Computational and logical aspects of infinite monoids

Von der Fakultät Informatik, Elektrotechnik und
Informationstechnik
der Universität Stuttgart
als Habilitationsschrift genehmigte Abhandlung

Vorgelegt von
Markus Lohrey

Stuttgart, 2003

Preface

The present work constitutes the *Habilitationsschrift* of the author written at the University of Stuttgart. It contains a treatise of several computational and logical aspects of infinite monoids.

The first chapter is devoted to the word problem for finitely generated monoids. In particular, the relationship between the the computational complexity of the word problem and the syntactical properties of monoid presentations is investigated. The second chapter studies Cayley-graphs of finitely generated monoids under a logical point of view. Cayley-graphs of groups play an important role in combinatorial group theory. We will study first-order and monadic second-order theories of Cayley-graphs for both groups and monoids. The third chapter deals with word equations over monoids. Using the graph product operation, which generalizes both the free and the direct product, we generalize the seminal decidability results of Makanin on free monoids and groups to larger classes of monoids.

Acknowledgments. It is a great pleasure for me to thank Professor V. Diekert for his enthusiastic and inspiring support over the last six years that I spent in his research group at the University of Stuttgart. Many ideas in this Habilitationsschrift have their origin in joint research efforts with Professor V. Diekert.

I am grateful to Professor Volker Claus (University of Stuttgart), Professor Erich Grädel (RWTH Aachen), and Professor Géraud Sénizergues (Université Bordeaux I Nouvelle) who were further referees of this Habilitationsschrift.

I am also greatly indebted to Dr. Dietrich Kuske for a fruitful research collaboration during the last few years. His insights had substantial influence on the content of this work.

Thanks to all members of the group *Theoretische Grundlagen der Informatik* at the University of Stuttgart for many fruitful discussions.

Finally, I acknowledge several motivating discussions with Professor Anca Muscholl.

Stuttgart, August 2003 Markus Lohrey

Contents

Chapter 1

Introduction

The origin of the material in this work goes back to the beginning of the 20th century. In his paper from 1910, Max Dehn introduced several concepts, which soon turned out be fundamental in combinatorial group theory [59]. One of his main attainments was the formulation of the word problem for finitely presented groups: Given a finitely presented group and two words over the generators of the group, do these two words represent the same group element? In the terminology of theoretical computer science, Dehn formulated the word problem as a decision problem, but one has to notice that the notion of decidability was formalized only more than twenty years later.

Four years after Dehn published his paper, Axel Thue introduced the word problem independently in another context [208]. Thue was interested in the problem whether two given strings can be transformed into each other by a given finite system of rules, which have to be applied in a local way. In modern terminology, Thue introduced the word problem for finitely presented monoids. Thue itself remarked in his paper that this problem in its general formulation seems to be extremely difficult, and he was only able to present solutions for a few special cases.

Maybe this difficulty of Thue's problem is one reason for the fact that his work was ignored for many years. Only when the foundations of recursion theory were developed in the 1930s, the importance of Thue's work was recognized by several logicans. Finally, Markov [135] and Post [163] proved independently that there exist finitely presented monoids with an undecidable word problem, thus also Thue's original problem is undecidable. This seems to be one the first undecidability results which touches "real mathematics".

The result of Markov and Post had no influence on Dehn's problem: Dehn was interested in the word problem for groups, but the monoids constructed by Markov and Post fail to be groups. In fact, it turned out to be much harder to encode computation steps faithfully into groups, but finally Novikov [155] and Boone [28] succeeded in proving that also the word problem for finitely presented groups is undecidable in general, see [172] for an excellent exposition.

Although the results of Markov, Post, Novikov, and Boone are negative statements with respect to effectiveness, they marked the starting point of a very fruitful and still ongoing line of research in the context of word problems. A few references from this vast field that are in particular relevant for the material in this work will be cited in the next paragraph.

The borderline between decidability and undecidability for monoid word problems was investigated in great detail by Adjan in [2]. On the decidability side let us mention monoids that can be presented by terminating and confluent rewriting systems. Inspired by the influential work of Knuth and Bendix [112] and the later development of the theory of term rewriting, this class received a lot of attention since the 1980s, see e.g. [16, 104, 117]. For the group case, algebraic restrictions on the class of groups often result in decidable word problems, see the contributions in [29] and the surveys [143, 199] for an overview. The development of computational complexity theory in the 1970s motivated a complexity oriented analysis of word problems. The results of [6, 41] roughly state that the complexity of the word problem for group presentations may reach every specified degree of complexity. Bauer and Otto [13] have shown that for every $n \geq 3$ there exists a finite, confluent, and terminating monoid presentation, for which the word problem is contained in the n-th but not in the $(n-1)$-th level of the Grzegorczyk hierarchy. Thus, also for confluent and terminating presentations, the complexity of the word problem may become extremely high. On the lower end of the spectrum of complexity classes let us mention the work of Waack [217, 218] and Robinson [171] on the parallel complexity of word problems for groups and the work of Lipton and Zalcstein [120] and Simon [194] on word problems within logarithmic space.

In Chapter 3 we investigate the complexity of the word problem for certain classes of monoid presentations. We will be interested in the interplay between the syntactical structure of presentations and the complexity of the associated word problem. It will turn out that many interesting complexity classes can be characterized in this way. This includes classes like P,

PSPACE, EXPTIME, and EXPSPACE, but also classes like SL (symmetric logspace) or LOGDCFL (the logspace closure of the class of deterministic context-free languages) for which only a few complete problems are known.

In his paper from 1910 [59], Dehn also considered the notion of the Cayley-graph of a group. Cayley-graphs for finite groups were already introduced by Cayley in 1878 [45]. The Cayley-graph of a group with respect to a generating set Γ has as vertices the group elements and an a-labeled edge from x to y if $y = xa$, where a is a generator from Γ. Only at the late 1960s, the real importance of the geometrical and combinatorial properties of Cayley-graphs was recognized. At that time, Lyndon developed the theory of cancellation diagrams [126], Serre established the foundations of the theory of groups acting on trees [60, 190], and Stallings proved his famous structure theorem of groups with more than one end [196]. All these developments use Cayley-graphs in an essential way. Another application of Cayley-graphs was found via their connection to formal language theory. The word problem of a group may be viewed as a formal language (the set of all words over the generators, which represent the identity of the group), hence one may classify groups according to the grammatical properties of their word problems. Questions of this kind were first investigated by Anisimov [4], who noticed that the word problem of a group is a regular language if and only if the group is finite. The much harder case of groups with context-free word problems was tackled by Letičevskiĭ and Smikun [118, 195] and later by Muller and Schupp [144, 145]. Together with later results of Dunwoody [75], Muller and Schupp have shown that the word problem of a group is context-free if and only if the group is virtually-free (i.e., has a free subgroup of finite index) [144] if and only if the Cayley-graph of the group is the transition graph of a pushdown automaton [145]. A group that has one of these three equivalent properties is called *context-free*. In [145], Muller and Schupp also gave a graph theoretical characterization of the transition graphs of pushdown automata, which they called *context-free graphs*. In the same paper, Muller and Schupp proved (by using a reduction to Rabin's famous result on the decidability of the monadic second-order theory of the complete binary tree [166]) that every context-free graph has a decidable monadic second-order theory. In particular, the monadic second-order theory of the Cayley-graph of a context-free group is decidable.

The latter result indicates a connection between the logical properties of Cayley-graphs and word problems. This connection will be explored in more detail in Chapter 4. We will present precise characterizations of

the finitely generated groups, whose Cayley-graphs have decidable monadic second-order theories (resp. first-order theories). It turns out that the Cayley-graph of a finitely generated group has a decidable monadic second-order theory (resp. first-order theory) if and only if the group is context-free (resp. has a decidable word problem). Moreover, we will extend this line of research also to Cayley-graphs of monoids, which are due to the missing symmetry much harder to analyze. We will deduce the following closure results: The class of finitely generated monoids whose Cayley-graphs have decidable first-order theories (resp. monadic second-order theories) is closed under graph products (resp. free products). The graph product construction is a well-known mathematical construction, which generalizes both free and direct products, see e.g. [87, 94, 211]. In order to obtain these results on Cayley-graphs, we will prove more general statements that apply to arbitrary structures. In particular, we obtain new results for general (infinite) graphs with decidable monadic second-order theories. Due to their potential application for model checking infinite state systems, such graphs received a lot of attention in the past, see [205] for an overview.

It should be remarked that the investigation of Cayley-graphs under the light of mathematical logic extends our previous work on word problems in a natural way: Whether two words over the generators represent the same element can be easily expressed as a logical first-order property of the Cayley-graph. On the other hand, our logical approach to Cayley-graphs has a natural generalization as well: when expressing properties of the Cayley-graph of a monoid $\mathcal{M}$ (generated by Γ) in the language of first-order logic, all atomic statements are of the form "there is an edge from x to y labeled with the generator $a \in \Gamma$", where x and y are variables that range over the monoid $\mathcal{M}$. In other words, we say that $y = xa$ in $\mathcal{M}$. This is a very simple example of a word equation. More generally, a word equation over $\mathcal{M}$ is an equation of the form $U = V$, where U and V are words consisting of variables (which take values from $\mathcal{M}$) and generators from Γ. The investigation of word equations over monoids and groups has its origin again in combinatorial group theory: In a paper from 1918, Nielsen studied the automorphism group of the free group generated by a and b [152]. Nielsen finally arrived at the word equation $xyx^{-1}y^{-1} = aba^{-1}b^{-1}$, and described all its solutions in the free group generated by a and b.

The investigation of word equations in free groups under a logical point of view was initiated by Tarski around 1945. Tarski asked whether the first-order theory of a free group F is decidable. In general, the theory of a monoid

$\mathcal{M}$ consists of all those true statements about $\mathcal{M}$ that are constructed from word equations over $\mathcal{M}$ using Boolean connectives and quantifications over variables (that appear in the word equations). Tarski's question became known as Tarski's problem and turned out be extremely difficult, see e.g. [127] for a discussion.

The investigation of word equations in free monoids was initiated by the Russian school around Markov in the late 1950's in connection with Hilbert's Tenth Problem. Markov noted that the undecidability of the problem whether a given word equation has a solution in a given free monoid would imply the undecidability of Hilbert's Tenth Problem. So Markov raised the conjecture that the former problem is indeed undecidable. But Markov's approach towards a solution of Hilbert's Tenth Problem could not succeed. In 1977 Makanin has shown that it is decidable whether a given word equation has a solution in a given free monoid [131] (whereas Hilbert's Tenth Problem was already shown to be undecidable in 1970 by Matiyasevich using a completely different approach, see [136]). In fact, Makanin has shown that the whole existential fragment of the theory of a free monoid is decidable. Together with the results of Marchenkov [134] and Durnev [76], which state that the $\forall\exists^3$-theory of the free monoid $\{a, b\}^*$ is undecidable, Makanin's result gives a quite sharp borderline for decidability questions concerning the theory of a free monoid. In 1982, Makanin was able to extend his approach for free monoids to free groups, and showed that the existential theory of a free group is decidable as well [132, 133]. This established a first step towards a positive solution of Tarski's Problem. A second step was taken again by Makanin. Based on results of Merzlyakov [141] he proved in 1984 that the positive theory of a free group is decidable as well [133]. This theory is defined in the same way as the whole first-order theory, except that only conjunctions and disjunctions (but no negations) are allowed as Boolean connectives. But a complete solution of Tarski's Problem remained open until recently, when Kharlampovich and Myasnikov announced a solution, see [108] for an overview, the complete solution is spread over a series of papers, of which not all appeared yet.

