While the theory of decision problems – and formally defining recursive and recursively enumerable sets – is vast, we shall limit our scope in this thesis. We shall instead follow the philosophy as presented by Lyndon & Schupp [298, p. 217]. They write: "many definitions [of recursivity] have been given: Turing machines, formal systems, λ-computability, etc. All the definitions proposed have been shown to be equivalent. The equivalence of these notions, which are formally quite different, has led logicians to the belief that the precisely defined concept of being recursive is an adequate formalisation of the intuitive notion of 'effective'. This philosophical position is called the "Church-Turing thesis". In this thesis we shall accept this philosophical position. However, doing so does not come without caveats: this position is not a rigorous one.[10] That is, this position is, in the words of Post [403, p. 105], in reality not much more than a *working hypothesis*; and, quoting Post further, "to mask this identification under a definition hides the fact that a fundamental discovery in the limitations of the mathematicizing power of Homo Sapiens has been made and blinds us to the need of its continual verification". This assumption is left plainly written for the benefit of any future Homo Sapiens (or otherwise), who will perhaps know more about the verification of which

[10] For discussions about the alternate view that the Church-Turing thesis *is* susceptible to being rigorously proved or disproved (a view famously held by Gödel), the reader may consult e.g. [272, 343, 455, 456, 155, 132, 266, 441, 204, 193].

Post speaks.[11] We implore the reader to consult Uspenskii & Semenov [481] and the remarks by Wittgenstein on how "Turing machines are *humans* who calculate" [495, §1096]; see [440].

Philosophical remarks aside, we shall not give a formal definition of the decision problems in terms of Turing machines (or equivalents). Instead, the important data for each problem which could be, if one would be so inclined, encoded into some such form, will for each relevant decision problem be presented in the following order: (1) What is the Input? (2) What is the Question? (3) What is the Output? These data will always be finite. For example, the input could take the form of a finite monoid presentation and a word over some finite alphabet; the question could be whether the word represents the identity element of the monoid defined by the monoid presentation; and the output could simply be the answer to this question. For all decision problems presented in this section, the output will simply be the answer to the question, and the output could take the form of a yes or a no.[12] As before, a presentation written as $\langle A \mid R \rangle$ rather than e.g. $\mathrm{Mon}\langle A \mid R \rangle$ below indicates that the problem is defined in the same way for all types of presentations; the notation F_A will in this case signify the free object of the corresponding type. The presentation is **not** implicitly taken to be part of the input. That is, on the next page of problems presented, we are not implicitly considering the *uniform* variants of the problems.

Before presenting our list of decision problems, there are two undefined terms to explain. First, a *rational subset* of a monoid is defined in §1.2. Second, an *equation* ϕ over a monoid $M = \mathrm{Mon}\langle A \mid R \rangle$ is defined as follows. Let Ω be a set of *variables*, disjoint from A. A *word equation over* M is a pair $U = V$, where $U, V \in (A \cup \Omega)^*$. A *system* of word equations is a finite set of word equations. A *solution* to a system of word equations $\{U_i = V_i \mid i \in I\}$ over M is given by a homomorphism $\sigma\colon (A \cup \Omega)^* \to M$ such that $\sigma(U_i) =_M \sigma(V_i)$ for all $i \in I$ and which fixes A. For example, the equation $xaa = aby$ over $M = \mathrm{Mon}\langle a, b \mid ab = ba \rangle$ has a solution given by (for example) $\sigma(x) = ba$ and $\sigma(y) = aa$. Equations over groups and semigroups are defined analogously. We shall not study equations in any depth in this thesis.

[11] There are recent and serious objections to the Church-Turing thesis coming from the idea of *supertasks* and *hypercomputations*. For example, Hogarth [210] shows that there are ways of computing non-Turing computable problems by using special relativity. Consider the following situation. A mathematician (say, Bob) is interested in a problem which is not decidable by a Turing machine – say, the halting problem – and leaves his poor graduate student to begin working on the problem. Bob then escapes in a rocket, begins orbiting Earth, and proceeds to accelerate closer to the speed of light (though never exceeding it) while his graduate student continues slavishly working on the problem, having promised to signal to Bob when the work is complete (which of course will take infinitely long). Hogarth shows that there is a way for Bob to accelerate to allow for infinite time to pass in the frame of reference of Bob's student, while Bob himself only experiences finitely much time passing in his own frame of reference. Thus, Bob will hear of the results in a finite amount of time and return to Earth, having thus solved the halting problem in finite time. There are similar solutions occurring in general relativity in orbits around rotating charged black holes, resulting in a phenomenon known as Malament–Hogarth spacetime (see especially [152]). The realities of the physical world which make such solutions impossible (e.g. the eventual heat-death of the universe, or running out of rocket fuel) are no greater obstructions than the impossibility of actually constructing a Turing machine – this requires arbitrarily large memory storage capabilities in the form of a tape which can be made arbitrarily large, another physical impossibility. Furthermore, while the explicit reference to a physical theory in the spacetime solution may at first appears as obvious grounds for dismissal, the same objection would dismiss Turing machines, as Turing machines cannot exist in e.g. a finite universe. Nevertheless, we shall remain grounded and leave the reader interested in tormenting their own graduate students to further reading on non-Turing computability in e.g. [211, 202, 294, 123, 487, 329, 330].

[12] Some authors require more. For example, an appropriate output corresponding to the question "does there exist a word X such that Y" might be yes, along with a word X such that Y. As all our presentations are finite, we are not concerned with finding such witnesses, as for all problems presented here, when the answer is yes, there is a procedure for verifying this.

The
Word Problem
for $M = \langle A \mid R \rangle$
–

Input : Two words $u, v \in F_A$.

Question : Is $u = v$ in M?

The
Conjugacy Problem
for $G = \mathrm{Gp}\langle A \mid R \rangle$
–

Input : Two words $u, v \in F_A$.

Question : $\exists x \in F_A$, $xux^{-1} = v$ in G?

The
Identity Problem
for $M = \langle A \mid R \rangle$
–

Input : A word $w \in F_A$.

Question : Is $w = 1$ in M?

The
Group Problem
for $M = \mathrm{Mon}\langle A \mid R \rangle$
–

Input : $\mathrm{Mon}\langle A \mid R \rangle$.

Question : Is M a group?

The
Left Divisibility Problem
for $M = \mathrm{Mon}\langle A \mid R \rangle$
–

Input : Words $u, v \in F_A$.

Question : $\exists w \in F_A$ s.t. $u = vw$?

The
Right Divisibility Problem
for $M = \mathrm{Mon}\langle A \mid R \rangle$
–

Input : Words $u, v \in F_A$.

Question : $\exists w \in F_A$ s.t. $u = wv$?

The
Subgroup Membership Problem
for $G = \mathrm{Gp}\langle A \mid R \rangle$
–

Input : Words $u_1, \dots, u_k, w \in F_A$.

Question : Is $\pi_R(w) \in \langle u_1, \dots, u_k \rangle_{\mathrm{gp}}$?

The
Submonoid Membership Problem
for $M = \langle A \mid R \rangle$
–

Input : Words $u_1, \dots, u_k, w \in F_A$.

Question : Is $\pi_R(w) \in \langle u_1, \dots, u_k \rangle$?

The
Rational Subset Membership Problem
for $M = \langle A \mid R \rangle$
–

Input : $X \in \mathrm{Rat}(F_A)$, $w \in F_A$.

Question : Is $\pi_R(w) \in \pi_R(X)$?

The
Diophantine Problem
for $M = \langle A \mid R \rangle$
–

Input : Equations $\{\phi_i\}_{i=1}^{n}$ over M.

Question : $\exists$ solutions for $\{\phi_i\}_i$ in M?

	Groups	**Special Monoids**	**Monoids**
	Gp$\langle A \mid r_i = 1 \ (i \in I)\rangle$	Mon$\langle A \mid r_i = 1 \ (i \in I)\rangle$	Mon$\langle A \mid u_i = v_i \ (i \in I)\rangle$
1	[301]	[4]	?
2	?	?	?
3	?	?	[337]
4	?	?	🟥
5	?	?	[309]
6	?	?	🟥
7	?	?	[477]
8	?	?	🟥
9	?	?	🟥
10	?	?	🟥
11	?	?	🟥
12	[73]	?	🟥
13	🟥	[73]	🟥

Table 1.1: A table showing how many defining relations suffices for undecidability of the word problem in groups, special monoids, resp. monoids. Green, red, and blue each indicates decidability, undecidability, resp. unknown. Thus, for example, it is known that there exists a 12-relator finitely presented group with undecidable word problem, the word problem for 1-relation *special* monoids (such monoids shall be presented in §1.3) is decidable, and the word problem for 1-relation monoids is (famously) open, see [388].

Most decision problems, including those just listed, are undecidable for finitely presented objects in general – this is the harsh reality of combinatorial (semi)group theory. Indeed, results such as Markov's theorem [333, 334] and the Adian–Rabin theorem [2, 413] demonstrate that "most" semigroup resp. group-theoretic properties cannot be algorithmically recognised from a given presentation. For example, it is not decidable whether a given finitely presented semigroup is commutative; or whether a given finitely presented group is trivial.

One of the few decidable properties of finitely presented groups is the rank of its abelianisation. That is, given a presentation $G = \text{Gp}\langle A \mid R\rangle$, one can compute its first homology group $H_1(G, \mathbb{Z}) \cong G/[G,G]$. The computation is simple: one simply adds all commutators of generators, and then reduces the relations to a standard presentation an easy analogue of the Smith Normal Form (see e.g. [306, §3.3, p.140]. This fundamental observation was made already by Poincaré (using more primitive language), and can at times be used to conclude that a group is infinite. For example, the group

$$G = \text{Gp}\langle a, b, c \mid a^{-1}ba = c^2, a^{-1}ca = c, c^{-1}a^3c = 1\rangle \quad \text{yields}$$

$$\begin{aligned} H_1(G,\mathbb{Z}) &\cong \text{Gp}\langle a, b, c \mid a^{-1}ba = c^2, a^{-1}ca = c, c^{-1}a^3c = 1, [a,b] = [b,c] = [a,c] = 1\rangle \\ &\cong \text{Gp}\langle a, b, c \mid b = c^2, c = c, a^3 = 1, [a,b] = [b,c] = [a,c] = 1\rangle \\ &\cong \text{Gp}\langle a, c \mid a^3 = 1, [a,c] = 1\rangle \cong \mathbb{Z} \times C_3. \end{aligned}$$

Hence G is infinite. Obviously, there are finitely presented infinite groups with finite abelianisation, and hence this technique cannot be used to decide in general whether a given group is infinite or not – this latter problem is even undecidable by the Adian-Rabin theorem

[2, 413]. It is also known that computing the *second* homology group $H_2(G, \mathbb{Z})$ (also known as the *Schur multiplier*) is, in general, an undecidable problem [164]. Thus the decidability of computing the abelianisation of a group appears a lonely beacon in a vast sea of undecidability.[13]

We finish this introduction to the decision problems of this thesis with a short discussion as to the pedigree and idea history of them. All problems presented in the list on the earlier page are decidable for free groups and monoids, and almost always trivially so.[14] This is a common theme, and highlights the combinatorial nature of such objects and, perhaps more importantly, the combinatorial nature of the decision problems we consider herein. Indeed, problems such as the membership problem grew out of Nielsen's new proof of the Nielsen-Schreier theorem, from which decidability of the subgroup membership problem in free groups falls out [377].

We wish to emphasise, which is not often done in the literature, the importance of membership problems on the development of combinatorial (semi)group theory. Membership problems can be tracked back to some of the earliest undecidability results in all of combinatorial algebra; Markov [332] essentially proved, in modern language, that the submonoid membership problem is undecidable in $\mathrm{SL}_4(\mathbb{Z})$ in the same year he proved that the word problem is undecidable in finitely presented semigroups. The first formal definition of the membership problem would not appear until the work by K. A. Mikhailova [345, 346, 347, 348], who defined the problem (which she called the *occurrence problem*) and proved by an ingenious construction that the subgroup (and hence also submonoid) membership problem is undecidable in the direct product $F_2 \times F_2$ of two free groups. Membership problems have also recently proved themselves to be key in understanding the word problem of certain inverse monoids, see e.g [230, 141], as well as having many fascinating cryptographic applications [499].

[13] However, abelianisation is a rather limited tool. For example, abelianisation cannot be used to tell knot groups apart; it was this difficulty that prompted Reidemeister [416] to invent his celebrated method, today known as the Reidemeister-Schreier method, which is an indispensable tool in the kit of a combinatorial group theorist.

[14] The sole exception to this is the Diophantine problem, which is exceptionally difficult (by comparison to other problems) in free groups and monoids. Makanin solved both these problems in two separate and exceedingly intricate papers [310, 311]. Especially the case of free groups remains an active research area (see [287, 126, 136]). See also [326, 327, 315, 316, 317, 318, 1, 319, 312, 313, 314, 328].

1.2 Formal language theory

In this section, we shall give another perspective on combinatorics on words, and give an overview of the various notions from *formal language theory* that will be required in the following chapters. The definitions given herein are all standard, and can be found, with minor modifications, in usual textbooks on the subject, e.g. [203, 217]. We will begin by describing the notion of a *class of languages*, which shall be of central importance in Chapters 2, 3, and 4. We then define and discuss *rewriting systems*, which are used throughout this thesis. Finally, we discuss how the *word problem* of a monoid or a group can be encoded into a formal language.

1.2.1 Classes of languages

We shall formalise the notion of *alphabets* and *languages* presently. To do this, we will extend the definition given earlier, in §1.1.1. We note, however, that this formalism is, in a sense, a proper extension of the former usage; that is, the two usages are consistent relative to one another, and this latter extension only serves to make some aspects of the former slightly more rigorous. We begin by fixing a countably infinite set A_ω, which will be called our *universal set of symbols*. Any commonly used symbol or English letter (such as $a, b, c, \dots, \#, a_1, a_2, \dots$) will be assumed to be in A_ω. In particular, every finite generating set of symbols encountered throughout this thesis appears as a subset of A_ω. Let $\mathrm{Fin}(A_\omega)$ denote the finite subsets of A_ω. We say that A is an *alphabet* if $A \in \mathrm{Fin}(A_\omega)$.

For an alphabet A, a *language over* A is any subset of A^*.[15] More generally, a *language* is any subset of A_ω^* such that there exists an alphabet $A \in \mathrm{Fin}(A_\omega)$ with $L \subseteq A^*$. An element of a language L is called a *word*. Given a language $L \subset A_\omega^*$, let $\mathrm{Alp}(L)$ denote the minimal (with respect to inclusion) element of $\mathrm{Fin}(A_\omega)$ such that $L \subseteq \mathrm{Alp}(L)^*$. We say that $\mathrm{Alp}(L)$ is the (minimal) *alphabet of* the language L. That is, $\mathrm{Alp}(L)$ is the set of symbols which occur in at least one word from L. If $L_1 \subseteq A^*$ and $L_2 \subseteq B^*$ are two languages, then we shall denote by L_1L_2 their concatenation in the free monoid $(A \cup B)^*$, into which we naturally embed A^* and B^*. In particular L_1L_2 is a well-defined language; as is $L_1 \cup L_2$ and $L_1 \cap L_2$.

A *context-free grammar* Γ consists of three finite sets X, A, R and a symbol $x_0 \in X$. The set X is called the set of *non-terminal symbols*; the set A is the set of *terminal symbols*; the set $R \subseteq X \times (X \cup A)^*$ is called the set of *productions*; and the symbol $x_0 \in X$ is called the *start* symbol of Γ. Thus the set of productions consists of pairs (x, w), usually written $x \to w$, where $x \in X$ and w is some word containing a mix of terminal and non-terminal symbols. For two words $u, v \in (X \cup A)^*$, we will write $u \Rightarrow v$ if there are: a rule $(\alpha \to \beta) \in R$ and words $u_1, u_2 \in (X \cup A)^*$ such that $u \equiv u_1\alpha u_2$ and $v \equiv u_1\beta u_2$. We define $\Rightarrow^*$ to be the reflexive transitive closure of $\Rightarrow$. For a context-free grammar Γ, we define the *language* of Γ to be

$$\mathcal{L}(\Gamma) = \{w \in A^* \mid x_0 \Rightarrow^* w\}.$$

A language $L \subseteq A^*$ is *context-free* if there exists a context-free grammar Γ such that $L = \mathcal{L}(\Gamma)$.

[15] Is it philosophically defensible to call such a simple object as a subset of A^* by such a complex term as *language*? Is the set $\{aghfj, \#ag, R\S g\}$ really a language? We assert that it is, on the basis that *language*, as used in the usual sense, is not a particularly well-defined notion. Indeed, quoting Wittgenstein, "man possesses the ability to construct languages capable of expressing every sense, without having any idea how each word has meaning or what its meaning is" [494, §4.002]. We find it easier to justify a broadly inclusive definition of *language* to mean something understood, rather than unjustifiably restrict the meaning of *language* to something not understood (or even not understandable).

A *class of languages* is a set $\mathcal{C} \subseteq 2^{A_\omega^*}$ of non-empty languages. Note that every language is a countable set, but not every class of languages is countable. The class $\mathcal{C}_{\text{reg}}$ of *regular languages* is the smallest subset of $2^{A_\omega^*}$ containing $\varnothing$, the singleton languages $\{a \mid a \in A_\omega\}$, and such that if $A, B \in \mathcal{C}_{\text{reg}}$, then (1) $A^* \in \mathcal{C}_{\text{reg}}$; (2) $A \cup B \in \mathcal{C}_{\text{reg}}$; (3) $AB \in \mathcal{C}_{\text{reg}}$. The class $\mathcal{C}_{\text{cf}}$ denotes the class of all context-free languages.[16] We use the notation $\mathcal{C}_{\text{rec}}$ and $\mathcal{C}_{\text{en}}$ for the class of *recursive* resp. *recursively enumerable* languages. The regular languages are sometimes called the *rational languages*. More generally, given a finitely generated monoid M, the class of *rational subsets* of M is the least class $\text{Rat}(M)$ of subsets of M containing $\varnothing$, the singleton sets, and such that if $A, B \in \text{Rat}(M)$, then (1) $A^* \in \text{Rat}(M)$; (2) $A \cup B \in \text{Rat}(M)$; (3) $AB \in \text{Rat}(M)$. Recall that $A^* = \langle A \rangle \leq M$. This definition coincides with the definition of regular languages when M is a free monoid. An alternative characterisation (see [148]) of rational subsets will be useful: let A be a finite generating set of M, and let $\pi \colon A^* \to M$ be the associated homomorphism. Then K is a rational subset of M if and only if there exists a regular language $L \subseteq A^*$ such that $K = \pi(L)$.

Example 1.2.1. Let a be a symbol and let D be the language

$$\{w \mid w \in \{a, a^{-1}\}^*, \sigma_a(w) = 0, \text{ and } \sigma_a(p) \geq 0 \text{ for every prefix } p \text{ of } w\}.$$

This is called the *Dyck language*, first studied by Chomsky & Schützenberger [109][17]. Then $D \notin \mathcal{C}_{\text{reg}}$, but $D \in \mathcal{C}_{\text{cf}}$. This is easy to prove once one is equipped with some closure properties of these classes of languages (see the following page); indeed, if D were regular, then $D \cap a^*(a^{-1})^*$ would be regular, for $\mathcal{C}_{\text{reg}}$ is closed under intersection with regular languages, but this language is $\{a^n(a^{-1})^n \mid n \geq 0\}$, a language which is straightforward to show not be regular. The Dyck language has many connections with combinatorics (via Catalan numbers, see e.g. [133]) and algebra. For example, Jantzen [235] calls certain rewriting systems (see §1.2.2) *Dyck systems* due to their similarity with the rewriting systems associated to free groups; these systems have strong algebraic properties, see [115]. △

Example 1.2.2. Let $G = \text{Gp}\langle A \mid R \rangle$ be finitely presented. Then the kernel of the natural homomorphism $\varrho \colon F_A \to G$ is a recursively enumerable language, i.e. lies in $\mathcal{C}_{\text{rec}}$. This was first formally observed by Maltsev [324], and the properties were more carefully investigated by Anīsīmov [17]. He proved that the kernel lies in $\mathcal{C}_{\text{reg}}$ if and only if G is finite. He also investigated the class of groups such that the kernel lies in $\mathcal{C}_{\text{cf}}$, showed that this is independent of choice of finite generating set and closed under free products, and furthermore that it does not contain $\mathbb{Z}^k$. Muller & Schupp [362] then completely characterised this class: the kernel lies in $\mathcal{C}_{\text{cf}}$ if and only if the group has a finitely generated free subgroup of finite index. We will revisit this latter theorem, and extend it to broad classes of monoids, in Chapter 3. △

The class $\mathcal{C}_{\text{ind}}$ of *indexed* languages is technical to define, and we shall only ever refer to closure properties of this class, rather than to any aspect of its definition whatsoever. For this reason, we refer the reader interested in the definition to [15].

[16] The class of context-free languages will occupy a special place in this thesis. We note that the class of context-*sensitive* languages, on the other hand, while capable of encoding more complicated structures (e.g. the prime numbers, see Brodda [78]), are less directly linked with the algebraic.

[17] One might ask why this language is called the *Dyck* language. A reasonable guess might be because of its connection with free groups, given that the obvious generalisation to more symbols than just one can be used to describe the kernel of the map from the free monoid on $A \cup A^{-1}$ to the free group on A. Famously, von Dyck (at the time only Dyck, not yet ennobled) was the first to properly study free groups in his 1882 and 1883 papers [145, 146]. However, von Dyck did not consider free monoids. I once asked Noam Chomsky what the original reasoning regarding his and M.-P. Schützenberger's naming of the Dyck language was. His response was "I wish I could help, but I have no idea".

An arbitrary class of languages will in general not enjoy many interesting properties.[18] We introduce some closure properties to alleviate this. We say that a class $\mathcal{C}$ is *closed under*

- *union*, if for all $L_1, L_2 \in \mathcal{C}$ we have $L_1 \cup L_2 \in \mathcal{C}$.
- *concatenation*, if for all $L_1, L_2 \in \mathcal{C}$ we have $L_1 L_2 \in \mathcal{C}$.
- *Kleene star*, if for all $L_1 \in \mathcal{C}$ we have $L_1^* \in \mathcal{C}$.
- *homomorphism*, if for all $L_1 \in \mathcal{C}$ and homomorphisms
$$\phi\colon \operatorname{Alp}(L_1)^* \to B^*,$$
where B is some alphabet, we have $\phi(L_1) \in \mathcal{C}$.
- *inverse homomorphism*, if for all $L_2 \in \mathcal{C}$ and homomorphisms
$$\phi\colon B^* \to \operatorname{Alp}(L_2)^*,$$
where B is some alphabet, we $\phi^{-1}(L_2) \in \mathcal{C}$.
- *intersection with regular languages*, if for all $L \in \mathcal{C}$ and all regular languages $R \subseteq \operatorname{Alp}(L)^*$ we have $L \cap R \in \mathcal{C}$.
- *reversal*, if for all $L \in \mathcal{C}$, we have $L^{\mathrm{rev}} \in \mathcal{C}$.

Our favourite classes of languages $\mathcal{C}_{\mathrm{reg}}$ and $\mathcal{C}_{\mathrm{cf}}$ are closed under all the above operations; this can be found proved in e.g. [217]. Note, however, that unlike $\mathcal{C}_{\mathrm{reg}}$ the class $\mathcal{C}_{\mathrm{cf}}$ is *not* closed under intersections (and hence also not under complementation). One way to prove this is showing that *every* recursively enumerable language is the homomorphic image of some intersection of two context-free languages [163].

A *rational transduction* ϱ from one finitely generated free monoid A^* to another B^* is a rational binary relation, i.e. a rational subset of the monoid $A^* \times B^*$. The *image* $\varrho(L)$ of a language $L \subseteq A^*$ under ϱ is the set $\{b \mid \exists a \in L\colon (a,b) \in \varrho\} \subseteq B^*$. For $w \in A^*$ we simply write $\varrho(w)$ rather than $\varrho(\{w\})$. We refer the reader to [144, §2.3] for further details; in particular, the union or product of rational transductions is again a rational transduction, and if ϱ is a rational transduction, then $\varrho^* := \langle \varrho \rangle$ is, too. In fact, one can show that a class of languages is closed under taking rational transductions if and only if it is closed under homomorphism, inverse homomorphism, and intersection with regular languages [217].

Example 1.2.3. Let $\phi\colon A^* \to B^*$ be any homomorphism for any alphabets A, B. We claim that (the graph of) ϕ is a rational transduction. Indeed, we have
$$\phi = \left(\bigcup_{a \in A} (a, \phi(a)) \right)^* \subseteq A^* \times B^*.$$
Thus ϕ is a rational subset of $A^* \times B^*$, as it is of the form X^* for a finite set $X \subseteq A^* \times B^*$. △

Example 1.2.4. Let A be a finite alphabet, and let # be a symbol disjoint from A. We consider words in $A^* \# A^*$. Let us say we want a rational transduction ϱ which when applied to an arbitrary word in $A^* \# A^*$ returns (the set containing) whichever word is to the left of the #-symbol. More specifically, we want to have, for any language $L \subseteq A^* \# A^*$, that
$$\varrho(L) = \{u \mid \exists v \in A^*\colon u \# v \in L\}. \tag{1.2.1}$$
This can certainly not be performed by a simple homomorphism. Instead, we employ a rational transduction. Our rational transduction will be a rational subset of the monoid
$$(A \cup \{\#\})^* \times A^*.$$

[18] Let $\mathcal{P}$ be a property of classes of languages. If almost every class of languages would satisfy $\mathcal{P}$, then $\mathcal{P}$ cannot be interesting, much as the property of "having two eyes", which almost every human satisfies, is rather uninteresting. Thus, if $\mathcal{P}$ is an interesting property, then almost every class of languages does not satisfy $\mathcal{P}$; so an arbitrary class of languages will almost certainly not satisfy any interesting property.

We define ϱ to be the relation

$$\varrho = \big(\bigcup_{a \in A} (a, a) \big)^* \cdot (\#, \varepsilon) \cdot \big(\bigcup_{a \in A} (a, \varepsilon) \big)^* \subseteq (A \cup \{\#\})^* \times A^*.$$

From this description, it is clear that ϱ is a rational subset of $(A \cup \{\#\})^* \times A^*$. Indeed, it is of the form $X^* Y Z^*$, where X, Y, Z are finite subsets of $(A \cup \{\#\})^* \times A^*$. But we clearly also have

$$\varrho = \{(u, u) \mid u \in A^*\}(\#, \varepsilon)\{(v, \varepsilon) \mid v \in A^*\} = \{(u\#v, u) \mid u, v \in A^*\},$$

and so in particular $\varrho(u\#v) = \{u\}$ for any $u, v \in A^*$. Thus (1.2.1) is satisfied, and we have our desired rational transduction. △

Most of our theorems will be stated for a special type of classes of languages. Such classes of languages are called *super*-AFL*s*, and were introduced by Greibach [180]. We mention, before giving any definitions, that the class of context-free languages (see [180]) and the class of *indexed* languages (see [150]) are super-AFLs. The reader who is only interested in such classes of languages may therefore substitute either of these classes whenever the word "super-AFL" appears in the sequel, if they wish.

We follow Book, Jantzen & Wrathall [69] in the following definitions. Let A be an alphabet. For each $a \in A$, let $\sigma(a)$ be a language (over whichever alphabet one desires); for every $x, y \in A^*$ let $\sigma(xy) = \sigma(x)\sigma(y)$; and for every $L \subseteq A^*$, let $\sigma(L) = \bigcup_{w \in L} \sigma(w)$. We then say that σ is a *substitution.* For a class $\mathcal{C}$ of languages, if for every $a \in A$ we have $\sigma(a) \in \mathcal{C}$, then we say that σ is a $\mathcal{C}$*-substitution.*

Let A be an alphabet, and σ a substitution on A. For every $a \in A$, let A_a denote $\operatorname{Alp}(\sigma(a))$, i.e. the smallest finite alphabet such that $\sigma(a) \subseteq A_a^*$. Extend σ to $A \cup (\bigcup_{a \in A} A_a)$ by defining $\sigma(b) = \{b\}$ whenever $b \in (\bigcup_{a \in A} A_a) \setminus A$. For $L \subseteq A^*$, let $\sigma^1(L) = \sigma(L)$, and let $\sigma^{n+1}(L) = \sigma(\sigma^n(L))$ for $n \geq 1$. Let $\sigma^\infty(L) = \bigcup_{n \geq 0} \sigma^n(L)$. Then we say that σ^∞ is an *iterated substitution.* If for every $b \in A \cup (\bigcup_a A_a)$ we have $b \in \sigma(b)$, then we say that σ^∞ is a *nested iterated substitution.* We say that $\mathcal{C}$ is closed under nested iterated substitution if for every $\mathcal{C}$-substitution σ and every $L \in \mathcal{C}$, we have that if σ^∞ is a nested iterated substitution then $\sigma^\infty(L) \in \mathcal{C}$.

Definition 1.2.5. Let $\mathcal{C}$ be a class of languages. We say that $\mathcal{C}$ is a *super*-AFL if it is an AFL (i.e. it is closed under homomorphism, inverse homomorphism, intersection with regular languages, union, concatenation, and the Kleene star) and if it closed under nested iterated substitution.

Some examples are given below.

Example 1.2.6. We give some examples and non-examples of super-AFLs.

(1) The class of context-free languages $\mathcal{C}_{\text{cf}}$ is a super-AFL; it is well known to be an AFL, and is closed under nested iterated substitution (see [265], [69, Theorem 2.2]).

(2) The class $\mathcal{C}_{\text{ind}}$ of indexed languages is a super-AFL; it is an AFL, as proved by Aho [14], and is closed under nested iterated substitution [150].

(3) The class $\mathcal{C}_{\text{reg}}$ of regular languages is *not* a super-AFL, e.g. by Lemma 1.2.7 below. △

For more examples and generalisations, we refer the reader to the so-called *hyper*-AFL*s* defined by Engelfriet [150], all of which are super-AFLs. We mention a useful property about super-AFLs.

Lemma 1.2.7 ([180, Theorem 2.2])**.** *Let* $\mathcal{C}$ *be a super*-AFL*. Then* $\mathcal{C}_{\text{cf}} \subseteq \mathcal{C}$*.*

We shall revisit super-AFLs in Chapter 2, and therein prove a characterisation of such languages based entirely on *rewriting systems* (rather than substitutions). Throughout Chapters 3 & 4, we will extensively make us of super-AFLs using this characterisation. Before we can do any of this, however, we need to define *rewriting systems*, and give a brief overview of some general results in this area.

1.2.2 Rewriting systems

The theory of rewriting systems gives nuance to the theory of presentations, and reveals some of their combinatorial nature. We now give the basic definitions regarding these systems. We refer the reader to the monographs by Jantzen [235] and Book & Otto [71] for a more thorough background on this topic, including its history. A *rewriting system* T (also called a *semi-Thue system*, named after the Norwegian mathematician A. Thue[19] [475]) on an alphabet A is a subset of $A^* \times A^*$. The elements of a rewriting system are called *rules*, and $(\ell, r) \in T$ will often be written $\ell \to r$.

At this point, before proceeding with definitions, we stop to address an issue of terminology. S. I. Adian [10, p. 14] writes: "at the end of the previous century, some authors started to call Thue systems 'Word Rewriting Systems' (shortly, WRS-systems). More recently, some other authors working in computer science, are attempting to change this to 'SRS-systems' because, for some unknown reason, they are now using (without any comment) the term 'string' instead of the classical term 'word.' We do not think that this practice is reasonable." We cannot agree with Adian in this dismissal of the term *string*. The term *string* for denoting a formal sequence of symbols, analogous to the term *word*, is present already in the 1918 book on symbolic logic by C. I. Lewis [279][20], published only four years after Thue's paper; furthermore, Thue himself (see [475, 407]) considered transformations not of *words* but rather of *Zeichenreihen* (Ger. *character-row* or *character-string*). Furthermore, E. Post – who certainly can be considered part of the "classical" school of thought! – uses the term *string*, rather than *word*, already in 1943 ([404, p. 197], see also [405]), four years before the publication of his proof of the undecidability of the word problem for finitely presented semigroups [406] (which also favours *string*). We could continue, but we think the point is clear: while we favour *word* over *string* in this thesis, there is nothing at all unreasonable about either using the word *string* or the string *word*.

A rewriting system T induces a relation $\to_T$ on A^* as follows: if $u, v \in A^*$, then $u \to_T v$ if and only if there exist $x, y \in A^*$ and some rule $(\ell, r) \in T$ such that $u \equiv x\ell y$ and $v \equiv xry$. The reflexive and transitive closure of $\to_T$ is denoted $\xrightarrow{*}_T$. We write $u \to_T^n v$ if there exists a sequence $u_0, u_1, \cdots, u_n \in A^*$ such that

$$u \equiv u_0 \to_T u_1 \to_T \cdots \to_T u_{n-1} \to_T u_n \equiv v.$$

The symmetric and transitive closure of $\xrightarrow{*}_T$ is denoted $\xleftrightarrow{*}_T$. If $(\ell, r) \in T$, then replacing an occurrence of ℓ by r (or vice versa) in some word $u \in A^*$ is called an *elementary transformation* in T. We urge the reader to see that this situation is entirely analogous to that of presentations.

A rewriting system T on A is called *terminating* (also sometimes called *Noetherian*) if there exists no infinite chain $u_1 \to_T u_2 \to_T \cdots$. The system is called *length-reducing* if $|\ell| > |r|$

[19] The name Thue is pronounced [tʉː]. That is, the e is silent. Unfortunately, this is not always adhered to; the name is often transcribed into Russian as *Tue*, with pronunciation [tʉɛ].

[20] Specifically, Lewis [279, p. 355] writes e.g. "A mathematical system is any set of strings of recognisable marks in which some of the strings are taken initially and the remainder derived from these by operations performed according to rules which are independent of any meaning assigned to the marks. That a system should consist of 'marks' instead of sounds or odours is immaterial."

for all rules $\ell \to r$ in T. It is called *length-preserving* if $|\ell| = |r|$ for all rules $\ell \to r$ in T. The system is called *locally confluent* if for all $u, v, w \in A^*$, we have $u \to_T v$ and $u \to_T w$ together imply that there exists some $z \in A^*$ such that $v \xrightarrow{*}_T z$ and $w \xrightarrow{*}_T z$. The system is called *confluent* if for all $u, v, w \in A^*$, we have $u \xrightarrow{*}_T v$ and $u \xrightarrow{*}_T w$ together imply that there exists some $z \in A^*$ such that $v \xrightarrow{*}_T z$ and $w \xrightarrow{*}_T z$. If a rewriting system T is terminating and confluent, then we say that T is *complete* (also sometimes called *convergent*). A word is $w \in A^*$ is *irreducible* (modulo T) if it does not contain any subword that is a left-hand side of some rule of T. The set of irreducible elements of T is denoted $\mathrm{Irr}(T)$. If T is terminating, we can for every word $w \in A^*$ find an element $w' \in \mathrm{Irr}(T)$ such that $w \xrightarrow{*}_T w'$ by "rewriting" w, i.e. continuously removing any left-hand sides of rules we find as subwords of w until this cannot be done any further. In a complete rewriting system, there exists a unique such w'. Hence any complete rewriting system has unique normal forms for all elements.

We say that a rewriting system T is *monadic* if for all rules $(\ell, r) \in T$ we have $|\ell| > |r|$, ℓ is non-trivial, and r is either a single letter or ε. We say that T is *$=$-monadic* if for all rules $(\ell, r) \in T$ we have $|\ell| \geq |r|$, ℓ is non-trivial, and r is either a single letter or ε. We say that T is *special* if $r \equiv \varepsilon$ and ℓ is non-empty for all rules $(\ell, r) \in T$. Hence any special rewriting system is monadic and $=$-monadic, and every monadic rewriting system is $=$-monadic. We remark that the distinction between monadic and $=$-monadic is only a minor technicality.[21] Indeed, from an algebraic point of view, the presence of length-preserving rules in a monadic system simply means identifying two generators in the monoid associated to the rewriting system, and Tietze transformations makes this very straightforward.

For a class $\mathcal{C}$ of languages, we say that T is a *$\mathcal{C}$-rewriting system* if for every distinct word r appearing as a right-hand side of some rule in T, the language $L_r = \{\ell \mid (\ell, r) \in T\}$ is in $\mathcal{C}$. For example, the rewriting system $\{((ab)^n, \varepsilon) \mid n \geq 1\}$ is a special $\mathcal{C}_{\text{reg}}$-rewriting system on $\{a, b\}$, whereas the rewriting system $\{(a^n b^n, a), (b^n a^n, b) \mid n \geq 1\}$ is a monadic $\mathcal{C}_{\text{cf}}$-rewriting system, but not a monadic $\mathcal{C}_{\text{reg}}$-rewriting system. We will generally forego hyphenation in the case that the class $\mathcal{C}$ has a standard English name, and speak of e.g. context-free rewriting systems rather than $\mathcal{C}_{\text{cf}}$-rewriting systems.

If $u, v \in A^*$ are such that $u \xrightarrow{*}_T v$, then we say that u is an *ancestor* of v modulo T, and v is a *descendant* of u. For $u \in A^*$, the set of all ancestors of u modulo T is denoted $\langle u \rangle_T$. Extending this notation, for $U \subseteq A^*$ we let $\langle U \rangle_T = \bigcup_{u \in U} \langle u \rangle_T$. If there is no v such that $u \to_T v$, then we say that u is *irreducible modulo* T. An irreducible element of an equivalence class is called a *normal form modulo* T for the equivalence class. For all these concepts, as long as the rewriting system in question is clear from context, we will generally suppress the "modulo T"-notation for brevity.

If $T \subseteq A^* \times A^*$ is a rewriting system, then $\xleftrightarrow{*}_T$ is a congruence on A^*. Thus $A^*/\xleftrightarrow{*}_T$ is a well-defined monoid. This will be called the *monoid associated with* T. Clearly the monoid associated with T is the same as – not just isomorphic to – the monoid defined by $\mathrm{Mon}\langle A \mid T \rangle$. If two words $u, v \in A^*$ are equal in $A^*/\xleftrightarrow{*}_T$, then we say that $u = v$ modulo T. We say that a monoid M *admits* the rewriting system $T \subseteq A^* \times A^*$ if it is isomorphic to $A^*/\xleftrightarrow{*}_T$. If M is a monoid generated by a finite set A, then we say that a rewriting system $\mathcal{R} \subseteq A^* \times A^*$ is *M-equivariant* if for every rule $(\ell \to r) \in \mathcal{R}$, we have $\ell =_M r$. It is easy to see, by induction

[21] For example, we quote the survey article [66]: "the definition of monadic Thue system requires that no rule in the system is length-preserving [...]. There are important differences between such systems and those that do possess length-preserving rules (see [53, 69]). However, for monadic systems [...] the restriction that no rule be length-preserving is made for technical convenience only".

on the number of rewriting steps, that if $\mathcal{R}$ is M-equivariant and $u \overset{*}{\leftrightarrow}_{\mathcal{R}} v$, then $u =_M v$. In other words, $\mathcal{R}$ is M-equivariant if and only if $\overset{*}{\leftrightarrow}_{\mathcal{R}} \subseteq \overset{*}{\leftrightarrow}_M$.

Example 1.2.8. We give some examples of rewriting systems and their connection with the algebraic theory of certain groups and monoids.

(1) The monoid $\mathbb{N} \times \mathbb{N}$ admits the finite complete rewriting system on $A = \{a, b\}$ with the single rule $ab \to ba$. The free group on A admits the finite special complete rewriting system on $A \cup A^{-1}$ with the rules $\{aa^{-1} \to \varepsilon, a^{-1}a \to \varepsilon \mid a \in A\}$. The infinite cyclic group $\mathbb{Z}$ admits the finite complete rewriting system on the alphabet $\{a, b\}$ with the rules $aba \to \varepsilon$, $ab \to ba$. Many more examples for infinite and finite groups have been presented by Le Chenadec [274, 275, 276], see also Bücken [84].

(2) An infinite group with an abelian subgroup of finite index admits a finite confluent length-reducing rewriting system if and only if the group is isomorphic to $\mathbb{Z}$ or isomorphic to the free product $C_2 * C_2$ [135].

(3) The monoid $\mathrm{Mon}\langle a, b \mid aba = bab \rangle$ does not admit a finite complete rewriting system T such that T is defined on an alphabet of only two letters; however, one can be find a finite complete rewriting system for the monoid which is defined on a larger alphabet [241]. It is a famous open problem whether every one-relation monoid (or indeed every one-relator group) admits a finite complete rewriting system.

(4) The word problem is easily decidable in any monoid which admits a finite complete rewriting system; however, the word problem can be of arbitrarily high complexity in a monoid defined by a finite complete rewriting system, see [395] for the precise statement. Furthermore, there exists a monoid M which admits a finite complete (and length-reducing) rewriting system such that the submonoid membership problem is undecidable for M, see [71, Corollary 5.2.2].

(5) Squier [460] proved that there exist finitely presented monoids with decidable word problem that cannot be defined by any finite complete rewriting system. As part of this proof, he showed that any monoid which admits a finite complete rewriting system satisfies the homological finiteness property FP_3 (see [460] for relevant definitions), which was later extended to FP_∞ [16]. Gray & Steinberg [171] recently showed that every one-relation monoid also satisfies FP_∞, thus lending credence to the conjecture that such monoids admit finite complete rewriting systems (see [388]). △

Rewriting systems and classes of languages are closely interlinked. This theme will be evident throughout this thesis. One of the many components of their intersection is the algebraic structure of monoids. We will now present this important component, which will be central in Chapters 3 and 4.

1.2.3 The word problem as a set

If G is a finitely generated group, with finite generating set A, then we mentioned in Example 1.2.2 that Anīsīmov studied the properties of the formal language consisting of elements in $A \cup A^{-1}$ which represent the identity element G. We formalise this here, and generalise this to semigroups and monoids. Let G be a group with finite (group) generating set A, with A^{-1} the set of inverses of the generators A. The language

$$\{w \mid w \in (A \cup A^{-1})^*, w =_G 1\}$$

is called the (group-theoretic) *word problem* for G. For a class of languages $\mathcal{C}$, we say that G has group-theoretic word problem *in* $\mathcal{C}$ (with respect to the generating set A) if the above set is in $\mathcal{C}$. It turns out that a good deal of algebraic information is encoded in this language. We have already mentioned that Anīsīmov proved that G is a finite group if and only if the above set is a regular language. Furthermore, we have the following remarkable theorem, commonly simply referred to as the *Muller-Schupp theorem.*

Theorem (Muller & Schupp, 1983)**.** *Let G be a finitely generated group. Then G has context-free word problem if and only if G is virtually free.*

On the other hand, let M be a monoid with a finite generating set A. Translating the above definition of the word problem directly to M does not, in general, yield much insight into the structure of M. That is, the language

$$\mathrm{IP}_A^M = \{w \in A^* \mid w =_M 1\}$$

which we will call the *identity problem* of M, does not contain much algebraic information about monoids (*special* monoids turn out to be exceptions to this, see Chapter 3, Corollary 3.4.3(3)).

Duncan & Gilman [144] instead introduced[22] a different generalisation of the group-theoretic word problem to all monoids. The *word problem of M with respect to A* is defined as the language

$$\mathrm{WP}_A^M := \{u\#v^{\mathrm{rev}} \mid u, v \in A^*, u =_M v\},$$

where $\# \in A_\omega \setminus A$ is a fixed symbol not in A.[23] For a class of languages $\mathcal{C}$, we say that M *has word problem in* $\mathcal{C}$ if $\mathrm{WP}_A^M \in \mathcal{C}$. If $\mathcal{C}$ is closed under inverse homomorphism, then one can show that whether or not M has word problem in $\mathcal{C}$ does not depend on the finite generating set chosen for M [144, Theorem 5.2]. Furthermore, if G is a group generated by a finite set A, then G has group-theoretic word problem in $\mathcal{C}$ if and only if $\mathrm{WP}_A^G \in \mathcal{C}$, see [144, Theorem 3]. That is, the word problem for monoids as defined above generalises the definition for groups.

Example 1.2.9. Let A be a finite non-empty alphabet, and A^* the free monoid on A. Then A^* has context-free word problem. For given any two words $u, v \in A^*$, by definition we have $u =_{A^*} v$ if and only if $u \equiv v$, and thus

$$\mathrm{WP}_A^{A^*} = \{u\#v^{\mathrm{rev}} \mid u, v \in A^*,\ u =_{A^*} v\} = \{w\#w^{\mathrm{rev}} \mid w \in A^*\}.$$

This last language is well-known and easy to show to be context-free. Furthermore, this language is certainly not regular, as can be easily verified (e.g. using the *pumping lemma* in [217]). Thus free monoids are context-free, but not regular. △

The above Example 1.2.9 shows that the word problem for any finitely generated free monoid is context-free. Thus we have the following useful consequence of Lemma 1.2.7.

Lemma 1.2.10. *Let $\mathcal{C}$ be a super-*AFL*. Then for every finite alphabet A we have $\mathrm{WP}_A^{A^*} \in \mathcal{C}$.*

While free monoids are context-free, and all context-free groups are classified by the Muller-Schupp theorem, the general problem of determining precisely which monoids have context-free word problem seems exceptionally difficult, and is wide open in general (see [144, Question 4]). We fully resolve this question for *special* monoids in Chapter 3. Before we can do this, however, we need to define what a *special* monoid is.

[22] Although Duncan & Gilman initialised the study of this language-theoretic word problem for monoids, the language in the definition was studied already by Book, Jantzen & Wrathall in 1982, see [66, Corollary 3.8]. This observation does not appear in the literature on the word problem for monoids.

[23] The reader should regard the symbol $\#$ as entirely disjoint from A, and as carrying no information other than purely as an "instruction" for the reader to turn on their head after reading it. In this manner, it is comparable to what Turing [478] calls a *symbol of the second kind*; symbols from A would be *of the first kind.*

1.3 Special monoids

Let $M = \text{Mon}\langle A \mid w_1 = 1, w_2 = 1, \dots, w_i = 1, \dots \rangle$. Then M is called *special*. That is, a monoid is special if it admits a presentation in which the right-hand side of every defining relation is the empty word. In this section we shall give an overview of the classical treatment of such monoids, and an overview of the main known results. This is in preparation for Chapters 3 & 5, in which special monoids will be studied in depth. Special monoids were first defined by G. S. Tseitin [477, p. 178], in his famous paper in which is demonstrated the existence of a semigroup with undecidable word problem and only seven defining relations. Tseitin called special monoids *associative systems of a special form*, but only used them as an intermediate stage in the proof of his main result. The first in-depth study of special monoids in their own right would instead be made by Adian [4], who stated several key results, including the decidability of the word and divisibility problems when $M = \text{Mon}\langle A \mid w = 1 \rangle$, i.e. the one-relation case. The proofs would later appear in his famous monograph [6]. Makanin, a student of Adian's, extended Adian's results in his 1966 Ph.D. thesis. A rewriting of these proofs and results were later made by Zhang and others in the early 1990s. In this section, we shall give an overview of the treatment given by these authors.

Recall that the *group of units* $U(M)$ is the maximal subgroup of M with respect to the idempotent 1, i.e. the subgroup of M consisting of all invertible elements. Special monoids have other maximal subgroups than $U(M)$, in which the identity element is some other idempotent; however, by a result of Malheiro all maximal subgroups of a special monoid are isomorphic to its group of units [325] (though this result can already be deduced from a result by McNaughton & Narendran [341, Theorem 5]). The importance of the group of units is one of the idiosyncratic features of special monoids; we shall see, for example, that for compressible monoids (in Chapter 4) this important rôle is instead played by the maximal subgroups that are not the group of units.

Let M be a k-relation special monoid. We shall give an overview of the following classical theorems from 1966 due to G. S. Makanin [309, 308]:

I. $U(M)$ is a k-relator group.
II. There exists a pseudo-algorithm for computing a k-relator presentation for $U(M)$.
III. The word and divisibility problems for M reduce to the word problem for $U(M)$.

For II, by a *pseudo-algorithm* we mean a procedure which does not always output an answer, but when it does, its output is always correct. We shall use the terms *procedure* and *pseudo-algorithm* interchangeably. The proofs of the above statements first appeared in Makanin's Ph.D.[24] thesis, written in Russian. After having a physical copy of this sent to me from Moscow, I produced an English translation of the thesis, which can be found online [387]. This material does not appear anywhere else in the literature on special monoids. We shall in part follow Makanin's notation, and in part follow Zhang [502], who produced a rewriting of the proof in terms of rewriting systems. This rewriting significantly compresses the statements of certain results, and makes the results significantly easier to parse in isolation from one another.

[24] In the Soviet system, there are two academic degrees similar in naming to the Western Ph.D., namely *candidate* and *doctor*. The former corresponds more or less directly to a Ph.D. degree, while the latter is closer to a *habilitation*. Thus, formally speaking, the results on special monoids were proved in Makanin's candidate thesis.

1.3.1 The invertible pieces

We will begin by treating the group of units of a special monoid. Thus, we will discuss invertible elements of special monoids. When doing so, there is a lemma of fundamental importance. This lemma is an overwhelmingly simple statement, but underpins the arguments used throughout this section, and indeed in the literature in general.[25] We shall primarily use it implicitly, except for a select few places; anything else would harm readability.

Fundamental Lemma. *If the words xy and yz are invertible in a monoid M, then all three of the words x, y, and z are invertible in M.*

This kind of "overlap" argument forms the basis of a (surprising) number of results. We will generally use it implicitly. As an application of a way that the fundamental lemma will be used is in the following way: if a word x is a prefix of an invertible word u, and a suffix of an invertible word v, then x is itself invertible.

We now return to special monoids. We shall in this section fix four special monoids, ordered by difficulty, which shall be illustrating examples for the types of difficult behaviours that can appear when computing the group of units:

- $M_1 = \mathrm{Mon}\langle a, b, c, d \mid abcdab = 1\rangle$.
- $M_2 = \mathrm{Mon}\langle a, b, c, d \mid abcdab = 1, acdcabdccddcabdcdacd = 1\rangle$.
- $M_3 = \mathrm{Mon}\langle a, b, c \mid abc = 1, b = 1\rangle$.
- $M_4 = \mathrm{Mon}\langle a, b, c, d \mid dba^4cd = 1, a^2 = 1\rangle$.

Let $M = \mathrm{Mon}\langle A \mid w_1 = 1, w_2 = 1, \dots, w_k = 1\rangle$ be an arbitrary finitely presented special monoid, which shall remain fixed throughout this section. The words $w_1, w_2, \dots, w_k$ will be called the *defining words* of the monoid. Deviating slightly from Zhang's definition into those used implicitly by Adian and Makanin [6, 308], we say that an invertible word $u \in A^+$ is *minimal* if none of its non-empty proper prefixes is invertible. For clarity, we will use the words *minimal word*, *minimal invertible word*, and *minimal invertible factor* interchangeably, swapping between them depending on the context. The set of minimal words forms an overlap-free code, and hence in particular also a biprefix code, as a subset of A^*.

Every defining word w_i for $1 \leq i \leq k$ is an invertible word, as $w_i =_M 1$. We shall now see that one can uniquely factorise every such word w_i into minimal words, by the following straightforward method. Of course, for every w_i there exists some shortest non-empty invertible prefix of w_i; say $w_i \equiv w_i' w_i''$, where w_i' is invertible, and no non-empty proper prefix of w_i' is invertible. That is, w_i' is a minimal word. Of course, it could be the case that w_i'' is empty, and that the entire word w_i is the shortest non-empty invertible prefix of w_i. In this case, we end our factorisation. If not, then as w_i and w_i' are invertible, so too – by the Fundamental Lemma – is w_i''. We may hence proceed to repeat the same process for w_i'', factorising this into smaller and smaller words, and eventually end up with an empty word, and have thus found a factorisation of w_i into minimal words. As the set of minimal words is a biprefix code, this is the unique way of factorising w_i into minimal words, as factorisation over a biprefix code is unique. Note that we have used no properties about the monoid M itself here. Thus the above description is not an effective one; saying that one can factorise the defining relation word w_i into minimal invertible factors is simply a property of the fact that

[25] For example, it appears in Adian [6, Lemma 3], Makanin [308, Lemma 3], McNaughton & Narendran [341, Lemma 2(2)], Nivat [378], Lallement [273], Perrin & Schupp [400, Lemme], Kobayashi [264, Corollary 3.3], Zhang [502, Proposition 2.1(4)], and many others.

w_i is invertible. This factorisation cannot in general be computed, and requires a solution to the word problem for the monoid.[26]

Thus, let us, for every $1 \leq i \leq k$, uniquely factorise $w_i \equiv w_{i,1} w_{i,2} \cdots w_{i,\ell_i}$, where $w_{i,j}$ for $1 \leq j \leq \ell_i$ is a minimal word. The set of all minimal words arising in this way shall be denoted Λ, and called the set of *presentation pieces* of M. That is,

$$\Lambda = \bigcup_{i=1}^{k} \bigcup_{j=1}^{\ell_i} \{w_{i,j}\} \subseteq A^*.$$

Example 1.3.1. Computing the presentation pieces of a special monoid is *a priori* a non-trivial task. Let us compute the presentation pieces of

$$M_1 = \operatorname{Mon}\langle a, b, c, d \mid abcdab = 1 \rangle.$$

As ab is a prefix of the defining word $abcdab$, ab is right invertible. Similarly, as ab is a suffix of the same word, ab is also left invertible. Thus ab is a proper non-empty invertible prefix of the defining word $abcdab$. Hence we find one factorisation of $abcdab$ into invertible factors as $(ab)(cdab)$. Now, $cdab$ is not minimal; since $cdab$ and ab are invertible, so too is cd. Thus $cdab$ has a proper non-empty invertible prefix, so our factorisation is refined as $(ab)(cd)(ab)$.

We claim that this factorisation is into minimal words. That is, we claim that neither a nor c are invertible. Indeed, let $\mathcal{R}_1 = \{(ababcd, 1), (cdab, abcd)\}$. We claim that $\mathcal{R}_1$ is a complete rewriting system which defines M_1. First, $\mathcal{R}_1$ is terminating as, when any rule is applied to a word, either the number of occurrences of ab in the word decreases or else an ab is strictly moved farther left). Second, $\mathcal{R}_1$ is locally confluent, as the only two non-trivial overlaps of rules (indicated by underlines) come from the words $abab\underline{cd}ab$ resp. $cd\underline{ab}abcd$; the first overlap resolves to ab regardless of which rule is applied first, while the second rewrites to cd if one first applies the rule $(ababcd, 1)$, and otherwise

$$cdababcd \rightarrow_{\mathcal{R}_1} abcdabcd \rightarrow_{\mathcal{R}_1} ababcdcd \rightarrow_{\mathcal{R}_1} cd.$$

Hence $\mathcal{R}_1$ is terminating and locally confluent, so it is complete. On the other hand, to see that $\mathcal{R}_1$ defines M_1, we have

$$ababcd =_{M_1} ababcd(abcdab) \equiv ab(abcdab)cdab =_{M_1} abcdab =_{M_1} 1,$$
$$cdab =_{M_1} (abcdab)cdab \equiv abcd(abcdab) =_{M_1} abcd,$$

so $\overset{*}{\leftrightarrow}_{\mathcal{R}_1} \subseteq \overset{*}{\leftrightarrow}_{M_1}$. On the other hand, $abcdab \rightarrow_{\mathcal{R}_1} ababcd \rightarrow_{\mathcal{R}_1} 1$, so $\overset{*}{\leftrightarrow}_{\mathcal{R}_1} \supseteq \overset{*}{\leftrightarrow}_{M_1}$. Hence $\mathcal{R}_1$ is a complete rewriting system which defines M_1.

Checking that neither a nor c are invertible is now trivial. For if a were invertible, then it is left invertible, so there is some $w \in \{a, b, c, d\}^*$ such that $wa =_{M_1} 1$, and hence $wa \xrightarrow{*}_{\mathcal{R}_1} 1$. But no rule of $\mathcal{R}_1$ ends in a, so this can never happen. Thus a is not invertible; checking that c is not invertible is entirely analogous.

Thus, the factorisation of $abcdab$ into minimal invertible factors is as $(ab)(cd)(ab)$, and the set of presentation pieces of M_1 is $\Lambda = \{ab, cd\}$. $\triangle$

[26] This final remark is not strictly true. There are only finitely many elements of Δ, so for a *fixed* M there certainly is a Turing machine which takes as input two words from Δ and outputs YES if they are equal, and otherwise NO. The problem with this is twofold: (1) there is no reason to expect that there is a uniform construction which starts with a presentation for a special monoid and outputs such a machine (in fact, and as we shall expand on later, one can show that there is no such construction); and (2) we have no reason to expect that we can recognise when we have such a Turing machine. This latter point is reminiscent of the idea that there is a Turing machine which solves the Riemann hypothesis; either the machine which always says YES is right, or the machine which always says NO is right – we just do not know which one, and at present we have no way of telling. Similarly, one can construct a finite set of candidate Turing machines, each one of which outputs one of the (finitely many) possible factorisations of the defining relation words into minimal invertible words; we just do not know which one is correct.

We now introduce another set, of critical importance. We let Δ denote the set of all minimal words $\delta \in A^*$ such that there exists some $w_{k,\ell}$ with $\delta =_M w_{k,\ell}$ and $|\delta| \leq |w_{k,\ell}|$. The set Δ is called the set of *invertible pieces* of the presentation. It is clear that no elements of Δ overlap non-trivially, for if $\delta_1, \delta_2 \in \Delta$ overlap non-trivially as $\delta_1 \equiv uv$ and $\delta_2 \equiv vw$ with $|v| \geq 1$ and $|uw| \geq 1$, then v is both left and right invertible by the fundamental lemma. Hence if $|w| \geq 1$, then v a proper non-empty invertible prefix of δ_2, which is a contradiction; or else if $|u| \geq 1$, then as uv and v are invertible, so too is u. Hence u is a proper non-empty invertible prefix of δ_1, which is again a contradiction. We conclude that no elements of Δ overlap non-trivially. In particular, Δ is an overlap-free code and a biprefix code. Furthermore, note that $\langle \Delta \rangle = \langle \Lambda \rangle$, as subgroups of M, and that $\Lambda \subseteq \Delta$.

Remark. The set Δ as defined by Zhang has one minor difference to our own, being the set of all minimal words δ such that there is some $w_{k,\ell}$ with $\delta =_M w_{k,\ell}$, and $|\delta| \leq \max_{i,j} |w_{i,j}|$. That is, the upper bound is global, rather than local to the specific piece that the minimal word δ is congruent to. The only time this "global" bound is used by Zhang is in the proof of the following proposition; thus we stress the point that by proving the following proposition by using the definition of Δ as given in this thesis, we are showing that any of Zhang's results can be used verbatim (even though his definition of Δ is slightly different).

Proposition 1.3.2 ([502, Prop 2.3]). *Let $x, y \in A^*$ and let $u, v \in \Delta^*$ such that $u = v$ in M and $|u| \geq |v|$. If $xuy \in \Delta^*$, then $xvy \in \Delta^*$.*

Proof. If $x, y \in \Delta^*$, then $xvy \in \Delta^*$.[27] If $x \notin \Delta^*$, then by [502, Prop 2.2][28] there exist $E, F \in A^+$ such that $x \in \Delta^* E, y \in F\Delta^*$, and $EuF \in \Delta$. Since $u = v$ in M, we have that[29] $EvF = EuF$ in M. Since $EuF \in \Delta$ and $v \in \Delta^*$, none of the proper prefixes of EvF is invertible in M. Thus EvF is minimal. As $EuF \in \Delta$, there is some $w_{i,j} \in \Lambda$ such that $EuF =_M w_{i,j}$ and $|EuF| \leq |w_{i,j}|$. Hence $EvF =_M w_{i,j}$, and as[30] $|EvF| \leq |EuF|$, we have $EvF \in \Delta$ by the definition of Δ. □

Thus the reader who may be accustomed to Zhang's approach should now be convinced that the two definitions of Δ are commensurable. We shall now describe a way of computing Δ in certain circumstances, and for computing a presentation for the group of units of M from Δ.

1.3.2 The group of units

We begin by making a rather important remark. The decision problem which takes as input a finite presentation of a special monoid M and outputs a presentation for its group of units $U(M)$, is in general undecidable [367]. This might already dampen one's spirits. However, we shall in this section show that this problem is rather tractable in many cases, and that in some special classes – e.g. in the one-relation case – the problem is decidable.

Now, note that it is possible for distinct pieces from Δ (or even Λ) to represent the same element of M. For example, if $M = \text{Mon}\langle a, b, c \mid ac = 1, ca = 1, bc = 1\rangle$, then all of a, b and c are invertible, so the minimal invertible pieces are $\Delta = \{a, b, c\}$, but as a is an inverse of c, and b is an inverse of c, we must (by uniqueness of inverses) have $a =_M b$. Far more

[27] Zhang misprints this as $xyv \in \Delta^*$.

[28] The proof of this cited proposition only uses the fact that Δ is a biprefix code. In particular the global upper bound is not used anywhere in the proof of that proposition.

[29] Zhang misprints this as $Evf = EuF$.

[30] It is here that the "global" bound is applied by Zhang. This ignores the fact that one can take the same piece $w_{i,j}$ to bound the length of EuF as EvF, as $EuF =_M EvF$.

complicated examples can be constructed. To remedy this, we will now partition Δ according to which elements of Δ represent the same elements of M. This partition of Δ, i.e. the partition of Δ induced by the equivalence relation $=_M$, will be denoted $\Delta_1 \cup \Delta_2 \cup \cdots \cup \Delta_\nu$. Again, we emphasise that this partition is not an effective one.

Let $X = \{x_1, \dots, x_\nu\}$, and let $\phi \colon \Delta^* \to X^*$ be the map induced by $\delta \mapsto x_i$ when $\delta \in \Delta_i$. This is a well-defined homomorphism, as Δ is a biprefix code. Then one can show (see [502, Theorem 3.7]) that

$$\mathrm{Gp}\langle X \mid \{\phi(w_i) = 1 \ (1 \leq i \leq k)\}\rangle$$

is a group presentation for the group of units $U(M)$ of M. Thus, as promised, $U(M)$ is a k-relator group. This yields the first (I) of the classical results.

It should perhaps come as no surprise that the set Δ is, in general, not effectively constructible. One of the naïve approaches one might attempt for actually computing the set Δ given a presentation is to successively find overlaps of the defining relations, then find overlaps of the overlaps, etc. until no more overlaps can be found. Every such overlap will be invertible by the Fundamental Lemma. This computation is made rigorous by the following Overlap Algorithm, which takes as input a non-empty set of words W and outputs an overlap-free code $C(W)$ which we shall call the *overlap-free code generated by W*.

The Overlap Algorithm

Input: A non-empty set of words $W \subseteq A^+$.

Output: An overlap-free code $C(W) \subseteq A^*$ such that $W \subseteq C(W)^+$.

1) Let $C_0 = W$.
2) If some two words $u, v \in C_0$ overlap non-trivially, i.e. we have $u \equiv u_0 w$ and $v \equiv w v_0$, where $w \in A^+$ and $|u_0 v_0| > \varepsilon$, then let C_1 be the set obtained by removing u, v from C_0 and adding those words of u_0, v_0 and w which are not empty.
3) Iterate step (2) with C_1 instead of C_0, and continue until no pair of words u, v can be found. This yields a finite sequence $C_0, C_1, \dots, C_K$ of sets.
4) We have $C(W) = C_K$.

We will sometimes substitute $C(w_1, \dots, w_k)$ for $C(\{w_1, \dots, w_k\})$ in the interest of ease of notation. It is clear from the definition that $C(W)$ is always a uniquely determined overlap-free code; see Nivat [378] or Lallement [273] for the (easy) full details. We remark that step (4) will eventually be reached after a finite number of steps, as every word $w \in W$ has only a finite number of prefixes and suffixes, and every prefix or suffix of w has strictly fewer prefixes resp. suffixes than w does.

Example 1.3.3. Let $W = C_0 = \{abcdab\}$. Then if $u \equiv abcdab$ and $v \equiv abcdab$, we have $u \equiv u_0 w$ and $v \equiv w v_0$ for $u_0 \equiv abcd$, $w \equiv ab$, and $v \equiv cdab$. Thus we can take

$$C_1 = \{abcd, ab, cdab\}.$$

Now if we take $u \equiv abcd$ and $v \equiv ab$, we can take $u_0 \equiv \varepsilon$, $w \equiv ab$, $v_0 \equiv cd$, and find

$$C_2 = \{ab, cd, cdab\}.$$

We take $u \equiv cdab$ and $v \equiv ab$. Thus we can take $u_0 \equiv cd$, $w \equiv ab$, and $v \equiv \varepsilon$, so

$$C_3 = \{ab, cd\}.$$

There are no non-trivial overlaps here; so $C(W) = C_3 = \{ab, cd\}$. △

Now, let $M = \text{Mon}\langle A \mid w_1 = 1, \dots, w_k = 1\rangle$ be a special monoid. Then Makanin calls the overlap-free code $C(\cup_i w_i)$ a set of *C-words* of the presentation, and in his notation this set would be written – rather confusingly, given our notation above – as "$\Delta[\cup_i w_i]$". Nevertheless, the use of Δ for the invertible pieces has become standard enough in subsequent literature that we shall maintain this. Now, if $M = \text{Mon}\langle A \mid w_i = 1 \ (1 \leq i \leq k)\rangle$, then we shall call $C(w_1, \dots, w_k)$ the *overlap pieces* of M.

Definition 1.3.4 (The overlap group of a special monoid)**.** Let M be the special monoid $\text{Mon}\langle A \mid w_i = 1 \ (1 \leq i \leq k)\rangle$. Let Y be a set in bijective correspondence with $C(\cup_i w_i)$ via a map $\psi \colon C(\cup_i w_i) \to Y$, which we extend to $\psi \colon C(\cup_i w_i)^* \to Y^*$. Then the *overlap group* $\mathcal{O}(M)$ is defined as the group with the presentation

$$\mathcal{O}(M) = \text{Gp}\langle Y \mid \psi(w_i) = 1 \ (1 \leq i \leq k)\rangle.$$

It is easy to see, by definition of $C(\cup_i w_i)$, that $\mathcal{O}(M)$ is isomorphic to the monoid defined by the *monoid* presentation with the same generators and defining relations, see [308, Lemma 8] for the full details.

Clearly there is a surjection $\mathcal{O}(M) \twoheadrightarrow \langle C(\cup_i w_i)\rangle_M \leq U(M)$, as every element of $C(\cup_i w_i)$ is invertible and must thus be subject to any relation in $U(M)$. Hence already $\mathcal{O}(M)$ gives some information regarding subgroups of the group of units of M. In general, however, the prescribed map is not injective (or indeed an isomorphism), i.e. $\mathcal{O}(M) \not\cong U(M)$. In general, we also have $\Delta \neq C(\cup_i w_i)$ (and there is generally not even any inclusion in either direction).

Example 1.3.5. Let $M_3 = \text{Mon}\langle a, b, c \mid abcac = 1, b = 1\rangle$. Then $C(abcac, b) = \{abcac, b\}$. Let $Y = \{y_1, y_2\}$ be in bijective correspondence with $C(abcac, b)$ via $abcac \mapsto y_1$ and $b \mapsto y_2$. Then the overlap group

$$\mathcal{O}(M_3) = \text{Gp}\langle y_1, y_2 \mid y_1 = 1, y_2 = 1\rangle = \text{Mon}\langle y_1, y_2 \mid y_1 = 1, y_2 = 1\rangle \cong 1$$

is trivial. Note that $abcac =_{M_3} acac$. Hence $(ac)^2 = 1$, and so ac is invertible. In fact, it is not hard to show that M_3 is equivalent to the monoid defined by the finite complete rewriting system with the two rules $\{(acac \to 1), (b \to 1)\}$, so $ac \neq 1$. In particular, $U(M)$ has a non-trivial element of finite order, and so $\mathcal{O}(M_3) \not\cong U(M)$. In fact, using this rewriting system, one easily shows that $\Delta = \{b, ac\}$ whereas $C(abcac, b) = \{b, abcac\}$. △

In general, it turns out that there is no algorithm for computing the set Δ given a presentation for M (we shall see this presently). On the other hand, in the one-relation case, something remarkable happens. This result is attributed to Adian by Lallement [273, p. 372].[31]

Theorem (Adian, 1960)**.** *Let $M = \text{Mon}\langle A \mid w = 1\rangle$ be a special one-relation monoid. Then $\Delta = \Lambda = C(w)$. Furthermore, $U(M) \cong \mathcal{O}(M)$.*

We have seen an example of this theorem in practice already, if we combine Example 1.3.1 and Example 1.3.3. As an aside, Otto & Zhang [396, Theorem 4.3] have proved that if M is given by a finite complete special rewriting system (see §1.2.2), then also $\mathcal{O}(M) \cong U(M)$.

[31] Adian did not deal explicitly with the one-relation case except as a very particular case at the end of a long treatise. The treatise gives Makanin's results in the particular case of "ℓ-homogeneous" k-relation special monoids; namely special monoids in which all k relations have the same length ℓ, where $\ell \in \mathbb{N}$. One-relation monoids are of course always ℓ-homogeneous for some ℓ. The proofs of these results in the isolated one-relation case were later simplified by Zhang [505] using the *Freiheitssatz* for one-relator groups, see §1.5.

1.3.3 Makanin's procedure

One can, in principle, from Zhang's work extract an algorithm for computing a presentation for $U(M)$, which takes as input only a k-relation presentation for M, under the assumption that the word problem for $U(M)$ is decidable. However, this algorithm is not constructive, and is certainly not implementable in practice. We shall instead give a *practical* procedure which computes the group of units in many cases, and which is far more applicable. This procedure takes as input a presentation for a special monoid, and is thus – in principle – something one can use in practice. This was described very implicitly by Makanin in his Ph.D. thesis, and we shall call it *Makanin's procedure*. The procedure is described below. Note that in stating the procedure, we temporarily forget the above definitions of M_1, M_2, M_3, and M_4 for ease of notation. One feature of the procedure is that while it can take an arbitrary special monoid as input, and often compute the group of units in a very efficient manner, it is generally not sufficient that the group of units be (by some oracle) known to have decidable word problem in order for the procedure to terminate. However, the procedure always produces explicit group presentations for which one must solve the word problem, and if one can do this the finitely many times one is asked to do so, then one finds a finite presentation for the group of units of the input special monoid.

Makanin's Procedure

Input: A k-relation special monoid M.

Output: If successful, a k-relator presentation for $U(M)$.

1) Compute the overlap group $\mathcal{O}(M) = G_0 = \text{Gp}\langle Y_0 \mid R_0 \rangle$, with overlap pieces Λ_0 and associated bijection $\psi_0 \colon \Lambda_0 \to Y_0$.
2) Solve the word problem for G_0. If this is not possible, then the procedure fails. If it is possible, then enumerate all non-graphical equalities $u =_{G_0} v$ of words with $u, v \in Y_0^*$ and such that $|\psi_0^{-1}(u)| \leq |\psi_0^{-1}(v)|$, and such that $\psi_0^{-1}(v)$ appears as a subword of some relation word from M.
3) For every $\lambda \in \Lambda_0$, using the equalities in (2), obtain the finite set S_λ consisting of all $\lambda' \in A^*$ such that $\lambda' \equiv h_1 u h_2$ and $\lambda \equiv h_1 v h_2$ for some $h_1, h_2 \in A^+$ and $u, v \in Y_0^*$ with $u =_{G_0} v$. Note that $\lambda \in S_\lambda$.
4) For every $\lambda \in \Lambda_0$, if there is some word $\lambda' \in S_\lambda$ with $|\lambda'| < |\lambda|$, then we replace every occurrence of λ in the defining relations of M by λ', resulting in a presentation of a monoid M_1. We have $M_1 \cong M_0$ by [308, Lemma 12].
5) Compute the overlap-free code generated by $\bigcup_\lambda S_\lambda$ using the Overlap Algorithm, and use these words to factor the defining relations of M_1. Let Λ_1 be the words which actually appear in this factorisation, and let G_1 be the group obtained from this factorisation in the same way the overlap group is obtained from a factorisation into overlap pieces.
6) Repeat the above steps for M_1, obtaining groups $G_0, G_1, G_2, \dots, G_k, \dots$. Eventually, by [308, Theorem 2], this stabilises, i.e. for some $K \geq 0$ we have $M_k = M_{k+1}$ and $G_k = G_{k+1}$ (as presentations) for all $k \geq K$.
7) We have $U(M) \cong G_K$.

Example 1.3.6. Consider the example

$$M_4 = \text{Mon}\langle a, b, c, d \mid dba^4cd = 1, abca = 1, a^2 = 1\rangle.$$

Then the overlap pieces are given as

$$\Lambda_{4,0} = \{d, a, bc, ba^4c\}$$

and so the overlap group for M_4 is

$$G_{4,0} = \text{Gp}\langle y_1, y_2, y_3, y_4 \mid y_1y_4y_1 = 1, y_2y_3y_2 = 1, y_2^2 = 1\rangle.$$

Now we enter Step 2 of the procedure. It is not hard to solve the word problem in this group, isomorphic to the free product $\mathbb{Z} * C_2$. Thus in M_4 we have the non-graphical equalities $a^4 = a^2, a^3 = a, a^2 = 1, bc = a^2$, and $bc = a^4$. In particular, we have that

$$\Lambda_{4,0} \ni b(a^4)c = b(a^2bc)c = b(bca^2)c = b(abca)c = b(bcbc)c = b(a^2)c = b(bc)c = bc,$$

in M_4, and therefore in Step 3 we obtain the sets

$$\begin{aligned}
S_d &= \{d\},\\
S_a &= \{a\},\\
S_{bc} &= \{bc\},\\
S_{ba^4c} &= \{b(a^4)c, b(a^2bc)c, b(bca^2)c, b(abca)c, b(bcbc)c, b(a^2)c, b(bc)c, bc\}.
\end{aligned}$$

Now performing Step 4, seeing that $bc \in S_{ba^4c}$ is the representative of shortest length, we find that $M_4 = M_{4,0}$ is isomorphic to the monoid obtained by replacing all occurrences of the piece ba^4c by bc, i.e. to

$$M_{4,1} = \text{Mon}\langle a, b, c, d \mid dbcd = 1, abca = 1, a^2 = 1\rangle.$$

Continuing with Step 5, we perform the overlap algorithm on $\bigcup_{\lambda\in\Lambda} S_\lambda$, which yields

$$C\left(\bigcup_{\lambda\in\Lambda} S_\lambda\right) = \{d, a, bc, b(a^4)c, b(a^2bc)c, b(bca^2)c, b(abca), b(bcbc)c, b(a^2)c, b(bc)c\}.$$

Now we can factor the defining relations of $M_{4,1}$ over this biprefix code; this factorisation becomes

$$M_{4,1} = \text{Mon}\langle a, b, c, d \mid (d)(bc)(d) = 1, (a)(bc)(a) = 1, (a)(a) = 1\rangle,$$

Thus we have

$$\Lambda_{4,1} = \{d, a, bc\},$$

and so the presentation for $G_{4,1}$ becomes

$$G_{4,1} = \text{Gp}\langle y_1, y_2, y_3 \mid y_1y_3y_1 = 1, y_2y_3y_2 = 1, y_2^2 = 1\rangle.$$

As $M_{4,1}$ is not given by the same presentation as $M_{4,0}$, we have not yet stabilised, and so perform the steps again. For Step (2), we find that the only non-trivial equality to consider is $a^2 = bc$, so

$$\begin{aligned}
S_d &= \{d\}\\
S_a &= \{a\}\\
S_{bc} &= \{bc, aa\}.
\end{aligned}$$

Now, the only non-trivial equality to consider is $bc = aa$, and as neither bc nor aa appear as a proper subword of some other element of $\Lambda_{4,1}$, nothing is done in Step (4). Furthermore, $C(\bigcup_{\lambda\in\Lambda} S_\lambda) = \{d, a, bc\} = \Lambda_{4,1}$, so the process has terminated. Thus $M_{4,1} = M_{4,2}$, and so

$$\begin{aligned}
U(M_4) \cong G_{4,1} &= \text{Gp}\langle y_1, y_2, y_3 \mid y_1y_3y_1 = 1, y_2y_3y_2 = 1, y_2^2 = 1\rangle\\
&\cong \text{Gp}\langle x, y \mid x^2 = 1, y^2 = 1\rangle \cong C_2 * C_2 = D_\infty,
\end{aligned}$$

the infinite dihedral group. △

A remarkable feature of the one-relation case is that steps (2)–(6) of Makanin's procedure are unnecessary if M has only a single defining relation. That this is unnecessary is a consequence of the *Freiheitssatz*, see [505] for further details. Of course, skipping these steps in that case simply consists of performing the overlap algorithm as presented by Adian. Note that the above procedure is only an *algorithm* in the case that the word problem is decidable in all the groups $G_0, G_1, \dots, G_K$ that one encounters. These are all k-relator groups, and the lengths of their defining relations are all less than or equal to $\max_i |w_i|$, giving an upper bound on the number of steps taken by the procedure before terminating.

We remark that Makanin's procedure also produces a "simple" presentation for any input special monoid M, provided the solution of the word problems for the groups one encounters along the way. For example, as noted, it detected the isomorphism

$$\begin{aligned} M_4 &= \text{Mon}\langle a, b, c, d \mid dba^4cd = 1, abca = 1, a^2 = 1\rangle \\ &\cong \text{Mon}\langle a, b, c, d \mid dbcd = 1, abca = 1, a^2 = 1\rangle. \end{aligned}$$

This notion of "simple" presentation, i.e. one which (informally speaking) minimises the presence of pieces appearing as subwords of other subwords, is quite closely connected to what Makanin calls a *distinguished*[32] presentation. However, this connection can only be taken so far, as distinguished presentations are not always what we might call the "simplest" presentation for a given monoid; for note that the relations $abca = 1$ and $a^2 = 1$ together imply $bc = 1$, and hence

$$M_4 \cong \text{Mon}\langle a, b, c, d \mid d^2 = 1, bc = 1, a^2 = 1\rangle \cong \text{Mon}\langle b, c \mid bc = 1\rangle * D_\infty.$$

This presentation is thus "simpler" than the one which was given by Makanin's procedure. Note that from this final presentation it becomes obvious that the group of units of M_4 has been correctly computed by Makanin's procedure.

The next example shows that in many other situations the group of units can be computed just as in the one-relator case, i.e. simply by considering the overlaps of words.

Example 1.3.7. Consider the example

$$M_2 = \text{Mon}\langle a, b, c, d \mid abcdab = 1, acdcabdccddcabdcdacd = 1\rangle.$$

Then the overlap pieces are given as

$$\Lambda_{2,0} = \{ab, cd, acd, cabd, ccdd\}$$

and so the overlap group for M_2 is

$$G_{2,0} = \text{Gp}\langle y_1, \dots, y_5 \mid y_1y_2y_1 = 1, y_3y_4y_5y_4y_3 = 1\rangle \cong F_3.$$

That is, $G_{2,0}$ is a free group on three generators. We can easily solve the word problem here. The only overlap pieces containing other pieces as proper subwords are $c(ab)d$ and $c(cd)d$. However, $\psi_{2,0}(ab) = y_1$ is not equal any element $w \in Y^*$ such that $|\psi_{2,0}^{-1}(w)| \leq |ab|$. Indeed, the only possibility for this would be if $\psi_{2,0}^{-1}(w)$ is one of $\{\varepsilon, ab, cd\}$. If ε, then $y_1 = 1$ in $G_{2,0}$, which is easily seen to not be the case; if cd, then $ab = cd$ in $M_{2,0}$, so $y_1 = y_2$ in $G_{2,0}$, which also not the case. Thus $S_{ab} = \{ab\}$. An analogous argument shows that $S_{cd} = \{cd\}$. Hence $S_\lambda = \{\lambda\}$ for every $\lambda \in \Lambda_{2,0}$, and so Makanin's procedure terminates, and we have that

$$U(M_2) = \text{Gp}\langle y_1, \dots, y_5 \mid y_1y_2y_1 = 1, y_3y_4y_5y_4y_3 = 1\rangle \cong F_3.$$

This shows the simplicity of Makanin's procedure in many cases. Without this procedure, computing the group of units of M_2 is not at all an obvious task (even given e.g. Zhang's paper). △

[32] Rather, this is my translation; the original Russian word is *vybrannoj*, which perhaps translates better as *selected*, but this latter word is not easily used in mathematical writing.

We will use some of Makanin's ideas about the set of pieces. However, Makanin does not explicitly define the set of pieces of a presentation. Instead, he defines the notion of c-*words*. We give a brief overview of this in language adapted to this thesis. The original can be found in [308]; in my English translation, this is [387, pp. 14–15]. He begins with any set into which the left-hand sides w_j of the defining relations $w_j = 1$ can be factored into minimal invertible factors[33], and calls this set

$$C = \{C_1, C_2, \dots, C_s\}.$$

That is, every w_j can be written as a product $C_{j_1} C_{j_2} \cdots C_{j_t} \in C^*$ of words from C. For our purposes, we will take $C = \Lambda$. Let $B = \{\beta_1, \beta_2, \dots, \beta_s\}$ be in bijective correspondence with C via the map induced by $\beta \colon C_j \mapsto \beta_j$, and extend this to an isomorphism $\beta \colon C^* \to B^*$. Let $G(C)$ be the group with the presentation

$$G(C) = \mathrm{Gp}\langle \beta_1, \dots, \beta_s \mid \beta(w_1) = 1, \dots, \beta(w_k) = 1 \rangle.$$

Now, there is always a surjection $G(C) \twoheadrightarrow U(M)$ induced by mapping $\beta_i \in B$ to $\phi(C_i) \in X$. Thus, if $u, v \in \Delta^*$ are such that $\beta(u) =_{G(C)} \beta(v)$, then $\phi(u) =_{U(M)} \phi(v)$ (but the converse does not always hold). If $\lambda \in \Lambda$, then we say that λ is obtained from itself by the *piece-generating operation*, and inductively, we say:

($*$) Suppose that $w \equiv h_1 \delta_{i,1} \delta_{i,2} \cdots \delta_{i,p} h_2$ is obtained from λ by the piece-generating operation, where $p \geq 0$ and h_1, h_2 are non-empty, and $\delta_{i,j} \in \Delta$ for every such $\delta_{i,j}$. Suppose then that $w' \equiv h_1 \delta_{j,1} \delta_{j,2} \cdots \delta_{j,t} h_2$, with $t \geq 0$, that $|w'| \leq |w|$, and that $\delta_{i,1} \delta_{i,2} \cdots \delta_{i,p} =_M \delta_{j,1} \delta_{j,2} \cdots \delta_{j,t}$. Then w' is also said to be obtained from λ by the piece-generating operation.

Any word obtainable from a C-word (that is, a presentation piece) C_i by the piece-generating operation will be called a c_i-word. The set of all c_i-words is called the c-*words* of the presentation. Now, if $w \in A^*$ can be obtained from λ by the piece-generating operation, then we denote this by $w \in [\lambda]^{\downarrow}$. It is easy to see that for any $w \in [\lambda]^{\downarrow}$, we have that w is a minimal word, i.e. $w \in \mathfrak{M}$ (see e.g. the second half of the proof of Lemma 3.2.4). We have $\lambda \in [\lambda]^{\downarrow}$, and $|w| \leq |\lambda|$ for every $w \in [\lambda]^{\downarrow}$, so in particular for every $\lambda \in \Lambda$ the set $[\lambda]^{\downarrow}$ is finite. In general, for $\lambda_1, \lambda_2 \in \Lambda$ we can have $[\lambda_1]^{\downarrow} \cap [\lambda_2]^{\downarrow} \neq \varnothing$ even when $\lambda_1 \not\equiv \lambda_2$. We remark the following useful fact: if $\lambda \in \Lambda$ and $\delta \in \Delta$ are such that $\lambda \equiv h_1 w h_2$ and $\delta \equiv h_1 w' h_2$, with $h_1, h_2 \in A^+$ and $w \xrightarrow{*}_S w'$, then $\delta \in [\lambda]^{\downarrow}$. The converse does not, in general, hold.

We shall make use of the above terminology in Chapter 3. This completes the exposition of the first two parts (I and II) of the classical results for special monoids.

1.3.4 Reducing decision problems

We now conclude our overview of the classical results by giving an overview of the fact that the word and divisibility problems for a special monoid M reduce to the word problem for $U(M)$. The fundamental idea behind this reduction can be heuristically explained as follows: suppose that we have a word w containing r_1 and r_2 as subwords, where $r_1 = 1$ and $r_2 = 1$ are some two defining relations of a special monoid. Suppose that these two occurrences have a

[33] Makanin begins in a more general situation by considering *any* biprefix code such that the defining relations can be factored over this code; using this, he then gives a procedure which produces the minimal invertible factors, under the assumption that the word problem in certain groups can be solved. This is of course the basis for Makanin's procedure.

non-trivial overlap, say in the word s. Then we can write $w \equiv w' r_1' s r_2'' w''$, where $r_1 \equiv r_1' s$ and $r_2 \equiv s r_2''$. As s is a suffix of r_1, it is left invertible, and as it is a prefix of r_2, it is right invertible. Hence any overlap of defining relations must be invertible: in particular, if we factor r_1 and r_2 (necessarily uniquely) into minimal invertible factors as $r_1 \equiv \delta_1 \delta_2 \cdots \delta_\kappa$ and $r_2 \equiv \delta_1' \delta_2' \cdots \delta_\ell'$, where $\delta_i, \delta_j' \in \Delta$ for all $1 \leq i \leq \kappa$ and $1 \leq j \leq \ell$, then we have

$$s \equiv \delta_i \delta_{i+1} \cdots \delta_\kappa \delta_1' \delta_2' \cdots \delta_j' \in \Delta^*,$$

for some $i, j \geq 1$. Hence, elementary transformations in a special monoid are controlled by invertible words; more specifically, they are controlled by words over Δ^*. Given the importance of resolving overlaps when solving the word problem in rewriting systems, this gives a heuristic for why a reduction of the word problem to the group of units is at all possible.

A little more rigorously, and using the rephrasing of Adian & Makanin's ideas done by Zhang [501, 502], we define the rewriting system

$$S = S(M) := \{(u, v) \mid u, v \in \Delta^* \colon u =_M v \text{ and } u >_s v\}.$$

Here $>_s$ denotes the shortlex ordering on A^*, as defined in §1.1.1, induced by (any) fixed ordering of A. Although S is generally infinite, Zhang proved that this system is complete and defines M [502, Proposition 3.2]. Thus, if one can effectively construct Δ and decide equalities of (necessarily invertible!) words over Δ^* in M, then one can solve the word problem for M by the following procedure: given two words $u, v \in A^*$, enumerate all finitely many equalities of words $w, w' \in \Delta^*$ with $\max\{|w|, |w'|\} \leq \max\{|u|, |v|\}$. Then, compute the finite set of descendants of u resp. v under $S(M)$ restricted to only the rules which involve one of the finitely many equalities computed. Denote these sets of descendants $\mathfrak{T}(u)$ and $\mathfrak{T}(v)$, respectively. Then $u = v$ in M if and only if $\mathfrak{T}(u) \cap \mathfrak{T}(v) \neq \varnothing$. This is exactly, up to being phrased in terms of rewriting systems, Theorem 3 in Makanin's Ph.D. thesis. Lallement [273] also describes this in the one-relation case. We have only described the reduction of the word problem for M to the word problem for $U(M)$. The reduction of the divisibility problems to the same problem requires no new ideas, and so this reduction is omitted (it can be found e.g. as [502, Theorem 5.3] or [308, Theorem 6]). In summary, we have:

Theorem (Makanin, 1966). *Let $M = \operatorname{Mon}\langle A \mid w_1 = 1, \dots, w_k = 1\rangle$. Then the word and divisibility problems for M reduce to the word problem for the k-relator group $U(M)$.*

One can in fact sharpen this result. Say that $v \in A^*$ is a *maximal invertible factor* of $u \in A^*$ if there exist $x, y \in A^*$ such that: (1) $u \equiv xvy$; (2) v is invertible; and (3) whenever $x \equiv x_1 x_2$ and $y \equiv y_1 y_2$ such that $|x_2 y_1| > 0$, then $x_2 v y_1$ is not invertible. We then have the following quite explicit way of comparing words from A^* in M, see Otto & Zhang [396, Theorem 5.2].

Lemma 1.3.8 (Otto & Zhang). *Let $M = \operatorname{Mon}\langle A \mid w_1 = 1, \dots, w_k = 1\rangle$, and let $u, v \in A^*$ be such that $u =_M v$. Then we can factorise u and v as*

$$u \equiv u_0 a_1 u_1 \cdots a_m u_m$$

$$v \equiv v_0 a_1 v_1 \cdots a_m v_m$$

respectively, where for all $0 \leq i \leq m$ we have $a_i \in A$ and

(1) $u_i =_M v_i$;

(2) u_i is a maximal invertible factor of u.

(3) v_i is a maximal invertible factor of v.

We note that some of the u_i and v_j may be empty. We end this section on special monoids by describing some particularly pleasant features of *one-relation* special monoids. Magnus (see

§1.5) proved that the word problem is decidable for any one-relator group $\mathrm{Gp}\langle A \mid w = 1\rangle$. We hence have the following centrally important theorem, proved already by Adian [3].

Corollary (Adian, 1960)**.** *The word and divisibility problems for every special one-relator monoid* $\mathrm{Mon}\langle A \mid w = 1\rangle$ *is decidable.*

This completes the classical treatment of special monoids. Before finishing, there are other results and papers on special monoids which bear mentioning. The earliest concern *identities* in special monoids by Adian [5]. We refer the reader to e.g. the survey by Shevrin & Volkov [443] or [88, Chapter II] for definitions and more information on identities. Adian's results were later used by e.g. Shneerson [444, 445][34] to completely classify the one-relation monoids satisfying non-trivial identities; this allows a complete characterisation of which one-relation monoids have decidable *first-order theory*, see [485]. Other sporadic work on special monoids has also appeared, particularly by Kashintsev [245, 246, 247, 248, 249], which is mostly related to embeddability and small cancellation results; we also refer the reader to [234, 67, 498, 503, 506] as well as [23, 83] for some further details on string rewriting and special monoids.

1.4 Graphs and geometry

In this section, we shall present the elements of graph theory as we shall need it in the sequel, particularly in Chapter 5. The following treatment follows [363] and [269] rather closely.

As before (cf. §1.2.1), an *alphabet* is a finite set of symbols. A *labelled (directed) graph* Γ consists of a set $V = V(\Gamma)$ of vertices, a *label alphabet* Σ, and a set E of (labelled) *edges*, where $E \subseteq V \times \Sigma \times V$. The cardinal of a graph is defined as the cardinal of its vertex set. For every edge $e \in E$, the projection to the first coordinate is called the *origin* $o(e) \in V$ of e, the projection to the second coordinate is called the *label* $\ell(e) \in \Sigma$ of e, and the projection to the third coordinate is called the *terminus* $t(e) \in V$ of e. For $\sigma \in \Sigma$, let $E_\sigma = E \cap (V \times \{\sigma\} \times V)$ be the set of edges labelled by σ. This definition of edge does not allow for multiple edges sharing all of origin, terminus, and label; but it does allow for two distinct edges sharing exactly two of the properties. We say that the graph has *bounded degree* if there exists some $K \geq 0$ such that for every vertex v each of the sets $\{e \mid e \in E(\Gamma) \text{ with } t(e) = v\}$ and $\{e \mid e \in E(\Gamma) \text{ with } o(e) = v\}$ has fewer than K elements. Many of the graphs we shall consider will be *rooted.* This is a simple notion: if Γ is a graph, then we root Γ by simply distinguishing a single vertex $v \in V(\Gamma)$, and call this the root of Γ. If Γ is a rooted graph, then $\mathtt{1}_\Gamma$ will denote its root.