Another recent breakthrough in the field of word equations concerns the complexity of solving word equations. Makanin's original approach for checking whether a word equation has a solution is quite time consuming, both for the case of free monoids and free groups. For the case of free groups, it can be even shown that Makanin's algorithm is not primitive recursive [113], whereas the best known upper bound for Makanin's algorithm for free

monoids is EXPSPACE [90]. Using a completely new approach, Plandowski has shown recently that the solvability of a word equation over a free monoid can be checked in PSPACE [162], this is currently the best known upper bound for this problem (the best known lower bound is NP, which follows from the NP-hardness of integer programming). Plandowski's PSPACE algorithm was adapted by Gutierrez to the case of free groups [91]. These two results can be easily extended to the whole existential theory of a free monoid (free group).

Of course, there is no reason to consider only word equations over free monoids and free groups: Unification theory can be seen as a generalization of the theory of word equations to arbitrary structures, see [7] for an overview. For the case of finitely generated monoids, the results of Makanin and Plandowski were generalized to several other classes of monoids and groups. The decidability of the existential theory of a free partially commutative monoid (trace monoid) was established in [68]. Analogous results were obtained in [168, 184] for torsion-free hyperbolic groups, in [69] for free partially commutative groups (graph groups), and in [67] for plain groups (free products of finite and free groups).

In Chapter 5 we will obtain new decidability results for existential and positive theories of monoids and groups. Our main attention is devoted to the graph product construction. We will prove that under some algebraic restriction, the decidability of the existential theory is preserved under graph products. As a corollary we obtain the decidability of the existential theory of a graph product of finite monoids, free monoids, and torsion-free hyperbolic groups. In general, our construction results in an exponential blow-up with respect to space complexity. For the special case of a graph product of finite monoids, free monoids, and free groups we will deduce a PSPACE upper bound. Concerning positive theories, we prove that the positive theory of a graph product of *groups* can be reduced to the positive theories of those factors that appear isolated in the underlying dependence graph of the graph product and the existential theories of the remaining factors. As a corollary we obtain that the positive theory of a graph product of finite groups and free groups is decidable. Our considerations will also include constraints for the variables, which means that we allow atomic formulas that restrict the value of a variable to some specified set. The concept of word equations with constraints was first considered in [180], further results were obtained in [64, 65, 67, 68, 69].

Chapter 2

Preliminaries

In this chapter we will briefly introduce several concepts that will occur repeatedly in this work. See the given references for more details.

2.1 General notations

For a binary relation $\to$ on some set, we denote by $\overset{+}{\to}$ ($\overset{*}{\to}$) the transitive (reflexive and transitive) closure of $\to$. Let A be an alphabet (finite or infinite). An *involution* ι on A is a function $\iota : A \to A$ such that $\iota(\iota(a)) = a$ for all $a \in A$. The empty word over A is denoted by ε. Let $s = a_1 a_2 \cdots a_n \in A^*$ be a word over A, where $a_i \in A$ for $1 \leq i \leq n$. We define $w^{\text{rev}} = a_n a_{n-1} \cdots a_1$. The *alphabet of* s is $\text{alph}(s) = \{a_1, \ldots, a_n\}$. The *length* of s is $|s| = n$. Furthermore for $a \in A$ we define $|s|_a = |\{i \mid a_i = a\}|$. For $1 \leq i \leq n$ let $s[i] = a_i$ and for $1 \leq i \leq j \leq n$ let $s[i, j] = a_i a_{i+1} \cdots a_j$. If $i > j$ we set $s[i, j] = \varepsilon$. Every word $t = s[1, i]$ with $i \geq 0$ is called a *prefix* of s, in this case we also write $t \preceq s$. If $t = s[1, i]$ for $i \geq 1$, then t is called a *nonempty prefix* of s. A *weight-function* is a homomorphism $f : A^* \to \mathbb{N}$ from the free monoid A^* with concatenation to the natural numbers with addition such that $f(s) = 0$ if and only if $s = \varepsilon$.

We assume some familiarity with computational complexity, see e.g. the textbook [159] for more details. In particular, the classes L (deterministic logarithmic space), NL (nondeterministic logarithmic space) P (deterministic polynomial time), PSPACE (polynomial space), EXPTIME (deterministic exponential time), and EXPSPACE (exponential space) will occur several times in this work. A few times, we will also need alternating classes. Let

ATIME$(a(n), t(n))$ denote the class of all problems that can be solved on an alternating Turing-machine in time $O(t(n))$ with at most $O(a(n))$ alternations [46]. It is well known that that ATIME$(*, t(n))$ is contained in NSPACE$(t(n))$.

2.2 Relational structures and logic

The material in this section will be needed in Chapter 4 and 5. For more details see e.g. [96].

The notion of a structure (or model) is defined as usual in logic. Here we only consider *relational structures*. Sometimes, we will also use constants, but a constant c can be always replaced by the unary relation $\{c\}$. Let us fix a relational structure $\mathcal{A} = (A, (R_i)_{i \in J})$, where $R_i \subseteq A^{n_i}$, $i \in J$. The *signature of* $\mathcal{A}$ contains the equality symbol $=$, and for each $i \in J$ it contains a relation symbol of arity n_i that we denote without risk of confusion by R_i as well. For $B \subseteq A$ we define the restriction $\mathcal{A}{\restriction}B = (B, (R_i \cap B^{n_i})_{i \in J})$, it is a structure over the same signature as $\mathcal{A}$. Let $\mathcal{A} \backslash B = \mathcal{A}{\restriction}(A \backslash B)$. Given further relations R_j, $j \in K$, $J \cap K = \emptyset$, we also write $(\mathcal{A}, (R_i)_{i \in K})$ for the structure $(A, (R_i)_{i \in J \cup K})$. With $\mathrm{Aut}(\mathcal{A})$ we denote the *automorphism group* of $\mathcal{A}$. On the universe A we define the equivalence relation $\sim$ by $a \sim b$ if there exists $f \in \mathrm{Aut}(\mathcal{A})$ with $f(a) = b$. The equivalence classes of $\sim$ are called the *orbits* of $\mathrm{Aut}(\mathcal{A})$ on $\mathcal{A}$.

Next, let us introduce *monadic second-order logic (MSO logic)*. Let $\mathbb{V}_1$ be a countably infinite set of *first-order variables*, which range over elements of the universe A. First-order variables are denoted by x, y, z, x', etc. Let $\mathbb{V}_2$ be a countably infinite set of *second-order variables*, which range over subsets of A. Variables from $\mathbb{V}_2$ are denoted by X, Y, Z, X', etc. *MSO formulas* over the signature of $\mathcal{A}$ are constructed from the atomic formulas $R_i(x_1, \ldots, x_{n_i})$, $x = y$, and $x \in X$ (where $i \in J$, $x_1, \ldots, x_{n_i}, x, y \in \mathbb{V}_1$, and $X \in \mathbb{V}_2$) using Boolean connectives and quantifications over variables from $\mathbb{V}_1$ and $\mathbb{V}_2$. The notion of a free variable is defined as usual. A formula without free variables is called a *sentence* . If $\varphi(x_1, \ldots, x_n, X_1, \ldots, X_m)$ is an MSO formula with free first-order variables among $x_1, \ldots, x_n$ and free second-order variables among $X_1, \ldots, X_m$, and $a_1, \ldots, a_n \in A$, $A_1, \ldots, A_m \subseteq A$, then $\mathcal{A} \models \varphi(a_1, \ldots, a_n, A_1, \ldots, A_m)$ means that φ evaluates to true in $\mathcal{A}$ if the free variable x_i (resp. X_j) evaluates to a_i (resp. A_j). The *MSO theory* of $\mathcal{A}$, denoted by $\mathrm{MSOTh}(\mathcal{A})$, is the set of all MSO sentences φ such that $\mathcal{A} \models \varphi$.

A first-order formula over the signature of $\mathcal{A}$ is an MSO formula that does not contain any occurrences of second-order variables. In particular, first-order formulas do not contain atomic subformulas of the form $x \in X$. The *first-order theory* $\mathrm{FOTh}(\mathcal{A})$ of $\mathcal{A}$ is the set of all first-order sentences φ such that $\mathcal{A} \models \varphi$. The *existential first-order theory* $\exists\mathrm{FOTh}(\mathcal{A})$ of $\mathcal{A}$ is the set of all sentences in $\mathrm{FOTh}(\mathcal{A})$ of the form $\exists x_1 \cdots \exists x_n : \varphi(x_1, \ldots, x_n)$, where $\varphi(x_1, \ldots, x_n)$ is a Boolean combination of atomic formulas.

An important method for proving the decidability of logical theories are interpretations. Let $\mathcal{B}$ be another relational structure with universe B. Then we say that $\mathcal{A}$ is *MSO-interpretable* (resp. *first-order interpretable*) in $\mathcal{B}$ if there exist MSO formulas (resp. first-order formulas) $\psi(x)$ and $\phi_i(\widetilde{x}_i)$ ($i \in J$, $\widetilde{x}_i$ is a tuple of first-order variables of length n_i) over the signature of $\mathcal{B}$ such that the structure $(\psi(x)^{\mathcal{B}}, (\phi_i(\widetilde{x}_i)^{\mathcal{B}})_{i \in J})$ is isomorphic to $\mathcal{A}$. Here $\psi(x)^{\mathcal{B}} = \{b \in B \mid \mathcal{B} \models \psi(b)\}$ and $\phi_i(\widetilde{x}_i) = \{\widetilde{c} \in B^{n_i} \mid \mathcal{B} \models \phi_i(\widetilde{c})\}$. It is easy to see that if $\mathcal{A}$ is MSO-interpretable (resp. first-order interpretable) in $\mathcal{B}$ and $\mathrm{MSOTh}(\mathcal{B})$ (resp. $\mathrm{FOTh}(\mathcal{B})$) is decidable, then also $\mathrm{MSOTh}(\mathcal{A})$ (resp. $\mathrm{FOTh}(\mathcal{A})$) is decidable.

2.3 Monoid presentations

Presentations for monoids are the basic concept of this work. In this section we will introduce the necessary definitions. For more details see [27, 104].

Let Γ be a (possibly infinite) alphabet. A *semi-Thue system* R over Γ is a (possibly infinite) subset $R \subseteq \Gamma^* \times \Gamma^*$, whose elements are called rules. A rule $(s,t) \in R$ will be also written as $s \to t$. The pair (Γ, R) is called a *monoid presentation*. We say that (Γ, R) is *finite*, if both Γ and R are finite. We say that (Γ, R) is *finitely generated*, if Γ is finite. The sets $\mathrm{dom}(R)$ of all left-hand sides and $\mathrm{ran}(R)$ of all right-hand sides are defined by $\mathrm{dom}(R) = \{s \mid \exists t : (s,t) \in R\}$ and $\mathrm{ran}(R) = \{t \mid \exists s : (s,t) \in R\}$, respectively. We define two binary relations $\to_R$ and $\leftrightarrow_R$ on Γ^* as follows:

- $s \to_R t$ if there exist $u, v \in \Gamma^*$ and $(\ell, r) \in R$ with $s = u\ell v$ and $t = urv$ (the *one-step rewrite relation*)
- $s \leftrightarrow_R t$ if ($s \to_R t$ or $t \to_R s$)

We also write $t \,{}_R\!\leftarrow s$ in case $s \to_R t$. For a homomorphism $h : \Gamma^* \to \Sigma^*$ we define the semi-Thue system $h(R) = \{(h(\ell), h(r)) \mid (\ell, r) \in R\}$. Clearly

$s \to_R t$ implies $h(s) \to_{h(R)} h(t)$. Let $\mathrm{RED}(R) = \Gamma^* \mathrm{dom}(R)\Gamma^*$ be the set of *reducible words* and $\mathrm{IRR}(R) = \Gamma^* \setminus \mathrm{RED}(R)$ be the set of *irreducible words* (with respect to R). The presentation (Γ, R) is *terminating* if there does not exist an infinite chain $s_1 \to_R s_2 \to_R s_3 \to_R \cdots$ in Γ^*. The presentation (Γ, R) is *confluent* if for all $s, t, u \in \Gamma^*$ with $t \,{}_R\!\overset{*}{\leftarrow} s \overset{*}{\to}_R u$ there exists $v \in \Gamma^*$ with $t \overset{*}{\to}_R v \,{}_R\!\overset{*}{\leftarrow} u$. It is well-known that (Γ, R) is confluent if and only if (Γ, R) is *Church-Rosser*, i.e., for all $s, t \in \Gamma^*$, if $s \overset{*}{\leftrightarrow}_R t$, then $s \overset{*}{\to}_R u \,{}_R\!\overset{*}{\leftarrow} t$ for some $u \in \Gamma^*$, see [27, p 12]. The presentation (Γ, R) is *locally confluent* if for all $s, t, u \in \Gamma^*$ with $t \,{}_R\!\leftarrow s \to_R u$ there exists $v \in \Gamma^*$ with $t \overset{*}{\to}_R v \,{}_R\!\overset{*}{\leftarrow} u$. By Newman's Lemma [151], a terminating presentation is confluent if and only if it is locally confluent. Moreover, if (Γ, R) is terminating and confluent, then for every $s \in \Gamma^*$ there exists a unique *normal form* $\mathrm{NF}_R(s) \in \mathrm{IRR}(R)$ such that $s \overset{*}{\to}_R \mathrm{NF}_R(s)$.

It is undecidable whether a given finite presentation is terminating [100], confluent [13], or locally confluent [13], respectively. On the other hand, for a finite and terminating presentation, local confluence (and hence by Newman's Lemma also confluence) can be checked using critical pairs [154]: A pair $(s_1, s_2) \in \Gamma^* \times \Gamma^*$ is a *critical pair* of (Γ, R) if there exist rules $(\ell_1, r_1), (\ell_2, r_2) \in R$ such that one of the following two cases holds:

- $\ell_1 = t\ell_2 u$, $s_1 = r_1$, and $s_2 = tr_2 u$ for some $t, u \in \Gamma^*$ (here the word $\ell_1 = t\ell_2 u$ is an overlapping of ℓ_1 and ℓ_2).

- $\ell_1 = ut$, $\ell_2 = tv$, $s_1 = r_1 v$, and $s_2 = ur_2$ for some $t, u, v \in \Gamma^*$ with $t \neq \varepsilon$ (here the word utv is an overlapping of ℓ_1 and ℓ_2).

Note that if (Γ, R) is finite, then there are only finitely many critical pairs. It can be shown that (Γ, R) is locally confluent if and only if for every critical pair (s_1, s_2) there exists s with $s_1 \overset{*}{\to}_R s \,{}_R\!\overset{*}{\leftarrow} s_2$. It follows that for finite and terminating presentations, local confluence and (by Newman's Lemma also) confluence are decidable.

In Chapter 3 we will consider the following classes of terminating presentations: A presentation (Γ, R) is

- *weight-lexicographic* if there exist a linear order $\succ$ on the alphabet Γ and a weight-function $f : \Gamma^* \to \mathbb{N}$ such that for all $(s, t) \in R$ we have either $f(s) > f(t)$ or $(f(s) = f(t)$ and there exists $1 \leq i \leq \min\{|s|, |t|\}$ with $s[1, i-1] = t[1, i-1]$ and $s[i] \succ t[i])$.

- *length-lexicographic* if there exists a linear order $\succ$ on the alphabet Γ such that for all $(s,t) \in R$ we have either $|s| > |t|$ or ($|s| = |t|$ and there exists $1 \leq i \leq \min\{|s|,|t|\}$ with $s[1,i-1] = t[1,i-1]$ and $s[i] \succ t[i]$).
- *weight-reducing* if there exists a weight-function $f : \Gamma^* \to \mathbb{N}$ such that $f(s) > f(t)$ for all $(s,t) \in R$.
- *length-reducing* if $|s| > |t|$ for all $(s,t) \in R$.
- *monadic* if it is length-reducing and $ran(R) \subseteq \Gamma \cup \{\varepsilon\}$.
- *erasing* if it is length-reducing and $\mathrm{ran}(R) = \{\varepsilon\}$.[1]
- *n-homogeneous* ($n \geq 1$) if it is erasing and $|\ell| = n$ for all $\ell \in \mathrm{dom}(R)$.

If Γ is clear from the context, then we also say briefly that the semi-Thue system R has one of the above properties.

The relation $\overset{*}{\leftrightarrow}_R$ is a congruence relation with respect to the concatenation of words, it is called the *Thue-congruence* generated by (Γ, R). Hence we can define the quotient monoid $\Gamma^*/\overset{*}{\leftrightarrow}_R$, which we denote by $\mathcal{M}(\Gamma, R)$. It is also called the *monoid presented by* (Γ, R). In case of a finite presentation (Γ, R) with $\Gamma = \{a_1, \ldots, a_n\}$ and $R = \{(s_1, t_1), \ldots, (s_m, t_m)\}$ we will also use the more readable notation $\mathcal{M}(a_1, \ldots, a_n \mid s_1 = t_1, \ldots, s_m = t_m)$ instead of $\mathcal{M}(\Gamma, R)$. The neutral element of the monoid $\mathcal{M}(\Gamma, R)$ will be usually denoted by 1, possibly with some index. Of course every monoid $\mathcal{M} = (M, \circ, 1)$ is of the form $\mathcal{M}(\Gamma, R)$ for some presentation (Γ, R): let $\Gamma = M \backslash \{1\}$ and $R = \{(ab, c) \mid a, b, c \in \Gamma, a \circ b = c\} \cup \{(ab, \varepsilon) \mid a, b \in \Gamma, a \circ b = 1\}$. Then $\mathcal{M} \cong \mathcal{M}(\Gamma, R)$, see e.g. [27, Thm. 7.1.7]. The presentation (Γ, R) is even monadic and confluent. On the other hand, in this work we are mainly interested in monoids of the form $\mathcal{M}(\Gamma, R)$ with (Γ, R) being finitely generated, i.e., Γ finite. In this case, we say that also $\mathcal{M}(\Gamma, R)$ is *finitely generated*. If (Γ, R) is finite, i.e., both Γ and R are finite, then $\mathcal{M}(\Gamma, R)$ is called *finitely presented*.

Example 2.3.1. *A class of finitely presented monoids that will occur several times in this work are free groups. Let Γ be a finite set. Let $\overline{\Gamma} = \{\overline{a} \mid a \in \Gamma\}$ be a disjoint copy of Γ. Then*

$$F(\Gamma) = \mathcal{M}(\Gamma \cup \overline{\Gamma}, \{(a\overline{a}, \varepsilon), (\overline{a}a, \varepsilon) \mid a \in \Gamma\})$$

[1] Erasing presentations are usually called special, but here we prefer the term "erasing".

is a group, it is called the free group generated by Γ *(as a monoid, $F(\Gamma)$ is of course generated by $\Gamma \cup \overline{\Gamma}$). In case $|\Gamma| = n$, we also write F_n for $F(\Gamma)$ and call F_n the* free group of rank n.

Since we will be interested in the complexity of decision problems, which have a finite monoid presentation as part of their input, we have to define the *size* $\|(\Gamma, R)\|$ *of a finite presentation* (Γ, R). Here, we choose $\|(\Gamma, R)\| = |\Gamma| + \sum_{(s,t)\in R} |st|$.[2]

2.4 The (uniform) word problem

A decision problem that is of fundamental importance in the theory of monoid presentations is the word problem. Let us first introduce a uniform variant of this problem. Let $\mathcal{P}$ be a class of *finite* monoid presentations. The *uniform word problem* for the class $\mathcal{P}$ is the following decision problem:

INPUT: A monoid presentation $(\Gamma, R) \in \mathcal{P}$ and two words $s, t \in \Gamma^*$.

QUESTION: Does $s \overset{*}{\leftrightarrow}_R t$ hold?

Here the length of the input is $\|(\Gamma, R)\| + |st|$.

For the class of all finite, terminating, and confluent presentations the uniform word problem is decidable [112]: In order to check whether $s \overset{*}{\leftrightarrow}_R t$, it suffices to check $\mathrm{NF}_R(s) = \mathrm{NF}_R(t)$.

It should be noted that the uniform word problem is a promise problem in the sense that it is promised that the input presentation indeed belongs to the class $\mathcal{P}$, but a priori this class might be undecidable, for instance this is the case for the class of all terminating and confluent presentations. On the other hand, in this work only decidable classes of presentations will occur.

Now let (Γ, R) be a fixed *finitely generated* presentation. The *word problem* for the presentation (Γ, R) is the following decision problem:

INPUT: Two words $s, t \in \Gamma^*$.

QUESTION: Does $s \overset{*}{\leftrightarrow}_R t$ hold?

In this case, the input size is $|st|$.

Assume that (Γ, R) and (Σ, S) are finitely generated presentations such that $\mathcal{M}(\Gamma, R) \cong \mathcal{M} \cong \mathcal{M}(\Sigma, S)$. Then for every $a \in \Gamma$ there exists a word $w_a \in \Sigma^*$ such that a and w_a represent the same element of $\mathcal{M}$. If we define the homomorphism $h : \Gamma^* \to \Sigma^*$ by $h(a) = w_a$ then for all $s, t \in \Gamma^*$ we have

[2] The binary coding of (Γ, R) has length $\log(|\Gamma|)\cdot\|(\Gamma, R)\|$, but this additional logarithmic factor will not play a crucial role for our considerations.

$s \overset{*}{\leftrightarrow}_R t$ if and only if $h(s) \overset{*}{\leftrightarrow}_S h(t)$. Thus, the word problem for (Γ, R) can be reduced to the word problem for (Σ, S) and vice versa. Moreover this reduction can be realized in deterministic logspace (it can be even realized in uTC^0, see Section 3.2.1). Thus, the decidability and complexity of the word problem does not depend on the chosen presentation. Hence, we may just speak of the word problem for the monoid $\mathcal{M}$.

Next assume that $\mathcal{M}(\Gamma, R)$ is a group $\mathcal{G}$. Let $W(\Gamma, R) = \{w \in \Gamma^* \mid w \overset{*}{\leftrightarrow}_R \varepsilon\}$ denote the set of all words over the generators that represent the identity of the group. Since $\mathcal{G}$ is a group, the set $W(\Gamma, R)$ may be identified with the word problem for $\mathcal{G}$, which is quite common. If (Σ, S) is another presentation for the group $\mathcal{G}$ and $\mathcal{C}$ is some class of languages that is closed under inverse morphisms, then $W(\Gamma, R) \in \mathcal{C}$ if and only if $W(\Sigma, S) \in \mathcal{C}$, see e.g. [93] for a proof. In particular, $W(\Gamma, R)$ is context-free if and only if $W(\Sigma, S)$ is context-free. If the latter holds, then, following [4], the group $\mathcal{G}$ is called *context-free*. The main result of [144] together with [75] implies that a group is context-free if and only if it is *virtually free*, i.e., has a free subgroup of finite index.