A *labelled undirected* graph is defined much like a labelled directed graph, with the key difference being that every edge is instead an ordered pair $(\{u, v\}, \sigma)$ of edges and a label σ. Finally, an *unlabelled* and *undirected* graph is simply one in which an edge is an unordered pair of vertices $\{u, v\}$, while (slightly abusively) allowing for loops $\{u, u\}$. When Γ is a labelled directed graph, we can associate to Γ the undirected and unlabelled graph

$$\mathrm{ud}(\Gamma) := \left(V, \bigcup_{\sigma \in \Sigma} \{\{u, v\} \mid u \neq v, (u, \sigma, v) \in E_\sigma \text{ or } (v, \sigma, u) \in E_\sigma\} \right).$$

If Σ is the label alphabet of the labelled directed graph Γ, then we associate an alphabet $\overline{\Sigma}$ in bijective correspondence with Σ, denoting this bijection by $\sigma \mapsto \overline{\sigma}$ for all $\sigma \in \Sigma$, with $\Sigma \cap \overline{\Sigma} = \varnothing$. There is a natural directed and labelled graph obtained from $\mathrm{ud}(\Gamma)$ in the following way:

[34] These papers are not readily available, even in Russian. However, I have recently translated these papers into English, and they will shortly be made available online by L. Shneerson. I thank L. Shneerson and M. Volkov for their aid in providing copies of these papers and their interest in my work.

for every undirected edge $\{u, v\}$ of ud(Γ), we add two edges (u, σ, v) and $(v, \overline{\sigma}, u)$ whenever (u, σ, v) is an edge of Γ. We denote the resulting labelled graph by lud(Γ). Note that if (u, σ, v) is an edge of $E(\Gamma)$, then in lud(Γ) there is now both an edge (u, σ, v) and an edge $(v, \overline{\sigma}, u)$. Thus lud$(\Gamma)$ can be considered as a "symmetrised" version of Γ.. In particular lud(lud$(\Gamma)) =$ lud(Γ). Note also that lud(Γ) is (essentially) a graph in the sense as defined by Serre [439].

The connected components of a graph Γ is defined as the connected components of udΓ. The *tree-width* of a graph Γ is the minimum width among all possible *tree decompositions* of Γ; this is the same notion of tree decompositions as appears originally in [421]. The reader need know nothing about tree-decompositions than what appears in the following sentence: a tree has tree-width 1, and if a graph Γ has tree-width $\leq k$, then any induced subgraph of Γ has tree-width $\leq k$. The reader may treat tree-width as a black box, as it is in this manner that we shall use it.[35]

We will frequently reference walks in graphs; if p is a walk

$$u_0 \xrightarrow{\sigma_1} u_1 \xrightarrow{\sigma_2} \cdots \xrightarrow{\sigma_n} u_n$$

where $u_i \xrightarrow{\sigma_{i+1}} u_{i+1}$ is meant to indicate that $(u_i, u_{i+1}) \in E_{\sigma_{i+1}}$, for $\sigma_i \in \Sigma$ for all $1 \leq i \leq n$, then we say that p has *walk label* $\ell(p) = \sigma_1 \cdots \sigma_n \in \Sigma^*$. If there is such a walk, then we write $p\colon u_0 \xrightarrow{\ell(p)} u_n$. Note, however, that there may be several distinct walks $u_0 \xrightarrow{\ell(p)} u_n$, so this notation is slightly abusive; we shall only use it when the existence of such a walk is the issue treated, or when context makes the chosen walk clear. The *length* $|p|$ of the walk p is the integer n. If all vertices u_i are pairwise distinct, except possibly u_0 and u_n, then we say that p is a *path*, and we will accordingly refer to its walk label as its *path label*. An *undirected walk* in Γ is a walk in lud(Γ).

1.4.1 Ends of graphs

A useful notion in the study of the coarse geometry of a graph is that of *ends*. For this, we follow [363]. Let Γ be a connected labelled graph of bounded degree. We will distinguish a vertex $\mathtt{1} \in V(\Gamma)$ and say that $\mathtt{1}$ is the *root* of Γ. If v is any vertex of Γ, then we use $|v|_\Gamma$ to denote the length of a shortest (undirected) walk from $\mathtt{1}$ to v in ud(Γ). By $\Gamma^{(n)}$ we mean the subgraph of Γ consisting of all the vertices and edges which are connected to $\mathtt{1}$ by an undirected walk of length less than n; in particular $\Gamma^{(0)}$ is empty, $\Gamma^{(1)}$ consists of $\mathtt{1}$, and $\Gamma^{(2)}$ consists of $\mathtt{1}$, its neighbours, and all edges connecting these vertices. As in the theory of ends in e.g. [118] or [414], the connected components of $\Gamma \setminus \Gamma^{(n)}$ will be the central objects of study. If C is a connected component of $\Gamma \setminus \Gamma^{(n)}$, then we say that a *frontier point* of C is a vertex u of C such that $|u|_{\mathrm{ud}\Gamma} = n$. If v is a vertex of Γ with $|v|_{\mathrm{ud}\Gamma} = n$, then we use $\Gamma(v)$ to denote the component of $\Gamma \setminus \Gamma^{(n)}$ which contains v. The set of frontier points of $\Gamma(v)$ will be denoted by $\Delta(v)$; this set is always finite (but possibly unbounded in v) as Γ has bounded degree. For example, in the infinite grid of Figure 1.2, the number of frontier points of $\Gamma(v)$ grows unboundedly as a function of $|v|$, as the number of frontier points in $\Gamma(v)$ grows as $\sim |v|^2_{\mathrm{ud}\Gamma}$.

[35] In slightly more detail, the way in which shall use it is the following: starting with a monoid M with a non-context-free Cayley graph Γ, we shall find inside it an induced subgraph Γ', closely related to the Cayley graph of a group G with $G \leq M$. As $G \leq M$, we will be able to conclude that G is not a context-free group; by deep theorems in geometric group theory, one can conclude that Γ' does not have finite tree-width. Therefore, via the black box, we conclude that Γ does not have finite tree-width. By certain general theorems on graphs, this will allows us to conclude that Γ is not quasi-isometric to a tree. Note that the method above is an instance of reducing a problem about monoids to a problem about groups; Sushkevič would likely be pleased of this application of his principle.

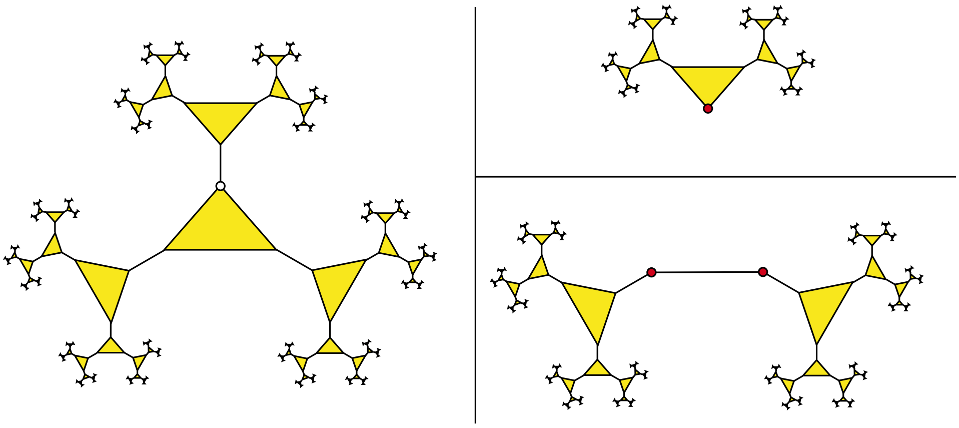


Figure 1.1: Left: A context-free graph Γ, with root 1 central and enlarged. Right: Two of the three end-isomorphism classes of Γ, with frontier points marked in red. The third end-isomorphism class is represented by $\Gamma(1)$, and is isomorphic to Γ. Note that all edge labels are intentionally suppressed; it is assumed that all triangles have the same labels, as do all single edges.

Let $u, v \in V(\Gamma)$. An *end-isomorphism* between the two subgraphs $\Gamma(u)$ and $\Gamma(v)$ is a mapping ψ between $\Gamma(u)$ and $\Gamma(v)$ such that

(1) ψ is a label-preserving graph isomorphism, and
(2) ψ maps $\Delta(u)$ onto $\Delta(v)$.

We will write $\Gamma(u) \sim \Gamma(v)$ if there exists some end-isomorphism $\psi : \Gamma(u) \to \Gamma(v)$.

Definition 1.4.1. A connected labelled graph Γ of bounded degree is said to be *context-free* if the set of end-isomorphism classes $\{\Gamma(v) \mid v \in V(\Gamma)\}/\sim$ is finite. Furthermore, a labelled graph Γ of bounded degree that is the union of finitely many connected graphs is said to be context-free if all its connected components are context-free.

We now give one example together with a non-example of a context-free graph, to illustrate some of the considerations of importance, as well as consolidate the definitions. The interested reader may construct many more examples of both kinds without much difficulty.

Example 1.4.2 (A context-free graph)**.** Let Γ be the graph obtained from the following procedure: take a triangle graph, and attach a single edge to each of its vertices. To each of these edges, attach an isomorphic copy of the original triangle, and repeat; the graph Γ is the colimit of the sequence of graphs obtained. See Figure 1.1, in which the resulting graph is drawn out without any edge labels. We note that this graph is very closely related to the Cayley graph of the virtually free group $\mathrm{Gp}\langle a, b \mid a^2 = b^3 = 1\rangle \cong C_2 * C_3$. △

Thus context-freeness captures the idea that if one traverses the graph from the identity, travelling outwards, one eventually encounters graphs which one already has seen.

Example 1.4.3 (A graph which is not context-free)**.** Let Γ be the infinite two-dimensional grid, with two different types of edge labels; see Figure 1.2. The root 1 is the central enlarged vertex. Some graphs $\Gamma^{(n)}$ are drawn out. In all of these graphs, the red vertices indicate vertices at distance n from the root 1. Then Γ is not a context-free graph, as the number of frontier points of

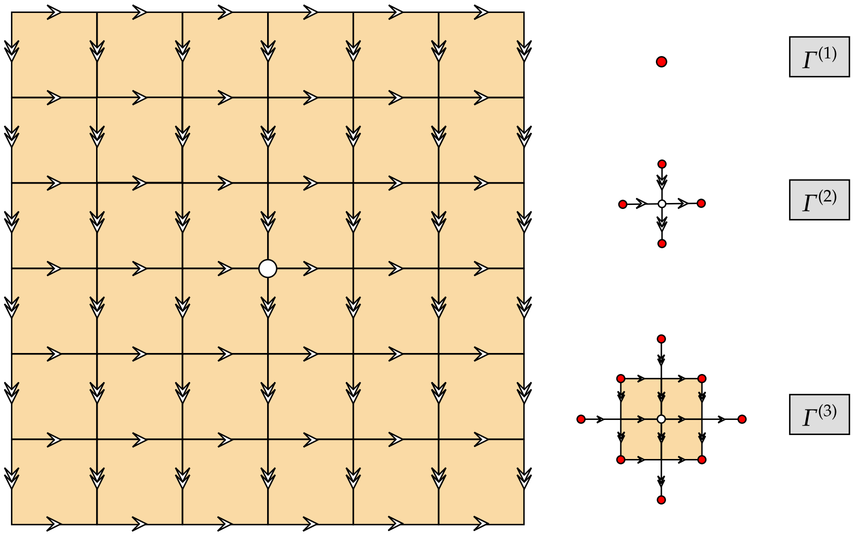


Figure 1.2: An illustration of why the infinite two-dimensional grid is not a context-free graph. We remark that the filled-in squares are purely to make the picture clear, and have no significance to the graph itself. Left: the graph Γ, with root 1 in the center. Right: three complements of ends of Γ, showing that the number of frontier points of $\Gamma \setminus \Gamma^{(n)}$ grows unboundedly in n.

$\Gamma \setminus \Gamma^{(n)}$ equals the number of vertices at distance n from 1. But this number grows quadratically, and hence unboundedly, in n, so there cannot be only finitely many end-isomorphism classes of ends of Γ; this is because any end-isomorphism must map the frontier points of one graph in an end-isomorphism class bijectively onto the frontier points of another. We note that this graph is isomorphic to the right Cayley graph of the group $\mathrm{Gp}\langle a, b \mid [a, b] = 1\rangle \cong \mathbb{Z} \times \mathbb{Z}$ with the generating set $\{a, b\}$, which is not a virtually free group. △

Note that as the two graph-theoretic tools used above for defining context-free graphs – namely connected components and distances – are used in the undirected sense, $\mathrm{lud}(\Gamma)$ is context-free if and only if Γ is context-free.

We note that by [363, Corollary 2.7] a context-free graph remains context-free independent of the choice of root; this result will often be implicitly assumed throughout. In particular, some statements will be easier to state without explicitly rooting the graphs involved; the reader can to such statements always add the sentence "independently of root chosen". Furthermore, if we speak of a context-free graph, then we implicitly assume this graph is labelled and has bounded degree.

A final definition which occasionally is useful is that of a *second-level subgraph* of a context-free graph, again following [363]. Let Γ be a context-free graph rooted at 1, and let $\{\Gamma_0, \Gamma_1, \dots, \Gamma_k\}$ be a complete list of representatives of the end-isomorphism classes of Γ. Assume without loss of generality that $\Gamma_0 = \Gamma(1)$. Let Γ_i be one of these representative graphs, with Δ_i as its set of frontier points. If the edges of Γ_i that are incident to some vertex of Δ_i are deleted, then there remains a finite union of connected subgraphs of Γ_i. These subgraphs are called the *second-level subgraphs* of Γ_i. It may be the case that Γ_0 is not the second-level subgraph of any Γ_i. If this is the case, then we declare Γ_0 to nevertheless be a second-level subgraph. We note that for every Γ_i there exists some Γ_j such that Γ_i is a second-level subgraph of Γ_j, and every second-level subgraph is isomorphic to some Γ_i (see

[363, p. 57]). Thus the notion of second-level subgraphs formalises the notion of "seeing" elements of $\{\Gamma_0, \Gamma_1, \dots, \Gamma_k\}$ as one travels in Γ outward from the root. We emphasise that a context-free graph only has finitely many second-level subgraphs, and that these graphs completely capture the structure of the original graph.

1.4.2 Pushdown automata and graphs

We will follow Muller & Schupp [363] more or less verbatim in their definition of a *pushdown automaton* (henceforth abbreviated as *pda*) and of concepts related to this for the theory of the ends of graphs. A pda is a 7-tuple $\mathcal{M} = (Q, A, Z, \delta, q_0, z_0, \hat{Q})$ consisting of a finite set Q of *states*, a finite *input alphabet* A, a finite *stack alphabet* Z, an *initial state* $q_0 \in Q$, a *start symbol* $z_0 \in Z \cup \{\varepsilon\}$, a set of *final states* $\hat{Q} \subseteq Q$, and a *transition function* δ. We allow pdas to run even when their stack is empty; hence δ is a mapping

$$\delta \colon Q \times (A \cup \{\varepsilon\}) \times (Z \cup \{\varepsilon\}) \to \mathrm{Fin}(Q \times Z^*).$$

We interpret

$$\delta(q, a, z) = \{(q_1, \zeta_1), \dots, (q_m, \zeta_m)\}$$

for $q_i \in Q, a \in A, z \in Z$, and every $\zeta_i \in Z^*$ for $1 \leq i \leq m$, as follows: when the pda is in state q, reading the input symbol a, and with z the top symbol on its stack, then the machine can, for any choice of $i = 1, \dots, m$, change to state q_i, replace z by ζ_i, and move the head reading the input tape one square to the right. We will consider the symbols in ζ_i as being placed on the stack from left to right; thus the rightmost symbol of ζ_i (or ζ_i itself if $\zeta_i \equiv \varepsilon$) is the top of the stack. If $\delta(q, a, z)$ is empty, then the machine halts.

We interpret

$$\delta(q, a, \varepsilon) = \{(q_1, \zeta_1), \dots, (q_m, \zeta_m)\}$$

as follows: when the pda is in state q, reading the input symbol a and the stack is empty, then the machine can change state to q_i and add ζ_i to the stack for a choice of i, $1 \leq i \leq m$.

We interpret

$$\delta(q, \varepsilon, z) = \{(q_1, \zeta_1), \dots, (q_m, \zeta_m)\}$$

as follows: when the pda is in state q with z as the top symbol of the stack, then – independently of the input symbol being scanned – the machine can change to state q_i and replace z by ζ_i. The input head is not moved. Such a transition of $\mathcal{M}$ will be called an ε-move. We also allow ε-moves when the stack is empty in the obvious way.

If $(q_i, \zeta_i) \in \delta(q, a, z)$, then we will write this graphically as $q \xrightarrow{a, z \mapsto \zeta_i} q_i$. This gives rise to the usual graphical interpretation of a pda as a finite machine with its states as the vertices, and with its transitions the edges of the machine.

A *total state*[36] of the pda $\mathcal{M}$ is a pair (q, ζ) where $q \in Q$ is the current state of $\mathcal{M}$, and $\zeta \in Z^*$ denotes the contents of the stack. We write $(q, \zeta) \vdash^a_{\mathcal{M}} (q', \zeta\zeta')$ if $\delta(q, a, z)$ contains (q', ζ'), where z is the rightmost symbol of ζ (or ε if ζ is empty). We write $\mathcal{M} \vdash^* (q, \zeta)$ if there exists some word $w \in A^*$ such that when $\mathcal{M}$ is started in its initial total state (q_0, z_0) with w written on the input tape, it is possible for $\mathcal{M}$ to be in the total state (q, ζ) after reading w.

We define the *graph* $\Gamma(\mathcal{M})$ *of the pda* $\mathcal{M}$ as follows. The vertex set of Γ is

$$V = \{(q, \zeta) \mid q \in Q, \zeta \in Z^*, \mathcal{M} \vdash^* (q, \zeta)\}$$

i.e. the set of possible total states that $\mathcal{M}$ can reach (generally, this is an infinite set). If (q, ζ) and

[36] Total states are also called *configurations* by e.g. Gray, Silva & Szakáks [170].

(q', ζ') are possible total states of $\mathcal{M}$ and $\mathcal{M}$ can go from (q, ζ) to (q', ζ') in a single transition, i.e. if $(q, \zeta) \vdash^a_{\mathcal{M}} (q', \zeta')$, then there is an edge e with label a from (q, ζ) to (q', ζ'). The following theorem, which is [363, Theorem 2.6], yields the main reason for studying such graphs.

Theorem (Muller & Schupp). *A graph Γ is context-free if and only if there exists some pda $\mathcal{M}$ such that $\Gamma = \Gamma(\mathcal{M})$.*

We shall find this characterisation of context-free graphs useful in Chapter 5.

1.4.3 Cayley graphs

Cayley graphs allow for a graphical representation of algebraic structures. As tools for studying infinite groups, they were first properly applied by Burnside [87, p.426], and particularly by Dehn [129], who used the term *Gruppenbild* (Ger. *picture of a group*).[37] The (right) *monoid Cayley graph* $\Gamma_M(M, A)$ of a monoid M together with a generating set A for M is the labelled graph with vertex set M and an edge $m_1 \xrightarrow{a} m_2$, i.e. $(m_1, m_2) \in E_a$, if and only if $m_1 \cdot \pi(a) = m_2$ in M. Here $\pi : A^* \to M$ denotes the natural homomorphism from the free monoid on A to M. Note that there is a dependency on the homomorphism π chosen, but we shall assume that this is fixed whenever we say that M is generated by A. The *group Cayley graph* $\Gamma_G(G, A)$ of a group G with a generating set A for G is defined analogously, but with two additional assumptions: that A be placed in involutive correspondence with an alphabet A^{-1} such that $A \cap A^{-1} = \varnothing$, and such that for every edge $(u, v) \in E_a$, there exists an edge $(v, u) \in E_{a^{-1}}$. In particular, the label alphabet of $\Gamma_G(G, A)$ is $A \sqcup A^{-1}$, and $\pi(a^{-1}) = \pi(a)^{-1}$. For an overview of material pertaining to Cayley graphs, particularly of general algebraic structures, we refer the reader to [251, 284, 93]. The following theorem is fundamental, and was proved in the same paper as context-free graphs were introduced.

Theorem (Muller & Schupp, 1985). *A finitely generated group is virtually free if and only if $\Gamma_G(G, A)$ is a context-free graph for some (any) choice of finite generating set A.*

If M is a monoid with finite generating set A, then the strongly connected component in $\Gamma_M(M, A)$ of the identity element is denoted $\mathfrak{R}_1$, and, following Stephen [467], is called the *Schützenberger graph of* 1. This will be an important object of study in Chapter 5. Of course, if M is a group, then $\Gamma_M(M, A) = \mathfrak{R}_1$. The vertex set of $\mathfrak{R}_1$ is the set of right invertible elements of M, but $\mathfrak{R}_1$ is not generally isomorphic to the right Cayley graph of the submonoid of right units of M. We shall revisit this theme in Chapter 5.

1.4.4 Logic of graphs

We shall assume some background from model theory and formal logic; all notions implicit here can be found in e.g. [412, 331, 258]. We can consider a labelled graph as a formal logical structure Γ with domain V (the vertex set of the graph) and a single relation, the edge relation. A *first-order predicate* in a graph is a predicate which can involve vertices, the edge relation, equality,

[37] Dehn's definition was not the one we present below. In fact, Dehn defined what is today usually referred to as *Stephen's procedure* – this observation does not appear to have been made anywhere in the literature. The *Gruppenbild* of a finitely presented group is obtained by attaching loops of all defining relations, including pairs aa^{-1} and $a^{-1}a$ for generators a, and folding together determinisable pairs. In the group case, the resulting limit graph results in the usual Cayley graph – but in the general monoid case, the same procedure only produces an induced subgraph of the full Cayley graph. Dehn's procedure is described in full by Chandler & Magnus [107, I.§5].

quantifiers ($\exists, \forall$) over vertices, and their boolean combinations ($\neg, \wedge, \vee, \rightarrow$). A *monadic second-order predicate* also allows quantification (both universal and existential) over sets of vertices; if such quantification is only allowed over finite sets of vertices, then this is known as *weak* monadic second-order predicates.

The *first-order (monadic second-order) theory* of a graph Γ is the collection of all first-order (monadic second-order) predicates ϕ with no free variables such that $\Gamma \models \phi$. We say that the first-order (monadic second-order) theory of a graph is *decidable* if, given any first-order (monadic second-order) predicate ϕ, there is an algorithm which decides whether or not $\Gamma \models \phi$. For more detailed background on these notions, see e.g. [363, 270]. A remarkable theorem due to Muller & Schupp [363] is that any context-free graph has decidable monadic second-order theory. The converse is certainly not true in general; for example, consider the graph Γ' with vertex set two disjoint copies of $\mathbb{N}$, where we denote the first copy as $\{0, 1, \dots\}$ and the second as $\{v_1, v_2, \dots, \}$. Let the root of Γ' be 0. Let the label alphabet be a singleton $\{a\}$ and the edges of Γ' be $(n, a, n+1)$ whenever $n \in \mathbb{N}$, and an edge $(\frac{1}{2}n(n+1), a, v_n)$. Thus Γ' has the appearance of $\mathbb{N}$ with a single strand of hair growing at every vertex of the form $\frac{1}{2}n(n+1)$. One can show with little difficulty that this graph has decidable monadic second-order theory (see [149] for much more general statements), but the graph is clearly not context-free: the distance between the individual hairs grows as one moves farther away from the root of $\mathbb{N}$.

Shifting our attention to (right) Cayley graphs, decidability of either the first-order or the monadic second-order theory of the Cayley graph of a finitely generated monoid M does not depend on the finite generating set chosen [270]. For this reason, we will generally omit reference to finite generating set below. There is a number of connections between decision problems for a given monoid M and different theories associated to the Cayley graph of M. This is most apparent in the group case: the first-order theory of the Cayley graph of a group is decidable if and only if the word problem for the group is decidable [269], and the monadic second-order theory of the Cayley graph of a group is decidable if and only if the group is virtually free [269, 363]. For monoids, by [270, Proposition 4], if the first-order theory of the Cayley graph of a finitely generated monoid is decidable, then the monoid has decidable word problem, but by [270, Proposition 5] there exists a monoid with word problem decidable even in linear time, but the Cayley graph of which nonetheless has undecidable first-order theory.

Thus studying the monadic second-order theory of the Cayley graphs of monoids seems, in general, a hopeless task; at least compared to the group case. On the other hand, in Chapter 5 we shall almost completely characterise the special monoids whose Cayley graphs have decidable monadic second order-theory.

1.4.5 Geometric (semi)group theory

Geometric methods in group and semigroup theory are vast in number, and the many results in these areas cannot possibly be covered in any justice in this introduction. Instead, in this section, we shall give a brief background on the importance and history of these methods, and give shallow definitions of deep concepts that shall be used later in the thesis, while including enough references to make expansion easy. Much of the exposition is taken from Howie [218]. See also the Notes on Literature for more detailed references.

Let (X, d) and (X', d') be metric spaces. An *isometry* $f\colon X \to X'$ is a map such that

$$d'(f(x), f(y)) = d(x, y) \quad \text{for all } x, y \in X.$$

Clearly any isometry is a continuous and injective map. If f is surjective, then f^{-1} is also an isometry; we then say that (X, d) and (X', d') are *isometric* – this is a "sameness" notion. A coarser notion of "sameness" is provided by *quasi-isometries.* Let $\lambda > \kappa \geq 0$ be real numbers. Then a map $f \colon X \to X'$ is a (λ, κ)-quasi-isometry if

$$\frac{1}{\lambda} d(x, y) - \kappa \leq d'(f(x), f(y)) \leq \lambda d(x, y) + \kappa \quad \text{for all } x, y \in X.$$

An isometry is a $(1, 0)$-quasi-isometry. A quasi-isometry f need not be continuous or injective. If f is *coarsely surjective*, i.e. if every point in X' is a bounded distance from some point in $\mathrm{im}(f)$, then there is a (λ', κ')-quasi-isometry $f' \colon X' \to X$ for some λ', κ'.[38] In this case, we say that the metric spaces (X, d) and (X', d') are *quasi-isometric.* A fundamental example of quasi-isometric but not isometric spaces are $(\mathbb{Z}^n, d)$ and $(\mathbb{R}^n, d)$ for $n \geq 1$, with the Euclidean metric d. There is an embedding $\mathbb{Z}^n \hookrightarrow \mathbb{R}^n$, which is an isometry, and while not surjective, it is coarsely surjective. The map $f \colon \mathbb{R}^n \to \mathbb{Z}^n$ defined by rounding every entry of the tuple $(x_1, \dots, x_n) \in \mathbb{R}^n$ to the nearest integer is a $(1, \frac{1}{2}\sqrt{n})$-quasi-isometry.

The *geometric realisation* of a graph is defined, following Serre [439], as follows: let Γ be an undirected and unlabelled graph. Form the topological space T which is the disjoint union of $V(\Gamma)$ and $E(\Gamma) \times [0, 1] \subseteq E(\Gamma) \times \mathbb{R}$, where the topology on $V(\Gamma)$ and $E(\Gamma)$ is the discrete topology. Let ς be the finest equivalence relation on T for which $(e, 0)\varsigma o(e)$ and $(e, 1)\varsigma t(e)$ for $e \in E(\Gamma)$. The quotient space T/ς is then the geometric realisation of Γ. We define a *quasi-isometry of undirected graphs* to be a quasi-isometry of the geometric realisation of the graphs in question. The distinction between a quasi-isometry of undirected graphs and general quasi-isometries is important, but to prevent cumbersome notation, we omit writing this explicitly in all places.

A *geodesic segment* of length ℓ in a metric space (X, d) from x to y is the image of an isometric embedding $i \colon [0, \ell] \to X$ with $i(0) = x$ and $i(\ell) = y$. We say that a metric space is *geodesic* if there exist geodesic segments between all pairs of points (thus discrete metric spaces are not geodesic). A *triangle* $\Delta(x, y, z)$ in X with vertices x, y, z is the union of the three geodesic segments from x to y, y to z, and z to x, respectively. A geodesic metric space (X, d) is *hyperbolic* if all its triangles are "thin", i.e. if there exists a constant $\delta \geq 0$ such that for all triangles $\Delta(x, y, z)$ in X, each edge of Δ is contained in the δ-neighbourhood of the union of the other two sides of Δ. Every tree is hyperbolic (take $\delta = 0$), and the hyperbolic plane $\mathbb{H}^2$ is hyperbolic – thankfully – e.g. by taking $\delta = 2 \log 3$. Crucially, if two geodesic metric spaces (X, d) and (X', d') are quasi-isometric, then one is hyperbolic if and only if the other is.

We now turn to groups. Let G be a group generated by some finite set S of (group) generators. Then every $g \in G$ can be expressed as a word

$$g = x_1^{\varepsilon_1} x_2^{\varepsilon_2} \cdots x_n^{\varepsilon_n}$$

where $x_i \in S$ and $\varepsilon_i = \pm 1$ for all $1 \leq i \leq n$. The number n here is of course the length of the word in the free group on S. If $g, h \in G$ then we define $d_S(g, h)$ to be the length of the shortest word representing $g^{-1}h$. It is not hard to verify that this is indeed a metric, and this metric d_S is called the *word metric* on G (with respect to S). One can show without much difficulty that if S and T are two finite generating sets for a group G, then (G, d_S) and (G, d_T) are quasi-isometric. Thus we can speak of two finitely generated groups being quasi-isometric without any ambiguity. For example, the group $\mathbb{Z}$ is quasi-isometric to $\mathbb{Z} \times C_n$ for any $n \geq 1$, where $C_n = \mathrm{Gp}\langle a \mid a^n = 1 \rangle$ is the cyclic group with n elements; a quasi-isometry is given, for example, by mapping $g \in \mathbb{Z}$ to $(g, 1) \in \mathbb{Z} \times C_n$. The exact constants λ and κ witnessing that

[38] This statement uses the axiom of choice; it can thus be quite hard to actually construct f'.

this is a quasi-isometry depend on the generating set chosen, but if one chooses the standard generating set $\{(1, 1), (0, a)\}$, then this map is in fact a $(1, 0)$-quasi-isometry, and the fact that it is coarsely surjective is witnessed by the fact that an arbitrary element (g, a^j) is at distance j from $(g, 1)$, as $j \leq n - 1$, this distance is globally bounded.

Now, let G, S be as above, and consider the Cayley graph $\Gamma(G, S)$, defined as in §1.4.3. Then we can consider $\Gamma(G, S)$ as a topological space in the usual way, and hence the geometric realisation of $\Gamma(G, S)$ into a geodesic metric space. The metric $d(x, y)$ on this space is just the length of the shortest path joining x to y. We leave as an exercise the observation that (G, d_S) and $(\Gamma(G, S), d)$, with d the metric above, are quasi-isometric spaces. We say that G is a *hyperbolic group* if $\Gamma(G, S)$ is a hyperbolic space in the above sense. Finite groups, free groups, and fundamental groups of compact 2-manifolds of genus $g \geq 2$ are all hyperbolic, as they are quasi-isometric to $\mathbb{H}^2$. The free abelian group $\mathbb{Z}^n$ is hyperbolic if and only if $n \leq 1$. An important theorem (cf. [122, Theorem 2.3]) is that any hyperbolic group is finitely presented.

On the other hand, defining *hyperbolic monoids* is more difficult. Certainly, we could define a hyperbolic monoid M, generated by a finite set S, to be one where the geometric realisation of the right Cayley graph $\Gamma(M, S)$ is hyperbolic; one can show that this property is independent of choice of generating set. However, this property is incredibly weak in other senses: for example, let G be any (!) finitely generated group. We define the monoid G^0 to be the group G together with an appended element 0 such that $00 = 0$ and $g0 = 0g = 0$ for all $g \in G$. Then it is easy to verify that G^0 is hyperbolic. This means that this definition of hyperbolic monoids is rather unstable, and many other definitions have been proposed to rectify this, cf. especially [28]. There have also been extensions of the notions of quasi-isometries and actions of monoids to this more general setting; we refer the interested reader to [165, 166]. We will focus on one extension of hyperbolicity in particular, proposed by Duncan & Gilman [144], which is especially pleasant in its link with the word problem.

Let M be a monoid finitely generated by A. Let $\#_1, \#_2$ be two new symbols. Then M is *word-hyperbolic* if there exists a regular language $L \subseteq A^*$ such that

$$M(L) := \{u\#_1 v\#_2 w^{\text{rev}} \mid u, v, w \in L, uv =_M w\}$$

is a context-free language. For groups the two notions of hyperbolicity can be shown to coincide [144], but there exist examples of hyperbolic monoids which are not word-hyperbolic. We remark that any monoid with context-free word problem, in the sense presented earlier in §1.2.3, clearly is word-hyperbolic. This is the monoid analogue of the fact that virtually free groups are hyperbolic. We shall see in Chapter 3 that any special monoid with virtually free group of units is word-hyperbolic.

This concludes our brief incursion into geometric group and semigroup theory; and with this, we have defined all concepts and properties necessary to understand the statements of all theorems and results in this thesis. The reader only interested in this new material may proceed directly to Chapter 2; the reader interested in understanding *one-relator groups* and, more generally, getting a (discursive) overview of the results known for various classes of groups regarding decision problems, finiteness properties, and much more, may remain in this chapter and carry on reading.

1.5 Case study: one-relator groups

The following section is slightly different in nature than the previous few. We will not need many theorems about one-relator groups in the sequel. Instead, the collection of results and historical material in this section are collected because of the striking way that one-relator groups have played – and continue to play – a central rôle in the development of combinatorial group theory. Some properties in this section (e.g. residual finiteness), which are of no further direct relevance to this thesis, are defined in §1.6.

A *one-relator group* is one which can be defined by a presentation $\mathrm{Gp}\langle A \mid w = 1\rangle$. All free groups are one-relator groups. One-relator groups have a special place in the theory of combinatorial group theory, and they occupy a special place in this thesis, as we shall see throughout. We shall here give a short overview of some of the important ideas in the area, and what is known about certain properties of one-relator groups.

The study of one-relator groups started in the latter part of the 19th century in the theory of functions of a complex variable, and is intricately linked especially with groups G of real linear fractional transformations of the upper half complex plane. Such groups G are called *Fuchsian* groups, and are subgroups of $\mathrm{PSL}_2(\mathbb{R})$. On the other hand, related to such functions is the study of the fundamental groups of 2-manifolds, also known as *surface groups*, which are one-relator groups. This was all joined together by Dehn [129, 131] whose geometric intuition proved crucial; he immediately realised the importance of the word problem for understanding the fundamental group of a surface, and the general combinatorial framework in which this takes place led him to conjecture that certain subgroups of one-relator groups are free. When his student Magnus was tasked with proving this, this resulted in the first, and arguably one of the most important, result for one-relator groups: the *Freiheitssatz* (Ger. *freedom* or *freeness* theorem).

Theorem (Dehn & Magnus'[39] *Freiheitssatz*[40])**.** *Let $G = \mathrm{Gp}\langle A \mid w = 1\rangle$, where w is cyclically reduced. Suppose that the letter $a \in A$ appears in w (either as a or a^{-1}). Then the subgroup of G generated by $A \setminus \{a\}$ is free and freely generated by this set.*

The *Freiheitssatz* can be seen as a partial extension of the Nielsen-Schreier theorem for free groups, and indicates that one-relator groups are, in a rather weak sense, "nearly free", in that the only relations that hold in the group are strongly dependent on the presence of the defining relation. In fact, one-relator groups contain an abundance of free subgroups, as the Tits alternative is true for one-relator groups, i.e. a subgroup of a one-relator group is either solvable or contains a free subgroup of rank two [244].[41] Short proofs, and extensions, of the theorem can also be found in e.g. [306, 298, 429, 307, 194, 153]. Using the *Freiheitssatz*, Magnus proved the following remarkable theorem two years later [301].

Theorem (Magnus)**.** *Let $G = \mathrm{Gp}\langle A \mid w = 1\rangle$. Then G has decidable word problem.*

[39] The importance of Max Dehn and his student Wilhelm Magnus to the development of combinatorial group theory cannot be overstated. We content ourselves by noting that there are essentially two textbooks on the subject: the book by Magnus, Karrass & Solitar [306] and the book by Lyndon & Schupp [298]. The first is dedicated to Dehn; the second to Magnus.

[40] The *Freiheitssatz* is usually ascribed solely to Magnus, as he published the first proof of it [300]. However, Chandler & Magnus [107, p. 114] clearly indicate that the theorem was well understood by Dehn via geometric intuition for a decade prior to Magnus' combinatorial proof.

[41] In fact, a one-relator group contains a non-abelian free group unless it is isomorphic to $\mathrm{BS}(1, m)$ or is cyclic [490].

As part proving this theorem, Magnus established a separate decidability result related to the subgroup membership problem, which we state for completeness and because of this connection. Let $G = \mathrm{Gp}\langle A \mid w = 1\rangle$, and let $X \subseteq A \cup A^{-1}$. The subgroup $\langle X\rangle_{\mathrm{gp}} \leq G$ is said to be a *Magnus subgroup* of G. Note that the *Freiheitssatz* tells us that any proper Magnus subgroup of a one-relator group is free.

Theorem (Magnus)**.** *Let $G = \mathrm{Gp}\langle A \mid w = 1\rangle$. Then membership in any Magnus subgroup of G is decidable.*[42,43]

We shall mention a result which is important throughout the remainder of this section. Magnus [302] proved the general result that if an n-generator k-relator group can be generated by $n - k$ elements, then it is freely generated by these elements. In particular if a one-relator group has a generating set with fewer than its original number of generators, then it is free. Thus all one-relator presentations of a given one-relator group have the same number of generators. We remark that it is decidable whether a one-relator group $\mathrm{Gp}\langle A \mid w = 1\rangle$ is free; it is free if and only if the cyclically reduced word w is trivial or a *primitive* word in the free group on A, see [298, Prop 5.10]. Here a *primitive* word is one that can be an element of some basis of the free group; this can be algorithmically checked with Whitehead's algorithm [488].

The theory of one-relator groups divides into two parts: the *torsion-free* and the *torsion* case, respectively. Note that a group is said to have torsion if it has non-trivial elements of finite order. A classical theorem (see [306, Theorem 4.12, p. 266]), which follows from the *Freiheitssatz*, shows that a one-relator group $\mathrm{Gp}\langle A \mid w = 1\rangle$ has torsion if and only if the word w is a proper power of another word. Thus $\mathrm{Gp}\langle a, b \mid (abba)^2 = 1\rangle$ has torsion, but the group $\mathrm{Gp}\langle a, t \mid [a, tat^{-1}] = 1\rangle$ is torsion-free. Although it may appear as if the presence of torsion elements would be a complicating factor, the opposite is true. One-relator groups with torsion have many structural properties which distinguish them from the torsion-free counterparts. The first major theorem demonstrating this was, without a doubt, the following, proved by B. B. Newman [372] in his Ph.D. thesis.

Theorem (The B. B. Newman Spelling Theorem)**.** *Let $G = \mathrm{Gp}\langle A \mid w^n = 1\rangle$ with w cyclically reduced and $n > 1$. Suppose that two words $u, v \in (A \cup A^{-1})^*$ define the same element, where u is a reduced word containing $a \in A$ non-trivially and v does not contain a. Then u contains a subword which is identical with a subword of w^n or w^{-n} of length greater than $(n - 1)/n$ times the length of w^n.*

I have described the historical context in which this theorem arose, along with the story of who the man behind the theorem was [384]. The theorem has subsequently been strengthened by a number of authors, see especially Gurevich [194], Brodskij [79], and Howie [220, 219]. The theorem implies, in the modern framework of geometric group theory, that one-relator groups with torsion are hyperbolic. For the purposes of studying groups from the point of view of decision problems, the following corollary is arguably the most important.

[42] This theorem has at times been misunderstood. The term *generalised word problem* or *extended word problem* (*erweiterten Identitätsproblem*) has been used to describe the problem in the statement of the theorem, but it has also been used to describe the subgroup membership problem, which is much harder, and remains open. This has led to some incorrect statements in the literature; for example, Stillwell [469] claims that Magnus solved the subgroup membership problem for one-relator groups.

[43] The problem of deciding membership in a subgroup generated by a subset of the generating set is also decidable for groups which satisfy the $C'(1/6)$-small cancellation condition [179].

Corollary (B. B. Newman). *Let $G = \mathrm{Gp}\langle A \mid w^n = 1\rangle$ with $n > 1$. Then G has decidable conjugacy problem.*

When $n > 2$, Pride [411] gave a significantly shorter proof of the same theorem. It is striking that nearly a century after Magnus' solution to the word problem for one-relator groups, the conjugacy problem remains open for torsion-free one-relator groups.[44] Another important reason for the structural strength of the torsion case is the presence of a malnormal subgroup structure. A subgroup $H \leq G$ is said to be *malnormal*[45] if for all $g \in G \setminus H$ we have $gHg^{-1} \cap H = 1$. Free factors are malnormal in a free product; and B. B. Newman proved (see [372, Lemma 2.3.1]) the remarkable theorem that Magnus subgroups are malnormal in one-relator groups with torsion. Using the Magnus-Moldavanskii breakdown procedure (see [298, II.§6] and [357, 339][46]), this allows for powerful results to be proved by induction. For example, this malnormality is crucial in applying Wise's Malnormal Quasiconvex Hierarchy Theorem (see [492, Theorem 11.2]) to one-relator groups with torsion. This allowed Wise to resolve an old conjecture due to Baumslag [34], as follows:

Theorem (Wise). *Let $G = \mathrm{Gp}\langle A \mid w^n = 1\rangle$ with $n > 1$. Then G is residually finite.*

Note in particular that this implies that any one-relator group with torsion is *Hopfian* and that its automorphism group is finitely generated.[47] These statements had already been proved by Pride [410] in the two-generator case. The above methods have little hope for extending to all one-relator groups, as Magnus subgroups need not be malnormal in torsion-free one-relator groups. Hence, the geometric methods employed by Wise cannot, in general, be applied in this case. The best one can say in this direction is a result due to Bagherzadeh [24], who proved that if G is a one-relator group, H is a Magnus subgroup of G, and $g \in G \setminus H$, then $gHg^{-1} \cap H$ is cyclic. Louder & Wilton [293] have recently extended the geometric methods for studying one-relator groups with torsion, and proved that they are *coherent*, i.e. that every finitely generated subgroup is finitely presented. This result is important in recent work by Gray & Ruskuč [169] on the structure of the group of units of one-relator special inverse monoids. It is an important open problem whether all torsion-free one-relator groups are coherent.

We now mention some results regarding the subgroup structure of one-relator groups. This is more difficult than it might at first appear. First of all, although subgroups of free groups are themselves free, it is certainly not the case that subgroups of one-relator groups are themselves always one-relator groups. In fact, as mentioned above, it is not even known whether finitely generated subgroups of one-relator groups are always finitely presented. Few general results are known about the subgroups of one-relator groups. Perhaps the strongest is that the finitely presented normal subgroups of one-relator groups are classified; they are either free or of finite index [60, 38]. General results of this form are far and few in between. We mention some. Although the abelian subgroups of torsion-free one-relator groups are easy to determine (by

[44] In some places, claims have appeared that the conjugacy problem for torsion-free one-relator groups has been proved decidable, see e.g [350, p. 37]. This is based on a sketch of a proof [237], and which was never completed. The sketch proof uses detailed small cancellation theory to study the problem, but the general consensus (e.g. at the wow workshop held at the University of East Anglia in 2018) is that the proof is incomplete and that the problem remains open.

[45] Malnormal subgroups were introduced and named by B. Baumslag in his Ph.D. thesis [29].

[46] I thank D. Moldavanskii for providing me a copy of his paper.

[47] Neither of these properties hold, in general, for torsion-free one-relator groups. Indeed, studying the automorphism groups of one-relator groups is a difficult task. Collins and Levin [119] proved that the automorphism group of a Baumslag-Solitar group need not be finitely generated. Even the special case of the automorphism group of $\mathrm{Gp}\langle a, b \mid [a^m, b^n] = 1\rangle$ is difficult, see Tieudjo & Moldavanskii [476].

a result on cohomological dimension due to Lyndon [297]) the corresponding problem was curiously harder in the torsion case. The abelian and solvable subgroups of one-relator groups with torsion were finally classified by B. B. Newman [373]. In the same paper, centralisers of elements in one-relator groups with torsion were also proved to always be cyclic; this result is key in the recent breakthrough result by Minasyan & Zalesskii [355] that one-relator groups with torsion are conjugacy separable. Many fascinating basic problems about understanding subgroups of one-relator groups remain, especially in the torsion-free case: for example, it is an open problem whether a one-relator group can contain a non-abelian simple subgroup. Furthermore, the lower central series of the *Baumslag-Solitar group* $\mathrm{BS}(m, n) = \mathrm{Gp}\langle a, b \mid ba^m b^{-1} = a^n\rangle$ was only very recently studied (cf. Bardakov [25].[48] The structure of the lower central series of general one-relator group currently seems far out of reach.

Associated to the subgroup structure is the subgroup membership problem, and more generally the submonoid membership problem. The subgroup membership problem remains open both in the torsion-free and torsion cases, and appears quite far out of reach of current methods. Some particular cases have been studied; for example, Bezverknij [57, 58] solved the problem in certain one-relator groups, including the Baumslag-Solitar groups $\mathrm{BS}(m, n)$. However, in 2020, Gray [168] gave the first known example of a one-relator group with undecidable submonoid membership problem. This group is defined as

$$G_B := \mathrm{Gp}\langle a, b \mid [a, bab^{-1}] = 1\rangle.$$

The proof is easy, and goes via embedding the right-angled Artin group $A(P_4)$ into G_B by an HNN-extension. This group has some remarkable properties, and has appeared in the literature on many occasions in the past. We mention a few of these fascinating properties, none of which were pointed out in [168]. In 1984, it was shown by Brunner, Burns & Solitar [82] that G_B is not subgroup separable. Burns, Karrass & Solitar [86] give G_B as the first example of a free-by-cyclic group that is not subgroup separable. Moreover, Gersten [160] prove that G_B is the fundamental group of a $\mathrm{CAT}(0)$ 2-complex, and some of its geometric properties are studied. Niblo & Wise [375] show that G_B virtually embeds in $A(P_4)$. Thus G_B and $A(P_4)$ share essentially all algorithmic properties. In particular G_B has decidable subgroup membership problem, in spite of it not being subgroup separable. Finally, Button [89] proved that G_B is a *large* group, in the sense that it virtually surjects a non-abelian free group; very few two-generator one-relator groups are known to have this property.[49] In particular, every countable group embeds in some quotient of G_B, as every large group has this property [371]. For comparison, Baumslag & Pride [32] conjectured that all one-relator groups with torsion are large, which was proved shortly thereafter by Stöhr [470]. Note that as $\mathbb{Z}^2 \cong \langle a, bab^{-1}\rangle < G_B$, it follows that G_B cannot be hyperbolic. This raises the following question.

Question 1.5.1. *Does there exist a hyperbolic one-relator group with undecidable submonoid membership problem?*

We end on a note on the problem of different presentations for the same one-relator group. We shall investigate this briefly in the context of special monoids in Chapter 3. Note that if ϕ is an automorphism of the free group F_A, then $\mathrm{Gp}\langle A \mid w = 1\rangle$ is obviously isomorphic to $\mathrm{Gp}\langle A \mid \phi(w) = 1\rangle$. Let r be a positive integer and G be an r-generator group. Two generating

[48] I thank V. G. Bardakov for providing me with a copy of this paper.

[49] Any n-generator one-relator group for $n > 2$ is large, by a celebrated result due to B. Baumslag & Pride [31]. Note that not every n-generator one-relator group surjects F_2 [463].

r-tuples

$$\mathbf{g} = (g_1, g_2, \dots, g_r) \quad \text{and} \quad \mathbf{g}' = (g_1', g_2', \dots, g_r')$$

are said to be *Nielsen equivalent* if there is an automorphism

$$x_i \mapsto Y_i(x_1, x_2, \dots, x_r), \quad i = 1, 2, \dots, r$$

of the free group on $x_1, x_2, \dots, x_r$ such that $g_i' = Y_i(g_1, g_2, \dots, g_r)$ for $i = 1, 2, \dots, r$. The r-tuples $\mathbf{g}$ and $\mathbf{g}'$ are said to lie in the same T*-system* if there is an automorphism ϱ of G such that $\mathbf{g}'$ is Nielsen equivalent to $\varrho(\mathbf{g})$. This notion of T-systems is important for classifying the different presentations a one-relator group might have. The importance comes from the following theorem: suppose that

$$\text{Gp}\langle x_1, x_2, \dots, x_r \mid R = 1 \rangle$$

is a presentation of the group G associated to the r-tuple $\mathbf{g}$, i.e. such that the kernel of the homomorphism $x_i \mapsto g_i$ for $i = 1, 2, \dots, r$ is the normal subgroup of the free group F_r on $x_1, \dots, x_r$ generated by R. Then it is not hard to see (see e.g. [409]) that $\mathbf{g}'$ is in the same T-system as $\mathbf{g}$ if and only if there is an automorphism ϕ of F_r such that $\text{Gp}\langle x_1, \dots, x_r \mid \phi(R) = 1 \rangle$ is a presentation of G associated with $\mathbf{g}'$. Of course, there may *a priori* be other presentations of G associated with $\mathbf{g}'$. However, in the one-relator case, the conjugacy theorem for one-relator groups (see [306, Theorem 4.11] states that if $\mathbf{g}, \mathbf{g}'$ as above are in the same T-system, and if

$$\text{Gp}\langle x_1, x_2, \dots, x_r \mid S = 1 \rangle$$

is a presentation associated to $\mathbf{g}'$, then S is a cyclic conjugate of $\phi(R)^{\pm 1}$ for some automorphism ϕ of F_r. If one knows that a certain one-relator group G has only a single T-system of generating r-tuples, for example, then this can be used to detect whether a one-relator group is isomorphic to G or not. Indeed, this is how Pride [410] solves the isomorphism problem for two-generator one-relator groups with torsion: such groups essentially only have a single T-system.

In fact, Magnus [306, p. 401] conjectured that *every* one-relator group only has a single T-system. However, this turned out to be false. Zieschang [507] and McCool & Pietrowski [338] provided the first counterexamples. In fact, Brunner [81] showed that $\text{Gp}\langle x, y \mid x^{x^y} = x^2 \rangle$ has infinitely many T-systems. To this effect, we present a striking theorem due to Pride [408], which we believe has received far less attention than it deserves.

Theorem 1.5.2 (Pride). *Let $w_1 \equiv x_1^3 x_2^{-1} x_1^{-2} x_2$ and $w_2 \equiv x_1 [x_1^{-1}, x_2]^2$. Then the groups*

$$\text{Gp}\langle x_1, x_2 \mid w_1^n = 1 \rangle \quad \text{and} \quad \text{Gp}\langle x_1, x_2 \mid w_2^n = 1 \rangle$$

are isomorphic if and only if $n = 1$, in which case the group has infinitely many T-systems.

This is a striking counterexample to the converse of the following theorem due to Magnus, Karrass & Solitar [306, Corollary 4.13.1]: let $n > 1$. If $\text{Gp}\langle A \mid w_1^n = 1 \rangle$ is isomorphic to $\text{Gp}\langle A \mid w_2^n = 1 \rangle$, then $\text{Gp}\langle A \mid w_1 = 1 \rangle$ is isomorphic to $\text{Gp}\langle A \mid w_2 = 1 \rangle$.

1.6 A table of properties

This section consists only of one object, together with some explanations of relevant terms. This object is a table of various classes of groups and their properties, especially centred on various finiteness properties and decision problems. The aim of the section is to provide a concise overview and collation to aid in referencing. Throughout, all objects are finitely presented.

We give some missing definitions and references necessary to understand some entries of the table. Recall the definition of a residually finite group from the very beginning of the chapter. Any finitely presented residually finite group has decidable word problem [340, 361]. For a stronger property than residual finiteness, we say that a group G is *subgroup separable* (or l. e. rf, i.e. *locally extended residually finite*) if for every finitely generated subgroup H and for every $g \in G \setminus H$, there is a finite quotient of G in which g has non-trivial image but H has trivial image. Any subgroup separable group is residually finite. By analogy with the proof of decidability of the word problem for finitely presented residually finite groups, any finitely presented subgroup separable group has decidable subgroup membership problem. A recent notion of separability is *conjugacy separability*, which we do not define here, see [354].

A group G is *Howson* if the intersection of any two finitely generated subgroups of G is again finitely generated. It is *Hopfian* if it is not isomorphic to any proper quotient of itself. A raag is a *right-angled Artin group*, which is defined as follows: let Γ be an undirected graph. Then the raag with the underlying graph Γ is the group with generators $V(\Gamma)$, and defining relations $[u, v] = 1$ whenever $u, v \in E(\Gamma)$. Because of the importance of raags in modern geometric group theory (see e.g. [196, 492, 493]), we have opted to include these in our table. For an introduction, see [397, 108, 257]. A $C'(\frac{1}{6})$-group (also known historically as a *sixth* group) is a group satisfying the *small cancellation condition* $C'(\frac{1}{6})$. We refer the reader to Schupp [428] for definitions. Such groups were introduced by Tartakovskiĭ [473, 472, 474], and studied in depth by Greendlinger [178, 176, 177, 179]. An *n-manifold group* is a group that is the fundamental group of some compact n-manifold. Any finitely presented group appears as a 4-manifold group, as first observed by Dehn. See [20] for a survey of 3-manifolds, and [469] for information on 2-manifolds (also known as *surface* groups). The *braid group B_n* is Artin's braid group on n strands, see [304]. *Automatic* groups are too technical to define here; see [151].

A green box indicates that the given class of groups has the property, or that the property is closed under whichever operation is indicated; a red box indicates that it does not; a blue box indicates that it is an open problem, or that the problem has not yet been explored anywhere in the literature (and does not appear to have an immediate answer). In many cases, a reference is given inside a box, which indicates the first (to the best of the author's knowledge) appearance of a proof of the claim in the literature, or the first proof of a theorem which directly implies the result. For example, Poincaré [402] was the first to prove the fundamental theorem of finitely generated abelian groups, which immediately implies that all abelian groups are coherent. In those boxes where no reference is given, the proof is trivial. The non-obvious abbreviations in the table are as follows: sgmp, smmp, rsmp – whether an object from this class always has decidable subgroup, submonoid, resp. rational subset membership problem[50]; Droms raag – a raag whose underlying graph has no induced subgraph isomorphic to the path or cycle on four vertices; *stable* – whether the property is preserved by the operation in question.

[50] We emphasise that we do **not** require there to be a uniform procedure for solving the problem for the entire class, only that for every object in the class, we can find an algorithm which solves the problem for the object. For example, there is no algorithm which solves the word problem for all finite groups [457]!

Table 1.2

	Linear	Coherent	Howson	Hopfian	Res. Fin.	L. E. RF	Conj. Sep.	Word Pr.	Conj. Pr.	SGMP	SMMP	RSMP	Dio. Pr.
Free groups	🟩 [260]	🟩 [376]	🟩 [225]	🟩 [376]	🟩 [426]	🟩 [197]	🟩 [465]	🟩 [426]	🟩 [376]	🟩 [376]	🟩 [48]	🟩 [48]	🟩 [311]
Abelian groups	🟩 [458]	🟩 [402]	🟩 [402]	🟩 [321]	🟩 [321]	🟩 [402]	🟩 [321]	🟩	🟩	🟩 [402]	🟩 [190]	🟩 [190]	🟩
Nilpotent groups	🟩 [21]	🟩 [207]	🟩 [207]	🟩 [207]	🟩 [207]	🟩 [323]	🟩 [63]	🟩 [322]	🟩 [63]	🟩 [323]	🟥 [424]	🟥 [423]	🟥 [422]
Polycyclic groups	🟩 [21]	🟩 [207]	🟩 [207]	🟩 [323]	🟩 [323]	🟩 [323]	🟩 [417]	🟩 [207]	🟩 [417]	🟩 [323]	🟦 ?	🟥 [423]	🟥 [422]
Solvable groups	🟥 [36]	🟥 [61]	🟥 [358]	🟥 [127]	🟥 [36]	🟥 [36]	🟥 [253]	🟥 [253]	🟥 [253]	🟥 [253]	🟥 [253]	🟥 [253]	🟥 [422]
Simple groups	🟥	🟥 [431]	🟥 [85]	🟩	🟥	🟥	🟥	🟩 [271]	🟥 [432]	🟥 [415]	🟥 [415]	🟥 [415]	🟥 [432]
Braid groups B_n	🟩 [62]	🟥 [59]	🟥 [320]	🟩 [303]	🟩 [303]	🟥 [59]	🟦 ?	🟩 [19]	🟩 [158]	🟥 [320]	🟥 [59]	🟥 [59]	🟦 ?
2-manifold/surface groups	🟩 [401]	🟩 [232]	🟩 [175]	🟩 [33]	🟩 [33]	🟩 [434]	🟩 [335]	🟩 [131]	🟩 [131]	🟩 [434]	🟦 ?	🟦 ?	🟩 [126]
3-manifold groups	🟦 ?	🟩 [433]	🟥 [233]	🟩 [399]	🟩 [399]	🟥 [86]	🟩 [200]	🟩 [399]	🟩 [200]	🟩 [156]	🟦 ?	🟦 ?	🟦 ?
4-manifold groups	🟥 [129]	🟥 [129]	🟥 [129]	🟥 [129]	🟥 [129]	🟥 [129]	🟥 [129]	🟥 [129]	🟥 [129]	🟥 [129]	🟥 [129]	🟥 [129]	🟥 [129]
$C'(\frac{1}{6})$-groups	🟩 [13]	🟥 [419]	🟥 [454]	🟩 [435]	🟩 [13]	🟥 [419]	🟩 [356]	🟩 [474]	🟩 [178]	🟥 [419]	🟥 [419]	🟥 [419]	🟩 [126]
Hyperbolic groups	🟥 [240]	🟥 [419]	🟥 [233]	🟩 [435]	🟦 ?	🟥 [419]	🟦 ?	🟩 [188]	🟩 [188]	🟥 [419]	🟥 [419]	🟥 [419]	🟩 [126]
Automatic groups	🟥 [491]	🟥 [347]	🟥 [233]	🟥 [491]	🟥 [491]	🟥 [491]	🟥 [491]	🟩 [151]	🟦 ?	🟥 [419]	🟥 [419]	🟥 [419]	🟦 ?
RAAGS	🟩 [226]	🟥 [40]	🟥 $F_2 \times \mathbb{Z}$	🟩 [226]	🟩 [226]	🟥 [347]	🟩 [354]	🟩	🟩 [282]	🟥 [347]	🟥 [347]	🟥 [347]	🟩 [138]
Droms RAAGS	🟩 [226]	🟩 [142]	🟩 $F_2 \times \mathbb{Z}$	🟩 [226]	🟩 [226]	🟩 [344]	🟩 [354]	🟩	🟩 [282]	🟩 [239]	🟩 [288]	🟩 [288]	🟩 [138]
BS(m,n)	🟥 [41]	🟩 [56]	🟥 [242]	🟥 [41]	🟥 [41]	🟥 [41]	🟥 [41]	🟩 [301]	🟩 [18]	🟩 [57]	🟦 ?	🟦 ?	🟦 ?
BS$(1,n)$	🟩 [41]	🟩 [56]	🟥 [242]	🟩 [41]	🟩 [41]	🟩 [496]	🟩 [359]	🟩 [301]	🟩 [18]	🟩 [57]	🟩 [92]	🟩 [92]	🟩 [252]
1-relator groups	🟥 [41]	🟦 ?	🟥 [242]	🟥 [41]	🟥 [41]	🟥 [41]	🟥 [41]	🟩 [301]	🟦 ?	🟦 ?	🟥 [168]	🟥 [168]	🟦 ?
1-relator groups w. torsion	🟩 [493]	🟩 [293]	🟦 ?	🟩 [493]	🟩 [493]	🟦 ?	🟩 [355]	🟩 [301]	🟩 [372]	🟦 ?	🟦 ?	🟦 ?	🟩 [126]
2-relator groups	🟥 [206]	🟥 [347]	🟥 [358]	🟥 [206]	🟥 [206]	🟥 [206]	🟥 [206]	🟦 ?	🟦 ?	🟥 [347]	🟥 [347]	🟥 [347]	🟦 ?
Subgroup stable?	🟩	🟩	🟩	🟥 [41]	🟩	🟩 [434]	🟥 [335]	🟩	🟥 [120]	🟩	🟩	🟩	🟥 [120]
Fin. ext. stable?	🟩	🟩 [426]	🟩	🟩 [208]	🟥 [99]	🟩 [434]	🟥 [336]	🟩	🟥 [120]	🟩 [190]	🟦 ?	🟩 [190]	🟦 ?
Free prod. stable?	🟩 [486]	🟩 [243]	🟩 [30]	🟩 [134]	🟩 [173]	🟩 [425]	🟩 [465]	🟩	🟩 [465]	🟩 [346]	🟦 ?	🟩 [238]	🟩 [102]
Direct prod. stable?	🟩	🟥 [461]	🟥 [358]	🟥 [124]	🟩	🟥 [347]	🟩	🟩	🟩	🟥 [347]	🟥 [347]	🟥 [347]	🟦 ?

Chapter 2

Alternating and Free Products

Synopsis

In this chapter, we first prove a new characterisation of super-AFLs, which is based entirely on monadic rewriting systems. The remainder of the chapter is concerned with two notions: free and alternating products. Using the latter, we study the former. The latter is an original notion, and captures the informal idea of interweaving two languages, each of whose structure is similar to the word problem of a finitely generated monoid. We will prove that an alternating product of two languages inherits certain properties from its factors (Corollary 2.2.7). We also introduce a notion of the (I_1, I_2)-ancestor L^{I_1,I_2} of a language L, where I_1, I_2 are arbitrary rewriting systems, and demonstrate that, under certain conditions on the rewriting systems, this operation preserves language-theoretic properties of L (Theorem 2.2.14). We then demonstrate that the word problem of a semigroup free product of two semigroups can be obtained as an alternating product of the word problem of the semigroups (Lemma 2.3.1). This allows us to show that if the class of languages $\mathcal{C}$ is a super-AFL, then the class of semigroups with word problem in $\mathcal{C}$ is closed under taking semigroup free products. Similarly, we subsequently demonstrate that, for suitably chosen rewriting systems I_1, I_2, the word problem of a monoid free product of two monoids can be obtained as an (I_1, I_2)-ancestor of an alternating product of the word problems of the factors (Lemma 2.3.5). From this, we conclude that if $\mathcal{C}$ is a super-AFL, then the class of monoids with word problem in $\mathcal{C}$ is closed under monoid free products. This result, and the aforementioned result for semigroup free products, generalises recent results by Brough, Cain & Pfeiffer, who proved these results when $\mathcal{C}$ is the class $\mathcal{C}_{\text{cf}}$ of context-free languages. This chapter is based on the forthcoming article [385].

Let $\mathcal{C}$ be a class of languages. Recall from §1.2.1 that $\mathcal{C}$ is a *super*-AFL if it is an AFL (i.e. it is closed under homomorphism, inverse homomorphism, intersection with regular languages, union, concatenation, and the Kleene star) and if it closed under nested iterated substitution. We will now introduce an original notion, which is related to substitution. Using this, we shall see that there is an equivalent definition of super-AFL which uses rewriting systems, rather than substitutions. This will prove useful in Chapters 3 and 4.

2.1 Super-AFLs and monadic ancestors

Let $\mathcal{C}$ be a class of languages. We say that a rewriting system $\mathcal{R} \subseteq A^* \times A^*$ is $\mathcal{C}$-*ancestry preserving* if for every $L \in \mathcal{C}$ with $L \subseteq A^*$, we have $\langle L \rangle_{\mathcal{R}} \in \mathcal{C}$.

Definition 2.1.1. A class of languages $\mathcal{C}$ has the *monadic ancestor property* if every monadic $\mathcal{C}$-rewriting system is $\mathcal{C}$-ancestry preserving.

Monadic ancestors and nested iterated substitutions have a superficial resemblance. We will presently show a straightforward proposition (namely Proposition 2.1.3), which connects the monadic ancestor property with the somewhat complicated notion of nested iterated substitutions. Before this, we need a slightly technical, but straightforward to prove, lemma, concerning monadic rewriting systems.

Lemma 2.1.2. *Let $\mathcal{R}^{(m)} \subseteq A^* \times A^*$ be a monadic rewriting system. Let $A_1 \subseteq A$, and let $\mathcal{R}^{(1)} \subseteq A^* \times A^*$ be the rewriting system with rules $\{(\varepsilon, a) \mid a \in A_1\}$. Let $\mathcal{R} = \mathcal{R}^{(1)} \cup \mathcal{R}^{(m)}$. Then for any $L \subseteq A^*$, we have $\langle L \rangle_{\mathcal{R}} = \langle \langle L \rangle_{\mathcal{R}^{(m)}} \rangle_{\mathcal{R}^{(1)}}$.*

Proof. For ease of notation, we denote $\rightarrow_{\mathcal{R}^{(1)}}$ by $\rightarrow_{(1)}$; we denote $\rightarrow_{\mathcal{R}^{(m)}}$ by $\rightarrow_{(m)}$; and $\rightarrow_{\mathcal{R}}$ by $\rightarrow_{(\cup)}$. The notation is extended to $\xrightarrow{*}_{(1)}$, $\xrightarrow{*}_{(m)}$, $\xrightarrow{*}_{(\cup)}$, etc.

Let $u, w \in A^*$ be such that $w \in \langle u \rangle_{\mathcal{R}}$. Then $w \xrightarrow{*}_{(\cup)} u$, so $w \rightarrow^k_{(\cup)} u$ for some $k \geq 0$. It suffices to show that there is some $v \in A^*$ such that $w \xrightarrow{*}_{(1)} v \xrightarrow{*}_{(m)} u$. We prove this claim by induction on k. The case $k = 0$ is trivial, for then we can take $v \equiv w (\equiv u)$. Suppose $k > 0$ and let $u_0, u_1, \dots, u_k \in A^*$ be such that

$$w \equiv u_0 \rightarrow_{(\cup)} u_1 \rightarrow_{(\cup)} \cdots \rightarrow_{(\cup)} u_{k-1} \rightarrow_{(\cup)} u_k \equiv u. \tag{2.1.1}$$

By the inductive hypothesis, there exists $v_1 \in A^*$ such that $u_1 \xrightarrow{*}_{(1)} v_1 \xrightarrow{*}_{(m)} u$. If the rewriting $u_0 \rightarrow_{(\cup)} u_1$ in (2.1.1) is a rewriting $u_0 \rightarrow_{(1)} u_1$, then we may take $v \equiv v_1$. Suppose, then, that $u_0 \rightarrow_{(m)} u_1$. Let $(s, a_j) \in \mathcal{R}^{(m)}$ be the rule applied to rewrite $u_0 \rightarrow_{(m)} u_1$, where $a_j \in A \cup \{\varepsilon\}$ and $s \in A^*$ is non-empty. Write $u_1 \equiv a_1 a_2 \cdots a_{j-1} a_j a_{j+1} \cdots a_n$ for some $n \geq 0$ (where $n = 0$ means $u_1 \equiv \varepsilon$) and $a_i \in A$ for $1 \leq i \leq n, i \neq j$. Then $u_0 \equiv a_1 a_2 \cdots a_{j-1} s a_{j+1} \cdots a_n$. Now, as $u_1 \xrightarrow{*}_{(1)} v_1$, we have that

$$v_1 \equiv s_1 a_1 s_2 a_2 \cdots s_j a_j s_{j+1} \cdots s_n a_n s_{n+1}$$

where the s_i are such that $\varepsilon \xrightarrow{*}_{(1)} s_i$ for $1 \leq i \leq n+1$. Let

$$v_1' \equiv s_1 a_1 s_2 a_2 \cdots s_j s s_{j+1} \cdots s_n a_n s_{n+1}.$$

Then $v_1' \rightarrow_{(m)} v_1$ by using the rule (s, a_j), so $v_1' \xrightarrow{*}_{(m)} u$. Furthermore, as $\varepsilon \xrightarrow{*}_{(1)} s_i$ for every $1 \leq i \leq n+1$, we have $u_0 \xrightarrow{*}_{(1)} v_1'$. That is, $w \xrightarrow{*}_{(1)} v_1' \xrightarrow{*}_{(m)} u$, so we can take $v \equiv v_1'$. □

We can now proceed with the main proposition of this section.

Proposition 2.1.3. *Let $\mathcal{C}$ be an* AFL. *Then $\mathcal{C}$ is closed under nested iterated substitution if and only if it has the monadic ancestor property.*

Proof. The proof of the forward implication is the same, *mutatis mutandis*, as the proof of [69, Theorem 2.2]. For the reverse, suppose $\mathcal{C}$ has the monadic ancestor property. Let σ be a $\mathcal{C}$-substitution such that σ^∞ is a nested iterated substitution on the alphabet A, and let for every $a \in A$ the language A_a be as earlier. Let $L \subseteq A^*$ be such that $L \in \mathcal{C}$. We must show that $\sigma^\infty(L) \in \mathcal{C}$. Let $\Sigma = A \cup (\bigcup_{a \in A} A_a)$, and define a rewriting system $\mathcal{R}_\sigma \subseteq \Sigma^* \times \Sigma^*$ as the system

$$\mathcal{R}_\sigma = \bigcup_{a \in \Sigma} \bigcup_{w \in \sigma(a)} \{(w, a)\}. \tag{2.1.2}$$

This is not, in general, a monadic system, as there can be rules of the form (ε, a) in $\mathcal{R}_\sigma$. Partition $\mathcal{R}_\sigma$ as $\mathcal{R}_\sigma^{(1)} \cup \mathcal{R}_\sigma^{(m)}$, where $\mathcal{R}_\sigma^{(1)}$ is the set of all rules of the form (ε, a) for some $a \in \Sigma$, and $\mathcal{R}_\sigma^{(m)} = \mathcal{R}_\sigma - \mathcal{R}_\sigma^{(1)}$. Then $\mathcal{R}_\sigma^{(m)}$ is a monadic rewriting system. Furthermore, $\mathcal{R}_\sigma^{(m)}$ is a $\mathcal{C}$-rewriting system, as for any $a \in \Sigma$ the language of left-hand sides of rules with right-hand side a is $\sigma(a)$ for every $a \in \Sigma$, and $\sigma(a) \in \mathcal{C}$ as σ is a $\mathcal{C}$-substitution. Let $\Sigma_1 \subseteq \Sigma$ be the set of letters $\{a \mid (\varepsilon, a) \in \mathcal{R}_\sigma^{(1)}\}$.

Let $\tau \subseteq \Sigma^* \times \Sigma^*$ be defined by $\tau = \bigcup_{a \in \Sigma_1} ((a, \varepsilon)^* \cup (a, a)^*)$. Then τ is a rational transduction. Furthermore, for any language $K \subseteq \Sigma^*$, it is easy to see that $\langle K \rangle_{\mathcal{R}_\sigma^{(1)}} = \tau(K)$. That is, informally speaking, any ancestor of a word w under $\mathcal{R}_\sigma^{(1)}$ can be obtained from w by deleting some number of letters from Σ_1.

We claim $\langle L \rangle_{\mathcal{R}_\sigma} = \sigma^\infty(L)$. This would complete the proof; indeed, by Lemma 2.1.2, we have $\langle L \rangle_{\mathcal{R}_\sigma} = \langle \langle L \rangle_{\mathcal{R}_\sigma^{(m)}} \rangle_{\mathcal{R}_\sigma^{(1)}}$. Hence

$$\langle L \rangle_{\mathcal{R}_\sigma} = \langle \langle L \rangle_{\mathcal{R}_\sigma^{(m)}} \rangle_{\mathcal{R}_\sigma^{(1)}} = \tau \langle L \rangle_{\mathcal{R}_\sigma^{(m)}}. \tag{2.1.3}$$

As $\mathcal{R}_\sigma^{(m)}$ is a monadic $\mathcal{C}$-rewriting system, and $\mathcal{C}$ has the monadic ancestor property, it follows that $\langle L \rangle_{\mathcal{R}_\sigma^{(m)}} \in \mathcal{C}$. As $\mathcal{C}$ is an AFL, it is closed under rational transduction, so the right-hand side, and thereby also the left-hand side $\langle L \rangle_{\mathcal{R}_\sigma}$, of (2.1.3) is in $\mathcal{C}$. We prove the claimed equality.

($\subseteq$) If $w \in \langle L \rangle_{\mathcal{R}_\sigma}$, there exists $u \in L$ and $n \geq 0$ such that $w \to_{\mathcal{R}_\sigma}^n u$. We prove that there exists $k \geq 1$ such that $w \in \sigma^k(L)$ by induction on this n. This would prove $w \in \sigma^\infty(L)$. If $n = 0$, then $w \equiv u$, and as σ^∞ is nested we have $L \subseteq \sigma(L) \subseteq \sigma^\infty(L)$, and we are done. Assume $n > 0$. Then there exists $w' \in \langle u \rangle_{\mathcal{R}_\sigma}$ such that $w \to_{\mathcal{R}_\sigma} w' \to_{\mathcal{R}_\sigma}^{n-1} u$. By the inductive hypothesis there exist $k' \geq 1$ such that $w' \in \sigma^{k'}(L)$. As $w \to_{\mathcal{R}_\sigma} w'$, there is a rule $(r, s) \in \mathcal{R}_\sigma$ such that $w \equiv w_0 r w_1$ and $w' \equiv w_0 s w_1$. But from (2.1.2), we have $s \in \Sigma$ and $r \in \sigma(s)$. Hence

$$w \equiv w_0 r w_1 \in w_0 \sigma(s) w_1 \subseteq \sigma(w_0)\sigma(s)\sigma(w_1) = \sigma(w_0 s w_1) = \sigma(w') \subseteq \sigma(\sigma^{k'}(L)).$$

Note that the inclusion $\{x\} \subseteq \sigma(x)$ for $x \in A^*$ follows from the fact that σ^∞ is nested. As $\sigma(\sigma^{k'}(L)) = \sigma^{k'+1}(L)$, we can take $k = k' + 1$.

($\supseteq$) Suppose $w \in \sigma^\infty(L)$. Then there exists $n \geq 0$ such that $w \in \sigma^n(L)$. We prove $w \in \langle L \rangle_{\mathcal{R}_\sigma}$ by induction on this n. If $n = 0$, then (by our convention), $w \in L$, so there is nothing to show. Assume $n > 0$. Then there exists some $u \in \sigma^{n-1}(L)$ such that $w \in \sigma(u)$.

By the inductive hypothesis $u \in \langle L \rangle_{\mathcal{R}_\sigma}$. Write $u \equiv a_1 a_2 \cdots a_k$, where $k \geq 1$ and $a_i \in A$ for

$1 \leq i \leq k$. Then $\sigma(u) = \sigma(a_1)\sigma(a_2)\cdots\sigma(a_k)$. Hence $w \equiv w_1 w_2 \cdots w_k$ for some $w_i \in \sigma(a_i)$ for $1 \leq i \leq k$. In particular, $(w_i, a_i) \in \mathcal{R}_\sigma$ for every $1 \leq i \leq k$. Hence we find

$$w \equiv w_1 w_2 \cdots w_k \xrightarrow{*}_{\mathcal{R}_\sigma} a_1 a_2 \cdots a_k \equiv u,$$

so $w \in \langle u \rangle_{\mathcal{R}_\sigma} \subseteq \langle \langle L \rangle_{\mathcal{R}_\sigma} \rangle_{\mathcal{R}_\sigma} = \langle L \rangle_{\mathcal{R}_\sigma}$, which is what was to be shown. □

We have an immediate corollary, as the class $\mathcal{C}_{\mathrm{cf}}$ is a super-AFL.

Corollary. *The class $\mathcal{C}_{\mathrm{cf}}$ of context-free languages has the monadic ancestor property.*

Furthermore, we have an equivalent definition for a super-AFL: a *super*-AFL is an AFL which has the monadic ancestor property. This is a far more directly combinatorial definition of super-AFLs. We shall in the sequel *exclusively* use this definition; no mention of nested iterated substitutions will henceforth be made in this thesis.

2.2 Alternating products

We shall in this section describe an operation on certain languages, which mimics the operation of the free product of monoids. Throughout this section, we will fix an alphabet A and let $\#$ be a symbol disjoint from A.

Definition 2.2.1. Let $L \subseteq A^*\#A^*$. We say that L is *concatenation-closed* (*with respect to* $\#$) if

$$u_1\#v_1 \in L \text{ and } u_2\#v_2 \in L \implies u_1u_2\#v_2v_1 \in L.$$

Note the order of the concatenation to the right of the $\#$ symbol. A typical example of a concatenation-closed language is the word problem for a finitely generated monoid.

Example 2.2.2. Let $L = \{u\#v \mid u, v \in \{a,b\}^* \mid \sigma_a(u) = \sigma_a(v), \sigma_b(u) = \sigma_b(v)\}$, and consider two elements $u_1\#v_1, u_2\#v_2 \in L$, where $u_1, u_2, v_1, v_2 \in \{a,b\}^*$ satisfy the given condition. In particular

$$\sigma_a(u_1u_2) = \sigma_a(u_1) + \sigma_a(u_2) = \sigma_a(v_1) + \sigma_a(v_2) = \sigma_a(v_1v_2) = \sigma_a(v_2v_1)$$

and hence $u_1u_2\#v_2v_1 \in L$. Thus L is concatenation-closed. Note that L is really the word problem of the free commutative monoid $\mathbb{N} \times \mathbb{N} = \mathrm{Mon}\langle a, b \mid ab = ba \rangle$.

Example 2.2.3. Let M be a finitely generated monoid with finite generating set A. If $u_1\#v_1$ and $u_2\#v_2$ are in WP_A^M, then $u_1 =_M v_1^{\mathrm{rev}}$ and $u_2 =_M v_2^{\mathrm{rev}}$. Thus

$$u_1u_2 =_M v_1^{\mathrm{rev}} v_2^{\mathrm{rev}} \equiv (v_2v_1)^{\mathrm{rev}}$$

from which it follows that

$$u_1u_2\#((v_2v_1)^{\mathrm{rev}})^{\mathrm{rev}} \equiv u_1u_2\#v_2v_1 \in \mathrm{WP}_A^M.$$

Hence WP_A^M is concatenation-closed.

Given two languages in $A^*\#A^*$, we will introduce a new and quite general operation for combining them into a single language.

Definition 2.2.4 (Alternating product)**.** Let $L_1, L_2 \subseteq A^*\#A^*$ be concatenation-closed languages. Then the *alternating product* of L_1 and L_2, denoted $L_1 \star L_2$, is defined as the language consisting of all words of the form $u_1u_2\cdots u_k\#v_k\cdots v_2v_1$, such that for all $i \geq 1$ we have that $u_i\#v_i \in L_{X(i)}$, where the parametrisation X is such that $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$.

Note that $L_1 \star L_2 \subseteq A^*\#A^*$. Note also that $L_1 \star L_2 = L_2 \star L_1$. Informally, $L_1 \star L_2$ is obtained from L_1 and L_2 by alternatingly writing terms from either language; the condition of the two languages being concatenation-closed, as we shall see, ensures that this alternating process does not break down if one chooses to write two terms from the same language after one another. We remark that the notation $\star$ is slightly abusive, as it suppresses any mention of the symbol $\#$, which is central to the definition; but this is always clear from context. Alternating products are modelled on the free product, as the following example suggests.

Example 2.2.5. Let $L_1 = \{a^n\#a^n \mid n \geq 0\}$ and $L_2 = \{b^n\#b^n \mid n \geq 0\}$. Then

$$\begin{aligned} L_1 \star L_2 &= \{a^{n_1}b^{n_2}\cdots a^{n_k}\#a^{n_k}\cdots b^{n_2}a^{n_1} \mid k \geq 0,\ n_1, n_2, \dots, n_{k-1} \geq 1, n_k \geq 0\} \\ &= \{w\#w^{\text{rev}} \mid w \in \{a,b\}^*\} \\ &= \mathrm{WP}_{\{a,b\}}^{\{a,b\}^*}. \end{aligned}$$

Clearly, $L_1 = \mathrm{WP}_{\{a\}}^{\{a\}^*}$ and $L_2 = \mathrm{WP}_{\{b\}}^{\{b\}^*}$. More generally, if A, B are disjoint alphabets and if $L_1 = \mathrm{WP}_A^{A^*}$ and $L_2 = \mathrm{WP}_B^{B^*}$, then $L_1 \star L_2 = \mathrm{WP}_{A\cup B}^{A^* * B^*}$, where $A^* * B^* = (A \cup B)^*$ denotes the monoid free product of the two free monoids A^* and B^*. △

Proposition 2.2.6. *Let $\mathcal{C}$ be a class of languages with the monadic ancestor property containing all singleton languages, and which is closed under union. Let $L_1, L_2 \subseteq A^*\#A^*$ be concatenation-closed languages. Then $L_1, L_2 \in \mathcal{C} \implies L_1 \star L_2 \in \mathcal{C}$.*

Proof. For $i = 1, 2$, let $\mathcal{R}_i = \{(w \to \#) \mid w \in L_i\}$. Then by assumption, $\mathcal{R}_i$ is a monadic $\mathcal{C}$-rewriting system. As $\mathcal{C}$ is closed under union, $\mathcal{R} := \mathcal{R}_1 \cup \mathcal{R}_2$ is also a monadic $\mathcal{C}$-rewriting system. As $\mathcal{C}$ has the monadic ancestor property and $\# \in \mathcal{C}$, the language $L = \langle\#\rangle_{\mathcal{R}}$ is also in $\mathcal{C}$. We will show that $L = (L_1 \star L_2) \cup \{\#\}$.

($\subseteq$). Suppose $w \in (L_1 \star L_2) \cup \{\#\}$. Clearly $\# \in L$, so suppose $w \in L_1 \star L_2$. Write $w \equiv u_1u_2\cdots u_k\#v_k\cdots v_2v_1$ where $u_i\#v_i \in L_{X(i)}$, and where $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$. We will prove that $w \in L$ by induction on k. If $k = 0$, then $w \equiv \#$, and there is nothing to show. Suppose that $k > 0$, and that the claim is true for all alternating products which can be written using fewer than k alternating factors. As $u_k\#v_k \in L_{X(k)}$, we have $(u_k\#v_k \to \#) \in \mathcal{R}_{X(k)} \subseteq \mathcal{R}$. In particular

$$w \equiv u_1u_2\cdots u_k\#v_k\cdots v_2v_1 \to_{\mathcal{R}} u_1u_2\cdots u_{k-1}\#v_{k-1}\cdots v_2v_1.$$

By the inductive hypothesis the right-hand side $u_1u_2\cdots u_{k-1}\#v_{k-1}\cdots v_2v_1$ lies in $\langle\#\rangle_{\mathcal{R}}$. Hence we also have $w \in \langle\#\rangle_{\mathcal{R}} = L$, which is what was to be shown.

($\supseteq$). Suppose $w \in L$. Then $w \xrightarrow{*}_{\mathcal{R}} \#$. We will prove by induction on the number of steps k in this rewriting that $w \in (L_1 \star L_2) \cup \{\#\}$. If $k = 0$, then $w \equiv \#$, and we are done. Suppose that $k > 0$ and that the claim is true for all words which rewrite in fewer than k steps. As

$w \rightarrow_{\mathcal{R}}^{k} \#$, we can find some word $w' \in A^*\#A^*$ such that

$$w \rightarrow_{\mathcal{R}} w' \rightarrow_{\mathcal{R}}^{k-1} \#.$$

By the inductive hypothesis, $w' \in (L_1 \star L_2) \cup \{\#\}$. Thus we can write

$$w' \equiv u_1 u_2 \cdots u_n \# v_n \cdots v_2 v_1,$$

where either all u_i, v_i are empty, or else $u_i \# v_i \in L_{X(i)}$ where $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$. Let $(u'\#v' \to \#)$ be the rule of $\mathcal{R}$ by which we rewrite $w \rightarrow_{\mathcal{R}} w'$. As $\#$ is disjoint from A, w' contains exactly one occurrence of $\#$, namely the one specified above. Thus

$$w \equiv u_1 u_2 \cdots u_n (u'\#v') v_n \cdots v_2 v_1.$$

As $(u'\#v' \to \#) \in \mathcal{R}$, either $u'\#v' \in L_{X(n+1)}$ or $u'\#v' \in L_{X(n)}$. In the former case, the above expression for w is clearly an alternating product, so $w \in L_1 \star L_2$. In the latter case, as $L_{X(n)}$ is concatenation-closed, we have that

$$u_n \# v_n, u'\#v' \in L_{X(n)} \quad \Longrightarrow \quad u_n u' \# v' v_n \in L_{X(n)}$$

and hence $w \in L_1 \star L_2$. As w was arbitrary, we conclude $L \subseteq (L_1 \star L_2) \cup \{\#\}$.

Hence we have proved that $L = (L_1 \star L_2) \cup \{\#\}$. Now as observed earlier, $L \in \mathcal{C}$, and hence if $\# \in L_1 \star L_2$, then $L = L_1 \star L_2$, and we are done. On the other hand, if $\# \not\in L_1 \star L_2$, then it follows that

$$L_1 \star L_2 = L \setminus \{\#\} = L \cap (A^*\#A^* \setminus \{\#\}) = L \cap \left((A^*\#A^+) \cup (A^+\#A^*) \right),$$

and as $\mathcal{C}$ is closed under intersection with regular languages, we thereby conclude from $L \in \mathcal{C}$ that $L_1 \star L_2 \in \mathcal{C}$, as desired. □

As super-AFLs have all the necessary closure properties to apply Proposition 2.2.6 (by definition, or by Proposition 2.1.3), we conclude:

Corollary 2.2.7. *Let $\mathcal{C}$ be a super-*AFL. *Let A be an alphabet, and $\#$ a symbol not in A. Let $L_1, L_2 \subseteq A^*\#A^*$ be concatenation-closed languages. Then $L_1, L_2 \in \mathcal{C} \implies L_1 \star L_2 \in \mathcal{C}$.*

This corollary will prove very useful; alternating products model semigroup free products, and we shall use them to good effect presently. Before this, we shall introduce another tool of a slightly more technical nature, which will be used in order to model monoid free products.

2.2.1 Ancestors

In some special cases, e.g. Example 2.2.5, the alternating product can be used to model the monoid free product, much the same as it models the semigroup free product. However, this example does not generalise very far. Indeed, if M, N are two monoids generated by two (disjoint) finite sets A, B, respectively, then in general $\mathrm{WP}_{A \cup B}^{M * N} \neq \mathrm{WP}_{A}^{M} \star \mathrm{WP}_{B}^{N}$. The reason for this is quite simple: there may be non-empty words in M or N representing the identity of M (resp. N), which thus – by the definition of the monoid free product – represents the identity of $M * N$. Concretely, we can take

$$M = \text{Mon}\langle a \mid a^2 = 1\rangle, \text{ and}$$
$$N = \text{Mon}\langle b \mid b^2 = 1\rangle,$$

in which case $M * N = \text{Mon}\langle a, b \mid a^2 = 1, b^2 = 1\rangle$. But now we have

$$bab^2a =_{M*N} b$$

and so

$$bab^2a\#b \in \text{WP}^{M*N}_{\{a,b\}}$$

despite the fact that $bab^2a\#b \notin \text{WP}^M_{\{a\}} \star \text{WP}^N_{\{b\}}$. In particular, we have, unlike for semigroup free products, that

$$\text{WP}^{M*N}_{\{a,b\}} \neq \text{WP}^M_{\{a\}} \star \text{WP}^N_{\{b\}}\ .$$

The problem is plainly that we may insert words equal to 1 wherever we please.

We will now remedy the above situation, by a general operation involving $\mathcal{C}$-ancestry preserving rewriting systems. The main idea is the following: for a language $L \subseteq A^*\#A^*$ and two rewriting systems I', I'', the language $L^{I',I''}$, which we will presently define, will be obtained by taking ancestors under I' to the left-hand side of the $\#$ in L, and ancestors under I'' to the right-hand side of the $\#$. Of course, in full generality, operations which only deal with one side of the $\#$ in such a language are not particularly well-behaved. For example, take reversal; given the language $\{w\#w^{\text{rev}} \mid w \in A^*\}$ of palindromes – the prototypical context-free language – if one were to reverse only on one side of the $\#$ in the language, one would find the language $\{w\#w \mid w \in A^*\}$ which is not context-free, which can be seen by applying the pumping lemma to the word $a^nb^n\#a^nb^n$ (the language is, however, context-sensitive). On the other hand, we shall prove (Lemma 2.2.13) that, as long as the rewriting systems I', I'' are well-behaved, the language $L^{I',I''}$ is, too, with respect to preserving the class of a given language. We now make this formal.

Definition 2.2.8. Let $I', I'' \subseteq A^* \times A^*$ be rewriting systems. Let $L \subseteq A^*\#A^*$. Then we define the language

$$L^{I',I''} = \{w_1\#w_2 \mid \exists u\#v \in L \text{ such that } w_1 \in \langle u\rangle_{I'}, w_2 \in \langle v\rangle_{I''}\},$$

and call this the (I', I'')-*ancestor* of L.

This definition is an interpretation in the language of rewriting systems of the possibility of the left-hand side and the right-hand side of the letter $\#$ to be altered independently of one another inside a language $L \subseteq A^*\#A^*$. We give an example to clarify what we mean.

Example 2.2.9. Let $A = \{a, b, c\}$, let $L = \{a\#a\}$, and

$$I' = \{(b^n \to a) \mid n \geq 1\}, I'' = \{(c^n \to a) \mid n \geq 1\}.$$

Then we find

$$L^{I',I''} = \{b^{n_1}\#c^{n_2} \mid n_1, n_2 \geq 1\} \cup \{b^{n_1}\#a \mid n_1 \geq 1\} \cup \{a\#c^{n_2} \mid n_2 \geq 1\} \cup \{a\#a\}.$$

Thus we have e.g. $a\#cc, bbb\#c \in L^{I',I''}$. This means that there is no need for the rewritings performed by I' and I'' to be "synchronised" with one another. $\triangle$

The reader more comfortable with rewriting, rather than taking ancestors, may of course rewrite the above definition to the equivalent

$$L^{I',I''} = \{w_1\#w_2 \mid \exists u\#v \in L \text{ such that } w_1 \xrightarrow{*}_{I'} u, w_2 \xrightarrow{*}_{I''} v\}.$$

Our main goal is to prove a general preservation property of taking ancestors. We will first prove a weak form of preservation, in which certain alphabets (A_1 and A_2) are disjoint. This will then be extended to the general case (i.e. Lemma 2.2.13), in which no assumption on disjointness will be needed. The proof of the general case will require the special case as a basis.

Lemma 2.2.10. *Let $\mathcal{C}$ be a class of languages. Let $\mathcal{R}_1 \subseteq A_1^* \times A_1^*$ and $\mathcal{R}_2 \subseteq A_2^* \times A_2^*$ be two rewriting systems with $A_1 \cap A_2 = \varnothing$. If $\mathcal{R}_1$ and $\mathcal{R}_2$ are $\mathcal{C}$-ancestry preserving, then for every language $L \subseteq A_1^*\#A_2^*$ we have $L \in \mathcal{C} \implies L^{\mathcal{R}_1,\mathcal{R}_2} \in \mathcal{C}$.*

Proof. We claim that $L^{\mathcal{R}_1,\mathcal{R}_2} = \langle\langle L\rangle_{\mathcal{R}_2}\rangle_{\mathcal{R}_1}$ (and indeed $\langle\langle L\rangle_{\mathcal{R}_1}\rangle_{\mathcal{R}_2}$). To prove this, first a word $w_1\#w_2$ is in $L^{\mathcal{R}_1,\mathcal{R}_2}$ if and only if there exists $u\#v \in L$ such that $w_1 \xrightarrow{*}_{\mathcal{R}_1} u$ and $w_2 \xrightarrow{*}_{\mathcal{R}_2} v$. As A_1 and A_2 are disjoint, and $\mathcal{R}_1 \subseteq A_1^* \times A_1^*$ and $\mathcal{R}_2 \subseteq A_2^* \times A_2^*$, it follows that

$$\begin{aligned} w_1\#w_2 \in L^{\mathcal{R}_1,\mathcal{R}_2} \quad &\iff \quad \exists u\#v \in L : \begin{cases} w_1\#w_2 \xrightarrow{*}_{\mathcal{R}_1} u\#w_2, \text{ and} \\ u\#w_2 \xrightarrow{*}_{\mathcal{R}_2} u\#v \end{cases} \\ &\iff \quad \exists u\#v \in L : w_1\#w_2 \xrightarrow{*}_{\mathcal{R}_1} u\#w_2 \xrightarrow{*}_{\mathcal{R}_2} u\#v \\ &\iff \quad w_1\#w_2 \in \langle\langle L\rangle_{\mathcal{R}_2}\rangle_{\mathcal{R}_1}. \end{aligned}$$

Thus $L^{\mathcal{R}_1,\mathcal{R}_2} = \langle\langle L\rangle_{\mathcal{R}_2}\rangle_{\mathcal{R}_1}$. As $\mathcal{R}_2$ is $\mathcal{C}$-ancestry preserving, we have $\langle L\rangle_{\mathcal{R}_2} \in \mathcal{C}$. As $\mathcal{R}_1$ is $\mathcal{C}$-ancestry preserving, thus also $L^{\mathcal{R}_1,\mathcal{R}_2} = \langle\langle L\rangle_{\mathcal{R}_2}\rangle_{\mathcal{R}_1} \in \mathcal{C}$. □

Example 2.2.11. Let $A = \{b, c, b', c'\}$, and let $\mathcal{R}_1 = \{(b' \to b)\}$ and $\mathcal{R}_2 = \{(c' \to c)\}$. Let $L = \{b\#c\}$. Then

$$L^{\mathcal{R}_1,\mathcal{R}_2} = \{b'\#c', b'\#c, b\#c', b\#c\},$$
$$\langle\langle L\rangle_{\mathcal{R}_2}\rangle_{\mathcal{R}_1} = \langle\{b\#c', b\#c\}\rangle_{\mathcal{R}_1} = \{b'\#c', b'\#c, b\#c', b\#c\}.$$

This gives an example of the proof of the lemma, taking $A_1 = \{b, b'\}$ and $A_2 = \{c, c'\}$.