2.5 Automatic structures

In this work the concept of an automatic structure will occur repeatedly. In this section we introduce the relevant definitions and state a few basic results, see [20, 109] for more details.

Let us fix $n \in \mathbb{N}$ and a finite alphabet Γ. Let $\# \notin \Gamma$ by an additional padding symbol. We define two encodings $\nu_r : (\Gamma^*)^n \to ((\Gamma \cup \{\#\})^n)^*$ and $\nu_\ell : (\Gamma^*)^n \to ((\Gamma \cup \{\#\})^n)^*$ as follows: Let $w_1, \ldots, w_n \in \Sigma^*$, $|w_i| = k_i$, and let $k = \max\{k_1, \ldots, k_n\}$. Define $u_i = w_i \#^{k-k_i}$, thus $|u_i| = k$. Then

$$\nu_r(w_1, \ldots, w_n) = (u_1[1], \ldots, u_n[1]) \cdots (u_1[k], \ldots, u_n[k]) \text{ and}$$
$$\nu_\ell(w_1, \ldots, w_n) = (\nu_r(w_1^{\mathrm{rev}}, \ldots, w_n^{\mathrm{rev}}))^{\mathrm{rev}}.$$

Thus, for instance

$$\nu_r(aba, bbabb) = (a,b)(b,b)(a,a)(\#,b)(\#,b) \text{ and}$$
$$\nu_\ell(aba, bbabb) = (\#,b)(\#,b)(a,a)(b,b)(a,b).$$

An n-ary relation $R \subseteq (\Gamma^*)^n$ is called α-automatic ($\alpha \in \{\ell, r\}$) if the language $\{\nu_\alpha(\widetilde{u}) \mid \widetilde{u} \in R\}$ is a regular language over the alphabet $(\Gamma \cup \{\#\})^n$.

A relation $R \subseteq (\Gamma^*)^n$ has *length-difference bounded by* $\gamma \in \mathbb{N}$ if for all $(u_1, \ldots, u_n) \in R$ and all $1 \leq i, j \leq n$, it holds $|(|u_i| - |u_j|)| \leq \gamma$ [84]. We say that R has *bounded length-difference* if for some γ it has *length-difference bounded by* γ. The following simple lemma will turn out to be useful later. Its simple proof is left to the reader.

Lemma 2.5.1. *Let $R, S \subseteq (\Gamma^*)^n$ have both bounded length-difference.*

- *R is ℓ-automatic if and only if R is r-automatic.*
- *If R and S are both α-automatic, then*

$$R \cdot S = \{(st, uv) \mid (s, u) \in R, (t, v) \in S\}$$

  is α-automatic as well.

Now let $\mathcal{A} = (A, (R_i)_{i \in J})$ be an arbitrary relational structure with finitely many relations, where $R_i \subseteq A^{n_i}$. Let $\alpha \in \{\ell, r\}$. A tuple (Γ, L, h) is called an *α-automatic presentation* for $\mathcal{A}$ if

- Γ is a finite alphabet,
- $L \subseteq \Gamma^*$ is a regular language,
- $h : L \to A$ is a surjective function,
- the relation $\{(u, v) \in L \times L \mid h(u) = h(v)\}$ is α-automatic, and
- the relation $\{(u_1, \ldots, u_{n_i}) \in L^{n_i} \mid (h(u_1), \ldots, h(u_{n_i})) \in R_i\}$ is α-automatic for every $i \in J$.

We say that $\mathcal{A}$ is *α-automatic* if there exists an α-automatic presentation for $\mathcal{A}$. Whereas an r-automatic presentation for $\mathcal{A}$ is not necessarily an ℓ-automatic presentation for $\mathcal{A}$ and vice versa, it is easy to see that $\mathcal{A}$ is r-automatic if and only if $\mathcal{A}$ is ℓ-automatic. If the latter holds, then we say that $\mathcal{A}$ is an *automatic structure*.[3]

[3]We introduced the concept of α-automatic presentations for both $\alpha = r$ and $\alpha = \ell$ because later we will need both variants in the context of automatic monoids.

Theorem 2.5.2 (cf [109]). *Let (Γ, L, h) be an α-automatic presentation for the structure $\mathcal{A}$ and let $\varphi(x_1, \ldots, x_n)$ be a first-order formula over the signature of $\mathcal{A}$. Then the relation*

$$\{(u_1, \ldots, u_n) \in L^n \mid \mathcal{A} \models \varphi(h(u_1), \ldots, h(u_n))\}$$

is α-automatic. Thus, (Γ, L, h) is also an α-automatic presentation for the structure $(\mathcal{A}, \{(a_1, \ldots, a_n) \mid \mathcal{A} \models \phi(a_1, \ldots, a_n)\})$.

The previous theorem implies that automatic structures are closed under first-order definitions. The following theorem is our main motivation for introducing automatic structures.

Theorem 2.5.3 (cf [109]). *If $\mathcal{A}$ is automatic, then* FOTh($\mathcal{A}$) *is decidable.*

In [20] it was shown that even the extension of first-order logic, which allows to say that there are infinitely many x with $\phi(x)$, is decidable.

2.6 Cayley-graphs

In Chapter 4 we will study Cayley-graphs of monoids and groups, but also for the investigation of word problems in Chapter 3, the concept of the Cayley-graph will turn out to be useful (see Section 3.8).

Let $\mathcal{M} = (M, \circ, 1)$ be a finitely generated monoid and Γ be a finite generating set of $\mathcal{M}$. With $\mathcal{M}$ and Γ we associate the following two relational structures:

$$\mathcal{C}_r(\mathcal{M}, \Gamma) = (M, (\{(u, v) \mid u \circ a = v\})_{a \in \Gamma})$$
$$\mathcal{C}_\ell(\mathcal{M}, \Gamma) = (M, (\{(u, v) \mid a \circ u = v\})_{a \in \Gamma})$$

Whenever we just write $\mathcal{C}(\mathcal{M}, \Gamma)$, we mean $\mathcal{C}_r(\mathcal{M}, \Gamma)$ and call this structure the *(right) Cayley-graph of $\mathcal{M}$ with respect to* Γ. It can be viewed as a directed graph, where every edge has a label from Γ, $\{(u, v) \mid u \circ a = v\}$ is the set of a-labeled edges. Note that since Γ generates $\mathcal{M}$, $\mathcal{C}(\mathcal{M}, \Gamma)$ is a connected graph. The "left Cayley-graph" $\mathcal{C}_\ell(\mathcal{M}, \Gamma)$ will be only needed in Section 3.8 and 4.5.2.

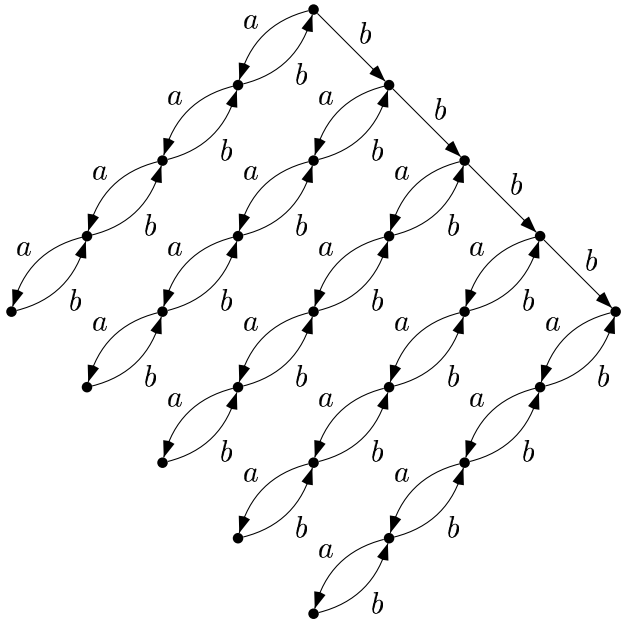
For a finitely generated group $\mathcal{G}$, we always may choose a generating set Γ, which is closed under taking inverses. Thus, for every a-labeled edge of

$\mathcal{C}(\mathcal{G}, \Gamma)$ ($a \in \Gamma$), there exists a reversed a^{-1}-edge. Moreover, for every two nodes $u, v \in \mathcal{G}$ there exists an automorphism of $\mathcal{C}(\mathcal{G}, \Gamma)$, which maps u to v.

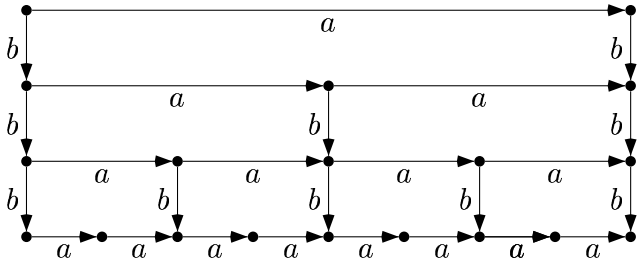
Cayley-graphs were mainly investigated for groups, in particular they play an important role in combinatorial group theory [127, 130], see also the surveys of Schupp [181] and Babai [8]. On the other hand, only a few papers deal with Cayley-graphs for monoids. In [192, 193], Cayley-graphs of automatic monoids are investigated. The work of Calbrix and Knapik on Thue-specifications [39, 111] covers Cayley-graphs of monoids with a terminating and confluent presentation as a special case.

Let us close this section with a few examples.

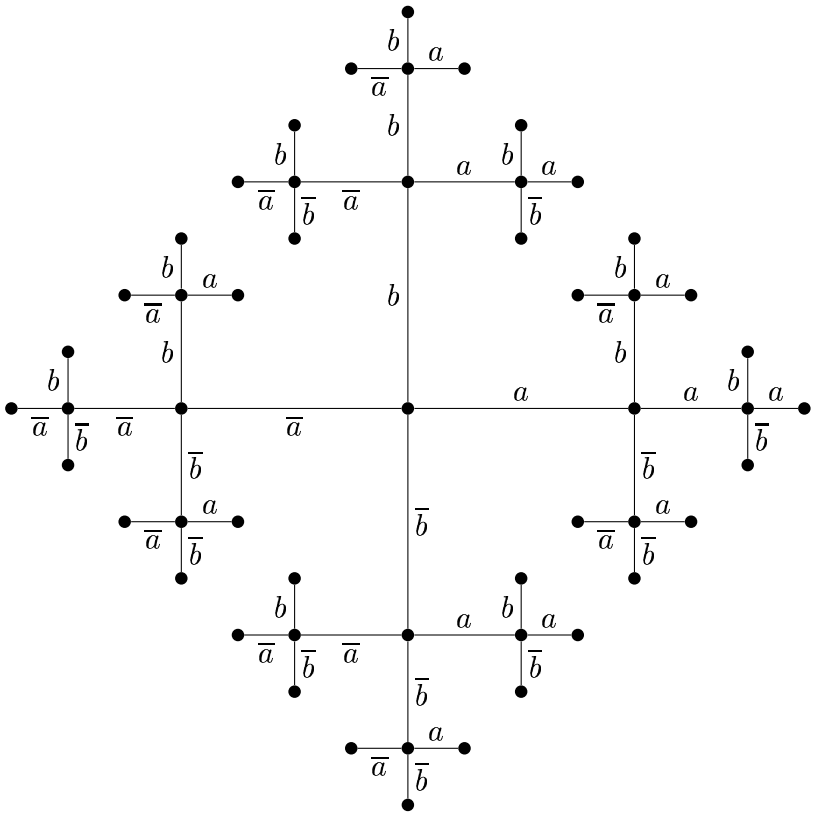
Example 2.6.1. *Let $\mathcal{M} = \mathcal{M}(a, b \mid ab = \varepsilon)$. A finite portion of the Cayley-graph of $\mathcal{M}$ with respect to $\{a, b\}$ looks as follows:*



Example 2.6.2. *Let $\mathcal{M} = \mathcal{M}(a, b \mid ab = baa)$. A finite portion of the Cayley-graph of $\mathcal{M}$ with respect to $\{a, b\}$ looks as follows:*



Example 2.6.3. *A finite portion of the Cayley-graph of F_2, the free group of rank 2, with respect to the generating set $\{a, \overline{a}, b, \overline{b}\}$ is shown below. Two edges which are reversed to each other are represented as a single undirected edge with the label of the edge that points away from the origin, which represents the identity.*



2.7 Mazurkiewicz traces

In Chapter 4 and 5 we will use several results from the theory of Mazurkiewicz traces. A detailed introduction to this field can be found in [70].

An *independence alphabet* is a pair (A, I), where A is a possibly infinite set and $I \subseteq A \times A$ is symmetric and irreflexive. The relation I is known as the *independence relation*, its complement $D = (A \times A) \setminus I$ is the *dependence*

relation. The pair (A, D) is called a *dependence alphabet.* For $a \in A$, we let $I(a) = \{b \in A \mid (a,b) \in I\}$ and $D(a) = \{b \in A \mid (a,b) \in D\} = A \setminus I(a)$. An (A, I)-*clique* is a subset $B \subseteq A$ such that $(a,b) \in I$ for all $a, b \in B$ with $a \neq b$. Let $\mathcal{F}(A, I)$ denote the set of all *finite* (A, I)-cliques. The *trace monoid (free partially commutative monoid)* $\mathbb{M}(A, I)$ associated to (A, I) is defined by

$$\mathbb{M}(A, I) = \mathcal{M}(A, \{(ab, ba) \mid (a,b) \in I\}),$$

its elements are called *traces.* Trace monoids will be one of the few examples of not necessarily finitely generated monoids in this work. Extreme cases are *free monoids* (if $D = A \times A$) and *free commutative monoids* (if $D = \{(a,a) \mid a \in A\}$). Trace monoids were first investigated by Cartier and Foata in combinatorics [43]. In computer science, traces appeared implicitly in the work of Keller [107]. Later, Mazurkiewicz [138] introduced them explicitly in computer science.

The trace represented by the word $s \in A^*$ is denoted by $[s]_I$. The neutral element of $\mathbb{M}(A, I)$ is the empty trace $[\varepsilon]_I$, briefly ε. An element $a \in A$ will be identified with the trace $[a]_I$. More generally, for a finite (A, I)-clique C, we can define a unique trace $[C] = [a_1 a_2 \cdots a_n]_I$, where $a_1, a_2, \ldots, a_n$ is an arbitrary enumeration of C.

Let $t = [s]_I \in \mathbb{M}(A, I)$. We define $|t| = |s|$ (the length of t), $|t|_a = |s|_a$ for $a \in A$, $\mathrm{alph}(t) = \mathrm{alph}(s)$, $\max(t) = \{a \in A \mid \exists u \in A^* : t = [ua]_I\}$, and $\min(t) = \{a \in A \mid \exists u \in A^* : t = [au]_I\}$. Note that $\min(t)$ and $\max(t)$ are (A, I)-cliques. For two traces $t, u \in \mathbb{M}(A, I)$ we write $(t, u) \in I$ if $\mathrm{alph}(t) \times \mathrm{alph}(u) \subseteq I$. For an n-ary relation R over A^*, we define its I-quotient

$$R/_I = \{([u_1]_I, \ldots, [u_n]_I) \mid (u_1, \ldots, u_n) \in R\}.$$

For instance, $\preceq/_I$ is the prefix order on traces. The set of all traces that have t as a prefix is denoted by $t\,\mathbb{M}(A, I) = \{tu \mid u \in \mathbb{M}(A, I)\}$. For instance, $t \in [\min(t)]\mathbb{M}(A, I)$.

A trace $t \in \mathbb{M}(A, I)$ can be visualized by its *dependence graph* D_t. To define D_t, choose an arbitrary word $w = a_1 a_2 \cdots a_n$, $a_i \in A$, with $t = [w]_I$ and define $D_t = (\{1, \ldots, n\}, E, \lambda)$, where $E = \{(i,j) \mid i < j, (a_i, a_j) \in D\}$ and $\lambda(i) = a_i$. If we identify isomorphic dependence graphs, then this definition is independent of the chosen word representing t. Moreover, the mapping $t \mapsto D_t$ is injective. Let $D_t = (V, E, \lambda)$ be the dependence graph of t. It is easy to see that the factors of t are in one-to-one correspondence with the subsets $U \subseteq V$ such that $i \xrightarrow{*} j \xrightarrow{*} k$ and $i, k \in U$ imply $j \in U$.

As a consequence of the representation of traces by dependence graphs, one obtains Levi's lemma for traces, see e.g. [70, p 74], which is one of the fundamental facts in trace theory. The formal statement is as follows.

Lemma 2.7.1. *Let* $u_1, \ldots, u_m, v_1, \ldots, v_n \in \mathbb{M}(A, I)$. *Then*

$$u_1 u_2 \cdots u_m = v_1 v_2 \cdots v_n$$

if and only if there exist $w_{i,j} \in \mathbb{M}(A, I)$ $(1 \leq i \leq m,\ 1 \leq j \leq n)$ *such that*

- $u_i = w_{i,1} w_{i,2} \cdots w_{i,n}$ *for all* $1 \leq i \leq m$,
- $v_j = w_{1,j} w_{2,j} \cdots w_{m,j}$ *for all* $1 \leq j \leq n$, *and*
- $(w_{i,j}, w_{k,\ell}) \in I$ *for all* $1 \leq i < k \leq m,\ n \geq j > \ell \geq 1$.

The situation in the lemma will be visualized by a diagram of the following kind. The i–th column corresponds to u_i, the j–th row corresponds to v_j, and the intersection of the i–th column and the j–th row represents $w_{i,j}$. Furthermore $w_{i,j}$ and $w_{k,\ell}$ are independent if one of them is left-above the other one.

v_n	$w_{1,n}$	$w_{2,n}$	$w_{3,n}$	$\ldots$	$w_{m,n}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
v_3	$w_{1,3}$	$w_{2,3}$	$w_{3,3}$	$\ldots$	$w_{m,3}$
v_2	$w_{1,2}$	$w_{2,2}$	$w_{3,2}$	$\ldots$	$w_{m,2}$
v_1	$w_{1,1}$	$w_{2,1}$	$w_{3,1}$	$\ldots$	$w_{m,1}$
	u_1	u_2	u_3	$\ldots$	u_m

A consequence of Levi's Lemma is that trace monoids are cancellative, i.e., $usv = utv$ implies $s = t$ for all traces $s, t, u, v \in \mathbb{M}(A, I)$.

The *Foata normal form* $\mathrm{FNF}(t)$ of $t \in \mathbb{M}(A, I)$ is a word over the set $\mathcal{F}(A, I)$ of finite (A, I)-cliques. It is defined inductively:

$$\mathrm{FNF}([\varepsilon]_I) = \varepsilon \quad \text{and} \quad \mathrm{FNF}(t) = \min(t)\ \mathrm{FNF}(s)$$

where s is the trace satisfying $t = [\min(t)]s$. Since $\mathbb{M}(A, I)$ is cancellative, s is given uniquely by this requirement. The *height* of t, briefly $\mathrm{height}(t)$, is the length of its Foata normal form $\mathrm{FNF}(t)$. Alternatively, $\mathrm{height}(t)$ can be defined as the number of nodes in a longest directed path in the dependence

graph D_t. Thus, $\mathrm{height}(st) \leq \mathrm{height}(s) + \mathrm{height}(t)$. The *reversed Foata normal form* of $t \in \mathbb{M}(A, I)$ is defined as follows:

$$\mathrm{rFNF}([\varepsilon]_I) = \varepsilon \quad \text{and} \quad \mathrm{rFNF}(t) = \mathrm{rFNF}(s)\ \max(t)$$

where s is the trace uniquely given by $s[\max(t)] = t$. Then $\mathrm{height}(t)$ also equals the length of $\mathrm{rFNF}(t)$. If the word $A_1 A_2 \cdots A_n$, where $A_i \in \mathcal{F}(A, I)$, is the (reversed) Foata normal form of t, then we say that the factorization $t = [A_1][A_2] \cdots [A_n]$ is *in (reversed) Foata normal form.*

We end this section with a brief discussion of *trace rewriting systems*, which generalize semi-Thue systems from words to traces. Formally, a trace rewriting system over $\mathbb{M}(A, I)$ is a subset $R \subseteq \mathbb{M}(A, I) \times \mathbb{M}(A, I)$. Completely analogous to semi-Thue systems we define the one-step rewrite relation $\to_R \subseteq \mathbb{M}(A, I) \times \mathbb{M}(A, I)$, the Thue congruence $\overset{*}{\leftrightarrow}_R \subseteq \mathbb{M}(A, I) \times \mathbb{M}(A, I)$, the set of reducible traces $\mathrm{RED}(R) \subseteq \mathbb{M}(A, I)$, the set of irreducible traces $\mathrm{IRR}(R) \subseteq \mathbb{M}(A, I)$, and terminating (resp. confluent, locally confluent, length-reducing) trace rewriting systems. In general, it is undecidable whether a finite length-reducing trace rewriting system is confluent, see [149]. This is in sharp contrast to semi-Thue systems, and makes confluence proofs challenging. However, if R is terminating and confluent, then still for every $s \in \mathbb{M}(A, I)$ there exists a unique normalform $\mathrm{NF}_R(s) \in \mathrm{IRR}(R)$ with $s \overset{*}{\to}_R \mathrm{NF}_R(s)$. Moreover, the set $\mathrm{IRR}(R)$ is in one-to-one correspondence with the quotient monoid $\mathbb{M}(A, I)/\overset{*}{\leftrightarrow}_R$.

2.8 Rational and recognizable sets

Let $\mathcal{M} = (M, \circ, 1)$ be a monoid. The *product* of two sets $L_1, L_2 \subseteq M$ is $L_1 \circ L_2 = \{a_1 \circ a_2 \mid a_1 \in L_1, a_2 \in L_2\}$. The *Kleene star* of $L \subseteq M$ is $L^* = \bigcup_{i \geq 0} L^i$, where $L^0 = \{1\}$ and $L^{i+1} = L \circ L^i$ for $i \geq 0$. The set $\mathrm{RAT}(\mathcal{M})$ of all *rational subsets* of M is the smallest class of subsets that contains every finite subset of $\mathcal{M}$ and that is closed under union, product, and Kleene star. A subset $L \subseteq M$ is called *recognizable* if there exists a finite monoid S and a monoid homomorphism $h : \mathcal{M} \to S$, which may be assumed to be surjective, such that $L = h^{-1}(h(L))$. The class of all recognizable subsets of M is denoted by $\mathrm{REC}(\mathcal{M})$.

The classes $\mathrm{REC}(\mathcal{M})$ and $\mathrm{RAT}(\mathcal{M})$ are classical, see e.g. [18]. If $\mathcal{M}$ is a finitely generated monoid, then $\mathrm{REC}(\mathcal{M}) \subseteq \mathrm{RAT}(\mathcal{M})$ [139]. In general, $\mathrm{REC}(\mathcal{M})$ is a proper subset of $\mathrm{RAT}(\mathcal{M})$, this holds, e.g., for infinite groups

or the trace monoid $\mathbb{N} \times \{a, b\}^*$. For a free monoid Γ^* we have $\mathrm{REC}(\Gamma^*) = \mathrm{RAT}(\Gamma^*)$ by Kleene's Theorem.