Before we can prove the key property of (I', I'')-ancestors, we introduce a convenient "alphabet–changing" procedure for languages in $A^*\#A^*$. Let $L \subseteq A^*\#A^*$. Let A_ℓ, A_r be two new alphabets in bijective correspondence with A via maps $a \mapsto a_\ell$ and $a \mapsto a_r$. Extend this to homomorphisms $\varphi_\ell \colon A^* \to A_\ell^*$ and $\varphi_r \colon A^* \to A_r^*$. For convenience, for $u \in A^*$ we will denote $\varphi_\ell(u)$ (resp. $\varphi_r(u)$) as u_ℓ (resp. u_r). We set $\varepsilon_\ell \equiv \varepsilon_r \equiv \varepsilon$. For a rewriting system $\mathcal{R} \subseteq A^* \times A^*$, we will extend this notation to let $\mathcal{R}_\ell$ denote the system

$$\mathcal{R}_\ell := \{(s_\ell \to t_\ell) \mid (s \to t) \in \mathcal{R}\},$$

and $\mathcal{R}_r$ is defined entirely analogously. The following is more or less obvious.

Lemma 2.2.12. *Let $\mathcal{C}$ be a class of languages closed under homomorphism. If $\mathcal{R}$ is $\mathcal{C}$-ancestry preserving, then so too are $\mathcal{R}_\ell$ and $\mathcal{R}_r$.*

Proof. We suppress the alphabets over which the rewriting systems are written for notational convenience. Let $L \in \mathcal{C}$. Then, as $\mathcal{R}$ is $\mathcal{C}$-ancestry preserving, we have that $\langle L\rangle_{\mathcal{R}} \in \mathcal{C}$. Recall that $\varphi_\ell \colon A^* \to A_\ell^*$ is an isomorphism of free monoids; in particular, we have

$$usv \to_{\mathcal{R}} utv,$$

if and only if

$$\varphi_\ell(usv) \equiv \varphi_\ell(u)s_\ell\varphi_\ell(v) \to_{\mathcal{R}_\ell} \varphi_\ell(u)t_\ell\varphi_\ell(v) \equiv \varphi_\ell(utv),$$

and hence we conclude, by induction on the number of rewritings, that

$$\langle L \rangle_{\mathcal{R}_\ell} = \varphi_\ell(\langle L \rangle_{\mathcal{R}}).$$

As $\langle L \rangle_{\mathcal{R}} \in \mathcal{C}$ and $\mathcal{C}$ is closed under homomorphisms, we conclude that $\langle L \rangle_{\mathcal{R}_\ell} \in \mathcal{C}$. As L was arbitrary, $\mathcal{R}_\ell$ is $\mathcal{C}$-ancestry preserving. The proof for $\mathcal{R}_r$ is entirely analogous. □

To generalise Lemma 2.2.10 to arbitrary languages $L^{I',I''}$ for less restrictive rewriting systems I', I'', we will consider the alphabets A_ℓ, A_r defined above and introduce a relation

$$\mu_{\ell,r} \subseteq (A \cup \{\#\})^* \times (A_\ell \cup A_r \cup \{\#\})^*$$

defined as

$$\mu_{\ell,r} = \left(\bigcup_{a \in A} (a, a_\ell) \right)^* (\#, \#) \left(\bigcup_{a \in A} (a, a_r) \right)^*.$$

Note the implicit dependency on A (and even $\#$) in the definition of $\mu_{\ell,r}$; indeed, we might more accurately denote it as $\mu_{\ell,r}^A$ or even $\mu_{\ell,r}^{A,\#}$, but this would quickly (as we shall soon see) become monstrously cumbersome. We therefore simply write $\mu_{\ell,r}$ for this map (the alphabet A and the symbol $\#$ will always be clear from context. Now, $\mu_{\ell,r}$ is clearly (compare with Example 1.2.4) a rational subset of $(A \cup \{\#\})^* \times (A_\ell \cup A_r \cup \{\#\})^*$. Indeed, it is of the form X^*xY^*, where x is the singleton $\{(\#, \#)\}$ and X, Y are the finite sets above. Therefore $\mu_{\ell,r}$ is a rational transduction.

When applied to a language in $A^*\#A^*$, this transduction has the effect of changing the alphabet to the left of the $\#$ to A_ℓ, and the alphabet to the right of the $\#$ to A_r (compare with Example 1.2.4). For example,

$$\mu_{\ell,r}(abc\#cc) = \{a_\ell b_\ell c_\ell \# c_r c_r\},$$
$$\mu_{\ell,r}(\{a^n\#a^n \mid n \geq 0\}) = \{a_\ell^n \# a_r^n \mid n \geq 0\}.$$

Analogously, there is a transduction, which we will for obvious reasons denote by $\mu_{\ell,r}^{-1}$, which when applied to a language in $A_\ell^*\#A_r^*$ has the effect of removing the $_r$ and $_\ell$-symbols on each letter. Note that $\mu_{\ell,r}$ is bijective on subsets of $A^* \times A^*$ and $\mu_{\ell,r}^{-1}$ is bijective on subsets of $A_\ell^*\#A_r^*$.

The rational transductions $\mu_{\ell,r}$ and $\mu_{\ell,r}^{-1}$ will be key in generalising Lemma 2.2.10 to the following lemma, which reveals the strength of (I', I'')-ancestors.

Lemma 2.2.13. *Let $\mathcal{C}$ be a class of languages closed under rational transductions. Let $L \subseteq A^*\#A^*$, and let $I', I'' \subseteq A^* \times A^*$ be rewriting systems. If I' and I'' are $\mathcal{C}$-ancestry preserving, then*

$$L \in \mathcal{C} \implies L^{I',I''} \in \mathcal{C}.$$

Proof. Define the language

$$\mathcal{L}_1 := \mu_{\ell,r}^{-1}\Big(\big(\mu_{\ell,r}(L)\big)^{I'_\ell, I''_r} \Big)$$

with the transductions $\mu_{\ell,r}, \mu_{\ell,r}^{-1}$ (and the alphabets A_ℓ, A_r) defined as above. We claim that

$$\mathcal{L}_1 = L^{I',I''}.$$

This would complete the proof, in view of the following argument: first, $\mu_{\ell,r}(L) \subseteq A_\ell^* \# A_r^*$, and furthermore, as $L \in \mathcal{C}$ and $\mathcal{C}$ is closed under rational transductions, we have that $\mu_{\ell,r}(L) \in \mathcal{C}$. As I', I'' are $\mathcal{C}$-ancestry preserving, so too are I'_ℓ and I''_r by Lemma 2.2.12. Hence, as $A_\ell \cap A_r = \varnothing$, and $\mu_{\ell,r}(L) \subseteq A_\ell^* \# A_r^*$, we have by Lemma 2.2.10 that $\mu_{\ell,r}(L)^{I'_\ell, I''_r} \in \mathcal{C}$. Finally, by another application of the closure of $\mathcal{C}$ under rational transductions, this time $\mu_{\ell,r}^{-1}$, we have $\mathcal{L}_1 \in \mathcal{C}$.

First, note that by the definition of (I', I'')-ancestors, we have

$$w_1 \# w_2 \in L^{I',I''} \qquad \Longleftrightarrow \qquad \exists u\#v \in L \text{ s.t. } \begin{cases} w_1 \xrightarrow{*}_{I'} u, \text{ and} \\ w_2 \xrightarrow{*}_{I''} v. \end{cases}$$

Now, as $\mu_{\ell,r}$ is bijective on subsets of $A^* \times A^*$, and $\mu_{\ell,r}(u\#v) = \{u_\ell \# v_\ell\}$, we have

$$\exists u\#v \in L \text{ s.t. } \begin{cases} w_1 \xrightarrow{*}_{I'} u \\ w_2 \xrightarrow{*}_{I''} v \end{cases} \qquad \Longleftrightarrow \qquad \exists u_\ell \# v_r \in \mu_{\ell,r}(L) \text{ s.t. } \begin{cases} w_1 \xrightarrow{*}_{I'} \varphi_\ell^{-1}(u_\ell) \\ w_2 \xrightarrow{*}_{I''} \varphi_r^{-1}(v_r). \end{cases}$$

As in the proof of Lemma 2.2.12, as φ_ℓ, φ_r are isomorphisms of free monoids, we find

$$\exists u_\ell \# v_r \in \mu_{\ell,r}(L) \text{ s.t. } \begin{cases} w_1 \xrightarrow{*}_{I'} \varphi_\ell^{-1}(u_\ell) \\ w_2 \xrightarrow{*}_{I''} \varphi_r^{-1}(v_r). \end{cases} \Longleftrightarrow \exists u_\ell \# v_r \in \mu_{\ell,r}(L) \text{ s.t. } \begin{cases} \varphi_\ell(w_1) \xrightarrow{*}_{I'_\ell} u_\ell \\ \varphi_r(w_2) \xrightarrow{*}_{I''_r} v_r. \end{cases}$$

But now this right-most condition is equivalent, by the definition of (I'_ℓ, I''_r)-ancestors, to $\varphi_\ell(w_1)\#\varphi_r(w_2) \in (\mu_{\ell,r}(L))^{I'_\ell, I''_r}$. As $\varphi_\ell(w_1)\#\varphi_r(w_2) \equiv \mu_{\ell,r}(w_1\#w_2)$, thus we have found, by combining all biconditionals above, that

$$w_1 \# w_2 \in L^{I',I''} \qquad \Longleftrightarrow \qquad \mu_{\ell,r}(w_1\#w_2) \in (\mu_{\ell,r}(L))^{I'_\ell, I''_r}.$$

As $\mu_{\ell,r}$ is bijective on subsets of $A^*\#A^*$, it follows that

$$w_1 \# w_2 \in L^{I',I''} \qquad \Longleftrightarrow \qquad w_1 \# w_2 \in \mu_{\ell,r}^{-1}\Big((\mu_{\ell,r}(L))^{I'_\ell, I''_r} \Big) = \mathcal{L}_1,$$

which is to say: $L^{I',I''} = \mathcal{L}_1$. □

Of course, as all monadic $\mathcal{C}$-rewriting systems are $\mathcal{C}$-ancestry preserving whenever $\mathcal{C}$ has the monadic ancestor property, we have proved the following.

Theorem 2.2.14. *Let $\mathcal{C}$ be a super-*AFL*. Let $L \subseteq A^*\#A^*$ be a language, and let I', I'' be monadic $\mathcal{C}$-rewriting systems. Then $L \in \mathcal{C} \implies L^{I',I''} \in \mathcal{C}$.*

We are now ready to apply the theory of (I', I'')-ancestors to monoid free products. We shall also revisit and pay tribute to these ancestors, as well as alternating products, in Chapter 4, when dealing with *weakly compressible* monoids.

2.3 Free products

In this section, we will extensively describe the language-theoretic properties of free products. We begin by considering the word problem – as a language – for the semigroup free product of two semigroups. We shall see that the alternating product of the word problems of M and N then completely captures the structure of the semigroup free product of M and N. To capture the structure of the monoid free product, a similar idea forms the basic layer, but we will additionally need ancestors to fully describe the language theoretic properties of such products.

2.3.1 Semigroup free products

All the hard work to describe the language theory of semigroup free products has already been carried out in the previous sections, in the theory of alternating products. To see this, we will begin by translating the algebraic structure of semigroup free products into language-theoretic terms.

Lemma 2.3.1. *Let S_1, S_2 be two semigroups, generated by some finite disjoint sets A_1 resp. A_2. Let $S_1 * S_2$ denote their semigroup free product. Then*

$$\mathrm{WP}_{A_1 \cup A_2}^{S_1 * S_2} = \mathrm{WP}_{A_1}^{S_1} \star \mathrm{WP}_{A_2}^{S_2}$$

Proof. First, we note that $\mathrm{WP}_{A_1}^{S_1}$ and $\mathrm{WP}_{A_2}^{S_2}$ are both concatenation-closed, so the language $\mathrm{WP}_{A_1}^{S_1} \star \mathrm{WP}_{A_2}^{S_2}$ is well-defined. Now an arbitrary word $w \equiv u \# v^{\mathrm{rev}}$ from $(A_1 \cup A_2)^+$ is in $\mathrm{WP}_{A_1 \cup A_2}^{S_1 * S_2}$ if and only if $u =_{S_1 * S_2} v$, so by Lemma 1.1.1, we have that there exist unique factorisations

$$u \equiv u_0 u_1 \cdots u_n$$
$$v \equiv v_0 v_1 \cdots v_n$$

such that $u_i, v_i \in A_{X(i)}^+$ and $u_i =_{S_{X(i)}} v_i$ for all $i \geq 0$, where $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$. But this means

$$w \equiv u_0 u_1 \cdots u_n \# (v_1 v_2 \cdots v_n)^{\mathrm{rev}} \equiv u_0 u_1 \cdots u_n \# v_n^{\mathrm{rev}} \cdots v_2^{\mathrm{rev}} v_1^{\mathrm{rev}}.$$

As $u_i =_{S_{X(i)}} v_i$ for all $i \geq 0$, we have that $u_i \# (v_i)^{\mathrm{rev}} \in \mathrm{WP}_{A_{X(i)}}^{S_{X(i)}}$ for all $i \geq 0$. We conclude, by definition of the alternating product, that this is equivalent to $w \in \mathrm{WP}_{A_1}^{S_1} \star \mathrm{WP}_{A_2}^{S_2}$, as was to be shown. □

Hence, as taking alternating products preserve language-theoretic properties of the languages by Corollary 2.2.7, we conclude the following.

Theorem 2.3.2. *Let $\mathcal{C}$ be a super-*AFL*. Then the class of semigroups with word problem in $\mathcal{C}$ is closed under taking (semigroup) free products.*

As the class $\mathcal{C}_{\mathrm{cf}}$ of context-free languages is a super-AFL, we conclude the following from Theorem 2.3.2; this result was originally proved using very different methods (pushdown automata) by Brough, Cain & Pfeiffer [80].

Corollary (Brough, Cain & Pfeiffer, 2019)**.** *The class of semigroups with context-free word problem is closed under taking semigroup free products.*

Of course, we obtain the following corollary, as the class $\mathcal{C}_{\text{ind}}$ of indexed languages is also a super-AFL which is closed under reversal.

Corollary 2.3.3. *The class of semigroups with indexed word problem is closed under taking semigroup free products.*

As far as the author is aware, this statement has not previously appeared in the literature. We are now ready to turn towards monoid free products, which shall require the use of ancestors.

2.3.2 Monoid free products

The identification of the identity elements of the factors in the monoid free product of two monoids means that we have little reason to conclude that the word problem of a monoid free product is the alternating product of its factors' word problems; indeed, it is not too hard to see that this is true if and only if no non-trivial word is equal to 1 in the respective factors (for example if both sides of all defining relations are non-empty). To resolve this, we use the alternating product with insertions. The insertions in the product $M_1 * M_2$ of two monoids M_1, M_2 generated by A_1, A_2 will come from the sets $\mathrm{IP}_{A_1}^{M_1}$ and $\mathrm{IP}_{A_2}^{M_2}$. Recall the definition

$$\mathrm{IP}_A^M = \{w \in A^* \mid w =_M 1\}$$

of the *identity problem* of M (see §1.2.3). We first need two quick properties.

Lemma 2.3.4. *Let $\mathcal{C}$ be a class of languages closed under homomorphism and intersection with regular languages. Let M be a monoid, finitely generated by A, with $\mathrm{WP}_A^M \in \mathcal{C}$. Then $\mathrm{IP}_A^M \in \mathcal{C}$.*

Proof. Note that $\mathrm{WP}_A^M \cap A^*\#$ is the language

$$L = \{w\# \mid w \in A^*, w =_M 1\}.$$

This language is in $\mathcal{C}$ as $\mathcal{C}$ is closed under intersection with regular languages. Let $\varphi\colon (A \cup \{\#\})^* \to (A \cup \{\#\})^*$ be the homomorphism defined by $a \mapsto a$ and $\# \mapsto \varepsilon$. Then $\varphi(L) = \mathrm{IP}_A^M$, so as $\mathcal{C}$ is closed under homomorphism, the result follows. □

As we shall mention it in the course of the subsequent proof, recall that if a monoid M is generated by a finite set A, then a rewriting system $\mathcal{R}$ with $\mathcal{R} \subseteq A^* \times A^*$ is said to be M-*equivariant* if $\xleftrightarrow{*}_R \subseteq \xleftrightarrow{*}_M$. We are now ready to state the key lemma for monoid free products.

Lemma 2.3.5. *Let M_1, M_2 be two monoids generated by the finite sets A_1, A_2, respectively. Let $M_1 * M_2$ denote their monoid free product. Let*

$$I' = \{(w \to 1) \quad \mid w \in \mathrm{IP}_{A_1}^{M_1} \cup \mathrm{IP}_{A_2}^{M_2}\} \subseteq (A_1 \cup A_2)^*,$$
$$I'' = \{(w^{rev} \to 1) \mid w \in \mathrm{IP}_{A_1}^{M_1} \cup \mathrm{IP}_{A_2}^{M_2}\} \subseteq (A_1 \cup A_2)^*.$$

Then we have

$$\mathrm{WP}_{A_1 \cup A_2}^{M_1 * M_2} = \left(\mathrm{WP}_{A_1}^{M_1} \star \mathrm{WP}_{A_2}^{M_2}\right)^{I', I''}.$$

Proof. We first note that $\mathrm{WP}_{A_1}^{M_1}$ and $\mathrm{WP}_{A_2}^{M_2}$ are both concatenation-closed. Hence the product $\mathrm{WP}_{A_1}^{M_1} \star \mathrm{WP}_{A_2}^{M_2}$ is well-defined. We now prove the claimed equality.

($\supseteq$) Choose an arbitrary word $w_1 \# w_2^{\mathrm{rev}} \in \left(\mathrm{WP}_{A_1}^{M_1} \star \mathrm{WP}_{A_2}^{M_2}\right)^{I',I''}$. Then, there exists by definition of (I', I'')-ancestors some $u\#v^{\mathrm{rev}} \in \mathrm{WP}_{A_1}^{M_1} \star \mathrm{WP}_{A_2}^{M_2}$ such that $w_1 \xrightarrow{*}_{I'} u$ and $w_2^{\mathrm{rev}} \xrightarrow{*}_{I''} v^{\mathrm{rev}}$. We claim that we have $w_1 =_{M_1 * M_2} u$ and $w_2 =_{M_1 * M_2} v$.

For the first claim, as I' is M_1-equivariant and M_2-equivariant, it is also $M_1 * M_2$-equivariant, as $\xleftrightarrow{*}_{M_i} \subseteq \xleftrightarrow{*}_{M_1 * M_2}$ for $i = 1, 2$. Hence, as $w_1 \xrightarrow{*}_{I'} u$, we have $w_1 =_{M_1 * M_2} u$, as required. For the second claim, we could use an analogous argument, or write it out in full; we opt for the latter. The proof is by induction on the number k of rewriting steps in $w_2^{\mathrm{rev}} \xrightarrow{*}_{I''} v^{\mathrm{rev}}$. If $k = 0$, then $w_2^{\mathrm{rev}} \equiv v^{\mathrm{rev}}$ and there is nothing to show, for then $w_2 \equiv v$. Assume $w_2^{\mathrm{rev}} \to_{I''}^{k} v^{\mathrm{rev}}$ with $k > 0$, and assume the claim holds for all rewritings of length shorter than k. Then there is some $y_2 \in A^*$ such that $w_2^{\mathrm{rev}} \to_{I''} y_2 \to_{I''}^{k-1} v^{\mathrm{rev}}$, and some $w_2', w_2'' \in A^*$ such that $w_2^{\mathrm{rev}} \equiv w_2' s w_2''$ and $y_2 \equiv w_2' w_2''$, where $(s \to 1)$ is a rule in I''. Now $s^{\mathrm{rev}} =_{M_1} 1$ or $s^{\mathrm{rev}} =_{M_2} 1$; in either case $s^{\mathrm{rev}} =_{M_1 * M_2} 1$ and so

$$\begin{aligned} w_2 \equiv (w_2^{\mathrm{rev}})^{\mathrm{rev}} \equiv (w_2' s w_2'')^{\mathrm{rev}} &\equiv (w_2'')^{\mathrm{rev}} s^{\mathrm{rev}} (w_2')^{\mathrm{rev}} \\ &=_{M_1 * M_2} (w_2'')^{\mathrm{rev}} (w_2')^{\mathrm{rev}} \equiv (w_2' w_2'')^{\mathrm{rev}} \equiv y_2^{\mathrm{rev}}. \end{aligned}$$

By the inductive hypothesis, as $y_2 \to_{I''}^{k-1} v^{\mathrm{rev}}$ we have $y_2^{\mathrm{rev}} =_{M_1 * M_2} v$. Thus we have our claim proved; that is, $w_2 =_{M_1 * M_2} v$.

Thus we have $w_1 =_{M_1 * M_2} u$ and $w_2 =_{M_1 * M_2} v$. As $u\#v^{\mathrm{rev}} \in \mathrm{WP}_{A_1}^{M_1} \star \mathrm{WP}_{A_2}^{M_2}$, we have, exactly as in the proof of Lemma 2.3.1, that $u\#v^{\mathrm{rev}} \in \mathrm{WP}_{A_1 \cup A_2}^{M_1 * M_2}$. In particular, $u =_{M_1 * M_2} v$. Hence $w_1 =_{M_1 * M_2} w_2$, i.e. $w_1 \# w_2^{\mathrm{rev}} \in \mathrm{WP}_{A_1 \cup A_2}^{M_1 * M_2}$. As $w_1 \# w_2^{\mathrm{rev}}$ was arbitrary, this completes the proof of the inclusion.

($\subseteq$) Now suppose $w \in \mathrm{WP}_{A_1 \cup A_2}^{M_1 * M_2}$. This implies that $w \equiv u\#v^{\mathrm{rev}}$ for some $u, v \in (A_1 \cup A_2)^*$ with $u =_{M_1 * M_2} v$. Let u', v' be any reduced forms of u, v, respectively, which we can by Proposition 1.1.2 choose to be such that $u \in \langle u' \rangle_{I'}$ and $v \in \langle v' \rangle_{I'}$. Let $u' \equiv u_0 u_1 \cdots u_m$ and $v' \equiv v_0 v_1 \cdots v_n$ be the normal forms of the reduced words u' and v'. By Lemma 1.1.3, we have $u =_{M_1 * M_2} v$ if and only if $n = m$, and $u_i, v_i \in A_{X(i)}^*$ with $u_i =_{M_{X(i)}} v_i$ for all $0 \leq i \leq n$, where $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$. Hence, as $u_i =_{M_{X(i)}} v_i$ for all $i \geq 0$, we have that $u_i \#(v_i)^{\mathrm{rev}} \in \mathrm{WP}_{A_{X(i)}}^{M_{X(i)}}$ for all $i \geq 0$. Furthermore, we clearly have

$$u'\#(v')^{\mathrm{rev}} \equiv u_0 u_1 \cdots u_n \#(v_1 v_2 \cdots v_n)^{\mathrm{rev}} \equiv u_0 u_1 \cdots u_n \# v_n^{\mathrm{rev}} \cdots v_2^{\mathrm{rev}} v_1^{\mathrm{rev}}.$$

We thereby conclude, by definition of the alternating product, that

$$u'\#(v')^{\mathrm{rev}} \in \mathrm{WP}_{A_1}^{M_1} \star \mathrm{WP}_{A_2}^{M_2}.$$

Now, $u \in \langle u' \rangle_{I'}$ and $v \in \langle v' \rangle_{I'}$. From the latter, we have $v^{\mathrm{rev}} \in \langle (v')^{\mathrm{rev}} \rangle_{I''}$. Thus, by definition of (I', I'')-ancestors, we conclude that $u\#v^{\mathrm{rev}} \in \left(\mathrm{WP}_{A_1}^{M_1} \star \mathrm{WP}_{A_2}^{M_2}\right)^{I',I''}$. As the word $w \equiv u\#v^{\mathrm{rev}}$ was arbitrary, we have proved the inclusion, as required. □

Combining the knowledge we gained of (I', I'')-ancestors from earlier, we can hence state the following theorem.

Theorem 2.3.6. *Let $\mathcal{C}$ be a super-*AFL *closed under reversal. Then the class of monoids with word problem in $\mathcal{C}$ is closed under taking monoid free products.*

Proof. Note that as $\mathcal{C}$ is closed under inverse homomorphism, there is no ambiguity in speaking of a monoid with word problem in $\mathcal{C}$ by [209, Proposition 8]. If M_1, M_2 are two monoids with word problem in $\mathcal{C}$, generated by disjoint finite sets A_1 resp. A_2, then by Lemma 2.3.4 we have that $\mathrm{IP}^{M_1}_{A_1}, \mathrm{IP}^{M_2}_{A_2} \in \mathcal{C}$. Let $I' = \mathrm{IP}^{M_1}_{A_1}$ and $I'' = \mathrm{IP}^{M_2}_{A_2}$. Then I', I'' are monadic $\mathcal{C}$-rewriting systems by Lemma 2.3.4 and the closure of $\mathcal{C}$ under reversal. By Corollary 2.2.7, as M_1 and M_2 both have word problem in $\mathcal{C}$, and $\mathcal{C}$ is a super-AFL, we have

$$\mathrm{WP}^{M_1}_{A_1} \star \mathrm{WP}^{M_2}_{A_2} \in \mathcal{C}.$$

Hence, by Theorem 2.2.14, we have

$$\left(\mathrm{WP}^{M_1}_{A_1} \star \mathrm{WP}^{M_2}_{A_2} \right)^{I', I''} \in \mathcal{C}.$$

But by Lemma 2.3.5, this language is $\mathrm{WP}^{M_1 * M_2}_{A_1 \cup A_2}$, and we are done. □

Again, as in the semigroup case, we recover the following result, originally due to Brough, Cain & Pfeiffer [80], as a direct corollary of Theorem 2.3.6.

Corollary (Brough, Cain & Pfeiffer)**.** *The class of monoids with context-free word problem is closed under taking monoid free products.*

We remark that although this result (and the semigroup analogue) was proved already by the named authors, these authors used entirely different methods specific to context-free languages (namely, they constructed explicit pushdown automata). The authors also only provided a sketch proof in the case of the monoid free product case, which is (as we have seen) significantly more complicated than the semigroup case. Now, the class of *deterministic* context-free languages is not an AFL (see [203] for a definition). Therefore, the methods in this chapter – which rely on e.g. the monadic ancestor property – are not directly useful for making progress towards resolving the conjecture of Brough, Cain & Pfeiffer [80] on whether the class of deterministic context-free monoids is closed under free products. For this latter question, we strongly suspect the answer is negative. On the other hand, we have:

Corollary 2.3.7. *The class of monoids with indexed word problem is closed under taking monoid free products.*

Indeed, as every group admits a (special) monoid presentation; and as the monoid free product of two such special monoid presentations coincides with the usual group-theoretic free product of the two groups; and as a group has language-theoretic word problem (in the above sense) in a class $\mathcal{C}$ closed under inverse homomorphism if and only if its identity problem is in $\mathcal{C}$; we conclude the following purely group-theoretic result.

Corollary 2.3.8. *The class of groups with indexed word problem is closed under free products.*

It is an open problem whether there exist groups with indexed word problem which are not context-free (see e.g. [162]). It is, however, known that groups with indexed word problems have decidable rational subset membership problem by a result of Lisovik (see [238, Corollary 3.6]).

Chapter 3

The Word Problem for Special Monoids

Synopsis

This chapter deals with special monoids, and studies the word problem (in the sense of Duncan & Gilman) for such monoids. In §3.1, we give an overview of the ideas involved in such a study via the example of the bicyclic monoid. In §3.2, we carry out a careful study of invertible words in a special monoid, including manipulating the presentations involved to gain control over the invertible pieces. In §3.3, we use this control to language-theoretically describe equalities between words representing invertible elements. In §3.4, we use this description together with a normal form lemma to completely understand the language-theoretic word problem of a special monoid modulo the same properties of its group of units. This results in a reduction theorem: for an appropriately restricted class of languages $\mathcal{C}$, a finitely presented special monoid has word problem in $\mathcal{C}$ if and only if its group of units has word problem in $\mathcal{C}$ (Theorem 3.4.1). In §3.5, we study the context-free case in detail. As a corollary, we find as a very particular case a full generalisation of the famous Muller-Schupp theorem; namely, a finitely presented special monoid has context-free word problem if and only if its group of units is virtually free (Theorem 3.5.1). We also discuss applications to the rational subset membership problem. In §3.6, we end with some open problems. This chapter is based on the pre-print [383]. This chapter answers broad generalisations of questions posed by Zhang in 1992, and by Book & Otto in 1993.

The goal of this chapter is to investigate the following question.

Question. *Let $\mathcal{C}$ be a class of languages. What is the algebraic structure of a special monoid with word problem in $\mathcal{C}$?*

In order for this Question to be well-posed, we assume that $\mathcal{C}$ is closed under inverse homomorphisms (as in §1.2.3). Prior to the work contained in this thesis, this question did not directly appear anywhere in the literature on special monoids; however, we note that the word problem of the bicyclic monoid $\operatorname{Mon}\langle b, c \mid bc = 1 \rangle$ is mentioned to be context-free by Brough, Cain, & Pfeiffer [80]. Book, Jantzen & Wrathall [69] proved that any monoid admitting a complete monadic context-free rewriting system has context-free word problem (Cain & Maltcev [95] later independently rediscovered this result). These methods, however, are sporadic at best in their application to special monoids, and we now therefore initiate the first systematic investigation of this problem.

Without developing any machinery, it is possible to answer the above question completely in the case that $\mathcal{C}$ is the class $\mathcal{C}_{\text{reg}}$ of regular languages. First, by the analogue of Anīsīmov's theorem, a monoid M has regular word problem if and only if M is finite (§1.2.3). It follows directly from work by Adian on identities that a special monoid is finite if and only if it is a finite group [3].[51] See also [341, Theorem 6] for a proof of this using rewriting techniques. Hence we have the following easy classification of special monoids with regular word problem.

Proposition. *Let $M = \operatorname{Mon}\langle A \mid w_i = 1 \, (1 \leq i \leq p) \rangle$ be a special monoid. Then M has regular word problem if and only if M is a finite group.*

This result has essentially been observed already by Bucher [83, Theorem 3.9], though using significantly more words. Our ultimate goal is to provide a language-theoretic result mirroring Makanin's reduction of the word problem for M to the same problem for $U(M)$. In particular, we wish to prove that, for a class of languages $\mathcal{C}$, we have:

$$M \text{ has word problem in } \mathcal{C} \iff U(M) \text{ has word problem in } \mathcal{C}. \tag{$*$}$$

In full generality, this is not possible. Let $\mathcal{C} = \mathcal{C}_{\text{reg}}$, the regular languages. The bicyclic monoid $B = \operatorname{Mon}\langle b, c \mid bc = 1 \rangle$ has trivial group of units $U(B)$. Thus the word problem for $U(B)$ is regular by Anīsīmov's theorem. If the above equivalence $(*)$ were to hold for $\mathcal{C}_{\text{reg}}$, then by the above proposition the bicyclic monoid would have to be a finite group – and, being neither finite nor a group, this is a lot to demand of it. Thus the above equivalence cannot hold for $\mathcal{C}_{\text{reg}}$.

We will thus need to provide certain restrictions on $\mathcal{C}$ to ensure that the equivalence $(*)$ does hold. This includes certain closure properties. Specifically, we will show that if $\mathcal{C}$ is a super-AFL closed under reversal, then $(*)$ holds for every special monoid M. The picture painted is that if $(*)$ is true for a class $\mathcal{C}$, then $\mathcal{C}$ should be "context-free-like". This is because – as we shall discover – the word problem of M is built up in a "context-free way" from the word problem of $U(M)$. This is the language-theoretic interpretation of the geometric idea that the Cayley graph of M is built up in a "tree-like way" from the Cayley graph of $U(M)$ (see Chapter 5).

[51] Specifically, Adian proved that a special monoid satisfies a non-trivial *identity* (see e.g. [443] for a definition) if and only if it is isomorphic to $\mathbb{N}$, the bicyclic monoid, or a group; and every finite special monoid satisfies the identity $x^n = 1$ for some $n \geq 1$.

3.1 Warm-up: the bicyclic monoid

Let $M = \text{Mon}\langle A \mid w_1 = 1, w_2 = 1, \dots, w_k = 1\rangle$. Then recall from §1.3 that M is called *special*, and that its group of units is denoted $U(M)$. By the results in that section, the word problem – as a decision problem – for M reduces to the same problem for $U(M)$. This reduction, however, is in general non-trivial. We give an example of this via the bicyclic monoid B, leading us into the difficulties that one might face along the way.

Let $B = \text{Mon}\langle b, c \mid bc = 1\rangle$ be the bicyclic monoid.[52] The group of units of B is the trivial group – but there is still non-triviality involved in comparing words in B. Worse still, even comparing whether two invertible words represent the same element is non-trivial! For example, the words $b^n c^n$ and bc represent the same element for any choice of $n \geq 0$; indeed, as can be observed by using the complete rewriting system $\{bc \to \varepsilon\}$, a word $w \in \{b, c\}^*$ represents the identity element in B if and only if it is an element of the Dyck language (see Example 1.2.1)

$$\{w \mid w \in \{b,c\}^*, \sigma_b(w) = \sigma_c(w), \text{ and } \sigma_b(p) \leq \sigma_c(p) \text{ for every prefix } p \text{ of } w\}.$$

This language is not regular, but it is context-free. Thus even though the group of units $U(B)$ is trivial, there is some "unfolding" in moving from the invertible elements of B to the invertible words. If we are to describe the equality of words in B (in order to describe the word problem of B), we must handle this. We give an overview of the general way this is handled in this simple case, but using only one particular property of the bicyclic monoid (we shall specify this property presently). The reader may instead choose their favourite special monoid M, compute the set Δ for it, and everywhere below substitute M for B, singing along with their own lyrics.

To control the invertible words in B, note that $\Delta = \{bc\}$ by the overlap algorithm. Let $X = \{x_1\}$, and let $\phi \colon \Delta^* \to X^*$ be the usual surjective homomorphism. Every invertible word is equal in B to a word over Δ^*, as $\langle \Delta \rangle = U(B)$. Thus, using the completeness of Zhang's rewriting system $S(B)$, if $w \in \{b,c\}^*$ is invertible, then there is an irreducible word in Δ^* in its equivalence class modulo $S(B)$. Thus there exists $W \in \Delta^*$ such that $w \xrightarrow{*}_{S(B)} W$. In our case, no word in Δ contains another piece as a subword. Thus, by induction on the length of the rewriting $w \xrightarrow{*}_{S(B)} W$, one can show (Lemma 3.2.20) that w is the ancestor of some element of Δ^* under the rewriting system $\mathcal{R} = \{V \to \varepsilon \mid V \in \Delta^*\}$. In our case, $\mathcal{R} = \{(bc)^n \to \varepsilon \mid n \geq 0\}$. That step was the difficult reduction – now we are almost done: suppose that $U(B)$ has word problem (with respect to the generating set X, say) in the class of languages $\mathcal{C}$, closed under inverse homomorphism. Then $U(B)$ has word problem with respect

[52] The bicyclic monoid admits only a single one-relation monoid presentation (up to renaming generators) – namely $\text{Mon}\langle b, c \mid bc = 1\rangle$. This was likely first proved by Shneerson [445]; the proof is not very long, but uses the *Freiheitssatz* for one-relation monoids and Adian's theory of (left/right) cycle-free presentations. Cain & Maltcev [94] independently rediscovered this result, and give a different (but less elementary) proof of the same fact. There are many natural questions along this line, which would be quite interesting to investigate, but which we have no space for herein; for example, are there any other special one-relation monoids which admit only a single special one-relation presentation? For every k, is there a special k-relation monoid admitting only one special k-relation monoid presentation? How many special inverse monoid presentations does the bicyclic monoid admit (at least three; note that $B \cong \text{Inv}\langle b, c \mid bc = 1\rangle$, $\text{Inv}\langle a \mid aa^{-1} = 1\rangle$, and $\text{Inv}\langle a, b \mid abb = 1\rangle$)?

to the generating set Δ in $\mathcal{C}$. Thus the language of words representing 1 in $U(B)$ is in $\mathcal{C}$ – so $\mathcal{R}$ is a $\mathcal{C}$-rewriting system! But w is an ancestor under $\mathcal{R}$ of the word ε. Hence, as every such ancestor is invertible, if $\mathcal{C}$ is a class of languages with the monadic ancestor property, then it follows that the language of invertible words (as w was arbitrary) is also in $\mathcal{C}$.[53]

The only limiting assumption we used about the bicyclic monoid above is that no piece appeared as a subword of another piece. In the more general case, some legwork is necessary to reach the conclusion that the language of invertible words is in $\mathcal{C}$. Suppose that we have the monoid $M = \operatorname{Mon}\langle a, b, c \mid a(bac)a = 1\rangle$. Then $\Delta = \{a, bac\}$ by the overlap algorithm. One can show that $a \neq_M bac$, either by constructing a finite complete rewriting system for M, or by appealing to the general (but non-trivial) fact that in a special one-relation monoid M', no two distinct words from Δ are equal in M' (see §1.3 or indeed [505]). In either case, we partition $\Delta = \Delta_1 \cup \Delta_2 = \{a\} \cup \{bac\}$, let $X = \{x_1, x_2\}$ and finally let $\phi(a) = x_1$ and $\phi(bac) = x_2$. Now there is a "piece in a piece", as a appears as a subword of bac. This complicates matters: unlike for the bicyclic monoid, there is no longer any reason to expect that every invertible word is the ancestor under a special rewriting system of a word in Δ^*. For example, let

$$\Pi = \operatorname{Mon}\langle \alpha, \beta, \gamma, \delta \mid \alpha\beta^2\gamma = 1, \beta\delta^2 = 1, \delta^2\beta = 1, \beta^3 = 1\rangle.$$

Then $\beta, \delta, \alpha\beta\gamma, \in \Delta$ by Makanin's procedure. As both β^2 and δ^2 are inverses of β, we must have $\beta^2 =_\Pi \delta^2$. Thus $\alpha\delta^2\gamma$ is an invertible word; it is equal in Π to $\alpha\beta^2\gamma$, which equals 1. But how could we, by removing subwords equal to 1, possibly rewrite $\alpha\delta^2\gamma$ to an element of Δ? Indeed, it is not hard to show (e.g. by constructing a finite complete rewriting system for Π) that none of the subwords of $\alpha\delta^2\gamma$, other than the entire word itself, is equal to 1! Of course, we would *like* to perform the rewriting $\delta^2 \to \beta^2$, but this is not a rule which could be a part of either a special or a monadic system, and there is little reason to expect that ancestors under such rewritings should preserve any language-theoretic properties.

We will remedy the situation in the following way. Let $M = \operatorname{Mon}\langle A \mid w_i = 1 \ (i \in I)\rangle$ be an arbitrary finitely presented special monoid with pieces Δ. Let w be an invertible word as before, and let $w \xrightarrow{*}_{S(M)} W$, where $W \in \Delta^*$. Then the rewriting can take place inside one of the pieces in W. However, if one can manipulate the presentation – and Δ – to ensure that the only pieces appearing inside pieces have length 1 – that is, such that

$$\delta_1, \delta_2 \in \Delta \text{ and } \delta_1 \text{ is a proper subword of } \delta_2 \quad \Longrightarrow \quad |\delta_1| = 1,$$

then w can be shown, by another induction, to be the ancestor of W under a monadic rewriting system. For if one would need to apply some rule (r, s) of $S(M)$, then either $|s| \leq 1$, or else one can reduce the number of steps in the rewriting by one – this is made precise in Lemma 3.2.20. It follows that the rewriting system

$$\mathcal{R}_\Delta = \{W \to \delta \mid W \in \Delta^*, \delta \in \Delta \cup \{\varepsilon\}, |\delta| = 1, W =_M \delta\}$$

is such that w is the ancestor under $\mathcal{R}_\Delta$ of some element of Δ^*. But if $\mathcal{C}$ is a class of languages closed under union, then $\mathcal{R}_\Delta$ is a monadic $\mathcal{C}$-rewriting system if $U(M)$ has word problem

[53] In the case of the bicyclic monoid, we can use this to conclude that the Dyck language is a context-free language by taking $\mathcal{C} = \mathcal{C}_{\mathrm{cf}}$. Note that $\mathcal{C}_{\mathrm{reg}}$ does not have the monadic ancestor property, so although the word problem for $U(B) = 1$ is a regular language, it does not follow that the language of invertible words is a regular language!

in $\mathcal{C}$ (with respect to Δ). Thus we can proceed as above, and conclude that the language of invertible words in M is also in $\mathcal{C}$; the dependency on specific generating set and Δ is removed by additionally assuming $\mathcal{C}$ is closed under inverse homomorphism.

Many details remain; for example, we formalise the study of "equality of invertible words" by introducing for M, generated by A, the *invertible word problem*

$$\mathrm{InvP}_A^M := \{u \# v^{\mathrm{rev}} \mid u, v \in A^*, \text{ such that } u =_M v \text{ and } u, v \text{ are invertible}\}.$$

For a sufficiently restricted class of languages $\mathcal{C}$, we prove that this language is in $\mathcal{C}$ if and only if $U(M)$ has word problem in $\mathcal{C}$. This connects invertible words with the group of units. The final step is then a very easy one: using the normal form lemma (Lemma 1.3.8) one shows more or less directly that M has word problem in $\mathcal{C}$ if and only if InvP_A^M is in $\mathcal{C}$. This completes the overview of the theorem.

We will in the next section begin with detailing the changes to the presentations needed to ensure that the "piece in a piece" difficulties can be resolved. We shall also see in Chapter 5 that pieces appearing inside pieces is a recurring theme for complicating matters in special monoids.

3.2 Invertible elements

Throughout this section, if not explicitly mentioned otherwise, we let

$$M = \mathrm{Mon}\langle A \mid w_1 = 1, w_2 = 1, \dots, w_k = 1\rangle$$

be a fixed special monoid. Let Δ be the set of invertible pieces of M, and let $\phi\colon \Delta^* \to X^*$ be the usual homomorphism. We know well (see §1.3) that the set Δ of pieces generates the invertible elements, in the sense that for every invertible word $w \in A^*$ there exists some $W \in \Delta^*$ such that $w =_M W$. However, as we have seen many times, it need not be true that an invertible word is an element of Δ^*. In other words, if $\pi\colon A^* \to M$ is the usual homomorphism, then in general

$$\pi(w) \in \pi(\Delta^*) \not\Longrightarrow w \in \Delta^*. \tag{$*$}$$

In fact, it is straightforward to see that the implication $(*)$ holds if and only if every element of Δ is a single letter, i.e. if and only if M is a free product of a free monoid by a group. Indeed, the "if" part is trivial. For the "only if" direction, one only needs observe that if some piece has length greater than 1, then one can repeatedly insert some word equal to 1 into this piece and obtain infinitely many minimal words (using an easy minimality argument), all of which are pairwise graphically distinct, all of which are invertible, but of which only finitely many are in Δ. For example, in the bicyclic monoid $\mathrm{Mon}\langle b, c \mid bc = 1\rangle$ we have $\Delta = \{bc\}$, whereas if $n \geq 2$, we have $b^n c^n \notin \{bc\}^*$, but obviously $b^n c^n =_M 1$ is invertible. The aim of this section is to obtain a deeper understanding of the relationship between the invertible elements and Δ^*.

3.2.1 Generalised pieces

We will begin the journey towards understanding invertible elements with the following simple lemma, which gives some insight into the relationship between invertible words and pieces.

Lemma 3.2.1. *If $w \in A^*$ is invertible and non-empty, then w contains a piece as a subword.*

Proof. As w is invertible, there is a word $w' \in A^*$ such that $ww' =_M 1$. As the rewriting system $S = S(M)$ is complete, and the empty word is irreducible modulo S, we have $ww' \xrightarrow{*}_S 1$. We may assume without loss of generality that w' is irreducible modulo S. We add a formal marker $|$ signifying the boundary of the words $w|w'$ for clarity. As w' is irreducible, when rewriting $w|w' \xrightarrow{*}_S 1$ we find that (as $S \subseteq \Delta^* \times \Delta^*$) the first application of a rule $(s_1, s_2) \in S$ must be such that s_1 either straddles the boundary of $|$, or else is a subword of w (as w' is irreducible). As $s_1 \not\equiv \varepsilon$, in the latter case we are done, as then w contains s_1 as a subword (and hence also a piece as a subword). In the former case, no single piece in s_1 can straddle the boundary $|$, as then it would have a left invertible prefix (as this prefix would be a suffix of the invertible word w). Thus the boundary $|$ splits s_1 as $s_1 \equiv s_1'|s_1''$ with $s_1', s_1'' \in \Delta^*$ and s_1' a subword of w; this subword s_1' is non-empty as w' is irreducible. Thus w contains a piece as a subword. □

The above lemma is of limited use in practice, and does not shed much light on what an arbitrary invertible word looks like. To understand such words, we will define a new approach. We now define a new set $\overline{\Delta}$ of words, being the set of all invertible words $w \in A^*$ such that:

(1) There exists some $W \in \Delta$ such that $w \xrightarrow{*}_S W$.
(2) No proper non-empty prefix of w is invertible.

We call $\overline{\Delta}$ the *closure of* Δ. We note that $\Delta \subseteq \overline{\Delta}$, and that condition (2) is easily seen to be equivalent to "no proper non-empty *suffix* of w is invertible", which will occasionally be useful.[54] Evidently, just as Δ is a biprefix code, so too is $\overline{\Delta}$ is a biprefix code (as a subset of A^*). Furthermore, any element of $\overline{\Delta}^*$ is clearly invertible.

Example 3.2.2. Let $B = \operatorname{Mon}\langle b, c \mid bc = 1\rangle$. Then $\Delta = \{bc\}$. As B is defined by the finite complete rewriting system $\mathcal{R}$ with the single rule $(bc \to 1)$, and since this rule is an element of $S(B)$ as $bc \in \Delta$, we have that for a word $w \in A^*$ condition (1) in the definition of $\overline{\Delta}$ is equivalent to: $w \xrightarrow{*}_{\mathcal{R}} bc$. Of course, the word $bcbc$ is such that $bcbc \xrightarrow{*}_{\mathcal{R}} bc$, but this does not satisfy condition (2). Hence $bcbc \not\in \overline{\Delta}$ (but note that $bcbc \in \Delta^* \subseteq \overline{\Delta}^*$). The word $bbcc$, on the other hand, is such that $bbcc \xrightarrow{*}_{\mathcal{R}} bc$, and none of its proper non-empty prefixes is invertible. Hence $bbcc \in \overline{\Delta}$. Indeed, $\overline{\Delta}$ can be seen to be precisely the context-free language $b\langle\varepsilon\rangle_{\mathcal{R}}c$. △

The following lemma is straightforward to prove, and its proof is entirely analogous to the proof of [502, Proposition 2.1], whose proof we follow directly.

Lemma 3.2.3. *Let $x, y, z \in A^*$. Then:*

(1) $xy, x \in \overline{\Delta}^*$ imply $y \in \overline{\Delta}^*$.
(2) $yz, z \in \overline{\Delta}^*$ imply $y \in \overline{\Delta}^*$.
(3) Suppose that $xy \in \overline{\Delta}^*$. If either x or y is invertible, then $x \in \overline{\Delta}^*$ and $y \in \overline{\Delta}^*$.
(4) Suppose that $xy \in \overline{\Delta}^*$ and $yz \in \overline{\Delta}^*$. Then $x, y, z \in \overline{\Delta}^*$.

Proof. Both statements (1) and (2) follow from the fact that $\overline{\Delta}$ is a biprefix code. For (3), suppose x is invertible, and let $xy \equiv x_1x_2\cdots x_m$ with $x_i \in \overline{\Delta}$ where $1 \leq i \leq m$. Then $x \equiv x_1x_2\cdots x_{\ell-1}E$, where E is a prefix of x_ℓ for some $\ell \leq m$. Let $x_\ell \equiv EF$, where $F \in A^*$. As x and $x_\ell \equiv EF$ are invertible, it follows that E is invertible (by the Fundamental Lemma). As $x_\ell \in \overline{\Delta}$, we must have that either E is empty, or else E is all of x_ℓ. In either case, $x \in \overline{\Delta}^*$. Thus by (1) also $y \in \overline{\Delta}^*$. Symmetrically, the results hold when y is invertible.[55] For (4), as we have $xy \in \overline{\Delta}^*$ and $yz \in \overline{\Delta}^*$, we have that xy and yz are invertible. Hence y is invertible, and so by (3) we have $x, y, z \in \overline{\Delta}^*$. □

Thus, by (4), the overlap-style arguments one can make for Δ can also be performed for $\overline{\Delta}$. We shall use the above lemma, and particularly case (4), implicitly throughout. The key property regarding $\overline{\Delta}$ is that just as any element of $\overline{\Delta}^*$ is invertible, the converse is also true.

Lemma 3.2.4. *A word $w \in A^*$ is invertible if and only if $w \in \overline{\Delta}^*$.*

Proof. Any element of $\overline{\Delta}^*$ is clearly invertible. For the converse, assume $w \in A^*$ is invertible. By [502, Lemma 3.4] there exists some least $n \geq 0$ and a $D \in \Delta^*$ such that $w \xrightarrow{*}{}^{n}_{S} D$. We will

[54] For, if w has an invertible prefix w', then writing $w \equiv w'w''$, we find that w'' is also invertible by the Fundamental Lemma. The symmetric argument holds.

[55] This case is misprinted by Zhang [502, p. 497] as "Symmetrically, the results hold when v is invertible".

prove the claim by induction on n. The base case $n = 0$ is clear, for then $w \equiv D \in \Delta^* \subseteq \overline{\Delta}^*$. Assume that the claim is true for some $n \geq 0$, and let w be such that $w \overset{*}{\to}{}_S^{n+1} D$. Then there exists some $w_1 \in A^*$ such that $w \to_S w_1$ and $w_1 \overset{*}{\to}{}_S^{n} D$. As $w =_M w_1$, the word w_1 is invertible and by the inductive hypothesis $w_1 \in \overline{\Delta}^*$. Write $w_1 \equiv \overline{\delta}_0\overline{\delta}_1 \cdots \overline{\delta}_k$ where $\overline{\delta}_i \in \overline{\Delta}$ for $0 \leq i \leq k$. As $w \to_S w_1$, there exists some $(\ell, r) \in S$ and words $u, v \in A^*$ such that $w \equiv u\ell v$ and $w_1 \equiv urv$. We subdivide into two cases, depending on whether r contains as a subword one of the $\overline{\delta}_i$ or not.

In the first case, we assume the fixed subword r of w_1 contains some $\overline{\delta}_i$, where $0 \leq i \leq k$, as a subword. Then, as $\overline{\Delta}$ is a biprefix code and $r \in \Delta^* \subseteq \overline{\Delta}^*$, we must have that $w_1 \equiv ErF$, where $E, F \in \overline{\Delta}^*$. Hence $w \equiv E\ell F$, and as $\ell \in \Delta^*$, we have $w \in \overline{\Delta}^*$. In the second case, this fixed subword r does not contain any $\overline{\delta}_i$ as a subword. We deal with two separate subcases, depending on whether r is empty or not.

First, if $r \equiv \varepsilon$, then there exists $0 \leq i \leq k$ we can write $u \equiv \overline{\delta}_0 \cdots \overline{\delta}_{i-1}\overline{\delta}'_i$ and $v \equiv \overline{\delta}''_i\overline{\delta}_{i+1} \cdots \overline{\delta}_k$, where $\overline{\delta}'_i, \overline{\delta}''_i \in A^*$ are such that $\overline{\delta}'_i\overline{\delta}''_i \equiv \overline{\delta}_i$; and such that

$$w \equiv u\ell v \equiv \overline{\delta}_0 \cdots \overline{\delta}_{i-1}(\overline{\delta}'_i\ell\overline{\delta}''_i)\overline{\delta}_{i+1} \cdots \overline{\delta}_k. \tag{3.2.1}$$

Assume $|\overline{\delta}'_i| \cdot |\overline{\delta}''_i| = 0$, i.e. at least one of $\overline{\delta}'_i, \overline{\delta}''_i$ is empty. If $\overline{\delta}'_i \equiv \varepsilon$, then $\overline{\delta}''_i \equiv \overline{\delta}_i \in \overline{\Delta}$, and as $\ell \in \overline{\Delta}^*$, we have $w \in \overline{\Delta}^*$ by (3.2.1). The case $\overline{\delta}''_i \equiv \varepsilon$ is entirely symmetrical. Thus assume instead that $|\overline{\delta}'_i| \cdot |\overline{\delta}''_i| > 0$. We claim that no non-empty prefix of $\overline{\delta}'_i\ell\overline{\delta}''_i$ is invertible, i.e. this word is minimal. By minimality of $\overline{\delta}'_i\overline{\delta}''_i \in \overline{\Delta}$, no proper non-empty prefix or suffix of this word is invertible; thus if some prefix of $\overline{\delta}'_i\ell\overline{\delta}''_i$ were invertible, then it is of the form $\overline{\delta}'_i\ell_1$, where $\ell_1 \in A^+$ is some non-empty proper prefix of ℓ. Thus ℓ_1 is left invertible, being a suffix of the invertible $\overline{\delta}'_i\ell_1$, but also right invertible, being a prefix of ℓ. It follows that ℓ_1 is invertible. As $\overline{\delta}'_i\ell_1$ is invertible, thus $\overline{\delta}'_i$ is invertible, which contradicts the minimality of $\overline{\delta}_i \equiv \overline{\delta}'_i\overline{\delta}''_i$ as $|\overline{\delta}''_i| > 0$. It follows that $\overline{\delta}'_i\ell\overline{\delta}''_i$ is minimal. As $\overline{\delta}'_i\ell\overline{\delta}''_i$ it is clearly invertible, being equal in M to $\overline{\delta}'_i\overline{\delta}''_i \in \overline{\Delta}$ by virtue of $\ell =_M 1$, we have that $\overline{\delta}'_i\ell\overline{\delta}''_i \in \overline{\Delta}$. By (3.2.1), we have $w \in \overline{\Delta}^*$.

The case $r \not\equiv \varepsilon$ uses very similar reasoning. Assume instead that $r \not\equiv \varepsilon$. Then there are two subcases, depending on whether r straddles the boundary of $\overline{\delta}_i\overline{\delta}_{i+1}$ for some $i \geq 0$, or whether r appears entirely as a subword of $\overline{\delta}_i$ for some $i \geq 0$. In the first case we can write $\overline{\delta}_i \equiv \overline{\delta}'_i\overline{\delta}''_i$ and $\overline{\delta}_{i+1} \equiv \overline{\delta}'_{i+1}\overline{\delta}''_{i+1}$ such that $r \equiv \overline{\delta}''_i\overline{\delta}'_{i+1}$, where $\overline{\delta}'_i, \overline{\delta}''_{i+1} \in A^+$ and $\overline{\delta}''_i, \overline{\delta}'_{i+1} \in A^*$. As $\overline{\delta}''_i$ is a suffix of the invertible word $\overline{\delta}_i$, it is left invertible. Furthermore, as a prefix of the invertible word r, it is right invertible; consequently, $\overline{\delta}''_i$ is invertible. As $\overline{\delta}_i \in \overline{\Delta}$, we necessarily have $\overline{\delta}''_i \equiv \varepsilon$. Symmetrically, we have $\overline{\delta}'_{i+1} \equiv \varepsilon$, and so $r \equiv \varepsilon$, a contradiction. In the second case, we can write $\overline{\delta}_i \equiv \overline{\delta}'_i r\overline{\delta}''_i$ for some $\overline{\delta}'_i, \overline{\delta}''_i \in A^*$. As r is invertible and non-empty, and $\overline{\delta}_i \in \overline{\Delta}$, we necessarily have that $\overline{\delta}'_i, \overline{\delta}''_i \in A^+$. We wish to show that $\overline{\delta}'_i\ell\overline{\delta}''_i \in \overline{\Delta}$, which would establish the claim. Now $\overline{\delta}'_i\ell\overline{\delta}''_i \to_S \overline{\delta}'_i r\overline{\delta}''_i \in \overline{\Delta}$, and hence it suffices to show that no non-empty proper prefix of $\overline{\delta}'_i\ell\overline{\delta}''_i$ is invertible. As $\overline{\delta}_i \in \overline{\Delta}$ and $|\overline{\delta}''_i| > 0$, no non-empty prefix of $\overline{\delta}'_i$ is invertible. Assume that $\overline{\delta}'_i\ell'$ is invertible, where $\ell \equiv \ell'\ell''$ for some $\ell', \ell'' \in A^*$. Then ℓ' is left invertible, being a suffix of $\overline{\delta}'_i\ell'$, and also right invertible, being a prefix of the invertible word ℓ. Thus ℓ' is invertible; hence as $\overline{\delta}'_i\ell'$ is invertible we also have that $\overline{\delta}'_i$ is invertible, a contradiction. Hence no non-empty prefix of $\overline{\delta}'_i\ell$ is invertible. Assume, finally, that there is a proper prefix $p \in A^*$ of

$\overline{\delta}_i''$ such that $\overline{\delta}_i'\ell p$ is invertible. Then $\overline{\delta}_i' rp$ is also invertible, being congruent to $\overline{\delta}_i'\ell p$. But then $\overline{\delta}_i' rp$ is an invertible non-empty proper prefix of $\overline{\delta}_i' r \overline{\delta}_i'' \in \overline{\Delta}$, a contradiction. This completes the subcase when $r \not\equiv \varepsilon$. □

Proposition 3.2.5. $\overline{\Delta}$ *is the set of minimal words of* M.

Proof. Clearly, if $\mathfrak{M}$ denotes the set of minimal words, then $\mathfrak{M}^*$ is the set of invertible words of M, so $\mathfrak{M}^* = \overline{\Delta}^*$ by Lemma 3.2.4. As $\mathfrak{M}$ and $\overline{\Delta}$ are both biprefix codes, we thus necessarily have $\overline{\Delta} = \mathfrak{M}$. □

We remark that this is a non-trivial equality. Of course, the equality suggests that we could have chosen to define $\overline{\Delta}$ as the set of minimal words, and instead proved that a word w is minimal if and only if no proper non-empty prefix of w is invertible and there is some piece $\delta \in \Delta$ with $w \xrightarrow{*}_S \delta$. In either case, we will not use the characterisation from the above equality directly, but we will instead use the explicit control over $\overline{\Delta}$ in terms of Δ given by the definition of $\overline{\Delta}$. We also remark that there is some similarity between Lemma 3.2.4 and [308, Lemma 19], a similarity discovered after the above proof was written.

Now, while the rewriting system S has rules in $\Delta^* \times \Delta^*$, it is not necessarily the case that if the *left-hand* of a rule is in Δ, then the *right-hand* side is also in Δ. This is demonstrated by the following example.

Example 3.2.6. Let $M = \operatorname{Mon}\langle a, b, c, p \mid (abc)^2 = 1, (abc)p^2 = 1, p^2(abc) = 1\rangle$. Then abc and p are invertible; one checks (using e.g. a finite complete rewriting system) that $\Delta = \{p, abc\}$. Now, as both abc and p^2 are inverses of abc, it follows that $abc =_M p^2$. Now $p^2 \in \Delta^*$ and $p^2 \notin \Delta$, yet we have the rule $abc \to_S p^2$ as $|abc| > |p^2|$ (and, of course, this is regardless of the ordering fixed on $A = \{a, b, c, p\}$). Thus the rule $(abc, p^2) \in S(M)$ is an element of $\Delta \times \Delta^2$, and so rewriting using $S(M)$ can increase the "Δ-length" of a word in Δ^* (i.e. the number of factors from Δ needed to write the word). △

In spite of this example, the following proposition has two purposes; it first indicates the kinds of arguments which become frequent when dealing with $\overline{\Delta}$; and second, it demonstrates that the type of behaviour indicated by Example 3.2.6 is rather controlled, in the (informal) sense that if a word in $\overline{\Delta}$ rewrites to an element from Δ^*, then this rewriting is in fact a rewriting into an element of Δ, followed by a rewriting of this element into the prescribed element of Δ^*.

Proposition 3.2.7. *Suppose* $u \in \overline{\Delta}$ *and* $v \in \overline{\Delta}^*$ *are such that* $u \to_S v$. *Then the following holds: if* $v \notin \overline{\Delta}$, *then* $v \in \Delta^*$ *and* $u \in \Delta$. *Consequently, if* $u \xrightarrow{*}_S W$ *where* $u \in \overline{\Delta}$ *and* $W \in \Delta^*$, *then there are* $W_0, \dots, W_\lambda \in \overline{\Delta}$ *and a piece* $\delta \in \Delta$ *such that*

$$u \equiv W_0 \to_S W_1 \to_S \cdots \to_S W_\lambda \to_S \delta \xrightarrow{*}_S W.$$

Proof. Suppose $u \equiv h_1 s_1 h_2$ and $v \equiv h_1 s_2 h_2$, where $(s_1, s_2) \in S(M)$ and $h_1, h_2 \in A^*$. We know that $s_1, s_2 \in \Delta^*$, and as $S(M)$ is ordered by length we know $s_1 \not\equiv \varepsilon$.

For the first part, suppose that $v \notin \overline{\Delta}$. Then as $v \in \overline{\Delta}^*$, we can write $v \equiv t_1 t_2 \cdots t_\mu$ uniquely, where $\mu \geq 2$ and $t_i \in \overline{\Delta}$ for all $1 \leq i \leq \mu$. Now if t_1 is a prefix of h_1, then some non-empty prefix of h_1 is invertible, and hence a non-empty proper (as $s_1 \not\equiv \varepsilon$) prefix of u is invertible, which contradicts $u \in \overline{\Delta}$. Thus t_1 contains h_1 as a proper prefix. Symmetrically, t_μ contains h_2 as a proper suffix.

Write $s_1 \equiv \delta_1 \cdots \delta_k$ for $\delta_i \in \Delta$, $1 \leq i \leq k$. Then, as $\mu \geq 2$, some prefix of s_1 must be equal to a suffix of t_1, i.e. $t_1 \equiv h_1\delta_1 \cdots \delta_j'$, where $\delta_j \equiv \delta_j'\delta_j''$ for some $1 \leq j \leq k$ and $|\delta_j'| > 0$. But then δ_j' is left invertible, being a suffix of t_1, and hence must be all of δ_j, as no proper prefix of a piece is left invertible. Thus $t_1 \equiv h_1\delta_1 \cdots \delta_j$. As t_1 and $\delta_1 \cdots \delta_j$ are both invertible, so too is h_1. As $u \in \overline{\Delta}$ has no invertible non-empty proper prefix, thus necessarily $h_1 \equiv \varepsilon$. Entirely symmetrically, one proves $t_\mu \equiv \delta_\ell \cdots \delta_k h_2$ for some $j < \ell \leq k$, and thus $h_2 \equiv \varepsilon$. Hence $v \equiv s_2 \in \Delta^*$, as needed. Furthermore, $u \equiv s_1 \in \Delta^*$, and as $u \in \overline{\Delta}$ we necessarily have $u \in \Delta$, as needed.

The final claim is an easy proof by induction on the number of steps λ in the rewriting $u \xrightarrow{*}_S W$. The base case $\lambda = 0$ is clear, for then $u \equiv W$, and consequently $W \in \Delta$. Thus we can take $\delta \equiv W$. Suppose that for some $\lambda > 0$ the claim is true for all such rewritings which take less than μ steps, and let $u \in \overline{\Delta}$ and $W \in \Delta^*$ be such that $u \xrightarrow{*}_S W$ in μ steps. Then there exist words $W_0, W_1, \cdots, W_\lambda \in A^*$ such that

$$u \equiv W_0 \rightarrow_S W_1 \rightarrow_S \cdots \rightarrow_S W_\lambda \equiv W.$$

As all of the W_i are invertible, thus $W_i \in \overline{\Delta}^*$ for all $0 \leq i \leq \lambda$ by Lemma 3.2.4. Now $W_0 \rightarrow_S W_1$ is a rewriting such that $W_0 \in \overline{\Delta}$ and $W_1 \in \overline{\Delta}^*$. If $W_1 \in \overline{\Delta}$, then as $W_1 \xrightarrow{*}_S W$ in $\mu - 1 < \mu$ steps, the claim follows by the inductive hypothesis. Assume then that $W_1 \notin \overline{\Delta}$. Then by the first part, $W_1 \in \Delta^*$ and $W_0 \in \Delta$. Hence we can take $\delta \equiv W_0$. □

The above lemma quite clearly captures the idea that the elements of $\overline{\Delta}$ are "generalised pieces". Indeed, it shows that any element of $\overline{\Delta}$ can be rewritten to an element of Δ by only passing through other elements of $\overline{\Delta}$ (rather than $\overline{\Delta}^*$). Note, however, that a given element of $\overline{\Delta}$ may be able to be rewritten to many distinct elements of Δ, as the following example shows.

Example 3.2.8. Let $M = \text{Mon}\langle a, b, c, p, q \mid abc = 1, apc = 1, p = 1, b = 1\rangle$. Then one can easily see that $\Delta = \{p, b, ac, abc, apc\}$, e.g. by using a finite complete rewriting system for M. Of course, among these pieces, we have plenty of equalities; for example, $abc =_M apc$. Furthermore, we have $abpc \in \overline{\Delta} \setminus \Delta$, as $abpc$ has no proper non-empty invertible prefix, and $abpc =_M ac$ is invertible. Thus there is some non-trivial rewriting of $abpc$ to an element of Δ. As $p =_M b =_M 1$, we have $(p \rightarrow 1), (b \rightarrow 1) \in S(M)$. Depending on which of these two rules we apply to $abpc$, we find either $abpc \rightarrow_S apc$ or $abpc \rightarrow_S abc$. However, $apc \not\equiv abc$ (obviously). Thus, rewriting an element of $\overline{\Delta}$ to an element of Δ does not always result in a unique element from Δ. △

Hence, an element of $\overline{\Delta}$ can intuitively be seen as an element of Δ with the "interior" altered.

3.2.2 Controlling the pieces

Throughout this section, fix a special monoid $M = \text{Mon}\langle A \mid w_1 = 1, \dots, w_k = 1\rangle$. For $w \in A^*$, let $\text{Rep}_A^M(w) \subseteq A^*$ denote the set of *representative words* of w, i.e. the set of words

$u \in A^*$ such that $w =_M u$. We will investigate when certain properties of the group of units of a special monoid guarantee certain properties of the set of representatives of invertible words in the monoid. Let $\delta \in \Delta$ be a piece. If $\delta \equiv h_1 w h_2$ for some non-trivial $h_1, h_2 \in A^+$ and a product of pieces $w \in \Delta^+$, then we say that w is a *subpiece* (of δ). If $|w| = 1$, then we say that w is a *small* subpiece. We will say that a special monoid presentation satisfies the *small subpiece condition* if all subpieces of pieces are small. In other words, the small subpiece condition states: if $\delta \in \Delta$ and $w \in \Delta^*$ are such that w is a proper subword of δ, then $|w| = 1$ (and $w \in \Delta$). We emphasise that a subpiece of a piece need not be a piece, but can be a product of several pieces.

Example 3.2.9. We give an example and a non-example of special monoids satisfying the small subpiece condition.

(1) Let $M = \text{Mon}\langle a, b, c \mid abc = 1, b = 1\rangle$. Then one can easily show that the pieces are $\Delta = \{abc, ac, b\}$. Thus M satisfies the small subpiece condition.

(2) Let $M' = \text{Mon}\langle a, b, c \mid ab^2c = 1, b = 1\rangle$. Then one can, using a finite complete rewriting system, easily see that $\Delta = \{ab^2c, abc, ac, b\}$. Then b^2 is a non-small subpiece of ab^2c. Thus M' does not satisfy the small subpiece condition.

In general, to determine given a special monoid presentation whether or not it satisfies the small subpiece condition seems likely to be an undecidable problem, as it appears that one needs to first compute the set Δ. △

The goal of this section is to prove the following statement:

Proposition 3.2.10. *Every finitely presented special monoid admits a presentation satisfying the small subpiece condition.*

Before proving this, we state a useful lemma, proved by Makanin [308, Lemma 12].

Lemma 3.2.11 ([308, Lemma 12])**.** *Let M be a special monoid given by the finite presentation*

$$M = \text{Mon}\langle A \mid w_1 = 1, \dots, w_k = 1\rangle.$$

with presentation pieces Λ. Fix some $1 \leq i \leq k$, and let $w_i \equiv \lambda_1\lambda_2 \cdots \lambda_\ell$ with $\lambda_j \in \Lambda$ for $1 \leq j \leq \ell$. Suppose that $\delta \in \Delta$ is such that $\delta \in [\lambda_p]^{\downarrow}$ for some fixed $1 \leq p \leq \ell$. Let $w_i' \equiv \lambda_1\lambda_2 \cdots \lambda_{p-1}\delta\lambda_{p+1} \cdots \lambda_\ell$, and let

$$M' = \text{Mon}\langle A \mid w_1 = 1, \dots, w_i' = 1, \dots, w_k = 1\rangle.$$

Then $M \cong M'$ via the identity map, i.e. the presentations define the same congruence on A^*. Furthermore, the factorisation in M' of the defining word w_i' into minimal invertible factors is obtained by replacing λ_p with δ in the factors of the factorisation of w_i, and the factorisation of the defining word w_j ($j \neq i$, $1 \leq j \leq k$) is identical to its factorisation in M.

The full proof can be found in Makanin's thesis, though using rather different language; it is rather lengthy, but uses the following key idea: if $\lambda_p =_M \delta$ via the generating operation, then there is a sequence of elementary transformations which transforms λ_p into δ, such that at no point does any elementary transformation actually involve deleting or inserting relation words which contain λ_p or δ as a subword. Thus, if two pieces are equal based on their "internal structure” (as δ and λ_p are in the statement of the lemma), then they are equal because of the equalities which hold between the other pieces, and so we may freely swap λ_p for δ.

Example 3.2.12. Let $M = \text{Mon}\langle a, b, c \mid abc = 1, b = 1\rangle$. Then $U(M) = 1$, and the pieces are given by $\Delta = \{abc, ac, b\}$, which can be verified using e.g. a finite complete rewriting system. Now ac is a c-word corresponding to the presentation piece abc. Thus we can replace abc by ac in the presentation, and see that

$$M \cong M' = \text{Mon}\langle a, b, c \mid ac = 1, b = 1\rangle,$$

which is obviously true. △

We will now give an example for how Makanin’s lemma can be applied to remove large subpieces. The idea in the example is the same as the general idea which will be used in the proof of Proposition 3.2.10. One crucial idea is the following. If we add a relation $u = v$ to a special monoid, then in general this cannot be replaced with some special relation(s) – equivalently, quotients of special monoids need not themselves be special monoids. On the other hand, if it happens that u or v is invertible in the special monoid, then the resulting quotient by the relation $u = v$ *is* special! For, if x is a word representing the inverse of, say, u, then the relation $u = v$ is equivalent to the relations $vx = xv = 1$.

Example 3.2.13. Let $M = \text{Mon}\langle a, b \mid abaabbab = 1\rangle$. By Adian’s algorithm, the defining word factors into invertible pieces as $(ab)(aabb)(ab)$, so $\Delta = \Lambda = \{ab, aabb\}$. Thus $ab \in \Delta^+$ is a large subpiece of $aabb \in \Delta$. We will replace this large subpiece ab by a small subpiece p.

Let p be any new symbol, and introduce the defining relation $p = ab$ to the presentation via a Tietze transformation, giving $\text{Mon}\langle a, b, p \mid abaabbab = 1, p = ab\rangle$. It is clear that $aabb \cdot ab$ is an inverse of ab, so from $p =_{M_3} ab$ it thus follows that M_3 is isomorphic to

$$\text{Mon}\langle a, b, p \mid abaabbab = 1, p = ab, p(aabb \cdot ab) = 1, (aabb \cdot ab)p = 1\rangle.$$

As the fact that both p and ab are inverses of $aabbab$ follows from the two added relations, the relation $p = ab$ follows from these relations; thus we can remove $p = ab$, and find that M_3 is isomorphic to the special monoid

$$\text{Mon}\langle a, b \mid (ab)(aabb)(ab) = 1, p(aabb)(ab) = 1, (aabb)(ab)p = 1\rangle. \tag{3.2.2}$$

It is clear that each of p, $aabb$, apb, and ab is a minimal invertible piece of this presentation. As $p =_{M_3} ab$, we have that the piece apc is obtained from $aabb$ by the piece-generating operation. Thus, by Makanin’s Lemma, we may replace $aabc$ by apc in (3.2.2) without changing the monoid defined by it. Thus

$$M_3 \cong \text{Mon}\langle a, b, p \mid (ab)(apb)(ab) = 1, p(apb)(ab) = 1, (apb)(ab)p = 1\rangle.$$

For this new presentation, $\Delta = \Lambda = \{p, ab, apb\}$, so it satisfies the small subpiece condition.

Example 3.2.14. Let $M = \text{Mon}\langle a, b \mid abaaabbbab = 1, aabbaabb = 1\rangle$. We do not expand on all the details involved. The reader is invited to verify that every step is as expected. We find that the defining relations factor into minimal invertible pieces as

$$(ab)(aaabbb)(ab) \quad \text{resp.} \quad (aabb)(aabb)$$

and hence $\Lambda = \{ab, aabb, aaabbb\}$. There is thus a horrifying amount of pieces appearing inside pieces. By introducing a generator $p = ab$, which we change into two relations $p(aaabbb)(ab) = 1$ and $(aaabbb)(ab)p = 1$, we may replace $aaabbb$ by its c-word $aapbb$, and thus find the presentation

$$M' = \text{Mon}\langle a, b, p \mid abaapbbab = 1, apbapb = 1, paapbbab = 1, aapbbabp = 1\rangle$$

with $M \cong M'$. Now the defining relations factor into minimal invertible pieces as

$$(ab)(aapbb)(ab), \quad (apb)(apb), \quad (p)(aapbb)(ab), \quad (aapbb)(ab)(p)$$

and thus the presentation pieces of this presentation for M is

$$\Lambda' = \{p, ab, apb, aapbb\}.$$

Now we have apb appearing as a subword of another piece $aapbb$, so we introduce $q = apb$, and as $(apb)^2 = 1$, we replace this relation by $qapb = 1, apbq = 1$. We then replace $aapbb$ by the piece aqb corresponding to it, finding

$$M'' = \text{Mon}\langle a, b, p, q \mid abaqbab = 1, apbapb = 1, paqbab = 1, aqbabp = 1, qapb = 1, apbq = 1\rangle$$

with $M'' \cong M' \cong M$. The relations factor as

$$(ab)(aqb)(ab), \quad (apb)(apb), \quad (p)(aqb)(ab), \quad (aqb)(ab)(p), \quad (q)(apb), \quad (apb)(q)$$

and hence

$$\Lambda'' = \{p, q, ab, apb, aqb\}.$$

Notice that

$$\sum_{\lambda \in \Lambda} (|\lambda| - 1) = (2-1) + (4-1) + (6-1) = 9,$$

$$\sum_{\lambda' \in \Lambda'} (|\lambda'| - 1) = (1-1) + (2-1) + (3-1) + (5-1) = 7,$$

and finally

$$\sum_{\lambda'' \in \Lambda''} (|\lambda''| - 1) = (1-1) + (1-1) + (2-1) + (3-1) + (3-1) = 5.$$

As $9 > 7 > 5$, we see that $\sum_{\lambda \in \Lambda}(|\lambda| - 1)$ strictly decreases when performing the above process; thus, this seems to be a good indicator for ensuring that the above process will always terminate (as this sum is always a positive integer). We shall see, in the proof of Proposition 3.2.10, that it indeed always decreases when performing the above procedure. △

We now generalise the above examples to the general case. First, for any set $S \subseteq A^*$, we let $\omega(S)$ denote the natural number $\sum_{s \in S}(|s| - 1)$. If $S = \Lambda$, where Λ is the set of presentation pieces of M, then in a loose sense $\omega(\Lambda)$ is a measure of the "complexity" of the pieces of the presentation – we remark that M is right cancellative if and only if $\omega(\Lambda) = 0$ by a result of Benois [49]. We shall, in the subsequent proof of Proposition 3.2.10, use several operations on the presentation for M, each of which reduces or does not increase $\omega(\Lambda)$. In particular, if a presentation with presentation pieces Λ does not satisfy the small subpiece condition, we will show that we find a new presentation with presentation pieces Λ' such that $\omega(\Lambda') < \omega(\Lambda)$. The proof will then be complete by induction.

Before we can realise the above idea in practice, we remark that it is not difficult to construct special monoids in which no presentation piece has a large subpiece, but there is some piece with a large subpiece. We begin with a lemma to remedy this, by showing that we can always find a presentation where large subpieces are "brought to light" in the presentation pieces.

Lemma 3.2.15. *Let M have presentation pieces Λ. Then M admits a special monoid presentation, with presentation pieces Λ', such that either this presentation satisfies the small subpiece condition; or else there is a presentation piece $\lambda \in \Lambda'$ containing a large subpiece, and $\omega(\Lambda') \leq \omega(\Lambda)$.*

Proof. If the given presentation for M satisfies the small subpiece condition, then we are done, so suppose that it does not. Let $\delta \in \Delta$ and $w \in \Delta^+$ be such that w is a large subpiece of δ. If $\delta \in \Lambda$, then we are done. If $\delta \notin \Lambda$, then there is some $\lambda_0 \in \Lambda$ such that $\delta =_M \lambda_0$. Fix such a λ_0. Then there are words $u_0, u_1, \dots, u_n \in A^*$ and a sequence

$$\delta \equiv u_0 \leftrightarrow_M u_1 \leftrightarrow_M \cdots \leftrightarrow_M u_{n-1} \leftrightarrow_M u_n \equiv \lambda_0. \tag{3.2.3}$$

In the rewriting (3.2.3), suppose (without loss of generality, by symmetry) that the first letter of δ is affected (in the sense of Novikov [381, I.§1] and Adian [6]) before the last letter is. Suppose the first time, if any, this happens is in the rewriting $u_i \to_M u_{i+1}$. Then $u_i \equiv v_0\delta' v_1$, where $v_0 =_M v_1 =_M 1$ and $\delta' =_M \delta$. The rewriting $u_i \to_M u_{i+1}$ affects the first letter of δ' (which is the same as the first letter of δ) by deleting a defining relation w_j ($1 \leq j \leq k$), and therefore must, by minimality of δ', and invertibility of v_0, v_1, be such that δ' is one of the minimal invertible pieces in the factorisation of this w_j; thus $\delta' \in \Lambda$.

We conclude that for our chosen δ, we can find a piece $\lambda \in \Lambda$ such that $\delta =_M \lambda$, and there is a rewriting $\delta \xleftrightarrow{*}_M \lambda$ which does not affect the first or last letter of δ. Indeed, we can take $\lambda \equiv \delta'$ as above if the first letter of δ is affected in (3.2.3); otherwise, we can take $\lambda \equiv \lambda_0$. In either case, pick the longest λ with the given property. Then there exist $h_1, h_2 \in A^+$ with $\delta \equiv h_1 w h_2$ and $\lambda \equiv h_1 w' h_2$ with $w =_M w'$. Thus there is some $W \in A^*$ with $w \xrightarrow{*}_S W$ and $w' \xrightarrow{*}_S W$. Hence $|w| \geq |W|$ and $|w'| \geq |W|$. As λ was chosen longest, we also have $|w| \leq |w'|$ (for otherwise $\delta \notin \Delta$).

Now, if $|w'| = |W|$, then also $|w| = |W|$. Thus the sequences of rules

$$(s_{1,1}, s_{1,2}), (s_{2,1}, s_{2,2}), \dots, (s_{m,1}, s_{m,2}) \in S$$
$$(s'_{1,1}, s'_{1,2}), (s'_{2,1}, s'_{2,2}), \dots, (s'_{\ell,1}, s'_{\ell,2}) \in S$$

transforming $w \xrightarrow{*}_S W$ resp. $w' \xrightarrow{*}_S W$ satisfies $|s_{i,1}| = |s_{i,2}|$ resp. $|s'_{j,1}| = |s'_{j,2}|$ for all $1 \leq i \leq m$ resp. $1 \leq j \leq \ell$. Thus, by composing the sequence of rules rewriting w' to W with the reverse of the sequence of rules rewriting w to W, we find a sequence of applications of the piece-generating operation rewriting $h_1 w' h_2$ to $h_1 w h_2$. In other words, $h_1 w h_2 \in [h_1 w' h_2]^{\downarrow}$, i.e. $\delta \in [\lambda]^{\downarrow}$. By Makanin's Lemma, we may everywhere replace λ by δ without changing the presentation; in the resulting presentation, whose presentation pieces will be denoted Λ', we have $\delta \in \Lambda'$, and δ contains a large subpiece. As $|\delta| = |\lambda'|$, we have $\omega(\Lambda') = \omega(\Lambda)$, and we are done.

Suppose instead that $|w'| > |W|$. We have $\delta' :\equiv h_1 W h_2 \in [\lambda]^{\downarrow}$. By Makanin's Lemma, we may everywhere replace λ with δ' in the given presentation for M without changing the monoid M. Let Λ' be the new presentation pieces of this presentation. Then, as λ was chosen longest and $|\delta'| < |\lambda|$, we have $\omega(\Lambda') < \omega(\Lambda)$, as $\Lambda' = (\Lambda - \{\lambda\}) \cup \{\delta'\}$ by the second part of Makanin's Lemma. We may thus repeat the above proof for the new presentation, and are done by induction. □

Proof of Proposition 3.2.10. Suppose M has presentation pieces Λ_0. Then M admits a presentation

$$\text{Mon}\langle A \mid w_1 = 1, w_2 = 1, \ldots, w_k = 1\rangle \tag{3.2.4}$$

satisfying the conclusions of Lemma 3.2.15, with presentation pieces Λ resp. pieces Δ, and such that $\omega(\Lambda) \leq \omega(\Lambda_0)$. If (3.2.4) satisfies the small subpiece condition, we are done, so assume the second part of Lemma 3.2.15 holds, and let $\lambda \in \Lambda$ be a presentation piece such that $w \in \Delta^+$ is a large subpiece of λ. Write $\lambda \equiv h_1 w h_2$ with $h_1, h_2 \in A^+$. Introduce a new symbol p, disjoint from A, and add by way of Tietze transformation the relation $p = w$ to the presentation (3.2.4). The resulting presentation is not special. However, as w is invertible, there exists some $w' \in \Lambda^+$ such that $ww' =_M w'w =_M 1$. Hence also $pw' =_M w'p =_M 1$. We add these relations to the presentation. As inverses in a group are unique, and as p and w are both invertible words, we find the relation $p = w$ redundant. We remove it by a Tietze transformation, resulting in a new special presentation:

$$M' = \text{Mon}\langle A \cup \{p\} \mid w_1 = 1, \ldots, w_k = 1, pw' = 1, w'p = 1\rangle. \tag{3.2.5}$$

Now the map induced by $a \mapsto a$ for all $a \in A$ extends to an isomorphism from M to M'. Thus the factorisation of w' and the w_i, for $1 \leq i \leq k$, into minimal invertible pieces is the same in M' as in M. Clearly, p is invertible. It follows that the set Λ' of presentation pieces of (3.2.5) is precisely $\Lambda' = \Lambda \cup \{p\}$.

From the presentation piece $\lambda \equiv h_1 w h_2 \in \Lambda \subset \Lambda'$ in M' we can by the piece-generating operation obtain the piece $\delta := h_1 p h_2$, as $p, w' \in (\Lambda')^*$ satisfy $p =_{M'} w$ and $|p| < |w|$. That is, $\delta \in [\lambda]^{\downarrow}$. By Makanin's Lemma, we can thus in the factorisations of the defining words in (3.2.5) replace λ by without changing the monoid. Let w_i' denote the word obtained by this replacement from w_i (for $1 \leq i \leq k$), and w'' the word from w'. We find a new presentation

$$M'' = \text{Mon}\langle A \cup \{p\} \mid w_1' = 1, \ldots, w_k' = 1, pw'' = 1, w''p = 1\rangle. \tag{3.2.6}$$

Let Λ'' denote the presentation pieces of (3.2.6). As $|p| = 1$ and $|w| > 1$, it follows that $\delta := h_1 p h_2$ satisfies $|\delta| < |\lambda|$. By the second part of Makanin's Lemma, $\delta \in \Lambda''$, and the other presentation pieces of (3.2.6) are presentation pieces of (3.2.5), i.e. in Λ'. Thus $\omega(\Lambda'') < \omega(\Lambda')$. In particular, we find

$$\omega(\Lambda'') < \omega(\Lambda') = \sum_{\lambda' \in \Lambda \cup \{p\}} (|\lambda'| - 1) = (1-1) + \sum_{\lambda' \in \Lambda} (|\lambda'| - 1) = \omega(\Lambda).$$

Thus, repeating the above proof starting with the presentation (3.2.6), either (3.2.6) satisfies the small subpiece condition, or else we obtain a presentation M''' with presentation pieces Λ''' satisfying $\omega(\Lambda''') < \omega(\Lambda'')$, etc. We conclude by induction on ω that there is some $n \geq 0$ and a presentation $M^{(n)}$ with pieces $\Delta^{(n)}$ such that no piece $\delta \in \Delta^{(n)}$ has a large subpiece; that is, $M^{(n)}$ satisfies the small subpiece condition, and defines M. □

Example 3.2.16. Consider the special monoid M with presentation

$$\text{Mon}\langle a, b, c \mid aabbacc = 1, abacab = 1\rangle.$$

Applying Makanin's procedure (see §1.3), we find the overlap group, and find that none of the overlap pieces are equal to some shorter word. Thus we factor the relation words as

$$\mathrm{Mon}\langle a, b, c \mid (aabbacc) = 1, (ab)(ac)(ab) = 1\rangle,$$

and thus easily find that $U(M) \cong \mathbb{Z}$ and $\Delta = \{aabbacc, ab, ac\}$. Evidently, the piece $aabbacc$ contains ab and ac as maximal invertible proper subwords. The inverse for ab in M can be represented by $acab$ (or indeed $abac$, as $abac =_M acab$). Thus, by adding a new generator p and a relation $p = ab$, we can make this relation redundant by adding the two relations $p(acab) = 1$ and $(acab)p = 1$. Hence M is isomorphic to the monoid defined by the presentation

$$\mathrm{Mon}\langle a, b, c, p \mid (aabbacc) = 1, (ab)(ac)(ab) = 1, p(ab)(ac) = 1, (ab)(ac)p = 1\rangle,$$

which has pieces $\Delta' = \{aabbacc, apbacc, ab, ac, p\}$. Now by Lemma 3.2.11, as $p =_M ab$ we may everywhere replace $a(ab)bacc$ by $a(p)bacc$ without changing the monoid; thus M is isomorphic to the monoid presented by

$$\mathrm{Mon}\langle a, b, c, p \mid (apbacc) = 1, (ab)(ac)(ab) = 1, p(ab)(ac) = 1, (ab)(ac)p = 1\rangle.$$

In this presentation, $aabbacc$ is no longer a piece, as $apbacc$ is the longest piece in its congruence class which appears in some relator of the presentation, and $|aabbacc| > |apbacc|$. Hence the pieces of this presentation are $\{p, ab, ac, apbacc\}$. If we now introduce the generator q and the relation $q = ac$, we can take as inverse of ac the word $abab$; then everywhere replace the piece $apbacc$ by $apbqc$, which leaves us with the presentation

$$\begin{aligned}\mathrm{Mon}\langle a, b, c, p, q \mid &(apbqc) = 1, (ab)(ac)(ab) = 1,\\ &(p)(ab)(ac) = 1, (ab)(ac)(p) = 1,\\ &(q)(ab)(ab) = 1, (ab)(ab)(q) = 1\rangle.\end{aligned}$$

The pieces of this presentation can now be checked, again, with Makanin's procedure to be $\Delta = \{p, q, ab, ac, apbqc\}$. This presentation satisfies the conclusions of the lemma. △

Example 3.2.17. Consider the special monoid M with presentation

$$\mathrm{Mon}\langle a, b, c \mid aaabccc = 1, aabccabcaabcc = 1\rangle.$$

Using Makanin's algorithm, we factor the relators into pieces as

$$\mathrm{Mon}\langle a, b, c \mid (aaabccc) = 1, (aabcc)(abc)(aabcc) = 1\rangle.$$

Evidently, the piece $aaabccc$ contains both $aabcc$ and abc as subwords. We choose $w \equiv aabcc$, as this is maximal. Let p be a new symbol and add the relation $p = aabcc$, giving

$$\mathrm{Mon}\langle a, b, c, p \mid (aaabccc) = 1, (aabcc)(abc)(aabcc) = 1, p = aabcc\rangle.$$

Now $aabcc$ has an inverse $(abc)(aabcc)$, and so adding the relations

$$p(abc)(aabcc) = 1 \qquad \text{and} \qquad (abc)(aabcc)p = 1$$

we see that $p = aabcc$ is now redundant, and hence M admits the presentation

$$\begin{aligned}\mathrm{Mon}\langle a, b, c, p \mid &(aaabccc) = 1, (aabcc)(abc)(aabcc) = 1,\\ &(p)(abc)(aabcc) = 1, (abc)(aabcc)(p) = 1\rangle.\end{aligned}$$

Now we have that the piece $aaabccc \equiv a(aabcc)c =_M apc$, and hence by Lemma 3.2.11 we may replace $aaabccc$ by apc in the above presentation without changing the monoid presented by it. Hence M is isomorphic to the monoid defined by the presentation

$$\begin{aligned}\operatorname{Mon}\langle a, b, c, p \mid (apc) = 1, (aabcc)(abc)(aabcc) = 1, \\ p(abc)(aabcc) = 1, (abc)(aabcc)p = 1\rangle.\end{aligned}$$

By carrying out the same procedure again, this time choosing the maximal word $w \equiv abc$, which appears in the piece $aabcc$, choosing the letter q to replace it, adding the necessary invertibility conditions and replacing $aabcc$ by aqc we obtain the presentation

$$\begin{aligned}\operatorname{Mon}\langle a, b, c, p \mid (apc) = 1, (aqc)(abc)(aqc) = 1, \\ (p)(abc)(aqc) = 1, (abc)(aqc)(p) = 1, \\ (q)(aqc)(aqc) = 1, (aqc)(aqc)(q) = 1\rangle\end{aligned}$$

which defines M, and in which all invertible subwords of any piece have length 1. $\triangle$

The fact that all pieces appearing inside pieces can be forced to have length 1 bodes well for applications of monadic rewriting systems, as we shall see in the sequel.[56]

3.2.3 Representatives of pieces

Recall that X is a set in bijective correspondence of cardinality the size ν of the partition of Δ as $\Delta_1 \cup \Delta_2 \cup \cdots \cup \Delta_\nu$ into pieces equal to each other in M, and that $\phi\colon \Delta^* \to X^*$ is the canonical surjective homomorphism. For a word $w \in \Delta^*$, we define the language of Δ-*representatives* of w as the set

$$\operatorname{Rep} \Delta_A^M(w) := \{v \in \Delta^* \mid v =_M w\} = \phi^{-1}\left(\operatorname{Rep}_X^{U(M)}(\phi(w))\right). \tag{3.2.7}$$

The following easy proposition is immediate by definition of $\operatorname{Rep} \Delta_A^M(w)$.