For every monoid $\mathcal{M}$, the class $\mathrm{REC}(\mathcal{M})$ is an effective Boolean algebra, but in general $\mathrm{REC}(\mathcal{M})$ is neither closed under products nor Kleene stars. On the other hand $\mathrm{RAT}(\mathcal{M})$ is not a Boolean algebra in general, for instance $\mathrm{RAT}(\mathbb{N} \times \{a, b\}^*)$ is not closed under intersection, see e.g. [70, Example 6.1.16].

For a trace monoid $\mathbb{M} = \mathbb{M}(A, I)$ with A finite, it is easy to see that $L \in \mathrm{REC}(\mathbb{M})$ if and only if the language $\{u \in A^* \mid [u]_I \in L\}$ is a regular subset of A^*, whereas $L \in \mathrm{RAT}(\mathbb{M})$ if and only if there is a regular language $K \subseteq A^*$ such that $L = \{[u]_I \mid u \in K\}$. Thus, every finite subset of $\mathbb{M}$ is recognizable. Moreover, $\mathrm{REC}(\mathbb{M})$ is closed under products and *connected Kleene stars* [156].[4] In particular, for a finite trace rewriting system R over a trace monoid $\mathbb{M}$, we have $\mathrm{RED}(R) \in \mathrm{REC}(\mathbb{M})$ and $\mathrm{IRR}(R) \in \mathrm{REC}(\mathbb{M})$.

[4] A Kleene star L^*, where $L \subseteq \mathbb{M}$, is called connected if every $t \in L$ is a connected trace, i.e., we cannot write $t = uv$ with $(u, v) \in I$ and $u \neq [\varepsilon]_I \neq v$.

Chapter 3

Word problems

3.1 Outline

This chapter is devoted to a complexity theoretical analysis of word problems for certain classes of monoid presentations.

Adjan obtained in [2] the remarkable result that there exists a fixed 3-homogeneous presentation with an undecidable word problem, whereas every 2-homogeneous presentation has a decidable word problem. Book [24] sharpened Adjan's decidability result by proving that the word problem for every 2-homogeneous presentation can be solved in linear time. In Section 3.2, we will study the space and circuit complexity of word problems for 2-homogeneous presentations. In Section 3.2.3 we will prove that the uniform word problem for the class of all confluent and 2-homogeneous presentations is in deterministic logarithmic space (L). From this result it follows easily that the word problem for every fixed (not necessarily confluent) 2-homogeneous presentation can be solved in L (Theorem 3.2.14), which improves Adjan's decidability result in another direction. Our proof is based on the important result of Lipton and Zalcstein [120] that the word problem for the free group of rank 2 is in L. Formally Theorem 3.2.14 generalizes the result of Lipton and Zalcstein. Furthermore our logarithmic space algorithm immediately shows that the word problem for an arbitrary 2-homogeneous presentation could be solved in uniform NC^1 (see Section 3.2.1 for a definition) if the word problem for the free group of rank 2 would belong to uniform NC^1. Whether the latter holds is one of the major open questions concerning uniform NC^1. In Section 3.2.5 we will consider the uniform word problem for the class of

all 2-homogeneous (not necessarily confluent) presentations. Based on our results from Section 3.2.3, we will show that the uniform word problem for 2-homogeneous presentations is complete for symmetric logarithmic space (SL, see Section 3.2.1 for a definition). This result is in particular interesting from the viewpoint of computational complexity, since there are quite few natural and nonobvious SL-complete problems outside graph theory, see [3].

As already mentioned in the introduction, Bauer and Otto [13] have shown that also for confluent and terminating presentations, the complexity of the word problem can reach every level of the Grzegorczyk hierarchy. A way to reduce the complexity of the word problem is to force bounds on the length of derivation sequences, e.g., by restricting to certain subclasses of terminating systems. For instance, Book has shown that the word problem for a confluent and length-reducing presentation can be solved in linear time [23]. Further results in this direction were obtained in [25]. In Section 3.3-3.6 we will continue the investigation of the complexity of the word problem for various classes of confluent and terminating presentations. Following [27, pp 41–42], we will study length-reducing systems, weight-reducing systems, length-lexicographic systems, and weight-lexicographic systems. For each of these classes we will investigate the word problem in its uniform and nonuniform variant. Moreover, we will also study the effect of fixing the underlying alphabet for the uniform word problem. Table 3.1 at page 61 summarizes the results.

Finally, in Section 3.8 we will prove that there exists a fixed automatic monoid with a P-complete word problem. Automatic monoids generalize automatic groups, which have attracted a lot of attention in combinatorial group theory since the middle 1980s, see [80] for a detailed introduction.

In this chapter we assume that all monoid presentations are finite. This assumption will not be mentioned any more. The results of this chapter are partly contained in [124, 125].

3.2 2-homogeneous presentations

3.2.1 Some complexity classes below P

In this section we will introduce several complexity classes that will turn out to be useful for the investigation of 2-homogeneous presentations. All these classes are contained in P.

Symmetric logarithmic space With *SL (symmetric logarithmic space)* we denote the class of all problems that can be solved in logarithmic space on a symmetric and nondeterministic Turing-machine. Basically a Turing-machine $\mathcal{T}$ is symmetric, if for every transition $c_1 \Rightarrow_{\mathcal{T}} c_2$ the machine can also make the transition $c_2 \Rightarrow_{\mathcal{T}} c_1$, see [119] for more details. We will never work with this original definition of SL from [119] (which has to be enriched by several quite technical additional conditions), but with a logical characterization of SL, see Section 3.2.5. Important results for SL are the closure of SL under logspace bounded Turing-reductions, i.e., $\mathrm{SL} = \mathrm{L}^{\mathrm{SL}}$ [153], and the fact that problems in SL can be solved in deterministic space $O(\log(n)^{\frac{4}{3}})$ [5]. A collection of SL-complete problems can be found in [3].

Circuit complexity For more details on circuit complexity see [215]. We will consider circuit families $(C_n)_{n\in\mathbb{N}}$, where C_n is a Boolean circuit with one output and n inputs that are linearly ordered. Such a circuit family accepts a language over $\{0,1\}$ in the obvious way, but of course this language may be even undecidable. In order to avoid such pathological cases, we assume a uniformity condition, which assures that given n, the n-th circuit C_n can be easily constructed. Here we assume *DLOGTIME-uniformity*, see e.g. [10, 37], which is defined via DLOGTIME Turing-machines. A DLOGTIME Turing-machine is a deterministic Turing-machine that operates in logarithmic time. In order to be able to do nontrivial computations, such a machine has a special input address tape of length $O(\log(n))$ that stores a position i of the input. At every step the machine has access to the input symbol at position i. Despite its seemingly restricted power, several nontrivial computations can be done with DLOGTIME machines, see [10]:

- determine the length of its input.
- add two binary coded numbers of length $O(\log(n))$.
- determine the logarithm of a binary coded number of length $O(\log(n))$.

A circuit family $(C_n)_{n\in\mathbb{N}}$ is called DLOGTIME-uniform if the set

$$\{(t,a,b,y) \mid a \text{ and } b \text{ are numbers of gates in } C_n,\ |y| = n,\\ b \text{ is a child of } a, \text{ and } t \text{ is the type of gate } a\}$$

(the direct connection language of the family) can be decided by a DLOGTIME machine. We will always skip the part of verifying DLOGTIME-uniformity, which is quite tedious but never causes any problems due to the simple structure of the circuit classes that we will present. The reader who is only interested in logspace upper bounds may verify logspace uniformity of our circuit classes, which is straight-forward.

DLOGTIME-uniform NC^k, briefly uNC^k,[1] is the class of all languages that can be accepted by a DLOGTIME-uniform family of circuits $(C_n)_{n\in\mathbb{N}}$ such that for every $n \geq 0$,

- the n-th circuit C_n has $n^{O(1)}$ many gates and depth $O(\log^k(n))$, and
- is built up from NOT-gates, and AND- and OR-gates of fan-in 2.

Finally $\mathrm{uNC} = \bigcup_{k\geq 1} \mathrm{uNC}^k$. Roughly speaking, uNC is the class of all problems in P that can be solved efficiently on a parallel machine. It is well known that uNC^1 corresponds to the class ALOGTIME (alternating logarithmic time) [176].

An important subclass of uNC^1 is *DLOGTIME-uniform* TC^0, briefly uTC^0.[2] It is the class of all languages that can be accepted by a DLOGTIME-uniform family of circuits $(C_n)_{n\in\mathbb{N}}$ such that

- the n-th circuit C_n has $n^{O(1)}$ many gates and depth $O(1)$, and
- is built up from NOT-gates, and AND-, OR-, and majority-gates of unbounded fan-in.

Recall that a majority-gate outputs 1 if more than half of its inputs receive 1, otherwise it outputs 0.

The following chain of inclusions holds between the classes inside P, introduced so far.

$$\mathrm{uTC}^0 \subseteq \mathrm{uNC}^1 \subseteq \mathrm{L} \subseteq \mathrm{SL} \subseteq \mathrm{NL} \subseteq \mathrm{uNC}^2 \subseteq \mathrm{uNC} \subseteq \mathrm{P}$$

When dealing with hardness of problems for classes below L, like uNC^1 or uTC^0, logspace-reductions lead to trivial statements. Thus, reductions of more restricted computational power are necessary. A popular choice in

[1] The mnemonic NC (Nick's class) was introduced by Cook [50] in recognition to the work of Nick Pippenger [161].

[2] The "T" in uTC^0 stands for "threshold".

this context are *DLOGTIME-reductions*, see e.g. [10, 37]: A function $f : \{0,1\}^* \to \{0,1\}^*$ can be calculated in DLOGTIME if $|f(x)| \in |x|^{O(1)}$ for every input x (i.e., f has only polynomial growth) and the set of all tuples (c, i, x) such that the i-th symbol of $f(x)$ is c belongs to DLOGTIME.

By allowing more than one output gate in circuits we can also speak of functions that can be calculated in uTC^0. But with this definition only functions $f : \{0,1\}^* \to \{0,1\}^*$, which satisfy the requirement that $|f(x)| = |f(y)|$ if $|x| = |y|$ could be computed. In order to overcome this restriction we define for a function $f : \{0,1\}^* \to \{0,1\}^*$ the function $\mathrm{pad}(f) : \{0,1\}^* \to \{0,1\}^*\{\#\}^*$ by $\mathrm{pad}(f)(x) = f(x)\#^{n-|f(x)|}$ with $n = \max\{|f(z)| \mid |x| = |z|\}$. Then we say that a function f can be calculated in uTC^0 if the function $\mathrm{pad}(f)$ can be calculated by a family of circuits that satisfy the restrictions for uTC^0 (where the alphabet $\{0,1,\#\}$ has to be encoded into the binary alphabet $\{0,1\}$). Typical examples of functions that can be calculated in uTC^0 are the following:

- the sum of a polynomial number of binary coded integers of polynomial length [10].
- the product of a polynomial number of binary coded integers of polynomial length [95].
- the quotient of two binary coded integers of polynomial length [95].