Proposition 3.2.18. *Let M be a finitely presented special monoid with finite generating set A, with Δ and X as usual. Let $\mathcal{C}$ be a class of languages closed under homomorphism and inverse homomorphism. Then for all $w \in \Delta^*$ we have*

$$\operatorname{Rep} \Delta_A^M(w) \in \mathcal{C} \quad \iff \quad \operatorname{Rep}_X^{U(M)}(\phi(w)) \in \mathcal{C}.$$

Proof. ($\Longrightarrow$) Suppose $\operatorname{Rep} \Delta_A^M(w) \in \mathcal{C}$. As ϕ is surjective, $\phi \circ \phi^{-1}$ is well-defined on all subsets of X^*, and is the identity function on such subsets. Thus

$$\begin{aligned}\operatorname{Rep}_X^{U(M)}(\phi(w)) &= (\phi \circ \phi^{-1})\left(\operatorname{Rep}_X^{U(M)}(\phi(w))\right) \\ &= \phi\left(\phi^{-1}\left(\operatorname{Rep}_X^{U(M)}(\phi(w))\right)\right) \\ &:= \phi\left(\operatorname{Rep} \Delta_A^M(w)\right)\end{aligned}$$

and the final term is in $\mathcal{C}$ as $\mathcal{C}$ is closed under homomorphism.

($\Longleftarrow$) Suppose that $\operatorname{Rep}_X^{U(M)}(\phi(w)) \in \mathcal{C}$. As $\operatorname{Rep}_X^{U(M)}(\phi(w)) = \phi^{-1}\left(\operatorname{Rep} \Delta_A^M(w)\right)$, it follows that $\operatorname{Rep}_X^{U(M)}(\phi(w)) \in \mathcal{C}$ as $\mathcal{C}$ is closed under inverse homomorphism. $\square$

[56] However, the above changing of presentations is not the only path towards proving the below theorems. If no manipulation of presentations were to be carried out, the proof works the same, up to reducing the problem to questions about a rewriting system in which the right-hand sides are all single elements from some biprefix code, and the left-hand sides are all words over this code (rather than a monadic rewriting system). It seems very conceivable that one could easily develop enough theory for such rewriting systems – which behave very much like generalised monadic rewriting systems – to prove any statements we might need. We have chosen this more conventional path, instead.

For example, in the bicyclic monoid $B = \mathrm{Mon}\langle b, c \mid bc = 1\rangle$, with $\Delta = \{bc\}$, $X = \{x_1\}$, and $U(B) = \mathrm{Mon}\langle x_1 \mid x_1 = 1\rangle$, we have

$$\mathrm{Rep}\,\Delta^B_{\{b,c\}}(1) = \phi^{-1}\left(\mathrm{Rep}^{U(B)}_X(\phi(1))\right) = \phi^{-1}(x_1^*) = (bc)^*.$$

Thus $\mathrm{Rep}\,\Delta^B_{\{b,c\}}(1)$ is regular, as $U(B)$ is a group with regular word problem.

The idea of this section is as follows. We will describe the language of representatives $\mathrm{Rep}^M_A(\delta)$ of a given piece $\delta \in \Delta$ as the set of ancestors of $\mathrm{Rep}\,\Delta^M_A(\delta)$ under a certain monadic rewriting system, which in turn is controlled by the group of units $U(M)$. Because $\mathrm{Rep}\,\Delta^M_A(\delta)$ can be understood in terms of $U(M)$ by (3.2.7), this gives an understanding of $\mathrm{Rep}^M_A(\delta)$ in terms of $U(M)$.

We will define the rewriting system

$$\mathcal{R}_\Delta = \bigcup_{\substack{p \in \Delta \cup \{\varepsilon\} \\ |p| \leq 1}} \left\{(W_p \to p) \mid W_p \in \mathrm{Rep}\,\Delta^M_A(p)\right\}. \tag{3.2.8}$$

In general, this is an infinite rewriting system. Furthermore, it is not in general a complete rewriting system. However, it has the following desirable property: let $\mathcal{C}$ be a class of languages closed under inverse homomorphism such that $U(M)$ has word problem in $\mathcal{C}$. Then for every $p \in \Delta \cup \{\varepsilon\}$ with $|p| \leq 1$ (i.e. for every right-hand side in $\mathcal{R}_\Delta$), the language $\mathrm{Rep}\,\Delta^M_A(p)$ is in $\mathcal{C}$, as $\mathrm{Rep}^{U(M)}_X(\phi(p)) \in \mathcal{C}$. Thus $\mathcal{R}_\Delta$ is a monadic $\mathcal{C}$-rewriting system.

Example 3.2.19. Consider the monoid $\mathrm{Mon}\langle b, c \mid abc = 1, b^2 = 1\rangle$, with pieces readily[57] computed by Makanin's procedure as $\Delta = \{abc, b\}$. Now the rules

$$(b^{2k+1} \to b), \quad (k \geq 0)$$

are all elements of $\mathcal{R}_\Delta$, for $\{b^{2k+1} \mid k \geq 0\} \subseteq \mathrm{Rep}\,\Delta^M_A(b)$. In fact, by considering all equalities in the group of units

$$U(M) \cong \mathcal{O}(M) \cong \mathrm{Mon}\langle b_1, b_2 \mid b_1 = 1, b_2^2 = 1\rangle$$

with the isomorphism induced by $abc \mapsto b_1$ and $b \mapsto b_2$, we can see that the full set $\mathrm{Rep}\,\Delta^M_A(b)$ is given by $\langle\{b^{2k+1} \mid k \geq 0\}\rangle_{I_1}$, where we define $I_1 = \{(abc)^+ \to \varepsilon, b^2 \to \varepsilon\}$. We note also that $I_1 \subseteq \mathcal{R}_\Delta$. $\triangle$

Before proving the lemma, we shall give understanding of why we need to prove it, some reasons to believe it might be true, and a simple case of when it is true. First, note that in groups, due to the presence of overwhelming symmetry it is sufficient to understand the words equal to 1 in order to understand the words equal to any given word. In monoids, this is not necessarily true.[58] On the other hand, in special monoids, we shall see that there is a similar kind of symmetry among words equal to a given *piece*, and that understanding the set of words equal to a single *fixed* piece $\delta \in \Delta$ gives full control over which invertible words are equal with one another (Theorem 3.3.2).

[57] We have $C(abc, b) = \{abc, b\}$. In the (finite!) overlap group $\mathcal{O}(M) \cong \mathrm{Gp}\langle b_1, b_2 \mid b_1 = 1, b_2^2 = 1\rangle$ it is easy to solve the word problem. One lifts this to words over the overlap pieces in M using ϕ, and sees that abc, being the only overlap piece containing another overlap piece (in this case, b), is not equal to a word ah_1c with $h_1 \in \Delta^*$ and $|h_1| < |b|$ in M, and so $\Delta = \{abc, b\}$ by Makanin's procedure.

[58] For example, in $\mathrm{Mon}\langle a, b \mid abaab = a\rangle$ only the empty word is equal to 1, but this is a very complicated one-relation monoid; the decidability of its word problem was even an open problem for some time [221].

Now, for an example to illustrate the forthcoming lemma, let $B = \text{Mon}\langle b, c \mid bc = 1\rangle$ be the (old faithful) bicyclic monoid. Let $A = \{b, c\}$ for ease of notation. We have that $\Delta = \{bc\}$, and this presentation satisfies the conclusions of Proposition 3.2.10. What are the elements of $\text{Rep}_A^B(\varepsilon)$? That is, what words in $\{b, c\}^*$ are equal to 1? Certainly any word in $\Delta^* = (bc)^*$ is equal to 1. Thus $(bc)^*$ is the set of words over the pieces which equal 1; in other words $\text{Rep}\,\Delta_A^B(\varepsilon) = (bc)^*$. Thus $\mathcal{R}_\Delta$ has as rules $((bc)^n \to \varepsilon)$ for all $n \geq 1$. Now, $\text{Rep}_A^B(bc)$ has more elements than just those from $\text{Rep}\,\Delta_A^B(bc)$. For example, any number of insertions of bc in any place will yield another element of $\text{Rep}_A^B(bc)$, such as $b(b(bc)c)c$, even though $bbbccc \notin \text{Rep}\,\Delta_A^B(bc)$. And yet, we notice that $bbbccc \xrightarrow{*}_{\mathcal{R}_\Delta} bc$, by removing precisely these insertions. Lemma 3.2.20 below will tell us that all words in $\text{Rep}_A^B(bc)$ can be obtained in precisely this way, by monadically rewriting down to an element of Δ^* using induction on the "depth" of the insertions.

In the more general case, we cannot simply remove the defining relations from inside pieces. For example, in

$$\Pi = \text{Mon}\langle p, q, b, c, d, e \mid (pq)b(pq) = 1, (pq)c(pq) = 1, b^2 = 1, dxe = 1, x = 1\rangle,$$

we can check that $\Delta = \{x, b, c, pq, de, dxe\}$, and that $b =_\Pi c$. Let $A = \{p, q, b, c, d, e, x\}$. Note that this presentation satisfies the small subpiece condition. Now

$$dxe =_\Pi d(pqcpq)e =_\Pi d(p(bcbc)qcpq)e$$

as $bc =_\Pi 1$. Hence $d(p(bcbc)qcpq)e \in \text{Rep}_A^\Pi(dxe)$. There is no way to go from $d(p(bcbc)qcpq)e$ to dxe by simply removing defining relations, as $d(p(bcbc)qcpq)e$ does not contain any defining relation as a subword! However, we can first remove $bcbc \in \Delta^*$ and replace it with the empty word, as $bcbc \in \text{Rep}\,\Delta_A^\Pi(\varepsilon)$; and then replace $pqcpq$ by x, as $pqcpq \in \text{Rep}\,\Delta_A^\Pi(x)$. Hence, performing these steps backwards, we have

$$d(p(bcbc)qcpq)e \to_{\mathcal{R}_\Delta} d(pqcpq)e \to_{\mathcal{R}_\Delta} dxe$$

as desired.

The key point to proving that can be done in general for an arbitrary special monoid M (with A, Δ, etc. as usual) satisfying the conditions of Proposition 3.2.10 is the following. If w is an invertible word, then $w \xrightarrow{*}_{S(M)} \delta_1\delta_2 \cdots \delta_\mu$. Now overlap arguments imply that applying a rule from $S(M)$ in reverse to $\delta_1\delta_2 \cdots \delta_\mu$ has exactly one of the following two effects: (1) either a subword $\delta_\lambda\delta_{\lambda+1} \cdots \delta_{\lambda'}$ is replaced by some other word from Δ^*, leaving us again with a word from Δ^*; or else (2) a word from Δ^* entirely contained *inside* one of the δ_i is replaced by a word from Δ^*. In case (1), we can shorten the rewriting and apply induction; and in (2), our control on the presentation gives that this is a monadic rewriting step, as any word from Δ^* inside δ_i must be a single letter! We then proceed by induction, and thereby show that all rewriting steps in $\xrightarrow{*}_{S(M)}$ can be replaced by rewriting steps in $\xrightarrow{*}_{\mathcal{R}_\Delta}$. We now give the actual proof in detail.[59]

Before finally showing the key technical lemma (Lemma 3.2.20), we introduce some useful

[59] The interested and diligent reader will find that some ideas of the proof presented below bear some resemblance to ideas given in the proofs of [309, Lemma 20 & Lemma 21]. This was discovered by the author some time after proving the below lemma. We also wish to emphasise that the language used in these aforementioned proofs is very different from the one given below.

terminology. For any word in $\overline{\Delta}^*$, we can obtain a word in Δ^* by successively removing left-hand sides of rules in $S(M)$, replacing them by their corresponding right-hand sides. We will consider this process in reverse, attributing the terminology of this idea to Cain & Maltcev [96]. First, let $u \in \Delta^*$ and factorise $u \equiv \delta_1\delta_2 \cdots \delta_n$ uniquely, where $\delta_i \in \Delta$ for $1 \leq i \leq n$. Then every non-empty subword of u of the form $\delta_j\delta_{j+1} \cdots \delta_\ell$ for $1 \leq j \leq \ell \leq n$ is called a *depth*-0 *inserted word.* Inductively, for $\mu \geq 0$, we define a depth-$(\mu + 1)$ insertion as follows: if (1) a right-hand side s_2 of a rule $(s_1 \to s_2) \in S(M)$ appearing as a proper non-prefix, non-suffix subword of some depth-μ inserted word D, with $D \in \Delta^*$, is replaced by s_1, then we call that occurrence of s_1 a *depth-*$(\mu + 1)$ *inserted word*, and the reversed rewriting $(s_2 \to s_1)$ is then called a depth-$(\mu + 1)$ *insertion*; but (2) if instead the specified occurrence of s_2 is a depth-μ inserted word $D \in \Delta^*$, or if $s_2 \equiv \varepsilon$ and does not satisfy the condition in (1), then the word s_1 is a depth-μ inserted word. The reversed rewriting $(s_2 \to s_1)$ is then called a depth-μ insertion.

We give a concrete example. If $\Delta = \{d, b, abc\}$, and (for simplicity) we have the rewriting system $\mathcal{T}$ with the rules $\{dbd \to b, abc \to \varepsilon\}$, then an ancestor of the word $u \equiv (abc)(abc)(b) \in \Delta^*$ modulo $\mathcal{T}$ might look like:

$$u' \equiv (adbdc)(abababccc)(dababccdbdd) \equiv (\underbrace{adbdc}_{\text{depth } 0})(ab\underbrace{ab\overbrace{abc}^{\text{depth } 2}c}_{\text{depth } 1}c)(\underbrace{dab\overbrace{abc}^{\text{depth } 1}cdbdd}_{\text{depth } 0}).$$

Thus, the word abc in the middle of the word u' is a depth-2 inserted word, and the rewriting of the leftmost term $abdbc$ to abc is via the reverse of a depth-0 insertion $(b \to dbd)$. Now, just as in [96, Example 4.2], it is clear by definition of insertions (using no properties of the rewriting systems involved) that since $u' \xrightarrow{*}_{\mathcal{T}} u$, we can perform this rewriting by first rewriting the depth-2 insertions in reverse, then the depth-1 insertions in reverse, and finally have a depth-0 inserted word in Δ^*, which is then rewritten to u.

Lemma 3.2.20. *Let M be a finitely presented special monoid, with finite generating set A and minimal invertible pieces Δ. If the presentation for M satisfies the small subpiece condition, then for every $\delta \in \Delta$ we have* $\operatorname{Rep}_A^M(\delta) = \langle \operatorname{Rep} \Delta_A^M(\delta) \rangle_{\mathcal{R}_\Delta}$.

Proof. Let us assume that M is given by the presentation

$$\operatorname{Mon}\langle A \mid w_1 = 1, w_2 = 1, \ldots, w_k = 1 \rangle \tag{3.2.9}$$

with pieces Δ, satisfying the small subpiece condition. We will prove that for this presentation $\operatorname{Rep}_A^M(\delta) = \langle \operatorname{Rep} \Delta_A^M(\delta) \rangle_{\mathcal{R}_\Delta}$.

($\supseteq$) Let $w \in \operatorname{Rep} \Delta_A^M(\delta)$ and $w' \in A^*$ be arbitrary words such that $w' \in \langle w \rangle_{\mathcal{R}_\Delta}$, i.e. $w' \xrightarrow{*}_{\mathcal{R}_\Delta} w$. Now, for every rule $(W_p, p) \in \mathcal{R}_\Delta$, we have by definition that $p =_M W_p$. Thus, by induction on the number of rules applied in rewriting w' to w, we have $w' =_M w$. As $w =_M \delta$, we have $w' \in \operatorname{Rep}_A^M(\delta)$.

($\subseteq$) Let $w \in \operatorname{Rep}_A^M(\delta)$. Then $w =_M \delta$, so w is invertible. In particular, $w \in \overline{\Delta}^*$ by Lemma 3.2.4, and there is some $u \in \Delta^*$ such that $w \xrightarrow{*}_S u$. By the earlier reasoning, we can thus obtain w from u by first performing all depth-0 insertions, then all depth-1 insertions, etc. until after performing a finite number insertions we obtain w. Let $\mu \geq 0$ be the highest depth of any such insertion performed.

We claim that $w \in \langle \operatorname{Rep} \Delta_A^M(u) \rangle_{\mathcal{R}_\Delta}$ by induction on this μ. The base case $\mu = 0$ is clear, for then $w \in \Delta^*$. As for every rule $(s_1 \to s_2) \in S(M)$ we have $s_1 =_M s_2$, it follows by induction on the number of rules applied in rewriting $w \xrightarrow{*}_S u$ that $w =_M u$. As $w \in \Delta^*$, we have $w \in \operatorname{Rep} \Delta_A^M(u) \subseteq \langle \operatorname{Rep} \Delta_A^M(u) \rangle_{\mathcal{R}_\Delta}$. Assume, then, for induction that the claim is true for some $\mu \geq 0$, and that w requires depth-$(\mu + 1)$ insertions (but no higher). In the fixed rewriting $w \xrightarrow{*}_S u$, let $u' \in A^*$ be such that $w \to_S u' \xrightarrow{*}_S u$. Then the rewriting $w \to_S u'$ is by replacing the depth-$(\mu + 1)$ inserted word $s_1 \in \Delta^*$ in w with the word s_2, where s_2 is a proper non-prefix non-suffix subword of either (I) a depth-$(\mu + 1)$ word $Q \in \Delta^*$, or (II) a depth-μ inserted piece $Q \in \Delta$; and where $(s_1 \to s_2) \in S(M)$ is the specified rule.

In case (I), write $Q \equiv Q_0 s_2 Q_1$, where necessarily $Q_0, Q_1 \in \Delta^*$. As Q is a depth-$(\mu + 1)$ inserted word in u', the word u' can be obtained from some word $u'' \in A^*$ by replacing a depth-$(\mu + 1)$ or a depth-μ inserted word $s_3 \in \Delta^*$ in u'' with Q. That is, there is some rule $(Q \to s_3) \in S(M)$, which rewrites $u' \to_S u''$. But as $Q_0 s_1 Q_1 =_M Q_0 s_2 Q_1 \equiv Q =_M s_3$, and $|Q_0 s_1 Q_2| \geq |Q_0 s_2 Q_1| = |Q| \geq |s_3|$, we have $(Q_0 s_1 Q_1 \to s_3) \in S(M)$. Hence, we can obtain u already from u'' by performing the insertion of replacing s_3 by s_1, i.e. $u \to_S u''$ by the rule $(Q_0 s_1 Q_1 \to s_3)$, thus reducing the rewriting $w \xrightarrow{*}_S u$ by one step; we may thus by another induction assume without loss of generality that w is obtained from u' as in case (II).

Thus, assume case (II), i.e. $Q \in \Delta$ is a depth-μ inserted word in u', and s_2 appears as a proper non-suffix non-prefix subword of the piece Q. As the presentation satisfies the small subpiece condition, it follows from $s_2 \in \Delta^*$ that the subpiece s_2 satisfies $|s_2| \leq 1$. Hence also $s_2 \in \Delta \cup \{\varepsilon\}$. As $s_1 =_M s_2$ and $s_1 \in \Delta^*$, we have $s_1 \in \operatorname{Rep} \Delta_A^M(s_2)$. Hence, $(s_1 \to s_2) \in \mathcal{R}_\Delta$. Thus, w can be rewritten to u' in a single application of a rule from $\mathcal{R}_\Delta$, and so, by repeating the same step for all depth-$(\mu + 1)$ insertions, we find that there is a word $w' \in A^*$ such that (1) $w \xrightarrow{*}_{\mathcal{R}_\Delta} w'$; and (2) w' can be obtained from the word u with at most depth-μ insertions. By the inductive hypothesis, thus $w' \in \langle u \rangle_{\mathcal{R}_\Delta}$, and hence also $w \in \langle u \rangle_{\mathcal{R}_\Delta}$. Now $u \in \operatorname{Rep} \Delta_A^M(\delta)$, as $u \in \Delta^*$ and $u =_M \delta$, and so we conclude that $w \in \langle \operatorname{Rep} \Delta_A^M(\delta) \rangle_{\mathcal{R}_\Delta}$, which is what was to be shown. □

Remark 3.2.1. The depth of inserted words appearing in the proof of the above lemma is closely associated to the depth of vertices of the Schützenberger graph $\mathfrak{R}_1$. We shall present this theory in full detail in Chapter 5. More specifically, and assuming some familiarity with the notions in that chapter, if w is an invertible word, and w' is a depth-λ inserted word, then for every prefix $w_0 w'$ of w, we have that the *depth* of the vertex $\pi(w_0 w')$ in $\mathfrak{R}_1$ is $\leq \lambda$. As mentioned, Makanin defines a set closely related to our $\overline{\Delta}$ as the set of *d-words* of a special monoid; for such words, say w, he defines the *rank* of w, and this is, with some minor modifications and translated, essentially the largest λ such that w contains a depth-λ inserted word, see [309, Definition 16, Definition 18].

We have now described all the words equal to a single piece δ as the set of monadic ancestors of the set of Δ-representatives of δ. This is a large part of the "context-free" increase in complexity of equality of words over Δ vs. the equality of invertible words which we described earlier in the chapter. Using the above lemma we can thus easily conclude the following, which can be interpreted as saying "if the set of Δ-representatives of pieces can be described using a class of languages $\mathcal{C}$, and $\mathcal{C}$ is well-behaved under taking monadic ancestors, then the set of *all* representatives of pieces can be described using $\mathcal{C}$".

Theorem 3.2.21. *Let M be a finitely presented special monoid, generated by a finite set A, with group of units $U(M)$ generated by some finite set X. Let $\mathcal{C}$ be a super-AFL. Then M admits a special presentation with pieces such that:*

$$\mathrm{WP}_X^{U(M)} \in \mathcal{C} \quad \implies \quad \mathrm{Rep}_A^M(\delta) \in \mathcal{C} \quad \text{for all } \delta \in \Delta,$$

with Δ the pieces of the presentation.

Proof. By Proposition 3.2.10, M admits a presentation satisfying the small subpiece condition. Let M be given by such a presentation, with A, Δ, etc. as usual. Suppose that $\mathrm{WP}_X^{U(M)} \in \mathcal{C}$. Then by Proposition 3.2.18, for all $p \in \Delta \cup \{\varepsilon\}$ with $|p| \leq 1$ we have $\mathrm{Rep}\, \Delta_A^M(p) \in \mathcal{C}$. As $\mathcal{C}$ is closed under finite unions, by the definition of $\mathcal{R}_\Delta$ we have that $\mathcal{R}_\Delta$ is a $\mathcal{C}$-=-monadic rewriting system. By Lemma 3.2.20, we have $\mathrm{Rep}_A^M(\delta) = \langle \mathrm{Rep}\, \Delta_A^M(\delta) \rangle_{\mathcal{R}_\Delta}$ for all $\delta \in \Delta$. As $\mathcal{C}$ has the monadic ancestor property and $\mathrm{Rep}\, \Delta_A^M(\delta) \in \mathcal{C}$, thus also $\mathrm{Rep}_A^M(\delta) \in \mathcal{C}$ for all $\delta \in \Delta$. □

Now that we have described the manner in which one passes from the group of units to the invertible elements – and, more importantly, described this language-theoretically – we are poised to use these in order to describe when invertible elements are equal to one another. This is, compared to the rather involved manipulations in the above section, rather straightforward, albeit technical at times.

3.3 The invertible word problem

The Adian-Makanin-Zhang theory of special monoids tells us that understanding equality of invertible words is crucial to understanding equality of all words. We thus pose the following natural question:

When are two invertible words equal in a special monoid M?

This is a non-trivial question; but we shall answer it here. Of course, some rôle must be played by the group of units; but in principle the structure of invertible words could be significantly more difficult than the structure of the group of units. We shall uncover that, language-theoretically, the structure of equality of invertible words is essentially that of the structure of equality of elements from $U(M)$, but with either side of the equality recursively iterated some finite number of steps via pieces appearing as subwords of other pieces. Fortunately, we have a tool for dealing with recursive statements, as we have seen before: the monadic ancestor property. This will be applied in a very similar way to passing from elements of $U(M)$ to representatives $\mathrm{Rep}_A^M(\delta)$ of single pieces in the proof of Theorem 3.2.21.

Motivated by the word problem WP_A^M for M, we define the set InvP_A^M, called the *invertible word problem* of M with respect to A, as

$$\mathrm{InvP}_A^M := \{w_1 \# w_2^{\mathrm{rev}} \mid w_1, w_2 \in A^* \text{ invertible}, w_1 =_M w_2\} = \mathrm{WP}_A^M \cap (\overline{\Delta}^* \# (\overline{\Delta}^{\mathrm{rev}})^*)$$

and note that the last equality is a direct consequence of Lemma 3.2.4.

What makes this language worthy of study? The Normal Form Lemma (Lemma 1.3.8) directly informs us that to understand when two words are equal, we must understand an "alternating product" (in the sense of Chapter 2) of a free monoid and InvP_A^M. But to understand such an alternating product, it suffices to understand equality of invertible words (as long as the monadic ancestor property is involved), as free monoids are well-behaved. The following lemma captures this idea, though without using the involved language of the alternating products introduced in Chapter 2.

Lemma 3.3.1. *Let $\mathcal{C}$ be a super-AFL. Then* $\mathrm{InvP}_A^M \in \mathcal{C} \implies \mathrm{WP}_A^M \in \mathcal{C}$.

Proof. The reader familiar with the alternating products introduced in Chapter 2 will find this lemma obvious. We give a direct proof instead. The idea of the proof is to from InvP_A^M construct a $\mathcal{C}$-monadic rewriting system $\mathcal{R}$, with the property that $\langle \# \rangle_{\mathcal{R}} = \mathrm{WP}_A^M$. As $\mathcal{C}$ has the monadic ancestor property, the result will follow.

The rules of the rewriting system $\mathcal{R}$, over the alphabet $A \cup \{\#\}$, will be

$$\{u\#v^{\mathrm{rev}} \to \# \mid u\#v^{\mathrm{rev}} \in \mathrm{InvP}_A^M\} \cup \{a_i \# a_i \to \# \mid a_i \in A\}.$$

That $\mathcal{R}$ is a $\mathcal{C}$-=-monadic rewriting system is clear; the language of left-hand sides in $\{u\#v^{\mathrm{rev}} \to \# \mid u\#v^{\mathrm{rev}} \in \mathrm{InvP}_A^M\}$ is precisely InvP_A^M, and hence in $\mathcal{C}$, and the language of left-hand sides in $\{a_i\#a_i \to \# \mid a_i \in A\}$ is a finite language (as M is finitely generated), and hence in $\mathcal{C}$. Thus, as $\mathcal{C}$ is closed under unions, $\mathcal{R}$ is indeed a $\mathcal{C}$-rewriting system, and it is of course $=$-monadic.

We will now show that $\langle \# \rangle_{\mathcal{R}} = \mathrm{WP}_A^M$.

($\subseteq$) Suppose that $w \in \langle \# \rangle_{\mathcal{R}}$. Then there exist some minimal $k \geq 0$ such that there exists w_i for $i = 0, 1, \dots, k$ with

$$w \equiv w_0 \to_{\mathcal{R}} w_1 \to_{\mathcal{R}} \cdots \to_{\mathcal{R}} w_{k-1} \to_{\mathcal{R}} w_k \equiv \#.$$

We prove by induction on k that $w \in \mathrm{WP}_A^M$. The base case $k = 0$ is trivial, for then $w \equiv \#$ and $\# \in \mathrm{WP}_A^M$ follows from the equality $1 =_M 1$. Suppose that the claim is true for some $n \geq 0$, and let w be as above with $k = n + 1$. Then w_1 rewrites to $\#$ in $k = n$ steps, and by the induction hypothesis we have $w_1 \in \mathrm{WP}_A^M$. Hence we can write $w_1 \equiv w_1' \# (w_1'')^{\text{rev}}$, where $w_1' =_M w_1''$. Suppose that the rule $r \in \mathcal{R}$ is such that w rewrites to w_1 by applying the rule r somewhere inside w. Then $r = (u \# v^{\text{rev}} \to \#)$ with $u \# v^{\text{rev}} \in \mathrm{InvP}_A^M$ or $u \equiv v \equiv a_i$ for some $a_i \in A$. As w_1 contains exactly one occurrence of $\#$, we must have that

$$w \equiv w_0 \equiv w_1'(u \# v^{\text{rev}})(w_1'')^{\text{rev}} \equiv w_1' u \# (w_1'' v)^{\text{rev}}.$$

Thus to show that $w \in \mathrm{WP}_A^M$, it suffices to show that $w_1' u =_M w_1'' v$. But as $u \# v^{\text{rev}} \in \mathrm{InvP}_A^M$ or $u \equiv v \equiv a_i$, we have $u =_M v$. As $w_1' =_M w_1''$, we hence have $w_1' u =_M w_1'' u =_M w_1'' v$ and the result follows.

($\supseteq$) Suppose that $w \in \mathrm{WP}_A^M$. Then $w \equiv u \# v^{\text{rev}}$ for some $u, v \in A^*$, with $u =_M v$. Hence, by Lemma 1.3.8, we can factorise u and v as

$$u \equiv u_0 a_1 u_1 \cdots a_m u_m$$
$$v \equiv v_0 a_1 v_1 \cdots a_m v_m$$

respectively, where for all $0 \leq i \leq m$ we have $a_i \in A$, $u_i =_M v_i$, and u_i (resp. v_i) is a maximal invertible factor of u (resp. v). As u_i, v_i are invertible, $u_i \# v_i^{\text{rev}} \in \mathrm{InvP}_A^M$ for all $0 \leq i \leq m$. Now

$$\begin{aligned} w &\equiv u_0 a_1 u_1 \cdots a_m u_m \# (v_0 a_1 v_1 \cdots a_m v_m)^{\text{rev}} \\ &\equiv u_0 a_1 u_1 \cdots a_m u_m \# v_m^{\text{rev}} a_m \cdots a_1 v_0^{\text{rev}}. \end{aligned}$$

But now clearly $w \in \langle \# \rangle_{\mathcal{R}}$, for $u_0 \# v_0^{\text{rev}} \in \mathrm{InvP}_A^M$ implies $(u_0 \# v_0^{\text{rev}} \to \#) \in \mathcal{R}$ and hence $u_0 \# v_0^{\text{rev}} \in \langle \# \rangle_{\mathcal{R}}$. Furthermore, we have $(a_1 \# a_1 \to \#) \in \mathcal{R}$, and hence $u_0 a_1 \# a_1 v_0^{\text{rev}} \in \langle \# \rangle_{\mathcal{R}}$. By this alternating process, we find that

$$\begin{aligned} &\# && \in \langle \# \rangle_{\mathcal{R}} \\ &(u_0 \# v_0^{\text{rev}}) && \in \langle \# \rangle_{\mathcal{R}} \\ &u_0 (a_1 \# a_1) v_0^{\text{rev}} && \in \langle \# \rangle_{\mathcal{R}} \\ &u_0 a_1 (u_1 \# v_1^{\text{rev}}) a_1 v_0^{\text{rev}} && \in \langle \# \rangle_{\mathcal{R}} \\ &\vdots && \\ w \equiv\ &u_0 a_1 u_1 \cdots a_m (u_m \# v_m^{\text{rev}}) a_m \cdots a_1 v_0^{\text{rev}} && \in \langle \# \rangle_{\mathcal{R}} \end{aligned}$$

Hence $w \in \langle \# \rangle_{\mathcal{R}}$, as desired. This completes the proof of Lemma 3.3.1. □

Thus we wish to understand InvP_A^M in terms of $U(M)$, for if we understand it well enough, then we understand WP_A^M in terms of $U(M)$. We will now use our understanding of the representative words of pieces and the properties of generalised pieces $\overline{\Delta}$ to present a full characterisation of InvP_A^M in terms of the word problem for $U(M)$. The reader will note that there is an assumption made on the presentation of the special monoid; it seems likely that the statement of this theorem holds true for all special monoids with no assumption on the presentation, but as we do not need this, we do not pursue it.

Theorem 3.3.2. *Let $\mathcal{C}$ be a super-AFL closed under reversal. Let M be a finitely presented special monoid. Then M admits a finite special monoid presentation with generating set A, and with pieces Δ and associated set X, such that:*

$$\mathrm{WP}_X^{U(M)} \in \mathcal{C} \quad \iff \quad \mathrm{InvP}_A^M \in \mathcal{C}.$$

Proof. Let M be a special monoid. Then as all assumptions about $\mathcal{C}$ in the statement of Theorem 3.2.21 hold, it follows that M admits a presentation, with pieces Δ (and X, ϕ as usual) satisfying the conclusions of Theorem 3.2.21. We assume M is given by this presentation. In particular, we have $\mathrm{WP}_X^{U(M)} \in \mathcal{C} \implies \mathrm{Rep}_A^M(\delta) \in \mathcal{C}$ for all $\delta \in \Delta$.

($\impliedby$) Assume that $\mathrm{InvP}_A^M \in \mathcal{C}$. For notational brevity, write $\Delta_r := \Delta^{\mathrm{rev}}$. We begin by noting that as Δ is a biprefix code, so too is Δ_r.[60] Now, $\bullet^{\mathrm{rev}}$ acts involutively on Δ, so in particular the map $\phi_r \colon \Delta_r \to X$ defined by $\delta^{\mathrm{rev}} \mapsto \phi(\delta)$ is well-defined. As Δ_r is a biprefix code, we can hence extend ϕ_r to a homomorphism (denoted by the same symbol) $\phi_r \colon \Delta_r^* \to X^*$. Note that $\phi_r = \phi \circ \bullet^{\mathrm{rev}}$ and hence ϕ_r is also surjective. Now, as Δ and Δ_r are both finite sets, it follows that $\Delta^* \# \Delta_r^*$ is a regular language. Hence, as $\mathcal{C}$ is closed under rational transductions, and thus in particular closed under intersection with regular languages, thus $K := \mathrm{InvP}_A^M \cap (\Delta^* \# \Delta_r^*)$ satisfies $K \in \mathcal{C}$.

Now we are almost done. Note that Δ is a finite generating set for $U(M)$. We claim that $\mathrm{WP}_\Delta^{U(M)} = K$. This would complete the proof, since $\mathcal{C}$ is closed under inverse homomorphism, and hence by [209, Proposition 8.2(a)] it would follow that $\mathrm{WP}_X^{U(M)} \in \mathcal{C}$. First, if $w \in K$, then $w \equiv u \# v^{\mathrm{rev}}$ for some $u, v \in \Delta^*$, as $w \in \Delta \# \Delta_r^*$. Furthermore, $u =_M v$, as $u \# v^{\mathrm{rev}} \in \mathrm{InvP}_A^M$. Hence also $u =_{U(M)} v$. Thus, by definition, $w \equiv u \# v^{\mathrm{rev}} \in \mathrm{WP}_\Delta^{U(M)}$. It follows that $K \subseteq \mathrm{WP}_\Delta^{U(M)}$. If instead $w \in \mathrm{WP}_\Delta^{U(M)}$, then $w \equiv u \# v^{\mathrm{rev}}$ for some $u, v \in \Delta^*$ such that $u =_{U(M)} v$. In particular $u \# v^{\mathrm{rev}} \in \Delta^* \# \Delta_r^*$. Also, $u =_{U(M)} v$ if and only if $u =_M v$, and as $u, v \in \Delta^* \subseteq \overline{\Delta}^*$, it follows that $u \# v^{\mathrm{rev}} \in \mathrm{InvP}_A^M$. Hence $w \equiv u \# v^{\mathrm{rev}} \in K$, and so $K = \mathrm{WP}_\Delta^{U(M)}$.

($\implies$) Assume that $\mathrm{WP}_X^{U(M)} \in \mathcal{C}$. For every $\delta \in \Delta$, let $\heartsuit_\delta, \widetilde{\heartsuit}_\delta$ be new symbols. Let $\heartsuit_\Delta = \{\heartsuit_\delta \mid \delta \in \Delta\}$ and $\widetilde{\heartsuit}_\Delta = \{\widetilde{\heartsuit}_\delta \mid \delta \in \Delta\}$. We will require that $\heartsuit_\Delta \cap \widetilde{\heartsuit}_\Delta = \varnothing$, which of course loses us no generality. Let R_δ be the rewriting system on the (finite) alphabet $A \cup \heartsuit_\Delta \cup \widetilde{\heartsuit}_\Delta$ with the rules

$$R_\delta := \bigcup_{\delta \in \Delta} \{(w, \heartsuit_\delta) \mid w \in \mathrm{Rep}_A^M(\delta)\}.$$

Informally, R_δ is the rewriting system which can replace any word $w \in A^*$ with the property that $w =_M \delta$ for some $\delta \in \Delta$ by the single symbol $\heartsuit_\delta$. Note, however, that there may be many such choices of δ and hence many choices of $\heartsuit_\delta$ to rewrite a given w into. Now, note

[60] We have that Δ_r is a suffix code, as if $u, v \in \Delta_r$ are such that $u \equiv wv$ for some $w \in A^+$, then $u^{\mathrm{rev}} \equiv v^{\mathrm{rev}} w^{\mathrm{rev}}$. As $u, v \in \Delta_r$, we have $u^{\mathrm{rev}}, v^{\mathrm{rev}} \in \Delta$, and as w^{rev} is non-trivial this contradicts the fact that Δ is a prefix code. Symmetrically, we prove Δ_r is a prefix code, as Δ is a suffix code.

that R_δ is a $=$-monadic rewriting system. Further, for every $\delta \in \Delta$ we have $\mathrm{Rep}_A^M(\delta) \in \mathcal{C}$ by Theorem 3.2.21. Hence, as $\mathcal{C}$ is closed under finite unions and Δ is finite, it follows that the set of left-hand sides $\bigcup_{\delta\in\Delta} \mathrm{Rep}_A^M(\delta)$ is in $\mathcal{C}$. Thus R_δ is a $\mathcal{C}$-$=$-monadic rewriting system. Let R_δ^{rev} be the rewriting system on the (finite) alphabet $A \cup \heartsuit_\Delta \cup \widetilde{\heartsuit}_\Delta$

$$R_\delta^{\mathrm{rev}} := \bigcup_{\delta\in\Delta} \{(w^{\mathrm{rev}}, \widetilde{\heartsuit}_\delta \mid w \in \mathrm{Rep}_A^M(\delta)\}.$$

Then, as $\mathcal{C}$ is closed under reversal, R_δ^{rev} is – for precisely the same reasons as R_δ – a $\mathcal{C}$-$=$-monadic rewriting system. As $\mathcal{C}$ is closed under finite unions, $R_\delta \cup R_\delta^{\mathrm{rev}}$ is a $\mathcal{C}$-$=$-monadic rewriting system.

Define the homomorphism

$$\varrho \colon (\heartsuit_\Delta \cup \widetilde{\heartsuit}_\Delta \cup \{\#\})^* \to (X \cup \{\#\})^*$$

as the homomorphism induced by $\heartsuit_\delta, \widetilde{\heartsuit}_\delta \mapsto \phi(\delta) \equiv b_\delta$, $\# \mapsto \#$ for all $\heartsuit_\delta \in \heartsuit_\Delta$ and $\widetilde{\heartsuit}_\delta \in \widetilde{\heartsuit}_\Delta$. Let now

$$T = \varrho^{-1}\left(\mathrm{WP}_X^{U(M)}\right) \cap \heartsuit_\Delta^* \# \widetilde{\heartsuit}_\Delta^*.$$

As $\mathrm{WP}_X^{U(M)} \in \mathcal{C}$, and $\mathcal{C}$ is closed under intersection with regular languages and inverse homomorphisms, it follows that $T \in \mathcal{C}$. Now, as $\mathcal{C}$ has the monadic ancestor property, and $R_\delta \cup R_\delta^{\mathrm{rev}}$ is a $\mathcal{C}$-$=$-monadic we have that $\langle T \rangle_{R_\delta \cup R_\delta^{\mathrm{rev}}} \in \mathcal{C}$. We claim that

$$\mathrm{InvP}_A^M = \langle T \rangle_{R_\delta \cup R_\delta^{\mathrm{rev}}} \cap A^* \# A^*.$$

($\subseteq$) Assume that $w \in \mathrm{InvP}_A^M$. Then there exist $w_1, w_2 \in \overline{\Delta}^*$ such that $w \equiv w_1 \# w_2^{\mathrm{rev}}$, and $w_1 =_M w_2$. Write $w_1 \equiv \vartheta_0\vartheta_1 \cdots \vartheta_n$ and $w_2 \equiv \vartheta_0'\vartheta_1' \cdots \vartheta_m'$ where $\vartheta_i, \vartheta_j' \in \overline{\Delta}$. Now by definition of $\overline{\Delta}$, for each ϑ_i there exists some $\delta_i \in \Delta$ such that $\vartheta_i =_M \delta_i$. Analogously, for each ϑ_j' there exists some $\delta_j' \in \Delta$ such that $\vartheta_j' =_M \delta_j'$. Hence

$$\delta_0\delta_1 \cdots \delta_n =_M w_1 =_M w_2 =_M \delta_0'\delta_1' \cdots \delta_m'.$$

This implies that

$$\phi(\delta_0\delta_1 \cdots \delta_n) =_{U(M)} \phi(\delta_0'\delta_1' \cdots \delta_m'),$$

and so it follows that

$$b_{\delta_0} b_{\delta_1} \cdots b_{\delta_n} \# (b_{\delta_0'} b_{\delta_1'} \cdots b_{\delta_m'})^{\mathrm{rev}} \equiv \phi(\delta_0\delta_1 \cdots \delta_n) \# \phi(\delta_0'\delta_1' \cdots \delta_m')^{\mathrm{rev}} \in \mathrm{WP}_X^{U(M)}.$$

Now let $W \equiv \heartsuit_{\delta_0}\heartsuit_{\delta_1} \cdots \heartsuit_{\delta_n} \# (\widetilde{\heartsuit}_{\delta_0'} \widetilde{\heartsuit}_{\delta_1'} \cdots \widetilde{\heartsuit}_{\delta_m'})^{\mathrm{rev}}$. Then

$$W \in \heartsuit_\Delta^* \# \widetilde{\heartsuit}_\Delta^* \subset (\heartsuit_\Delta \cup \widetilde{\heartsuit}_\Delta \cup \{\#\})^*.$$

Thus, applying ϱ to W, we find

$$\begin{aligned} \varrho(W) &\equiv \phi(\delta_0)\phi(\delta_1) \cdots \phi(\delta_n) \# \phi(\delta_m') \cdots \phi(\delta_1')\phi(\delta_0') \\ &\equiv \phi(\delta_0\delta_1 \cdots \delta_n) \# \phi(\delta_0'\delta_1' \cdots \delta_m')^{\mathrm{rev}} \\ &\equiv b_{\delta_0} b_{\delta_1} \cdots b_{\delta_n} \# (b_{\delta_0'} b_{\delta_1'} \cdots b_{\delta_m'})^{\mathrm{rev}}. \end{aligned}$$

Hence as $b_{\delta_0} b_{\delta_1} \cdots b_{\delta_n} \# (b_{\delta_0'} b_{\delta_1'} \cdots b_{\delta_m'})^{\mathrm{rev}} \in \mathrm{WP}_X^{U(M)}$, it follows that

$$W \in \varrho^{-1}(\mathrm{WP}_X^{U(M)}) \cap \heartsuit_\Delta^* \# \widetilde{\heartsuit}_\Delta^* = T.$$

Now we are nearly finished. We claim that we can rewrite $w \equiv w_1 \# w_2^{\mathrm{rev}}$ into W, i.e. that

$$w \equiv \vartheta_0\vartheta_1 \cdots \vartheta_n \# (\vartheta_0'\vartheta_1' \cdots \vartheta_m')^{\mathrm{rev}} \xrightarrow{*}_{R_\delta \cup R_\delta^{\mathrm{rev}}} W.$$

From this follows $w \in \langle T \rangle_{R_\delta \cup R_\delta^{\mathrm{rev}}}$. Note that as $\vartheta_i =_M \delta_i$, we have that $(\vartheta_i, \heartsuit_{\delta_i}) \in R_\delta$, for all $0 \leq i \leq n$. Analogously, as $\vartheta'_j =_M \delta'_j$, we have that $((\vartheta'_j)^{\mathrm{rev}}, \widetilde{\heartsuit}_{\delta'_j}) \in R_\delta^{\mathrm{rev}}$ for all $0 \leq j \leq m$. Hence

$$\begin{aligned}
w \equiv w_1 \# w_2^{\mathrm{rev}} &\equiv \vartheta_0\vartheta_1 \cdots \vartheta_n \# (\vartheta'_0\vartheta'_1 \cdots \vartheta'_m)^{\mathrm{rev}} \\
&\equiv \vartheta_0\vartheta_1 \cdots \vartheta_n \# (\vartheta'_m)^{\mathrm{rev}}(\vartheta'_{m-1})^{\mathrm{rev}} \cdots (\vartheta'_1)^{\mathrm{rev}} \\
&\to_{R_\delta \cup R_\delta^{\mathrm{rev}}} \heartsuit_{\delta_0}\vartheta_1 \cdots \vartheta_n \# (\vartheta'_m)^{\mathrm{rev}}(\vartheta'_{m-1})^{\mathrm{rev}} \cdots (\vartheta'_1)^{\mathrm{rev}} \\
&\to_{R_\delta \cup R_\delta^{\mathrm{rev}}} \heartsuit_{\delta_0}\heartsuit_{\delta_1} \cdots \vartheta_n \# (\vartheta'_m)^{\mathrm{rev}}(\vartheta'_{m-1})^{\mathrm{rev}} \cdots (\vartheta'_1)^{\mathrm{rev}} \\
&\to_{R_\delta \cup R_\delta^{\mathrm{rev}}} \cdots \\
&\to_{R_\delta \cup R_\delta^{\mathrm{rev}}} \heartsuit_{\delta_0}\heartsuit_{\delta_1} \cdots \heartsuit_{\delta_n} \# (\vartheta'_m)^{\mathrm{rev}}(\vartheta'_{m-1})^{\mathrm{rev}} \cdots (\vartheta'_1)^{\mathrm{rev}} \\
&\to_{R_\delta \cup R_\delta^{\mathrm{rev}}} \heartsuit_{\delta_0}\heartsuit_{\delta_1} \cdots \heartsuit_{\delta_n} \# \widetilde{\heartsuit}_{\delta'_m}(\vartheta'_{m-1})^{\mathrm{rev}} \cdots (\vartheta'_1)^{\mathrm{rev}} \\
&\to_{R_\delta \cup R_\delta^{\mathrm{rev}}} \heartsuit_{\delta_m}\heartsuit_{\delta_1} \cdots \heartsuit_{\delta_n} \# \widetilde{\heartsuit}_{\delta'_m}\widetilde{\heartsuit}_{\delta'_{m-1}} \cdots (\vartheta'_1)^{\mathrm{rev}} \\
&\to_{R_\delta \cup R_\delta^{\mathrm{rev}}} \cdots \\
&\to_{R_\delta \cup R_\delta^{\mathrm{rev}}} \heartsuit_{\delta_m}\heartsuit_{\delta_1} \cdots \heartsuit_{\delta_n} \# \widetilde{\heartsuit}_{\delta'_m}\widetilde{\heartsuit}_{\delta'_{m-1}} \cdots \widetilde{\heartsuit}_{\delta'_0} \\
&\to_{R_\delta \cup R_\delta^{\mathrm{rev}}} \heartsuit_{\delta_m}\heartsuit_{\delta_1} \cdots \heartsuit_{\delta_n} \# (\widetilde{\heartsuit}_{\delta'_0}\widetilde{\heartsuit}_{\delta'_1} \cdots \widetilde{\heartsuit}_{\delta'_m})^{\mathrm{rev}} \equiv W.
\end{aligned}$$

Thus we have proved that $w \equiv w_1 \# w_2^{\mathrm{rev}} \in \langle T \rangle_{R_\delta \cup R_\delta^{\mathrm{rev}}}$. As $w \in \mathrm{InvP}_A^M$, we also have *a fortiori* that $w \in A^*\#A^*$. Hence, as w was arbitrary, it follows that $\mathrm{InvP}_A^M \subseteq \langle T \rangle_{R_\delta \cup R_\delta^{\mathrm{rev}}} \cap A^*\#A^*$.

$(\supseteq)$ Let $w \in \langle T \rangle_{R_\delta \cup R_\delta^{\mathrm{rev}}} \cap A^*\#A^*$. Then in particular, $w \in \langle T \rangle_{R_\delta \cup R_\delta^{\mathrm{rev}}}$, so there exists $W \in T$ such that $w \xrightarrow{*}_{R_\delta \cup R_\delta^{\mathrm{rev}}} W$. As $W \in T$, on the one hand we have that $W \in \varrho^{-1}(\mathrm{WP}_X^{U(M)})$. Thus there must exist $b_i, b'_j \in X$ with $(0 \leq i \leq n,\ 0 \leq j \leq m)$ such that

$$W \equiv \varrho(b_0 b_1 \cdots b_n \# b'_m b'_{m-1} \cdots b'_0) \equiv \varrho(b_0 b_1 \cdots b_n) \# \varrho(b'_m b'_{m-1} \cdots b'_0),$$

with $b_0 b_1 \cdots b_n =_{U(M)} b'_0 b'_1 \cdots b'_m$. What is the pre-image under ϱ of the word $b_0 b_1 \cdots b_n$? As ϱ is length-preserving, it is precisely those words $h_0 h_1 \cdots h_n$ with $h_i \in \heartsuit_\Delta \cup \widetilde{\heartsuit}_\Delta$ such that $\varrho(h_i) \equiv b_i$. But

$$\varrho(h_i) \equiv b_i \qquad \Longleftrightarrow \qquad h_i \in \{\heartsuit_\delta, \widetilde{\heartsuit}_\delta \mid \phi(\delta) = b_i\}.$$

The analogous statement is, of course, also true for $b'_0 b'_1 \cdots b'_m$. Hence, it follows that there exist $h_i, h'_m \in \heartsuit_\Delta \cup \widetilde{\heartsuit}_\Delta$ such that

$$W \equiv h_0 h_1 \cdots h_n \# (h'_m h'_{m-1} \cdots h_0).$$

But as $W \in T$, we also have that $W \in \heartsuit_\Delta^* \# \widetilde{\heartsuit}_\Delta^*$. Hence $h_i \in \heartsuit_\Delta$, and $h'_j \in \widetilde{\heartsuit}_\Delta$, for all $0 \leq i \leq n$ and $0 \leq j \leq m$. For $0 \leq i \leq n$, let $\delta_i \in \Delta$ be the piece such that $h_i \equiv \heartsuit_{\delta_i}$, and analogously for $0 \leq j \leq m$ let $\delta'_j \in \Delta$ be the piece such that $h_j \equiv \widetilde{\heartsuit}_{\delta_j}$. In particular $\phi(\delta_i) \equiv b_i$, and $\phi(\delta'_j) \equiv b'_j$. Then we can write

$$W \equiv \heartsuit_{\delta_0}\heartsuit_{\delta_1} \cdots \heartsuit_{\delta_n} \# \widetilde{\heartsuit}_{\delta'_m}\heartsuit_{\delta'_{m-1}} \cdots \widetilde{\heartsuit}_{\delta'_0},$$

Note that as $b_0 b_1 \cdots b_n =_{U(M)} b'_0 b'_1 \cdots b'_m$, we also have that $\delta_0\delta_1 \cdots \delta_n =_M \delta'_0\delta'_1 \cdots \delta'_m$. Before proceeding with the proof, we first require the following lemma.

Lemma 3.3.3. *Any ancestor of W under $R_\delta \cup R_\delta^{\mathrm{rev}}$ has the form*

$$\alpha_0\alpha_1 \cdots \alpha_n \# \alpha'_m \alpha'_{m-1} \cdots \alpha'_0$$

where, for $0 \leq i \leq n$ and $0 \leq j \leq m$, we have that α_i is either (1) $\heartsuit_{\delta_i}$, or else (2) w_{δ_i}, where

$w_{\delta_i} \in \mathrm{Rep}_A^M(\delta_i)$*; and similarly* α'_j *is either (1)* $\widetilde{\heartsuit}_{\delta'_j}$*, or else (2)* $w_{\delta'_j}^{rev}$*, where* $w_{\delta'_j} \in \mathrm{Rep}_A^M(\delta'_j)$*.*

Proof. The proof is, of course, by induction on the number of rewriting steps. Let $u \in A^*$ be such that $u \xrightarrow{*}_{R_\delta \cup R_\delta^{\text{rev}}} W$ in $\lambda \geq 0$ steps. If $\lambda = 0$, i.e. $u \equiv W$, then the claim is obviously true, as W is already of the desired form. Suppose that for some $\lambda > 0$ the claim is true for all words which can be rewritten in $< \lambda$ steps. As $u \rightarrow^{\lambda}_{R_\delta \cup R_\delta^{\text{rev}}} W$, there exists some $u' \in A^*$ such that

$$u \rightarrow_{R_\delta \cup R_\delta^{\text{rev}}} u' \rightarrow^{\lambda-1}_{R_\delta \cup R_\delta^{\text{rev}}} W.$$

By the inductive hypothesis, we can write

$$u' \equiv \alpha_0 \alpha_1 \cdots \alpha_n \# \alpha'_m \alpha'_{m-1} \cdots \alpha'_0$$

where the α_i, α'_j are as specified above. The only right-hand sides of rules from $R_\delta \cup R_\delta^{\text{rev}}$ in u' are now the α_i (resp. α'_j) which are of the form $\heartsuit_{\delta_i}$ (resp. $\widetilde{\heartsuit}_{\delta'_j}$). Then the rewriting $u \rightarrow_{R_\delta \cup R_\delta^{\text{rev}}} u'$ was either via a rule which replaced a word β by some specific $\alpha_k \equiv \heartsuit_{\delta_k}$ for some $0 \leq k \leq n$; or else via a rule which replaced a word β by some specific $\alpha'_k \equiv \widetilde{\heartsuit}_{\delta'_k}$ for some $0 \leq k \leq n$. In the first case, $\beta \in \mathrm{Rep}_A^M(\delta_k)$, as the only rules of $R_\delta \cup R_\delta^{\text{rev}}$ are of this form; thus

$$u \equiv \alpha_0 \alpha_1 \cdots \alpha_{k-1} \beta \alpha_{k+1} \cdots \alpha_n \# \alpha'_m \alpha'_{m-1} \cdots \alpha'_0$$

which is of the specified form, and we are done by induction. In the second case, $\beta^{\text{rev}} \in \mathrm{Rep}_A^M(\delta'_k)$, and it follows that

$$u \equiv \alpha_0 \alpha_1 \cdots \alpha_n \# \alpha'_m \alpha'_{m-1} \cdots \alpha'_{k+1} \beta^{\text{rev}} \alpha'_{k-1} \cdots \alpha'_0$$

which is, of course, also of the specified form. This completes the proof of the lemma. □

We are now nearly done. We know that $w \xrightarrow{*}_{R_\delta \cup R_\delta^{\text{rev}}} W$, and that $w \in A^* \# A^*$. We conclude by Lemma 3.3.3 that we can write

$$\begin{aligned} u &\equiv w_{\delta_0} w_{\delta_1} \cdots w_{\delta_n} \# w_{\delta'_m}^{\text{rev}} w_{\delta'_{m-1}}^{\text{rev}} \cdots w_{\delta'_0}^{\text{rev}} \\ &\equiv w_{\delta_0} w_{\delta_1} \cdots w_{\delta_n} \# (w_{\delta'_0} w_{\delta'_1} \cdots w_{\delta'_m})^{\text{rev}} \end{aligned}$$

where $w_{\delta_i} \in \mathrm{Rep}_A^M(\delta_i)$ and $w_{\delta_j} \in \mathrm{Rep}_A^M(\delta'_j)$ for all $0 \leq i \leq n$ and $0 \leq j \leq m$. Thus, with the same indexing, $w_{\delta_i} =_M \delta_i$ and $w_{\delta_j} =_M \delta'_j$. Hence

$$\begin{aligned} w_{\delta_0} w_{\delta_1} \cdots w_{\delta_n} &=_M \delta_0 \delta_1 \cdots \delta_n \\ &=_M \delta'_0 \delta'_1 \cdots \delta'_m \\ &=_M w_{\delta'_0} w_{\delta'_1} \cdots w_{\delta'_m}. \end{aligned}$$

Hence, as all w_{δ_i} and $w_{\delta'_j}$ are invertible, being equal to $\delta_i \in \Delta$ resp. $\delta'_j \in \Delta$, it follows – at last! – that $w \in \mathrm{InvP}_A^M$. As $w \in \langle T \rangle_{R_\delta \cup R_\delta^{\text{rev}}} \cap A^* \# A^*$ was arbitrary, we conclude that $\langle T \rangle_{R_\delta \cup R_\delta^{\text{rev}}} \cap A^* \# A^* \subseteq \mathrm{InvP}_A^M$, and hence that

$$\mathrm{InvP}_A^M = \langle T \rangle_{R_\delta \cup R_\delta^{\text{rev}}} \cap A^* \# A^* \in \mathcal{C},$$

as required. This completes the proof of Theorem 3.3.2. □

Remark 3.3.1. The proof of Theorem 3.3.2 relies on the presence of the symbol $\#$ "of the second kind" (see §1.2.3) within WP_A^M to separate which pieces are written backwards, and which are not.

Example 3.3.4. The following classes of languages are all closed under finite unions, intersection with regular languages, inverse homomorphisms, and reversal: $\mathcal{C}_{\text{reg}}, \mathcal{C}_{\text{cf}}$, the class of indexed languages, the class of context-sensitive languages, and $\mathcal{C}_{\text{en}}$. On the other hand, the class $\mathcal{C}_{\text{dcf}}$ of deterministic context-free languages is not closed under reversal. △

Example 3.3.5. Consider the bicyclic monoid $M = \text{Mon}\langle b, c \mid bc = 1\rangle$. Then $\Delta = \{bc\}$, and $U(M) \cong 1$. Hence $\text{Rep}^M_{\{b,c\}}(bc) = [bc]_M = [1]_M$ is the set of all words equal in M to the identity element. This is the Dyck language on b and c, which is well-known to be context-free. As $U(M)$ is trivial, it has context-free word problem. As we shall see in the following section, $\mathcal{C}_{\text{cf}}$ has the monadic ancestor property. Hence it follows from Theorem 3.3.2 that $\text{InvP}^M_{\{b,c\}} \in \mathcal{C}_{\text{cf}}$. △

Example 3.3.6. Consider $M = \text{Mon}\langle a, b, c \mid abc = 1, b^2 = 1\rangle$. Then $\Delta = \{abc, b\}$; let $X = \{x_1, x_2\}$ and $\phi(abc) = x_1, \phi(b) = x_2$. Let $G := U(M) = \text{Mon}\langle x_1, x_2 \mid x_1 = 1, x_2^2 = 1\rangle$. Then

$$R_\delta = \{w \to \heartsuit_{abc} \mid w \in A^*, w =_M abc\} \cup \{w \to \heartsuit_b \mid w \in A^*, w =_M b\}.$$

In particular, the rule $ab^{17}c \to \heartsuit_{abc}$ is in R_δ. The system R_δ^{rev} is similarly defined; we have that $b^2cb^3a \to \widetilde{\heartsuit}_{abc}$ is a rule in R_δ^{rev}, as $ab^3cb^2 =_M abc$.

Now as $G \cong C_2$ is a finite group, it has context-free word problem. Furthermore, as this presentation satisfies the small subpiece condition, it follows by Theorem 3.2.21 that $\text{Rep}^M_A(abc)$ is context-free; and that $\text{Rep}^M_A(b)$ is context-free. Hence R_δ is a context-free rewriting system, as is R_δ^{rev}. It just remains – essentially – to show that any invertible word can be obtained as an ancestor by using the context-free system $R_\delta \cup R_\delta^{\text{rev}}$. In general, one would need to prove this as in the above proof of Theorem 3.3.2, but in this case we can see this in an easier manner. For if w is invertible, then as the monadic finite rewriting system with the two rules $(abc \to 1), (b^2 \to 1)$, is easily seen to be complete[61], and as the only two irreducible invertible words are 1 and b, it follows that w rewrites under this system either to 1 or to b. As w was arbitrary, it follows that the set of invertible words is the set of ancestors under a finite monadic rewriting system of a finite set; and hence is context-free. △

Thus, we conclude that if we understand the structure of $U(M)$, then we understand the structure of InvP^M_A. In one sense, this means that we can state that we completely understand the invertible elements of special monoids. However, the dependency on the presentation having certain properties for Theorem 3.3.2 to be true – which, as mentioned, is not an issue, given that every special monoid M admits a presentation of the desired form by Proposition 3.2.10 together with Lemma 3.2.20 – is nevertheless rather unsightly. We therefore pose the following natural question.

Question 3.3.7. *Let $\mathcal{C}$ be a super-AFL. Let M be a special monoid, with pieces Δ and associated set X. Is it always the case that*

$$\text{WP}^{U(M)}_X \in \mathcal{C} \quad \iff \quad \text{InvP}^M_A \in \mathcal{C}?$$

[61] The only overlap of left-hand sides of rules is $bb \to 1$ with itself when applied to the word bbb; but either way, this word rewrites to b. As the system is length-reducing, it is terminating; as it is locally confluent and terminating, it is complete.

The author conjectures a positive answer to this question. If $\mathcal{C}$ is additionally closed under intersection (which, for example, the class $\mathcal{C}_{\text{cf}}$ of context-free languages is not, despite being a super-AFL) then the answer may be somewhat simplified by the fact that $\text{InvP}_A^M = \text{WP}_A^M \cap (\overline{\Delta}^* \# \overline{\Delta}^*)$ and the fact that $\text{WP}_A^M \in \mathcal{C}$ if and only if $U(M)$ has word problem in $\mathcal{C}$

We remark that the conjecture can be shown to be true (i.e. the question has an affirmative answer) for many classes of special monoids (and conjecturally for all special monoids, see Theorem 5.5.4) when $\mathcal{C} = \mathcal{C}_{\text{cf}}$, by using the results from Chapter 5. We do not expand much on this, as this is far beyond our current scope, leaving the (very) interested reader to pursue this on their own. We will assume the notation and terminology of pushdown automata and related material directly from [363]. The rough outline is as follows. If $U(M)$ is context-free, then in many classes (and conjecturally in all, see Claim $(*)$ in Theorem 5.5.4) there exists a finite deterministic pushdown automaton $\mathcal{P}$ such that the right Cayley graph $\Gamma_M(M, A)$ is the graph $\Gamma(\mathcal{P})$ of $\mathcal{P}$. If $w \in A^*$ represents the element $m \in M$, and (q_m, z_m) is the total state corresponding to m, then we may easily modify $\mathcal{P}$ to accept the input word $u \in A^*$ if and only if the modified automaton is in the state q_0 and the stack contains the word z_m. Thus u is accepted if and only if $w =_M u$. This is a direct adaptation of the easy direction of [363, Theorem 2.9], i.e. that if M is a group and has a context-free Cayley graph, then the set[62] $\text{Rep}_A^M(1)$ is context-free. Thus one can show by geometric means that if $U(M)$ is a context-free group, then $\text{Rep}_A^M(w)$ is a context-free language for all $w \in A^*$. As this is irrespective of the presentation by which M is given, one bypasses the need to use Makanin's techniques for manipulating the presentation; one can thus with little difficulty, going through the statements and proofs, conclude that InvP_A^M and WP_A^M are both context-free languages if $U(M)$ is a context-free group, regardless of the presentation by which M is given, as long as the hypothesis of benignity in Theorem 5.5.4 holds. Thus the above question has an affirmative answer for the class $\mathcal{C} = \mathcal{C}_{\text{cf}}$ if Claim $(*)$ of Chapter 5 holds. One can say more, but we are not – at present – particularly interested in doing so, and leave this to the interested reader.

In any case, we have now understood the step from $\text{WP}_X^{U(M)}$ to InvP_A^M. As we have also understood the step from InvP_A^M to WP_A^M we are now essentially done, and all that remains is to write the theorem out.

[62] Recall that if M is a group, then the set $\text{Rep}_A^M(1)$ is commonly called "the word problem for the group M" in the literature, which is quite confusing. To be clear: it is the set of elements representing the identity in M.

3.4 The word problem

In this section, we will again utilise the extraordinary properties of monadic rewriting systems to show that the word problem and the invertible word problem for a given special monoid are intricately connected. In particular, we will completely describe when two words are equal in a special monoid M, modulo the abstract properties of $U(M)$. This is entirely in line with Sushkevič's principle of reducing semigroup-theoretic problems to group-theoretic problems. Combining all of the above results, we arrive at the following full description of WP_A^M modulo the word problem for $U(M)$. We emphasise that although some results used (e.g. Theorem 3.3.2) depend on the presentation for M, this final theorem does not – whether or not WP_A^M is in a class of languages closed under inverse homomorphism does not depend on the presentation.

Theorem 3.4.1. *Let M be a finitely presented special monoid. Let $\mathcal{C}$ be a super-AFL closed under reversal. Then M has word problem in $\mathcal{C}$ if and only if $U(M)$ has word problem in $\mathcal{C}$.*

Proof. ($\implies$) As M is finitely presented we have that $U(M)$ is finitely generated (see §1.3). If M has word problem in $\mathcal{C}$, then by [209, Proposition 8(a)] any finitely generated submonoid of M has word problem in $\mathcal{C}$. Hence $U(M)$ has word problem in $\mathcal{C}$.

($\impliedby$) By Theorem 3.2.21, M admits a special monoid presentation with generating set A' which satisfies the conclusions of Theorem 3.2.21. Let X' be the finite generating set for $U(M)$ obtained from this presentation. Then as $U(M)$ has word problem in $\mathcal{C}$ (with respect to X), it follows that since $\mathcal{C}$ is closed under inverse homomorphisms we also have $\mathrm{WP}_{X'}^{U(M)} \in \mathcal{C}$ by [209, Proposition 8(b)]. Hence, by Theorem 3.3.2 we have that $\mathrm{InvP}_{A'}^M \in \mathcal{C}$. By Lemma 3.3.1, we thus have that $\mathrm{WP}_{A'}^M \in \mathcal{C}$. By another application of [209, Proposition 8(b)] we also have $\mathrm{WP}_A^M \in \mathcal{C}$. □

Remark 3.4.1. If $U(M)$ has regular word problem, then M will not generally have regular word problem. This is due to Proposition 3, i.e. a special monoid has regular word problem if and only if it is a finite group. Of course, the failure comes from the fact that the class $\mathcal{C}_{\mathrm{reg}}$ of regular languages does not have the monadic ancestor property, and is hence not a super-AFL; indeed, any super-AFL contains $\mathcal{C}_{\mathrm{cf}}$. Hence the assumption on $\mathcal{C}$ to have the monadic ancestor property cannot be significantly weakened in any obvious way.

Thus we have proved the main result of this chapter. We will now describe some corollaries. For our first corollary, which informally says that the "group-theoretic word problem" and "Duncan-Gilman word problem" have the same language-theoretic properties for special monoids, we will first require the following simple lemma. The lemma will also have a convenient corollary regarding the rational subset membership problem.

Lemma 3.4.2. *Let M be a monoid generated by some finite set A, with associated homomorphism $\pi\colon A^* \to M$. Let $\mathcal{C}$ be a class of languages such that $\mathcal{C}$ is closed under rational transductions. Suppose M has word problem in $\mathcal{C}$. Then for every rational subset $R \subseteq M$, we have $\pi^{-1}(R) \in \mathcal{C}$.*

Proof. Let $R \subseteq M$ be a rational subset, given to us as a regular language $K \subseteq A^*$ with $\pi(K) = R$, and let $w \in A^*$. As K is regular, so is K^{rev} and hence also $A^*\#K^{\text{rev}}$, being a concatenation of three regular languages. As K is regular, so is K^{rev}. As $\mathrm{WP}_A^M \in \mathcal{C}$, and $\mathcal{C}$ is closed under intersection with regular languages, we hence have that

$$L = \mathrm{WP}_A^M \cap (A^*\#K^{\text{rev}}) \in \mathcal{C}.$$

Let A' be an alphabet disjoint from $A \cup \{\#\}$, and in bijective correspondence with a via a map $a \mapsto a'$ for all $a \in A$. Let $\varrho_1 \colon (A \cup A' \cup \{\#\})^* \to A^*$ be the homomorphism defined by $a \mapsto a, a' \mapsto a$, and $\# \mapsto \#$, for all $a \in A$ and $a' \in A'$. Define the homomorphism $\varrho_2 \colon A^*\#(A')^* \to A^*$ by $a \mapsto a$, $\# \mapsto \varepsilon$, and $a' \mapsto \varepsilon$ for all $a \in A$ and $a' \in A'$.

We claim that

$$\pi^{-1}(R) = \varrho_2(\varrho_1^{-1}(L) \cap A^*\#(A')^*).$$

As $\mathcal{C}$ is closed under rational transductions, from this claim it would follow that $\pi^{-1}(R) \in \mathcal{C}$, as desired.

($\subseteq$) Let $w \in \pi^{-1}(R)$. Then $\pi(w) \in R$. As $\pi(K) = R$, there exists some $u \in K$ such that $w =_M u$. Thus $w\#u^{\text{rev}} \in \mathrm{WP}_A^M$. As $w \in A^*$, we have that $w\#u^{\text{rev}} \in A^*\#K^{\text{rev}}$, and hence

$$w\#u^{\text{rev}} \in \mathrm{WP}_A^M \cap (A^*\#K^{\text{rev}}) = L.$$

Write $u \equiv a_1a_2\cdots a_p$ for some (not necessarily distinct) generators $a_i \in A$ $(1 \leq i \leq p)$. Let $u' \equiv a_1'a_2'\cdots a_p'$, where now $a_i' \in A'$ $(1 \leq i \leq p)$. Now we have

$$\begin{aligned}\varrho_1(w\#(u')^{\text{rev}}) \equiv \varrho_1(w\#(a_1'a_2'\cdots a_p')^{\text{rev}}) &= w\#\varrho_1'(a_1'a_2'\cdots a_p')^{\text{rev}})\\ &= w\#\varrho_1'(a_p')\varrho_1'(a_{p-1}')\cdots\varrho_1'(a_1')\\ &= w\#a_pa_{p-1}\cdots a_1\\ &\equiv w\#u^{\text{rev}} \in L.\end{aligned}$$

Hence $w\#(u')^{\text{rev}} \in \varrho_1^{-1}(L)$. Clearly $w\#(u')^{\text{rev}} \in A^*\#(A')^*$. Hence

$$w\#(u')^{\text{rev}} \in \varrho_1^{-1}(L) \cap (A^*\#(A')^*).$$

We claim that $\varrho_2(w\#(u')^{\text{rev}}) = w$. Indeed,

$$\begin{aligned}\varrho_2(w\#(a_1'a_2'\cdots a_p')^{\text{rev}}) &= w\varrho_2(\#)\varrho_2(a_p')\varrho_2(a_{p-1}')\cdots\varrho_2(a_2')\varrho_2(a_1')\\ &= w\cdot\varepsilon\cdot\varepsilon\cdot\cdots\cdot\varepsilon \equiv w.\end{aligned}$$

Thus $w \in \varrho_2(\varrho_1^{-1}(L) \cap (A^*\#(A')^*))$. As w was arbitrary, it follows that

$$\pi^{-1}(R) \subseteq \varrho_2(\varrho_1^{-1}(L) \cap (A^*\#(A')^*))$$

as desired.

($\supseteq$) Suppose that $w \in \varrho_2(\varrho_1^{-1}(L) \cap (A^*\#(A')^*))$. By definition, there hence exists some $w' \in \varrho_1^{-1}(L) \cap (A^*\#(A')^*)$ such that $w = \varrho_2(w')$. Thus there exists $w'' \in L$ such that $w'' = \varrho_1(w')$. As $w'' \in L$, it follows that $w'' \in \mathrm{WP}_A^M$ and $w'' \in A^*\#K^{\text{rev}}$. From $w'' \in \mathrm{WP}_A^M$, we find that $w'' \equiv u\#v^{\text{rev}}$ for some $u, v \in A^*$ such that $u =_M v$. From $w'' \in A^*\#K^{\text{rev}}$ we additionally have that $v^{\text{rev}} \in K^{\text{rev}}$, i.e. $v \in K$.

Now from the above structure of w'', and the fact that $w'' = \varrho_1(w')$, we can deduce quite a bit about the structure of w'. Suppose $u \equiv a_1a_2\cdots a_p$ and $v \equiv b_1b_2\cdots b_q$ for some (not necessarily distinct) $a_i, b_j \in A$, with $1 \leq i \leq p$ and $1 \leq j \leq q$. As $\varrho_1^{-1}(\#) = \{\#\}$, i.e. the only symbol which maps to $\#$ is $\#$, it necessarily follows that

$$w' \in \varrho_1^{-1}(u)\#\varrho_1^{-1}(v^{\text{rev}})$$

i.e. that

$$w' \in \varrho_1^{-1}(a_1a_2\cdots a_p)\#\varrho_1^{-1}(b_qb_{q-1}\cdots b_1).$$

But as both $\varrho_1^{-1}(a_1a_2\cdots a_p)$ and $\varrho_1^{-1}(b_qb_{q-1}$ are subsets of $(A \cup A')^*$, and furthermore that $w' \in A^*\#(A')^*$, we must have that

$$\varrho_1^{-1}(a_1a_2\cdots a_p) \subseteq A^*,$$
$$\varrho_1^{-1}(b_qb_{q-1}\cdots b_1) \subseteq (A')^*$$

and as ϱ_1 restricts to a bijection on A^* (resp. $(A')^*$) it follows that

$$\varrho_1^{-1}(a_1a_2\cdots a_p) = \{a_1a_2\cdots a_p\},$$
$$\varrho_1^{-1}(b_qb_{q-1}\cdots b_1) = \{b'_qb'_{q-1}\cdots b'_1\}.$$

In particular

$$w' \in \{a_1a_2\cdots a_p\}\#\{b'_qb'_{q-1}\cdots b'_1\},$$

i.e. $w' \equiv a_1a_2\cdots a_p\#b'_qb'_{q-1}\cdots b'_1$. If we denote $v' \equiv b'_1b'_2\cdots b'_q$, then $w' \equiv u\#(v')^{\text{rev}}$.

Now $w = \varrho_2(w')$. Thus

$$\begin{aligned} w &= \varrho_2(a_1a_2\cdots a_p\#b'_qb'_{q-1}\cdots b'_1) \\ &= \varrho_2(a_1)\varrho_2(a_2)\cdots\varrho_2(a_p)\varrho_2(\#)\varrho_2(b'_q)\varrho_2(b'_{q-1})\cdots\varrho_2(b'_1) \\ &= a_1a_2\cdots a_p\cdot\varepsilon\cdot\varepsilon\cdot\cdot\varepsilon \equiv u. \end{aligned}$$

Hence our arbitrary word w is, in fact, a word $u \in A^*$ such that there exists $v \in K$ with $u =_M v$. That is, $\pi(u) \in \pi(K) = R$. Hence $w \equiv u \in \pi^{-1}(R)$, which is what was to be shown, and which completes the proof of the entire lemma. □

Now, recall that for groups, to talk about e.g. a "context-free" or "regular" group, it is sufficient to talk about the class of words representing a single element of the group – or indeed just the identity element. That is, to understand equality of words in a group G, it is sufficient to consider the set

$$\{w \mid w =_G 1\}$$

of words equal to 1 (where for clarity any reference to a specific generating set is deliberately suppressed). Recall that, in the literature, this set is often referred to as "the word problem" of the group G. The main reason for this is that two words u, v satisfy $u =_G v$ if and only if $uv^{-1} =_G 1$, so the set is a recursive language if and only if the word problem (as a decision problem) is decidable for G.

On the other hand, for general monoids, the set of words equal to the identity is hopelessly terrible at describing equality of words. For example, in $\text{Mon}\langle a, b \mid abaab = a\rangle$, only the empty word is equal to the identity (!), and yet the problem of whether this particular monoid has decidable word problem was, for some time, an open problem [221]. However, the following corollary shows that for special monoids, we do not fall into such dangers: one need only consider the "group-theoretic word problem", i.e. the set of words equal to 1. This holds true for the same classes of languages as the main theorem (Theorem 3.4.1), and can be regarded as a corollary of the same.

Corollary 3.4.3. *Let M be a finitely presented special monoid with finite generating set A, and let $\mathcal{C}$ be a super-AFL closed under reversal. Then the following are equivalent:*

(1) M has word problem in $\mathcal{C}$.
(2) For every word $w \in A^*$, $\mathrm{Rep}_A^M(w) \in \mathcal{C}$.
(3) $\mathrm{Rep}_A^M(1) := \{w \mid w =_M 1, w \in A^*\} \in \mathcal{C}$.

Proof. (1) $\implies$ (2) Suppose $\mathrm{WP}_A^M \in \mathcal{C}$. Then for all $w \in A^*$ the set $\mathrm{Rep}_A^M(w) = \pi^{-1}(\{\pi(w)\})$ is the pre-image of a rational (even finite) subset of M. Hence, by Lemma 3.4.2, it follows that $\mathrm{Rep}_A^M(w) \in \mathcal{C}$.

(2) $\implies$ (3) Obvious.

(3) $\implies$ (1) Suppose $\mathrm{Rep}_A^M(1) \in \mathcal{C}$. Then, as $\mathcal{C}$ is closed under rational transductions, it is in particular closed under intersection with regular languages. Hence $\mathrm{Rep}_A^M(1) \cap \Delta^* \in \mathcal{C}$, where Δ is the (finite) set of pieces of the given presentation. Let X, ϕ be as usual. As $\mathcal{C}$ is closed under homomorphism (being closed under rational transductions), the set

$$\begin{aligned}\phi(\mathrm{Rep}_A^M(1) \cap \Delta^*) &= \phi\left(\{w \in \Delta^* \mid w =_M 1\}\right) \\ &= \quad \{w \in X^* \mid w =_{U(M)} 1\}\end{aligned}$$

is also in $\mathcal{C}$. But $U(M)$ is a group; thus it has word problem in $\mathcal{C}$ if and only if the set of words in X^* equal to the identity element is in $\mathcal{C}$. This is the statement of [144, Theorem 5.3]. Hence $U(M)$ has word problem in $\mathcal{C}$. By Theorem 3.4.1 thus M also has word problem in $\mathcal{C}$. □

Corollary 3.4.3 tells us something rather beautiful: special monoids and groups behave very similarly from a formal language theoretic point of view. In particular, the two definitions of "the word problem" for general monoids – which are known to coincide in the case of groups – are also equivalent for special monoids. Thus, while we may speak of monoids as being "generalised groups", the above shows that the (a?) first step in generalising groups is perhaps most naturally taken by passing through the special monoids. We shall see this theme repeated in Chapter 5, when some geometric aspects of special monoids will be shown to be similar to those of groups.

We will now turn from general classes $\mathcal{C}$ to specific classes. As mentioned in §1.2, the class $\mathcal{C}_{\text{ind}}$ of indexed languages is a super-AFL, and is closed under reversal (see e.g. [15]). Thus:

Theorem 3.4.4. *Let M be a finitely presented special monoid. Then M has indexed word problem if and only if the group of units $U(M)$ has indexed word problem.*

We now elaborate our discussion from this class to the class $\mathcal{C}_{\text{cf}}$ of context-free language. This allows us, in particular, to write down a generalisation of the Muller-Schupp theorem to special monoids, and to investigate some concrete decision problems.

3.5 Context-free special monoids

We shall begin by recalling that $\mathcal{C}_{\mathrm{cf}}$ has the monadic ancestor property (see e.g. §1.2 or Chapter 2). Furthermore, as mentioned in §1.2, $\mathcal{C}_{\mathrm{cf}}$ is closed under finite unions, rational transductions, and reversal; and obviously every finite language is context-free. Thus, as we have recalled before, $\mathcal{C}_{\mathrm{cf}}$ is a super-AFL closed under reversal, and so we can apply Theorem 3.4.1 with $\mathcal{C} = \mathcal{C}_{\mathrm{cf}}$. The following theorem for groups has already been mentioned in §1.2; we restate it for completeness.

Theorem (Muller & Schupp, 1983)**.** *Let G be a finitely generated group. Then G has context-free word problem if and only if G is virtually free.*

As a finitely presented special monoid has finitely generated group of units, we can combine this with Theorem 3.4.1 in order to obtain a full algebraic characterisation of the special monoids with context-free word problem.

Theorem 3.5.1. *Let M be a finitely presented special monoid. Then M has context-free word problem if and only if the group of units $U(M)$ of M is virtually free.*

Every group is a special monoid. Thus the Muller-Schupp theorem is a special case of Theorem 3.5.1; indeed, if M is a group, then of course $U(M) = M$ and we obtain the statement of the Muller-Schupp theorem, as any context-free group is finitely presented. This gives a complete and algebraic characterisation of the finitely presented special monoids with context-free word problem. In 2004, Duncan & Gilman asked for a characterisation of monoids with context-free word problem [144, Question 4]. To this end, Hoffmann, Holt, Owens & Thomas [209] write "the depth of the Muller-Schupp result and its reliance on the geometrical structure of Cayley graphs of groups suggests that a generalization to semigroups could be very hard to obtain". The above Theorem 3.4.1 is nevertheless such a generalisation of the Muller-Schupp theorem, free from reliance on any geometric structure. See Chapter 5 for the geometric aspects of special monoids.

We make a remark about some curious behaviour: a context-free special monoid need not be finitely presentable. For example, the monoid

$$M = \mathrm{Mon}\langle a, b, c \mid ab^i c = 1 \ (i \geq 1)\rangle$$

is not finitely presentable[63], but the rewriting system with rules $\{ab^i c \to 1 \mid i \geq 1\}$ is context-free (even regular), monadic, and complete, and hence M has context-free word problem by [69, Corollary 3.8]. Thus:

Proposition 3.5.2. *There exist (fin. gen.) non-finitely related context-free special monoids.*

This is one of few results distinguishing special monoids from groups. It would be interesting to investigate what the structure of finitely generated context-free monoids are,

[63] If it were, then only finitely many of the relations on the same alphabet would suffice to define the monoid (this result is essentially due to B. H. Neumann [39, III.Theorem 12]). Let $\mathcal{R}$ be such a finite set of relations, and let n be the largest i such that $ab^i c \in \mathcal{R}$. Then as $\mathcal{R}$ is clearly a complete rewriting system, as no rules overlap, it would follow that $ab^{n+1}c \xrightarrow{*}_{\mathcal{R}} 1$, which is impossible, as $ab^{n+1}c$ contains no subword of the form $ab^i c$ for $1 \leq i \leq n$, i.e. $ab^{n+1}c$ is irreducible mod $\mathcal{R}$.

beyond the full characterisation given in the finitely presented case. One initial approach might be to understand the relation between the group of units and such a monoid M. For example, the minimal invertible pieces of $\operatorname{Mon}\langle a, b, c \mid ab^i c = 1 \ (i \geq 1)\rangle$ can be checked elementarily to be $\Delta = \{ab^i c \mid i \geq 1\}$, which is an infinite set; thus at first glance, one might expect the group of units to be non-finitely generated. However, as all pieces are equal to 1, the group of units $U(M)$ is trivial, and hence finitely generated (by a different generating set). The monoid

$$\Pi_2 = \operatorname{Mon}\langle a, b, c \mid (ab^i c)^2 = 1 \ (i \geq 1)\rangle$$

on the other hand, has a non-trivial group of units. Furthermore, this group of units is not finitely generated, despite the fact that Π_2 has context-free word problem. This answered a question of Brough, Cain & Pfeiffer of whether the group of units of a context-free monoid is always finitely generated. We refer the reader to Nyberg-Brodda [386] for the proofs of these facts, where this example is studied in depth.

The assumption of finite presentability in Theorem 3.5.1 is necessary, as the following example demonstrates.

Example 3.5.3 (A finitely generated, not context-free, special monoid with context-free group of units)**.** Let $M = \operatorname{Mon}\langle a, b, c, d \mid ab^i c^j d = 1 \ (i, j \geq 2, j = i^2)\rangle$, and let $A = \{a, b, c, d\}$. Then the rewriting system $\mathcal{R} \subseteq A^* \times A^*$ with the rules $\{(ab^i c^j d \to 1) \mid i, j \geq 2, j = i^2\}$ is quick to check is complete; it is locally confluent, as none of the rules overlap with one another, and as it is obviously terminating, the system is complete by Newman's lemma. Thus, as ε is irreducible modulo this system, we have that $\operatorname{Rep}_A^M(1) = \langle \varepsilon \rangle_{\mathcal{R}}$. We claim that this is not a context-free language. Indeed, it is clear that as applying every rule of $\mathcal{R}$ to a word w reduces the number of occurrences of the letter a in w by one, we have

$$\operatorname{Rep}_A^M(1) \cap ab^*c^*d = \{ab^i c^j d \mid i, j \geq 2, j = i^2\}$$

but the right-hand side is well-known to not be a context-free language, which can easily be checked using e.g. the pumping lemma. Thus, as $\mathcal{C}_{\mathrm{cf}}$ is closed under intersection with regular languages, we have that $\operatorname{Rep}_A^M(1)$ cannot be a context-free language. By Corollary 3.4.3, as $\operatorname{Rep}_A^M(1)$ is not context-free, we hence have that M does not have context-free word problem. On the other hand, using the complete rewriting system $\mathcal{R}$, it is easy to check that $ab^k c^\ell$ is not invertible for any $k, \ell \geq 0$. It follows that the factorisation into minimal invertible pieces of the defining word $ab^i c^j d$ is necessarily as $(ab^i c^j d)$. Thus the minimal invertible pieces of M are all equal to 1, and hence $U(M) = \langle \lambda \rangle_M = 1$, so M has trivial (and therefore context-free) group of units. △

Now, the study of monoids in which the congruence class of every word is a context-free language has received some deal of attention; see e.g. [235]. For this reason, we write out the following corollary of Theorem 3.4.3 and Theorem 3.4.1, using the language of rewriting.

Theorem 3.5.4. *Let $T \subseteq A^* \times A^*$ be a finite special rewriting system on a finite alphabet A. Then the congruence class $[w]_{\overset{*}{\leftrightarrow}_T}$ of every word $w \in A^*$ is context-free if and only if the maximal group congruence T_G contained in T is such that A^*/T_G is a virtually free group.*

This answers a question posed in 1992 by Zhang [502, Problem 1] in the positive. This question was also asked by Book & Otto [71] in 1993. We finish with a note on word-hyperbolicity. It is not difficult to show (see e.g. [95]) that any monoid with context-free word problem is *word-hyperbolic* (in the sense of Duncan & Gilman [144]). Hence, we have the following corollary.

Corollary 3.5.5. *Let M be a finitely presented special monoid such that the group of units $U(M)$ is virtually free. Then M is word-hyperbolic.*

The author conjectures that a much stronger connection is true; this will be expanded on in future work. If this connection holds true, then it would yield an affirmative answer to the following question.

Question 3.5.6. *Let M be a finitely presented special monoid such that its group of units $U(M)$ is a hyperbolic group. Is M word-hyperbolic?*

This question was posed by Garreta & Gray [157], who showed that a finitely presented special monoid with hyperbolic group of units is itself hyperbolic (in the undirected sense, see §1.4). However, the properties of being hyperbolic and word-hyperbolic, while equivalent for groups, are in general independent of one another for monoids.

3.5.1 Some decision problems

We mention some decision problems for context-free special monoids, and related concepts. The first one is straightforward using known results in combinatorial group theory.

Theorem 3.5.7. *It is decidable whether a one-relator special monoid* $\text{Mon}\langle A \mid w = 1\rangle$ *has context-free word problem.*

Proof. Let $M = \text{Mon}\langle A \mid w = 1\rangle$. By Theorem 3.5.1, it suffices to decide whether the group of units $U(M)$ is virtually free. By using Adian's overlap algorithm (as in [6]), we can compute a presentation $U(M) = \text{Gp}\langle X \mid v = 1\rangle$ for the group of units of M. But it is well-known that it is decidable whether a one-relator group is virtually free or not; namely, $U(M)$ is virtually free if and only if the word v is a power of a primitive word in the free group on X. This can be decided using Whitehead's algorithm (see Theorem 4.3.6 for a detailed proof). □

In 1992, Zhang [502, Problem 3] asked if it is decidable whether a one-rule special rewriting system is context-free, in the sense that the congruence class of every word is a context-free language. The above Corollary 3.4.3 and Theorem 3.5.7 thus show that the answer to this question is affirmative.

For two words $u, v \in A^*$, to solve the word problem Zhang [502] provides an algorithm which is exponential in $f(|u| + |v|)$, where f is the complexity of the word problem of the group of units.[64] Thus for context-free special monoids his approach gives decidability which is

[64] As mentioned in §1.3, this solution to the word problem is not in general constructive if one considers the presentation as part of the input.

exponential in $|u| + |v|$, as hyperbolic groups have decidable word problem in linear time. We can improve this to a polynomial bound.

Theorem 3.5.8. *Let M be a finitely presented special monoid, generated by A, with virtually free group of units. Then the word problem with input $u, v \in A^*$ is decidable in $O(n^{2.3728639})$-time, where $n = |u| + |v|$.*

Proof. Let Γ be a context-free grammar generating the word problem for the group $U(M)$. Then by Theorem3.5.1, it follows that there exists a context-free grammar Γ_M generating WP_A^M; in fact, one can check that all the steps in the proof are even effective (though this is not required for the current proof). Thus the word problem for M has time complexity at most the time complexity of the membership problem for $\mathcal{L}(\Gamma_M)$. If $n = |u| + |v|$ denotes the length of the input, then the problem of membership in the language of a context-free grammar is well-known, by a result of Valiant [482], to be reducible to the problem of the multiplication of $n \times n$-matrices with entries in $\mathbb{F}_2$, for which the best algorithm is currently $O(n^{2.3728639})$, due to Le Gall [277], see also Williams [489]. □

As mentioned earlier, any context-free monoid is word-hyperbolic and it is known that word-hyperbolic semigroups have word problem decidable in polynomial time; however, unlike for the case of hyperbolic groups (in which the word problem is decidable in linear time) no upper bound on the degree of the polynomial is known to exist; the best current known algorithm cannot give better than $O(n^5 \log n)$, see [98]. Note that word-hyperbolic *groups* are well-known to have word problem decidable in linear time. Matrix multiplication cannot be faster than $O(n^2)$, and is the conjectured best time. Thus, the following conjecture seems natural.

Conjecture 3.5.9. *Let M be a finitely presented special monoid with virtually free (resp. word-hyperbolic) group of units. Then the word problem with input $u, v \in A^*$ is decidable in $O(n^2)$-time, where $n = |u| + |v|$.*

Finally, the following broad conjecture, which currently seems out of reach, presents itself naturally; it is closely related to the above.

Conjecture 3.5.10. *Let M be a finitely presented special monoid. Suppose the word problem for the group of units of M is decidable in $O(f(n))$. Then there exists a polynomial $g(n)$ such that the word problem for M is decidable in $O(g(f(n)))$. In particular, if the word problem for the group of units is in* PTIME, *then the word problem for M is in* PTIME.

Now, the rational subset membership problem (see §1.1.4) for *groups* has been relatively well-studied; see especially the recent survey by Lohrey [285]. For monoids, however, it does not appear to have been studied in any great depth, which is affirmed by Lohrey.[65] We begin by noting that the word problem for any monoid is equivalent to deciding membership in singleton subsets of the monoid, and such sets are rational subsets. Hence the word problem reduces to the rational subset membership problem. The latter is certainly much harder in general,

[65] The survey dedicates 19 pages of material on the problem for groups, and yet a single paragraph summarises all then known material on the problem for monoids (ibid. §12).

as decidability of the rational subset membership problem clearly implies decidability of the divisibility problems. We first prove a general theorem, of independent interest.

Theorem 3.5.11. *Let M be a finitely generated monoid. If M has context-free word problem, then the rational subset membership problem for M is decidable.*

Proof. Suppose M is generated by the finite set A. Let $R \subseteq M$ be a rational subset, given to us as a regular language $K \subseteq A^*$ with $\pi(K) = R$, and let $w \in A^*$. We wish to decide if $\pi(w) \in R$. Let $L = \mathrm{WP}_A^M$. We may suppose we are given a context-free grammar generating L. Then, given any rational transduction ϱ, one can effectively compute a context-free grammar whose language is precisely $\varrho(L)$, see e.g. [217]. Now by Lemma 3.4.2 we have that $\pi^{-1}(R)$ is a rational transduction of L. Thus, we can effectively compute a context-free grammar Γ whose language is $\pi^{-1}(R)$. Now $\pi(w) \in R$ if and only if $w \in \pi^{-1}(R)$, if and only if $w \in \mathcal{L}(\Gamma)$. But the membership problem in context-free languages is uniformly decidable, using e.g. the cyk algorithm (see e.g. Younger [500]). Thus there is an algorithm which takes as input K and w, produces Γ, and decides if $w \in \mathcal{L}(\Gamma)$, i.e. decides if $\pi(w) \in \pi(K)$. In other words, the rational subset membership problem is decidable. □

This generalises the fact that virtually free groups have decidable rational subset membership problem. We now have the following, by combining Theorem 3.5.1 and Theorem 3.5.11.

Corollary 3.5.12. *Let M be a finitely presented special monoid. If the group of units $U(M)$ is virtually free, then the rational subset membership problem for M is decidable.*

This corollary broadly generalises the result by Kambites-Render [418] that the rational subset membership problem is decidable in the case of the bicyclic monoid $\mathrm{Mon}\langle b, c \mid bc = 1 \rangle$, for which the group of units is trivial (and hence certainly virtually free). We also remark that Kambites, Silva & Steinberg [238, Corollary 3.5] show that virtually free groups have decidable rational subset membership problem via (rather different) language-theoretic methods. We remark that we can also obtain Corollary 3.5.12 using the monadic second-order logic of graphs, see Chapter 5 (specifically Corollary 5.5.8).

Note that decidability of the rational subset membership problem clearly implies decidability of the word problem for *any* finitely generated monoid M; for $u, v \in A^*$, we have $u =_M v$ if and only if $u \in \{v\}$. On the other hand, whereas decidability of the submonoid membership problem implies decidability of the word problem for groups (as this latter problem is equivalent to deciding membership in the trivial submonoid), there is in general no obvious reduction of the word problem to the submonoid membership problem. However, for special monoids, one has the easy following result, which we only write out as it does not appear to have been observed anywhere in the literature.

Theorem 3.5.13. *Let M be a finitely presented special monoid. If the submonoid membership problem for M is decidable, then the word problem for M is decidable.*

Proof. As M has decidable submonoid membership problem, so too does its group of units $U(M)$. Thus $U(M)$, being a group, has decidable word problem; and hence, M has decidable word problem by Makanin's theorem. □

In line with such reduction-style results, we finally turn towards the decidability of the Diophantine problem for special monoids. We do not have much do say on this topic. Garreta & Gray [157] posed the following reduction-style question.

Question 3.5.14. *Let M be a special monoid such that its group of units $U(M)$ has decidable Diophantine problem. Does M have decidable Diophantine problem?*

We remark that Makanin [309] claimed to have solved a simplified version of this problem in a bulletin article; however, no proof was ever published, despite the fact that all other claims in the bulletin article were proved in his subsequent Ph.D. thesis [308].[66] For this reason, we should consider the above question as open. Diekert & Lohrey [137], by a quick reduction to the decidability of Presburger arithmetic, proved that the first-order theory of the bicyclic monoid $\mathrm{Mon}\langle b, c \mid bc = 1\rangle$ is decidable. In particular, this implies decidability of the Diophantine problem for the bicyclic monoid. Gray & Garreta ask whether the Diophantine problem is decidable for the next easiest special monoid $\mathrm{Mon}\langle a, b, c \mid abc = 1\rangle$, which has trivial group of units. We suspect the Diophantine problem for the monoids $\mathrm{Mon}\langle a, b \mid a^n b = 1\rangle$ where $n \geq 1$ is decidable; this seems easier than the example by Gray & Garreta, as the submonoid of right units is here free of rank 1, rather than of rank 2.

In the general case it is in any case clear that new techniques will be needed, rather than passing via the first-order theory. We remark that while it is tempting to wish to deduce the property of decidability of the Diophantine problem for $\mathrm{Mon}\langle b, c \mid bc = 1\rangle$ from the fact that the bicyclic monoid is defined by a finite complete special rewriting system, this cannot be done; Otto [394] has shown that there exists a monoid M defined by a finite complete special rewriting system such that the Diophantine problem for M is undecidable.

[66] No trace of a proof of the claim can be found in any of his other publications either, despite the fact that his (arguably) most famous works were concerned with solving the Diophantine problem in free monoids and free groups. Makanin passed away in 2017, so whatever solution he may have had, if any, will remain a mystery.

3.6 Some open problems

By [273, Theorem 1.2] any *one-relation* special monoid not isomorphic to the free product of a group and a free monoid contains a submonoid isomorphic to the bicyclic monoid. However, the proof does not rely on any particular properties of one-relation monoids, but rather only on properties of invertible pieces – Lallement simply was not aware of the possibility of extending the notion of invertible pieces to general special monoids. We include a proof modelled entirely on Lallement's.

Lemma 3.6.1. *If the special monoid $M = \operatorname{Mon}\langle A \mid w_i = 1 \; (1 \leq i \leq p)\rangle$ is not isomorphic to a free product of a free monoid and a group, then M contains a submonoid isomorphic to the bicyclic monoid.*

Proof. If M is not isomorphic to a free product of a free monoid and a group, then there exists some $w \in \Delta$ such that $|w| > 1$. Write $w \equiv w_1 w_2$ for some non-empty $w_1, w_2 \in A^+$. As w is invertible, there exists some w' such that $ww' =_M 1$. Let $u \equiv w_2 w'$. Then $w_1 u =_M 1$. If $uw_1 =_M 1$, then $w_2 w' w_1 =_M 1$ whence w_1 would be invertible, contradicting the minimality of w. Thus $uw_1 \neq_M 1$. By [112, Lemma 1.31], $\langle u, w_1 \rangle$ is isomorphic to the bicyclic monoid. □

The following conjecture thus becomes natural.

Conjecture 3.6.2. *Let $M = \operatorname{Mon}\langle A \mid w_i = 1 \; (1 \leq i \leq p)\rangle$ be a special monoid. Then M has deterministic context-free word problem if and only if it is isomorphic to the free product $G * F$ of a context-free group G by a free monoid F.*

Brough, Cain, and Pfeiffer [80] conjectured that the bicyclic monoid does not have deterministic context-free word problem, which was recently answered in the affirmative by Kambites (unpublished). The forward implication of Conjecture 3.6.2 is equivalent to this conjecture by Lemma 3.6.1, as the class of monoids with deterministic context-free word problem is closed under taking finitely generated submonoids; hence the forward implication of Conjecture 3.6.2 is true. Furthermore, whereas the class of monoids with context-free word problem is closed under free products, it is not known whether the class of *deterministic* context-free word problem is, see [80, Question 6.2]. Hence Conjecture 3.6.2 is closely related to the conjecture that the class of monoids with deterministic context-free word problem is closed under free products.

One might reasonably ask to what extent the theorems of this chapter (e.g. Theorem 3.4.1) apply to classes beyond super-AFLs. A natural target might at first seem the class of *context-sensitive languages* $\mathcal{C}_{\text{cs}}$. This class contains but is significantly larger than $\mathcal{C}_{\text{cf}}$, and has many applications throughout various branches of mathematics; for example, there is a context-sensitive grammar which generates the language $\{a^p \mid p \text{ is prime}\}$, whereas this language is certainly not context-free [78]. However, $\mathcal{C}_{\text{cs}}$ is not a super-AFL, as it is not closed under homomorphism. Other techniques would need to be developed to approach this class.

There are many questions remaining open for special monoids, particularly pertaining to their presentations. We have presented some of these throughout this thesis. We end with a few more. These are related to the number of presentations a given special monoid admits. The bicyclic monoid admits exactly one one-relation monoid presentation, namely the usual one $\text{Mon}\langle b, c \mid bc = 1\rangle$. This was proved by Shneerson [444][67] and independently rediscovered by Cain & Maltcev [94, Proposition 22]. More generally, if $M = \text{Mon}\langle A \mid w = 1\rangle$, then if M admits another one-relation presentation $\text{Mon}\langle B \mid u = v\rangle$, it is obviously the case that this presentation is special. But there is some flexibility. For example, if $M = \text{Mon}\langle a, b \mid aba = 1\rangle$, then $M \cong \mathbb{Z}$, via the isomorphism induced by $a \mapsto 1$ and $b \mapsto -2$. But for all $i \geq 1$, we have that

$$\text{Mon}\langle a, b \mid a^i ba^i = 1\rangle \cong \mathbb{Z}$$

via the isomorphism $a \mapsto 1$ and $b \mapsto -2i$. Thus M admits infinitely many distinct special one-relation presentations. More generally, if $M = \text{Mon}\langle a_1, \dots, a_n \mid w = 1\rangle$ is a group, then M admits a presentation of the form

$$\text{Mon}\langle a_1, \dots, a_n \mid a_1 a_2 \cdots a_n w' a_n \cdots a_2 a_1 = 1\rangle,$$

for some word w', see e.g. the manipulations made by Perrin & Schupp [400]. Now as M is a group, M is also isomorphic to

$$\text{Gp}\langle a_1, \dots, a_n \mid a_1 a_2 \cdots a_n w' a_n \cdots a_2 a_1 = 1\rangle.$$

Let φ be the automorphism of the free group on $\{a_1, \dots, a_n\}$ defined by $a_i \mapsto a_i$ for $1 \leq i < n$ and $a_n \mapsto a_n a_1$. Then for $k \geq 1$, we have

$$\varphi^k(a_1 a_2 \cdots a_n w' a_n \cdots a_2 a_1) = a_1 a_2 \cdots a_n a_1^k \varphi(w') a_n a_1^k a_{n-1} \cdots a_2 a_1.$$

As φ is an automorphism of the free group on the generators $a_1, \dots, a_n$, we have that φ^k is also such an automorphism; hence it follows that

$$M \cong M_k := \text{Gp}\langle a_1, \dots, a_n \mid a_1 a_2 \cdots a_{n-1} a_n a_1^k \varphi(w') a_n a_1^k a_{n-1} \cdots a_2 a_1 = 1\rangle.$$

We now prove the following easy lemma, which shows that the above presentations also give rise to monoid presentations for M. In similarity to the naming of *der Freiheitssatz* for one-relator groups, we might call this lemma *der Gruppenhilfssatz* – the "group lemma" – for special monoids.

Lemma 3.6.3. *Let $M = \text{Mon}\langle a_1, a_2, \dots, a_n \mid w_1 = 1, \dots, w_k = 1\rangle$. If the generators $a_1, \dots, a_{n-1}$ are invertible and a_n appears in some defining relation w_i, then M is a group.*

Proof. Let $\Delta = \{\delta_1, \dots, \delta_m\}$ be the set of minimal invertible pieces of M. As a_n appears in some w_i, it appears in some δ_j. Consider the left-most occurrence of a_n in δ_j, and correspondingly write $\delta_j \equiv \delta_j' a_n \delta_j''$ for some $\delta_j', \delta_j'' \in \{a_1, \dots, a_n\}^*$. Now as a_n is the left-most occurrence of a_n in δ_j, we have that δ_j' does not contain a_n, and hence it is invertible. Being a proper prefix of the piece δ_j, it thus follows that δ_j' is empty. Hence δ_j begins with a_n, and so a_n is right invertible. By considering the right-most occurrence of a_n in δ_j, it follows symmetrically that δ_j ends with a_n, and so a_n is left invertible. Hence a_n is invertible, and M is a group. □

[67] I recently translated this and another paper into English; I wish to thank Lev Shneerson and Mikhail Volkov for providing me copies of the originals.

As a special case, we mention that if $\Pi = \mathrm{Mon}\langle a, b \mid r = 1\rangle$ is such that a and b both appear in r, then in Π the letter a is invertible if and only if b is invertible. Note now that, for all $k \geq 1$, in the monoids given by the presentations

$$M'_k := \mathrm{Mon}\langle a_1, \dots, a_n \mid a_1 a_2 \cdots a_{n-1} a_n a_1^k \varphi(w') a_n a_1^k a_{n-1} \cdots a_2 a_1 = 1\rangle$$

the generators $a_1, \dots, a_{n-1}$ are all invertible. Thus a_n is also invertible by the *Gruppenhilfssatz*. Hence M'_k is a group, and so $M'_k \cong M_k \cong M$. All the presentations M'_k are obviously distinct (as the length of the defining relation is strictly increasing as k increases). Thus we have proved: if M is a one-relation special monoid which is a group, then M admits infinitely many distinct one-relation special monoid presentations. Hence, we have two extremes: the bicyclic monoid (which has a single special one-relation presentation), and positive one-relator groups (which have infinitely many). We remark that there is nothing particular about the one-relation case – indeed, an easy modification shows that k-relation special monoids which are groups also admit infinitely many distinct k-relation special monoid presentations. Thus we ask: what can happen in the middle?

Question 3.6.4. *Let $n > 1$. Does there exist a special k-relation monoid which admits exactly n distinct special k-relation monoid presentations?*

The answer to the question appears to the author as almost certainly negative, but a direct proof appears difficult. Adian's overlap algorithm might yield some insight in the one-relation case when attempting to answer the question above.

Throughout this chapter, we have seen the importance of – and headaches caused by – pieces appearing as subwords of other pieces. We shall find these headaches repeated in Chapter 5. The fact that pieces in pieces cause difficulties does not seem to appear anywhere in the literature on special monoids. Garreta & Gray [157] appear to be the first to have indirectly recognised that the case in which no piece appears as a subword of another piece is easier; their theorems in [157] include this condition on the presentation.[68] However, no elaboration is made on the topic, nor is any discussion included on whether this condition can always be satisfied.

Question 3.6.5. *Does every special monoid admit a presentation with pieces Δ in which no piece appears as a subword of another piece (that is, in which Δ is an infix code)?*

It is clear from the definition of the piece-generating operation in §1.3 that Δ is an infix code if and only if Λ is an infix code. That is, if there is a piece inside a piece, then there is a witness for this in the factorisation of some defining relation. We will say that a presentation in which Δ is an infix code is an *infix presentation*. Thus the above question asks: does every special monoid admit an infix presentation? We conjecture that the answer to this question is negative; and yet the question resists direct efforts to attack it. One of the difficulties in approaching this question is that the property of pieces appearing as subwords of other pieces is not a particularly algebraic property – that is, when passing from words to elements of a monoid, the subtleties of pieces appearing as subwords of other pieces are "smoothed out". We present the only result in this direction.

[68] They call this condition (C1), see e.g. their Theorem A.

Proposition 3.6.6. *Any special monoid with trivial group of units admits an infix presentation.*

Proof. The result is almost immediate if one uses Lemma 3.2.11, but we instead present an elementary approach. Suppose that $M = \mathrm{Mon}\langle A \mid w_1 = 1, \dots, w_k = 1\rangle$ is such that $U(M) = 1$. Let Δ, Λ be the pieces resp. presentation pieces of this presentation. If some w_i is a product $\delta_1 \cdots \delta_n$ of pieces $\delta_j \in \Lambda$ and $n > 1$, then, as $U(M) = 1$, it follows that $\delta_j =_M 1$ for $1 \leq j \leq n$. Thus we may replace w_i by the n special relations $\delta_1 = 1, \delta_2 = 1, \dots, \delta_n = 1$ without changing the monoid, i.e. M is still defined by the resulting presentation. Doing this to all defining relations, we obtain a presentation

$$M' = \mathrm{Mon}\langle A \mid \delta_1 = 1, \dots, \delta_\kappa = 1\rangle$$

where $\delta_i \in \Lambda$ for all $1 \leq i \leq \kappa$, and $\kappa \geq 0$. Now, the pieces resp. presentation pieces of M' are identical with the corresponding sets for M. Furthermore, for all pieces $\delta \in \Delta \setminus \Lambda$, as $\delta =_M 1$ we have $\delta =_{M'} 1$, and so we can add the relation $\delta = 1$ to the presentation above. Doing this for every $\delta \in \Delta \setminus \Lambda$, we obtain a presentation

$$M'' = \mathrm{Mon}\langle A \mid \delta_1 = 1, \dots, \delta_{\kappa'} = 1\rangle$$

for which the pieces Δ'' and presentation pieces Λ'' coincide. If Δ'' is not an infix code, then there is some $\delta_i, \delta_j \in \Lambda$ such that $\delta_i \equiv h_1 \delta_j h_2$ and h_1, h_2 are non-trivial. By a sequence of Tietze transformations, we may replace δ_i by $h_1 h_2$ without changing the monoid, as $\delta_j =_{M''} 1$ follows from $U(M'') = U(M') = U(M) = 1$. We can of course only repeat this step finitely many times, at which point no pieces appear inside pieces; the resulting presentation will thus be an infix presentation. □

Example 3.6.7. Let $M = \mathrm{Mon}\langle a, b, c \mid abc = 1, b = 1\rangle$. Then it follows from Makanin's procedure that $U(M) = 1$, and the pieces are given by $\Delta = \{abc, ac, b\}$. Then $M \cong M' = \mathrm{Mon}\langle a, b, c \mid ac = 1, b = 1\rangle$, with pieces $\Delta' = \{ac, b\}$, an infix code. △

A candidate for a special monoid which seems unlikely to admit an infix presentation is $\mathrm{Mon}\langle a, b \mid abaabbab = 1\rangle$. The pieces of this presentation are obviously $\Delta = \{ab, aabb\}$, and the group of units is infinite cyclic. There seems nothing particular about the one-relation case with regards to the question of infix presentations.

Chapter 4
Weakly Compressible Monoids

Synopsis

This chapter will study another class of monoids defined by a combinatorial condition on their presentations. This class – of *weakly compressible monoids* – and the methods used in studying them, first discovered by Lallement and Adian & Oganesian, form an important part of the theory of one-relation monoids. Associated to any weakly compressible monoid M is a compressed monoid $L(M)$. It is known that decidability of the word problem for M is equivalent to that for $L(M)$. Here, we prove that also the language-theoretic properties of M and $L(M)$ are closely related. In particular, we will prove that if $\mathcal{C}$ is a super-AFL, then M has word problem in $\mathcal{C}$ if and only if $L(M)$ has word problem in $\mathcal{C}$ (Theorem 4.2.14). From this, we deduce several corollaries of importance to the theory of one-relation monoids. Specifically, we solve the rational subset membership problem for many classes of one-relation monoids (Corollary 4.3.5), and show that given a one-relation monoid M containing a non-trivial idempotent, it is decidable whether or not M has context-free word problem. Several of the proofs involve the alternating products and ancestors of Chapter 2. This chapter is primarily based on work in the preprint [382], and contains material included in a recent survey of the word problem for one-relation monoids [388].

4.1 Weak compression

In this section, we shall give the necessary background of weak compression as we shall need it. The exposition given here is original, but is composed of an amalgamation of the approaches by several authors; in particular, the papers by Lallement [273], Adian & Oganesian [12], Zhang [504], Kobayashi [264], Gray & Steinberg [172], as well as in §3.1 of the survey [388]. The exposition below combines aspects from several of these sources, and has been designed to be maximally amenable to language-theoretic analysis.

Let A be an alphabet. We say that a pair (u, v) of words is *sealed* by the word $w \in A^+$ if $u, v \in wA^* \cap A^*w$. If a pair is sealed by some word, then it is easy to see that it is also sealed by some unique self-overlap free word α. For example, the pair $(xyxpxyx, xyxqxyx)$ is sealed by xyx, but it is also sealed by the self-overlap free word x; and the pair $(ababab, abab)$ is sealed by $abab$, but is also sealed by the self-overlap free word ab. We will consider finitely presented monoids for which there exists some self-overlap free word α such that all defining relations $u_i = v_i$ are such that (u_i, v_i) is sealed by α. Such monoids are called *weakly compressible.* A monoid which is not weakly compressible is called *incompressible.*

Example 4.1.1. We give some examples and non-examples of weakly compressible monoids.

(1) $\text{Mon}\langle x, y \mid xyxyx = xyx, xyxxxyx = x\rangle$ is weakly compressible with $\alpha \equiv x$.

(2) $\text{Mon}\langle a, b \mid abab = ab, abaabaab = ab\rangle$ is weakly compressible with $\alpha \equiv ab$.

(3) $\text{Mon}\langle p, q \mid pqq = q, qpp = p\rangle$ is incompressible.

(4) $\text{Mon}\langle b, c \mid bc = 1\rangle$ is incompressible.

Any special monoid is incompressible (by default). Checking whether a given monoid is weakly compressible is thus straightforward and decidable. We remark that, of course, it is not the monoids themselves which are weakly compressible (or incompressible), but rather their presentations. This is similar to the fact that the definition of a *special* monoid is really a definition about presentations rather than monoids. For the remainder of this section, we will fix a finitely presented weakly compressible monoid $M = \text{Mon}\langle A \mid u_i = v_i \ (i \in I)\rangle$, with $\alpha \in A^+$ the unique self-overlap free word sealing all defining relations. We will assume that $|A| > 1$; otherwise M is just a finite monoid, and all results herein are vacuously true.

4.1.1 Conjugators and the left monoid

A word $w \in A^*$ is called a *left α-conjugator* if $w \in \alpha A^*$. Let $\Sigma_*(\alpha)$ be the set of left α-conjugators which contain exactly one occurrence of α. That is, $\Sigma_*(\alpha) = \alpha(A^* \setminus A^*\alpha A^*)$, and hence $\Sigma_*(\alpha)$ is a regular language. Furthermore, $\Sigma_*(\alpha)^+$ is the set of all left α-conjugators, and $\Sigma_*(\alpha)$ is easily seen to be a suffix code, see e.g. [264, Lemma 3.4].[69]

[69] One may more generally define left α-conjugators in the same way even when α is not self-overlap free (Zhang [504] denote this set as $S_L(\alpha)$). This language still turns out to be generated by a suffix code, which can be shown, with more work, to also be a regular language. This is the general approach taken by Lallement [273] and Zhang [504]. In principle their treatment can be translated into language-theoretic terms, but this is significantly more convoluted, and, as we shall see, it turns out to be no more general than the self-overlap free case for language-theoretic purposes.

Now, as $\Sigma_*(\alpha)$ is regular, it is countable. Indeed, it is not hard to see (as $|A| > 1$) that it has cardinality $\aleph_0$. Write $\Sigma_*(\alpha) = \{w_1, w_2, \dots, \}$, and fix a countably infinite set of symbols $\Gamma_*(\alpha) = \{\gamma_{w_1}, \gamma_{w_2}, \dots, \}$, in bijective correspondence with $\Sigma_*(\alpha)$ via the map

$$\begin{aligned}\phi \colon \Sigma_*(\alpha) &\to \Gamma_*(\alpha)\\ w_i &\mapsto \gamma_{w_i}.\end{aligned}$$

As $\Sigma_*(\alpha)$ is a (suffix) code, we can extend the map ϕ to an isomorphism (of free monoids) $\phi^* \colon \Sigma_*(\alpha)^* \to \Gamma_*(\alpha)^*$. For ease of notation we write ϕ rather than ϕ^* for this map, too.

Example 4.1.2. If we take $A = \{x, y\}$, then considering the self-overlap free word $\alpha := x$, we have that the set of left x-conjugators is

$$\alpha A^* = x\{x, y\}^* = \{x, xy, xx, xyx, xyy, xyxx, xyxy, \dots\}$$

which is generated by the suffix code

$$\Sigma_*(x) = \{x, xy, xyy, \dots\} = \{xy^i \mid i \geq 0\}.$$

Now any word in $x\{x, y\}^*$ can be uniquely factored into a product of elements from $\Sigma_*(x)$; for example, we have

$$xxyxyxyxyxyyx \equiv (x)(xy)(xy)(xy)(xy)(xyy)(x).$$

Now, we can fix a set of symbols $\Gamma_*(x) = \{\gamma_x, \gamma_{xy}, \gamma_{xy^2}, \dots\}$ in bijective correspondence with $\Sigma_*(x)$ via the bijection ϕ defined by $xy^i \mapsto \gamma_{xy^i}$. Thus, continuing the above example, we have

$$\phi(xxyxyxyxyxyyx) \equiv \phi\big((x)(xy)(xy)(xy)(xy)(xyy)(x)\big) \equiv \gamma_x \gamma_{xy}^4 \gamma_{xyy} \gamma_x.$$

Thus computing the map ϕ is not particularly difficult. △

Note that the above discussion only uses (1) the alphabet A; and (2) a self-overlap free word α. Thus, in particular we have not yet made any (direct) reference to the defining relations of the monoid M. We do this now. Every defining relation (u_i, v_i) of M is sealed by α, and so in particular we have $u_i, v_i \in A^*\alpha$. Hence we can first distinguish the right-most appearance of the self-overlap free word α in u_i resp. v_i, denoting this as $u_i \equiv u_i' \boxed{\alpha}$ resp. $v_i \equiv v_i' \boxed{\alpha}$ (here the box is only used as a notational device). As $u_i, v_i \in \alpha A^*$, it follows that $u_i', v_i' \in \alpha A^* \cup \{\varepsilon\}$. As $\Sigma_*(\alpha)$ is a suffix code generating αA^*, any word in αA^* has a *unique* factorisation into words from $\Sigma_*(\alpha)$. Thus we may uniquely factor these words u_i', v_i' into (possibly empty) products of elements from $\Sigma_*(\alpha)$. This yields a factorisation

$$\begin{aligned}u_i &\equiv w_{i,1} w_{i,2} \cdots w_{i,k_i} \boxed{\alpha}\\ v_i &\equiv w_{i,1}' w_{i,2}' \cdots w_{i,n_i}' \boxed{\alpha}\end{aligned}$$

where $w_{i,j}, w_{i,j}' \in \Sigma_*(\alpha)$. We remark that this factorisation can be trivial, i.e. we could have e.g. $u_i \equiv \boxed{\alpha}$. Any word $w_{i,j}$ or $w_{i,j}'$ arising in a factorisation of the above form is called a *left piece* of M, and the set of all left pieces of M is denoted $\Sigma(\alpha)$. We let $\Gamma(\alpha) := \phi(\Sigma(\alpha))$ be the set of symbols from $\Gamma_*(\alpha)$ corresponding to these left pieces.

For clarity, we remark on the distinction between $\Sigma(\alpha)$ and $\Sigma_*(\alpha)$. As M is finitely presented, we have that $\Sigma(\alpha)$ is finite (and hence also $\Gamma(\alpha)$ is finite). On the other hand, as $|A| > 1$, we have that $\Sigma_*(\alpha)$ (and hence also $\Gamma_*(\alpha)$) is countably infinite.

Example 4.1.3. Let $M_1 = \text{Mon}\langle x, y \mid xyxyyx = xyyyx \rangle = \text{Mon}\langle x, y \mid u = v \rangle$. The defining relation is sealed by $\alpha \equiv x$, and hence $\Sigma_*(\alpha) = \Sigma_*(x) = \{xy^i \mid i \geq 0\}$. We can factor

$$\begin{aligned} u &\equiv xyxyyx \equiv (xy)(xyy)\boxed{x} \\ v &\equiv xyyyx \quad \equiv (xyyy)\boxed{x} \end{aligned}$$

and hence $\Sigma(x) = \{xy, xyy, xyyy\}$ and $\Gamma(x) = \{\gamma_{xy}, \gamma_{xyy}, \gamma_{xyyy}\}$. △

Example 4.1.4. Let $M_2 = \text{Mon}\langle x, y \mid xyyxxxyxxyyxxxy = xy \rangle$. The defining relation is sealed by $\alpha \equiv xy$. Thus we find $\Sigma_*(\alpha) = \Sigma_*(xy)$ to be given by

$$\Sigma_*(xy) = \{xyw \mid w \in \{x, y\}^* \text{ does not contain } xy\} = xy\left(\{x, y\}^* \setminus \{x, y\}^* xy \{x, y\}^*\right).$$

We can uniquely factor the words in the defining relation over this suffix code as

$$\begin{aligned} xyyxxxyxxyyxxxy &\equiv (xyyxx)(xyx)(xyyxx)\boxed{xy} \\ xy &\equiv \boxed{xy}\,. \end{aligned}$$

Thus $\Sigma(xy) = \{xyx, xyyxx\}$ are the left pieces of the presentation. △

From the factorisation of the defining relations into left pieces, we define a new presentation

$$L_*(M) := \text{Mon}\langle \Gamma_*(\alpha) \mid \phi(w_{i,1}w_{i,2}\cdots w_{i,k_i}) = \phi(w'_{i,1}w'_{i,2}\cdots w'_{i,n_i}),\ i \in I \rangle,$$

which we shall call the *extended left monoid* of M. Note that in general, this is an infinitely generated monoid. We call the submonoid of $L_*(M)$ generated by $\Gamma(\alpha)$ the *left monoid* of M, and denote this $L(M)$. Clearly $L(M)$ has the same defining relations as $L_*(M)$, as all relations of $L_*(M)$ are words over $\Gamma(\alpha)$. That is,

$$L(M) := \text{Mon}\langle \Gamma(\alpha) \mid \phi(w_{i,1}w_{i,2}\cdots w_{i,k_i}) = \phi(w'_{i,1}w'_{i,2}\cdots w'_{i,n_i}),\ i \in I \rangle,$$

and $L_*(M) \cong \mathcal{F} * L(M)$, where $\mathcal{F}$ is a free monoid of countably infinite rank (here $*$ is the *monoid* free product). As M is finitely presented, so is $L(M)$. We remark that the sum of the lengths of the defining words of $L(M)$ is strictly less than the same sum for M. Therefore, there is a uniquely defined incompressible monoid $L(L(\cdots(L(M))\cdots))$ associated to any weakly incompressible monoid M.

Example 4.1.5. We continue with the monoid M_1 from Example 4.1.3. By the given factorisation of the relation, we have that the extended left monoid is given by

$$L_*(M_1) = \text{Mon}\langle \gamma_{xy^i}\ (i \geq 1) \mid \gamma_{xy}\gamma_{xyy} = \gamma_{xyyy} \rangle,$$

and accordingly, the left monoid is given by

$$L(M_1) = \text{Mon}\langle \gamma_{xy}, \gamma_{xyy}, \gamma_{xyyy} \mid \gamma_{xy}\gamma_{xyy} = \gamma_{xyyy} \rangle \cong \text{Mon}\langle x_1, x_2, x_3 \mid x_1x_2 = x_3 \rangle.$$

This monoid is easy to understand: it is the free monoid on two free generators. △

Example 4.1.6. We continue with the monoid M_2 from Example 4.1.4. By the given factorisation of the relation, we have that the extended left monoid is given by

$$L_*(M_2) = \text{Mon}\langle \Gamma_*(xy) \mid \gamma_{xyyxx}\gamma_{xyx}\gamma_{xyyxx} = 1 \rangle,$$

and accordingly, the left monoid is given by

$$L(M_2) = \text{Mon}\langle \gamma_{xyyxx}, \gamma_{xyx} \mid \gamma_{xyyxx}\gamma_{xyx}\gamma_{xyyxx} = 1 \rangle \cong \text{Mon}\langle a, b \mid aba = 1 \rangle.$$

In this final presentation, we may observe that the map induced by $a \mapsto -1$ and $b \mapsto 2$ is an isomorphism $\text{Mon}\langle a, b \mid aba = 1 \rangle \to \mathbb{Z}$, and hence $L(M_2) \cong \mathbb{Z}$. △

4.1.2 Normal form results

We present the classical results of Lallement [273] and Adian & Oganesian [12]. We begin with an easy observation. Given any word u not containing α (i.e. a word $u \in A^* \setminus A^*\alpha A^*$), it is not possible to apply any defining relation to u, as both sides of every defining relation of M contain α as a subword. Similarly, if a word does contain α, then any word equal to it in M also contains α as a subword, by induction on the number of elementary transformations. Hence:

Lemma 4.1.7. *If $u \in A^* \setminus A^*\alpha A^*$ and $v \in A^*$, then $u =_M v$ if and only if $u \equiv v$. In particular, if two words are equal in M, then one contains α as a subword if and only if the other does.*

Given any word containing α, we can always factor it uniquely into three parts as follows: a prefix with no occurrence of α; followed by a word in $\alpha A^* \cap A^*\alpha$; and lastly followed by a suffix with no occurrence of α. That is, if $u \in A^*\alpha A^*$, then there exist unique $u', u'' \in A^* \setminus A^*\alpha A^*$ and $u^\dagger \in \alpha A^* \cap A^*\alpha$ such that $u \equiv u'u^\dagger u''$. We call this factorisation the *canonical form* of u. We call $u^\dagger$ the *α-part* of u. The following is then easy to prove (see [12, Theorem 3]).

Lemma 4.1.8 (Adian & Oganesian). *Let $u, v \in A^*\alpha A^*$, with canonical forms $u \equiv u'u^\dagger u''$ and $v \equiv v'v^\dagger v''$, respectively. Then $u =_M v$ if and only if $u' \equiv v'$, $u'' \equiv v''$, and $u^\dagger =_M v^\dagger$.*

Now, given the α-part $u^\dagger$, we have $u^\dagger \in \alpha A^* \cap A^*\alpha$ and hence $u^\dagger \in \Sigma_*(\alpha)^*\alpha$. Write

$$u^\dagger \equiv u_0 u_1 \cdots u_n \boxed{\alpha}$$

uniquely, where $u_i \in \Sigma_*(\alpha)$ for $1 \leq i \leq n$. We can thus apply the previously constructed map $\phi\colon \Sigma_*(\alpha)^* \to \Gamma_*(\alpha)^*$ to $u_0 u_1 \cdots u_n$, which we write as

$$\phi(u_0 u_1 \cdots u_n) = \gamma_{u_0}\gamma_{u_1} \cdots \gamma_{u_n},$$

where $\phi(u_i) = \gamma_{u_i} \in \Gamma_*(\alpha)$ for $1 \leq i \leq n$. We call $\gamma_{u_0}\gamma_{u_1} \cdots \gamma_{u_n}$ the *γ-part* of u. Note that the γ-part of u can be empty. This is the case if and only if $u \equiv \alpha$.

Example 4.1.9. Let $M = \mathrm{Mon}\langle x, y \mid xyxyx = xyx\rangle$. Then M is weakly compressible with $\alpha \equiv x$. Let $u \equiv yyyxyyxyxyy$. We can write the canonical form of u as $(yyy)(xyyxyx)(yy)$, i.e. $u' \equiv yyy$, $u^\dagger \equiv xyyxyx$, and $u'' \equiv yy$. The x-part of u is $u^\dagger$, and if we write $A = \{x, y\}$, then $u^\dagger \in xA^* \cap A^*x$. Note that $u^\dagger \equiv (xyy)(xy)\boxed{x}$, so the γ-part of u is $\gamma_{xyy}\gamma_{xy}$. △

Lallement [273, Lemmas 3.1, 3.2] and Adian & Oganesian [12, Theorem 3] proved that the properties in $L_*(M)$ of the γ-part of a word controls the properties of the word.

Theorem 4.1.10 (Lallement, Adian & Oganesian). *Let $u, v \in A^*\alpha A^*$ have canonical forms $u'u^\dagger u''$ and $v'v^\dagger v''$, respectively. If the γ-parts of u and v are $\gamma_0\gamma_1 \cdots \gamma_n$ and $\gamma_0'\gamma_1' \cdots \gamma_m'$, respectively, then*

$$u =_M v \qquad \Longleftrightarrow \qquad \begin{cases} u' \equiv v', & \textit{and} \\ \gamma_0\gamma_1 \cdots \gamma_n = \gamma_0'\gamma_1' \cdots \gamma_m' \textit{ in } L_*(M), & \textit{and} \\ u'' \equiv v''. & \end{cases}$$

See [502, Proposition 4.1] for a proof of the above using rewriting systems. As $L_*(M) \cong \mathcal{F} * L(M)$, an immediate corollary of Theorem 4.1.10 is that the word problem for M is decidable if and only if it is decidable for $L(M)$. In the remainder of this chapter, we now study the language-theoretic properties of weak compression using these normal form results.

4.2 Language-theoretic compression

As in the previous section, we will throughout the remainder fix a finitely presented weakly compressible monoid $M = \text{Mon}\langle A \mid u_i = v_i \ (i \in I)\rangle$, with $|A| > 1$ and with the self-overlap free word $\alpha \in A^+$ sealing all defining relations. The overall aim of this section is to reduce the language-theoretic properties of M to $L(M)$, and vice versa. In other words, we will deduce properties of WP_A^M from properties of $\text{WP}_{\Gamma(\alpha)}^{L(M)}$, and vice versa We will generally phrase the statements of the results of this section in terms of super-AFLs. For some results, this assumption will be unduly strong, and only some of the properties of super-AFLs will be needed. However, when assembling the many statements, we will require the union of all these assumptions; this union will be the definition of a super-AFL. To write out the exact assumptions on the classes of languages in every result, which are of very limited interest on their own, would hence merely serve as a source of confusion.

One of the desired reductions is not difficult to show; namely, we can show that if M has word problem in some (sufficiently restrictive) class of languages, then so too does $L(M)$. For the proof of this direction, we will consider $\Sigma(\alpha)$, rather than $\Gamma(\alpha)$, as a generating set for $L(M)$. By [209, Proposition 8(b)], $L(M)$ has word problem in a class of languages, closed under inverse homomorphism, with respect to one finite generating set if and only if it does with respect to any finite generating set. In particular, if $\pi_\Gamma \colon \Gamma(\alpha)^* \to L(M)$ denotes the surjective homomorphism associated to the generating set $\Gamma(\alpha)$, then we can define $\pi_\Sigma \colon \Sigma(\alpha)^* \to L(M)$ by $\pi_\Sigma = \pi_\Gamma \circ \phi$, and note that this is also surjective, as ϕ is an isomorphism, so $\Sigma(\alpha)$ can be taken as a finite generating set for $L(M)$. Note that for all $w_1, w_2 \in \Sigma(\alpha)^*$,

$$w_1\alpha =_M w_2\alpha \iff \pi_\Sigma(w_1) = \pi_\Sigma(w_2).$$

The key idea for the following theorem is that $\text{WP}_{\Sigma(\alpha)}^{L(M)}$ behaves very similarly to WP_A^M intersected with a regular language.

Theorem 4.2.1. *Let M be a weakly compressible monoid, and let $L(M)$ be its left monoid. Let $\mathcal{C}$ be a super-*AFL*. If M has word problem in $\mathcal{C}$, then $L(M)$ has word problem in $\mathcal{C}$.*

Proof. Let A be the finite generating set for M, and let α be the self-overlap free word sealing all defining relations of M. For ease of notation, in this proof we will write $\Sigma_1 = \Sigma(\alpha)$ and $\Gamma_1 = \Gamma(\alpha)$ (we shall return to this notation in the sequel). Let $\varphi_{\#,\alpha}$ be the homomorphism

$$\varphi_{\#,\alpha} \colon (A \cup \{\#\})^* \to (A \cup \{\#\})^*$$

defined by $\varphi_{\#,\alpha}(\#) = \alpha\#\alpha^{\text{rev}}$ and $\varphi_{\#,\alpha}(a) = a$ for $a \in A$. We have

$$\begin{aligned}
\varphi_{\#,\alpha}\left(\text{WP}_{\Sigma_1}^{L(M)}\right) &:= \{\varphi_{\#,\alpha}(w_1\#w_2^{\text{rev}}) \mid w_1, w_2 \in \Sigma_1^*, \pi_\Sigma(w_1) = \pi_\Sigma(w_2)\} \\
&= \{w_1(\alpha\#\alpha^{\text{rev}})w_2^{\text{rev}} \mid w_1, w_2 \in \Sigma_1^*, \pi_\Sigma(w_1) = \pi_\Sigma(w_2)\} \\
&= \{w_1\alpha\#(w_2\alpha)^{\text{rev}} \mid w_1, w_2 \in \Sigma_1^*, \pi_\Sigma(w_1) = \pi_\Sigma(w_2)\} \\
&= \{w_1\alpha\#(w_2\alpha)^{\text{rev}} \mid w_1, w_2 \in \Sigma_1^*, w_1\alpha =_M w_2\alpha\} \\
&= \{u\#v^{\text{rev}} \mid u, v \in \Sigma_1^*\alpha, u =_M v\} \\
&= \text{WP}_A^M \cap \left(\Sigma_1^*\alpha\#(\alpha^{\text{rev}}\Sigma_1^{\text{rev}})^*\right),
\end{aligned}$$

where the antepenultimate (and the only slightly non-trivial) equality follows from the aforementioned fact that $w_1\alpha =_M w_2\alpha \iff \pi_\Sigma(w_1) = \pi_\Sigma(w_2)$. Note that, for every $w \in A^*\#A^*$ containing only a single $\#$, we have that the homomorphism $\varphi_{\#,\alpha}$ is injective, in the sense that

$$\varphi^{-1}_{\#,\alpha} \circ \varphi_{\#,\alpha}(w) = \{w\}.$$

Hence, more generally, for languages $L \subseteq (A \cup \{\#\})^*$ with $L \subseteq A^*\#A^*$, we have

$$\varphi^{-1}_{\#,\alpha} \circ \varphi_{\#,\alpha}(L) = L.$$

Thus, as $\mathrm{WP}^{L(M)}_{\Sigma_1} \subseteq A^*\#A^*$, we have from the above that

$$\mathrm{WP}^{L(M)}_{\Sigma_1} = \varphi^{-1}_{\#,\alpha} \circ \varphi_{\#,\alpha}\left(\mathrm{WP}^{L(M)}_{\Sigma_1}\right) = \varphi^{-1}_{\#,\alpha}\left(\mathrm{WP}^M_A \cap \left(\Sigma_1^*\alpha\#(\alpha^{\mathrm{rev}}\Sigma_1^{\mathrm{rev}})^*\right)\right).$$

Now assume M has word problem in $\mathcal{C}$. As $\mathcal{C}$ is closed under inverse homomorphism, M has word problem in $\mathcal{C}$ with respect to any finite generating set, so $\mathrm{WP}^M_A \in \mathcal{C}$. As $\mathcal{C}$ is closed under intersection with regular languages and inverse homomorphism, it follows that the right-hand side of the above equality is in $\mathcal{C}$. Thus also $\mathrm{WP}^{L(M)}_{\Sigma_1} \in \mathcal{C}$, so $L(M)$ has word problem in $\mathcal{C}$. □

Having proved one (and the easy) direction of the main theorem, we now turn toward proving the converse. This will require more technical tools; in particular, we will make use of the ancestors and alternating products of Chapter 2.

4.2.1 Initial reductions

We begin with a language-theoretic interpretation of the earlier observation (Lemma 4.1.7) regarding equality in M.

Lemma 4.2.2. *The language* WP^M_A *is a union* $\mathrm{WP}[\alpha]^M_A \cup \mathrm{W}^-_\alpha$ *of two disjoint languages*

$$\mathrm{WP}[\alpha]^M_A := \{w_1\#w_2^{\mathit{rev}} \mid w_1, w_2 \in A^*\alpha A^* \textit{ such that } w_1 =_M w_2\}$$
$$\mathrm{W}^-_\alpha := \{w\#w^{\mathit{rev}} \mid w \in A^* \setminus A^*\alpha A^*\}.$$

Proof. Let $w_1, w_2 \in A^*$ be arbitrary, and suppose $w_1\#w_2^{\mathrm{rev}} \in \mathrm{WP}^M_A$. This is equivalent to $w_1 =_M w_2$. If w_1 does not contain α, then by Lemma 4.1.7 $w_1 =_M w_2$ is equivalent to $w_1 \equiv w_2$. Hence in this case $w_1 =_M w_2$ is equivalent to $w_1\#w_2^{\mathrm{rev}} \in \mathrm{W}^-_\alpha$. On the other hand, if w_1 does contain α, then so does w_2 by another application of Lemma 4.1.7, as $w_1 =_M w_2$. Thus $w_1\#w_2^{\mathrm{rev}} \in \mathrm{WP}[\alpha]^M_A$. The two cases are disjoint; thus also the languages. □

The second of the two languages, W^-_α, appearing in the statement of Lemma 4.2.2 is very easy to describe in language-theoretic terms; it is just the intersection of the context-free language $\{w\#w^{\mathrm{rev}} \mid w \in A^*\}$, which is $\mathrm{WP}^{A^*}_A$, with the regular language $(A \setminus A^*\alpha A^*)\#(A \setminus A^*\alpha A^*)^{\mathrm{rev}}$. Thus W^-_α is a context-free language. As every super-AFL contains every context-free language by Lemma 1.2.7, and as every super-AFL is closed under union, we conclude:

Lemma 4.2.3. *Let* $\mathcal{C}$ *be a super-*AFL. *Then* $\mathrm{WP}[\alpha]^M_A \in \mathcal{C} \implies \mathrm{WP}^M_A \in \mathcal{C}$.

Proof. Any super-AFL is closed under finite unions, and as noted above $\mathrm{W}^-_\alpha \in \mathcal{C}$. Hence $\mathrm{WP}[\alpha]^M_A \in \mathcal{C}$ implies that $\mathrm{W}^-_\alpha \cup \mathrm{WP}[\alpha]^M_A \in \mathcal{C}$, and this union is WP^M_A by Lemma 4.2.2. □

Hence we have made a small amount of progress: we have reduced the properties of WP_A^M to those of $\mathrm{WP}[\alpha]_A^M$. We shall now reduce the properties of $\mathrm{WP}[\alpha]_A^M$ to those of a language $\mathrm{WP}[\alpha \sqcap \alpha]_A^M$, whose properties are then in turn easily reducible to the properties of $\mathrm{WP}_{\Gamma(\alpha)}^{L(M)}$. This will yield our main reduction theorem.

We give an informal overview of the idea. The language $\mathrm{WP}[\alpha]_A^M$ encodes equality of words over $A^*\alpha A^*$. Equality of words over $A^*\alpha A^*$ can then easily be understood in terms of equality of words over $\alpha A^* \cap A^*\alpha$ by using canonical forms. But by Theorem 4.1.10 equality of words over $\alpha A^* \cap A^*\alpha$ corresponds in a very natural way to equality of words in $L_*(M)$. It thus suffices to understand equality of words in $L_*(M)$ – although this is an infinitely generated monoid, it has the structure of a (monoid) free product $\mathcal{F} * L(M)$ of an infinitely generated free monoid by $L(M)$, where the free monoid corresponds directly to words over the suffix code $\Sigma_*(\alpha) \setminus \Sigma(\alpha)$. This means that although we cannot speak of "$\mathrm{WP}_{\Gamma_*(\alpha)}^{L_*(M)}$" (as $L_*(M)$ is not finitely generated!), we *can* speak of properties of an alternating product (from §2.2) of (1) the "word problem" of graphically equal words over $\Sigma_*(\alpha) \setminus \Sigma(\alpha)$ and (2) $\mathrm{WP}_{\Gamma(\alpha)}^{L(M)}$. We then, similarly to how we dispatched of arbitrary insertions in monoid free products in Chapter 2, will see that an ancestor of this alternating product is equal to $\mathrm{WP}[\alpha \sqcap \alpha]_A^M$. Using the properties of alternating products and ancestors proved in §2.2, this will complete our reduction to $\mathrm{WP}_{\Gamma(\alpha)}^{L(M)}$. We now prove this formally.

We first show how to use canonical forms to reduce the language-theoretic properties of $\mathrm{WP}[\alpha]_A^M$ to those of equalities of words over $\alpha A^* \cap A^*\alpha$. Let

$$\mathrm{WP}[\alpha \sqcap \alpha]_A^M := \{w_1 \# w_2^{\mathrm{rev}} \mid w_1, w_2 \in \alpha A^* \cap A^*\alpha \text{ such that } w_1 =_M w_2\}.$$

We pronounce this set as the $\alpha\alpha$-*word problem* of M. Note that

$$\mathrm{WP}[\alpha \sqcap \alpha]_A^M \subset \mathrm{WP}[\alpha]_A^M \subset \mathrm{WP}_A^M .$$

The set $\mathrm{WP}[\alpha \sqcap \alpha]_A^M$ encodes equalities of words over $\alpha A^* \cap A^*\alpha$. We give a concrete example below, showing that the idea behind $\mathrm{WP}[\alpha \sqcap \alpha]_A^M$ is not complicated.

Example 4.2.4. Let $M = \mathrm{Mon}\langle x, y \mid xyxxyyxy = xy\rangle$. Let $A = \{x, y\}$. Then M is compressible with respect to $\alpha = xy$. We factor the defining relation as

$$(xyx)(xyy)\boxed{xy} = \boxed{xy}$$

where the box is, as always, only for clarity. Thus we find the bicyclic monoid

$$L(M) = \mathrm{Mon}\langle \gamma_{xyx}, \gamma_{xyy} \mid \gamma_{xyx}\gamma_{xyy} = 1\rangle \cong \mathrm{Mon}\langle \gamma_1, \gamma_2 \mid \gamma_1\gamma_2 = 1\rangle.$$

Now $\mathrm{WP}[\alpha \sqcap \alpha]_A^M$ encodes equalities of words in $\alpha A^* \cap A^*\alpha$. For example, as

$$\gamma_1\gamma_1\gamma_2 =_{L(M)} \gamma_1,$$

we have that

$$(xyx)(xyx)(xyy)\boxed{xy} =_M (xyx)\boxed{xy}.$$

But now $(xyx)(xyx)(xyy)(xy)$ and $(xyx)(xy)$ are both in $xyA^* \cap A^*xy$, so we have

$$xyxxyxxyyxy\#(xyxxy)^{\mathrm{rev}} \in \mathrm{WP}[\alpha \sqcap \alpha]_A^M.$$

Thus, if one were to peek into the set $\mathrm{WP}[\alpha \sqcap \alpha]_A^M$, with no awareness of any reversal or monoids, one would be able to find the word $xyxxyxxyyxy\#yxxyx$ in there. △

We now define the rewriting system

$$\mathcal{R}_\alpha = \left\{ (w_1 \# w_2^{\text{rev}} \to \#) \;\middle|\; w_1 \# w_2^{\text{rev}} \in \mathrm{WP}[\alpha \sqcap \alpha]_A^M \cup \mathrm{WP}_A^{A^*} \right\}.$$

Then $\mathcal{R}_\alpha$ is a monadic rewriting system. This is not a particularly complicated system; informally, taking ancestors in this system does not "break equality" in M, in the sense of the equalities encoded by WP_A^M. Formally, we have:

Lemma 4.2.5. *If $w \in \langle \mathrm{WP}_A^{A^*} \rangle_{\mathcal{R}_\alpha}$ then $w \equiv u\#v^{\text{rev}}$ for some $u, v \in A^*$ with $u =_M v$.*

Proof. The proof is, of course, by induction on the number $k \geq 0$ of $\mathcal{R}_\alpha$-rewritings required to rewrite w to an element of $\mathrm{WP}_A^{A^*}$. If $k = 0$, then $w \in \mathrm{WP}_A^{A^*}$, in which case $w \equiv u\#u^{\text{rev}}$ for some $u \in A^*$, so the statement is true in this case. Suppose the statement is true for some $\kappa \geq 0$, and suppose $k = \kappa + 1$. Then there exists some $w' \in \langle \mathrm{WP}_A^{A^*} \rangle_{\mathcal{R}_\alpha}$ such that

$$w \to_{\mathcal{R}_\alpha} w'$$

and w' rewrites in κ steps of $\mathcal{R}_\alpha$ to some element of $\mathrm{WP}_A^{A^*}$. By the inductive hypothesis, $w' \equiv u_0\#v_0^{\text{rev}}$ for some $u_0, v_0 \in A^*$ with $u_0 =_M v_0$. As w' only contains $\#$ in a single place, and $w \to_{\mathcal{R}_\alpha} w'$, it follows that $w \equiv u_0(w_1\#w_2^{\text{rev}})v_0^{\text{rev}}$ where $w_1\#w_2^{\text{rev}} \in \mathrm{WP}[\alpha \sqcap \alpha]_A^M \cup \mathrm{WP}_A^{A^*}$. In particular $w_1 =_M w_2$. Hence $u_0 w_1 =_M v_0 w_2$, and as $u_0 w_1 \# w_2^{\text{rev}} v_0^{\text{rev}} \equiv u_0 w_1 \# (v_0 w_2)^{\text{rev}}$, we can take $u \equiv u_0 w_1$ and $v \equiv v_0 w_2$, and we are done by induction. □

Using this lemma, together with the normal form lemma, we are without much difficulty able to prove the following reduction; this uses no alternating products or ancestors, but relies only on interpreting the fact that for a word $u \in A^*\alpha A^*$ with canonical form $u'u^\dagger u''$, the crucial part for understanding how u behaves in terms of equality is $u^\dagger$.

Lemma 4.2.6. *Let $\mathcal{C}$ be a super-*AFL*. Then* $\mathrm{WP}[\alpha \sqcap \alpha]_A^M \in \mathcal{C} \implies \mathrm{WP}[\alpha]_A^M \in \mathcal{C}$.

Proof. Indeed, we claim that

$$\mathrm{WP}[\alpha]_A^M = \langle \mathrm{WP}_A^{A^*} \rangle_{\mathcal{R}_\alpha} \cap \big(A^*\alpha A^* \# A^* \alpha^{\text{rev}} A^*\big).$$

This would yield the conclusion, for as $\mathcal{C}$ is a super-AFL, we have $\mathrm{WP}_A^{A^*} \in \mathcal{C}$ by Lemma 1.2.10. As $\mathrm{WP}[\alpha \sqcap \alpha]_A^M \in \mathcal{C}$, we have that $\mathcal{R}_\alpha$ is a $\mathcal{C}$-monadic rewriting system, as the set of left-hand sides of $\#$ in $\mathcal{R}_\alpha$ is $\mathrm{WP}[\alpha \sqcap \alpha]_A^M \cup \mathrm{WP}_A^{A^*}$, and $\mathcal{C}$ is closed under finite unions. As $\mathcal{C}$ is a super-AFL, it has the monadic ancestor property by Proposition 2.1.3, and so $\langle \mathrm{WP}_A^{A^*} \rangle_{\mathcal{R}_\alpha} \in \mathcal{C}$. Finally, as $\mathcal{C}$ is a super-AFL, it is closed under intersections with regular languages; thus, if we can establish the above equality, we can conclude $\mathrm{WP}[\alpha]_A^M \in \mathcal{C}$. Let us do so.

($\supseteq$) Let $w \in \langle \mathrm{WP}_A^{A^*} \rangle_{\mathcal{R}_\alpha} \cap \big(A^*\alpha A^* \# A^* \alpha^{\text{rev}} A^*\big)$ be arbitrary. As $w \in \langle \mathrm{WP}_A^{A^*} \rangle_{\mathcal{R}_\alpha}$, by Lemma 4.2.5, we have that $w \equiv u\#v^{\text{rev}}$ for some $u, v \in A^*$ with $u =_M v$. As we also have $w \in \big(A^*\alpha A^* \# A^* \alpha^{\text{rev}} A^*\big)$, it follows that $u, v \in A^*\alpha A^*$. Hence w is of the form $u\#v^{\text{rev}}$ for two words u, v, equal in M, which both contain α. By definition, hence $w \in \mathrm{WP}[\alpha]_A^M$, and as w was arbitrary, we are done.

($\subseteq$) Suppose $w \in \mathrm{WP}[\alpha]_A^M$. Then $w \equiv u\#v^{\text{rev}}$ for some $u, v \in A^*\alpha A^*$ with $u =_M v$. Let $u \equiv u'u^\dagger u''$ and $v \equiv v'v^\dagger v''$ be the canonical forms of the respective words. Then by Lemma 4.1.8, as $u =_M v$ we have $u' \equiv v'$, $u'' \equiv v''$, and $u^\dagger =_M v^\dagger$. Hence

$$u'\#(v')^{\text{rev}}, u''\#(v'')^{\text{rev}} \in \mathrm{WP}_A^{A^*},$$

and as $u^\dagger, v^\dagger \in \alpha A^* \cap A^*\alpha$ we have $u^\dagger \#(v^\dagger)^{\text{rev}} \in \text{WP}[\alpha \sqcap \alpha]_A^M$. Hence the three rules

$$\begin{gathered}\big(u'\#(v')^{\text{rev}} \to \#\big),\\ \big((u'')\#(v'')^{\text{rev}} \to \#\big), \text{ and}\\ \big(u^\dagger\#(v^\dagger)^{\text{rev}} \to \#\big)\end{gathered}$$

are all in $\mathcal{R}_\alpha$. Hence:

$$\begin{aligned}w \equiv u\#v^{\text{rev}} \equiv u'u^\dagger u''\#(v'v^\dagger v'')^{\text{rev}} &\equiv u'u^\dagger u''\#(v'')^{\text{rev}}(v^\dagger)^{\text{rev}}(v')^{\text{rev}}\\ &\to_{\mathcal{R}_\alpha} u'u^\dagger\#(v^\dagger)^{\text{rev}}(v')^{\text{rev}}\\ &\to_{\mathcal{R}_\alpha} u'\#(v')^{\text{rev}}\\ &\to_{\mathcal{R}_\alpha} \# \in \text{WP}_A^{A^*}\end{aligned}$$

and so $w \in \langle \text{WP}_A^{A^*}\rangle_{\mathcal{R}_\alpha}$. As u and v both contain α, we also have that

$$w \equiv u\#v^{\text{rev}} \in \big(A^*\alpha A^*\#A^*\alpha^{\text{rev}}A^*\big).$$

We conclude that $w \in \langle \text{WP}_A^{A^*}\rangle_{\mathcal{R}_\alpha} \cap \big(A^*\alpha A^*\#A^*\alpha^{\text{rev}}A^*\big)$. As w was arbitrary, have thus showed that $\text{WP}[\alpha]_A^M \subseteq \langle \text{WP}_A^{A^*}\rangle_{\mathcal{R}_\alpha} \cap \big(A^*\alpha A^*\#A^*\alpha^{\text{rev}}A^*\big)$. Hence also equality holds. □

Thus, we have reduced understanding $\text{WP}[\alpha]_A^M$ to understanding the language $\text{WP}[\alpha\sqcap\alpha]_A^M$. We will now show that $\text{WP}[\alpha \sqcap \alpha]_A^M$ has the structure of an alternating product, together with some ancestry, in the sense of §2.2.

4.2.2 Utilising alternating products

We will, for ease of writing, introduce the following notation:

$$\begin{aligned}\Sigma_1 &:= \Sigma(\alpha),\\ \Sigma_2 &:= \Sigma_*(\alpha) \setminus \Sigma(\alpha).\end{aligned}$$

Obviously $\Sigma_1 \cap \Sigma_2 = \varnothing$. Let $\Gamma_i = \phi(\Sigma_i)$ for $i = 1, 2$, in which case $\Gamma_1 \cap \Gamma_2 = \varnothing$. We also have $\phi(\Sigma_i^*) = \phi(\Sigma_i)^* = \Gamma_i^*$ as Σ_i is a suffix code $(i = 1, 2)$. The following two statements are essentially just rephrasings of Theorem 4.1.10 into more directly usable forms for our purposes.

Lemma 4.2.7. *Let $u, v \in \Sigma_1^*$. Then $u\alpha =_M v\alpha$ if and only if $\phi(u) =_{L(M)} \phi(v)$.*

Proof. Note that $u\alpha =_M v\alpha$ if and only if $\phi(u) =_{L_*(M)} \phi(v)$ by Theorem 4.1.10. As $\phi(u), \phi(v) \in \Gamma_1^*$, $\phi(u) =_{L_*(M)} \phi(v)$ is equivalent to $\phi(u)$ and $\phi(v)$ being equal in the submonoid of $L_*(M)$ generated by Γ_1; and that submonoid is $L(M)$. □

Symmetrically, we also have the following easy lemma.

Lemma 4.2.8. *Let $u, v \in \Sigma_2^*$. Then $u\alpha =_M v\alpha$ if and only if $u \equiv v$.*

Proof. Note that $u\alpha =_M v\alpha$ if and only if $\phi(u) =_{L_*(M)} \phi(v)$ by Theorem 4.1.10. But $\phi(u), \phi(v) \in \Gamma_2^*$, and no defining relations of $L_*(M)$ involve letters from Γ_2. Therefore $\phi(u) =_{L_*(M)} \phi(v)$ if and only if $\phi(u) \equiv \phi(v)$. As ϕ is an isomorphism of free monoids, this is equivalent to $u \equiv v$. □

As mentioned earlier, $L_*(M)$ has the structure of a monoid free product $L(M) * \mathcal{F}$, where $L(M)$ is generated by Γ_1 and $\mathcal{F}$ is generated by Γ_2. For ease of stating the theorems below, let

$$M_1 := \langle \Gamma_1 \rangle_{L_*(M)} = L(M),$$
$$M_2 := \langle \Gamma_2 \rangle_{L_*(M)} = \mathcal{F}.$$

Let $u \equiv u_0 u_1 \cdots u_n \in (\Gamma_1 \cup \Gamma_2)^*$ be arbitrary, where $u_i \in \Gamma^*_{X(i)}$ for all $0 \leq i \leq n$ with $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$; and recall further that we say that u is *reduced* if it is empty or else $u_i \neq 1$ in $M_{X(i)}$ for all $0 \leq i \leq n$; and that furthermore every word admits a reduced form.

We now extend the notion of reduced words to words in $(\Sigma_1 \cup \Sigma_2)^*$, by a very similar definition. Let $u \in (\Sigma_1 \cup \Sigma_2)^*$ be arbitrary, factorised uniquely as $u \equiv u_0 u_1 \cdots u_n$ where $u_i \in \Sigma^*_{X(i)}$ for all $0 \leq i \leq n$ with $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$. We say that u is *reduced* if $u \equiv \alpha$ or $u \equiv \varepsilon$; or if $u_i\alpha \neq \alpha$ in M for all $0 \leq i \leq n$. We remark that the case $u \equiv \varepsilon$ will be important.

We will seek to express $\mathrm{WP}[\alpha \sqcap \alpha]^M_A$ as an alternating product of languages, with ancestors from two rewriting systems. We will first introduce the rewriting systems that we will use for this purpose. Let

$$I_\alpha = \{(w\alpha \to \alpha) \mid w \in \Sigma_1^+ : w\alpha =_M \alpha\},$$
$$I_\alpha^{\mathrm{rev}} = \{((w\alpha)^{\mathrm{rev}} \to \alpha^{\mathrm{rev}}) \mid w \in \Sigma_1^+ : w\alpha =_M \alpha\}.$$

Before proceeding to some non-trivial properties of I_α and I_α^{rev}, we show that I_α is useful for discussing reduced words, in the following sense:

Lemma 4.2.9. *Let $u \in (\Sigma_1 \cup \Sigma_2)^*$ be arbitrary, factorised uniquely as $u \equiv u_0 u_1 \cdots u_n$ where $u_i \in \Sigma^*_{X(i)}$ for all $0 \leq i \leq n$ with $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$. Then u is reduced if and only if it is irreducible modulo I_α.*

Proof. By definition, $u \equiv u_0 u_1 \cdots u_n$ is reduced if and only if it none of the u_i are such that $u_i\alpha =_M \alpha$. But this latter condition is equivalent to not having $\phi(u_i) =_{L_*(M)} 1$, which is equivalent to $\phi(u_i) =_{L(M)} 1$ and $u_i \in \Sigma_1^*$. Hence have that u is reduced if and only if none of its factors u_i (with $0 \leq i \leq n$) are such that $u_i\alpha =_M \alpha$ and $u_i \in \Sigma_1^*$. Thus $u \equiv u_0 u_1 \cdots u_n$ is reduced if and only if it is irreducible modulo I_α. □

We say that u' is a *reduced form* of u as above if (1) $u \xrightarrow{*}_{I_\alpha} u'$; and (2) u' is irreducible modulo I_α. In particular, every word $u \in (\Sigma_1 \cup \Sigma_2)^*$ has a reduced form (though this is generally not unique). Furthermore, as I_α is M-equivariant (i.e. $\xleftrightarrow{*}_{I_\alpha} \subseteq \xleftrightarrow{*}_M$), it follows that if u' is any reduced form of u, then $u' =_M u$. Given a reduced word u', we can uniquely factor it as $u'_0 u'_1 \cdots u'_n$, where the u'_i come alternatingly from Σ_1^* and Σ_2^*. We call the factorisation $u'_0 u'_1 \cdots u'_n$ the *normal form* of the reduced word w'.

The following normal form lemma is essentially a direct restatement of the usual normal form lemma for monoid free products (Lemma 1.1.3); there is an important distinction, however, in (2), as the equality there is in M, rather than the free factors of $L_*(M)$.

Lemma 4.2.10. *Let $u, v \in \Sigma_*(\alpha)^*$. Let u' resp. v' be any reduced forms of u resp. v, with normal forms*

$$u' \equiv u_0' u_1' \cdots u_m'$$
$$v' \equiv v_0' v_1' \cdots v_n'.$$

Then we have $u\alpha =_M v\alpha$ if and only if

(1) *$n = m$, and*
(2) *$u_i', v_i' \in \Sigma_{X(i)}^*$ and $u_i'\alpha =_M v_i'\alpha$ for all $0 \leq i \leq n$,*

where $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$.

Proof. As $u' \in (\Sigma_1 \cup \Sigma_2)^*$, we have that $\phi(u') \in (\Gamma_1 \cup \Gamma_2)^*$. We claim that

$$\phi(u_0')\phi(u_1') \cdots \phi(u_m')$$

is reduced with respect to the free product $L(M) * \mathcal{F} = M_1 * M_2$, i.e. $\phi(u_i') \neq 1$ in M_1 (if $\phi(u_i') \in \Gamma_1^*$) resp. M_2 (if $\phi(u_i') \in \Gamma_2^*$). But this is immediate; $\phi(u_i') = 1$ in M_1 (or M_2) if and only if $u_i'\alpha =_M \alpha$ by Theorem 4.1.10, and u' is reduced, so this latter equality never holds. Therefore $\phi(u_0')\phi(u_1') \cdots \phi(u_m')$ is a reduced form of $\phi(u')$, and this factorisation is the normal form of $\phi(u')$. Entirely symmetrically, $\phi(v_0')\phi(v_1') \cdots \phi(v_n')$ is a reduced form of $\phi(v')$, and this factorisation is the normal form of $\phi(v')$.

Now $u\alpha =_M v\alpha$ if and only if $u'\alpha =_M v'\alpha$, which is equivalent to $\phi(u') =_M \phi(v')$ by Theorem 4.1.10. Thus, by the normal form theorem for monoid free products (Lemma 1.1.3), we have (1) $n = m$; (2) $\phi(u_i'), \phi(v_i') \in \Gamma_{X(i)}^*$, and $\phi(u_i') =_{M_{X(i)}} \phi(v_i')$ for every $1 \leq i \leq n$; where $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$. As $\phi(u_i') \in \Gamma_{X(i)}^*$ if and only if $u_i' \in \Sigma_{X(i)}^*$, and $\phi(u_i') =_{M_{X(i)}} \phi(v_i')$ if and only if $u_i'\alpha =_M v_i'\alpha$, the lemma now follows. □

We are almost prepared with the setup; the following lemma will prove very important.

Lemma 4.2.11. *Let $\mathcal{C}$ be a super-*AFL *closed under reversal. Suppose $L(M)$ has word problem in $\mathcal{C}$. Then I_α and I_α^{rev} are $\mathcal{C}$-ancestry preserving.*

Proof. We begin with an informal note on the idea of the slightly technical proof in the case of I_α. Note that I_α is "almost monadic", in the sense that (1) α is self-overlap free; and (2) the right-hand side of every rule in I_α is α. Furthermore, if it were monadic, it would be a $\mathcal{C}$-rewriting system, as the set of left-hand sides of α is essentially (a transduction of) $\mathrm{IP}_{\Gamma_1}^{L(M)}$, which is in $\mathcal{C}$ by the fact that $L(M)$ has word problem in $\mathcal{C}$ and Lemma 2.3.4. By the monadic ancestor property of $\mathcal{C}$, the result would follow. This "almost monadicity" of I_α can be malleated via a transduction to show that I_α is $\mathcal{C}$-ancestry preserving.

Let $L \in \mathcal{C}$ be arbitrary. To prove that I_α is $\mathcal{C}$-ancestry preserving it suffices (by the fact that L is arbitrary), to show that $\langle L \rangle_{I_\alpha} \in \mathcal{C}$. First, let I_α^+ be the set of left-hand sides of rules in I_α. We claim $I_\alpha^+ \in \mathcal{C}$. Indeed, $w \in \Sigma_1^*$ is such that $w\alpha =_M \alpha$ if and only if $\phi(w) =_{L(M)} 1$ by Lemma 4.2.7. That is,

$$I_\alpha^+ = \left(\phi^{-1}(\mathrm{IP}_{\Gamma_1}^{L(M)})\alpha\right) \setminus \{\alpha\} = \left(\phi^{-1}(\mathrm{IP}_{\Gamma_1}^{L(M)})\alpha\right) \cap \Sigma_1^+$$

As $\mathcal{C}$ is a super-AFL, $\mathcal{C}$ is closed under rational transductions, as well as intersection and concatenation with regular languages. It follows that $I_\alpha^+ \in \mathcal{C}$.

Let now $\Diamond$ be a new symbol. Set $A_\Diamond = A \cup \{\Diamond\}$, and define a homomorphism $\varphi_\Diamond \colon A_\Diamond^* \to A^*$ by $a \mapsto a$ for all $a \in A$, and $\Diamond \mapsto \alpha$. Define a new rewriting system

$$I_\alpha^\Diamond := \{(W \to \Diamond) \mid W \in \varphi_\Diamond^{-1}(I_\alpha^+) \cap \big(A_\Diamond^* \setminus A_\Diamond^* \alpha A_\Diamond^*\big)\}.$$

Clearly $I_\alpha^\Diamond$ is a monadic rewriting system. Furthermore the language of all left-hand sides of $\Diamond$ in $I_\alpha^\Diamond$ is, of course, the intersection $\varphi_\Diamond^{-1}(I_\alpha^+) \cap \big(A_\Diamond^* \setminus A_\Diamond^* \alpha A_\Diamond^*\big)$. Since $I_\alpha^+ \in \mathcal{C}$, as shown earlier, and $\mathcal{C}$ is a super-AFL, and hence closed under inverse homomorphism and intersection with regular languages, it follows that the language of all left-hand sides of $\Diamond$ is in $\mathcal{C}$. Thus $I_\alpha^\Diamond$ is a monadic $\mathcal{C}$-rewriting system.

Now, as α is self-overlap free, reasoning (as we have done many times) by suffix codes, we have that for any word $u \in A^* \alpha A^*$, i.e. for any word u containing α, we can *uniquely* factor u as $u \equiv u_0 \alpha u_1 \alpha \cdots \alpha u_k$, with $u_i \in A^* \setminus A^* \alpha A^*$ not containing α for every $1 \leq i \leq k$. Thus we have

$$\varphi_\Diamond^{-1}(u_0 \alpha u_1 \alpha \cdots \alpha u_k) \cap \big(A_\Diamond^* \setminus A_\Diamond^* \alpha A_\Diamond^*\big) = \{u_0 \Diamond u_1 \Diamond \cdots \Diamond u_k\}.$$

That is, there is exactly one word in $A_\Diamond^*$ which (1) does not contain α; and (2) maps to $u_0 \alpha u_1 \alpha \cdots \alpha u_k$ under $\varphi_\Diamond$; that word is $u_0 \Diamond u_1 \Diamond \cdots \Diamond u_k$.[70] Denote this word by $u_\Diamond$. Of course, if u does not contain α, then $u' \equiv u$ is the unique word in $A_\Diamond^*$ such that $\varphi_\Diamond(u') = u$. In this case, we set $u_\Diamond \equiv u$. Thus $u_\Diamond$ is now defined for all words $u \in A^*$. Let $L_\Diamond$ be the set $\{w_\Diamond \mid w \in L\}$. Then, by the above argument,

$$L_\Diamond = \varphi_\Diamond^{-1}(L) \cap \big(A_\Diamond^* \setminus A_\Diamond^* \alpha A_\Diamond^*\big)$$

so in particular $L_\Diamond \in \mathcal{C}$, as $\mathcal{C}$ is closed under rational transductions.

We now claim that

$$\langle L \rangle_{I_\alpha} = \varphi_\Diamond(\langle L_\Diamond \rangle_{I_\alpha^\Diamond}).$$

This would complete the proof, by the following argument: $L_\Diamond \in \mathcal{C}$, and $I_\alpha^\Diamond$ is a monadic $\mathcal{C}$-rewriting system. Thus, as $\mathcal{C}$ is a super-AFL, $\mathcal{C}$ has the monadic ancestor property, and so it follows that $\langle L_\Diamond \rangle_{I_\alpha^\Diamond} \in \mathcal{C}$. As $\mathcal{C}$ is a super-AFL, it is closed under homomorphisms; therefore $\varphi_\Diamond(\langle L_\Diamond \rangle_{I_\alpha^\Diamond}) \in \mathcal{C}$. Thus, if the equality holds, we have $\langle L \rangle_{I_\alpha} \in \mathcal{C}$, which is what we needed to show. We now prove the desired equality.

The equality would follow if we can prove: if $w \in A^*, u \in L$ are arbitrary, then we have $w \xrightarrow{*}_{I_\alpha} u$ if and only if $w_\Diamond \xrightarrow{*}_{I_\alpha^\Diamond} u_\Diamond$. To show this it suffices to show that $(x\alpha \to \alpha) \in I_\alpha$ if and only if $((x\alpha)_\Diamond \to \Diamond) \in I_\alpha^\Diamond$. This is almost obvious, but we write out the proof for completeness.

Let $(x\alpha \to \alpha) \in I_\alpha$ be an arbitrary rule. Note that $x \in \Sigma_1^*$, so $x \in \alpha A^*$. We factor, uniquely, $x\alpha \equiv \alpha x_0 \alpha \cdots \alpha x_m \alpha$ such that for every $0 \leq i \leq m$, x_i does not contain α. Then $(x\alpha)_\Diamond \equiv \Diamond x_0 \Diamond \cdots \Diamond x_m \Diamond$, so $(x\alpha)_\Diamond \in A_\Diamond^* \setminus A_\Diamond^* \alpha A_\Diamond^*$. Furthermore, as

$$\varphi_\Diamond((x\alpha)_\Diamond) \equiv \varphi_\Diamond(\Diamond x_0 \Diamond \cdots \Diamond x_m \Diamond) = \alpha x_0 \alpha \cdots \alpha x_m \alpha \equiv x\alpha \in I_\alpha^+,$$

[70] The corresponding statement is not true if α is not self-overlap free. For example, if $\alpha \equiv xyx$, then $\varphi_\Diamond(xy\Diamond) = \varphi_\Diamond(\Diamond yx) = xyxyx$. This is perhaps the most direct appearance of the difficulties mentioned in the introduction regarding studying the language-theoretic aspects of compressing with respect to arbitrary words.

we hence have that $(x\alpha)_\Diamond \in \varphi_\Diamond^{-1}(I_\alpha^+)$, and we conclude

$$(x\alpha)_\Diamond \in \varphi_\Diamond^{-1}(I_\alpha^+) \cap (A_\Diamond^* \setminus A_\Diamond^* \alpha A_\Diamond^*).$$

Thus, if $(x\alpha \to \alpha) \in I_\alpha$, then $((x\alpha)_\Diamond \to \Diamond) \in I_\alpha^\Diamond$. To prove the converse, we simply note that if $((x\alpha)_\Diamond \to \Diamond) \in I_\alpha^\Diamond$, then $(\varphi_\Diamond((x\alpha)_\Diamond) \to \varphi_\Diamond(\Diamond)) \in I_\alpha$; but $\varphi_\Diamond(\Diamond) = \alpha$ and $\varphi_\Diamond((x\alpha)_\Diamond) = x\alpha$. Thus $(x\alpha \to \alpha) \in I_\alpha$. This completes the proof of the claim; hence I_α is $\mathcal{C}$-ancestry preserving.

The case for I_α^{rev} is a close copy of the proof for I_α, and we do not write it out; the only two non-trivial observations to make is that α is self-overlap free if and only if α^{rev} is; and that Σ_2^{rev} is a prefix code rather than a suffix code (factorisation is still unique). □

We are making good progress towards understanding $\mathrm{WP}[\alpha \sqcap \alpha]_A^M$, as we shall soon discover. Some more steps are needed. Let $\mathcal{P}_{\Sigma_2} = \{w\#w^{\text{rev}} \mid w \in \Sigma_2^*\}$. The $\mathcal{P}$ here stands for "palindrome", as every word in the language is a palindrome. We note that

$$\begin{aligned} \mathrm{WP}_A^M \cap \Sigma_2^* \# (\Sigma_2^{\text{rev}})^* &= \{w_1\#w_2^{\text{rev}} \mid w_1, w_2 \in \Sigma_2^*, w_1 =_M w_2\} \\ &= \{w_1\#w_2^{\text{rev}} \mid w_1, w_2 \in \Sigma_2^*, w_1 \equiv w_2\} \\ &= \{w\#w^{\text{rev}} \mid w \in \Sigma_2^*\} \\ &= \mathcal{P}_{\Sigma_2}, \end{aligned}$$

where the second (and only non-trivial) equality is by Lemma 4.2.8. Thus $\mathcal{P}_{\Sigma_2}$ can be seen, informally speaking, as encoding equality of words over Σ_2^*, much as $\mathrm{WP}_{\Sigma_1}^{L(M)}$ encodes equality of words over Σ_1^*. Note that $\mathcal{P}_{\Sigma_2}$ is a context-free language, being the intersection of the word problem of the free monoid $\mathrm{WP}_A^{A^*}$ with the regular language $\Sigma_2^* \cap (\Sigma_2^{\text{rev}})^*$.

The remaining material left to show before we can conclude the main theorem is rather technical. An informal overview of the idea is as follows: by Lemma 4.2.10 equality in M behaves much like a monoid free product of words over Σ_1^* and Σ_2^*, with identity α. Thus, it behaves like an alternating product of "equalities of words over Σ_1^*" and "equalities of words over Σ_2^*", with insertions of words equal to α possible at any place one spots an α. But equalities of words over Σ_1^* are entirely described by $\mathrm{WP}_{\Sigma_1}^{L(M)}$; and equalities of words over Σ_2^* are entirely described by $\mathcal{P}_{\Sigma_2}$; and insertions of words equal to α are captured by the rewriting system I_α. All these steps are thus, by now, well understood and well behaved; so it remains only to assemble them.

We now present the above argument formally, which is the final assembly step. Recall the definition of I_α and I_α^{rev} from earlier as

$$\begin{aligned} I_\alpha &= \{(w\alpha \to \alpha) \mid w \in \Sigma_1^* : w\alpha =_M \alpha\}, \\ I_\alpha^{\text{rev}} &= \{((w\alpha)^{\text{rev}} \to \alpha^{\text{rev}}) \mid w \in \Sigma_1^* : w\alpha =_M \alpha\}. \end{aligned}$$

Let $\varphi_{\#,\alpha}$ be the homomorphism sending $\#$ to $\alpha\#\alpha^{\text{rev}}$. We define the language $\mathcal{L}_\alpha$ as the $(I_\alpha, I_\alpha^{\text{rev}})$-ancestor of the image under $\varphi_{\#,\alpha}$ of the alternating product of $\mathrm{WP}_{\Sigma_1}^{L(M)}$ by $\mathcal{P}_{\Sigma_2}$. That is, we set

$$\mathcal{L}_\alpha = \left(\varphi_{\#,\alpha}\left(\mathrm{WP}_{\Sigma_1}^{L(M)} \star \mathcal{P}_{\Sigma_2}\right)\right)^{I_\alpha, I_\alpha^{\text{rev}}}.$$

First, note that $\mathcal{L}_\alpha \subseteq A^*\#A^*$. The following lemma is immediate by the properties of the terms involved in defining $\mathcal{L}_\alpha$, and some previous lemmas.

Lemma 4.2.12. *Let* $\mathcal{C}$ *be a super-*AFL. *Then* $\mathrm{WP}_{\Sigma_1}^{L(M)} \in \mathcal{C} \implies \mathcal{L}_\alpha \in \mathcal{C}$.

Proof. First, note that $\mathcal{P}_{\Sigma_2} \in \mathcal{C}$, as $\mathcal{P}_{\Sigma_2}$ is a context-free language, being equal to the intersection $\mathrm{WP}_A^{A^*} \cap (\Sigma_2^* \cap (\Sigma_2^{\mathrm{rev}})^*)$, and every context-free language is an element of every super-AFL by Lemma 1.2.7. Furthermore, $\mathcal{P}_{\Sigma_2}$ is clearly concatenation-closed, as if $w_1 \# w_1^{\mathrm{rev}}, w_2 \# w_2^{\mathrm{rev}} \in \mathcal{P}_{\Sigma_2}$, then

$$w_1 w_2 \# w_2^{\mathrm{rev}} w_1^{\mathrm{rev}} \equiv w_1 w_2 \# (w_1 w_2)^{\mathrm{rev}}$$

is in $\mathcal{P}_{\Sigma_2}$. Now, as $\mathrm{WP}_{\Sigma_1}^{L(M)} \in \mathcal{C}$ by assumption, and $\mathrm{WP}_{\Sigma_1}^{L(M)}$ is concatenation-closed, it follows by Corollary 2.2.7 that $\mathrm{WP}_{\Sigma_1}^{L(M)} \star \mathcal{P}_{\Sigma_2} \in \mathcal{C}$. As $\mathcal{C}$ is closed under homomorphism,

$$\varphi_{\#,\alpha}\left(\mathrm{WP}_{\Sigma_1}^{L(M)} \star \mathcal{P}_{\Sigma_2}\right) \in \mathcal{C}.$$

By Lemma 4.2.11, the rewriting systems $I_\alpha, I_\alpha^{\mathrm{rev}}$ are $\mathcal{C}$-ancestry preserving. Hence, by Lemma 2.2.13, we finally have that

$$\left(\varphi_{\#,\alpha}\left(\mathrm{WP}_{\Sigma_1}^{L(M)} \star \mathcal{P}_{\Sigma_2}\right)\right)^{I_\alpha, I_\alpha^{\mathrm{rev}}} \in \mathcal{C},$$

and this is precisely what was to be shown. □

Having deduced the properties of $\mathcal{L}_\alpha$, we are ready to reveal its true identity.

Lemma 4.2.13. $\mathrm{WP}[\alpha \sqcap \alpha]_A^M = \mathcal{L}_\alpha$.

Proof. We prove the claim one inclusion at a time.

($\subseteq$) Suppose $u \# v^{\mathrm{rev}} \in \mathrm{WP}[\alpha \sqcap \alpha]_A^M$ is arbitrary. Then $u, v \in \alpha A^* \cap A^* \alpha$ and $u =_M v$. We factor u and v as words over $(\Sigma_1 \cup \Sigma_2)^* \alpha$ as

$$u \equiv u_0 u_1 \cdots u_m \alpha,$$

$$v \equiv v_0 v_1 \cdots v_n \alpha,$$

where $u_i, v_j \in \Sigma_1^* \cup \Sigma_2^*$ for all $0 \leq i \leq m$ and $0 \leq j \leq n$. Let u', v' be any irreducible descendant of u resp. v under I_α. Then $u \xrightarrow{*}_{I_\alpha} u'$ and $v \xrightarrow{*}_{I_\alpha} v'$, so $v^{\mathrm{rev}} \xrightarrow{*}_{I_\alpha^{\mathrm{rev}}} (v')^{\mathrm{rev}}$. Furthermore, as I_α is M-equivariant, we have $u =_M u'$ and $v =_M v'$.

An easy induction on the number of rules applied shows that $u', v' \in \alpha A^* \cap A^* \alpha$. We factor u' and v' uniquely over $(\Sigma_1 \cup \Sigma_2)^*$ as

$$u' \equiv u'_0 u'_1 \cdots u'_s \alpha,$$

$$v' \equiv v'_0 v'_1 \cdots v'_t \alpha,$$

such that $u'_i, v'_j \in \Sigma_1^* \cup \Sigma_2^*$ for all $0 \leq i \leq s$ and $0 \leq j \leq t$, and such that none of the u'_i for $0 \leq i \leq s$ (resp. v'_j for $0 \leq j \leq t$) are empty unless all of them are. By Lemma 4.2.9, as u' and v' are irreducible modulo I_α, both u' and v' are reduced, so these factorisations are normal forms for u' and v'.

Now $u =_M v$, so $u' =_M u =_M v =_M v'$. That is

$$u'_0 u'_1 \cdots u'_s \boxed{\alpha} =_M v'_0 v'_1 \cdots v'_t \boxed{\alpha},$$

where the boxed α is only used to aid in seeing how to apply the normal form lemma. Indeed, by that lemma, i.e. Lemma 4.2.10, it follows that (1) $s = t$, and (2) $u'_i, v'_i \in \Sigma_{X(i)}^*$ and $u'_i \alpha =_M v'_i \alpha$ for all $0 \leq i \leq s$; where $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$.

Note that for those i such that $X(i) = 1$, we have that $u_i'\alpha =_M v_i'\alpha$ implies $u_i' =_{L(M)} v_i'$ when Σ_1 is, as before, considered as a generating set for $L(M)$. That is, for those i such that $X(i) = 1$, we have $u_i'\#(v_i')^{\mathrm{rev}} \in \mathrm{WP}_{\Sigma_1}^{L(M)}$. On the other hand, for those i such that $X(i) = 2$, we have that $u_i'\alpha =_M v_i'\alpha$ implies $u_i' \equiv v_i'$ by Lemma 4.2.8, as $u_i', v_i' \in \Sigma_2^*$. In other words, for those i such that $X(i) = 2$, we have $u_i'\#(v_i')^{\mathrm{rev}} \in \mathcal{P}_{\Sigma_2}$.

Thus, by the definition of the alternating product, we have

$$(u_0'u_1' \cdots u_s')\#(v_0'v_1' \cdots v_t')^{\mathrm{rev}} \in \mathrm{WP}_{\Sigma_1}^{L(M)} \star \mathcal{P}_{\Sigma_2}.$$

Applying $\varphi_{\#,\alpha}$, we hence have

$$(u_0'u_1' \cdots u_s')(\alpha\#\alpha^{\mathrm{rev}})(v_0'v_1' \cdots v_t')^{\mathrm{rev}} \in \varphi_{\#,\alpha}\left(\mathrm{WP}_{\Sigma_1}^{L(M)} \star \mathcal{P}_{\Sigma_2}\right)$$

which is to say

$$(u_0'u_1' \cdots u_s'\alpha)\#(v_0'v_1' \cdots v_t'\alpha)^{\mathrm{rev}} \in \varphi_{\#,\alpha}\left(\mathrm{WP}_{\Sigma_1}^{L(M)} \star \mathcal{P}_{\Sigma_2}\right)$$

Now, as mentioned earlier, we have $u \xrightarrow{*}_{I_\alpha} u'$ and $v^{\mathrm{rev}} \xrightarrow{*}_{I_\alpha^{\mathrm{rev}}} (v')^{\mathrm{rev}}$, i.e.

$$u \xrightarrow{*}_{I_\alpha} u_0'u_1' \cdots u_s'\alpha, \quad \text{and}$$

$$v^{\mathrm{rev}} \xrightarrow{*}_{I_\alpha^{\mathrm{rev}}} (v_0'v_1' \cdots v_t'\alpha)^{\mathrm{rev}}.$$

It follows by definition of $(I_\alpha, I_\alpha^{\mathrm{rev}})$-ancestors that we hence have

$$u\#v^{\mathrm{rev}} \in \left(\varphi_{\#,\alpha}\left(\mathrm{WP}_{\Sigma_1}^{L(M)} \star \mathcal{P}_{\Sigma_2}\right)\right)^{I_\alpha, I_\alpha^{\mathrm{rev}}} = \mathcal{L}_\alpha$$

which proves the forward inclusion, as $u\#v^{\mathrm{rev}}$ was arbitrary.

($\supseteq$) Let $u\#v \in \mathcal{L}_\alpha$ be arbitrary. We will prove that (1) we have $u, v^{\mathrm{rev}} \in \alpha A^* \cap A^*\alpha$; and (2) that $u =_M v^{\mathrm{rev}}$. We would then be able to conclude that $u\#v \in \mathrm{WP}[\alpha \sqcap \alpha]_A^M$, and hence we would have the required inclusion. For ease of notation, we will abbreviate $\mathcal{W}_1 = \mathrm{WP}_{\Sigma_1}^{L(M)}$ and $\mathcal{W}_2 = \mathcal{P}_{\Sigma_2}$.

As $u\#v$ is in $\mathcal{L}_\alpha$, there exists some $u'\#v' \in \mathcal{W}_1 \star \mathcal{W}_2$ such that

$$u \xrightarrow{*}_{I_\alpha} u'\alpha,$$

$$v \xrightarrow{*}_{I_\alpha^{\mathrm{rev}}} \alpha^{\mathrm{rev}}v',$$

where we have used the fact that $\varphi_{\#,\alpha}(u'\#v') = (u'\alpha)\#(\alpha^{\mathrm{rev}}v')$. As I_α is M-equivariant, it follows that $u =_M u'\alpha$. Similarly, as $v \xrightarrow{*}_{I_\alpha^{\mathrm{rev}}} \alpha^{\mathrm{rev}}v'$ implies that

$$v^{\mathrm{rev}} \xrightarrow{*}_{I_\alpha} (v')^{\mathrm{rev}}\alpha$$

we can conclude that $v^{\mathrm{rev}} =_M (v')^{\mathrm{rev}}\alpha$. Thus

$$u =_M u'\alpha, \qquad (v')^{\mathrm{rev}}\alpha =_M v^{\mathrm{rev}},$$

and so to show $u =_M v^{\mathrm{rev}}$ we must only show the missing equality $u'\alpha =_M (v')^{\mathrm{rev}}\alpha$.

Now, as u', v' are in $\mathrm{WP}_{\Sigma_1}^{L(M)} \star \mathcal{P}_{\Sigma_2}$, we can write

$$u'\#v' \equiv u_0u_1 \cdots u_n\#v_n \cdots v_1v_0$$

such that $u_i\#v_i \in \mathcal{W}_{X(i)}$ for all $0 \leq i \leq n$, where $X(2j) = 1$ and $X(2j+1) = 2$, or else $X(2j) = 2$ and $X(2j+1) = 1$. If $X(i) = 1$, then $u_i\#v_i \in \mathcal{W}_1 = \mathrm{WP}_{\Sigma_1}^{L(M)}$, so $u_i =_{L(M)} v_i^{\mathrm{rev}}$ and $u_i, v_i^{\mathrm{rev}} \in \Sigma_1^*$. On the other hand, if $X(i) = 2$, then $u_i\#v_i \in \mathcal{W}_1 = \mathcal{P}_{\Sigma_2}$, and so $u_i \equiv v_i^{\mathrm{rev}}$ and $u_i, v_i^{\mathrm{rev}} \in \Sigma_2^*$. Thus, we conclude $u_i\alpha =_M v_i^{\mathrm{rev}}\alpha$ for all $0 \leq i \leq n$,

Hence we conclude

$$u'\alpha \equiv u_0 u_1 \cdots u_n \alpha =_M v_0^{\mathrm{rev}} v_1^{\mathrm{rev}} \cdots v_n^{\mathrm{rev}} \alpha \equiv (v_n \cdots v_1 v_0)^{\mathrm{rev}} \alpha \equiv (v')^{\mathrm{rev}} \alpha,$$

which was the missing equality; we conclude $u =_M v^{\mathrm{rev}}$.

It remains to show that $u, v^{\mathrm{rev}} \in \alpha A^* \cap A^* \alpha$. As $u_0, v_0^{\mathrm{rev}} \in \Sigma_{X(i)}^* \subset \alpha A^*$, it follows that $u', (v')^{\mathrm{rev}} \in \alpha A^*$. Furthermore, as every rule $(\ell \to \alpha)$ in I_α is such that $\ell \in A^*\alpha$, we conclude from $u \xrightarrow{*}_{I_\alpha} u'\alpha$ that $u \in A^*\alpha$. Similarly, we conclude from $v \xrightarrow{*}_{I_\alpha^{\mathrm{rev}}} \alpha^{\mathrm{rev}} v'$ that $v^{\mathrm{rev}} \xrightarrow{*}_{I_\alpha} (v')^{\mathrm{rev}}\alpha$, and hence also that $v^{\mathrm{rev}} \in A^*\alpha$. Thus we have proved

$$u, v^{\mathrm{rev}} \in \alpha A^* \cap A^* \alpha,$$

and $u =_M v^{\mathrm{rev}}$. Thus $u\#v \in \mathrm{WP}[\alpha \sqcap \alpha]_A^M$, and we are done. □

With this, we are ready to assemble the main theorem.