The above definition of a functional equivalent to uTC^0 immediately leads to a notion of uTC^0 many-one reducibility. More generally we say that a language A is uTC^0-*reducible* to a language B if A can be recognized by a DLOGTIME-uniform circuit family $(C_n)_{n\in\mathbb{N}}$ such that

- the n-th circuit C_n has $n^{O(1)}$ many gates and depth $O(1)$, and
- is built up from NOT-gates, and AND-, OR-, majority-, and oracle-gates for the language B of unbounded fan-in.

This notion of reducibility is a special case of Cook's NC^1-reducibility [51]. In particular [51, Proposition 4.1] immediately implies that L is closed under uTC^0-reductions. Moreover also uNC^1 and uTC^0 are closed under uTC^0-reductions and uTC^0-reducibility is transitive.

Remark 3.2.1. *A typical example of a* uTC^0*-reduction is a homomorphism between two free monoids.*[3] *In particular, it follows that if* (Γ, R) *and* (Σ, S) *are two presentations of the same monoid, then the word problem for* (Γ, R) *is* uTC^0*-reducible to the word problem for* (Σ, S) *and vice versa. Thus, it makes sense to say that the word problem for a finitely generated monoid belongs to* uTC^0 *or* uNC^1*, respectively.*

3.2.2 Some known results for 2-homogeneous presentations

Most of the results in Section 3.2 are based on the following result of Lipton and Zalcstein.

Theorem 3.2.2 (cf [120]). *The word problem for the free group* F_2 *of rank 2 is in L.*

Lipton and Zalcstein obtained this result as a special case of a more general result on finitely generated linear groups. We will need a uniform variant of Theorem 3.2.2. For this let us choose the canonical presentation (Δ_n, S_n) of the free group F_n of rank n, where

$$\Delta_n = \{c_1, \ldots, c_n, \overline{c}_1, \ldots, \overline{c}_n\}$$
$$S_n = \{c_i\overline{c}_i \to \varepsilon,\ \overline{c}_i c_i \to \varepsilon \mid 1 \leq i \leq n\}.$$

We obtain the following result:

Corollary 3.2.3. *The uniform word problem for the class* $\{(\Delta_n, S_n) \mid n \geq 1\}$ *is* uTC^0*-reducible to the word problem for* F_2*, and therefore is also in L.*

Proof. The group homomorphism $\varphi_n : F_n \to F_2$ defined by $c_i \mapsto \overline{c}_1^i c_2 c_1^i$ is injective, see e.g. [127, Proposition 3.1]. Furthermore $\varphi_n(w)$ can be calculated from w and (Δ_n, S_n) in uTC^0. The second statement of the theorem follows with Theorem 3.2.2. □

Let us also mention the following result of Robinson [171].

Theorem 3.2.4 (cf [171]). *The word problem for the free group* F_2 *of rank 2 is* uNC^1*-hard under DLOGTIME-reductions.*

[3] For this we have to use the fact that the sum of a polynomial number of integers can be computed in uTC^0.

The Dyck-language over 2 bracket pairs is the set of words over the alphabet $\{a, b, \overline{a}, \overline{b}\}$ that are well-bracketed, where $\overline{a}$ (resp. $\overline{b}$) is the right bracket that corresponds to the left bracket a (resp. b). Using the fact that the number of 1's in a word over $\{0,1\}$ can be calculated in uTC^0 [10], the following result was shown in [9].

Theorem 3.2.5 (cf [9]). *The Dyck-language over 2 bracket pairs is in* uTC^0.

3.2.3 The confluent case

In this section we investigate the uniform word problem for the class of all confluent and 2-homogeneous presentations. For the rest of this section let (Γ, R) be a confluent and 2-homogeneous presentation. W.l.o.g. we may assume that every symbol in Γ appears in some rule of R.

Lemma 3.2.6. *There exist pairwise disjoint sets* $\Gamma_\ell, \Gamma_r, \Delta \subseteq \Gamma$, *an involution* $\overline{\ }: \Delta \to \Delta$, *and a semi-Thue system* $D \subseteq \{(ab, \varepsilon) \mid a \in \Gamma_\ell,\ b \in \Gamma_r\}$ *such that* $\Gamma = \Gamma_\ell \cup \Gamma_r \cup \Delta$ *and* $R = D \cup \{(a\overline{a}, \varepsilon) \mid a \in \Delta\}$. *Furthermore, given* R *and* $a \in \Gamma$ *we can decide in* uTC^0 *whether* a *belongs to* Γ_ℓ, Γ_r, *or* Δ.

Proof. Define $\Sigma_1, \Sigma_2 \subseteq \Gamma$ by $\Sigma_1 = \{a \in \Gamma \mid \exists b \in \Gamma : (ab, \varepsilon) \in R\}$ and $\Sigma_2 = \{a \in \Gamma \mid \exists b \in \Gamma : (ba, \varepsilon) \in R\}$. Let $\Gamma_\ell = \Sigma_1 \backslash \Sigma_2$, $\Gamma_r = \Sigma_2 \backslash \Sigma_1$, and $\Delta = \Sigma_1 \cap \Sigma_2$. Obviously Γ_ℓ, Γ_r, and Δ are pairwise disjoint and $\Gamma = \Gamma_\ell \cup \Gamma_r \cup \Delta$. Now let $a \in \Delta$. Then there exist $b, c \in \Gamma$ with $(ab, \varepsilon), (ca, \varepsilon) \in R$. It follows $b \,{}_R\!\leftarrow cab \to_R c$. Since R is confluent we get $b = c$, i.e, $(ab, \varepsilon), (ba, \varepsilon) \in R$ and thus $b \in \Delta$. Now assume that also $(ab', \varepsilon) \in R$ for some $b' \neq b$. Then $b \,{}_R\!\leftarrow bab' \to_R b'$, which contradicts the confluence of R. Similarly there cannot exist $b' \neq b$ with $(b'a, \varepsilon) \in R$. Thus, we can define an involution $\overline{\ }: \Delta \to \Delta$ by $\overline{a} = b$ if $(ab, \varepsilon), (ba, \varepsilon) \in R$. The lemma follows easily. □

Note that the involution $\overline{\ }: \Delta \to \Delta$ may have fixed points, i.e., $\overline{a} = a$ for some $a \in \Delta$. For the rest of this section it is helpful to eliminate these fixed points. Let $a \in \Delta$ such that $\overline{a} = a$. Take a new symbol a' and redefine the involution $\overline{\ }$ on the alphabet $\Gamma \cup \{a'\}$ by $\overline{a} = a'$ and $\overline{a'} = a$. Let $R' = (R \cup \{(aa', \varepsilon), (a'a, \varepsilon)\}) \backslash \{(aa, \varepsilon)\}$, which is still confluent. Furthermore for $w \in \Gamma^*$ let $w' \in (\Gamma \cup \{a'\})^*$ be the word that results from s by replacing the i-th occurrence of a in w by a' if i is odd and leaving all other occurrences of symbols unchanged. Then for all $s, t \in \Gamma^*$ we have $s \overset{*}{\leftrightarrow}_R t$ if and only if $s' \overset{*}{\leftrightarrow}_{R'} t'$. Since counting modulo 2 can be done in uTC^0, s', t', and R' can

be calculated from s, t, and R in uTC^0. In this way we can eliminate all fixed points of the involution $\overline{\ }$ in uTC^0 (we can do the above replacement in parallel for every fixed point of $\overline{\ }$). Thus, for the rest of the section we may assume that $a \neq \overline{a}$ for all $a \in \Delta$. Let $S = \{(a\overline{a}, \varepsilon) \mid a \in \Delta\} \subseteq R$. Then $\mathcal{M}(\Delta, S)$ is the free group of rank $|\Delta|/2$.

Define the homomorphism $\pi : \Gamma^* \to \{\langle, \rangle\}^*$ by $\pi(a) = \langle$ for $a \in \Gamma_\ell$, $\pi(b) = \rangle$ for $b \in \Gamma_r$, and $\pi(c) = \varepsilon$ for $c \in \Delta$. We say that a word $w \in \Gamma^*$ is *well-bracketed* if the word $\pi(w)$ is well-bracketed. It is easy to see that if $w \xrightarrow{*}_R \varepsilon$, then w is well-bracketed. Furthermore, Theorem 3.2.5 implies that for a word w and two positions $i, j \in \{1, \ldots, |w|\}$ we can check in uTC^0 whether the factor $w[i, j]$ is well-bracketed. We say that two positions $i, j \in \{1, \ldots, |w|\}$ are *corresponding brackets in* w, briefly $\mathrm{co}_w(i, j)$, if $i < j$, $w[i] \in \Gamma_\ell$, $w[j] \in \Gamma_r$, $w[i, j]$ is well-bracketed, and $w[i, k]$ is not well-bracketed for all k with $i < k < j$. Again it can be checked in uTC^0, whether two positions are corresponding brackets. If w is well-bracketed, then we can factorize w uniquely as $w = s_0 w[i_1, j_1] s_1 \cdots w[i_n, j_n] s_n$, where $n \geq 0$, $\mathrm{co}_w(i_k, j_k)$ for all $k \in \{1, \ldots, n\}$, and $s_k \in \Delta^*$ for all $k \in \{0, \ldots, n\}$. We define $\delta(w) = s_0 s_1 \cdots s_n \in \Delta^*$.

Lemma 3.2.7. *The partial function $\delta : \Gamma^* \to \Delta^*$ (which is only defined on well-bracketed words) can be calculated in* uTC^0.

Proof. First, in parallel for every $m \in \{1, \ldots, |w|\}$ we calculate in uTC^0 the value $f_m \in \{0, 1\}$, where $f_m = 0$ if and only if there exist positions $i \leq m \leq j$ such that $\mathrm{co}_w(i, j)$. Next we calculate in parallel for every $m \in \{1, \ldots, |w|\}$ the sum $F_m = \sum_{i=1}^{m} f_i$, which is possible in uTC^0 by [10]. If $F_{|w|} < m \leq |w|$, then the m-th output block is set to the binary coding of the padding symbol $\#$. If $m \leq F_{|w|}$ and $i \in \{1, \ldots, |w|\}$ is such that $f_i = 1$ and $F_i = m$, then the m-th output block is set to the binary coding of $w[i]$. □

Lemma 3.2.8. *Let $w \in \Gamma^*$. If $w \xrightarrow{*}_R s$ for some $s \in \Delta^*$, then w is well-bracketed. Moreover, if we write $w = s_0 w[i_1, j_1] s_1 \cdots w[i_n, j_n] s_n$, where $n \geq 0$, $\mathrm{co}_w(i_k, j_k)$ for all $1 \leq k \leq n$, and $s_k \in \Delta^*$ for all $0 \leq k \leq n$, then:*

- $(w[i_k] w[j_k], \varepsilon) \in R$ *and*
- $w[i_k + 1, j_k - 1] \xrightarrow{*}_R \varepsilon$ *for all* $1 \leq k \leq n$.

In particular, $w \xrightarrow{*}_R s_0 s_1 \cdots s_n = \delta(w)$.