4.2.3 Main theorem

The proof of the main theorem now requires no additional information than the results proved in the previous section. We write out the steps in detail.

Theorem 4.2.14. *Let M be a weakly compressible monoid, and let $L(M)$ be its left monoid. Let $\mathcal{C}$ be a super-*AFL *closed under reversal. Then M has word problem in $\mathcal{C}$ if and only if $L(M)$ does.*

Proof. ($\implies$). This is the statement of Theorem 4.2.1.

($\impliedby$). If $L(M)$ has word problem in $\mathcal{C}$, then by Lemma 4.2.12 we have $\mathcal{L}_\alpha \in \mathcal{C}$. By Lemma 4.2.13, $\mathcal{L}_\alpha = \mathrm{WP}[\alpha \sqcap \alpha]_A^M$, so $\mathrm{WP}[\alpha \sqcap \alpha]_A^M \in \mathcal{C}$. By Lemma 4.2.6, it follows that $\mathrm{WP}[\alpha]_A^M \in \mathcal{C}$, and hence, by Lemma 4.2.3, we conclude $\mathrm{WP}_A^M \in \mathcal{C}$. □

Thus we have proved the main theorem of this chapter. By iterating compression until we arrive at an incompressible monoid, this hence gives a complete characterisation of the word problem of weakly compressible monoids in terms of the word problem of incompressible monoids. Recalling that the classes of context-free and indexed languages are super-AFLs closed under reversal, we find:

Corollary 4.2.15. *Let M be a weakly compressible monoid, and let $L(M)$ be its left monoid. Then M has context-free (resp. indexed) word problem if and only if $L(M)$ has context-free (resp. indexed) word problem.*

Thus, if we turn our attention to context-free monoids, in order to completely characterise the compressible monoids with context-free word problem, it suffices to completely characterise which incompressible monoids do. However, this latter problem seems currently vastly out of reach. Certainly some examples are easy to construct.

Example 4.2.16. Let $M = \mathrm{Mon}\langle a_1, a_2, \dots, a_n \mid a_1^{\beta_1} a_2^{\beta_2} \cdots a_n^{\beta_n} = a_k \rangle$ be a one-relation monoid with $n > 1$, such that $1 \leq k \leq n$ and $\beta_i \geq 1$ for all $1 \leq i \leq n$. Then M is incompressible, for at most one of the first or the last letters of each side of the defining relation can be the same. The finite rewriting system with the single rule $(a_1^{\beta_1} a_2^{\beta_2} \cdots a_n^{\beta_n} \to a_k)$ is complete and monadic, as $a_1^{\beta_1}a_2^{\beta_2}\cdots a_n^{\beta_n}$ has no self-overlap. Thus by [66, Corollary 3.8], M has context-free word problem. Some concrete examples of these types of incompressible monoids with context-free word problem are given below.

(1) $\mathrm{Mon}\langle a, b \mid ab = b\rangle$.
(2) $\mathrm{Mon}\langle a, b, c \mid aabbcc = b\rangle$.

Some concrete examples of compressible monoids which compress to the examples above (in order), as well as the self-overlap free word α with respect to which it compresses, are given below. Thus the below monoids have context-free word problem by Corollary 4.2.15:

(1) $\mathrm{Mon}\langle x, y \mid xyxyyxy = xyyxy\rangle$ with $\alpha = xy$.
(2) $\mathrm{Mon}\langle x, y \mid xyxyxxxyyxyyx = xx\rangle$ with $\alpha = x$.

We note that the compressible monoids in these examples are both 2-generated, whether or not they compress to a 2-generated monoid. Having more generators often makes it easier to devise a finite complete rewriting system defining the monoid, which means showing that such monoids are context-free occasionally becomes simplified. △

We give an application of the above result to the rational subset membership problem. The rational subset membership problem for groups has been relatively well-studied. We have already seen that having context-free word problem implies having a decidable rational subset membership problem (Theorem 3.5.11). Thus we have the following corollary.

Corollary 4.2.17. *Let M be a weakly compressible monoid, and let $L(M)$ be the left monoid associated to M. If $L(M)$ has context-free word problem, then the rational subset membership problem for M is decidable.*

Recall that for general monoids there is no obvious reduction of the word problem to the submonoid membership problem, although there is one for groups (obvious) and special monoids (Theorem 3.5.13). The following becomes natural.

Question 4.2.18. *Let M be a weakly compressible monoid. If M has decidable submonoid membership problem, is the word problem decidable for M? If $L(M)$ has decidable submonoid membership problem, does M have decidable submonoid membership problem?*

We conjecture that the former of the two questions can be answered negatively, and that the latter can be answered positively. The reason for our pessimism in the first question is that there is no reason to expect decidability of the word problem to reduce to the submonoid membership problem for general monoids; and the class of weakly compressible monoids behaves more or less as the class of general monoids. The second question, on the other hand, asks nothing about decidability in such a class, giving rise to some optimism. We will now turn our attention to a very particular case of weakly compressible monoids.

4.3 Subspecial monoids

One particular type of weakly compressible monoid occurs in the one-relation case. A one-relation monoid $\mathrm{Mon}\langle A \mid u = v\rangle$, with $|u| \geq |v|$, is called *subspecial* if it is special or if $u \in vA^* \cap A^*v$. We begin with a brief overview of the general theory of subspecial monoids, and refer the reader to the article by Gray & Steinberg [172] for a full algebraic investigation of as well as recent topological results on subspecial monoids.

The interest in subspecial monoids originated with the following theorem, proved by Lallement [273], which completely characterises the one-relation monoids M with a non-trivial idempotent, i.e. an element $m \in M$ with $m^2 = m$, but $m \neq 1$.

Theorem (Lallement)**.** *Let $M = \mathrm{Mon}\langle A \mid u = v\rangle$ be a one-relation monoid with $|u| > |v| > 0$. Then M has a non-trivial idempotent if and only if M is subspecial.*

Note that if $|u| = |v|$ then it is already clear that $\mathrm{Mon}\langle A \mid u = v\rangle$ does not have any non-trivial idempotent. Furthermore, if $|v| = 0$, then it is not hard to see that the special monoid $M = \mathrm{Mon}\langle A \mid u = 1\rangle$ has a non-trivial idempotent unless every invertible piece of M is a letter (recall this terminology from Chapter 3). This is easily decided by using Adian's overlap algorithm [3]. Thus, we conclude that it is decidable whether a one-relation monoid has a non-trivial idempotent.

Given any subspecial one-relation monoid $\mathrm{Mon}\langle A \mid u = v\rangle$ where $u \in vA^* \cap A^*v$, it is clear that v seals (u, v). What is perhaps not as immediately clear is that such a presentation is weakly compressible, for v may not be self-overlap free. With some afterthought, however, one finds a self-overlap free word which seals (u, v). With further thought, it is also rather clear that the resulting compressed monoid is itself subspecial. We illustrate this with an example.

Example 4.3.1. Let $M = \mathrm{Mon}\langle x, y \mid xyxxyxxxyxxyxyx = xyx\rangle$. Then the reader may at a glance note that xyx seals this defining relation. Thus x also seals this defining relation, as x is a prefix and a suffix of xyx. Compressing this monoid with respect to $\alpha := x$, we thus obtain

$$L(M) = \mathrm{Mon}\langle \gamma_\varepsilon, \gamma_y \mid \gamma_y\gamma_\varepsilon\gamma_y\gamma_\varepsilon\gamma_\varepsilon\gamma_y\gamma_\varepsilon\gamma_y\gamma_y = \gamma_y\rangle \cong \mathrm{Mon}\langle a, b \mid babaababb = b\rangle.$$

But now this monoid is itself subspecial, and hence weakly compressible, this time with respect to $\alpha := b$. Compressing this, we obtain the special monoid

$$L(L(M)) \cong \mathrm{Mon}\langle \gamma_\varepsilon, \gamma_a, \gamma_{a^2} \mid \gamma_a\gamma_{a^2}\gamma_a\gamma_\varepsilon = 1\rangle \cong \mathrm{Mon}\langle a, b, c \mid abac = 1\rangle.$$

By Adian's algorithm, this special monoid has trivial group of units, and hence by Corollary 3.5.1 it follows that $L(L(M))$ has context-free word problem. This can also be witnessed by the fact that the rewriting system with the single rule $(abac \to \varepsilon)$ is complete, monadic, and defines $L(L(M))$, which implies that its word problem is context-free [66, Corollary 3.8]. In either case, it follows by Corollary 4.2.15 that $L(M)$ has context-free word problem, and hence, by another application of the same corollary, that M has context-free word problem. △

The following is essentially a restatement of [264, Lemma 5.4] and is implicitly used in [12]; the reader is directed to the former reference for a full proof.

Proposition. *Suppose that the one-relation monoid $M = \text{Mon}\langle A \mid u = v\rangle$ is a subspecial, but not special, monoid. Then there exists some self-overlap free word $\alpha \in A^+$ which seals (u, v), so M is weakly compressible. Furthermore, $L(M)$ is itself subspecial.*

Hence for every (non-special) subspecial monoid M there exists some special monoid obtained by first compressing M, then compressing $L(M)$, etc. until one obtains an incompressible and special monoid. This process, which is effective, is guaranteed to terminate, as the defining relation in a compressed monoid has total length shorter than that of the original monoid. Furthermore, this incompressible monoid is unique by [172, Corollary 4.3]. We denote the special monoid obtained by iteratively compressing M in this manner by $L_s(M)$, and call it the *left special monoid* of M. If M is special itself, we set $L_s(M) := M$. We remark that $L_s(M)$ is always a one-relation special monoid. In particular, $L_s(M)$ has decidable word problem, as its group of units is a one-relator group, see [3]. As weak compression reduces decidability of the word problem to the compressed monoid by Theorem 4.1.10, it follows by repeated application of the same lemma that any subspecial one-relation monoid has decidable word problem.

4.3.1 The word problem for subspecial monoids

By the earlier discussion and the definition of the left special monoid of a subspecial monoid, together with repeated application of Theorem 4.2.14, we find the following reduction of the word problem for subspecial monoids to the special case.

Theorem 4.3.2. *Let $\mathcal{C}$ be a super-*AFL. *Let M be a subspecial one-relation monoid. Then M has word problem in $\mathcal{C}$ if and only if $L_s(M)$ has word problem in $\mathcal{C}$.*

We shall now make this theorem algebraic in nature, and omit any mention of the compressed monoid $L_s(M)$. Recall that in Chapter 3, it was shown that if $\mathcal{C}$ is a super-AFL closed under reversal, then any finitely presented special monoid has word problem in $\mathcal{C}$ if and only if its group of units has word problem in $\mathcal{C}$. We can utilise this result, together with some structural results on the maximal subgroups of subspecial monoids, due to Gray & Steinberg, and of special monoids, due to Malheiro, to obtain an algebraic variation of Theorem 4.3.2 which does not mention compression.

Corollary 4.3.3. *Let $\mathcal{C}$ be a super-*AFL *closed under reversal. Let M be a subspecial one-relation monoid. Then M has word problem in $\mathcal{C}$ if and only if all of its maximal subgroups have word problem in $\mathcal{C}$.*

Proof. Let $M = \text{Mon}\langle A \mid u = v\rangle$ be a subspecial one-relation monoid, with $|u| \geq |v|$, and $u \in vA^* \cap A^*v$. Recall that by Theorem 3.4.1, if $\mathcal{C}$ is as in the statement of the present theorem, then if Π is any special monoid, it follows that Π has word problem in $\mathcal{C}$ if and only if its group of units $U(\Pi)$ has word problem in $\mathcal{C}$.

First, the maximal subgroups of a special monoid are all isomorphic to its group of units [325, Theorem 4.6], and hence if $v \equiv \varepsilon$, i.e. if M is special, then the result follows. Thus assume

$v \not\equiv \varepsilon$. Then by [172, Lemma 5.2], the maximal subgroups of M are all isomorphic to the group of units of $L_s(M)$, with the exception of the group of units of M, which is trivial.

($\implies$) The maximal subgroups of M are all finitely presented, being isomorphic to the group of units of a finitely presented special monoid, which is always finitely presented [309]. Thus, as $\mathcal{C}$ is closed under inverse homomorphism, the property of M having word problem in $\mathcal{C}$ is inherited by finitely generated submonoids of M by [209, Proposition 8(b)]. Hence if M has word problem in $\mathcal{C}$, then all its maximal subgroups have word problem in $\mathcal{C}$.

($\impliedby$) As the maximal subgroups of M all have word problem in $\mathcal{C}$, in particular the group of units of $L_s(M)$ has word problem in $\mathcal{C}$. As $\mathcal{C}$ is a super-AFL closed under reversal, it follows by Theorem 3.4.1 that $L_s(M)$ also has word problem in $\mathcal{C}$. By Theorem 4.3.2, M has word problem in $\mathcal{C}$. □

As we did in §3.5, we will now make our results specific to the context-free case.

4.3.2 Context-free subspecial monoids

As the class of context-free languages is a class which satisfies all the conditions to apply the above corollary, i.e. $\mathcal{C}_{\text{cf}}$ is a super-AFL closed under reversal, we immediately have the following, by combining Corollary 4.3.3 with the Muller-Schupp theorem.

Theorem 4.3.4. *A subspecial one-relation monoid has context-free word problem if and only if all of its maximal subgroups are virtually free.*

Remark 4.3.1. By [172, Lemma 5.2], all non-trivial maximal subgroups of a subspecial monoid are pairwise isomorphic.

This gives a complete and completely algebraic characterisation of subspecial monoids with context-free word problem, completely and algebraically answering for the class of subspecial monoids the question from 2004 of Duncan & Gilman on the structure of monoids with a context-free word problem [144, Question 4]. We mention that as all special monoids are subspecial, Theorem 4.3.4 generalises Theorem 3.5.1 for the class of one-relation monoids.

The following result is little more than a restatement of Corollary 4.2.17 in the setting of subspecial monoids, when combined with Theorem 4.3.4.

Corollary 4.3.5. *Let M be a subspecial one-relation monoid such that all of its non-trivial maximal subgroups are virtually free. Then M has decidable rational subset membership problem.*

We finish the chapter with two algorithmic results.

Theorem 4.3.6. *Given a presentation* $\operatorname{Mon}\langle A \mid u = v \rangle$ *of a subspecial monoid M, it is decidable whether or not the word problem of M is context-free.*

Proof. Given a presentation $\operatorname{Mon}\langle A \mid u = v \rangle$ as input, a presentation for $L_s(M)$ is effectively computable, as a single step of compression is of course easily computable. Then $L_s(M)$ is a one-relation special monoid, given by some special presentation $\operatorname{Mon}\langle \Gamma \mid w = 1 \rangle$. Thus, the group of units $\mathcal{U} = U(L_s(M))$ is a one-relator group (see §1.3). Furthermore, we can effectively compute a presentation for this group by Adian's algorithm. As all non-trivial maximal subgroups of M are isomorphic to $\mathcal{U}$ by [172, Lemma 5.2], it follows by Theorem 4.3.4 that M has context-free word problem if and only if $\mathcal{U}$ is virtually free.

Thus as $U(L_s(M))$ is a *positive* one-relator group (i.e. a one-relator group $\mathrm{Gp}\langle A \mid w = 1\rangle$ where $w \in A^*$), in the sense of Baumslag [35], to complete the proof it suffices to exhibit an algorithm for deciding if an arbitrary positive one-relator group is virtually free, given a positive one-relator presentation as input. We show a stronger claim, namely: there exists an algorithm for deciding if an arbitrary one-relator group is virtually free, given a one-relator presentation for this group as input. This result is well-known, but a direct reference is hard to come by.

Let $H = \mathrm{Gp}\langle A \mid w = 1\rangle$ be a one-relator group. Assume without loss of generality that w is cyclically reduced. Then H has torsion if and only if w is graphically a proper power u^k of some cyclically reduced word u which is not a proper power, for some $k > 1$, see [154]. Thus it is decidable, by checking all possible decompositions of w for one of this form, whether H is torsion-free or not. A torsion-free group is virtually free if and only if it is free [462]. A one-relator group given by a presentation $\mathrm{Gp}\langle A \mid w = 1\rangle$ is free if and only if w is empty or a primitive element of F_A, where F_A denotes the free group on A, see [488, Theorem 4]. By Whitehead's algorithm, it is decidable whether an element of a finitely generated free group is primitive [488]. Hence if H is torsion-free it is decidable whether H is free. Assume instead that H has torsion. Then we can uniquely decompose $w \equiv u^k$ as above. Now H is virtually free if and only if u is a primitive element in F_A by [154, Theorem 3]. Thus, if H has torsion, it is again decidable whether or not it is virtually free. Hence it is decidable whether a given one-relator group is virtually free, given a one-relator presentation for this group. □

Corollary 4.3.7. *Let $M = \mathrm{Mon}\langle A \mid u = v\rangle$ be a one-relation monoid containing a non-trivial idempotent. Then it is decidable whether M has context-free word problem.*

Proof. By Lallement's theorem and the subsequent discussion, M has a non-trivial idempotent if and only if (1) $|u| > |v| > 0$ and M is subspecial; or (2) M is special, and not every invertible piece of M is a letter, i.e. there is some piece of length greater than 1. It is easy to check directly whether we are in case (1). If not, then M is a special monoid. By using Adian's overlap algorithm we can check whether every invertible piece of M is a letter, and thereby check whether we are in case (2). To decide if M has context-free word problem, we thus first decide which case we are in. If we are in case (1), then it can be decided whether or not M has context-free word problem by Theorem 4.3.6. On the other hand, if we are in case (2), then as M is special (and in particular subspecial), we can also decide whether M has context-free word problem by another application of Theorem 4.3.6. □

Recall that Zhang [502, Problem 3] asked in 1992 if it is decidable whether a special one-relation monoid has context-free word problem. We answered this question affirmatively in Chapter 3 as Theorem 3.5.7. As any special one-relation monoid M contains a non-trivial idempotent whenever M is not a free product of a group by a free monoid, it follows that

Corollary 4.3.7 answers a generalisation of Zhang's problem affirmatively.

We remark that, in general, it is not decidable whether a weakly compressible monoid has context-free word problem. The argument is easy. Let G be any finitely presented group, given by a presentation $\mathrm{Gp}\langle A \mid r_i = 1 \ (i \in I)\rangle$. Then G admits some finite special monoid presentation $\mathrm{Mon}\langle B \mid r_i' = 1 \ (i \in I)\rangle$ by introducing a new set A^{-1} of formal inverse symbols a^{-1} for every generator $a \in A$, and setting $B = A \cup A^{-1}$; the relations r_i can then be rewritten over B in the obvious way. Let x be a symbol not in B, and let $\vartheta_x \colon B^* \to (B \cup \{x\})^*$ be the function which sends any word $w \in B^*$ to the same word, but in which each letter b_i has been replaced by xb_i. For example, $\vartheta_x(bcd) = (xb)(xc)(xd)$. Let now

$$M = \mathrm{Mon}\langle B \cup \{x\} \mid \vartheta_x(r_i')x = x\rangle.$$

Then M is weakly compressible with respect to x, and $L(M) \cong G$. Thus, by Theorem 4.2.14, M has context-free word problem if and only if G has context-free word problem. But G is context-free if and only if it is virtually free, and it is undecidable in general whether an arbitrary finitely presented group is virtually free; being virtually free is a Markov property, so this latter claim follows by the Adian-Rabin theorem [2, 413]. The result follows.

This final result paints a broader (albeit informal) picture: the property of being weakly compressible is not powerful on its own; indeed, any exotic and rowdy behaviour whatsoever can be exhibited in a rather straightforward manner in a weakly compressible monoid. Instead, the advantage of weakly compressible monoids comes from their transient behaviour, in the following sense: if one studies a general monoid M, and by some manner of investigation finds oneself reducing a problem about M to a problem about a weakly compressible monoid M', then there is a good chance that one can reduce the original problem to some problem for the compressed monoid $L(M')$. At the very least, this is true for language theoretic and some decision theoretic problems (as we have just seen). It would be interesting to study what other problems this adage is applicable to.

Chapter 5

Context-free Graphs and Special Monoids

Synopsis

In this chapter, we will discuss graph-theoretic constructions and their relation to monoids. In particular, we will be interested in how the underlying algebraic structure of a special monoid is reflected in its right Cayley graph. In §5.1, we will describe a rather general tree-like construction for constructing new context-free graphs from a given context-free graph. We will then, in §5.2, study the submonoid of right units of special monoids, and gain qualitative information on the properties of the embedding of this submonoid. In §5.3, we will introduce the *Schützenberger graph of the units* $\mathfrak{U}$ of a special monoid, which is a graph-theoretic representation of the embedding of the group of units into a special monoid. In §5.4, we will use the aforementioned results to construct the Schützenberger graph of 1 for special monoids; this is divided into two parts. In the first part, which deals with the simple class of *no-folding* special monoids, all details of the proof are written out in full (§5.4.1). In the second part, which deals with special monoids in general, we introduce a notion of a *benign* special monoid. We show that no-folding special monoids are benign. We claim that all special monoids are benign, and provide a sketch proof of this claim (§5.4.2). In §5.5, we show that the right Cayley graph of a benign special monoid M is a context-free graph if and only if the group of units $U(M)$ of M is virtually free (Theorem 5.5.4). This also gives a classification of when the monadic second-order logic of the right Cayley graph of a special monoid is decidable. As a corollary, we deduce that if M is benign and $U(M)$ is virtually free, then M has decidable rational subset membership problem (Corollary 5.5.8). This gives an entirely different proof of the same statement (Corollary 3.5.12) from Chapter 3, which has no assumption of benignity. Finally, we discuss further directions, including growth rate. We characterise the special monoids with subexponential growth, and prove that a special monoid with intermediate growth is a group.

5.1 Trees of copies

For ease of notation, throughout this section all graphs will be labelled with alphabet A. Let Γ be a connected labelled graph rooted at 1 with bounded degree. Let $R = \{\Gamma_0, \Gamma_1, \dots\}$ be a (necessarily at most countable) set of representatives of the end-isomorphism classes of Γ. We will require (which loses no generality) that $\Gamma(1) \sim \Gamma_i$ if and only if $i = 0$. For every i appearing in this list, let Δ_i be the set of frontier points of Γ_i, and let $F(\Gamma) = \cup_i \Delta_i$, where the union is taken over all i appearing in the earlier list. We assume 1 is the representative for Γ_0, so $1 \in F(\Gamma)$. For example, in Figure 5.1, we can take $F(\Gamma)$ as a set of four vertices: the root 1, together with the three vertices around any triangle other than the central one. Then, for every $v \in V(\Gamma)$ there is some $v' \in F(\Gamma)$ such that there is an end-isomorphism $\Phi\colon \Gamma(v) \to \Gamma(v')$ with $\Phi(v) = \Phi(v')$. That is, $F(\Gamma)$ is a set of representatives of frontier points of the ends of Γ. In particular, as every representative Γ_i of an end of Γ has finitely many frontier points, we conclude that $F(\Gamma)$ can (and will) always be chosen to be finite if Γ is context-free.

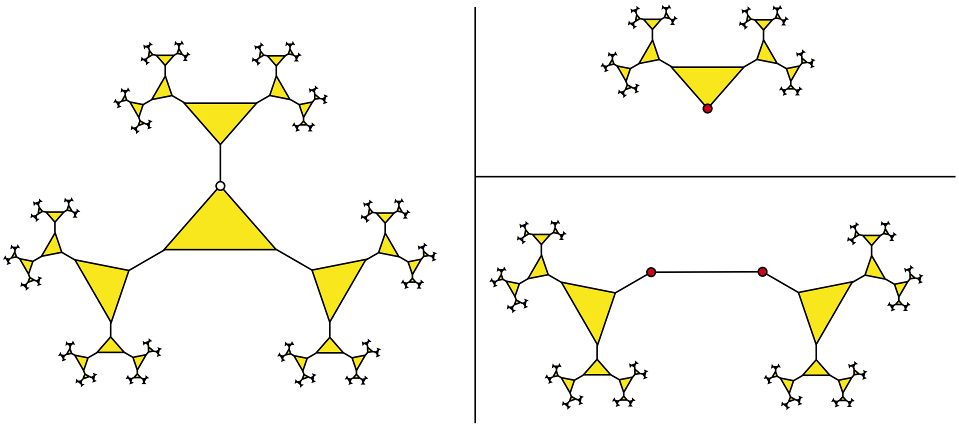


Figure 5.1: For the graph Γ to the left in this figure, we may take $F(\Gamma)$ as the root 1 (the central vertex of Γ) together with the three red vertices whose locations and end-isomorphism classes are indicated to the right. See Figure 1.1 for a detailed caption of this figure.

We remark that when we will in the sequel write e.g. "let Γ be a graph, then $F(\Gamma)$..." is slightly abusive; of course, for a given graph Γ, we can pick whichever set of representatives of end-isomorphism classes we want, and each such set could yield a different $F(\Gamma)$. A more correct notation, which would yield a uniquely determined set, would be $F_{\Gamma_0,\Gamma_1,\dots,}(\Gamma)$ or $F_R(\Gamma)$, where R is a fixed set of representatives. We shall not use this. Indeed, the particular choice made will either not be important (up to it being e.g. finite), or else it will be clear from the context.

For any subset $S \subseteq F(\Gamma)$, we will now define a graph $\textsc{Tree}(\Gamma, S)$, and will show that if Γ is a context-free graph then so too is $\textsc{Tree}(\Gamma, S)$, for any (non-trivial) choice of S. We first give some intuition. Informally, the graph $\textsc{Tree}(\Gamma, S)$ will capture the idea of "a tree of copies of Γ", where the "copying" or "branching" of the tree takes place only on vertices which are end-equivalent to a vertex of S. If one were to "branch" at completely arbitrary places in Γ and thus construct a tree of copies, then it seems likely that the resulting graph could be highly non-context-free, even if Γ itself was chosen to be context-free. However, the condition that we only "branch" at vertices which are end-equivalent to a vertex from S, which is necessarily a finite set if Γ is context-free, means that the resulting tree of copies will only have finitely many different types of "branching behaviours", which gives some initial justification for the claim that this tree of copies should be a context-free graph.

We now make this formal. Fix some $S \subseteq F(\Gamma)$ with $\mathtt{1} \not\in S$. By definition of $F(\Gamma)$, every vertex $v \in V(\Gamma)$ is such that there exists a unique i and a vertex $f \in \Delta_i \subseteq F$ together with an end-isomorphism $\psi : \Gamma(v) \to \Gamma_i$ with $\psi(v) = f$. In particular, there exists a set $V_S \subseteq V(\Gamma)$ which is the collection of all $v \in V(\Gamma)$ with an end-isomorphism such that $\psi(v) \in S$. That is, V_S is the equivalence class under end-isomorphism of the finite set S. We remark that if Γ is context-free and infinite, then whereas S is necessarily finite, the set V_S will in general be infinite, unless $S = \varnothing$. Furthermore, $\mathtt{1} \not\in V_S$ by our specification that $\Gamma(\mathtt{1}) \sim \Gamma_i$ if and only if $i = 0$, and the fact that a vertex of Γ is the frontier point of at most one Γ_i.

We inductively define graphs $\textsc{Tree}_n(\Gamma, S)$ and associated distinguished sets of vertices $V_n \subseteq V(\textsc{Tree}_n(\Gamma, S))$. Define $\textsc{Tree}_0(\Gamma, S) := \Gamma$, and $V_0 := V_S$. Assume that for some $n \geq 0$ the graphs $\textsc{Tree}_k(\Gamma, S)$ and the sets V_k have been defined for $0 \leq k \leq n$. Then $\textsc{Tree}_{n+1}(\Gamma, S)$ is defined as the graph obtained from attaching a copy Γ_v of Γ to every vertex $v \in V_n$ inside $\textsc{Tree}_n(\Gamma, S)$, identifying the root of Γ_v with v. The root of $\textsc{Tree}_{n-1}(\Gamma, S)$ is defined as the root of $\textsc{Tree}_n(\Gamma, S)$. Denote by $V_{v,S}$ the copy of the subset V_S inside $V(\Gamma_v)$. Then we define

$$V_{n+1} := \bigcup_{v \in V_n} V_{v,S} \subseteq V(\textsc{Tree}_{n+1}(\Gamma, S)).$$

In other words, V_{n+1} is defined by taking the union of all the newly added copies of V_0. Note that the graphs

$$\textsc{Tree}_0(\Gamma, S) \subseteq \textsc{Tree}_1(\Gamma, S) \subseteq \cdots \subseteq \textsc{Tree}_k(\Gamma, S) \subseteq \cdots$$

form a directed system, and in particular their directed colimit (in the category of rooted, directed, labelled graphs) exists. If S is non-empty, then the inclusions are strict.

Definition 5.1.1. For a graph Γ and a set $S \subseteq F(\Gamma)$ of vertices of Γ, we define the *tree of copies of* Γ *with respect to* S as

$$\textsc{Tree}(\Gamma, S) := \varinjlim_{i} \textsc{Tree}_i(\Gamma, S) = \bigcup_{i \geq 0} \textsc{Tree}_i(\Gamma, S).$$

The set $\bigcup_i V_i$ is called the set of *branch points* of $\textsc{Tree}(\Gamma, S)$.

An example of $\textsc{Tree}(\Gamma, S)$ for when Γ is a triangle-shaped graph and S consisting of the two non-root vertices is shown in Figure 5.2; an example with the same Γ but S consisting of only a single vertex is shown in Figure 5.3.

Note that if we were to permit $\mathtt{1} \in S$, then $\textsc{Tree}(\Gamma, S)$ would in general not have bounded degree, as we would then in the above definitions have that $\mathtt{1} \in V_n$ for all $n \geq 0$, and infinitely many copies of Γ will be attached to $\mathtt{1}$. However, without this restriction, it is easy to see that $\textsc{Tree}(\Gamma, S)$ has bounded degree. Thus, to restrict ourselves to this latter case, we will call a subset S of $F(\Gamma)$ a set of *attachment points* (of Γ) if we have $\mathtt{1} \not\in S$. We note again that if Γ is context-free, then S can be chosen to be a finite set, as $F(\Gamma)$ can be. Hence if Γ is a context-free

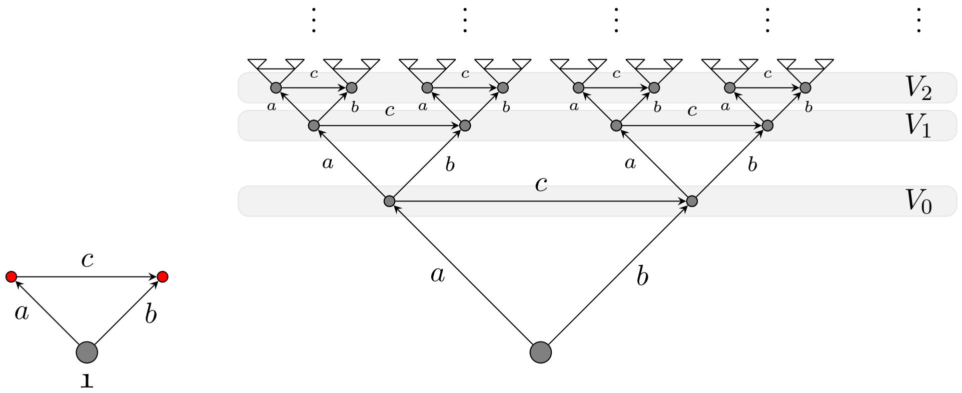


Figure 5.2: A triangle graph Γ (left), and a tree of copies of Γ (right). The set S of attachment points consists of two frontier points (onto which copies are attached), and are marked as red vertices. It is clear that the tree of copies is context-free, having only two distinct end-spaces; that of the root vertex, and the union of two such spaces, connected by an edge. For all $i \geq 0$, the set V_i (considered as a subset of $\textsc{Tree}(\Gamma, S)$) consists of the 2^{i+1} vertices in the ith "layer" of the figure, as indicated by the rounded rectangles.

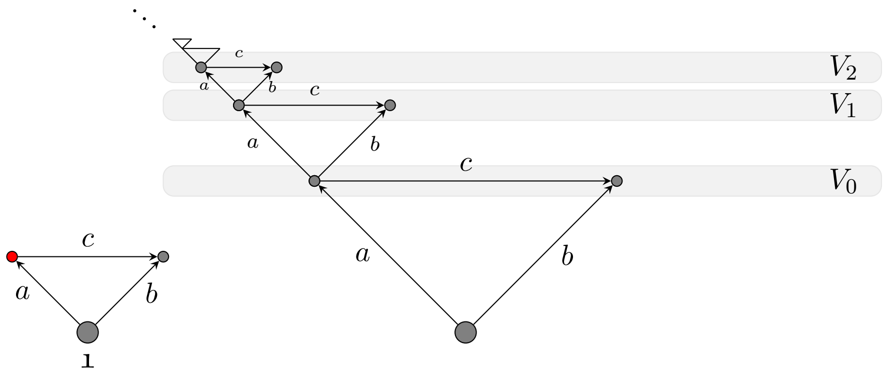


Figure 5.3: A triangle graph Γ (left), and a tree of copies $\textsc{Tree}(\Gamma, S)$ of Γ (right). The set S of attachment points consists of a single frontier point (onto which copies are attached), which is marked as a red vertex. It is clear that the tree of copies is a context-free graph, having only two distinct end-spaces; namely $\Gamma(\mathbb{1})$, and one isomorphic to $\Gamma(\mathbb{1})$ with a vertex adjoined and an edge labelled by c going out from $\mathbb{1}$ to this vertex. For all $i > 0$, the set V_i (considered as a subset of $\textsc{Tree}(\Gamma, S)$) consists of the 2^{i+1} vertices at the ith "layer" of the figure, as indicated by the rounded rectangles.

graph, then $\textsc{Tree}(\Gamma, S)$ can be encoded by a finite amount of information, viz. the ends of Γ together with S. In fact, we will show that in this setting, $\textsc{Tree}(\Gamma, S)$ is actually a context-free graph. Before this, we make a simple observation about $\textsc{Tree}(\Gamma, S)$.

Proposition 5.1.2. *Let Γ be a graph, and $S \subseteq F(\Gamma)$ a set of attachment points. Then the set of vertices of $\textsc{Tree}(\Gamma, S)$ is in bijective correspondence with the set of finite tuples*

$$(u_0, u_1, \dots, u_k)$$

where $u_i \in V_0$ for $0 \leq i < k$; and where $u_k \in V(\Gamma) \setminus \{\mathbb{1}\}$ if $k > 0$, or else $u_k \in V(\Gamma)$. Furthermore, for two vertices $u, w \in \textsc{Tree}(\Gamma, S)$ with corresponding tuples $(u_0, u_1, \dots, u_k)$ and $(w_0, w_1, \dots, w_n)$, respectively, there is an edge $u \xrightarrow{a} w$ if and only if one of the following holds:

(1) $n = k$, $u_i = w_i$ for all $0 \leq i \leq k-1$, and $(u_k \xrightarrow{a} w_n) \in E(\Gamma)$; or

(2) $n = k+1$, $u_i = w_i$ for all $0 \leq i \leq k-1$, $u_k \in V_0$, and $(\mathbb{1} \xrightarrow{a} w_n) \in E(\Gamma)$.

Proof. Let $v \in \textsc{Tree}(\Gamma, S)$. Then there exists a minimal $k \geq 0$ such that $v \in V(\textsc{Tree}_k(\Gamma, S))$.

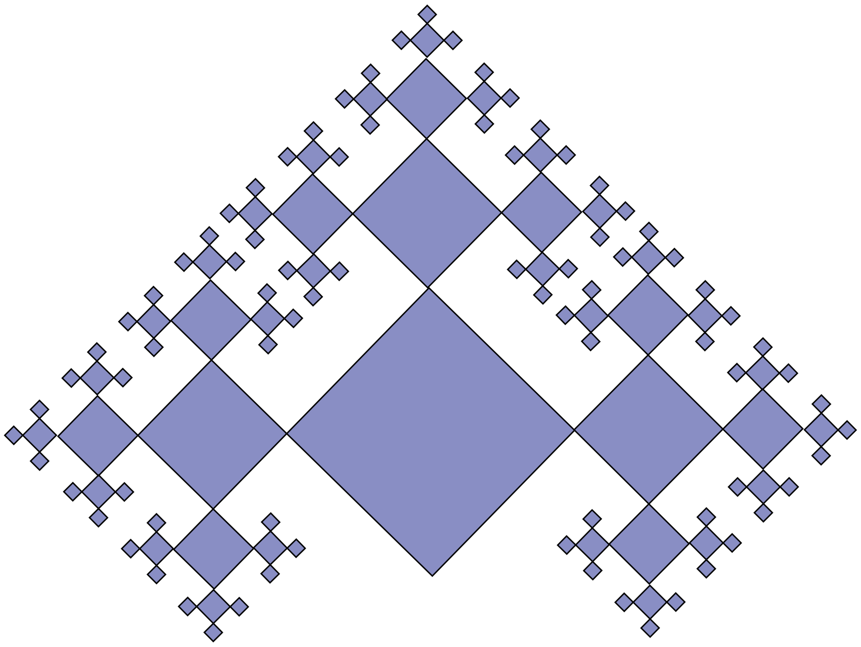


Figure 5.4: A tree of copies of a square graph, with branch points everywhere except the root vertex (the only vertex of degree 2). This graph is context-free (regardless of edge labels) as it has two distinct end-spaces; that of the root, and that of either of the two vertices at distance 1 from the root. Both of these repeat infinitely often.

Hence, we will for all $k \geq 0$ define a bijection ϕ_k from the set of vertices of $\textsc{Tree}_k(\Gamma, S)$ to the set of tuples of the above form of length at most k. This bijection will then extend to a bijection ϕ from the set of vertices of $\textsc{Tree}(\Gamma, S)$ of the desired form when taking the directed colimit.

If $k = 0$, then $v \in V(\textsc{Tree}_0(\Gamma))$; this set is in bijective correspondence with $V(\Gamma)$ by definition, and we will define ϕ_0 to be this bijection.

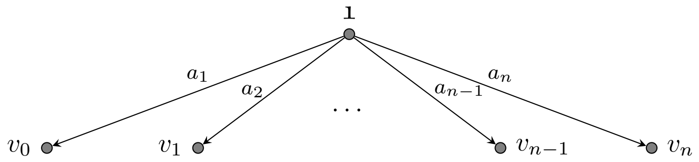
Assume, for strong induction, that ϕ_k is defined for all $0 \leq k \leq n$ for some $n \geq 0$. Let $v \in \textsc{Tree}(\Gamma, S)$ be such that $v \in V(\textsc{Tree}_{n+1}(\Gamma, S))$ with $n + 1$ minimal. Since $n + 1$ is minimal and $1 \notin S$, we have that there exists a unique $u_n \in V_n$ such that $v \in \Gamma_{u_n}$, and furthermore since Γ_{u_n} is a copy of Γ, there exists a vertex $v' \in V(\Gamma)$ uniquely determined by $v \in V(\Gamma_{u_n})$. By strong induction, we can find a unique n-tuple $\phi_n(u_n)$ representing u_n. We then define $\phi_{n+1}(v) := (\phi_n(u_n), v')$. In this resulting tuple, the first n entries will be chosen as elements of V_0, and the final will be any element of $V(\Gamma)$, as needed. This completes the claim concerning the vertices by induction.

Now, for the edges, it is clear that the only possible adjacent vertices are either such that they belong to the same added copy of Γ; or one is an element of V_n for some $n \geq 0$, and the other is connected to the root of the copy of Γ attached to the former. But clearly, the first case happens if and only if criterion (1) of the statement of the proposition holds; the second case happens if and only if criterion (2) holds. This completes the proof. □

Example 5.1.3. The above proposition gives an interpretation of $\textsc{Tree}(\Gamma, S)$ as a "free" graph with basis Γ; of course, the set S plays an important rôle in the construction, so this analogy only works in full if V_S is all of $V(\Gamma) \setminus \{1\}$. For example, if Γ is a graph with three vertices $\{1, v_1, v_2\}$

and two edges $(\mathtt{1} \xrightarrow{a_1} v_1)$ and $(\mathtt{1} \xrightarrow{a_2} v_2)$, then letting $S = \{v_1, v_2\}$ we find that $\textsc{Tree}(\Gamma, S)$ is isomorphic to the right Cayley graph of the free monoid $\{a_1, a_2\}^*$ with generating set $\{a_1, a_2\}$. The isomorphism is given by $(v_{i_1}, v_{i_2}, \dots, v_{i_n}) \mapsto a_{i_1} a_{i_2} \cdots a_{i_n}$, where $i_j \in \{1, 2\}$ for all $1 \leq j \leq n$. Under this isomorphism, the set V_k corresponds to the 2^{k+1} words over $\{a_1, a_2\}$ of length $\leq k + 1$.

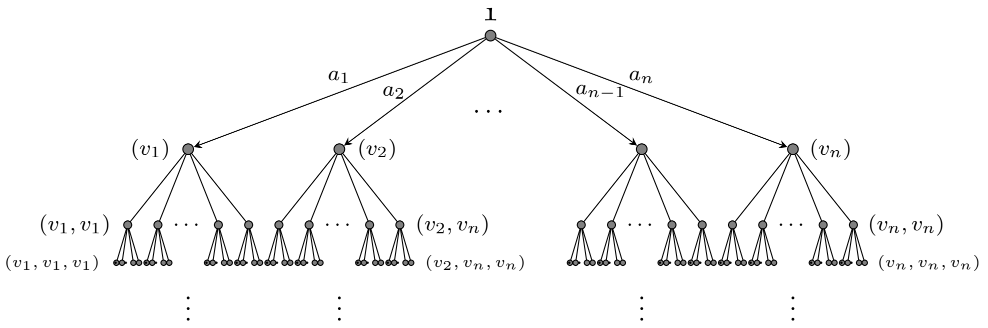
Indeed, we can generalise this to arbitrary free monoids as follows. Let A be a finite set, enumerated as $A = \{a_1, a_2, \dots, a_n\}$. Consider the graph Γ below:



As Γ is finite, it is context-free. If we set

$$\{\Gamma_0, \Gamma_1, \dots, \Gamma_n, \Gamma_{n+1}\} = \{\Gamma(\mathtt{1}), \Gamma(v_0), \Gamma(v_1), \cdots, \Gamma(v_n)\}$$

i.e. such that $\Gamma_i = \Gamma(v_{i-1})$ for $1 \leq i \leq n + 1$, then, with this choice, $F(\Gamma) = V(\Gamma)$, as v_i is a frontier point of Γ_{i+1}, and $\mathtt{1}$ is (obviously) a frontier point of $\Gamma(\mathtt{1})$. Thus $F(\Gamma)$ is a set of representatives of frontier points. Let $S = \{v_1, \dots, v_n\}$. Then S is a set of attachment points, and the graph $\textsc{Tree}(\Gamma, S)$ will be the infinite n-regular graph



where the edge labels and directions repeat in each layer as in the first layer. We have placed labels adjacent to some vertices to indicate the tuple associated to them by Proposition 5.1.2. $\triangle$

Because of this proposition, we will often consider this above prescribed bijection as invisible, i.e. we will say that vertices of $\textsc{Tree}(\Gamma, S)$ with $S \neq \varnothing$ are equal to a tuple of elements of the above prescribed form. A natural definition in light of the proposition is the following.

Definition 5.1.4. The *depth function* $d \colon V(\textsc{Tree}(\Gamma, S)) \to \mathbb{N}$ is defined for $v \in V(\textsc{Tree}(\Gamma, S))$ as follows: if $(u_0, u_1, \dots, u_k)$ is the tuple associated to v by Proposition 5.1.2, then $d(v) := k$.

Alternatively, $d(v)$ is the least integer n such that $v \in V(\textsc{Tree}_n(\Gamma, S))$. We note that the edge case, and indeed the base cases for many inductive arguments on the depth, that $d(u) = 0$ for all $u \in V(\Gamma)$, where we canonically identify u with the 1-tuple (u). In particular, $d(\mathtt{1}) = 0$. We are now ready to show the proposition which is the main reason for using trees of copies.

Proposition 5.1.5. *If* Γ *is context-free and* S *is a set of attachment points, then* $\textsc{Tree}(\Gamma, S)$ *is context-free.*

Proof. Assume that Γ is context-free. We may then without loss of generality assume S is finite. If S is empty, then $\textsc{Tree}(\Gamma, S) = \textsc{Tree}(\Gamma, \varnothing) = \Gamma$, and there is nothing to show. Hence, assume that S is non-empty. Since $1 \notin S$, it is clear that $\textsc{Tree}(\Gamma, S)$ has bounded degree, as every vertex has a copy of Γ attached at most once. For ease of notation, we will denote $\mathcal{T} := \textsc{Tree}(\Gamma, S)$ and, for every $k \geq 0$, let $\mathcal{T}_k := \textsc{Tree}_k(\Gamma, S)$. We will now show that for any $v \in V(\mathcal{T})$ there exists $v' \in V(\mathcal{T}_0)$ such that $\mathcal{T}(v) \sim \mathcal{T}(v')$.

To establish this claim, let $v \in V(\textsc{Tree}(\Gamma, S))$ be any vertex. Let $n = d(v)$. If $n = 0$, then we may take $v' = v$, and there is nothing to show. Hence, assume that $n > 0$. By Proposition 5.1.2, there exists a unique tuple $(u_0, u_1, \dots, u_n)$ with $u_i \in V_0$ for $0 \leq i < n$, and $u_n \in V(\Gamma)$. We claim that we can take $v' = u_n$.

First, note that v belongs to a copy $\Gamma_{u_{n-1}}$ of Γ with a labelled graph isomorphism $\phi : \Gamma_{u_{n-1}} \to \Gamma$ taking v to u_n, where ϕ is simply the identity mapping. This isomorphism hence extends to an end-isomorphism $\phi' : \Gamma_{u_{n-1}}(v) \to \Gamma(u_n)$. Since all walks in $\textsc{Tree}(\Gamma, S)$ from 1 to vertices $w \in V(\Gamma_{u_{n-1}})$ must pass through u_{n-1}, applying Proposition 5.1.2 to such w, the tuple corresponding to w must be $(u_0, u_1, \dots, u_{n-1}, w_n)$ for some $w_n \in V(\Gamma)$. In particular, if u' denotes the vertex corresponding to $(u_0, u_1, \dots, u_{n-1})$, then

$$|w|_{\textsc{Tree}(\Gamma,S)} = |u'|_{\textsc{Tree}(\Gamma,S)} + |w_n|_\Gamma.$$

In particular, our end-isomorphism ϕ' can be extended to an end-isomorphism ϕ'' between the embedded $\Gamma_{u_{n-1}}(v)$ and $\Gamma(u_n)$ inside the graph $\textsc{Tree}(\Gamma, S)$, where Γ is now identified with the canonically embedded copy of Γ inside $\textsc{Tree}(\Gamma, S)$, i.e. Γ_1.

We now extend this end-isomorphism one final time. If w_s is a vertex of the embedded $\Gamma_{u_{n-1}}(v)$ such that $\Gamma_{u_{n-1}}(w_s)$ is end-isomorphic to $\Gamma_{u_{n-1}}(x)$ for a vertex $x \in \phi^{-1}(V_S)$, then $\Gamma(\phi''(w_s))$ is end-isomorphic to $\Gamma(x')$ for some $x' \in V_S$, as S is a set of representatives of frontier points. This is a key step. In particular, the second-level subgraphs of w_s and $\phi''(w_s)$ are isomorphic in $\mathcal{T}$, as a copy of Γ is attached to both vertices, and ϕ'' is an end-isomorphism. If w_s is instead such that $\Gamma_{u_{n-1}}(w_s)$ is end-isomorphic to $\Gamma_{u_{n-1}}(x)$ for a vertex $x \notin \phi^{-1}(V_S)$, then again the second-level subgraphs of w_s and $\phi''(w_s)$ are isomorphic in $\mathcal{T}$. Thus ϕ'' extends to an end-isomorphism $\phi^* : \mathcal{T}(v) \to \mathcal{T}(u_n)$, and since certainly $u_n \in \mathcal{T}_0$, we have our claim.

To show the proposition it now suffices, as $\mathcal{T}_0 = \Gamma$ is context-free, to show that if $u, v \in V(\mathcal{T}_0)$, then $\mathcal{T}(u) \sim \mathcal{T}(v)$. But this follows from a near-identical argument to the final extension of the end-isomorphism above, as we only need to show that the end-isomorphism sends elements of S to S, and elements not in S to elements not in S. Thus every vertex in $\mathcal{T}$ has an end-isomorphism representative in $\mathcal{T}_0$, and there are only finitely many end-isomorphism classes of such vertices; hence $\mathcal{T}$ is context-free. □

We remark that this result could also be obtained by directly constructing a pushdown automaton for $\textsc{Tree}(\Gamma, S)$ from a pushdown automaton defining the graph Γ, by adding certain new transitions to, informally speaking, the states corresponding to the vertices of V_S. We will revisit this idea in the proof of Lemma 5.4.7, so do not expand on it here.

Corollary 5.1.6. *Let Γ be a finite graph, and let S be any set of vertices with $1 \not\in S$. Then* $\textsc{Tree}(\Gamma, S)$ *is context-free.*

Proof. The only verification needed is that S is a set of attachment points; indeed, if $V(\Gamma) = \{1, v_1, \dots, v_n\}$, then we can take as representatives of the end-isomorphism classes of Γ the set $\{\Gamma(1), \Gamma(v_1), \dots, \Gamma(v_n)\}$.[71] As v_i is a frontier point of $\Gamma(v_i)$, we have that $F(\Gamma)$, being the union of all frontier points of our choice of representatives, is all of $V(\Gamma)$. Thus $S \subseteq F(\Gamma)$ and $1 \not\in S$; so S is a set of attachment points. □

The tree of copies construction can be used to describe free algebraic structures, as we have hinted at in Example 5.1.3. The following corollary of Proposition 5.1.5 is nearly trivial to verify directly without using trees of copies, but it makes for a neat first application of the construction.

Corollary 5.1.7. *Let A be a finite set, and let A^* be the free monoid on A. Then the right Cayley graph $\Gamma_M(A^*, A)$ of A^* is context-free.*

Proof. As seen in Example 5.1.3, $\Gamma_M(A^*, A)$ is isomorphic to $\textsc{Tree}(\Gamma, S)$ where Γ is a finite graph, and where S is a set of attachment points. By Proposition 5.1.5, $\Gamma_M(A^*, A)$ is context-free (indeed, it has exactly two end-isomorphism classes!). □

Having understood these rather general constructions, we shall now begin an investigation into the structure of the right Cayley graphs of special monoids. It will turn out that in some cases, for a special monoid M, this can be realised as a tree of copies of a graph which is itself a tree of copies of a graph $\mathfrak{U}$; and this latter graph $\mathfrak{U}$ will be context-free if the group of units of M is a context-free group. This will reveal much geometric structure of special monoids.

[71] Of course, this set of representatives will likely contain some redundancies; for example, in Example 5.1.3 all n of the $\Gamma(v_i)$ so chosen were pairwise end-isomorphic! We have not, however, at any stage demanded of our representatives that they be unique, and hence this is no cause for concern.

5.2 The right units of a special monoid

Let $M = \mathrm{Mon}\langle A \mid w_i = 1 (i \in I)\rangle$ be an arbitrary finitely presented special monoid. Recall that $U(M)$ denotes the group of units of M. We will let $U_r(M)$ denote the submonoid consisting of all right invertible elements of M. Let Δ be, as usual, the set of minimal invertible pieces, and let I be the set of non-empty prefixes of elements of Δ. In particular $\Delta \subseteq I$. We set $I_c = I \setminus I^2$. Then I_c is a suffix code, i.e. $I_c \cap I_c A^+ = \varnothing$.[72] Note that $\Delta \subseteq I_c$. Set $\Pi = I_c \setminus \Delta$. Then Π is the set of proper prefixes of pieces which are not a product of two (or more) prefixes. We will at times make use of the following lemma, which we have essentially already mentioned in §1.3 as Proposition 1.3.2.

Lemma 5.2.1 (See [502, Proposition 2.3]). *Let $x, y \in A^*$, and let $u, v \in \Delta^*$ such that $u \xrightarrow{*}_S v$. If $xuy \in \Delta^*$, then $xvy \in \Delta^*$.*

We can use this to deduce the following lemma.

Lemma 5.2.2. *Suppose $\xi \in \Pi$ has a proper suffix equal to some y' with $y' \in I^*$ and $y' \in \mathrm{Irr}(S)$. Then there exists $\xi' \in \Pi$ such that $\xi' \equiv xy'$ with $x \in A^*$.*

Proof. Suppose $\xi \equiv xy$ with $y =_M y'$, and with $x \in A^+$. Then $y \xrightarrow{*}_{S(M)} y'$, as $y' \in \mathrm{Irr}(S)$. Let $z \in A^+$ be such that $\xi z \in \Delta$. Then $xyz \in \Delta$. By Lemma 5.2.1, as $xyz \in \Delta$, we have $xy'z \in \Delta^*$. But it is easy to verify that $xy'z$ has no proper non-empty invertible prefix, as xyz does not; thus $xy'z \in \Delta$. Hence $\xi' \equiv xy' \in \Pi$. □

As a corollary, we find the following, as subwords of irreducible words are irreducible.

Lemma 5.2.3. *Suppose $\xi \in \Pi$ is irreducible and has a proper suffix equal to some y' with $y' \in I^*$. Then $\xi \equiv xy'$ for some $x \in A^*$.*

Let $I_0 = \mathrm{Irr}(S) \cap I_c$. By [502, Lemma 4.1], for every right invertible word $u \in A^*$, there exists an element $v \in I_0^*$ with $u =_M v$. Partition $\Delta = \Delta_1 \cup \Delta_2 \cup \cdots \cup \Delta_\kappa$, where $\delta_1, \delta_2 \in \Delta$ are in the same set of the partition if and only if $\delta_1 =_M \delta_2$. By [502, Lemma 4.2], for every $1 \leq i \leq \kappa$, either there exists exactly one word in $I_0 \cap \Delta_i$, or else $\delta =_M 1$ for every $\delta \in \Delta_i$.[73] We then denote by Δ_0 the set $I_0 \cap \Delta$; if $\delta, \delta' \in \Delta_0$, then $\delta =_M \delta'$ if and only if $\delta \equiv \delta'$. Note further that for every $\delta \in \Delta_0$, we have $\delta \neq_M 1$. Let $\Pi_0 = I_0 \setminus \Delta_0$. Then $\Pi_0 = \Pi \cap \mathrm{Irr}(S)$, so Π_0 consists of the irreducible proper non-empty prefixes of words from Δ.

Example 5.2.4. Let $M = \mathrm{Mon}\langle a, c, \beta, p, q \mid (pq)^2 = 1, \beta pq = pq\beta = 1, a\beta pc = 1\rangle$. Then as β is an inverse of pq, it follows that $\beta =_M pq$. It is not hard, e.g. by using a finite complete

[72] This condition of being a suffix code is misprinted in Zhang [502]; he states that a set of words X is a suffix code if and only if $X \cap XA^* = \varnothing$, which is absurd; for *any* set of words X, we have $X \cap XA^* = X$.

[73] Note that [502, Lemma 4.2] by Zhang is false as stated. Indeed, in the proof of the same, Zhang states "from the definition of Δ_i, Δ_i is closed under the induction $\xrightarrow{*}_S$" [502, p. 502], which is only true if no word in Δ_i is equal to 1. Indeed his statement fails even for the easy case of the bicyclic monoid $\mathrm{Mon}\langle b, c \mid bc = 1\rangle$, where $\Delta = \Delta_1 = \{bc\}$. Clearly Δ_1 is *not* closed under $\xrightarrow{*}_S$, as $bc \xrightarrow{*}_S 1$. Our formulation is correct, as the remainder of Zhang's proof is unaffected.

rewriting system for M, to find

$$\Delta = \{\beta, pq, a\beta pc\}, \text{ whence}$$
$$I = \{\beta, p, a, a\beta, pq, a\beta p, a\beta pc\}.$$

We can partition the pieces, as usual, into equivalence classes with respect to $\xleftrightarrow{*}_M$ as

$$\Delta = \Delta_1 \cup \Delta_2 = \{\beta, pq\} \cup \{a\beta pc\}.$$

Now $I_c = I \setminus I^2 = \{\beta, p, a, pq, a\beta pc\}$; there is a rule $pq \xrightarrow{*}_S \beta$, so $I_0 = \{\beta, p, a, a\beta pc\}$. Then $\Delta_1 \cap I_0 = \{\beta\}$ and $\Delta_2 \cap I_0 = a\beta pc$. Thus we find

$$\Delta_0 = \Delta \cap I_0 = \{\beta, a\beta pc\},$$
$$\Pi_0 = I_0 \setminus \Delta_0 = \{p, a\}.$$

Thus, provided a solution to the word problem, finding Δ_0, Π_0, and I_0 is not difficult. △

Now by [502, Lemma 4.1], every right invertible word in M is equal to some element from $(\Delta_0 \cup \Pi_0)^* = I_0^*$. Hence, let $u \in I_0^*$ be arbitrary. Then we can *uniquely* factorise

$$u \equiv u_0\xi_0u_1\xi_1 \cdots u_k\xi_k$$

where $u_i \in \Delta_0^*$ for every $0 \leq i \leq k$, and $\xi_i \in \Pi_0$ for $0 \leq i < k$, and $\xi_k \in \Pi_0 \cup \{\varepsilon\}$. We say that this factorisation $u \equiv u_0\xi_0u_1\xi_1 \cdots u_k\xi_k$ is a *normal form* for the word u. Normal forms are not unique (as there can be many different representations of e.g. each u_i). However, they are close to being unique, in the following sense; the statement of the proposition a restatement of [502, Proposition 4.3 & Theorem 4.4].

Proposition 5.2.5. *Let $u \equiv u_0\xi_0u_1\xi_1 \cdots u_k\xi_k$ and $u' \equiv u_0'\xi_0'u_1'\xi_1' \cdots u_n'\xi_n'$ be normal forms. Then $u =_M u'$ if and only if $n = k$, $u_i =_M u_i'$, and $\xi_i \equiv \xi_i'$ for every $0 \leq i \leq k$.*

Note the graphical equality between the ξ_i, rather than equality in M. One consequence of this proposition is that given a right invertible word $u \in A^*$, and given two words $u', u'' \in I_0^*$ representing u, we have that u' and u'' have the same length as an alternating product of words from Δ_0 and Π_0.

Definition 5.2.6 (Depth of a right invertible word)**.** The *depth* $d(u)$ of a right invertible word $u \in A^*$ is the unique k such that $u =_M u_0\xi_0 \cdots u_k\xi_k$ for some $u_0\xi_0 \cdots u_k\xi_k \in I_0^*$.

Note that if M is a group, then $d(u) = 0$ for every $u \in A^*$. As the depth of a word as defined above is the same for all words equal to each other in M, and as every right invertible element of M has a normal form, we may also extend the depth function to $d\colon U_r(M) \to \mathbb{N}$ and speak of the depth of a right invertible element, which shall be convenient.

Stated otherwise, the depth $d(u)$ of the right invertible $u \in A^*$ is the smallest k such that u can be written as a product $u_0\xi_0 \cdots u_k\xi_k$ where $u_i \in \Delta^*$ and $\xi_i \in \Pi$. Note that – quite unlike the case of a normal form! – while in certain easy cases (such as the case when $I^2 \cap I = \varnothing$, which we shall call the *no-folding* case presently) there is only one k such that u can be written as a product $u_0\xi_0 \cdots u_k\xi_k$ where $u_i \in \Delta^*$ and $\xi_i \in \Pi^*$. However, this is not always the case; there may be many representations of different lengths even of a given word $u \in I^*$.

Example 5.2.7. Let $M = \text{Mon}\langle a, b \mid aab = 1\rangle$. Then $\Delta = \{aab\}, I = \{a, aa, aab\}$. In particular $U(M) \cong 1$. We have $\Delta_0 = \varnothing$ and $I_0 = \{a\}$. Then a^2 is right invertible, but we can write is as

$$a^2 \equiv (\varepsilon)(a)(\varepsilon)(a) \equiv u_0\xi_0u_1\xi_1$$
$$a^2 \equiv (\varepsilon)(a^2) \qquad \equiv u_0'\xi_0'$$

i.e. where $u_0 \equiv u_1 \equiv u_0' \equiv \varepsilon \in \Delta_0^*$, $\xi_0 \equiv \xi_1 \equiv a$, and $\xi_0' \equiv aa$. The first form is a normal form, whereas the second is not (as ξ_0' has a proper suffix in I); therefore $d(a^2) = 1$. △

To deal with this potential increase in complextiy which can result from writing a word in I^* as a word over I_0^*, we shall, informally speaking, bound this increase; or, stated otherwise, we shall introduce a bound on how long a representation of a given normal form can be in terms of the length of that normal form. We define the *prefix complexity constant* Ω of M to be $\max_{p\in\Pi} |p|$.

Example 5.2.8. Let $M = \text{Mon}\langle a, b, c \mid babcb = 1\rangle$. Then $\Delta = \{b, abc\}$ and $I = \{b, a, ab, abc\}$, so $\Pi = \{a, ab\}$ and $\Omega = \max_{p\in\Pi} |p| = \max\{|a|, |ab|\} = 2$. △

The reason for the name *complexity constant* are the following two results.

Lemma 5.2.9. *Let $u \equiv u_0\xi_0 \cdots u_k\xi_k \in I_0^*$, and let $a \in A$. Let $k' = \max(0, k - \Omega)$. If ua is right invertible, then a normal form of ua is*

$$(u_0\xi_0 \cdots u_{k'-1}\xi_{k'-1})u'$$

where $u' \in I_0^*$ is a normal form of $u_{k'}\xi_{k'} \cdots u_k\xi_k a$. Note that $d(u') \leq \Omega$.

Proof. As ua is right invertible, we can rewrite it to an element of I_0^* using S. We may assume without loss of generality that u is irreducible modulo S. Suppose then that $ua \xrightarrow{*}_S w$ for some $w \in I_0^*$. If a is not irreducible mod S, then either there exists some letter $b \in A$ such that b is irreducible mod S and $a =_M b$, or else $a =_M 1$. In the latter case, $u =_M ua$, in which case the statement is vacuously true.

Thus, assume a is irreducible. As u and a are irreducible, it follows by [505, Theorem 2.2(2)] that either $ua \equiv w$, or else $ua \to_S w$. If $ua \equiv w$, then $ua \in I_0^*$. Thus we can uniquely factor ua into words from I_0 from the left; it is then clear that either $u_k\xi_k a \in \Delta_0^*$, or else $\xi_k a \in \Pi_0$. In either case, the claim clearly holds. Suppose instead that $ua \to_S w$, say via an application of the rule $(\ell, r) \in S$. As u and a are irreducible, the specified occurrence of ℓ in ua must straddle the boundary between u and a; thus, we must have $ua \equiv v\ell$ for some $v \in A^*$. Suppose $\ell \equiv \ell_0\ell_1 \cdots \ell_s$ for some $s \geq 0$ with $\ell_i \in \Delta$ for all $0 \leq i \leq s$. Consider the specified occurrence of ℓ_0 in ua. As Δ is a biprefix code, we must have $\ell_0 \equiv \xi_i u_{i+1} \cdots u_j\xi_j$ for some $0 \leq i \leq j \leq k$, or else $\ell_0 \equiv \ell \equiv \xi_i u_{i+1}\xi_{i+1} \cdots u_k\xi_k a$. The former case is impossible, as then ξ_j is a proper suffix of the invertible word ℓ_0, and hence left invertible; but ξ_j is right invertible, so it is invertible, a contradiction to $\xi_j \in \Pi_0$.

Thus we have the latter case, i.e. $\ell_0 \equiv \xi_i u_{i+1}\xi_{i+1} \cdots u_k\xi_k a$. Now $\ell_0 \in \Delta$, so $|\ell_0| \leq \Omega + 1$. As $|\xi_j| \geq 1$ for all $0 \leq j \leq k$, we hence have

$$\Omega + 1 \geq |\ell_0| \geq \sum_{j=i}^{k} |\xi_k| \geq \sum_{j=i}^{k} 1 = k - i + 1.$$

Hence $k - i \leq \Omega$, so $i \geq k - \Omega$. As $i \geq 0$, we have $i \geq k'$. Hence, we have

$$w \equiv u_0\xi_0 u_1\xi_1 \cdots u_{k'-1}\xi_{k'-1}(u_{k'}\xi_{k'} \cdots u_{i-1}\xi_{i-1}r) \in I_0^*.$$

Now, factoring w from the left over the suffix code I_0, we hence find that a normal form for w begins with $u_0\xi_0 u_1\xi_1 \cdots u_{k'-1}\xi_{k'-1}$ and ends with a normal form for $u_{k'}\xi_{k'} \cdots u_{i-1}\xi_{i-1}r$; as we have

$$u_{k'}\xi_{k'} \cdots u_{i-1}\xi_{i-1}r =_M u_{k'}\xi_{k'} \cdots u_{i-1}\xi_{i-1}\ell \equiv u_{k'}\xi_{k'} \cdots u_k\xi_k a$$

this is what was to be shown.

□

Thus, informally speaking, given a normal form $u \in I_0^*$, then to understand ua for a single letter $a \in A$, it suffices to understand what right multiplication by a does to the last Ω components of the alternating product. This gives a significant degree of uniformity to the right invertible elements, which we shall exploit when describing the Schützenberger graph of 1. The following proposition is not distantly related.

Proposition 5.2.10. *Let* $u \equiv u_0\zeta_0 u_1\zeta_1 \cdots u_k\zeta_k$ *with* $u_i \in \Delta^*$ *and* $\zeta_i \in \Pi$ *for* $0 \leq i \leq k$*, except* $\zeta_k \in \{\varepsilon\} \cup \Pi$*. Then* u *has depth less than* $(k+1)\Omega$*.*

Proof. The proof is by induction on k. Suppose $k = 0$, and $u \equiv u_0\zeta_0$. Suppose the depth of u is $d(u) = n$, and let $v_0\xi_0 v_1\xi_1 \cdots v_n\xi_n$ be a normal form for u. It follows that $u_0 =_M v_0$, and so

$$\zeta_0 =_M \xi_0 v_1\xi_1 \cdots v_n\xi_n.$$

As $v_0\xi_0 v_1\xi_1 \cdots v_n\xi_n$ is a normal form, then it is obvious that the maximal invertible factors of this word are precisely the v_i for $0 \leq i \leq n$, together with the non-prefix non-suffix maximal invertible subwords of the ξ_i. It follows by the normal form lemma that ζ_0 contains a subword graphically equal to the first letter (and the last letter) of ξ_i for every $0 \leq i \leq n$, and these subwords are pairwise non-overlapping. As each of the ξ_i is non-empty, we have $|\xi_i| \geq 1$, so

$$n + 1 = (n+1) \cdot 1 \leq \sum_{i=0}^{n} |\xi_i| \leq |\zeta_0| \leq \max_{p \in \Pi} |p| = \Omega,$$

and we have $n \leq (0+1)\Omega$, as desired. Suppose for induction that the claim is true for all values of k less than ℓ for some $\ell > 0$, and suppose we have $k = \ell$. Let $v \equiv v_0\xi_0 v_1\xi_1 \cdots v_n\xi_n$ be a normal form for u. Clearly, if

$$v' \equiv v_0'\xi_0' \cdots v_{n'}'\xi_{n'}'$$

is a normal form for $u_0\zeta_0 \cdots u_{k-1}\zeta_{k-1}$ and $\zeta_{k-1} \not\equiv \varepsilon$, and

$$v'' \equiv v_0''\xi_0'' \cdots v_{n''}''\xi_{n''}'$$

is a normal form for $u_k\zeta_k$, then $v'v''$ is a normal form for u. Hence $n = n' + n''$. By the inductive hypothesis, $n' \leq k\Omega$ and $n'' \leq \Omega$. Thus $n \leq (k+1)\Omega$, and we are done by induction. □

In particular, given some $u \equiv u_0 \in \Delta^*$, there is no product of the form $v \equiv v_0\zeta_0 v_1\zeta_1 \cdots v_n\zeta_n$ with $u_i \in \Delta^*$ and $\zeta_j \in \Pi$ with $n > \Omega$ such that $u =_M v$. This can be viewed as a type of distortion measure for the embedding of the group of units inside M. In particular, the map which maps an element of $U(M)$ to its normal form is a quasi-isometric embedding of $U(M)$ into M. We shall now seek to understand the graphical representation of the group of units.

5.3 The Schützenberger graph of the units

Throughout this section, fix a finitely presented special monoid $M = \text{Mon}\langle A \mid w_i = 1\,(i \in I)\rangle$. We will make some minor modifications to our notation used in e.g. Chapter 3, for reasons that will become clear. First, let $\pi\colon A^* \to M$ denote the canonical surjective homomorphism. Let $B = \{\beta_{\delta_1}, \beta_{\delta_2}, \dots\}$ be a (finite) set of symbols in bijective correspondence with the set of pieces Δ of M via the bijection $\phi_B\colon \delta_i \mapsto \beta_{\delta_i}$. For a relation word w_i, define the relations

$$B_c(w_i) = \{(W = 1) \mid W \text{ is a cyclic permutation of } \phi_B(w_i)\}.$$

Thus, for the easy example of $M = \text{Mon}\langle a, b \mid abca = 1\rangle$, we have, by Adian's algorithm, $\Delta = \{a, bc\}$, and if $B = \{\beta_a, \beta_{bc}\}$, then $B_c(abca) = \{\beta_a\beta_a\beta_{bc}, \beta_a\beta_{bc}\beta_a, \beta_{bc}\beta_a\beta_a\}$. Let $U_B(M)$ denote the monoid defined by the monoid presentation

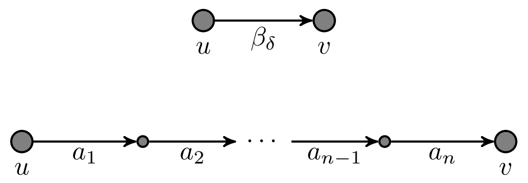
$$\text{Mon}\langle B \mid \{\beta_{\delta_i} = \beta_{\delta_j} \text{ whenever } \delta_i =_M \delta_j \text{ for some } 1 \leq i < j\} \cup \bigcup_{i \in I} B_c(w_i)\rangle.$$

It follows from §1.3 that $U_B(M) \cong U(M)$. We note that we can easily change the above monoid presentation to a special monoid presentation, as every β_{δ_i} is invertible in $U_B(M)$. Let $\Gamma_M(U_B(M), B)$ denote the right monoid Cayley graph of $U(M)$ with respect to the generating set B, and let $\Gamma_G(U_B(M), B)$ denote the group Cayley graph of $U(M)$ with respect to the (group) generating set B. These two graphs are very closely related. Indeed, recall that the graph lud(Γ) is obtained from the labelled graph Γ by replacing every edge $u \xrightarrow{a} v$ by the two edges $u \xrightarrow{a} v$ and $v \xrightarrow{\overline{a}} u$, where $a \in A$ for the label alphabet A of Γ, and $\overline{A}$ is a set in bijective correspondence with A via $a \mapsto \overline{a}$ and with $\overline{A} \cap A = \varnothing$. We have the following general and essentially obvious claim, by identifying every label $\overline{a}$ with the label a^{-1}:

Proposition 5.3.1. *Let G be a group generated as a monoid by the set A. Then there is an isomorphism of labelled, rooted graphs* lud$(\Gamma_M(G, A)) \cong \Gamma_G(G, A)$.

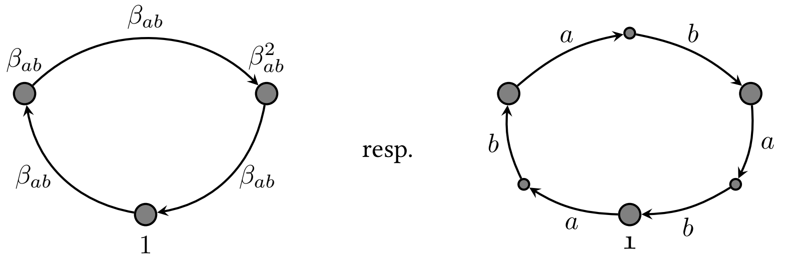
We shall now work towards defining a new graph $\mathfrak{U}$ from the monoid Cayley graph $\Gamma_M(U_B(M), B)$, and we will do so in a way to ensure the existence of a properly discontinuous and co-compact action of $U(M)$ on $\mathfrak{U}$. By a straightforward application of the Svarč-Milnor lemma, we will thus be able to conclude that the graph $\mathfrak{U}$ is quasi-isometric to the Cayley graph of $U(M)$ (as undirected graphs). Along the way, we will also capture many of the algebraic properties of $U(M)$ in the algebraic properties of $\mathfrak{U}$. This graph $\mathfrak{U}$ will be called the *Schützenberger graph of the units of* M, for reasons that will become clear.

Starting with $\Gamma_M(U_B(M), B)$, we first replace every edge labelled β_δ for $\delta \in \Delta$ by a directed path of length $n = |\delta|$ with path label δ, as indicated below (the reason for the relative sizes of the vertices shall be made clear presently):



$(\delta \equiv a_1 a_2 \cdots a_n)$

We denote the resulting graph $\mathfrak{U}_0$, root it at the vertex $\mathbb{1}$ naturally corresponding to 1, and note that its labelling alphabet is now A. For example, if $M = \text{Mon}\langle a, b \mid (ab)^3 = 1\rangle$, then $U_B(M) = \text{Mon}\langle \beta_{ab} \mid \beta_{ab}^3 = 1\rangle$, and the graphs $\Gamma_M(U_B(M), B)$ resp. $\mathfrak{U}_0$ for this example are:



Some subset of the vertices of $\mathfrak{U}_0$ will have been present already in $\Gamma_M(U_B(M), B)$ – we call this subset the *locally invertible* vertices. In the above example of the monoid $M = \text{Mon}\langle a, b \mid (ab)^3 = 1\rangle$, the three enlarged vertices are the locally invertible vertices. Every vertex of $\mathfrak{U}_0$ is locally invertible if and only if every piece in Δ has length 1, i.e. if and only if M is a free product of a free monoid by a group. For every vertex v of $\mathfrak{U}_0$, there exists a unique locally invertible vertex $\widetilde{v}$ such that v is reachable from $\widetilde{v}$ by a path with path label in $\Pi = I \setminus \Delta$. This follows directly from the fact that no non-empty prefix of an element of Δ appears as a proper prefix of another element of Δ. We then write $i(v) := \widetilde{v}$, letting $i : V(\mathfrak{U}_0) \to V(\mathfrak{U}_0)$ denote the function which, on input $v \in V(\mathfrak{U}_0)$, returns the *locally invertible vertex associated to* v. Note that the prescribed path from $i(v)$ to v is unique; that is, if one considers (slightly informally) the image of the out-neighbourhood of a single vertex of $\Gamma_M(U_B(M), B)$ in $\mathfrak{U}_0$, then the resulting graph is a directed tree rooted at the image of the chosen vertex, and every other vertex in this tree can be reached by a unique path from the root.[74]

Recall that Π is the set of non-empty prefixes of elements from Δ. Denote by Π_ε the set $\Pi \cup \{\varepsilon\}$. Let similarly Δ_ε denote $\Delta \cup \{\varepsilon\}$. Then the set of vertices of $\mathfrak{U}_0$ is in bijective correspondence with the set of all triples

$$(m, \xi, \delta) \in U(M) \times \Pi_\varepsilon \times \Delta_\varepsilon,$$

with the property that ξ is a proper prefix of δ, via the bijection: a vertex v of $\mathfrak{U}_0$ is mapped to the triple $(i_m(v), \xi, \delta)$, where $i_m(v)$ is the vertex of $\Gamma_M(U_B(M), B)$ corresponding uniquely to the locally invertible vertex $i(v)$; and ξ is the path label of the unique path from $i(v)$ to v; and δ is either ε, if $i(v) = v$, or else is the unique piece which, when subdivided, gave rise to the path of which the path from $i(v)$ to v is an initial segment. The locally invertible vertices are then precisely those associated to triples of the form $(m, \varepsilon, \varepsilon)$. We will henceforth not make a distinction between a vertex and the triple associated to it. We remark that the information contained in ξ and δ cannot be condensed into just providing ξ, as ξ may be a prefix of many different pieces. We refer the reader to Figure 5.5 for an example of this bijection.

As $U(M)$ acts vertex-transitively on $\Gamma_M(U_B(M), B)$, it clearly acts also on $\mathfrak{U}_0$ via the action $m.(m', \xi, \delta) = (mm', \xi, \delta)$. This action is vertex-transitive on the locally invertible vertices. We will now define an equivalence relation $\sim_M$ on $\mathfrak{U}_0$. This equivalence relation

[74] Such a graph is called an *arborescence* by Tutte [480].

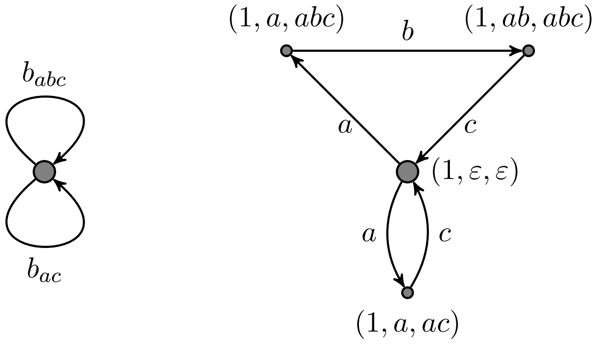


Figure 5.5: Left: The Cayley graph $\Gamma_M(U_B(M), B)$ when $M = \mathrm{Mon}\langle a, b, c \mid abc = 1, ac = 1\rangle$. Right: $\mathfrak{U}_0$ of the same example.

will, informally speaking, capture the fact that some vertices of $\mathfrak{U}_0$ represent "the same elements of M"; indeed, one may initially note that the locally invertible vertices can be considered as representing distinct elements of $U(M)$. We define $\sim_M$ on the vertices of $\mathfrak{U}_0$ by

$$(m_1, \xi, \xi') \sim_M (m_2, \zeta, \zeta') \iff m_1 \cdot \pi(\xi) = m_2 \cdot \pi(\zeta)$$

Note, for example, that in Figure 5.5, we have that the vertices labelled by $(1, a, abc)$ and $(1, a, ac)$ satisfy $(1, a, abc) \sim_M (1, a, ac)$.

It turns out that $\sim_M$-equivalence is highly controlled and is uniform throughout $\mathfrak{U}_0$. In order to apply rewriting techniques to demonstrate this, we will introduce a piece of notation for the vertices of $\mathfrak{U}_0$. If $u = (m_1, \xi, \delta) \in V(\mathfrak{U}_0)$, then let $u_0 \in \Delta^*$ be any word over the (presentation) pieces such that $\pi(u_0) = m_1$. We then say that a *literal form* of u is the triple (u_0, ξ, δ). Of course, literal forms are not unique (there are, in general, infinitely many words over Δ^* representing a given invertible element). Note that any literal form of a vertex is an element of $A^* \times A^* \times A^*$, whereas the vertices of $\mathfrak{U}_0$ are elements of $M \times A^* \times A^*$. If $u, v \in V(\mathfrak{U}_0)$ have literal forms (u_0, ξ, δ) and (v_0, ζ, δ'), respectively, then we have that

$$u \sim_M v \qquad \iff \qquad u_0\xi =_M v_0\zeta.$$

The following proposition shows that to consider $\sim_M$-equivalence of two triples is to consider equality in $U(M)$ and equality of the second element of the triples in M. In particular, as we shall see, the third element of the triple becomes entirely redundant in the context of $\sim_M$.

Proposition 5.3.2. *Let $u, v \in V(\mathfrak{U}_0)$ with $u = (m_1, \xi, \delta_1)$ and $v = (m_2, \zeta, \delta_2)$. Then*

$$u \sim_M v \iff m_1 = m_2 \quad \text{and} \quad \xi =_M \zeta$$

In particular, if u and v are distinct vertices such that $u \sim_M v$, then neither u nor v are locally invertible.

Proof. Let u, v have literal forms (u_0, ξ, δ_1) and (v_0, ζ, δ_2), respectively. We will begin by proving the biconditional.

($\Longleftarrow$) If $m_1 = m_2$, then $u_0 =_M v_0$. Consequently if also $\xi =_M \zeta$, then we must have $u_0\xi =_M v_0\zeta$. Hence $u \sim_M v$.

($\Longrightarrow$) We make use of the normal form lemma, i.e. Lemma 1.3.8. Assume for the contrapositive that $u_0 \neq_M v_0$ or $\xi \neq_M \zeta$, and assume for contradiction that $u_0\xi =_M v_0\zeta$. Let

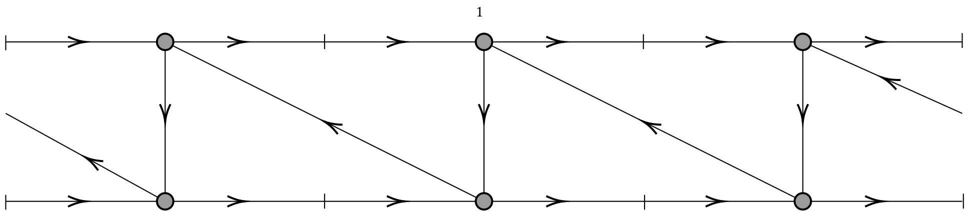


Figure 5.6: A portion of the graph $\mathfrak{U}'$ for the monoid $M = \mathrm{Mon}\langle a, b, c \mid a(bc)a = 1\rangle$. Note that $U(M) \cong \mathbb{Z}$. In the graph, horizontal movement corresponds to reading *bc*, whereas vertical and diagonal movement corresponds to reading a. The gray and enlarged vertices are the locally invertible vertices, and the small vertical dashes are the non-locally invertible vertices.

v_0^{-1} be any word representing the inverse of m_2 in M. Then $v_0^{-1}u_0\xi =_M \zeta$. Now no non-empty maximal invertible factor of $v_0^{-1}u_0\xi$ can overlap with ξ, as otherwise some non-empty prefix of ξ would be left invertible and hence also invertible, contradicting the minimality of elements of Δ. Hence any factorisation of $u_0\xi$ into maximal invertible factors must contain $v_0^{-1}u_0$ as the first maximal invertible factor. Analogously, any factorisation of ζ must begin with an empty maximal invertible factor. Thus, by the normal form lemma, we have that $v_0^{-1}u_0 =_M 1$, i.e. $u_0 =_M v_0$. Hence, by our assumption, we must have had that $\xi \neq_M \zeta$. But

$$\xi =_M 1 \cdot \xi = v_0^{-1}u_0\xi =_M v_0^{-1}v_0\zeta =_M \zeta,$$

a contradiction. This proves the biconditional.

We now prove the particular consequence. Let u, v be two distinct vertices with $u \sim_M v$. If one of u and v is locally invertible, say (without loss of generality) u, then $\xi \equiv \varepsilon$, so $\xi =_M 1$. Thus $\zeta =_M 1$ by the above biconditional. In particular ζ is invertible. As ζ is a proper prefix of δ_2 and as δ_2 is a minimal word or empty, it follows that ζ is necessarily empty; hence v is locally invertible. By symmetry, u is locally invertible if and only if v is. Suppose then that u and v are distinct and $u \sim_M v$; if either of the two were locally invertible, then both would be, in which case $u = (m_1, \varepsilon, \varepsilon) = (m_2, \varepsilon, \varepsilon) = v$, which is a contradiction to the fact that u and v are distinct. Thus neither of u, v are locally invertible. □

We set $\mathfrak{U}' := \mathfrak{U}_0 / \sim_M$. The vertices of $\mathfrak{U}'$ are thus equivalence classes under $\sim_M$ of pairs (m, ξ) for $m \in U(M)$ and $\xi \in \Pi_\varepsilon$ (proper prefixes of pieces), where we have an equivalence $(m_1, \xi) \sim_M (m_2, \zeta)$ if and only if $m_1\pi(\xi) = m_2\pi(\zeta)$, which by Proposition 5.3.2 is equivalent to $m_1 = m_2$ and $\pi(\xi) = \pi(\zeta)$. We will denote vertices of $\mathfrak{U}'$ using the notation $[m, \xi] := [(m, \xi)]_{\sim_M}$ for notational brevity. Again, as in the case of $\mathfrak{U}_0$, to enable the use of rewriting techniques, we will define a *literal form* of a vertex $[m, \xi]$ to be an expression of the form $[u_0; \xi]$, where $u_0 \in \Delta^*$ is any word such that $\pi(u_0) = m$. The use of a semicolon in writing the pair $[u_0; \xi]$ is to emphasise that it is not the same object as $[m, \xi]$. Indeed, note that $u_0 \in \Delta^* \subseteq A^*$ while $m \in U(M) \subseteq M$.

Under stronger assumptions on the presentations involved, the graph $\mathfrak{U}'$ would already be an induced subgraph of the Cayley graph of M via the map $[m, \xi] \mapsto m \cdot \pi(\xi)$.[75] In general,

[75] This happens, for example, if the set of proper non-empty prefixes Π of pieces forms a code as a subset of A^*. This is the "no-folding case".

however, certain edges will be "missing", in the sense that although the aforementioned map is injective, it need not be faithful.[76] The goal of this subsection is then to describe what edges are "missing" from $\mathfrak{U}'$ in this sense, and that these are evenly distributed across $\mathfrak{U}'$. In fact, this description is rather simple, and follows from the normal form lemma.

Proposition 5.3.3. *Let $u_0, v_0 \in A^*$ be invertible, and ξ, ζ be proper prefixes of pieces. Let $a \in A$ be a letter. Then $(u_0\xi) \cdot a =_M v_0\zeta$ if and only if exactly one of the following occurs:*

(1) $\zeta \equiv \varepsilon$, there is some piece $\delta \in \Delta$ such that $\xi \cdot a =_M \delta$, and $u_0 \cdot \delta =_M v_0$.

(2) $\zeta \not\equiv \varepsilon$, $\xi \cdot a =_M \zeta$, and $u_0 =_M v_0$.

Proof. ($\Longleftarrow$) In either case (1) or (2), it is clear that $(u_0\xi) \cdot a =_M v_0\zeta$.

($\Longrightarrow$) We know that we have the equality

$$u_0\xi a =_M v_0\zeta$$

and, seeking to use the normal form lemma (Lemma 1.3.8), we will first detect the maximal invertible factors. The left-most maximal invertible factor in the right-hand side is v_0, as if this invertible factor were to extend into ζ, then some prefix p of ζ will be a suffix of an invertible factor. If this prefix p were empty, then $\zeta \equiv \varepsilon$, and we have that both sides are invertible; thus we are in case (1) above after observing that an overlap argument yields that $\xi \cdot a$ cannot possibly be congruent to a product of more than one piece. Now, if this prefix p were instead non-empty, then by another overlap argument we have a contradiction. Thus, in this case, the left-most maximal invertible factor of the right-hand side must be v_0. As a consequence, ζ is not invertible, and in particular $\zeta \not\equiv \varepsilon$; as ζ is a non-empty prefix of a piece, it follows that $v_0\zeta$ is not invertible, so neither is $u_0\xi a$. Applying the same argument to $u_0\xi a$, we see that u_0 must thus be the left-most maximal invertible factor of $u_0\xi a$, using the fact that $u_0\xi a$ is not invertible. By the normal form lemma, we hence have $u_0 =_M v_0$ and $\xi \cdot a =_M \zeta$, and we are in case (2). □

Definition 5.3.4. [The Schützenberger graph of the units] We define $\mathfrak{U} = (V(\mathfrak{U}), E(\mathfrak{U}))$ to be the (labelled, directed) graph obtained from $\mathfrak{U}'$ in the following way: if there exist prefixes $\xi, \zeta \in \Pi_\varepsilon$ and $a \in A$ such that $\xi \cdot a =_M \zeta$, then for all $m \in U(M)$ we add an edge $[m, \xi] \xrightarrow{a} [m, \zeta]$ if one does not already exist. We call $\mathfrak{U}$ the *Schützenberger graph of the units of M*.

Note that such an edge as is added to $\mathfrak{U}'$ in the above definition exists for *some* $m \in U(M)$, if and only if it exists for *any* $m \in U(M)$, by case (2) of Proposition 5.3.3, and is hence added in all these cases. Furthermore, if ξ and ζ are both empty, then a is congruent to a piece δ. Hence, as a is a single letter we have $a \in \Delta$, so as $u_0 \cdot a =_M v_0$ it follows that there is already an edge

[76] We remark that such behaviour ultimately comes down to being a consequence of the fact that whereas $\Delta\Pi \cap \Pi = \varnothing$ by overlaps, it might happen that $\Pi\Delta \cap \Pi \neq \varnothing$. That is, there can be pieces appearing as proper subwords of other pieces; this situation, which caused such difficulties in Chapter 3, again rears its ugly head. It would seem that most of the difficulties arising for special monoids is caused by this behaviour, which reiterates just how tantalising is the lack of any counterexample to the possibility that every special monoid admits a presentation with no pieces appearing as proper subwords of other pieces. If there were no such counterexample, then all proofs in this chapter regarding special monoids would be quite significantly simplified.

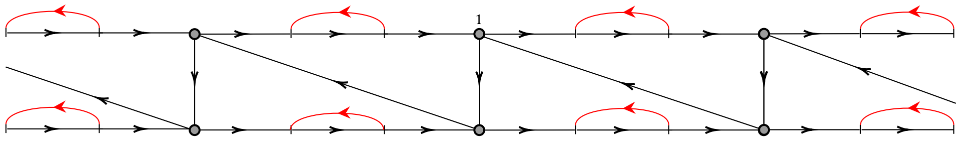


Figure 5.7: An example of the missing edges added by Definition 5.3.4. This particular example comes from the monoid $M = \mathrm{Mon}\langle a, b, c, d \mid b(abc)b = 1, bd = 1\rangle$. Horizontal movement corresponds to reading abc, whereas vertical and diagonal movement both correspond to reading b. The added red edges correspond to reading d. We note that we have omitted adding the loops corresponding to $bd = 1$ to each locally invertible vertex for ease of drawing.

$[u_0; \xi] \xrightarrow{a} [v_0; \zeta]$. Thus the added edges always either originate in or terminate in a non-locally invertible vertex. We illustrate this in Figure 5.7.

For the following proposition, we recall the definition of $\mathfrak{R}_1$ as the connected component of the identity element in $\Gamma_M(M, A)$, and that a graph homomorphism $\phi\colon G \to H$ is *faithful* if $\phi(G)$ is an induced subgraph of H. Note that a faithful injective graph homomorphism is equivalent to a *full* injective graph homomorphism, i.e. an injective graph homomorphism in which $\phi(u)$ is adjacent to $\phi(v)$ if and only if u is adjacent to v. Furthermore, we recall that Π is the set of proper prefixes of elements of Δ, and that $\Pi_\varepsilon = \Pi \cup \{\varepsilon\}$.

Proposition 5.3.5. *The map*

$$\begin{aligned}\phi : V(\mathfrak{U}) &\to \{m \cdot \pi(\xi) \mid m \in U(M), \xi \in \Pi_\varepsilon\} \subseteq V(\mathfrak{R}_1)\\ [m, \xi] &\mapsto m \cdot \pi(\xi)\end{aligned}$$

extends to a faithful injective labelled graph homomorphism $\phi : \mathfrak{U} \hookrightarrow \mathfrak{R}_1$. In other words, $\mathfrak{U}$ is isomorphic to the subgraph of $\mathfrak{R}_1$ induced on the set of vertices of the form $m \cdot \pi(\xi)$, where $m \in U(M)$ and $\xi \in \Pi_\varepsilon$.

Proof. We set $V_0 := \{m \cdot \pi(\xi) \mid m \in U(M), \xi \in \Pi_\varepsilon\}$. It is clear that ϕ is injective on the set of vertices, by definition of $\mathfrak{U}'$ as a quotient of $\mathfrak{U}_0$ modulo $\sim_M$-equivalence; for, $\phi([m, \xi]) = \phi([n, \zeta])$ implies $m \cdot \pi(\xi) =_M n \cdot \pi(\zeta)$, which is saying $(m, \xi) \sim_M (n, \zeta)$, i.e. $[m, \xi] = [n, \zeta]$.

We first show ϕ is injective on the set of edges. Let $u = [m_1, \xi]$ and $v = [m_2, \zeta]$ be arbitrary vertices of $\mathfrak{U}$. If $e = (u, a, v)$ is an edge of $\mathfrak{U}$ then we will show that there is an edge in $\mathfrak{R}_1$ from $\phi(u)$ to $\phi(v)$ labelled by a, i.e. that $\phi(u) \cdot \pi(a) = \phi(v)$. Hence assume $e = (u, a, v)$ is an edge of $\mathfrak{U}$. Suppose u and v have literal forms $[u_0; \xi]$ and $[v_0; \zeta]$, respectively. Then by definition $\phi(u) = \pi(u_0\xi)$ and $\phi(v) = \pi(v_0\zeta)$. Since e is an edge of $\mathfrak{U}$ we must have that either ξ and ζ are prefixes of some piece such that $\xi a \equiv \zeta$, in which case $u_0 =_M v_0$, or a is the final letter of a piece ξa, in which case $v_0 =_M u_0\xi \cdot a$. In either case, we have

$$u_0\xi \cdot a =_M v_0\zeta$$

whence $\phi(u) \cdot \pi(a) = \phi(v)$, and we conclude that ϕ is an injective graph homomorphism.

To see that it is faithful, it suffices to note that if $\pi(u_0\xi) \xrightarrow{a} \pi(v_0\zeta)$ is an edge of $\mathfrak{R}_1$, then we are in exactly one of the cases (1) or (2) of Proposition 5.3.3. In case (1), then the edge $[m_1, \xi] \xrightarrow{a} [m_2, \zeta]$ is clearly present in $\mathfrak{U}'$ by construction of $\mathfrak{U}'$ from $\mathfrak{U}_0$, and hence is also an edge of $\mathfrak{U}$. In case (2), we have that $[m_1, \xi] \xrightarrow{a} [m_2, \zeta]$ is an edge added to $\mathfrak{U}$ by Definition 5.3.4. Thus, in both cases ϕ is faithful, and hence ϕ has all the desired properties. □

As a convention, we will henceforth refer to elements of $V(\mathfrak{U})$ by specifying literal forms

$[u_0; \xi]$, where $u_0 \in \Delta^*$ and $\xi \in \Pi_\varepsilon$. The construction of $\mathfrak{U}$ guarantees that this specifies precisely as much information as a pair $[m, \xi]$, where $m \in U(M)$. We will now turn towards studying the properties of $\mathfrak{U}$ as a graph with respect to the properties of the group of units $U(M)$. In particular, we will see that the context-freeness of $\mathfrak{U}$ is closely connected to that of $U(M)$.

We will investigate the context-free properties of $\mathfrak{U}$. To begin, we will use the strong symmetry exhibited by $\mathfrak{U}$ with regards to the group action; this will allow us to borrow results from group theory and apply them to this monoid setting. The group $U(M)$ acts properly discontinuously and transitively via left multiplication on $\Gamma_M(U_B(M), B)$, and this action, as we have already noted, hence extends to a properly discontinuous action of $U(M)$ on $\mathfrak{U}_0$. Explicitly, this action is defined for $m \in U(M)$ by setting $m.[m', \xi] = [mm', \xi]$ for $[m', \xi] \in V(\mathfrak{U})$. As $(m', \xi') \sim_M (m'', \xi'')$ only if $m' =_M m''$, in which case by the second part of Proposition 5.3.2 this identification commutes with the action of $U(M)$; and since the addition of edges to $\mathfrak{U}$ done by Definition 5.3.4 also commutes with this action by the discussion following Definition 5.3.4, we hence have that the above action is a well-defined, properly discontinuous action of $U(M)$ on $\mathfrak{U}$. By considering the associated undirected and unlabelled graphs $\mathrm{ud}(\mathfrak{U})$ and $\mathrm{ud}(\Gamma_M(U_B(M), B))$, we hence obtain the following by the Svarč-Milnor lemma.

Lemma 5.3.6. *There exists a quasi-isometry* $\mathrm{ud}(\Gamma_M(U_B(M), B)) \xrightarrow{\sim} \mathrm{ud}(\mathfrak{U})$. *That is, the group of units of M is quasi-isometric to $\mathfrak{U}$.*

If $R(M)$ denotes the submonoid of right invertible elements of M, then the lemma above is well worth comparing with the following due to Garreta & Gray [157]; this sheds some light on the rôle played by $\mathfrak{U}$ in the context of M.

Lemma 5.3.7 ([157])**.** *There exists a quasi-isometry* $\mathrm{ud}(\Gamma_M(R(M), I_0)) \xrightarrow{\sim} \mathrm{ud}(\mathfrak{R}_1)$. *That is, the submonoid of right units of M is quasi-isometric to $\mathfrak{R}_1$.*

We can now state the central theorem about the graph $\mathfrak{U}$. If $\mathfrak{U}$ were replaced with the right Cayley graph of a group, then the equivalence is already well-known. On the other hand, there is generally no reason to expect that some similar statement need be true for general right Cayley graphs of monoids.

Theorem 5.3.8. *Let M be a finitely presented special monoid, with group of units $U(M)$. Let $\Gamma_M(U_B(M), B)$ denote the right Cayley graph of $U_B(M) \cong U(M)$, and let $\mathfrak{U}$ be the Schützenberger graph of the units of M. Then the following are equivalent:*

(1) $U(M)$ *is virtually free.*
(2) $\Gamma_M(U_B(M), B)$ *is a context-free graph.*
(3) $\mathrm{ud}(\Gamma_M(U_B(M), B))$ *is quasi-isometric to a tree.*
(4) $\mathrm{ud}(\mathfrak{U})$ *is quasi-isometric to a tree.*
(5) $\mathrm{ud}(\mathfrak{U})$ *has finite tree-width.*
(6) $\mathfrak{U}$ *is a context-free graph.*

Proof. Let $M = \text{Mon}\langle A \mid R_i = 1 \ (i \in I)\rangle$ be a finitely presented special monoid. We begin by noting that it follows from e.g. [502] that the group of units of a finitely presented special monoid is finitely generated, so we will throughout the proof assume that $U(M)$ is finitely generated.

($1 \iff 2$) A finitely generated group G is virtually free if and only if $\Gamma_G(G, S)$ is a context-free graph for some (any) finite set S which generates G as a group; this is precisely the statement of Theorem 2.9 of [363]. Thus, when additionally S generates G as a monoid, then by Proposition 5.3.1, $\Gamma_G(G, S)$ is context-free if and only if $\Gamma_M(G, S)$ is context-free. Since B generates $U(M)$ as a monoid, we are done.

($1 \iff 3$) A group is virtually free if and only if its group Cayley graph with respect to (any) finite generating set is quasi-isometric to a tree; the forward implication is obvious, and the converse is well-known, e.g. as Theorem 7.19 of [161]. Clearly, the undirected monoid Cayley graph of a group is quasi-isometric to its group Cayley graph, and we have the equivalence.[77]

($3 \implies 4$) Immediate by Lemma 5.3.6.

($4 \implies 5$) We begin with a well-known proposition, which will serve as a black box: let Γ and Γ' be graphs such that Γ has finite tree-width and such that the degrees of Γ and Γ' are both bounded by some constant d. Then Proposition 5.17 of [139] says that if Γ' is quasi-isometric to Γ, then Γ' has finite tree-width too. Hence if ud($\mathfrak{U}$) is quasi-isometric to a tree, then as ud($\mathfrak{U}$) has bounded degree by finiteness of Δ, and as trees clearly have finite tree-width, we may use the aforementioned proposition to conclude that ud($\mathfrak{U}$) has finite tree-width.

($5 \implies 3$) Assume that ud($\mathfrak{U}$) has finite tree-width. Then as ud($\mathfrak{U}$) has bounded degree and is quasi-isometric to ud($\Gamma_M(U_B(M), B)$), we have that ud($\Gamma_M(U_B(M), B)$) also has finite tree-width, again by the aforementioned Proposition. But ud($\Gamma_M(U_B(M), B)$) is a vertex-transitive connected graph of bounded degree. Hence by Theorem 3.10 of [269] we have that $\Gamma_M(U_B(M), B))$ is context-free.

($5 \iff 6$) By Theorem 3.10 of [269], if Γ is a connected graph of bounded degree such that Aut(Γ) has only finitely many orbits on Γ, then ud(Γ) having finite tree-width is equivalent to Γ being context-free. Hence it suffices to show that $\mathfrak{U}$ has only finitely many orbits under automorphism, as $\mathfrak{U}$ is connected and has bounded degree since Δ is finite. But a co-compact (and hence co-finite) action of $U(M)$ by automorphisms on $\mathfrak{U}$ was constructed earlier when proving Lemma 5.3.6, and we are done. □

Structurally, $\mathfrak{U}$ has a rather close connection with $\mathscr{H}_1$, the Green's $\mathscr{H}$-class of the identity element. First, note that $\mathscr{H}_1 = U(M)$, and thus taking the subgraph of $\Gamma_M(M, A)$ induced on $\mathscr{H}_1$ will in general produce a disconnected graph. To remedy this, for a word $w \in A^*$, let Γ_w denote the subgraph of $\Gamma_M(M, A)$ induced on the set of vertices $\{\pi(w') \mid \exists w'' \in A^*, w \equiv w'w''\}$. For example, if $v \in A^*$ is a word with $\pi(v) = 1$, then Γ_v is a walk from 1 to 1, visiting a number of $\mathscr{H}$-classes inside the $\mathscr{R}$-class of 1. For a right invertible word $v \in A^*$, let H_v (note that this is distinct from $\mathscr{H}_v$) denote the set of all $\mathscr{H}$-classes visited by Γ_v. Then, by a

[77] It may be noted that any context-free graph is quasi-isometric to a tree, i.e. we always have the implication ($6 \implies 4$); this statement appears at the end of the proof of Lemma 8.4 of [106]. Thus, we also have ($6 \implies 5$) for any graph with bounded degree.

standard application of Green's Lemma via the left action of $U(M)$ on $\mathscr{R}_1$, it follows that $\mathfrak{U}$ is isomorphic to $\bigcup_{\delta\in\Delta} H_\delta$. We note further that Δ is a finite set, and that $\bigcup_{\delta\in\Delta} H_\delta \cong \bigcup_{i=1}^{|\Delta|} \mathscr{H}_1$, yielding another proof of Lemma 5.3.6. Of course, this proof does not give any insight into the structure of $\mathfrak{U}$, which we shall need presently.

We now give one more lemma regarding $\mathfrak{U}$ which shall come in handy.

Lemma 5.3.9. *For all $w_1, w_2 \in V(\mathfrak{U})$ we have*

$$\mathfrak{U}(w_1) \sim \mathfrak{U}(w_2) \implies (w_1 \text{ locally invertible } \iff w_2 \text{ locally invertible})$$

In particular any set $F = F(\mathfrak{U})$ consisting of representatives of all frontier points of $\mathfrak{U}$ can be written as a union $I_F \cup N_F$ of the locally invertible and not locally invertible representatives, respectively, with $I_F \cap N_F = \varnothing$.

Proof. Assume that $\mathfrak{U}(w_1) \sim \mathfrak{U}(w_2)$. Assume that w_1 is locally invertible and that w_2 is not locally invertible. Then in ud$(\mathfrak{U}(w_2))$ there is a set of undirected non-trivial paths from w_2 to a locally invertible vertex passing through no other locally invertible vertices. Choose one such path. This path will be labelled either by a suffix s of a piece, or by the involution of a prefix p of a piece, denoted $\bar{p}$. In either case, since w_2 is not locally invertible, s or p, respectively, will be proper. Since $\mathfrak{U}(w_1) \sim \mathfrak{U}(w_2)$, this same path can be found from w_2 in ud$(\mathfrak{U}(w_2))$. But in the first case, s will now be readable from a locally invertible vertex and hence a prefix of a piece; thus s is invertible, and since it is a proper suffix of a piece we have a contradiction. In the second case, $\bar{p}$ will be readable from a locally invertible vertex, and hence p is also a suffix of a piece; thus p is invertible, and since it is a proper prefix of a piece we have a contradiction. Thus w_2 must be locally invertible. By symmetry, we have the claim. □

Let now $\mathcal{N}$ be the set of all non-locally invertible vertices of $\mathfrak{U}$, and let $\mathcal{I}$ be the set of all locally invertible vertices of $\mathfrak{U}$. Of course, $\mathcal{N} \cap \mathcal{I} = \varnothing$. Then there exists by Lemma 5.3.9 a set of representatives $S \subseteq N_F \subset F(\mathfrak{U})$ such that $V_S = \mathcal{N}$. Fix such a set S, and call it $N(\mathfrak{U})$. Of course, if every vertex is locally invertible, i.e. if M is a free product of a free monoid by a group, then $N(\mathfrak{U}) = \varnothing$. We shall be interested in the graph $\textsc{Tree}(\mathfrak{U}, N(\mathfrak{U}))$ in certain situations. This completes our description of $\mathfrak{U}$. Using this graph, which captures a great deal of graphical information regarding the group of units of M, we will now turn to a description of the right units.

5.4 The Schützenberger graph of 1

We will now describe the Schützenberger graph of 1 of a special monoid M. Recall that this graph $\mathfrak{R}_1$ is the subgraph of the right Cayley graph of M induced on the right invertible elements. We shall begin by describing it in a very easy case; namely, the "no-folding case". It turns out that in this case $\mathfrak{R}_1$ is isomorphic to a tree of copies of $\mathfrak{U}$. In other words, the fact that there is an isomorphism $U_r(M) \cong U(M) * F$ for some finitely generated free monoid F is translated without any difficulty into graph-theoretic language. On the other hand, in the more general case, which we shall deal with after the no-folding case, this translation is harder. In part, this is because whereas the *abstract* generators of (say) the free monoid F are entirely disjoint from those of $U(M)$ in the above isomorphism, this is not the case when passing to the graph-theoretic interpretation; indeed, in the Cayley graph, the abstract generators of $U(M)$ are instead translated to words in Δ, and the abstract generators of F are translated to prefixes of words from Δ. Thus there can be a significant amount of "overlap" between these words. Some of this was dealt with in the construction of $\mathfrak{U}$; we now, informally speaking, deal with the rest.

5.4.1 The no-folding case

Suppose $I \cap I^2 = \varnothing$. Then we say that M satisfies the *no-folding condition.* That is, if M satisfies this condition, then no non-empty prefix of any piece appears as a proper subword of any prefix. For example, the bicyclic monoid $\text{Mon}\langle b, c \mid (bc) = 1 \rangle$ and the rather simple monoid $\text{Mon}\langle a, b, c, d \mid (ab)(cd)(ab) = 1 \rangle$ both satisfy the no-folding condition, but neither $\text{Mon}\langle a, b \mid (aab) = 1 \rangle$ nor $\text{Mon}\langle a, b \mid (b)(abc)(b) = 1 \rangle$ satisfies the condition. Here the parentheses are used to indicate the factorisation of the defining relations into minimal invertible pieces, as provided by Adian's overlap algorithm. We shall see that the no-folding condition makes it exceptionally easy to prove various graphs (including the right Cayley graphs) related to M are context-free. We shall, in this section, deal with this case in great detail, before extending it to the general case by using pushdown automata. Indeed, this section can be regarded as a "warm-up" for the general case.

Lemma 5.4.1. *Suppose the no-folding condition holds for the special monoid M. Then no proper subword of any piece is a piece. That is, M is given by an infix presentation.*

Proof. Suppose not, and let $\delta_1, \delta_2 \in \Delta$ are such that $\delta_1 \equiv h_1\delta_2 h_2$, with $h_1, h_2 \in A^+$. Then $h_1, h_1\delta_2 \in I$. As $\delta_2 \in I$, this means $h_1\delta_2 \in I \cap I^2$. This is a contradiction. □

It is very easy to understand the structure of the Schützenberger graph of 1 in the no-folding case. We write out the following lemma in a very explicit form, to ensure the reader that the behaviour of "right multiplication by a single letter" is easily understood; in fact, we shall see that it is so easily understood as to be simulatable into the graph of a pushdown automaton, whose total states correspond to the normal forms.

Lemma 5.4.2. *Suppose the no-folding condition holds. Suppose $u \in I_c^*$ is a normal form*

$$u \equiv u_0\xi_0u_1\xi_1 \cdots u_k\xi_k.$$

If $u \cdot a$ is right invertible for some letter $a \in A$, then exactly one of the following is true:

(i) $\xi_k a \in \Pi$,

(ii) $\xi_k a \in \Delta$,

(iii) $a \in \Pi$, *or*

(iv) $a \in \Delta$.

In case (i), the word

$$u_0\xi_0u_1\xi_1 \cdots u_k\xi_k',$$

where $\xi_k' \equiv \xi_k a \in \Pi$, is a normal form for ua; in case (ii) the word

$$u_0\xi_0u_1\xi_1 \cdots u_k',$$

where $u_k' \equiv u_k\xi_k a \in \Delta^*$, is a normal form for ua; in case (iii) the word

$$u_0\xi_0u_1\xi_1 \cdots u_k\xi_k u_{k+1}\xi_{k+1},$$

where $u_{k+1} \equiv \varepsilon, \xi_{k+1} \equiv a$, is a normal form for ua; and in case (iv) the word

$$u_0\xi_0u_1\xi_1 \cdots u_k\xi_k u_{k+1},$$

where $u_{k+1} \equiv a$, is a normal form for ua.

Proof. It is easy to see that $I \cap I^2 = \varnothing$ and $\Delta \cap \Pi = \varnothing$ imply that the four cases are mutually exclusive. We will factorise

$$u_0\xi_0u_1\xi_1 \cdots u_k\xi_k a$$

into maximal invertible factors. Suppose that the maximal invertible factor which includes u_k as a subword were to be of the form

$$u_i\xi_iu_{i+1}\xi_{i+1} \cdots u_{k-1}\xi_{k-1}u_kw'$$

where $0 \leq i < k$ and w' is some prefix of $\xi_k a$. However, as $I \cap I^2 = \varnothing$, it follows easily that if uvw is invertible and v is invertible, then u and w are also invertible. Hence it follows that $u_i\xi_i$ is invertible, so $\xi_i \equiv \varepsilon$; a contradiction, as then

$$u_0\xi_0u_1\xi_1 \cdots u_iu_{i+1}\xi_{i+1} \cdots u_k\xi_k$$

would be normal form for u with depth $k-1$. Thus there is no such i. Hence the maximal invertible factor which includes u_k as a subword is of the form u_kw', where w' is some prefix of $\xi_k a$. In particular w' is invertible. As ξ_k has no non-empty invertible prefix, we either have (a) $w' \equiv \xi_k a$ is invertible; or (b) $w' \equiv \varepsilon$.

Suppose we are in case (a). Then $\xi_k a \xrightarrow{*}_S u$ for some $u \in \Delta^*$. Then as no non-empty subword of ξ_k is invertible, and ξ_k contains no piece as a subword and is hence irreducible modulo S, any application of a rule to $\xi_k a$ must involve the entire word at once; hence $\xi_k a \in \Delta$.

Suppose we are in case (b). We then factorise $\xi_k a$ into maximal invertible factors. As ξ_k has no invertible suffix, since the given factorisation of u is a normal form, we must have that any maximal invertible subword of $\xi_k a$ containing a is a itself. But using $S(M)$ one immediately sees that any invertible letter is in Δ. Thus $a \in \Delta$, and we are in case (iv). If no maximal invertible subword of $\xi_k a$ contains a, then a is not invertible. Thus, when we find the normal form for ua by factorising from the left, we find that either $a \in \Pi$, or else $\xi_k a \in \Pi$. This is case (iii) resp. (i). □

From this lemma, we immediately deduce the following structural result.

Theorem 5.4.3. *Suppose $I \cap I^2 = \varnothing$. Let $u_0\xi_0u_1\xi_1 \cdots u_k\xi_k$ and $v_0\zeta_0v_1\zeta_1 \cdots v_n\zeta_n$ be any normal forms of right invertible elements in M. Then there is an edge*

$$u_0\xi_0u_1\xi_1 \cdots u_k\xi_k \xrightarrow{a} v_0\zeta_0v_1\zeta_1 \cdots v_n\zeta_n$$

if and only if one of the following holds:

(1) $n = k$, $u_i\xi_i =_M v_i\zeta_i$ for all $0 \leq i \leq k-1$, and $u_k\xi_k \cdot a =_M v_n\zeta_n$; or

(2) $n = k+1$, $u_i\xi_i =_M v_i\zeta_i$ for all $0 \leq i \leq k-1$, $\xi_k \not\equiv \varepsilon$, and $a \equiv v_n\zeta_n$.

Proof. The reverse implication is immediate; for the forward implication, we apply Lemma 5.4.2 and notice that case (1) corresponds to cases (i) and (ii), and case (2) corresponds to cases (iii) and (iv). □

Hence, we immediately deduce the following by Proposition 5.1.2.

Corollary 5.4.4. *Suppose $I \cap I^2 = \varnothing$. Then $\mathfrak{R}_1 \cong \textsc{Tree}(\mathfrak{U}, N(\mathfrak{U}))$.*

As a corollary, in the no-folding case, we have that if $\mathfrak{U}$ is context-free, then so too is $\mathfrak{R}_1$ by Proposition 5.1.5. We shall now give an interpretation of this in terms of pushdown automata, which will subsequently be the general approach used in the general case.

Suppose $U(M)$ is virtually free, and let $\mathfrak{U}$ be the graph constructed from the right Cayley graph of $U(M)$ as before. Then $\mathfrak{U}$ is a context-free graph with label alphabet A. Let $\mathcal{M} = \mathcal{M}(U) = (Q, A, Z, \delta, q_0, z_0, \hat{Q})$ be a pda such that $\Gamma(\mathcal{M}) \cong \mathfrak{U}$. Indeed, such a pda always exists by the main theorem (Theorem 2.6) of [363]. We will make some assumptions on $\mathcal{M}$, which are easy to see do not lose us any generality:

(1) We will assume that no transition of $\mathcal{M}$ changes the length of the stack by more than one, cf. e.g. [363, Lemma 2.4(i)]).

(2) We will assume that $\mathcal{M}$ never empties its stack except for when it is in the total state (q_0, ε), cf. e.g. [363, Lemma 2.4(ii)]).

(3) We will assume there is a distinguished symbol $\sigma \in Z$ such that every transition from q_0 is of the form $q_0 \xrightarrow{a, \varepsilon \mapsto \sigma} q$ for some $a \in A$ and $q \in Q$. Note that (2) ensures that this means that every transition *into* q_0 is of the form $q \xrightarrow{b, \sigma \mapsto \varepsilon} q_0$ for some $b \in A, q \in Q$.

(4) Most importantly, we will assume that the set of states Q is partitioned into two disjoint sets $Q = Q_I \cup Q_N$ such that every locally invertible vertex v of $\mathfrak{U}$ corresponds to a total state of the form (q_I, ζ) with $q_I \in Q_I$, and every non-locally invertible vertex v' of $\mathfrak{U}$ corresponds to a total state of the form (q_N, ζ') with $q_N \in Q_N$.

Of course, for arbitrary partitions of the set of vertices of $\mathfrak{U}$, one can easily construct examples which makes the analogous statement of (4) false. However, Lemma 5.3.9 guarantees that for this particular choice of partition this is possible; and, as conditions (1)–(3) can always be satisfied for any given machine with only very basic modifications, the conditions can easily be made simultaneously satisfied. One obvious consequence is also the following.

Lemma 5.4.5. *If $\mathcal{M} \vdash^* (q, \zeta)$, then $\zeta \equiv \sigma\zeta'$ for some $\zeta' \in Z^*$ or else $\zeta \equiv \varepsilon$ and $q = q_0$.*

We shall call Q_I resp. Q_N the *locally invertible* resp. the *non-locally invertible states* of $\mathcal{M}$. Note that as a corollary of (1), (2), and (3), we have that all edges going out from the initial state are of the form $q_0 \xrightarrow{a,x\mapsto xz} q$ for some $a \in A$, $x, z \in Z \cup \{\varepsilon\}$, and $q \in Q$, and all edges going into the initial state are of the form $q \xrightarrow{a,x\mapsto\varepsilon} q_0$ for some $a \in A, x \in Z \cup \{\varepsilon\}$, and $q \in Q$. Furthermore, from (1) and (2) it follows that, in fact, we may assume without loss of generality that all edges going out from the initial state are of the form $q_0 \xrightarrow{a,\varepsilon\mapsto z} q$, for edges of any other form will never be utilised.

We will now describe an extension operation E_1 on $\mathcal{M}$ and obtain a new pda $E_1(\mathcal{M})$. This pda will completely describe the right invertible elements of M. First, for every non-locally invertible state $q \in Q_N$, we add a new symbol z_q to the stack alphabet. We then perform the following two operations:

(1) For every non-locally invertible state $q \in Q_N$, and for every edge $q_0 \xrightarrow{a,\varepsilon\mapsto z} q'$, where $a \in A, z \in Z$, and $q' \in Q$, we add an edge $q \xrightarrow{a,\varepsilon\mapsto z_q} q'$.

(2) For every non-locally invertible state $q \in Q_N$, and for every edge $q' \xrightarrow{a,x\mapsto\varepsilon} q_0$, where $a \in A, x \in Z \cup \{\varepsilon\}$, and $q' \in Q$, we add an edge $q' \xrightarrow{a,z_q\mapsto\varepsilon} q$.

This gives a new pda $E_1(\mathcal{M})$. We call this pda the *first expansion* of $\mathcal{M}$. Let $Z_q = \bigcup_{q\in Q_N} z_q$ be the new symbols added, which, together with Z, forms the stack alphabet of $E_1(\mathcal{M})$. The following lemma is then immediate by induction on the number of transitions.

Lemma 5.4.6. *If $\zeta \in Z^*$, then $E_1(\mathcal{M}) \vdash^* (q, \zeta)$ if and only if $\mathcal{M} \vdash^* (q, \zeta)$. Furthermore, for $\zeta_1, \zeta_2 \in Z^*$, we have $(q_1, \zeta_1) \vdash^*_{E_1(\mathcal{M})} (q_2, \zeta_2)$ if and only if $(q_1, \zeta_1) \vdash^*_{\mathcal{M}} (q_2, \zeta_2)$.*

Thus the set of total states of $E_1(\mathcal{M})$ contains the set of total states of $\mathcal{M}$, and furthermore the subgraph of $\Gamma(E_1(\mathcal{M}))$ induced on the total states of $\mathcal{M}$ is isomorphic to $\Gamma(\mathcal{M})$; but this latter graph is – by definition – isomorphic to $\mathfrak{U}$. We will let $[q_0, \zeta]$, for $q_0 \in Q$ and $\zeta \in Z^*$, denote the vertex of $\mathfrak{U}$ corresponding, under this isomorphism, to the total state (q_0, ζ) of $E_1(\mathcal{M})$.

Lemma 5.4.7. *Suppose $I \cap I^2 = \varnothing$. Then $\Gamma(E_1(\mathcal{M})) \cong \textsc{Tree}(\mathfrak{U}, N(\mathfrak{U}))$.*

Proof. Let (q_0, ζ) be some total state of $E_1(\mathcal{M})$. Then as $\zeta \in (Z \cup Z_q)^*$, and $Z \cap Z_q = \varnothing$, either $\zeta \in Z^*$, or else we can point out the last occurrence in ζ of a symbol from Z_q, say as $\zeta \equiv \zeta_0 z_n \zeta_1$, where $\zeta_0 \in (Z \cup Z_q)^*, \zeta_1 \in Z^*$, and $z_n \in Z_q$ corresponding to the non-locally invertible state $q_n \in Q_N$. We define a map $\varphi\colon \Gamma(E_1(\mathcal{M})) \to \textsc{Tree}(\mathfrak{U}, N(\mathfrak{U}))$ inductively defined as:

$$\varphi(q, \zeta) = \begin{cases} ([q, \zeta]), & \text{if } \zeta \in Z^* \\ \varphi(q_n, \zeta_0) \cdot ([q, \sigma\zeta_1]) & \text{otherwise,} \end{cases}$$

where $\cdot$ denotes concatenation of sequences. We will prove by induction on $k \geq 0$ that $\textsc{Tree}_k(\mathfrak{U}, N(\mathfrak{U}))$ is isomorphic to subgraph of $\Gamma(E_1(\mathcal{M}))$ induced on the set of total states whose stack contains at most k occurrences of symbols from Z_q, via the above map. By passing to direct limits, this will yield a proof of the lemma. Let, for ease of notation, S_k denote the set of total states of $E_1(\mathcal{M})$ whose stack contains at most k occurrences of symbols from Z_q.

If $k = 0$, then

$$\begin{aligned} S_k := &\{(q, \zeta) \mid (q, \zeta) \text{ is a total state of } E_1(\mathcal{M}), \zeta \in Z^*\} \\ = &\{(q, \zeta) \mid (q, \zeta) \text{ is a total state of } \mathcal{M}, \zeta \in Z^*\} \end{aligned}$$

by Lemma 5.4.6. But this latter set is, by definition, $V(\Gamma(\mathcal{M}))$. As, again by definition, $\Gamma(\mathcal{M})$ is is isomorphic to $\mathfrak{U}$ via the map $(q, \zeta) \mapsto [q, \zeta]$, we are done.

Suppose $k > 0$ and that the claim is true for all $0 \leq i < k$. Let $v \in V(\textsc{Tree}_k(\mathfrak{U}, N(\mathfrak{U}))) \setminus V(\textsc{Tree}_{k-1}(\mathfrak{U}, N(\mathfrak{U})))$, let $\mathfrak{U}_v$ denote the unique copy of $\mathfrak{U}$ which v is a vertex of, and let $v_{k-1} \in V(\textsc{Tree}_{k-1}(\mathfrak{U}, N(\mathfrak{U}))) \cap V_{k-1}$ be the branch point onto which $\mathfrak{U}_v$ is attached. Now $v \neq 1_{\mathfrak{U}_v}$, as otherwise $v \in V(\textsc{Tree}_{k-1}(\mathfrak{U}, N(\mathfrak{U})))$. Therefore, we have $v = (v_0, v_1, \dots, v_{k-1}, v_k)$, where $v_i \in V_{N(\mathfrak{U})}$ for $0 \leq i < k$ and $v_k \in V(\mathfrak{U}) \setminus \{1_{\mathfrak{U}}\}$. Let (q, ζ) be the (uniquely determined) total state of $\Gamma(\mathcal{M})$ corresponding to v_k, i.e. $(q, \zeta) = \varphi^{-1}(v_k)$ and $[q, \zeta] = v_k$. Then, by our assumption, it follows from $v_k \neq 1_{\mathfrak{U}}$ that $\zeta \not\equiv \varepsilon$ and $q \neq q_0$. Furthermore, of course $\zeta \in Z^*$. Now as (q, ζ) is a total state of $\mathcal{M}$, we have $\mathcal{M} \vdash^* (q, \zeta)$, so there is a sequence of transitions

$$(q_0, \varepsilon) \xrightarrow{a_0, \varepsilon \mapsto \sigma} (q_1, z) \to \cdots \to \cdots (q_j, y_j) \xrightarrow{a_m, w_j \mapsto w'_j} (q, \zeta), \tag{5.4.1}$$

of minimal length, where the $\xrightarrow{a_i, w_i \mapsto w'_i}$ denotes arbitrary transitions in $\mathcal{M}$, $y_i \in Z^*$ for $0 \leq i \leq j$, and $\sigma \in Z$ is the distinguished single stack symbol from our assumptions on the pda. Furthermore, by our assumption on the pda, every y_i begins with the letter $\sigma \in Z$, as otherwise the stack would have been emptied, and we would have thus revisited (q_0, ε), contradicting the minimality of the sequence. Write $y_i \equiv \sigma y'_i$ for every $0 \leq i \leq j$, and $\zeta \equiv \sigma \zeta_i$.

Now as $(v_0, v_1, \dots, v_{k-1})$ is a vertex of $\textsc{Tree}_{k-1}(\mathfrak{U}, N(\mathfrak{U}))$ it follows by the inductive hypothesis that $\varphi^{-1}(v_0, v_1, \dots, v_{k-1})$ is a uniquely defined total state in S_{k-1}. Let (q', ξ) denote this total state, where $q' \in Q$ and $\xi \in (Z \cup Z_n)^*$. As v_{k-1} is an attachment point, i.e. $v_{k-1} \in V_{N(\mathfrak{U})}$, we have $q' \in Q_N$. Hence, by replacing the first transition in the above sequence (5.4.1) by one of the edges added in the first operation of the definition of $E_1(\mathcal{M})$, we find the following sequence of transitions in $E_1(\mathcal{M})$:

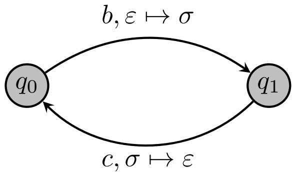
$$(q', \xi) \xrightarrow{a_0, \varepsilon \mapsto z_{q'}} (q_1, \xi z'_q) \to \cdots \to \cdots (q_j, \xi z_{q'} y'_j) \xrightarrow{a_m, w_j \mapsto w'_j} (q, \xi z_{q'} \zeta'). \tag{5.4.2}$$

In particular, again by the inductive hypothesis, we have $E_1(\mathcal{M}) \vdash^* (q, \xi z_{q'} \zeta'$. As all our choices were uniquely determined, the total state $(q, \xi z_{q'} \zeta')$ is uniquely determined by v. Furthermore, the distinguished occurrence of the letter $z_{q'}$ in $\xi z_{q'} \zeta'$ is the rightmost such occurrence, as $\zeta' \in Z^*$. But we have

$$\begin{aligned} v = (v_0, v_1, \dots, v_k) &= (v_0, v_1, \dots, v_{k-1}) \cdot (v_k) \\ &= \varphi(q', \xi) \cdot (v_k) \\ &= \varphi(q', \xi) \cdot ([q, \zeta]) \\ &= \varphi(q', \xi) \cdot ([q, \sigma\zeta']) \\ &:= \varphi(q, \xi z_{q'} \zeta'). \end{aligned}$$

Thus φ is a bijection on the vertices of $\textsc{Tree}_n(\mathfrak{U}, N(\mathfrak{U}))$; using this bijection, it is immediately clear that by considering adjacencies in $\mathfrak{U}$ and $\textsc{Tree}(\mathfrak{U}, N(\mathfrak{U}))$, and using Lemma 5.4.6 that for any two total states $(p_0, \xi_0), (p_1, \xi_1)$ of $\Gamma(E_1(\mathcal{M}))$, we have $(p_0, \xi_0) \vdash^a_{E_1(\mathcal{M})} (p_1, \xi_1)$ if and only if $\varphi(p_0, \xi_0) \xrightarrow{a} \varphi(p_1, \xi_1)$ in $\textsc{Tree}(\mathfrak{U}, N(\mathfrak{U}))$. Thus φ extends to a graph isomorphism, and we are done. □

Example 5.4.8. Let $M = \mathrm{Mon}\langle b, c \mid bc = 1\rangle$. Then $\Delta = \{bc\}$ and $I = \{b, bc\}$, so $I \cap I^2 = \varnothing$ and hence M satisfies the no-folding condition. Now $\mathfrak{U}$ consists of two vertices $[1, \varepsilon]$ and $[1, b]$, with an edge $[1, \varepsilon] \xrightarrow{b} [1, b]$ and an edge $[1, b] \xrightarrow{c} [1, \varepsilon]$. Thus a pda $\mathcal{M}$ such that $\Gamma(\mathcal{M}) \cong \mathfrak{U}$ is easily constructed; indeed, one can informally speaking simply take the states to be the vertices of $\mathfrak{U}$, and the transitions to correspond to the edges. Formally, we can take $\mathcal{M}$ to be the pda:


$b, \varepsilon \mapsto \sigma$
q_0
q_1
$c, \sigma \mapsto \varepsilon$

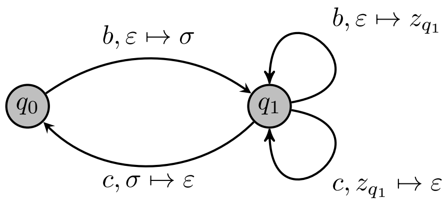
(5.4.3)

As we are only considering total states, we have not specified the final states of the pda, as these are not relevant. Clearly the total states of $\mathcal{M}$ are

$$T(\mathcal{M}) = \{(q_0, \varepsilon), (q_1, \sigma)\},$$

and as the only transitions are $(q_0, \varepsilon) \vdash^b_{\mathcal{M}} (q_1, \sigma)$ and $(q_1, \sigma) \vdash^c_{\mathcal{M}} (q_0, \varepsilon)$, it follows that $\Gamma(\mathcal{M}) \cong \mathfrak{U}$ via the isomorphism given by $(q_0, \varepsilon) \mapsto [1, \varepsilon]$ and $(q_1, \sigma) \mapsto [1, b]$.

Note that the non-locally invertible states are $Q_N = \{q_1\}$, and the invertible states $Q_I = \{q_0\}$. The first expansion $E_1(\mathcal{M})$ of $\mathcal{M}$ is thus:

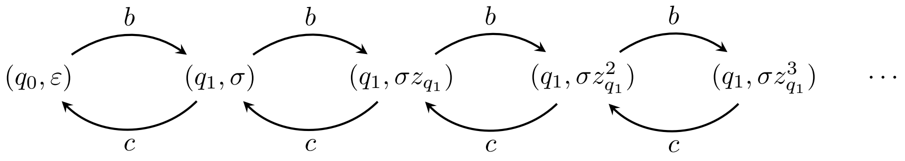

$b, \varepsilon \mapsto \sigma$
$b, \varepsilon \mapsto z_{q_1}$
q_0
q_1
$c, \sigma \mapsto \varepsilon$
$c, z_{q_1} \mapsto \varepsilon$

(5.4.4)

Now the total states of the first expansion of $\mathcal{M}$ is easily seen to be

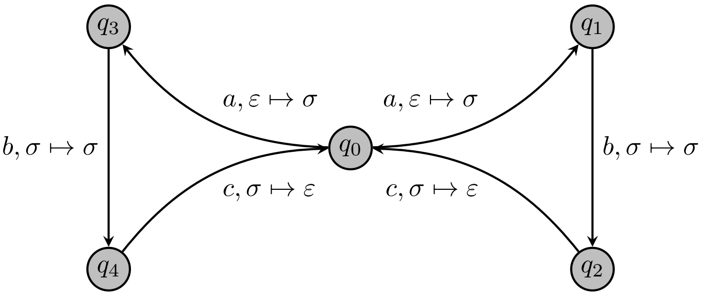
$$T(E_1(\mathcal{M})) = \{(q_0, \varepsilon), (q_1, \sigma z^i_{q_1}) \mid i \geq 0\},$$

and the graph $\Gamma(E_1(\mathcal{M}))$ of $E_1(\mathcal{M})$ is



This is clearly isomorphic to the Schützenberger graph $\mathfrak{R}_1$ of the bicyclic monoid, which is the subgraph of the right Cayley graph of the bicyclic monoid induced on the set of vertices $\{\pi(b^i) \mid i \geq 0\}$. △

Example 5.4.9. Let $M = \text{Mon}\langle a, b, c, d, e, f \mid abc = 1, def = 1\rangle$. Then $\Delta = \{abc, def\}$ and $U(M) = 1$. We have that $I \cap I^2 = \varnothing$, so M satisfies the no-folding condition. Then $\mathfrak{U}$ is a graph consisting of two triangles, one of whose boundary is labelled by abc, and the other by def. The only locally invertible vertex in this graph is the shared vertex, i.e. the root 1. We have that $\textsc{Tree}(\mathfrak{U}, N(\mathfrak{U})) \cong \mathfrak{R}_1$, as M satisfies the no-folding condition. A pda whose graph is isomorphic to $\mathfrak{R}_1$ is obtained as the first expansion of the following pda:



q_3
q_1
$a, \varepsilon \mapsto \sigma$
$a, \varepsilon \mapsto \sigma$
$b, \sigma \mapsto \sigma$
q_0
$b, \sigma \mapsto \sigma$
$c, \sigma \mapsto \varepsilon$
$c, \sigma \mapsto \varepsilon$
q_4
q_2

(5.4.5)

The non-locally invertible states of this pda are $Q_N = \{q_1, q_2, q_3, q_4\}$. Drawing the numerous edges resulting from the first expansion of the above pda would obscure the idea. The graph $\mathfrak{R}_1$ is shown, rather stylised, in Figure 5.8. △

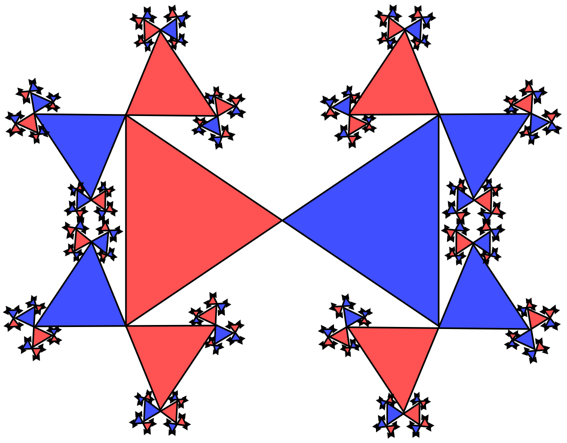


Figure 5.8: The (context-free) Schützenberger graph of 1 (with 1 being the central vertex) $\mathfrak{R}_1$ of the special monoid M defined by the presentation $\text{Mon}\langle a, b, c, d, e, f \mid abc = 1, def = 1\rangle$. The blue triangle corresponds to reading abc, and the red to def. The graph $\mathfrak{U}$ is isomorphic to the graph with one red and one blue triangle attached to the root. We have $\mathfrak{R}_1 \cong \textsc{Tree}(\mathfrak{U}, N(\mathfrak{U}))$.

Thus we have an automata-theoretic proof that $\mathfrak{R}_1$ is a context-free graph whenever $\mathfrak{U}$ is a context-free graph, provided that M satisfies the no-folding condition. We will now use the “bounded behaviour” described earlier to give a sketch proof that an automata-theoretic approach works in the general case, too.

5.4.2 The general case

For a fixed special monoid M, with no assumption on the presentation other than that it is finite, we shall now describe a pushdown automaton whose graph is $\mathfrak{R}_1$ when $U(M)$ is virtually free. This will be the graphical interpretation of Proposition 5.2.10. For a pda $\mathcal{M}$, we will let $T(\mathcal{M})$ denote the set of possible total states of $\mathcal{M}$.

Definition 5.4.10. We say that a (presentation of a) special monoid M is *benign* if the following statement holds for M: if $U(M)$ is context-free, then there exist a pda $\mathcal{P} = (Q, A', Z, \delta_{\mathcal{P}}, q_0, z_0, \widehat{Q})$ and a bijection

$$\varrho_{\mathfrak{R}} \colon I_0^*/\pi \to T(\mathcal{P}),$$

such that for every $u \equiv u_0\xi_0 \cdots u_k\xi_k \in I_0^*$ and $v \equiv v_0\zeta_0 \cdots v_n\zeta_n \in I_0^*$, where $u_i, v_j \in \Delta_0^*$ and $\xi_i, \zeta_j \in \Pi_0$, for every $0 \leq i \leq k$ and $0 \leq j \leq n$, except $\xi_k, \zeta_n \in \Pi_0 \cup \{\varepsilon\}$; and for every $a \in A$, we have

$$u_0\xi_0 \cdots u_k\xi_k \cdot a =_M v_0\zeta_0 \cdots v_n\zeta_n \quad \iff \quad \varrho_{\mathfrak{R}}(\pi(u)) \vdash_{\mathcal{P}}^{a} \varrho_{\mathfrak{R}}(\pi(v)). \tag{5.4.6}$$

In particular, $\mathfrak{R}_1 \cong \Gamma(\mathcal{P})$.

Thus, roughly speaking, a benign special monoid is one which does not hinder context-free behaviour of the group of units from being “blown up” to a context-free structure on $\mathfrak{R}_1$.

Theorem 5.4.11. *If M satisfies the no-folding condition, then M is benign.*

Proof. If $U(M)$ is not context-free, then M is vacuously benign. Assume that $U(M)$ is context-free. Then we may take $\mathcal{P}$ to be the pda $E_1(\mathcal{M})$ constructed in Lemma 5.4.7. □

In fact, we claim that all (finitely presented) special monoids are benign; at present, we are only able to provide a sketch proof of this claim.

Claim $(*)$**.** Every special monoid is benign.

Proof sketch of Claim $(*)$. If $U(M)$ is not context-free, then M is vacuously benign. Assume that $U(M)$ is context-free. Let u, v be as above. Recall that $\Omega = \max_{p\in\Pi} |p|$. Let $k' = \max(0, k - \Omega)$ and $n' = \max(0, n - \Omega + 1)$. Then it follows from Lemma 5.2.9 that

$$u_0\xi_0u_1\xi_1 \cdots u_k\xi_k \cdot a =_M v_0\zeta_0v_1\zeta_1 \cdots v_n\zeta_n$$

$$\iff$$

$$u_0\xi_0u_1\xi_1 \cdots u_{k'-1}\xi_{k'-1} =_M v_0\zeta_0v_1\zeta_1 \cdots v_{n'-1}\zeta_{n'-1} \quad \text{and} \tag{5.4.7}$$

$$u_{k'}\xi_{k'}u_{k'+1}\xi_{k'+1} \cdots u_k\xi_k \cdot a =_M v_{n'}\zeta_{n'}v_{n'+1}\zeta_{n'+1} \cdots v_n\zeta_n. \tag{5.4.8}$$

Thus, if $\varrho_{\mathfrak{R}}$ satisfies (5.4.6) on normal forms from I_0^* with $\max(k, n) \leq \Omega$, i.e. when (5.4.7) is vacuously true, then we can extend $\varrho_{\mathfrak{R}}$ to satisfy (5.4.6) for all normal forms in I_0^*.

For any $w \in I_0^*$, let $R(w) \subseteq I^*$ be a maximal, with respect to set inclusion, set of words satisfying the following: (1) for every $w' \in R(w)$, we can write $w' \equiv w_0' p_0' \cdots w_{t'}' p_{t'}$ for some $w_i' \in \Delta^*$, and $p_i' \in \Pi$ for all $0 \leq i \leq t'$, except possibly $p_{t'}' \in \Pi_\varepsilon$; and that furthermore (2) $w' =_M w$ for every $w' \in R(w)$; and (3) for every distinct $w', w'' \in R(w)$ with $t' = t''$, we have $w_i' \neq_M w_i''$ and $p_i' \not\equiv p_i''$ for every $0 \leq i \leq t' = t''$. By Proposition 5.2.10, for every $w' \in R(w)$, we have $d(w') \leq (d(w) + 1)\Omega$. Furthermore, it follows directly from [502, Proposition 4.3 & Theorem 4.4] that there are at most $\Omega d(w)$ words in $R(w)$, and that by maximality we have $w' \in I^*$ is such that $w' =_M w$ if and only if $w' \in R(w)$. Thus $R(w)$ is a finite and complete set of representatives for w as alternating products of words in Δ^* and Π.

Now every word $w' \equiv w_0' p_0' w_1' p_1' \cdots w_t' p_t' \in R(w)$ uniquely determines a vertex of $\textsc{Tree}(\mathfrak{U}, N(\mathfrak{U}))$; namely the vertex $([w_0'; p_0'], [w_1'; p_1'], \dots, [w_k'; p_k'])$. Property (3) now guarantees that, for two distinct $w', w'' \in R(w)$, the resulting vertices remain. Let $R'(w)$ be the collection of vertices obtained this way. Using this, we can prove the following.

Lemma 5.4.12. *There exists a surjection* $\textsc{Tree}(\mathfrak{U}, N(\mathfrak{U})) \to \mathfrak{R}_1$. *Furthermore, the preimage of any element of* $\mathfrak{R}_1$ *under this surjection is finite.*

Proof. For ease of notation, let $\mathrm{T}\mathfrak{U} = \textsc{Tree}(\mathfrak{U}, N(\mathfrak{U}))$. There is a natural map $\phi \colon \mathrm{T}\mathfrak{U} \to \mathfrak{R}_1$ given as follows. Let $(\hat{u}_0, \hat{u}_1, \dots, \hat{u}_k) \in V(\mathrm{T}\mathfrak{U})$ be an arbitrary element of $V(\mathrm{T}\mathfrak{U})$, where $\hat{u}_i \in \mathcal{N}$ for every $0 \leq i < k$ and $\hat{u}_i \in V(\mathfrak{U})$, using the bijection given by Proposition 5.1.2. As $\hat{u}_i \in \mathcal{N}$ for every $0 \leq i < k$, we can write $\hat{u}_i = [m_i, \xi_i]$ with $m_i \in U(M)$ and $\xi_i \in \Pi$ for $0 \leq i < k$ and $\xi_k \in \Pi_\varepsilon$. For every $0 \leq i \leq k$, let $[u_i; \xi_i]$ be a literal form of $[m_i, \xi_i]$, where $u_i \in \Delta^*$. We then set

$$\tau(\hat{u}_0, \hat{u}_1, \dots, \hat{u}_k) := \pi(u_0 \xi_0 u_1 \xi_1 \cdots u_k \xi_k).$$

This map is well-defined, as $\pi(u_i) = m_i$ for all $0 \leq i \leq k$ independently of literal form chosen. Now, every right invertible element of M, and therefore every vertex of $\mathfrak{R}_1$, is equal to some word from $I^* = (\Delta \cup \Pi)$ by [501, Lemma 4.1]; thus every such element is equal to some word of the form $u_0 \xi_0 u_1 \xi_1 \cdots u_n \xi_n$, where $u_i \in \Delta^*$ and $\xi_i \in \Pi^*$. It follows that τ is surjective. Now, given an element $m \in \mathfrak{R}_1$, represented by an element $w \in I_0^*$, it follows by maximality of $R(w)$ that $R'(w)$ is precisely the set of pre-images of m under τ; and $|R'(w)| \leq \Omega \cdot d(w) < \infty$. □

As $U(M)$ is context-free, so too is $\mathfrak{U}$. Hence, it is easy to construct a pda $\mathcal{U}'$ whose graph is $T_{\Omega(\Omega+1)}(\mathfrak{U}, N(\mathfrak{U}))$. Note that $R'(w)$ consists precisely of those vertices x of $T_{\Omega(\Omega+1)}(\mathfrak{U}, N(\mathfrak{U}))$ for which there is a walk from 1 to x with walk label w. Now, let as above $u' \equiv u_{k'} \xi_{k'} u_{k'+1} \xi_{k'+1} \cdots u_k \xi_k$ and $v' \equiv v_{n'} \zeta_{n'} v_{n'+1} \zeta_{n'+1} \cdots v_n \zeta_n$. We first wish to bound the depth of any word from I^* equal to u'. Note that as u' is a normal form, we have $d(u') = k - k' + 1 \leq \Omega$. Hence, if $u'' \in I^*$ is such that $u'' =_M u'$, we have by Proposition 5.2.10 that $d(u'') \leq \Omega(\Omega + 1)$. Furthermore, $u'' \in R(u')$, and $|R(u')| \leq \Omega^2(\Omega + 1)$. It follows by Lemma 5.4.12 that there is an edge from $\pi(u')$ to $\pi(v')$ in $\mathfrak{R}_1$ labelled by a if and only if there is an edge from some $u'' \in R'(u')$ to some $v'' \in R'(v')$ in $\textsc{Tree}(\mathfrak{U}, N(\mathfrak{U}))$ labelled by a. But as, for every such u'' and v'' we have $d(u'') \leq \Omega(\Omega + 1)$ and $d(v'') \leq \Omega(\Omega + 1)$, as $d(u') \leq \Omega(\Omega + 1)$ and $d(v') \leq \Omega(\Omega + 1)$, there is hence such an edge in $\textsc{Tree}(\mathfrak{U}, N(\mathfrak{U}))$ if and only if there is such an edge in $T_{\Omega(\Omega+1)}(\mathfrak{U}, N(\mathfrak{U}))$. Hence, the presence of an edge (in $\mathfrak{R}_1$) labelled by a from $\pi(u')$ to $\pi(v')$ is entirely determined by the structure of the graph $T_{\Omega(\Omega+1)}(\mathfrak{U}, N(\mathfrak{U}))$, and this structure is entirely determined by a pda. Thus, by a suitable encoding of this structure into a pda, we can construct a pda $\mathcal{X}$ whose set of total states is in bijective correspondence with the set of pairwise non-equal (in M) normal forms in I_0^* – that is, $V(\mathfrak{R}_1)$ – and where for every $(q_i, z_i), (q_i', z_i') \in T(\mathcal{X})$, we have that the pda transitions from (q_i, z_i) to (q_i', z_i') after reading $a \in A$ (i.e. $(q_i, z_i) \vdash_{\mathcal{X}}^{a} (q_i', z_i')$) if and only if there is an edge labelled by a from the vertex of $\mathfrak{R}_1$ corresponding to (q_i, z_i) to the vertex of $\mathfrak{R}_1$ corresponding to (q_i', z_i'). But this is precisely what we needed; indeed, we can take $\varrho_{\mathfrak{R}}$ to be the inverse of this bijection.

This completes the proof sketch of Claim $(*)$. □

In either case, we certainly have the following; indeed, we can regard the above sketch proof of Claim $(*)$ as a sketch proof of the following theorem with the assumption of benignity dropped.

Theorem 5.4.13. *Let M be a finitely presented benign special monoid. If $U(M)$ is virtually free, then $\mathfrak{R}_1$ is a context-free graph.*

Proof. ($\Longleftarrow$) Assume $U(M)$ is virtually free. Then $U(M)$ is context-free, so by definition of benignity, we may construct a pushdown automaton $\mathcal{P}$ from $\mathfrak{U}$ such that $\Gamma(\mathcal{P}) \cong \mathfrak{R}_1$. The result follows.

($\Longrightarrow$) Assume that $\mathfrak{R}_1$ is context-free. Since M is finitely presented, so is $U(M)$ by [309, Theorem 5]. Therefore, by the Muller-Schupp theorem, it suffices to show that $U(M)$ is context-free. By [398, Proposition 25] the automorphism group Aut($\mathfrak{R}_1$) of $\mathfrak{R}_1$ as a labelled graph is virtually free, as $\mathfrak{R}_1$ is deterministic, and context-free by assumption. Furthermore, by [468, Theorem 3], the automorphism group of $\mathfrak{R}_1$ is isomorphic to the Schützenberger group of $\mathscr{H}_1$. As $\mathscr{H}_1$ is a group $\mathscr{H}$-class, the Schützenberger group of $\mathscr{H}_1$ is hence isomorphic to $\mathscr{H}_1$ itself [430, 112]. Since $\mathscr{H}_1 = U(M)$, we have that $U(M) \cong$ Aut($\mathfrak{R}_1$) is a context-free group; hence it is virtually free by the Muller-Schupp theorem. □

Remark 5.4.1. Note that in the forward claim of the above result, the assumption that M is benign is not used. Furthermore, in the forward claim the assumption that M be finitely presented is only used to make the results easier to state, as we are only generally interested in the finitely presented case. However, [398, Proposition 25] actually proves that the automorphism group of any deterministic context-free graph is context-free. Hence we always have the implication ($\mathfrak{R}_1$ context-free $\implies$ $U(M)$ context-free), even dropping the assumptions of benignity and finite presentability. In fact, there are many cases when $\mathfrak{R}_1$ is a context-free graph even when M is not finitely presented; e.g. for $M = \text{Mon}\langle a, b, c \mid ab^ic = 1\ i \geq 1\rangle$. Then $\mathfrak{R}_1$ of M is a context-free graph, and $U(M)$ is therefore context-free (indeed $U(M) = 1$). The key to this fact is that while the submonoid of right units of M is not finitely presented (it is isomorphic to a free monoid of countable rank), it is generated by $\{a, ab, ab^2, \dots, ab^i, \dots, \}$, and the subgraph of the right Cayley graph of M

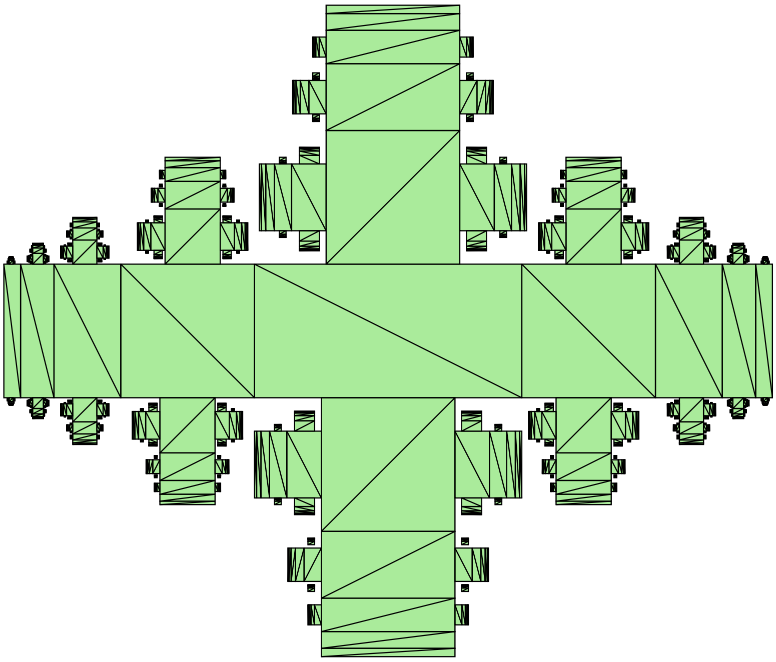


Figure 5.9: The Schützenberger graph of 1 (with edge labels suppressed) for the special monoid in Example 5.4.15. This representation is not entirely accurate, as, say, the vertical copy of $\mathfrak{U}$ in the upper centre continues "behind" the central strip infinitely far (but remains distinct from the copy in the lower centre). The reader might imagine that any copy of $\mathfrak{U}$ except the central one is folded in half along the line to which it is attached to the previous copy. We note that $\mathfrak{U}$ is embedded inside $\mathfrak{R}_1$ as the central horizontal strip; these are the vertices of depth 0.

induced on this set is easily described: it consists of the infinite binary tree with labels a and b, with a single edge labelled a entering the root. This is clearly a context-free graph – it has only two end-spaces. We remark on the similarity between this situation and the situation posed in Chapter 4 in dealing with an infinitely generated free monoid by regarding it as a formal language over a finite alphabet.

As a finitely generated group is context-free with respect to any finite generating set, and as the group of units of M is independent of the generating set of M, we additionally have the following corollary.

Corollary 5.4.14. *Assume Claim* $(*)$*. Let* M *be a special monoid, and let* A, B *be two finite generating sets for* M*. Then* $\mathfrak{R}_1(A)$ *is a context-free graph if and only if* $\mathfrak{R}_1(B)$ *is context-free.*

In particular, assuming Claim $(*)$, for any finitely presented special monoid M we may with impunity say that $\mathfrak{R}_1$ of M is a context-free graph or that it is not, without regard to the finite generating set chosen. We end this section with an example of a context-free Schützenberger graph of a special monoid which does not satisfy the no-folding condition.

Example 5.4.15. Let $M = \mathrm{Mon}\langle a, b, c \mid b(abc)b = 1\rangle$. Then $\Delta = \{b, abc\}$, and we have $U(M) \cong \mathbb{Z}$. Then $\mathfrak{R}_1$ is shown in Figure 5.9. △

We will now perform a far easier construction; namely, we shall use $\mathfrak{R}_1$ to construct the right Cayley graph of M in a treelike manner, using results due to Gray & Steinberg.

5.5 The right Cayley graph of special monoids

For the entirety of this section, let $M = \mathrm{Mon}\langle A \mid R_i = 1 \; (i \in I)\rangle$ denote a fixed finitely presented special monoid. This section concerns the building of the full monoid right Cayley graph of M from the Schützenberger graph of 1, and relies on the structural theorems of [171]. Let $m \in M$ be an arbitrary element of M. Let $\mathcal{T} \subseteq A^*$ be the set of words irreducible mod $S(T)$ with no right invertible suffix; the notation $\mathcal{T}$ suggests that $\mathcal{T}$ is a form of *transversal* of the $\mathscr{R}$-classes of M, which we will see below. By [171, Proposition 3.7], we can then uniquely write $m = [m_0]m'$ where $m_0 \in \mathcal{T}$ and $m' \in R(M)$. The following appears as [171, Corollary 3.12].

Theorem 5.5.1 (Gray & Steinberg '18). *Let $m \in M$. Then there is an isomorphism of labelled graphs $\mathfrak{R}_1 \to \mathfrak{R}_m$ sending 1 to $[m_0]$. If Γ_m is the induced subgraph of $\Gamma_M(M, A)$ consisting of all vertices accessible from m, then Γ_m is isomorphic to $\Gamma_M(M, A)$ as labelled graphs via an isomorphism taking 1 to $[m_0]$.*

This is already a great deal of structural information about $\Gamma_M(M, A)$ modulo $\mathfrak{R}_1$. However, for context-freeness, it does not give a whole lot; it is easy to construct an example of a graph whose strongly connected components are all pairwise isomorphic and context-free, and all Γ_m (retaining the notation from the above theorem) are isomorphic, and yet the resulting graph is not context-free. Thus we need more structure. We say that an edge of a digraph *enters* a strong component C of the graph if its initial vertex is not in C and its terminal vertex is in C. Dually, we say that an edge *leaves* if its terminal vertex is not in C, and its initial vertex is in C. The following is [171, Proposition 3.13].

Proposition 5.5.2 ([171]). *Let $m \in \pi(\mathcal{T}) \setminus \{1\}$ (so $m = [m_0]$). Then if $m_0 \equiv x \cdot a$ with $a \in A$, we have $[x] >_{\mathscr{R}} m$, and $[a] \notin U_r$, and $[x] \to_a m$ is the unique edge entering $\mathfrak{R}_m$.*

This is a great deal more information. The final piece of the puzzle comes from the information of how $\Gamma_M(M, A)$ is built up from $\mathfrak{R}_1$. First, recall that $\mathfrak{R}_1$ is the strongly connected component of the identity element in $\Gamma_M(M, A)$, or equivalently the subgraph of $\Gamma_M(M, A)$ induced on the set of right invertible elements $\mathscr{R}_1$. Let Γ be the directed graph obtained from $\Gamma_M(M, A)$ by collapsing each strongly connected component (and its internal edges) to a point. The vertex set of Γ is $M/\mathscr{R}$, and there is an edge (m, a) from the $\mathscr{R}$-class $\mathscr{R}_m$ of m to the $\mathscr{R}$-class $\mathscr{R}_{ma}$ of ma if $m \in M, a \in A$, and $\mathscr{R}_m \neq \mathscr{R}_{ma}$. The following is [171, Theorem 3.14].

Theorem 5.5.3 ([171]). *The graph Γ is isomorphic as a digraph to the Hasse diagram of $M/\mathscr{R}$ ordered by $\geq_{\mathscr{R}}$. This graph is a regular rooted tree with root $\mathfrak{R}_1$.*

The tree in the above theorem can be of infinite degree. Armed with these results, we now just need one more lemma before we can prove the main theorem. First, note that $\mathfrak{R}_1$ is an induced subgraph of $\Gamma_M(M, A)$. However, since $\mathfrak{R}_1$ need not be a regular graph in general, there are inside $\Gamma_M(M, A)$ some edges leaving $\mathfrak{R}_1$. In principle, if this leaving were not in any way controlled, it might be the case that no end-isomorphism inside $\mathfrak{R}_1$ will extend to one between the same vertices inside $\Gamma_M(M, A)$. On the other hand, if we do know that it is controlled, in the sense that such end-isomorphisms always do extend, then Proposition 5.5.2 together with Theorem 5.5.3 and Theorem 5.5.1 easily combine to show that $\Gamma_M(M, A)$ is a context-free graph. For, if $m \in M$, then by Theorem 5.5.3 we have that $\Gamma_M(M, A)(m)$ consists only of vertices $n \in M$ with $n \leq_{\mathscr{R}} m$, and by Proposition 5.5.2 together with Theorem 5.5.1, we must hence have that $\Gamma_M(M, A)(m) \sim \Gamma_M(M, A)(m')$, where $m = [m_0]m'$ as above. Hence, since each end-isomorphism between vertices inside $\mathfrak{R}_1$ extends to one between the same vertices in $\Gamma_M(M, A)$, we therefore would have that there are only finitely many end-isomorphism classes of vertices in $\Gamma_M(M, A)$.

But it is an easy matter to describe the edges of $\mathfrak{R}_1$ which leave $\mathfrak{R}_1$ using a pda; indeed, the pda $\mathcal{P}$ constructed in §5.4.2 with $T(\mathcal{P}) \cong \mathfrak{R}_1$ was constructed using a full description, via a pda $\mathcal{X}$, of which edges are *not* leaving. Thus, let $\mathfrak{R}_1'$ be the graph constructed in the following manner: if there is a vertex $m \in V(\mathfrak{R}_1)$ and an $a \in A$ such that there is an edge $m \xrightarrow{a} m'$ leaving $\mathfrak{R}_1$, where $m' \in M$, then attach this single edge (a "strand of hair") to m. Do this for every vertex of $\mathfrak{R}_1$. The resulting graph is denoted $\mathfrak{R}_1'$; considering $V(\mathfrak{R}_1)$ as a subset of $V(\mathfrak{R}_1')$ the subgraph of $\mathfrak{R}_1'$ induced on $V(\mathfrak{R}_1)$ is, of course, just $\mathfrak{R}_1$. Furthermore, for every $a \in A$, in $\mathfrak{R}_1'$ every vertex in $V(\mathfrak{R}_1)$ has exactly one edge labelled by a going out. Let $\mathcal{H} \subseteq V(\mathfrak{R}_1')$ denote the set of newly added vertices (the "ends of the hairs"). Using an entirely analogous procedure as in §5.4, it is now straightforward to construct a pda $\mathcal{P}'$ and a bijection

$$\varrho_{\mathfrak{R}'} \colon V(\mathfrak{R}_1') \to T(\mathcal{P}')$$

such that for every $v_1, v_2 \in V(\mathfrak{R}_1')$, we have that $v_1 \xrightarrow{a} v_2$ is an edge in $\mathfrak{R}_1'$ if and only if $\varrho_{\mathfrak{R}'}(v_1) \vdash^a_{\mathcal{P}'} \varrho_{\mathfrak{R}}(v_2)$. Thus $\mathfrak{R}_1'$ is a context-free graph, and it is not hard to see that there exists a set S of representatives of frontier points in $\mathfrak{R}_1'$ such that $V_S = \mathcal{H}$. By construction and Theorem 5.5.3, we now have that the right Cayley graph of M is isomorphic to a tree of copies of $\mathfrak{R}_1'$, where the branch points are the "ends of hairs" $\mathcal{H}$; that is, $\Gamma_M(M, A) \cong \textsc{Tree}(\mathfrak{R}_1', \mathcal{H})$. Thus, if $U(M)$ is context-free, then $\mathfrak{U}$ is context-free, and so too is $\mathfrak{R}_1$ and $\mathfrak{R}_1'$ by Theorem 5.4.13 resp. the above discussion. Hence, by Proposition 5.1.5, we have that $\textsc{Tree}(\mathfrak{R}_1', \mathcal{H})$ is context-free. This yields the main theorem of this chapter:

Theorem 5.5.4. *Let $M = \operatorname{Mon}\langle A \mid w_i = 1 \ (i \in I)\rangle$ be a finitely presented benign special monoid. Then the Cayley graph $\Gamma_M(M, A)$ is context-free if and only if $U(M)$ is virtually free.*

Proof. A finitely generated group is context-free if and only if it is virtually free; thus the reverse direction of the theorem is proved above. If $\Gamma_M(M, A)$ is, on the other hand, assumed context-free, then, as discussed in the proof of Theorem 5.3.8, it follows that $\Gamma_M(M, A)$ has finite tree-width. Thus $\mathfrak{R}_1$, being an induced subgraph of $\Gamma_M(M, A)$, also has finite tree-width; and thus $\mathfrak{U}$, being an induced subgraph of $\mathfrak{R}_1$ by Theorem 5.3.5, also has finite tree-width. By Theorem 5.3.8, this is equivalent to $U(M)$ being virtually free, and we are done. □

As any context-free graph is quasi-isometric to a tree by [106, Lemma 8.4], we have the following corollary, of independent interest.

Corollary 5.5.5. *Let $M = \mathrm{Mon}\langle A \mid w_i = 1 \ (i \in I)\rangle$ be a finitely presented benign special monoid. Then the right Cayley graph of M is quasi-isometric to a tree (as undirected graphs) if and only if $U(M)$ is virtually free.*

Proof. As noted above, if $U(M)$ is virtually free, then $\Gamma_M(M, A)$ is context-free, and hence quasi-isometric to a tree, cf. [106, Lemma 8.4]. On the other hand, if $\Gamma_M(M, A)$ is quasi-isometric to a tree, then since M has uniformly bounded degree, it follows that $\Gamma_M(M, A)$ has finite tree-width, as discussed in the proof of Theorem 5.3.8. The proof now proceeds just as the proof of Theorem 5.5.4, and we find that $U(M)$ is virtually free. □

This corollary could also be proved by noting that Lemma 5.3.7 reduces the problem of classifying when $\mathfrak{R}_1$ is quasi-isometric to a tree to the problem of classifying when the Cayley graph of $F * G$ is quasi-isometric to a tree, where F is a finitely generated free monoid and G is a group. This free product should be relatively straightforward to approach directly to show that this is equivalent to having G be virtually free; the result would then follow by, following Gray-Steinberg, building the Cayley graph of M in a tree-like way from $\mathfrak{R}_1$.

5.5.1 Logic of Cayley graphs

We will use the understanding of the right Cayley graph of a special monoid to deduce some theorems regarding the logic of such graphs. Logically, we can consider a labelled graph as a formal logical structure Γ with domain V (the vertex set of the graph) and a single relation, the edge relation. A predicate of first-order logic in a graph involves vertices, the edge relation, equality, quantifiers $(\exists, \forall)$, and their boolean combinations $(\neg, \wedge, \vee, \rightarrow)$. Monadic second-order logic also allows quantification (both universal and existential) over subsets of the vertices; if such quantification is only allowed over finite subsets, then this is known as weak monadic second-order logic.

The *first-order (monadic second-order) theory* of a graph Γ is the collection of all first-order (monadic second-order) predicates ϕ with no free variables such that $\Gamma \models \phi$. We say that the first-order (monadic second-order) theory of a graph is *decidable* if, given any first-order (monadic second-order) predicate ϕ, there is an algorithm which decides whether or not $\Gamma \models \phi$. For more detailed background on these notions, we refer the reader to e.g. [363, 270, 427].

Decidability of either the first-order or the monadic second-order theory of the Cayley graph of a finitely generated monoid M does not depend on the finite generating set chosen [270]. For this reason, we will generally omit reference to finite generating set below. There is a number of connections between decision problems for a given monoid M and different theories associated to the Cayley graph of M. This is most apparent in the group case: the first-order theory of the Cayley graph of a group is decidable if and only if the word problem for the group is decidable [269], and the monadic second-order theory of the Cayley graph of a group is decidable if and only if the group is virtually free [269, 363].

However, the same need not be true for monoids in general; we only have implications in one direction. By [270, Proposition 4], if the first-order theory of the Cayley graph of a finitely generated monoid is decidable, then the monoid has decidable word problem, but by [270, Proposition 5] there exists a monoid with word problem decidable even in linear time, but the Cayley graph of which nonetheless has undecidable first-order theory. One of the main consequences of Theorem 5.5.4 is the following, which lends credence to the adage that special monoids are, in many respects, very similarly behaved to groups.

Theorem 5.5.6. *Let $M = \operatorname{Mon}\langle A \mid R_i = 1 \ (i \in I)\rangle$ be a finitely presented benign special monoid. Then the right Cayley graph of M has decidable monadic second order theory if and only if the Cayley graph of the group of units of M has decidable monadic second order theory.*

Proof. ($\Longleftarrow$) The Cayley graph of a group has decidable monadic second order theory if and only if the group is virtually free [269, 363]. Thus $U(M)$ is virtually free; by Theorem 5.5.4 $\Gamma_M(M, A)$ is context-free. By [363, Theorem 4.4], the monadic second-order theory of $\Gamma_M(M, A)$ is hence decidable.

($\Longrightarrow$) Note that $\mathfrak{R}_1$ is the connected component of the root of $\Gamma_M(M, A)$, and so $\mathfrak{R}_1$ is *MSO-interpretable* in $\Gamma_M(M, A)$ (see [270] for this notion); hence if $\Gamma_M(M, A)$ has decidable monadic second order-theory, so too does $\mathfrak{R}_1$ (cf. [270, p. 3]). On the other hand, by Proposition 5.3.5, $\mathfrak{U}$ is the subgraph of $\mathfrak{R}_1$ induced on precisely the set of vertices v for which there exists a walk from $1_{\mathfrak{R}_1}$ to v with label in $\Delta^*\Pi$. Thus we easily see that the structure $\mathfrak{U}$ is MSO-interpretable in $\mathfrak{R}_1$, and so if $\mathfrak{R}_1$ has decidable monadic second order-theory, so too does $\mathfrak{U}$. It follows that $\mathfrak{U}$ has decidable monadic second-order theory; as $U(M)$ acts almost transitively on $\mathfrak{U}$, it follows by [268, Theorem 3.1], that $\mathfrak{U}$ is context-free, and so by Theorem 5.3.8 we have that $U(M)$ is context-free. Hence, by the Muller-Schupp theorem, the Cayley graph of $U(M)$ is context-free, and so has decidable monadic second-order theory by [363, Theorem 4.4]. □

The question of characterising in general which monoids have a right Cayley graph with decidable monadic second order theory was posed by Kuske & Lohrey in [270]. While the fully general case remains (wide) open, the above theorem completely answers this question in the (benign) special case. One application of the above theorem is to the rational subset membership problem. The following proposition has not appeared explicitly in the literature (to the best of the author's knowledge), but the proof is similar to that of [286, Theorem 17].

Proposition 5.5.7. *Let M be a monoid generated by a finite set A. If the Cayley graph $\Gamma_M(M, A)$ has decidable monadic second-order theory, then M has decidable rational subset membership problem.*

Proof. Let $L \subseteq A^*$ be a regular language. Let $\mathbf{reach}_L(x, y)$ be the predicate with variables $x, y \in M$ denoting whether the vertex y can be reached from x by some sequence of edges $E_{\sigma_1}, \dots, E_{\sigma_n}$ such that E_{σ_i} is an edge labelled by $\sigma_i \in A$, and $\sigma_1 \cdots \sigma_n \in L$. Clearly, for a word $w \in A^*$, we have that $\pi(w) \in \pi(L)$ if and only if $\Gamma_M(M, A) \models \mathbf{reach}_L(1, \pi(w))$. Thus, to establish the claim, it suffices to show that $\mathbf{reach}_L(x, y)$ is a monadic second-order predicate. In fact, it is even a weak monadic second-order predicate; this can be shown (see e.g. [427]) by induction on the operations of a regular expression for L. □

Remark 5.5.1. We do not require the full monadic second-order logic in the above proposition. The fragment, called FO(Reg) in [427], consisting of first-order logic together with the reachability predicates $\mathbf{reach}_L(x, y)$, is obviously sufficient; that it is not all of monadic second-order logic is not as obvious.

Of course, an immediate corollary of the above is the following.

Corollary 5.5.8. *Let $M = \mathrm{Mon}\langle A \mid R_i = 1 \ (i \in I)\rangle$ be a finitely presented benign special monoid with virtually free group of units. Then the rational subset membership problem for M is decidable.*

Recall that we have already obtained this corollary without the assumption of benignity via the vastly different methods in Chapter 3, namely as Corollary 3.5.12.

5.6 Some questions and miscellany

There are many natural questions that arise in the study of the geometry of special monoids. Perhaps the most natural is the following.

Question 5.6.1. *Let M be a (finitely presented) special monoid with context-free word problem. Does M have a context-free right Cayley graph? Conversely, if M has a context-free right Cayley graph, does M have context-free word problem?*

We make three remarks: (1) there is no dependency on the finite generating set chosen for a finitely presented special monoid to have context-free word problem; (2) assuming all special monoids are benign, there is no dependency on the finite generating set chosen for a finitely presented special monoid to have a context-free right Cayley graph; and (3) for groups, the answer to the question is positive, as proved by Muller & Schupp [362, 363].

Other connections between context-free right Cayley graphs and other forms of geometry are abundant; for example, Khukhro [254] has very recently proved that, for a group G, being finitely generated virtually free is equivalent to the Cayley graph being minor excluded, i.e. to there existing some finite graph which is not a minor of the Cayley graph of G. Is the analogous statement also true for special monoids? That is, if M is a special monoid, is the right Cayley graph of M minor excluded if and only if $U(M)$ is virtually free? This seems like a difficult question.

Recall for a special monoid M, it follows from §5.3 that the right Cayley graph of $U(M)$ embeds quasi-isometrically in the right Cayley graph of M. The following becomes natural.

Question 5.6.2. *Is there a natural notion of quasi-isometry for special monoids, which is stronger than the usual undirected one? In particular, develop a notion of quasi-convexity for submonoids of special monoids.*

There is also a strong notion of quasi-convexity for this submonoid, in that – informally speaking – the geodesics between elements of $U(M)$ in the right Cayley graph of M will almost always pass entirely within $\mathfrak{U}$; if they do not, then the length of the geodesic outside $\mathfrak{U}$ will be bounded by Ω, the length of the longest piece of M. Any natural notion of quasi-convexity of submonoids of special monoids would thus be one in which $U(M)$ is a quasi-convex submonoid of M. Similarly, the right units $U_r(M)$ should also be quasi-convex in M, by the results of Gray & Steinberg [171].

Another idea for future work comes from the fact that there are many natural notions of "ends" for groups and monoids; we have studied one in particular in this chapter. Other notions are similarly important, however, and while most coincide for groups, this need not be true for monoids in general; we refer the reader to [231, 256]. Studying these notions of ends for special monoids should be quite tractable. Using a notion of ends based on ends of digraphs introduced by [508], Craik et al [125] proved that a left cancellative monoid has either $0, 1, 2$, or $\geq \aleph_0$ such ends. This mirrors the situation that a group either has $0, 1, 2$ or $2^{\aleph_0}$ ends. It should not be difficult to use Theorem 5.6.8 to prove that any special monoid has either $0, 1, 2, \aleph_0$, or $2^{\aleph_0}$

ends (in this specific sense), and that, unless it is the bicyclic monoid, a group, or $\mathbb{N}$, it has $2^{\aleph_0}$ ends. We leave this for future work, or the interested reader.

5.6.1 Self-avoiding walks

The following short section is based on a small observation, which we present below as Theorem 5.6.4. It would be interesting to investigate it further. Let M be a monoid with finite generating set A. We say that a word $w \in A^*$ is *self-avoiding* if for all distinct prefixes p, q of w, we have $p \neq_M q$. Of course, a word is self-avoiding if and only if it is the label of a (unique) walk starting in 1 of the right Cayley graph of M with label alphabet A, and such that the walk does not visit the same vertex twice.

Example 5.6.3. Let $M = \operatorname{Mon}\langle b, c \mid bc = 1 \rangle$. The only self-avoiding words are those of the form $c^i b^j$ for some $i, j \geq 0$. For example, $ccbcc$ is not self-avoiding, as $ccbc =_M cc$. △

We let $\theta_M \colon \mathbb{N} \to \mathbb{N}$ denote the function such that $\theta_M(n)$ is the number of self-avoiding words of length n in M (we suppress the reference to generating set chosen). Of course, $\theta_M(0) = 1$. We are interested in the generating function of the sequence $\theta_M(0), \theta_M(1), \dots$. In general, this seems like a very difficult question. It turns out that for special one-relation monoids, quite a lot can be said already.

Theorem 5.6.4. *Let $M = \operatorname{Mon}\langle A \mid w = 1 \rangle$ be a special one-relation monoid such that the group of units $U(M)$ is trivial. Then $\sum_{n \geq 0} \theta_M(n) x^n$ is a rational function, i.e. a quotient of two polynomials $P(x)/Q(x)$. Moreover, these polynomials can be effectively computed from w.*

Proof. It follows by Adian's overlap algorithm that the $U(M)$ is trivial if and only if w is self-overlap free. Thus the rewriting system $\mathcal{R}$ with the single rule $(w \to 1)$ is complete and defines M. Let $u \in A^*$, with $u \equiv a_0 a_1 \cdots a_k$ for $a_i \in A$.

We claim that u is self-avoiding if and only if it is irreducible modulo $\mathcal{R}$. On the one hand, if u is not irreducible modulo $\mathcal{R}$, then $u \equiv u'wu''$, for some $u', u'' \in A^*$, so in particular $u'w =_M u'$ and u is not self-avoiding. Conversely, if u is not self-avoiding, then for some $1 \leq i < j \leq k$ we have

$$a_1 a_2 \cdots a_i =_M a_1 a_2 \cdots a_j.$$

As $\mathcal{R}$ is confluent and defines M, it follows that there is some word $x \in A^*$ such that $a_1 a_2 \cdots a_i \xrightarrow{*}_{\mathcal{R}} x$ and $a_1 a_2 \cdots a_j \xrightarrow{*}_{\mathcal{R}} x$. Thus either $a_1 a_2 \cdots a_i$ or $a_1 a_2 \cdots a_j$ is reducible modulo $\mathcal{R}$, or else $a_1 a_2 \cdots a_i \equiv a_1 a_2 \cdots a_j$. The latter is impossible, as $i < j$. In the former case, $a_1 a_2 \cdots a_j$ must be reducible as $i < j$, so in particular u is reducible. This completes the proof of the claim.

Now u is irreducible modulo $\mathcal{R}$ if and only if u does not contain w as a subword, i.e. if and only if $u \in A^* \setminus A^* w A^*$. This latter set is a regular language; and it is well-known that the generating function for the number of words of length n in a regular language is a rational function, effectively computable from a regular expression for the language (see [464, §4.7]). This generating function is hence precisely the generating function for θ_M. □

Example 5.6.5. Let $M = \text{Mon}\langle a, b \mid ababb = 1 \rangle$. Then $U(M) = 1$, and it is not hard to see that θ_M satisfies the recurrence relation with $\theta_M(n) = 2^n$ for $n < 5$ and

$$\theta_M(n) = 2\theta_M(n-1) - \theta_M(n-5)$$

for $n \geq 5$. In particular, we find

$$\sum_{n=0}^{\infty} \theta_M(n) x^n = \frac{1}{1 - 2x + x^5},$$

which is a rational function, as predicted by the theorem. △

We do not know much about self-avoiding words for special monoids. Investigating this question would be a rather interesting mix of enumerative combinatorics and techniques of combinatorial semigroup theory (as the proof of the above theorem demonstrates), and suitable for an introductory project in combinatorial semigroup theory. For example, for the special monoid $M = \text{Mon}\langle a, b \mid aba = 1 \rangle \cong \mathbb{Z}$, we have that $\theta_M(n)$ grows roughly linearly (as can be expected from considering the Cayley graph), and finding an explicit recursion for this function would be rather a nice exercise.

Question 5.6.6. *What is the generating function for the number of self-avoiding words in the special monoid $M = \text{Mon}\langle a, b, c \mid babcb = 1 \rangle$? Is it rational?*

Note that for M as in the above question, we have $U(M) \cong \mathbb{Z}$. Unlike for the earlier example of $\mathbb{Z}$, however, the function $\theta_M(n)$ clearly grows exponentially. We make the following bold conjecture, with little concrete support, except some minor experimental results.

Conjecture 5.6.7. *Let M be a (one-relation) special monoid. Then the generating function for $\theta_M(n)$ is a rational function of some polynomial transform of $\theta_{U(M)}(n)$.*

The interest in this conjecture would come primarily from the recent interest in self-avoiding words for groups, which has some curious links to e.g. amenability, see [185, 186], and also [227, 228, 280, 281, 195]

5.6.2 Growth functions of special monoids

Let M be a monoid with finite generating set A. The *length function* $\ell_A \colon M \to \mathbb{N}$ of M with respect to A is defined as follows: $\ell_A(m)$ is the smallest length of a word in the elements of A representing m. This is closely related to the word metric discussed in §1.4.5. The *growth function* $\gamma_A^M \colon M \to \mathbb{N}$ is defined as

$$\gamma_A^M(n) = |\{m \in M \mid \ell_A(m) \leq n\}|.$$

The function $\gamma := \gamma_A^M$ is said to have *exponential growth* if there exist constants $c > 1$ and $K > 0$ so that $\gamma_A^M(n) \geq Kc^n$ for all $n \geq 1$. Otherwise, we say that γ is *subexponential.* Note that as M is finitely generated it is clear that γ_A^M grows at most exponentially.[78] We say that γ grows (at most) *polynomially* if there exist constants $d, L > 0$ such that $\gamma_A^M(n) \leq Ln^d$ for

[78] Informally speaking, this is because the addition of relations in M means $\gamma := \gamma_A^M$ grows slower than $\gamma_A^{A^*}$, which is clearly exponential. Formally, as for all $m, n \in M$, we have $\ell_A(mn) \leq \ell_A(m) + \ell_A(n)$, it follows that $\gamma(m+n) \leq \gamma(m)\gamma(n)$, so $\gamma(n) \leq \gamma(1)^n = c^n$.

all $n \geq 1$. We analogously define *linear* growth. A function which grows subexponentially, but is not of polynomial growth, is said to have *intermediate growth.* Two growth functions γ and λ are *asymptotically equivalent* if there exist constants P, Q such that $\lambda(n) \leq \gamma(Pn)$ and $\gamma(n) \leq \lambda(Qn)$ for all $n \geq 1$.

Importantly, if A, B are two finite generating sets for a monoid M, then γ_A^M and γ_B^M are equivalent (see e.g. [187]). Thus, we may speak e.g. of a monoid with *polynomial growth.* Growth functions were introduced by Milnor [352, 353]. Early work by Bass [26] and Guivarc'h [192] connected growth with algebraic properties; Gromov [187] subsequently proved his famous theorem that a finitely generated group has polynomial growth if and only if it is virtually nilpotent. See especially [229]. Bergman [27] showed that an infinite finitely generated monoid cannot have sublinear growth. Bell & Zelmanov [47] recently characterised which functions can appear as the growth function of some semigroup. Kobayashi [263] used growth rate crucially to prove that there exists a finitely presented monoid with decidable word problem, but which does not admit any regular complete rewriting system. For further papers on growth, see [446, 369, 447, 448, 449, 451, 147, 370, 450, 452]. We prove the following rather quick and pleasant result.

Theorem 5.6.8. *Let M be a finitely presented special monoid with subexponential growth. Then M is either the bicyclic monoid, $\mathbb{N}$, or a group.*

Proof. The submonoid of right units of M is isomorphic to $F_r * U(M)$, where F_r is the free monoid of finite rank $|\Pi_0| = r \geq 0$. If $r = 0$, then $|\Pi_0| = 0$ implies that every piece is a letter; hence M is itself isomorphic to a free product of a free monoid F_m by a group G. If $m = 0$, then $M = G$ is a group. Now, if $m > 1$, then M cannot have subexponential growth, as it would contain the free monoid F_2, which has exponential growth. Suppose $m = 1$. If G has some element of infinite order, then G is infinite and contains a submonoid isomorphic to $\mathbb{N}$, so M contains a submonoid isomorphic to $\mathbb{N} * \mathbb{N} \cong F_2$, and so M has exponential growth, a contradiction. On the other hand, if every element of G has finite order, then G contains some element of order $1 \leq n \leq \infty$, so it contains as a submonoid the monoid $\mathbb{N} * C_n$ for the finite cyclic group C_n. If $n > 1$, then one readily checks that $\mathbb{N} * C_n$ has exponential growth; for example, $\mathbb{N} * C_2$ is easily seen to grow (in k) as φ^k, where $\varphi = \frac{1+\sqrt{5}}{2}$ is the golden ratio.[79] Thus necessarily $n = 1$ for every element of finite order n in G, so G is trivial. Thus $M \cong \mathbb{N}$ if $r = 0$.

On the other hand, if $r > 0$, then necessarily $r = 1$, as otherwise M contains the free monoid F_2 as a submonoid. If $U(M)$ is non-trivial, then by the same argument as above, it follows that M, containing $\mathbb{N}*U(M)$, must have exponential growth; thus $U(M) = 1$. We may thus assume without loss of generality that M is given by an infix presentation, i.e. no piece is contained within another piece, by Proposition 3.6.6. Let Δ be the pieces of the presentation of M. Now $|\Pi_0| = r = 1$, so there is one prefix of pieces which is irreducible modulo $S(M)$; if this prefix

[79] Let $\mathbb{N} * C_2 = \text{Mon}\langle a, b \mid b^2 = 1\rangle$. Then the growth function $\gamma(k)$ counts the number of words S_k of length k of as and bs which contains no two consecutive bs. Any element of S_k either starts with a or b. Every word in the former is obtained by concatenating a with some word from S_{k-1}. For every word in the latter, the b must be succeeded by an a, at which point any word from S_{k-2} can be appended. All words in S_k appear in this way. Hence $\gamma(k) = \gamma(k-1) + \gamma(k-2)$ for $k \geq 2$. Note that $\gamma(0) = |S_0| = 1$, and $\gamma(1) = |S_1| = \{a, b\} = 2$. Thus $\gamma(k) = F_k \sim \varphi^k$, where F_k is the kth Fibonacci number.

had length > 1 then its first letter would be irreducible, and so that letter would also be in Π_0, a contradiction. Thus, suppose $\Pi_0 = \{a\}$, where $a \in A$.

As no piece contains any piece as a subword, it follows that no proper prefix of a piece contains any piece as a subword. Thus every proper prefix of every piece is irreducible modulo $S(M)$. Hence every proper prefix of every piece is a power of a. The pieces of M are thus $a^{i_1}a_1, a^{i_2}a_2, \dots, a^{i_n}a_n$, where $i_j \geq 0$ and $a_j \in A$ for all $1 \leq j \leq n$. By an entirely symmetric argument, reasoning instead using the left units of M, it follows that every proper *suffix* of every piece is a power of some letter $b \in A$. the pieces of M are hence, written in the same order as above, graphically equal to $b_1 b^{k_1}, b_2 b^{k_2}, \dots, b_n b^{k_n}$, where $k_j \geq 0$ and $b_j \in A$ for all $1 \leq j \leq n$. For a fixed $1 \leq j \leq n$, the only way this is possible is if either (1) $b_j \equiv a^{j_1}$, $b^{k_j} \equiv a_j$, and $a \not\equiv b$, or else (2) $a^{i_j} \equiv b^{k_j} \equiv \varepsilon$ and $a_j \equiv b_j$. It follows that for every $\delta \in \Delta$, we have (i) $\delta \equiv ab$; or else (ii) $\delta \equiv a \equiv b$.

If at least one piece is of the form (ii), then all pieces are, and as $a \equiv b =_M 1$, these generators may be removed by Tietze transformations, leaving a presentation with no invertible pieces; that is, if $A' = A \setminus \{a, b\}$, then $M \cong \operatorname{Mon}\langle A' \mid \varnothing \rangle$. As M has subexponential growth, we must have $|A'| < 2$ in order to avoid containing a free submonoid of rank 2, so either $A' = \varnothing$ and M is the trivial group, or else $|A'| = 1$ and $M \cong \mathbb{N}$.

If no piece is of the form (ii), then all pieces are of the form $\delta \equiv ab$, i.e. $\Delta = \{ab\}$. Thus

$$M \cong \operatorname{Mon}\langle A \mid (ab)^{n_1} = 1, (ab)^{n_2} = 1, \dots, (ab)^{n_\ell} = 1 \rangle$$

where $n_1, n_2, \dots, n_\ell \geq 1$. Letting $d = \gcd(n_1, n_2, \dots, n_\ell) \geq 1$, we find that

$$M \cong \operatorname{Mon}\langle A \mid (ab)^d = 1 \rangle$$

is a one-relation monoid. Now it follows from Adian's theorem on the group of units of a special one-relation monoid that $U(M) \cong C_d$, a finite cyclic group. As $U(M) \cong 1$, thus $d = 1$, so

$$M \cong \operatorname{Mon}\langle A \mid ab = 1 \rangle \cong \operatorname{Mon}\langle A' \mid \varnothing \rangle * \operatorname{Mon}\langle a, b \mid ab = 1 \rangle.$$

where $A' \equiv A \setminus \{a, b\}$. If now $|A'| > 0$, then M contains a submonoid isomorphic to $\mathbb{N} * \mathbb{N}$. Thus $|A'| = 0$, so $M \cong \operatorname{Mon}\langle a, b \mid ab = 1 \rangle$ is isomorphic to the bicyclic monoid. □

Thus, as finite monoids have subexponential growth, we recover as, a corollary of Theorem 5.6.8, Adian's theorem: a special monoid is finite if and only if it is a finite group.

We end with a connection to a famous open problem. For some time, it was an open problem (posed in 1968 by Milnor [351]) whether there exist finitely generated groups with intermediate growth. Examples of such groups where proved to exist in 1984 by Grigorchuk [182] (see also [184, 183]). It remains an open problem whether there exist finitely *presented* groups with intermediate growth; the most common conjecture appears to be that such groups do not exist, see e.g. [181, Conjecture 11.3]. Whether or not this is the case, the following immediate corollary of Theorem 5.6.8 shows that special monoids are rather tame.

Corollary 5.6.9. *Any finitely presented special monoid with intermediate growth is a group.*

Hence, if one is interested in hunting for finitely presented special monoids with intermediate growth, then one might as well become a group theorist. We are in no position to speak about such matters; and whereof we cannot speak, thereof we must be silent [494, Proposition 7].

Notes on Literature

As will be clear from reading the bibliography, many references have been used as part of the writing of this thesis. However, the usage of some references has been indirectly, rather than directly; this means that, were this not addressed, many relevant references would be excluded (through no fault of their own) from the bibliography. In this section, we address this issue.

For general semigroup theory, we have already mentioned the referential works by Ljapin, Clifford & Preston, and others. We mention a few more. The reader will no doubt find Higgins's book [205] useful in practice, particularly if they are interested in geometric semigroup theory (a subject only tangentially mentioned in this thesis). To this end, the monograph by Guba & Sapir [191] is an excellent starting point, as is the Ph. D. thesis of Kilibarda [255] (these ideas were treated implicitly already by Adian [7]). For a delightful and extraordinarily deep dive into combinatorics on words, one has the books by Lothaire [289, 290, 291, 292].

The reader interested in general decision problems for groups may consult two excellent pieces of literature by Miller: the rather terse monograph [349], and the pleasantly readable and accessible survey [350]. The reader may also find the survey by Vazhenin [484] very worthwhile consulting, and indeed the short [8]. The survey by Adian & Durnev [11] is excellent and treats a broad range of topics in-depth (though there are some number of omissions and inaccuracies in this survey, and the quality of the English translation suffers). We also refer the reader to the monograph by Bokut [65] as well as Bokut & Kukin [64] for slightly different perspectives on algorithmic problems in algebra. See also [198, 364, 72]. On a more philosophical note, the reader interested in the very early history of computability and the human attitude towards it, we cannot recommend Berkeley's 1949 remarkable book [52] higher. The reader who found the (sizable) footnote in §1.1.4 enjoyable may also enjoy the articles [459, 46, 45, 117, 42, 43, 44], which connect physical systems (e.g. ones arising in Newtonian mechanics) with decidability.

For an introduction to many of the topics in the theory of rewriting systems, we cannot recommend higher the monograph by Book & Otto [71], which is very readable for anyone with even a modicum of interest in the subject. The monograph by Jantzen [235] is also recommended, especially for connections with formal language theory via e.g. ancestors and descendants (not entirely unrelated to the work done in Chapter 2 of this present thesis). However, Jantzen's style is, at times, terse. We have, in this thesis, only made some connections between rewriting systems and formal language theory. This area on its own is vast, having seen its inception sometime in the 1970s, and cannot be adequately summarised in a few articles. We give some pointers. One particularly early and influential article by Benois [48] recognised the decidability of the rational subset membership problem for free groups via formal language theory. Other important early, and highly readable, articles include [116, 379, 380, 90, 91, 114]. For more specialised techniques of rewriting (e.g. via simulating Turing machines), we refer the reader to begin with [22, 51, 50, 66, 70, 68, 71, 101, 104, 105, 110, 111, 235, 342, 299, 368, 365, 366, 389, 390, 391, 396, 392, 393, 436, 437, 438]. We also refer the reader to the beautiful articles by Adian [9, 10].

Regarding the geometric notions in §1.4, we have only discussed a rather particular notion of ends of monoids. For other notions of ends of monoids and semigroups, and related topics, we refer the reader primarily to [140, 508, 231, 251, 165, 167, 166, 256, 125, 97] as well as [250, §2.4]. Standard references on geometric group theory include the books by Bridson & Haefliger [77] and the monograph by Gromov [188]. A particularly inviting reference is Gromov [189]. Other standard references include the memoir by Bowditch [74], the book by de la Harpe [128], and the recent monograph by Druţu & Kapovich [143]. We can also recommend the reader to consult [159, 100, 76, 75]. We refer the brave reader to the monograph by Wise [493].

If the reader is interested in the history of combinatorial group theory, then they will delight in learning of the existence of a monograph dedicated to this precise topic – the 1982 masterpiece by Chandler & Magnus [107]. Lyndon [296] has also written an elegant and philosophising survey of combinatorial group theory. For a more general history of group theory, the reader may consult the at times rather esoteric book by Wussing [497]. Kleiner [262] also gives an exquisite account of the history of group theory, although the reader is directed to read Wussing first, in order to gain a slightly broader view of the subject; for example, Kleiner does not deal with semigroups in any depth, and repeats Clifford & Preston's mistake to identify de Séguier's *demi-groupes* with *semi-groupes.* The lectures by Klein [261], which includes the development of group theory, are also invaluable. A more indirect reference is the collection of Mathematical Reviews on infinite groups collated by Baumslag [37], which gives a rather naked presentation of combinatorial group theory.

The reader interested in the history of semigroup theory will, by contrast, be somewhat disappointed. There is one major historical account of the development of algebraic semigroups, published only a few years ago; this is the book by Hollings [214], which deals (roughly) with the twentieth century up to 1970. The reader is also referred to [213, 215, 216] for further reading, and, as mentioned in the preface, [212] for a detailed overview of the history of semigroups. These sources present a fairly complete picture; however, they lack a detailed study of the epistemological grounds of generalisations of groups (or indeed other objects) and the initial motivations for these generalisations, which – from a certain point of view – can be seen as the true foundation of semigroup theory. This thesis is not the place for this latter discussion.

We remark, finally, that every letter in the English alphabet appears as the first letter of at least one reference. The author thanks A. Yamamura for sending him an email when only the letter Y remained, and thereby indirectly and unknowingly helping to resolve this matter.

Bibliography

[1] H. Abdulrab, P. Goralčík, and G. S. Makanin. Towards parametrizing word equations. *Theor. Inform. Appl.*, 35(4):331–350, 2001.

[2] S. I. Adian. Algorithmic unsolvability of problems of recognition of certain properties of groups. *Dokl. Akad. Nauk SSSR (N.S.)*, 103:533–535, 1955.

[3] S. I. Adian. On the embeddability of semigroups in groups. *Soviet Math. Dokl.*, 1:819–821, 1960.

[4] S. I. Adian. The problem of identity in associative systems of a special form. *Soviet Math. Dokl.*, 1:1360–1363, 1960.

[5] S. I. Adian. Identities in special semigroups. *Dokl. Akad. Nauk SSSR*, 143:499–502, 1962.

[6] S. I. Adian. *Defining relations and algorithmic problems for groups and semigroups.* Proceedings of the Steklov Institute of Mathematics, No. 85 (1966). Translated from the Russian by M. Greendlinger. American Mathematical Society, Providence, R.I., 1966.

[7] S. I. Adian. Word transformations in a semigroup that is given by a system of defining relations. *Algebra i Logika*, 15(6):611–621, 743, 1976.

[8] S. I. Adian. On some algorithmic problems for groups and monoids. In *Rewriting techniques and applications (Montreal, PQ, 1993)*, volume 690 of *Lecture Notes in Comput. Sci.*, pages 289–300. Springer, Berlin, 1993.

[9] S. I. Adian. Estimation of derivational complexity in a rewriting system. *Dokl. Akad. Nauk*, 428(3):295–299, 2009.

[10] S. I. Adian. On a method for proving exact bounds on derivational complexity in Thue systems. *Math. Notes*, 92(1-2):3–15, 2012. Translation of Mat. Zametki **92** (2012), no. 1, 3–18.

[11] S. I. Adian and V. G. Durnev. Algorithmic problems for groups and semigroups. *Uspekhi Mat. Nauk*, 55(2(332)):3–94, 2000.

[12] S. I. Adian and G. U. Oganesian. On problems of equality and divisibility in semigroups with a single defining relation. *Math. USSR, Izv.*, 12:207–212, 1978. [Izv. Akad. Nauk SSSR Ser. Mat. **42**, 2 (1978)].

[13] I. Agol. The virtual Haken conjecture. *Doc. Math.*, 18:1045–1087, 2013. With an appendix by Agol, Daniel Groves, and Jason Manning.

[14] A. V. Aho. Indexed grammars—an extension of context-free grammars. *J. A. Comp. Mach.*, 15:647–671, 1968.

[15] A. V. Aho, R. Sethi, and J. D. Ullman. Code optimization and finite Church-Rosser systems. In *Design and optimization of compilers (Courant Comput. Sci. Sympos. 5, New York Univ., New York, 1971)*, pages 89–105. Prentice–Hall Series in Automat. Comput. 1972.

[16] D. J. Anick. On the homology of associative algebras. *Trans. Amer. Math. Soc.*, 296(2):641–659, 1986.

[17] A. V. Anīsīmov. The group languages. *Kibernetika (Kiev)*, (4):18–24, 1971.

[18] M. Anshel and P. Stebe. The solvability of the conjugacy problem for certain HNN groups. *Bull. Amer. Math. Soc.*, 80:266–270, 1974.

[19] E. Artin. Theorie der Zöpfe. *Abh. Math. Sem. Univ. Hamburg*, 4(1):47–72, 1925.

[20] M. Aschenbrenner, S. Friedl, and H. Wilton. *3-manifold groups.* EMS Series of Lectures in Mathematics. European Mathematical Society (EMS), Zürich, 2015.

[21] L. Auslander. On a problem of Philip Hall. *Ann. of Math. (2)*, 86:112–116, 1967.

[22] J.-M. Autebert and L. Boasson. The equivalence of pre-NTS grammars is decidable. *Math. Systems Theory*, 25(1):61–74, 1992.

[23] J.-M. Autebert, L. Boasson, and G. Sénizergues. Groups and NTS languages. *J. Comput. System Sci.*, 35(2):243–267, 1987.

[24] G. H. Bagherzadeh. Commutativity in one-relator groups. *J. London Math. Soc. (2)*, 13(3):459–471, 1976.

[25] V. G. Bardakov and M. V. Neshchadim. Lower central series of baumslag–solitar groups. *Algebra and Logic*, 59(4):281–294, 2020.

[26] H. Bass. The degree of polynomial growth of finitely generated nilpotent groups. *Proc. London Math. Soc. (3)*, 25:603–614, 1972.

[27] H. Bass. A note on growth functions of algebras and semigroups, 1978. Unpublished lecture notes, University of California, Berkeley.

[28] C. J. K. Batty and Y. Tomilov. Quasi-hyperbolic semigroups. *J. Funct. Anal.*, 258(11):3855–3878, 2010.

[29] B. Baumslag. *Residually free groups and related topics.* PhD thesis, University of London, 1965.

[30] B. Baumslag. Intersections of finitely generated subgroups in free products. *J. London Math. Soc.*, 41:673–679, 1966.

[31] B. Baumslag and S. J. Pride. Groups with two more generators than relators. *J. London Math. Soc. (2)*, 17(3):425–426, 1978.

[32] B. Baumslag and S. J. Pride. Groups with one more generator than relators. *Math. Z.*, 167(3):279–281, 1979.

[33] G. Baumslag. On generalised free products. *Math. Z.*, 78:423–438, 1962.

[34] G. Baumslag. Residually finite one-relator groups. *Bull. Amer. Math. Soc.*, 73:618–620, 1967.

[35] G. Baumslag. Positive one-relator groups. *Trans. Amer. Math. Soc.*, 156:165–183, 1971.

[36] G. Baumslag. A finitely presented solvable group that is not residually finite. *Math. Z.*, 133:125–127, 1973.

[37] G. Baumslag. *Reviews on Infinite Groups: As Printed in Mathematical Reviews 1940 Through 1970, Volumes 1-40 Inclusive.* Number pt. 1 in Reviews on Infinite Groups: As Printed in Mathematical Reviews 1940 Through 1970, Volumes 1-40 Inclusive. American Mathematical Society, 1974.

[38] G. Baumslag. A survey of groups with a single defining relation. In *Proceedings of groups—St. Andrews 1985*, volume 121 of *London Math. Soc. Lecture Note Ser.*, pages 30–58. Cambridge Univ. Press, Cambridge, 1986.

[39] G. Baumslag. *Topics in combinatorial group theory.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 1993.

[40] G. Baumslag and J. E. Roseblade. Subgroups of direct products of free groups. *J. London Math. Soc. (2)*, 30(1):44–52, 1984.

[41] G. Baumslag and D. Solitar. Some two-generator one-relator non-Hopfian groups. *Bull. Amer. Math. Soc.*, 68:199–201, 1962.

[42] E. Beggs, J. F. Costa, and J. V. Tucker. Physical experiments as oracles. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, (97):137–151, 2009.

[43] E. J. Beggs, J. F. Costa, and J. V. Tucker. Physical oracles: the Turing machine and the Wheatstone bridge. *Studia Logica*, 95(1-2):279–300, 2010.

[44] E. J. Beggs, J. F. Costa, and J. V. Tucker. The impact of models of a physical oracle on computational power. *Math. Structures Comput. Sci.*, 22(5):853–879, 2012.

[45] E. J. Beggs and J. V. Tucker. Can Newtonian systems, bounded in space, time, mass and energy compute all functions? *Theoret. Comput. Sci.*, 371(1-2):4–19, 2007.

[46] E. J. Beggs and J. V. Tucker. Experimental computation of real numbers by Newtonian machines. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 463(2082):1541–1561, 2007.

[47] J. Bell and E. Zelmanov. On the growth of algebras, semigroups, and hereditary languages. *Invent. Math.*, 224(2):683–697, 2021.

[48] M. Benois. Parties rationnelles du groupe libre. *C. R. Acad. Sci. Paris Sér. A-B*, 269:A1188–A1190, 1969.

[49] M. Benois. Simplifiabilité et plongement dans un groupe des monoïdes quotients d'un monoïde libre par une congruence de Thue unitaire. *C. R. Acad. Sci. Paris Sér. A-B*, 276:A665–A668, 1973.

[50] M. Benois. Descendants of regular language in a class of rewriting systems: algorithm and complexity of an automata construction. In *Rewriting techniques and applications (Bordeaux, 1987)*, volume 256 of *Lecture Notes in Comput. Sci.*, pages 121–132. Springer, Berlin, 1987.

[51] M. Benois and J. Sakarovitch. On the complexity of some extended word problems defined by cancellation rules. *Inform. Process. Lett.*, 23(6):281–287, 1986.

[52] E. C. Berkeley. *Giant Brains or Machines That Think.* John Wiley & Sons, Inc., New York, N. Y., 1949.

[53] J. Berstel. Congruences plus que parfaites et langages algébriques. *Seminaire d'informatique Théorique*, pages 123–147, 1977.

[54] J. Berstel and D. Perrin. *Theory of codes*, volume 117 of *Pure and Applied Mathematics.* Academic Press, Inc., Orlando, FL, 1985.

[55] J. Berstel, D. Perrin, and C. Reutenauer. *Codes and automata*, volume 129. Cambridge University Press, 2010.

[56] M. Bestvina, M. Feighn, and M. Handel. The Tits alternative for $\mathrm{Out}(F_n)$. I. Dynamics of exponentially-growing automorphisms. *Ann. of Math. (2)*, 151(2):517–623, 2000.

[57] V. N. Bezverkhniĭ. Solvability of the inclusion problem in a class of HNN-groups. In *Algorithmic problems of the*

theory of groups and semigroups, pages 20–62. Tulsk. Gos. Ped. Inst., Tula, 1981.

[58] V. N. Bezverkhniĭ. Solution of the occurrence problem in some classes of groups with one defining relation. In *Algorithmic problems in the theory of groups and semigroups (Russian)*, pages 3–21, 126. Tul sk. Gos. Ped. Inst., Tula, 1986.

[59] V. N. Bezverkhniĭ and I. V. Dobrynina. Undecidability of the conjugacy problem for subgroups in the colored braid group R_5. *Mat. Zametki*, 65(1):15–22, 1999.

[60] R. Bieri. Normal subgroups in duality groups and in groups of cohomological dimension 2. *J. Pure Appl. Algebra*, 7(1):35–51, 1976.

[61] R. Bieri and R. Strebel. Valuations and finitely presented metabelian groups. *Proc. London Math. Soc. (3)*, 41(3):439–464, 1980.

[62] S. J. Bigelow. Braid groups are linear. *J. Amer. Math. Soc.*, 14(2):471–486, 2001.

[63] N. Blackburn. Conjugacy in nilpotent groups. *Proc. Amer. Math. Soc.*, 16:143–148, 1965.

[64] L. A. Bokut and G. P. Kukin. Undecidable algorithmic problems for semigroups, groups and rings. In *Algebra. Topology. Geometry, Vol. 25 (Russian)*, Itogi Nauki i Tekhniki, pages 3–66. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1987. Translated in J. Soviet Math. **45** (1989), no. 1, 871–911.

[65] L. A. Bokut and G. P. Kukin. *Algorithmic and combinatorial algebra*, volume 255 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994.

[66] R. V. Book. Confluent and other types of Thue systems. *J. Assoc. Comput. Mach.*, 29(1):171–182, 1982.

[67] R. V. Book. Decidable sentences of Church-Rosser congruences. *Theoret. Comput. Sci.*, 24(3):301–312, 1983.

[68] R. V. Book. Thue systems as rewriting systems. volume 3, pages 39–68. 1987. Rewriting techniques and applications (Dijon, 1985).

[69] R. V. Book, M. Jantzen, and C. Wrathall. Monadic Thue systems. *Theoret. Comput. Sci.*, 19(3):231–251, 1982.

[70] R. V. Book and F. Otto. Cancellation rules and extended word problems. *Inform. Process. Lett.*, 20(1):5–11, 1985.

[71] R. V. Book and F. Otto. *String-rewriting systems.* Texts and Monographs in Computer Science. Springer-Verlag, New York, 1993.

[72] W. W. Boone, D. J. Collins, and Y. V. Matijasevič. Embeddings into semigroups with only a few defining relations. In *Proceedings of the Second Scandinavian Logic Symposium (Univ. Oslo, Oslo, 1970)*, pages 27–40. Studies in Logic and the Foundations of Mathematics, Vol. 63, 1971.

[73] V. V. Borisov. Simple examples of groups with unsolvable word problem. *Mat. Zametki*, 6:521–532, 1969.

[74] B. H. Bowditch. *A course on geometric group theory*, volume 16 of *MSJ Memoirs.* Mathematical Society of Japan, Tokyo, 2006.

[75] N. Brady, T. Riley, and H. Short. *The geometry of the word problem for finitely generated groups.* Advanced Courses in Mathematics. CRM Barcelona. Birkhäuser Verlag, Basel, 2007. Papers from the Advanced Course held in Barcelona, July 5–15, 2005.

[76] M. R. Bridson. The geometry of the word problem. In *Invitations to geometry and topology*, volume 7 of *Oxf. Grad. Texts Math.*, pages 29–91. Oxford Univ. Press, Oxford, 2002.

[77] M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 1999.

[78] B. Brodda. A variant of the Friant grammar for generating prime numbers. *Bull. Math. Soc. Sci. Math. R. S. Roumanie (N.S.)*, 12(60)(4):19–23 (1969), 1968.

[79] S. D. Brodskiĭ and J. Howie. One-relator products of torsion-free groups. *Glasgow Math. J.*, 35(1):99–104, 1993.

[80] T. Brough, A. J. Cain, and M. Pfeiffer. Context-free word problem semigroups. In *Developments in language theory*, volume 11647 of *Lecture Notes in Comput. Sci.*, pages 292–305. Springer, Cham, 2019.

[81] A. M. Brunner. A group with an infinite number of Nielsen inequivalent one-relator presentations. *J. Algebra*, 42(1):81–84, 1976.

[82] A. M. Brunner, R. G. Burns, and D. Solitar. The subgroup separability of free products of two free groups with cyclic amalgamation. In *Contributions to group theory*, volume 33 of *Contemp. Math.*, pages 90–115. Amer. Math. Soc., Providence, RI, 1984.

[83] W. Bucher. A note on regular classes in special Thue systems. *Discrete Appl. Math.*, 21(3):199–205, 1988.

[84] H. Bücken. Reduction systems and small cancellation theory. In *Proc. 4th Workshop on Automated Deduction*, pages 53–59, 1979.

[85] M. Burger and S. Mozes. Finitely presented simple groups and products of trees. *C. R. Acad. Sci. Paris Sér. I Math.*, 324(7):747–752, 1997.

[86] R. G. Burns, A. Karrass, and D. Solitar. A note on groups with separable finitely generated subgroups. *Bull. Austral. Math. Soc.*, 36(1):153–160, 1987.

[87] W. Burnside. *Theory of groups of finite order.* Cambridge University Press, 1911. 1st ed.

[88] S. Burris and H. P. Sankappanavar. *A course in universal algebra*, volume 78 of *Graduate Texts in Mathematics.* Springer-Verlag, New York-Berlin, 1981.

[89] J. O. Button. Large groups of deficiency 1. *Israel J. Math.*, 167:111–140, 2008.

[90] P. Butzbach. Une famille de congruénces de Thue pour lesquelles le problème de l'équivalence est décidable. Application à l'équivalence des grammaires séparées. In *Automata, languages and programming (Proc. Sympos., Rocquencourt, 1972)*, pages 3–12, 1973.

[91] P. Butzbach. Sur l'équivalence des grammaires simples. In *Langages algébriques (Proc. First Meeting, Information Theory, Bonascre, 1973)*, pages 223–245. École Nat. Sup. Tech. Avancées, Paris, 1978.

[92] M. Cadilhac, D. Chistikov, and G. Zetzsche. Rational Subsets of Baumslag-Solitar Groups. In A. Czumaj, A. Dawar, and E. Merelli, editors, *47th International Colloquium on Automata, Languages, and Programming (ICALP 2020)*, volume 168 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 116:1–116:16, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[93] J.-y. Cai. Parallel computation over hyperbolic groups. In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '92, page 106–115, New York, NY, USA, 1992. Association for Computing Machinery.

[94] A. J. Cain and V. Maltcev. Decision problems for finitely presented and one-relation semigroups and monoids. *Internat. J. Algebra Comput.*, 19(6):747–770, 2009.

[95] A. J. Cain and V. Maltcev. Context-free rewriting systems and word-hyperbolic structures with uniqueness. *Internat. J. Algebra Comput.*, 22(7), 2012.

[96] A. J. Cain and V. Maltcev. Hopfian and co-Hopfian subsemigroups and extensions. *Demonstr. Math.*, 47(4):791–804, 2014.

[97] A. J. Cain and V. Maltcev. Growths of endomorphisms of finitely generated semigroups. *J. Aust. Math. Soc.*, 102(2):163–184, 2017.

[98] A. J. Cain and M. Pfeiffer. Decision problems for word-hyperbolic semigroups. *J. Algebra*, 465:287–321, 2016.

[99] R. Campbell. Residual finiteness results. *Residually Finite Groups (available at http://www.math.umbc.edu/∼campbell/CombGpThy/RF_Thesis/2_RF_Results.html, accessed 14 May 2021)*, 1989.

[100] J. W. Cannon. Geometric group theory. In *Handbook of geometric topology*, pages 261–305. North-Holland, Amsterdam, 2002.

[101] A.-C. Caron. Linear bounded automata and rewrite systems: influence of initial configurations on decision properties. In *TAPSOFT '91, Vol. 1 (Brighton, 1991)*, volume 493 of *Lecture Notes in Comput. Sci.*, pages 74–89. Springer, Berlin, 1991.

[102] M. Casals-Ruiz and I. V. Kazachkov. On systems of equations over free products of groups. *J. Algebra*, 333:368–426, 2011.

[103] J. Cassaigne, T. Harju, and J. Karhumäki. On the undecidability of freeness of matrix semigroups. volume 9, pages 295–305. 1999. Dedicated to the memory of Marcel-Paul Schützenberger.

[104] D. Caucal. A fast algorithm to decide on simple grammars equivalence. In *Optimal algorithms (Varna, 1989)*, volume 401 of *Lecture Notes in Comput. Sci.*, pages 66–85. Springer, Berlin, 1989.

[105] D. Caucal. A fast algorithm to decide on the equivalence of stateless DPDA. *RAIRO Inform. Théor. Appl.*, 27(1):23–48, 1993.

[106] J. Chalopin and V. Chepoi. A counterexample to Thiagarajan's conjecture on regular event structures. In *44th International Colloquium on Automata, Languages, and Programming*, volume 80 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 101, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017.

[107] B. Chandler and W. Magnus. *The history of combinatorial group theory*, volume 9 of *Studies in the History of Mathematics and Physical Sciences.* Springer-Verlag, New York, 1982. A case study in the history of ideas.

[108] R. Charney. An introduction to right-angled Artin groups. *Geom. Dedicata*, 125:141–158, 2007.

[109] N. Chomsky and M. P. Schützenberger. The algebraic theory of context-free languages. In *Computer programming and formal systems*, pages 118–161. North-Holland, Amsterdam, 1963.

[110] L. Chottin. Strict deterministic languages and controlled rewriting systems. In *Automata, languages and programming (Sixth Colloq., Graz, 1979)*, volume 71 of *Lecture Notes in Comput. Sci.*, pages 104–117. Springer, Berlin-New York, 1979.

[111] L. Chottin. Langages algébriques et systèmes de réécriture rationnels. *RAIRO Inform. Théor.*, 16(2):93–112, 1982.

[112] A. H. Clifford and G. B. Preston. *The algebraic theory of semigroups. Vol. I.* Mathematical Surveys, No. 7. American Mathematical Society, Providence, R.I., 1961.

[113] A. H. Clifford and G. B. Preston. *The algebraic theory of semigroups. Vol. II.* Mathematical Surveys, No. 7. American Mathematical Society, Providence, R.I., 1967.

[114] Y. Cochet. Langages définis par des congruences. In *Groupe d'Étude d'Algèbre (Marie-Paule Malliavin), 1re année (1975/76)*, pages Exp. No. 16, 7. Secrétariat Math., Paris, 1978.

[115] Y. Cochet. Church-Rosser congruences on free semigroups. In *Algebraic theory of semigroups (Proc. Sixth Algebraic Conf., Szeged, 1976)*, volume 20 of *Colloq. Math. Soc. János Bolyai*, pages 51–60. North-Holland,

Amsterdam-New York, 1979.

[116] Y. Cochet and M. Nivat. Une généralisation des ensembles de Dyck. *Israel J. Math.*, 9:389–395, 1971.

[117] P. Cockshott, L. Mackenzie, and G. Michaelson. Physical constraints on hypercomputation. *Theoret. Comput. Sci.*, 394(3):159–174, 2008.

[118] D. E. Cohen. *Groups of cohomological dimension one.* Lecture Notes in Mathematics, Vol. 245. Springer-Verlag, Berlin-New York, 1972.

[119] D. J. Collins and F. Levin. Automorphisms and Hopficity of certain Baumslag-Solitar groups. *Arch. Math. (Basel)*, 40(5):385–400, 1983.

[120] D. J. Collins and C. F. Miller, III. The conjugacy problem and subgroups of finite index. *Proc. London Math. Soc. (3)*, 34(3):535–556, 1977.

[121] A. Connes, A. Lichnerowicz, and M. P. Schützenberger. *Triangle of thoughts.* American Mathematical Society, Providence, RI, 2001. Translated from the 2000 French original by Jennifer Gage.

[122] M. Coornaert, T. Delzant, and A. Papadopoulos. *Géométrie et théorie des groupes*, volume 1441 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1990. Les groupes hyperboliques de Gromov. [Gromov hyperbolic groups], With an English summary.

[123] B. J. Copeland. Hypercomputation: philosophical issues. *Theoret. Comput. Sci.*, 317(1-3):251–267, 2004.

[124] A. L. S. Corner. Three examples of hopficity in torsion-free abelian groups. *Acta Math. Acad. Sci. Hungar.*, 16:303–310, 1965.

[125] S. Craik, R. Gray, V. Kilibarda, J. D. Mitchell, and N. Ruškuc. Ends of semigroups. *Semigroup Forum*, 93(2):330–346, 2016.

[126] F. Dahmani and V. Guirardel. Foliations for solving equations in groups: free, virtually free, and hyperbolic groups. *J. Topol.*, 3(2):343–404, 2010.

[127] Y. de Cornulier. Finitely presentable, non-Hopfian groups with Kazhdan's property (T) and infinite outer automorphism group. *Proc. Amer. Math. Soc.*, 135(4):951–959, 2007.

[128] P. de la Harpe. *Topics in geometric group theory.* Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 2000.

[129] M. Dehn. Über die Topologie des dreidimensionalen Raumes. *Math. Ann.*, 69(1):137–168, 1910.

[130] M. Dehn. Über unendliche diskontinuierliche Gruppen. *Math. Ann.*, 71(1):116–144, 1911.

[131] M. Dehn. Transformation der Kurven auf zweiseitigen Flächen. *Math. Ann.*, 72(3):413–421, 1912.

[132] N. Dershowitz and Y. Gurevich. A natural axiomatization of computability and proof of Church's Thesis. *Bull. Symbolic Logic*, 14(3):299–350, 2008.

[133] E. Deutsch. Dyck path enumeration. *Discrete Math.*, 204(1-3):167–202, 1999.

[134] I. M. S. Dey and H. Neumann. The Hopf property of free products. *Math. Z.*, 117:325–339, 1970.

[135] V. Diekert. Some remarks on presentations by finite Church-Rosser Thue systems. In *STACS 87 (Passau, 1987)*, volume 247 of *Lecture Notes in Comput. Sci.*, pages 272–285. Springer, Berlin, 1987.

[136] V. Diekert and M. Elder. Solutions of twisted word equations, EDT0L languages, and context-free groups. In *44th International Colloquium on Automata, Languages, and Programming*, volume 80 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 96, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017.

[137] V. Diekert and M. Lohrey. Word equations over graph products. *Internat. J. Algebra Comput.*, 18(3):493–533, 2008.

[138] V. Diekert and A. Muscholl. Solvability of equations in free partially commutative groups is decidable. In *Automata, languages and programming*, volume 2076 of *Lecture Notes in Comput. Sci.*, pages 543–554. Springer, Berlin, 2001.

[139] V. Diekert and A. Weiß. *Context-Free Groups and Bass–Serre Theory*, pages 43–110. Springer International Publishing, Cham, 2017.

[140] R. Diestel. The end structure of a graph: recent results and open problems. volume 100, pages 313–327. 1992. Special volume to mark the centennial of Julius Petersen's "Die Theorie der regulären Graphs", Part I.

[141] I. Dolinka and R. D. Gray. New results on the prefix membership problem for one-relator groups. *Trans. Amer. Math. Soc.*, 374(6):4309–4358, 2021.

[142] C. Droms. Graph groups, coherence, and three-manifolds. *J. Algebra*, 106(2):484–489, 1987.

[143] C. Druţu and M. Kapovich. *Geometric group theory*, volume 63 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2018. With an appendix by Bogdan Nica.

[144] A. Duncan and R. H. Gilman. Word hyperbolic semigroups. *Math. Proc. Cambridge Philos. Soc.*, 136(3):513–524, 2004.

[145] W. Dyck. Gruppentheoretische Studien. *Math. Ann.*, 20(1):1–44, 1882.

[146] W. Dyck. Gruppentheoretische Studien. II. Ueber die Zusammensetzung einer Gruppe discreter Operationen, über ihre Primitivität und Transitivität. *Math. Ann.*, 22(1):70–108, 1883.

[147] D. Easdown and L. M. Shneerson. Growth of Rees quotients of free inverse semigroups defined by small numbers

of relators. *Internat. J. Algebra Comput.*, 23(3):521–545, 2013.

[148] S. Eilenberg and M. P. Schützenberger. Rational sets in commutative monoids. *J. Algebra*, 13:173–191, 1969.

[149] C. C. Elgot and M. O. Rabin. Decidability and undecidability of extensions of second (first) order theory of (generalized) successor. *The Journal of Symbolic Logic*, 31(2):169–181, 1966.

[150] J. Engelfriet. Hierarchies of hyper-AFLs. *J. Comput. System Sci.*, 30(1):86–115, 1985.

[151] D. B. A. Epstein, J. W. Cannon, D. F. Holt, S. V. F. Levy, M. S. Paterson, and W. P. Thurston. *Word processing in groups.* Jones and Bartlett Publishers, Boston, MA, 1992.

[152] G. Etesi and I. Németi. Non-Turing computations via Malament-Hogarth space-times. *Internat. J. Theoret. Phys.*, 41(2):341–370, 2002.

[153] B. Fine and G. Rosenberger. The Freiheitssatz and its extensions. In *The mathematical legacy of Wilhelm Magnus: groups, geometry and special functions (Brooklyn, NY, 1992)*, volume 169 of *Contemp. Math.*, pages 213–252. Amer. Math. Soc., Providence, RI, 1994.

[154] J. Fischer, A. Karrass, and D. Solitar. On one-relator groups having elements of finite order. *Proc. Amer. Math. Soc.*, 33:297–301, 1972.

[155] J. Folina. Church's thesis and the variety of mathematical justifications. In *Church's thesis after 70 years*, volume 1 of *Ontos Math. Log.*, pages 220–241. Ontos Verlag, Heusenstamm, 2006.

[156] S. Friedl and H. Wilton. The membership problem for 3-manifold groups is solvable. *Algebr. Geom. Topol.*, 16(4):1827–1850, 2016.

[157] A. Garreta and R. D. Gray. Equations and first-order theory of one-relator and word-hyperbolic monoids, 2019.

[158] F. A. Garside. The braid group and other groups. *Quart. J. Math. Oxford Ser. (2)*, 20:235–254, 1969.

[159] S. M. Gersten. Isoperimetric and isodiametric functions of finite presentations. In *Geometric group theory, Vol. 1 (Sussex, 1991)*, volume 181 of *London Math. Soc. Lecture Note Ser.*, pages 79–96. Cambridge Univ. Press, Cambridge, 1993.

[160] S. M. Gersten. Quadratic divergence of geodesics in $\mathrm{CAT}(0)$ spaces. *Geom. Funct. Anal.*, 4(1):37–51, 1994.

[161] E. Ghys and P. de la Harpe, editors. *Sur les groupes hyperboliques d'après Mikhael Gromov*, volume 83 of *Progress in Mathematics.* Birkhäuser Boston, Inc., Boston, MA, 1990. Papers from the Swiss Seminar on Hyperbolic Groups held in Bern, 1988.

[162] R. H. Gilman, R. P. Kropholler, and S. Schleimer. Groups whose word problems are not semilinear. *Groups Complex. Cryptol.*, 10(2):53–62, 2018.

[163] S. Ginsburg, S. A. Greibach, and M. A. Harrison. One-way stack automata. *J. Assoc. Comput. Mach.*, 14:389–418, 1967.

[164] C. M. Gordon. Some embedding theorems and undecidability questions for groups. In *Combinatorial and geometric group theory (Edinburgh, 1993)*, volume 204 of *London Math. Soc. Lecture Note Ser.*, pages 105–110. Cambridge Univ. Press, Cambridge, 1995.

[165] R. Gray and M. Kambites. A Švarc-Milnor lemma for monoids acting by isometric embeddings. *Internat. J. Algebra Comput.*, 21(7):1135–1147, 2011.

[166] R. Gray and M. Kambites. Groups acting on semimetric spaces and quasi-isometries of monoids. *Trans. Amer. Math. Soc.*, 365(2):555–578, 2013.

[167] R. Gray and N. Ruškuc. Generators and relations for subsemigroups via boundaries in Cayley graphs. *J. Pure Appl. Algebra*, 215(11):2761–2779, 2011.

[168] R. D. Gray. Undecidability of the word problem for one-relator inverse monoids via right-angled Artin subgroups of one-relator groups. *Invent. Math.*, 219(3):987–1008, 2020.

[169] R. D. Gray and N. Ruškuc. On groups of units of special and one-relator inverse monoids. *Pre-print*, 2021. Available at arXiv:2103.02995.

[170] R. D. Gray, P. V. Silva, and N. Szakács. Algorithmic properties of inverse monoids with hyperbolic and tree-like Schützenberger graphs. *Pre-print*, 2021. Available at arXiv:1912.00950.

[171] R. D. Gray and B. Steinberg. Topological finiteness properties of monoids. part 2: special monoids, one-relator monoids, amalgamated free products, and HNN extensions. *Pre-print*, 2018. Available at arXiv:1805.03413.

[172] R. D. Gray and B. Steinberg. A Lyndon's identity theorem for one-relator monoids. *Pre-print*, 2019. Available at arXiv:1910.09914.

[173] E. Green. *Graph products.* PhD thesis, Univ. of Warwick, 1991.

[174] J. A. Green. On the structure of semigroups. *Ann. of Math. (2)*, 54:163–172, 1951.

[175] L. Greenberg. Discrete groups of motions. *Canadian J. Math.*, 12:415–426, 1960.

[176] M. Greendlinger. *Dehn's algorithm for the word problem.* ProQuest LLC, Ann Arbor, MI, 1960. Thesis (Ph.D.)–New York University.

[177] M. Greendlinger. Dehn's algorithm for the word problem. *Comm. Pure Appl. Math.*, 13:67–83, 1960.

[178] M. Greendlinger. On Dehn's algorithms for the conjugacy and word problems, with applications. *Comm. Pure Appl. Math.*, 13:641–677, 1960.

[179] M. D. Greendlinger. On Magnus's generalized word problem. *Sibirsk. Mat. Ž.*, 5:955–957, 1964.

[180] S. A. Greibach. Full AFLs and nested iterated substitution. *Information and Control*, 16:7–35, 1970.

[181] R. Grigorchuk and I. Pak. Groups of intermediate growth: an introduction. *Enseign. Math. (2)*, 54(3-4):251–272, 2008.

[182] R. I. Grigorchuk. Degrees of growth of finitely generated groups and the theory of invariant means. *Izv. Akad. Nauk SSSR Ser. Mat.*, 48(5):939–985, 1984.

[183] R. I. Grigorchuk. Degrees of growth of p-groups and torsion-free groups. *Mat. Sb. (N.S.)*, 126(168)(2):194–214, 286, 1985.

[184] R. I. Grigorčuk. On Burnside's problem on periodic groups. *Funktsional. Anal. i Prilozhen.*, 14(1):53–54, 1980.

[185] G. R. Grimmett and Z. Li. Connective constants and height functions for Cayley graphs. *Trans. Amer. Math. Soc.*, 369(8):5961–5980, 2017.

[186] G. R. Grimmett and Z. Li. Self-avoiding walks and amenability. *Electron. J. Combin.*, 24(4):Paper No. 4.38, 24, 2017.

[187] M. Gromov. Groups of polynomial growth and expanding maps. *Inst. Hautes Études Sci. Publ. Math.*, (53):53–73, 1981.

[188] M. Gromov. Hyperbolic groups. In *Essays in group theory*, volume 8 of *Math. Sci. Res. Inst. Publ.*, pages 75–263. Springer, New York, 1987.

[189] M. Gromov. Asymptotic invariants of infinite groups. In *Geometric group theory, Vol. 2 (Sussex, 1991)*, volume 182 of *London Math. Soc. Lecture Note Ser.*, pages 1–295. Cambridge Univ. Press, Cambridge, 1993.

[190] Z. Grunschlag. *Algorithms in geometric group theory.* ProQuest LLC, Ann Arbor, MI, 1999. Thesis (Ph.D.)–University of California, Berkeley.

[191] V. Guba and M. Sapir. Diagram groups. *Mem. Amer. Math. Soc.*, 130(620):viii+117, 1997.

[192] Y. Guivarc'h. Croissance polynomiale et périodes des fonctions harmoniques. *Bull. Soc. Math. France*, 101:333–379, 1973.

[193] Y. Gurevich. Unconstrained Church-Turing thesis cannot possibly be true. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, (127):46–59, 2019.

[194] G. A. Gurevič. On the conjugacy problem for groups with one defining relation. *Dokl. Akad. Nauk SSSR*, 207:18–20, 1972.

[195] E. Gwynne and J. Miller. Convergence of the self-avoiding walk on random quadrangulations to $\mathrm{SLE}_{8/3}$ on $\sqrt{8/3}$-Liouville quantum gravity. *Ann. Sci. Éc. Norm. Supér. (4)*, 54(2):305–405, 2021.

[196] F. Haglund and D. T. Wise. Special cube complexes. *Geom. Funct. Anal.*, 17(5):1551–1620, 2008.

[197] M. Hall, Jr. Coset representations in free groups. *Trans. Amer. Math. Soc.*, 67:421–432, 1949.

[198] M. Hall, Jr. The word problem for semigroups with two generators. *J. Symbolic Logic*, 14:115–118, 1949.

[199] P. Hall. Some word-problems. *J. London Math. Soc.*, 33:482–496, 1958.

[200] E. Hamilton, H. Wilton, and P. A. Zalesskii. Separability of double cosets and conjugacy classes in 3-manifold groups. *J. Lond. Math. Soc. (2)*, 87(1):269–288, 2013.

[201] W. R. Hamilton. LVI. memorandum respecting a new system of roots of unity. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 12(81):446–446, 1856.

[202] J. D. Hamkins and A. Lewis. Infinite time Turing machines. *J. Symbolic Logic*, 65(2):567–604, 2000.

[203] M. A. Harrison. *Introduction to formal language theory.* Addison-Wesley Publishing Co., Reading, Mass., 1978.

[204] A. Hernández-Espinosa, F. Hernández-Quiroz, and H. Zenil. Is there any real substance to the claims for a 'new computationalism'? In *Unveiling dynamics and complexity*, volume 10307 of *Lecture Notes in Comput. Sci.*, pages 14–23. Springer, Cham, 2017.

[205] P. M. Higgins. *Techniques of semigroup theory.* Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1992. With a foreword by G. B. Preston.

[206] G. Higman. A finitely related group with an isomorphic proper factor group. *J. London Math. Soc.*, 26:59–61, 1951.

[207] K. A. Hirsch. Über lokal-nilpotente Gruppen. *Math. Z.*, 63:290–294, 1955.

[208] R. Hirshon. Some theorems on hopficity. *Trans. Amer. Math. Soc.*, 141:229–244, 1969.

[209] M. Hoffmann, D. F. Holt, M. D. Owens, and R. M. Thomas. Semigroups with a context-free word problem. In *Developments in language theory*, volume 7410 of *Lecture Notes in Comput. Sci.*, pages 97–108. Springer, Heidelberg, 2012.

[210] M. Hogarth. Non-Turing Computers and Non-Turing Computability. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1(1):126–138, 1994.

[211] M. L. Hogarth. Does general relativity allow an observer to view an eternity in a finite time? *Found. Phys. Lett.*, 5(2):173–181, 1992.

[212] C. Hollings. The early development of the algebraic theory of semigroups. *Arch. Hist. Exact Sci.*, 63(5):497–536, 2009.

[213] C. Hollings. Investigating a claim for Russian priority in the abstract definition of a ring. *BSHM Bull.*, 29(2):111–119, 2014.

[214] C. Hollings. *Mathematics across the Iron Curtain*, volume 41 of *History of Mathematics*. American Mathematical Society, Providence, RI, 2014. A history of the algebraic theory of semigroups.

[215] C. Hollings. The acceptance of abstract algebra in the USSR, as viewed through periodic surveys of the progress of Soviet mathematical science. *Historia Math.*, 42(2):193–222, 2015.

[216] C. D. Hollings. 'Nobody could possibly misunderstand what a group is': a study in early twentieth-century group axiomatics. *Arch. Hist. Exact Sci.*, 71(5):409–481, 2017.

[217] J. E. Hopcroft and J. D. Ullman. *Introduction to automata theory, languages, and computation.* Addison-Wesley Publishing Co., Reading, Mass., 1979. Addison-Wesley Series in Computer Science.

[218] J. Howie. Lecture notes on hyperbolic groups, 1999. Lectures given at the summer school 'Groups and Applications'.

[219] J. Howie. Erratum: "Some results on one-relator surface groups" [Bol. Soc. Mat. Mexicana (3) **10** (2004), Special Issue, 255–262; mr2199352]. *Bol. Soc. Mat. Mexicana (3)*, 10(Special Issue):545–546, 2004.

[220] J. Howie. Some results on one-relator surface groups. *Bol. Soc. Mat. Mexicana (3)*, 10(Special Issue):255–262, 2004.

[221] J. Howie and S. J. Pride. The word problem for one-relator semigroups. *Math. Proc. Cambridge Philos. Soc.*, 99(1):33–44, 1986.

[222] J. M. Howie. *An introduction to semigroup theory.* L. M. S. Monographs, No. 7. Academic Press [Harcourt Brace Jovanovich, Publishers], London-New York, 1976.

[223] J. M. Howie. Why study semigroups? *Math. Chronicle*, 16:1–14, 1987.

[224] J. M. Howie. *Fundamentals of semigroup theory*, volume 12 of *London Mathematical Society Monographs. New Series.* The Clarendon Press, Oxford University Press, New York, 1995. Oxford Science Publications.

[225] A. G. Howson. On the intersection of finitely generated free groups. *J. London Math. Soc.*, 29:428–434, 1954.

[226] S. P. Humphries. On representations of Artin groups and the Tits conjecture. *J. Algebra*, 169(3):847–862, 1994.

[227] T. Hutchcroft. Self-avoiding walk on nonunimodular transitive graphs. *Ann. Probab.*, 47(5):2801–2829, 2019.

[228] T. Hutchcroft. Statistical physics on a product of trees. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(2):1001–1010, 2019.

[229] W. Imrich and N. Seifter. A bound for groups of linear growth. *Arch. Math. (Basel)*, 48(2):100–104, 1987.

[230] S. V. Ivanov, S. W. Margolis, and J. C. Meakin. On one-relator inverse monoids and one-relator groups. *J. Pure Appl. Algebra*, 159(1):83–111, 2001.

[231] D. A. Jackson and V. Kilibarda. Ends for monoids and semigroups. *J. Aust. Math. Soc.*, 87(1):101–127, 2009.

[232] W. Jaco. On certain subgroups of the fundamental group of a closed surface. *Proc. Cambridge Philos. Soc.*, 67:17–18, 1970.

[233] W. Jaco. *Lectures on three-manifold topology*, volume 43 of *CBMS Regional Conference Series in Mathematics.* American Mathematical Society, Providence, R.I., 1980.

[234] M. Jantzen. On a special monoid with a single defining relation. *Theoret. Comput. Sci.*, 16(1):61–73, 1981.

[235] M. Jantzen. *Confluent string rewriting*, volume 14 of *EATCS Monographs on Theoretical Computer Science.* Springer-Verlag, Berlin, 1988.

[236] J. Jaynes. *The origin of consciousness in the breakdown of the bicameral mind.* Houghton Mifflin, Boston, 1976.

[237] A. Juhász. Solution of the conjugacy problem in one-relator groups. In *Algorithms and classification in combinatorial group theory (Berkeley, CA, 1989)*, volume 23 of *Math. Sci. Res. Inst. Publ.*, pages 69–81. Springer, New York, 1992.

[238] M. Kambites, P. V. Silva, and B. Steinberg. On the rational subset problem for groups. *J. Algebra*, 309(2):622–639, 2007.

[239] I. Kapovich, R. Weidmann, and A. Miasnikov. Foldings, graphs of groups and the membership problem. *Internat. J. Algebra Comput.*, 15(1):95–128, 2005.

[240] M. Kapovich. Representations of polygons of finite groups. *Geom. Topol.*, 9:1915–1951, 2005.

[241] D. Kapur and P. Narendran. A finite Thue system with decidable word problem and without equivalent finite canonical system. *Theoret. Comput. Sci.*, 35(2-3):337–344, 1985.

[242] A. Karrass and D. Solitar. On the failure of the Howson property for a group with a single defining relation. *Math. Z.*, 108:235–236, 1969.

[243] A. Karrass and D. Solitar. The subgroups of a free product of two groups with an amalgamated subgroup. *Trans. Amer. Math. Soc.*, 150:227–255, 1970.

[244] A. Karrass and D. Solitar. Subgroups of HNN groups and groups with one defining relation. *Canadian J. Math.*, 23:627–643, 1971.

[245] E. V. Kashintsev. On the word problem. *Tul. Gos. Ped. Inst. Učen. Zap. Mat. Kaf.*, (Vyp. 2 Geometr. i Algebra):185–214, 1970.

[246] E. V. Kashintsev. An algorithm for the solution of the conjugacy problem for certain semigroups. In *Recursive functions (Russian)*, pages 18–26. Ivanov. Gos. Univ., Ivanovo, 1978.

[247] E. V. Kashintsev. On the word problem for special semigroups. *Izv. Akad. Nauk SSSR Ser. Mat.*, 42(6):1401–1416, 1440, 1978.

[248] E. V. Kashintsev. Small cancellation conditions and embeddability of semigroups in groups. *Internat. J. Algebra Comput.*, 2(4):433–441, 1992.

[249] E. V. Kashintsev. On the satisfiability of the conditions $C'(\frac{1}{3})$ and $C(4)$ for special homogeneous semigroups with defining words-degrees. *Mat. Zametki*, 54(3):40–47, 158, 1993.

[250] A. Kelarev. *Graph algebras and automata*, volume 257 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker, Inc., New York, 2003.

[251] A. Kelarev, J. Ryan, and J. Yearwood. Cayley graphs as classifiers for data mining: the influence of asymmetries. *Discrete Math.*, 309(17):5360–5369, 2009.

[252] O. Kharlampovich, L. Lopez, and A. Miasnikov. Diophantine problem in some metabelian groups. *Pre-print*, 2019. Available at arXiv:1903.10068.

[253] O. G. Kharlampovič. A finitely presented solvable group with unsolvable word problem. *Izv. Akad. Nauk SSSR Ser. Mat.*, 45(4):852–873, 928, 1981.

[254] A. Khukhro. A characterisation of virtually free groups via minor exclusion, 2020. Preprint available at arXiv:2006.16918.

[255] V. Kilibarda. *On the algebra of semigroup diagrams*. PhD thesis, 1994. Thesis (Ph.D.)–The University of Nebraska - Lincoln.

[256] V. Kilibarda, V. Maltcev, and S. Craik. Ends for subsemigroups of finite index. *Semigroup Forum*, 91(2):401–414, 2015.

[257] S.-h. Kim and T. Koberda. Embedability between right-angled Artin groups. *Geom. Topol.*, 17(1):493–530, 2013.

[258] J. Kirby. *An invitation to model theory*. Cambridge University Press, Cambridge, 2019.

[259] D. A. Klarner, J.-C. Birget, and W. Satterfield. On the undecidability of the freeness of integer matrix semigroups. *Internat. J. Algebra Comput.*, 1(2):223–226, 1991.

[260] F. Klein. Ueber eindeutige Functionen mit linearen Transformationen in sich. *Math. Ann.*, 19(4):565–568, 1882.

[261] F. Klein. *Lectures on the development of mathematics in the XIXth century*. Springer, Berlin, 1926.

[262] I. Kleiner. The evolution of group theory: a brief survey. *Math. Mag.*, 59(4):195–215, 1986.

[263] Y. Kobayashi. A finitely presented monoid which has solvable word problem but has no regular complete presentation. *Theoret. Comput. Sci.*, 146(1-2):321–329, 1995.

[264] Y. Kobayashi. Finite homotopy bases of one-relator monoids. *J. Algebra*, 229(2):547–569, 2000.

[265] J. Král. A modification of a substitution theorem and some necessary and sufficient conditions for sets to be context-free. *Math. Systems Theory*, 4:129–139, 1970.

[266] S. A. Kripke. The Church-Turing "thesis" as a special corollary of Gödel's completeness theorem. In *Computability—Turing, Gödel, Church, and beyond*, pages 77–104. MIT Press, Cambridge, MA, 2013.

[267] A. G. Kurosh. *Teoriya Grupp*. OGIZ, Moscow-Leningrad, Russia, 1944.

[268] D. Kuske and M. Lohrey. Decidable theories of Cayley-graphs. In *STACS 2003*, volume 2607 of *Lecture Notes in Comput. Sci.*, pages 463–474. Springer, Berlin, 2003.

[269] D. Kuske and M. Lohrey. Logical aspects of Cayley-graphs: the group case. *Ann. Pure Appl. Logic*, 131(1-3):263–286, 2005.

[270] D. Kuske and M. Lohrey. Logical aspects of Cayley-graphs: the monoid case. *Internat. J. Algebra Comput.*, 16(2):307–340, 2006.

[271] A. V. Kuznetsov. Algorithms as operations in algebraic systems. *Izv. Akad. Nauk SSSR Ser. Mat.*, 1983.

[272] I. Lakatos. *Philosophical papers. Vol. 2*. Cambridge University Press, Cambridge-New York, 1978. Mathematics, science and epistemology, Edited and with an introduction by John Worrall and Gregory Currie.

[273] G. Lallement. On monoids presented by a single relation. *J. Algebra*, 32:370–388, 1974.

[274] P. Le Chenadec. Canonical forms in finitely presented algebras. In *7th international conference on automated deduction (Napa, Calif., 1984)*, volume 170 of *Lecture Notes in Comput. Sci.*, pages 142–165. Springer, Berlin, 1984.

[275] P. Le Chenadec. *Canonical forms in finitely presented algebras*. Research Notes in Theoretical Computer Science. Pitman Publishing, Ltd., London; John Wiley & Sons, Inc., New York, 1986.

[276] P. Le Chenadec. A catalogue of complete group presentations. *J. Symbolic Comput.*, 2(4):363–381, 1986.

[277] F. Le Gall. Powers of tensors and fast matrix multiplication. In *ISSAC 2014—Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation*, pages 296–303. ACM, New York, 2014.

[278] A. Lentin. Équations dans les monoïdes libres. *Math. Sci. Humaines*, 31:5–16, 1970.

[279] C. I. Lewis. *A survey of symbolic logic*. Berkeley University of California Press, 1918.

[280] Z. Li. Positive speed self-avoiding walks on graphs with more than one end. *J. Combin. Theory Ser. A*, 175:105257, 47, 2020.

[281] C. Lindorfer. A general bridge theorem for self-avoiding walks. *Discrete Math.*, 343(12):112092, 11, 2020.

[282] H.-N. Liu, C. Wrathall, and K. Zeger. Efficient solution of some problems in free partially commutative monoids. *Inform. and Comput.*, 89(2):180–198, 1990.

[283] E. S. Ljapin. *Semigroups.* Gosudarstv. Izdat. Fiz.-Mat. Lit., Moscow, 1960.

[284] M. Lohrey. Decidability and complexity in automatic monoids. *Internat. J. Found. Comput. Sci.*, 16(4):707–722, 2005.

[285] M. Lohrey. The rational subset membership problem for groups: a survey. In *Groups St Andrews 2013*, volume 422 of *London Math. Soc. Lecture Note Ser.*, pages 368–389. Cambridge Univ. Press, Cambridge, 2015.

[286] M. Lohrey and N. Ondrusch. Inverse monoids: decidability and complexity of algebraic questions. *Inform. and Comput.*, 205(8):1212–1234, 2007.

[287] M. Lohrey and G. Sénizergues. Theories of HNN-extensions and amalgamated products. In *Automata, languages and programming. Part II*, volume 4052 of *Lecture Notes in Comput. Sci.*, pages 504–515. Springer, Berlin, 2006.

[288] M. Lohrey and B. Steinberg. The submonoid and rational subset membership problems for graph groups. *J. Algebra*, 320(2):728–755, 2008.

[289] M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics and its Applications.* Addison-Wesley Publishing Co., Reading, Mass., 1983. A collective work by Dominique Perrin, Jean Berstel, Christian Choffrut, Robert Cori, Dominique Foata, Jean Eric Pin, Guiseppe Pirillo, Christophe Reutenauer, Marcel-P. Schützenberger, Jacques Sakarovitch and Imre Simon, With a foreword by Roger Lyndon, Edited and with a preface by Perrin.

[290] M. Lothaire. *Mots.* Langue, Raisonnement, Calcul. [Language, Reasoning, Computation]. Editions Hermès, Paris, 1990. Mélanges offerts à M.-P. Schützenberger. [Miscellany offered to M.-P. Schützenberger].

[291] M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 2002. A collective work by Jean Berstel, Dominique Perrin, Patrice Seebold, Julien Cassaigne, Aldo De Luca, Steffano Varricchio, Alain Lascoux, Bernard Leclerc, Jean-Yves Thibon, Veronique Bruyere, Christiane Frougny, Filippo Mignosi, Antonio Restivo, Christophe Reutenauer, Dominique Foata, Guo-Niu Han, Jacques Desarmenien, Volker Diekert, Tero Harju, Juhani Karhumaki and Wojciech Plandowski, With a preface by Berstel and Perrin.

[292] M. Lothaire. *Applied combinatorics on words*, volume 105 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 2005. A collective work by Jean Berstel, Dominique Perrin, Maxime Crochemore, Eric Laporte, Mehryar Mohri, Nadia Pisanti, Marie-France Sagot, Gesine Reinert, Sophie Schbath, Michael Waterman, Philippe Jacquet, Wojciech Szpankowski, Dominique Poulalhon, Gilles Schaeffer, Roman Kolpakov, Gregory Koucherov, Jean-Paul Allouche and Valérie Berthé, With a preface by Berstel and Perrin.

[293] L. Louder and H. Wilton. One-relator groups with torsion are coherent. *Math. Res. Lett.*, 27(5):1499–1512, 2020.

[294] B. Löwe. Revision sequences and computers with an infinite amount of time. *J. Logic Comput.*, 11(1):25–40, 2001.

[295] Lucretius. *De rerum natura.* I, verses 823–826; II, verses 688–694; III, verses 1013–1018.

[296] R. Lyndon. Problems in combinatorial group theory. In *Combinatorial group theory and topology (Alta, Utah, 1984)*, volume 111 of *Ann. of Math. Stud.*, pages 3–33. Princeton Univ. Press, Princeton, NJ, 1987.

[297] R. C. Lyndon. Cohomology theory of groups with a single defining relation. *Ann. of Math. (2)*, 52:650–665, 1950.

[298] R. C. Lyndon and P. E. Schupp. *Combinatorial group theory.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 89. Springer-Verlag, Berlin-New York, 1977.

[299] K. Madlener, P. Narendran, F. Otto, and L. Zhang. On weakly confluent monadic string-rewriting systems. volume 113, pages 119–165. 1993. 8th Annual Symposium on Theoretical Aspects of Computer Science (STACS 91) (Hamburg, 1991).

[300] W. Magnus. Über diskontinuierliche Gruppen mit einer definierenden Relation. (Der Freiheitssatz). *J. Reine Angew. Math.*, 163:141–165, 1930.

[301] W. Magnus. Das Identitätsproblem für Gruppen mit einer definierenden Relation. *Math. Ann.*, 106(1):295–307, 1932.

[302] W. Magnus. Über freie Faktorgruppen und freie Untergruppen gegebener Gruppen. *Monatsh. Math. Phys.*, 47(1):307–313, 1939.

[303] W. Magnus. Residually finite groups. *Bull. Amer. Math. Soc.*, 75:305–316, 1969.

[304] W. Magnus. Braid groups: a survey. In *Proceedings of the Second International Conference on the Theory of Groups (Australian Nat. Univ., Canberra, 1973)*, pages 463–487. Lecture Notes in Math., Vol. 372, 1974.

[305] W. Magnus. The significance of mathematics: the mathematicians' share in the general human condition. *Amer. Math. Monthly*, 104(3):261–269, 1997.

[306] W. Magnus, A. Karrass, and D. Solitar. *Combinatorial group theory: Presentations of groups in terms of generators and relations.* Interscience Publishers [John Wiley & Sons, Inc.], New York-London-Sydney, 1966.

[307] S. Majumdar. A generalized Freiheitssatz. *International Atomic Energy Agency (IAEA)*, 88:245), 1988.

[308] G. S. Makanin. *On the Identity Problem for Finitely Presented Groups and Semigroups.* PhD thesis, Steklov

Mathematical Institute, Moscow, 1966.

[309] G. S. Makanin. On the identity problem in finitely defined semigroups. *Dokl. Akad. Nauk SSSR*, 171:285–287, 1966.

[310] G. S. Makanin. The problem of the solvability of equations in a free semigroup. *Mat. Sb. (N.S.)*, 103(145)(2):147–236, 319, 1977.

[311] G. S. Makanin. Equations in a free group. *Izv. Akad. Nauk SSSR Ser. Mat.*, 46(6):1199–1273, 1344, 1982.

[312] G. S. Makanin. Finite parametrization of solutions of equations in a free monoid. I. *Mat. Sb.*, 195(2):41–90, 2004.

[313] G. S. Makanin. Finite parametrization of solutions of equations in a free monoid. II. *Mat. Sb.*, 195(4):65–96, 2004.

[314] G. S. Makanin. Parametrization of solutions of the equation $x_1x_2\ldots x_{n-1}x_n = x_nx_{n-1}\ldots x_2x_1$ in a free monoid. *Mat. Zametki*, 89(6):879–884, 2011.

[315] G. S. Makanin, H. Abdulrab, and P. Goralcik. Functions for the general solution of parametric word equations. In *Logical foundations of computer science (Yaroslavl, 1997)*, volume 1234 of *Lecture Notes in Comput. Sci.*, pages 189–202. Springer, Berlin, 1997.

[316] G. S. Makanin and T. A. Makanina. Parametrization of solutions of some equations of squares in a free monoid. *Diskret. Mat.*, 11(3):133–148, 1999.

[317] G. S. Makanin and T. A. Makanina. Functions for parametrization of solutions of an equation in a free monoid. *Trans. Amer. Math. Soc.*, 352(1):1–54, 2000.

[318] G. S. Makanin and T. A. Makanina. Parametrisation of solutions of parametric equation in free monoid. *Theoret. Comput. Sci.*, 242(1-2):403–475, 2000.

[319] G. S. Makanin and A. G. Savushkina. An equation in a free group that defines colored braids. *Mat. Zametki*, 70(4):591–602, 2001.

[320] T. A. Makanina. The occurrence problem for the braid group $\mathfrak{B}_{n+1}$ when $n+1 \geq 5$. *Mat. Zametki*, 29(1):31–33, 154, 1981.

[321] A. Malcev. On isomorphic matrix representations of infinite groups. *Rec. Math. [Mat. Sbornik] N.S.*, 8 (50):405–422, 1940.

[322] A. I. Mal'cev. Two remarks on nilpotent groups. *Mat. Sb. N.S.*, 37(79):567–572, 1955.

[323] A. I. Mal'cev. Homomorphisms onto finite groups (Russian). *Ivanov gosudarst. ped. Inst., utsen. Zap. fiz-mat. Nauk*, 18:49–60, 1958.

[324] A. I. Malcev. *Algoritmy i rekursivnye funktsii.* Izdat. "Nauka", Moscow, 1965.

[325] A. Malheiro. Complete rewriting systems for codified submonoids. *Internat. J. Algebra Comput.*, 15(2):207–216, 2005.

[326] A. S. Malkhasyan. The degree of periodicity of solutions of equations in a free group. *Mat. Zametki*, 39(3):413–423, 462, 1986.

[327] A. S. Malkhasyan. Equations in a free group with coefficients depending on parameters. *Sibirsk. Mat. Zh.*, 27(5):116–127, 206, 1986.

[328] A. S. Malkhasyan. Pseudosymmetric equations in a free monoid. *Math. Notes*, 95(3-4):520–529, 2014. Translation of Mat. Zametki **95** (2014), no. 4, 577–589.

[329] J. B. Manchak. On the possibility of supertasks in general relativity. *Found. Phys.*, 40(3):276–288, 2010.

[330] J. B. Manchak. Malament-Hogarth machines. *British J. Philos. Sci.*, 71(3):1143–1153, 2020.

[331] D. Marker. *Model theory*, volume 217 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 2002. An introduction.

[332] A. A. Markov. On certain insoluble problems concerning matrices. *Doklady Akad. Nauk SSSR (N. S.)*, 57:539–542, 1947.

[333] A. A. Markov. The impossibility of certain algorithms in the theory of associative systems. *Doklady Akad. Nauk SSSR (N.S.)*, 77:19–20, 1951.

[334] A. A. Markov. *Teoriya algorifmov.* Izdat. Akad. Nauk SSSR, Moscow, 1954. Trudy Mat. Inst. Steklov. no. 42.

[335] A. Martino. A proof that all Seifert 3-manifold groups and all virtual surface groups are conjugacy separable. *J. Algebra*, 313(2):773–781, 2007.

[336] A. Martino and A. Minasyan. Conjugacy in normal subgroups of hyperbolic groups. *Forum Math.*, 24(5):889–910, 2012.

[337] J. V. Matijasevič. Simple examples of unsolvable canonical calculi. *Trudy Mat. Inst. Steklov*, 93:50–88, 1967.

[338] J. McCool and A. Pietrowski. On recognising certain one relation presentations. *Proc. Amer. Math. Soc.*, 36:31–33, 1972.

[339] J. McCool and P. E. Schupp. On one relator groups and HNN extensions. *J. Austral. Math. Soc.*, 16:249–256, 1973. Collection of articles dedicated to the memory of Hanna Neumann, II.

[340] J. C. C. McKinsey. The decision problem for some classes of sentences without quantifiers. *J. Symbolic Logic*, 8:61–76, 1943.

[341] R. McNaughton and P. Narendran. Special monoids and special Thue systems. *J. Algebra*, 108(1):248–255, 1987.

[342] R. McNaughton, P. Narendran, and F. Otto. Church-Rosser Thue systems and formal languages. *J. Assoc. Comput. Mach.*, 35(2):324–344, 1988.

[343] E. Mendelson. Second thoughts about Church's thesis and mathematical proofs. *J. Philos.*, 87(5):225–233, 1990.

[344] V. Metaftsis and E. Raptis. On the profinite topology of right-angled Artin groups. *J. Algebra*, 320(3):1174–1181, 2008.

[345] K. A. Mikhaĭlova. The occurrence problem for direct products of groups. *Dokl. Akad. Nauk SSSR*, 119:1103–1105, 1958.

[346] K. A. Mikhaĭlova. The occurrence problem for free products of groups. *Dokl. Akad. Nauk SSSR*, 127:746–748, 1959.

[347] K. A. Mikhaĭlova. The occurrence problem for direct products of groups. *Mat. Sb. (N.S.)*, 70 (112):241–251, 1966.

[348] K. A. Mikhaĭlova. The occurrence problem for free products of groups. *Mat. Sb. (N.S.)*, 75 (117):199–210, 1968.

[349] C. F. Miller, III. *On group-theoretic decision problems and their classification.* Annals of Mathematics Studies, No. 68. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1971.

[350] C. F. Miller, III. Decision problems for groups—survey and reflections. In *Algorithms and classification in combinatorial group theory (Berkeley, CA, 1989)*, volume 23 of *Math. Sci. Res. Inst. Publ.*, pages 1–59. Springer, New York, 1992.

[351] J. Milnor. Growth of finitely generated solvable groups. *J. Differential Geometry*, 2:447–449, 1968.

[352] J. Milnor. A note on curvature and fundamental group. *J. Differential Geometry*, 2:1–7, 1968.

[353] J. Milnor. Problem 5603. *Amer. Math. Monthly*, 75:685–686, 1968.

[354] A. Minasyan. Hereditary conjugacy separability of right-angled Artin groups and its applications. *Groups Geom. Dyn.*, 6(2):335–388, 2012.

[355] A. Minasyan and P. Zalesskii. One-relator groups with torsion are conjugacy separable. *J. Algebra*, 382:39–45, 2013.

[356] A. Minasyan and P. Zalesskii. Virtually compact special hyperbolic groups are conjugacy separable. *Comment. Math. Helv.*, 91(4):609–627, 2016.

[357] D. I. Moldavanskiĭ. Certain subgroups of groups with one defining relation. *Sibirsk. Mat. Ž.*, 8:1370–1384, 1967.

[358] D. I. Moldavanskiĭ. The intersection of finitely generated subgroups. *Sibirsk. Mat. Ž.*, 9:1422–1426, 1968.

[359] D. I. Moldavanskiĭ. On the residual properties of Baumslag-Solitar groups. *Comm. Algebra*, 46(9):3766–3778, 2018.

[360] E. H. Moore. Concerning the Abstract Groups of Order k ! and 1/2k ! Holohedrically Isomorphic with the Symmetric and the Alternating Substitution-Groups on k Letters. *Proc. Lond. Math. Soc.*, 28:357–366, 1896/97.

[361] A. W. o. Mostowski. On the decidability of some problems in special classes of groups. *Fund. Math.*, 59:123–135, 1966.

[362] D. E. Muller and P. E. Schupp. Groups, the theory of ends, and context-free languages. *J. Comput. System Sci.*, 26(3):295–310, 1983.

[363] D. E. Muller and P. E. Schupp. The theory of ends, pushdown automata, and second-order logic. *Theoret. Comput. Sci.*, 37(1):51–75, 1985.

[364] V. L. Murskiĭ. Isomorphic imbeddability of a semigroup with an enumerable set of defining relations into a finitely presented semigroup. *Mat. Zametki*, 1:217–224, 1967.

[365] P. Narendran. On the equivalence problem for regular Thue systems. *Theoret. Comput. Sci.*, 44(2):237–245, 1986.

[366] P. Narendran. It is decidable whether a monadic Thue system is canonical over a regular set. *Math. Systems Theory*, 23(4):245–254, 1990.

[367] P. Narendran, C. Ó'Dúnlaing, and F. Otto. It is undecidable whether a finite special string-rewriting system presents a group. *Discrete Math.*, 98(2):153–159, 1991.

[368] P. Narendran, C. Ó'Dúnlaing, and H. Rolletschek. Complexity of certain decision problems about congruential languages. *J. Comput. System Sci.*, 30(3):343–358, 1985.

[369] M. B. Nathanson. Number theory and semigroups of intermediate growth. *Amer. Math. Monthly*, 106(7):666–669, 1999.

[370] M. B. Nathanson. Additive number theory and linear semigroups with intermediate growth. In *Combinatorial and additive number theory—CANT 2011 and 2012*, volume 101 of *Springer Proc. Math. Stat.*, pages 175–194. Springer, New York, 2014.

[371] P. M. Neumann. The SQ-universality of some finitely presented groups. *J. Austral. Math. Soc.*, 16:1–6, 1973. Collection of articles dedicated to the memory of Hanna Neumann, I.

[372] B. B. Newman. *Some aspects of one-relator groups.* PhD thesis, James Cook University, 1968.

[373] B. B. Newman. The soluble subgroups of a one-relator group with torsion. *J. Austral. Math. Soc.*, 16:278–285, 1973. Collection of articles dedicated to the memory of Hanna Neumann, III.

[374] M. H. A. Newman. On theories with a combinatorial definition of "equivalence.". *Ann. of Math. (2)*, 43:223–243, 1942.

[375] G. A. Niblo and D. T. Wise. Subgroup separability, knot groups and graph manifolds. *Proc. Amer. Math. Soc.*, 129(3):685–693, 2001.

[376] J. Nielsen. Om regning med ikke-kommutative faktorer og dens anvendelse i gruppeteorien. *Math. Tids.*, B:78–94, 1921.

[377] J. Nielsen. A basis for subgroups of free groups. *Math. Scand.*, 3:31–43, 1955.

[378] M. Nivat. Éléments de la théorie générale des codes. In *Automata Theory*, pages 278–294. Academic Press, New York, 1966.

[379] M. Nivat. Congruences de Thue et t-langages. *Studia Sci. Math. Hungar.*, 6:243–249, 1971.

[380] M. Nivat. Congruences parfaites et quasi-parfaites. In *Séminaire P. Dubreil, 25e année (1971/72), Algèbre, Fasc. 1, Exp. No. 7*, page 9. 1973.

[381] P. S. Novikov. *Ob algoritmičeskoĭ nerazrešimosti problemy toždestva slov v teorii grupp.* Trudy Mat. Inst. im. Steklov. no. 44. Izdat. Akad. Nauk SSSR, Moscow, 1955.

[382] C.-F. Nyberg-Brodda. On the word problem for compressible monoids. *Pre-print (submitted)*, 2020. Available at arXiv:2012.01402.

[383] C.-F. Nyberg-Brodda. On the word problem for special monoids. *Pre-print (submitted)*, 2020. Available at arXiv:2011.09466.

[384] C.-F. Nyberg-Brodda. The B. B. Newman spelling theorem. *The British Journal for the History of Mathematics*, 36(**2**):117–131, 2021.

[385] C.-F. Nyberg-Brodda. The language theory of monoid and semigroup free products. 2021. In preparation.

[386] C.-F. Nyberg-Brodda. Non-finitely generated maximal subgroups of context-free monoids. *Pre-print (submitted)*, 2021. Available at arXiv:2107.12861.

[387] C.-F. Nyberg-Brodda. On the Identity Problem for Finitely Presented groups and Semigroups, 2021. English translation of "K Probleme Tozhdestva v Konechno-opredelennyh Gruppah i Polugruppah", Ph.D. Thesis by G. S. Makanin (1966) (Available online at arXiv:2102.00745).

[388] C.-F. Nyberg-Brodda. The word problem for one-relation monoids: a survey. *Semigroup Forum*, 103(**2**):297–355, 2021.

[389] C. Ó'Dúnlaing. Infinite regular Thue systems. *Theoret. Comput. Sci.*, 25(2):171–192, 1983.

[390] F. Otto. Some undecidability results for nonmonadic Church-Rosser Thue systems. *Theoret. Comput. Sci.*, 33(2-3):261–278, 1984.

[391] F. Otto. On deciding the confluence of a finite string-rewriting system on a given congruence class. *J. Comput. System Sci.*, 35(3):285–310, 1987.

[392] F. Otto. Completing a finite special string-rewriting system on the congruence class of the empty word. *Appl. Algebra Engrg. Comm. Comput.*, 2(4):257–274, 1992.

[393] F. Otto. The problem of deciding confluence on a given congruence class is tractable for finite special string-rewriting systems. *Math. Systems Theory*, 25(4):241–251, 1992.

[394] F. Otto. Solvability of word equations modulo finite special and confluent string-rewriting systems is undecidable in general. *Inform. Process. Lett.*, 53(5):237–242, 1995.

[395] F. Otto and Y. Kobayashi. Properties of monoids that are presented by finite convergent string-rewriting systems—a survey. In *Advances in algorithms, languages, and complexity*, pages 225–266. Kluwer Acad. Publ., Dordrecht, 1997.

[396] F. Otto and L. Zhang. Decision problems for finite special string-rewriting systems that are confluent on some congruence class. *Acta Inform.*, 28(5):477–510, 1991.

[397] S. Papadima and A. I. Suciu. Algebraic invariants for right-angled Artin groups. *Math. Ann.*, 334(3):533–555, 2006.

[398] L. Pélecq. Automorphism groups of context-free graphs. *Theoret. Comput. Sci.*, 165(2):275–293, 1996.

[399] G. Perelman. Ricci flow with surgery on three-manifolds, 2003.

[400] D. Perrin and P. Schupp. Sur les monoïdes à un relateur qui sont des groupes. *Theoret. Comput. Sci.*, 33(2-3):331–334, 1984.

[401] H. Poincaré. Théorie des groupes fuchsiens. *Acta Math.*, 1(1):1–76, 1882.

[402] H. Poincaré. Second Complement a l'Analysis Situs. *Proc. Lond. Math. Soc.*, 32:277–308, 1900.

[403] E. L. Post. Finite combinatory processes—formulation. *Journal of Symbolic Logic*, 1(3):103–105, 1936.

[404] E. L. Post. Formal reductions of the general combinatorial decision problem. *Amer. J. Math.*, 65:197–215, 1943.

[405] E. L. Post. Recursively enumerable sets of positive integers and their decision problems. *Bull. Amer. Math. Soc.*, 50:284–316, 1944.

[406] E. L. Post. Recursive unsolvability of a problem of Thue. *J. Symbolic Logic*, 12:1–11, 1947.

[407] J. F. Power. Thue's 1914 paper: a translation. *arXiv*, 2013. Available at arXiv:1308.5858.

[408] S. J. Pride. *Residual Properties of Free Groups.* PhD thesis, Australian National University, 1974.

[409] S. J. Pride. On the generation of one-relator groups. *Trans. Amer. Math. Soc.*, 210:331–364, 1975.

[410] S. J. Pride. The isomorphism problem for two-generator one-relator groups with torsion is solvable. *Trans. Amer. Math. Soc.*, 227:109–139, 1977.

[411] S. J. Pride. Small cancellation conditions satisfied by one-relator groups. *Math. Z.*, 184(2):283–286, 1983.

[412] W. V. O. Quine. *Mathematical logic.* Harvard University Press, Cambridge, Mass., 1951. Revised ed.

[413] M. O. Rabin. Recursive unsolvability of group theoretic problems. *Ann. of Math. (2)*, 67:172–194, 1958.

[414] M. O. Rabin. Decidability of second-order theories and automata on infinite trees. *Trans. Amer. Math. Soc.*, 141:1–35, 1969.

[415] D. Rattaggi. An incoherent simple group, 2005. Available at arXiv:0507363.

[416] K. Reidemeister. Knoten und Gruppen. *Abh. Math. Sem. Univ. Hamburg*, 5(1):7–23, 1927.

[417] V. N. Remeslennikov. Conjugacy in polycyclic groups. *Algebra i Logika*, 8:712–725, 1969.

[418] E. Render and M. Kambites. Rational subsets of polycyclic monoids and valence automata. *Inform. and Comput.*, 207(11):1329–1339, 2009.

[419] E. Rips. Subgroups of small cancellation groups. *Bull. London Math. Soc.*, 14(1):45–47, 1982.

[420] E. F. Robertson, N. Ruškuc, and J. Wiegold. Generators and relations of direct products of semigroups. *Trans. Amer. Math. Soc.*, 350(7):2665–2685, 1998.

[421] N. Robertson and P. D. Seymour. Graph minors. II. Algorithmic aspects of tree-width. *J. Algorithms*, 7(3):309–322, 1986.

[422] V. A. Romankov. Equations in free metabelian groups. *Sibirsk. Mat. Zh.*, 20(3):671–673, 694, 1979.

[423] V. A. Romankov. On the occurrence problem for rational subsets of a group. In *International Conference on Combinatorial and Computaitonal Methods in Mathematics*, pages 76–81, 1999.

[424] V. A. Romankov. Two problems on solvable and nilpotent groups. *Algebra Logika*, 59(6), 2020.

[425] N. S. Romanovskiĭ. On the residual finiteness of free products with respect to subgroups. *Izv. Akad. Nauk SSSR Ser. Mat.*, 33:1324–1329, 1969.

[426] O. Schreier. Die Untergruppen der freien Gruppen. *Abh. Math. Sem. Univ. Hamburg*, 5(1):161–183, 1927.

[427] S. Schulz. First-order logic with reachability predicates on infinite systems. In *30th International Conference on Foundations of Software Technology and Theoretical Computer Science*, volume 8 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages 493–504. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2010.

[428] P. E. Schupp. A survey of small cancellation theory. In *Word problems: decision problems and the Burnside problem in group theory (Conf. on Decision Problems in Group Theory, Univ. California, Irvine, Calif. 1969; dedicated to Hanna Neumann)*, pages 569–589. Studies in Logic and the Foundations of Math., Vol. 71. 1973.

[429] P. E. Schupp. A strengthened Freiheitssatz. *Math. Ann.*, 221(1):73–80, 1976.

[430] M. P. Schützenberger. $\overline{\mathscr{D}}$ représentation des demi-groupes. *C. R. Acad. Sci. Paris*, 244:1994–1996, 1957.

[431] E. A. Scott. The embedding of certain linear and abelian groups in finitely presented simple groups. *J. Algebra*, 90(2):323–332, 1984.

[432] E. A. Scott. A finitely presented simple group with unsolvable conjugacy problem. *J. Algebra*, 90(2):333–353, 1984.

[433] G. P. Scott. Compact submanifolds of 3-manifolds. *J. London Math. Soc. (2)*, 7:246–250, 1973.

[434] P. Scott. Subgroups of surface groups are almost geometric. *J. London Math. Soc. (2)*, 17(3):555–565, 1978.

[435] Z. Sela. Endomorphisms of hyperbolic groups. I. The Hopf property. *Topology*, 38(2):301–321, 1999.

[436] G. Sénizergues. A new class of C.F.L. for which the equivalence is decidable. *Inform. Process. Lett.*, 13(1):30–34, 1981.

[437] G. Sénizergues. The equivalence and inclusion problems for NTS languages. *J. Comput. System Sci.*, 31(3):303–331, 1985.

[438] G. Sénizergues. On the rational subsets of the free group. *Acta Inform.*, 33(3):281–296, 1996.

[439] J.-P. Serre. *Arbres, amalgames,* SL_2. Astérisque, No. 46. Société Mathématique de France, Paris, 1977. Avec un sommaire anglais, Rédigé avec la collaboration de Hyman Bass.

[440] S. G. Shanker. Wittgenstein versus Turing on the nature of Church's thesis. *Notre Dame J. Formal Logic*, 28(4):615–650, 1987.

[441] S. Shapiro. Proving things about the informal. In *Turing's revolution*, pages 283–296. Birkhäuser/Springer, Cham, 2015.

[442] S. Shelah. On a problem of Kurosh, Jónsson groups, and applications. In *Word problems, II (Conf. on Decision Problems in Algebra, Oxford, 1976)*, volume 95 of *Stud. Logic Foundations Math.*, pages 373–394. North-Holland, Amsterdam-New York, 1980.

[443] L. N. Shevrin and M. V. Volkov. Identities of semigroups. *Soviet Math. (Iz. VUZ)*, (29):1–64, 1985. [Izv. Vyssh. Uchebn. Zaved. Mat. **11** (1985)].

[444] L. M. Shneerson. Identities in semigroups with one defining relation. *Logic, Algebra, and Computational Mathematics, Ivanovo Pedagogical Institute*, 1(1–2):139–156, 1972.

[445] L. M. Shneerson. Identities in semigroups with one defining relation. II. *Logic, Algebra, and Computational*

Mathematics, Ivanovo Pedagogical Institute, 1(3–4):112–124, 1972.

[446] L. M. Shneerson. Identities in finitely generated semigroups of polynomial growth. *J. Algebra*, 154(1):67–85, 1993.

[447] L. M. Shneerson. On semigroups of intermediate growth. *Comm. Algebra*, 32(5):1793–1803, 2004.

[448] L. M. Shneerson. Types of growth and identities of semigroups. *Internat. J. Algebra Comput.*, 15(5-6):1189–1204, 2005.

[449] L. M. Shneerson. Polynomial growth in semigroup varieties. *J. Algebra*, 320(6):2218–2279, 2008.

[450] L. M. Shneerson. On growth, identities and free subsemigroups for inverse semigroups of deficiency one. *Internat. J. Algebra Comput.*, 25(1-2):233–258, 2015.

[451] L. M. Shneerson and D. Easdown. Growth of finitely presented Rees quotients of free inverse semigroups. *Internat. J. Algebra Comput.*, 21(1-2):315–328, 2011.

[452] L. M. Shneerson and D. Easdown. On finite presentations of inverse semigroups with zero having polynomial growth. *Semigroup Forum*, 99(2):391–446, 2019.

[453] L. M. Shneerson and M. V. Volkov. The identities of the free product of two trivial semigroups. *Semigroup Forum*, 95(1):245–250, 2017.

[454] H. Short. Quasiconvexity and a theorem of Howson's. In *Group theory from a geometrical viewpoint (Trieste, 1990)*, pages 168–176. World Sci. Publ., River Edge, NJ, 1991.

[455] W. Sieg. Mechanical procedures and mathematical experience. In *Mathematics and mind (Amherst, MA, 1991)*, Logic Comput. Philos., pages 71–117. Oxford Univ. Press, New York, 1994.

[456] W. Sieg. Calculations by man and machine: conceptual analysis. In *Reflections on the foundations of mathematics (Stanford, CA, 1998)*, volume 15 of *Lect. Notes Log.*, pages 390–409. Assoc. Symbol. Logic, Urbana, IL, 2002.

[457] A. M. Slobodskoĭ. Undecidability of the universal theory of finite groups. *Algebra i Logika*, 20(2):207–230, 251, 1981.

[458] H. J. S. Smith. On systems of linear indeterminate equations and congruences. *Philosophical Transactions of the Royal Society of London*, 151:293–326, 1861.

[459] W. D. Smith. Church's thesis meets the N-body problem. *Appl. Math. Comput.*, 178(1):154–183, 2006.

[460] C. C. Squier. Units of special Church-Rosser monoids. *Theoret. Comput. Sci.*, 49(1):13–22, 1987.

[461] J. Stallings. Coherence of 3-manifold fundamental groups. In *Séminaire Bourbaki, Vol. 1975/76, 28 ème année, Exp. No. 481*, pages 167–173. Lecture Notes in Math., Vol. 567. 1977.

[462] J. R. Stallings. On torsion-free groups with infinitely many ends. *Ann. of Math. (2)*, 88:312–334, 1968.

[463] J. R. Stallings. Quotients of the powers of the augmentation ideal in a group ring. In *Knots, groups, and* 3-*manifolds (Papers dedicated to the memory of R. H. Fox)*, pages 101–118. Ann. of Math. Studies, No. 84. 1975.

[464] R. P. Stanley. *Enumerative combinatorics. Vol. 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997. With a foreword by Gian-Carlo Rota, Corrected reprint of the 1986 original.

[465] P. F. Stebe. A residual property of certain groups. *Proc. Amer. Math. Soc.*, 26:37–42, 1970.

[466] L. A. Steen, editor. *Mathematics tomorrow*. Springer-Verlag, New York-Berlin, 1981.

[467] J. B. Stephen. *Applications of automata theory to presentations of monoids and inverse monoids*. ProQuest LLC, Ann Arbor, MI, 1987. Thesis (Ph.D.)–The University of Nebraska - Lincoln.

[468] J. B. Stephen. The automorphism group of the graph of an $\mathscr{R}$ class. *Semigroup Forum*, 53(3):387–389, 1996.

[469] J. Stillwell. The occurrence problem for mapping class groups. *Proc. Amer. Math. Soc.*, 101(3):411–416, 1987.

[470] R. Stöhr. Groups with one more generator than relators. *Math. Z.*, 182(1):45–47, 1983.

[471] A. K. Sushkevich. *The Theory of Operations as the General Theory of Groups*. PhD thesis, Voronezh State University, 1922.

[472] V. Tartakovskiĭ. On the problem of equivalence for certain types of groups. *Doklady Akad. Nauk SSSR (N.S.)*, 58:1909–1910, 1947.

[473] V. Tartakovskiĭ. On the process of extinction. *Doklady Akad. Nauk SSSR (N.S.)*, 58:1605–1608, 1947.

[474] V. A. Tartakovskiĭ. Solution of the word problem for groups with a k-reduced basis for $k > 6$. *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, 13:483–494, 1949.

[475] A. Thue. Problem über Veränderungen von Zeichenreihen nach gegebenen Regeln. *Christiana Videnskaps-Selskabs Skrifter, I. Math. naturv. Klasse*, 10, 1914.

[476] D. Tieudjo and D. I. Moldavanskii. On the automorphisms of some one-relator groups. *Comm. Algebra*, 34(11):3975–3983, 2006.

[477] G. S. Tseitin. An associative calculus with an insoluble problem of equivalence. *Trudy Mat. Inst. Steklov.*, 52:172–189, 1958.

[478] A. M. Turing. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proc. London Math. Soc. (2)*, 42(3):230–265, 1936.

[479] A. M. Turing. The word problem in semi-groups with cancellation. *Ann. of Math. (2)*, 52:491–505, 1950.

[480] W. T. Tutte. *Graph theory*, volume 21 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Company, Advanced Book Program, Reading, MA, 1984. With a foreword by C. St. J. A. Nash-Williams.

[481] V. A. Uspenskii and A. L. Semenov. What are the gains of the theory of algorithms: basic developments connected with the concept of algorithm and with its application in mathematics. In *Algorithms in modern mathematics and computer science (Urgench, 1979)*, volume 122 of *Lecture Notes in Comput. Sci.*, pages 100–234. Springer, Berlin-New York, 1981.

[482] L. G. Valiant. General context-free recognition in less than cubic time. *J. Comput. System Sci.*, 10:308–315, 1975.

[483] B. L. van der Waerden. Free products of groups. *Amer. J. Math.*, 70:527–528, 1948.

[484] J. M. Vazhenin. Sur la liaison entre problèmes combinatoires et algorithmiques. *Theoret. Comput. Sci.*, 16(1):33–41, 1981.

[485] Y. M. Vazhenin. Semigroups with one defining relation whose elementary theories are decidable. *Sib. Math. J.*, 24:33–41, 1983. [Sibirsk. Mat. Zh. **24** (1983)].

[486] B. A. F. Wehrfritz. Generalized free products of linear groups. *Proc. London Math. Soc. (3)*, 27:402–424, 1973.

[487] P. D. Welch. The extent of computation in Malament-Hogarth spacetimes. *British J. Philos. Sci.*, 59(4):659–674, 2008.

[488] J. H. C. Whitehead. On equivalent sets of elements in a free group. *Ann. of Math. (2)*, 37(4):782–800, 1936.

[489] V. V. Williams. Multiplying matrices faster than Coppersmith-Winograd [extended abstract]. In *STOC'12—Proceedings of the 2012 ACM Symposium on Theory of Computing*, pages 887–898. ACM, New York, 2012.

[490] J. S. Wilson. Soluble groups of deficiency 1. *Bull. London Math. Soc.*, 28(5):476–480, 1996.

[491] D. T. Wise. A non-Hopfian automatic group. *J. Algebra*, 180(3):845–847, 1996.

[492] D. T. Wise. Research announcement: the structure of groups with a quasiconvex hierarchy. *Electron. Res. Announc. Math. Sci.*, 16:44–55, 2009.

[493] D. T. Wise. *From riches to raags: 3-manifolds, right-angled Artin groups, and cubical geometry*, volume 117 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2012.

[494] L. Wittgenstein. *Tractatus Logico-Philosophicus.* Routledge, London, 1922.

[495] L. Wittgenstein. *Remarks on the Philosophy of Psychology, Volume I.* Basil Blackwell, Oxford, 1980.

[496] P. C. Wong. Subgroup separability of certain HNN extensions. *Rocky Mountain J. Math.*, 23(1):391–394, 1993.

[497] H. Wussing. *Die Genesis des abstrakten Gruppenbegriffes. Ein Beitrag zur Entstehungsgeschichte der abstrakten Gruppentheorie.* VEB Deutscher Verlag der Wissenschaften, Berlin, 1969.

[498] Z. L. Xin. An efficient algorithm to decide whether a monoid presented by a regular Church-Rosser Thue system is a group. *Theoret. Comput. Sci.*, 66(1):55–63, 1989.

[499] A. Yamamura and T. Saito. Subgroup membership problem and its applications to information security. volume 57, pages 25–41. 2003. Japanese Association of Mathematical Sciences 2001 Annual Meeting (Tennoji).

[500] D. H. Younger. Recognition and parsing of context-free languages in time n3. *Information and Control*, 10(2):189–208, 1967.

[501] L. Zhang. Conjugacy in special monoids. *J. Algebra*, 143(2):487–497, 1991.

[502] L. Zhang. Applying rewriting methods to special monoids. *Math. Proc. Cambridge Philos. Soc.*, 112(3):495–505, 1992.

[503] L. Zhang. Congruential languages specified by special string-rewriting systems. In *Words, languages and combinatorics (Kyoto, 1990)*, pages 551–563. World Sci. Publ., River Edge, NJ, 1992.

[504] L. Zhang. On the conjugacy problem for one-relator monoids with elements of finite order. *Internat. J. Algebra Comput.*, 2(2):209–220, 1992.

[505] L. Zhang. A short proof of a theorem of Adjan. *Proc. Amer. Math. Soc.*, 116(1):1–3, 1992.

[506] L. Zhang. Some properties of finite special string-rewriting systems. *J. Symbolic Comput.*, 14(4):359–369, 1992.

[507] H. Zieschang. Über die Nielsensche Kürzungsmethode in freien Produkten mit Amalgam. *Invent. Math.*, 10:4–37, 1970.

[508] J. Zuther. Ends in digraphs. *Discrete Math.*, 184(1-3):225–244, 1998.