Proof. By projecting the derivation $w \xrightarrow{*}_R s \in \Delta^*$ onto the alphabet $\Gamma \setminus \Delta$, it becomes obvious that w must be well-bracketed. Hence, we can write $w = s_0 w[i_1, j_1] s_1 \cdots w[i_n, j_n] s_n$, where $n \geq 0$, $\mathrm{co}_w(i_k, j_k)$ for all $1 \leq k \leq n$, and $s_k \in \Delta^*$ for all $0 \leq k \leq n$. We prove $(w[i_k]w[j_k], \varepsilon) \in R$ and $w[i_k{+}1, j_k{-}1] \xrightarrow{*}_R \varepsilon$ $(1 \leq k \leq n)$ by induction on the length of the derivation $w \xrightarrow{*}_R s \in \Delta^*$. The case that this derivation has length 0, i.e., $w \in \Delta^*$, is trivial. If the removed occurrence of ℓ in w lies completely within one of the factors s_k $(0 \leq k \leq n)$ or $w[i_k + 1, j_k - 1]$ $(1 \leq k \leq n)$ of w, then we can directly apply the induction hypothesis to $w_1 w_2$. On the other hand, if the removed occurrence of ℓ contains one of the positions i_k or j_k $(1 \leq k \leq n)$, then, since $\mathrm{co}_w(i_k, j_k)$, we must have $\ell = w[i_k]w[j_k]$, $w[i_k + 1, j_k - 1] = \varepsilon$, and $w_1 w_2 = s_0 w[i_1, j_1] s_1 \cdots w[i_{k-1}, j_{k-1}] s_{k-1} s_k w[i_{k+1}, j_{k+1}] s_{k+1} \cdots w[i_n, j_n] s_n \xrightarrow{*}_R s$. Again, we can conclude by using the induction hypothesis. □

Lemma 3.2.9. *Let $w \in \Gamma^*$. Then $w \xrightarrow{*}_R \varepsilon$ if and only if*

- *w is well-bracketed,*
- *$\delta(w) \xrightarrow{*}_S \varepsilon$, and*
- *for all $i, j \in \{1, \ldots, |w|\}$ with $\mathrm{co}_w(i, j)$ we have: $(w[i]w[j], \varepsilon) \in R$ and $\delta(w[i + 1, j - 1]) \xrightarrow{*}_S \varepsilon$.*

Proof. The only if-direction can be shown by an induction on $|w|$ as follows. Let $w \xrightarrow{*}_R \varepsilon$. Then w must be well-bracketed, thus we can factorize w as $w = s_0 w[i_1, j_1] s_1 \cdots w[i_n, j_n] s_n$, where $n \geq 0$, $\mathrm{co}_w(i_k, j_k)$ for all $k \in \{1, \ldots, n\}$, and $s_k \in \Delta^*$ for all $k \in \{0, \ldots, n\}$. By Lemma 3.2.8, we obtain $(w[i_k]w[j_k], \varepsilon) \in R$ and $w[i_k + 1, j_k - 1] \xrightarrow{*}_R \varepsilon$ for all $k \in \{1, \ldots, n\}$. Moreover, $w \xrightarrow{*}_R \delta(w)$. Thus, since R is confluent, $\delta(w) \xrightarrow{*}_S \varepsilon$. Since $|w[i_k + 1, j_k - 1]| < |w|$, we can apply the induction hypothesis to each of the words $w[i_k + 1, j_k - 1]$, which proves the only if-direction. For the other direction assume that w satisfies all three assumptions from the lemma. Then, by induction on $j - i$, we can show that $w[i, j] \xrightarrow{*}_R \varepsilon$ for all $i, j \in \{1, \ldots, |w|\}$ with $\mathrm{co}_w(i, j)$. Together with $\delta(w) \xrightarrow{*}_S \varepsilon$ we get $w \xrightarrow{*}_R \varepsilon$. □

The previous lemma implies the following partial result.

Lemma 3.2.10. *The following problem is* uTC^0*-reducible to the word problem for F_2:*

INPUT: A confluent and 2-homogeneous presentation (Γ, R) and $w \in \Gamma^*$.
QUESTION: Does $w \xrightarrow{*}_R \varepsilon$ (or equivalently $w \overset{*}{\leftrightarrow}_R \varepsilon$) hold?

Proof. A circuit with oracle gates for the word problem for F_2 that on input (Γ, R, w) determines whether $w \xrightarrow{*}_R \varepsilon$ can be easily build using Lemma 3.2.9. The quantification over all pairs $i, j \in \{1, \ldots, |w|\}$ in Lemma 3.2.9 corresponds to an AND-gate of unbounded fan-in. In order to check whether $\delta(w[i,j]) \xrightarrow{*}_S \varepsilon$ for two positions i and j, we first calculate in uTC^0 the word $\delta(w[i,j])$ using Lemma 3.2.7. Next we apply Corollary 3.2.3, and finally we use an oracle gate for the word problem for F_2. □

For $w \in \Gamma^*$ we define the set $\Pi(w)$ as the set of all positions $i \in \{1, \ldots, |w|\}$ such that $w[i] \in \Gamma_\ell \cup \Gamma_r$ and furthermore there does not exist a position $k > i$ with $w[i,k] \xrightarrow{*}_R \varepsilon$ and there does not exist a position $k < i$ with $w[k,i] \xrightarrow{*}_R \varepsilon$. Thus, $\Pi(w)$ is the set of all positions in w whose corresponding symbols are from $\Gamma_\ell \cup \Gamma_r$ but which cannot be deleted in any derivation starting from w. The following lemma should be compared with [158, Lem. 5.4] which makes a similar statement for arbitrary erasing presentations.

Lemma 3.2.11. *Let $u, v \in \Gamma^*$, $\Pi(u) = \{i_1, \ldots, i_m\}$ and $\Pi(v) = \{j_1, \ldots, j_n\}$, where $i_1 < i_2 < \cdots < i_m$ and $j_1 < j_2 < \cdots < j_n$. Define $i_0 = j_0 = 0$, $i_{m+1} = |u| + 1$, and $j_{n+1} = |v| + 1$. Then $u \xleftrightarrow{*}_R v$ if and only if*

- *$m = n$,*
- *$u[i_k] = v[j_k]$ for $1 \leq k \leq n$, and*
- *$\delta(u[i_k + 1, i_{k+1} - 1]) \xleftrightarrow{*}_S \delta(v[j_k + 1, j_{k+1} - 1])$ for $0 \leq k \leq n$.*

Proof. First we show the following statement:

$$\text{If } \Pi(w) = \emptyset \text{ for } w \in \Gamma^*, \text{ then } w \text{ is well-bracketed and } w \xrightarrow{*}_R \delta(w). \quad (3.1)$$

Since $\Pi(w) = \emptyset$ and R is confluent, there must exist $s \in \Delta^*$ with $w \xrightarrow{*}_R s$. Thus, (3.1) follows from Lemma 3.2.8.

Now we prove the lemma. Consider a factor $u_k := u[i_{k-1} + 1, i_k - 1]$ of u. Let $i_{k-1} < i < i_k$ such that $u[i] \in \Gamma_\ell$. Then $i \notin \Pi(u)$, hence there exists $j > i$ such that $u[i,j] \xrightarrow{*}_R \varepsilon$. But since $i_k \in \Pi(u)$ (or $i_k = |u|+1$) we must have $j < i_k$. A similar argument applies if $u[i] \in \Gamma_r$, hence $\Pi(u_k) = \emptyset$ and thus $u_k \xrightarrow{*}_R \delta(u_k)$ by (3.1). We obtain $u \xrightarrow{*}_R \delta(u_1)u[i_1]\delta(u_2)u[i_2] \cdots \delta(u_m)u[i_m]\delta(u_{m+1}) =: u'$ and similarly $v \xrightarrow{*}_R \delta(v_1)v[j_1]\delta(v_2)v[j_2] \cdots \delta(v_n)v[j_n]\delta(v_{n+1}) =: v'$. Thus, $u \xleftrightarrow{*}_R v$ if and only if $u' \xleftrightarrow{*}_R v'$ if and only if u' and v' can be reduced to a common word. But only the factors $\delta(u_k)$ and $\delta(v_k)$ of u' and v', respectively, are reducible. The lemma follows easily. □

From the previous lemma we can infer the following theorem.

Theorem 3.2.12. *The uniform word problem for the class of all confluent and 2-homogeneous presentations is* uTC^0*-reducible to the word problem for the free group* F_2 *of rank 2.*

Proof. Let (Γ, R) be confluent and 2-homogeneous and $u, v \in \Gamma^*$. First we calculate in parallel for all $i, j \in \{1, \ldots, |u|\}$ with $i < j$ the Boolean value $e_{i,j}$, which is FALSE if and only if $u[i,j] \xrightarrow{*}_R \varepsilon$. Next we calculate in parallel for all $i \in \{1, \ldots, |u|\}$ the number $g_i \in \{0, 1\}$ by

$$g_i = \left\{ \begin{array}{ll} 1 & \text{if } u[i] \in \Gamma_\ell \cup \Gamma_r \ \wedge \bigwedge_{k=1}^{i-1} e_{k,i} \ \wedge \bigwedge_{k=i+1}^{|u|} e_{i,k} \\ 0 & \text{else} \end{array} \right\}.$$

Thus, $g_i = 1$ if and only if $i \in \Pi(u)$. Similarly we calculate for all $j \in \{1, \ldots, |v|\}$ the number $h_j \in \{0, 1\}$, which is 1 if and only if $j \in \Pi(v)$. W.l.o.g. we assume that $g_1 = g_{|u|} = h_1 = h_{|v|} = 1$, this can be enforced by appending suitable symbols to the left and right end of u and v, respectively. Now we calculate in parallel for all $i \in \{1, \ldots, |u|\}$ and all $j \in \{1, \ldots, |v|\}$ the sums $G_i = \sum_{k=1}^{i} g_k$ and $H_j = \sum_{k=1}^{j} h_k$, which can be done in uTC^0 by [10]. Finally by Lemma 3.2.11, we have $u \overset{*}{\leftrightarrow}_R v$ if and only if $G_{|u|} = H_{|v|}$ and furthermore for all $i_1, i_2 \in \{1, \ldots, |u|\}$ and all $j_1, j_2 \in \{1, \ldots, |v|\}$ such that $(g_{i_1} = g_{i_2} = h_{j_1} = h_{j_2} = 1, G_{i_1} = H_{j_1}$, and $G_{i_2} = H_{j_2} = G_{i_1} + 1)$ we have

- $u[i_1] = v[j_1]$,
- $u[i_2] = v[j_2]$, and
- $\delta(u[i_1 + 1, i_2 - 1]) \overset{*}{\leftrightarrow}_S \delta(v[j_1 + 1, j_2 - 1])$.

Using Corollary 3.2.3, Lemma 3.2.7, and Lemma 3.2.10 the above description can be easily converted into a uTC^0-reduction to the word problem for F_2. □

Corollary 3.2.13. *The uniform word problem for 2-homogeneous and confluent presentations is in L. Furthermore if the word problem for* F_2 *is in* uNC^1*, then the uniform word problem for the class of all 2-homogeneous and confluent presentations is in* uNC^1.

3.2.4 The nonuniform case

Book has shown in [24] that for every 2-homogeneous presentation there exists a confluent and 2-homogeneous presentation, which presents the same monoid (see also Lemma 3.2.23). Together with Corollary 3.2.13 and Remark 3.2.1 we obtain the following result:

Theorem 3.2.14. *Let (Γ, R) be a fixed 2-homogeneous presentation. Then the word problem for $\mathcal{M}(\Gamma, R)$ is in L. Furthermore if the word problem for F_2 is in* uNC^1*, then also the word problem for $\mathcal{M}(\Gamma, R)$ is in* uNC^1*.*

In the next theorem we present lower bounds for word problems for 2-homogeneous presentations. It deals w.l.o.g. only with confluent and 2-homogeneous presentations. We use the notation from Lemma 3.2.6.

Theorem 3.2.15. *Let (Γ, R) be a confluent and 2-homogeneous presentation, where $\Gamma = \Gamma_\ell \cup \Gamma_r \cup \Delta$. Let $|\Delta| = 2 \cdot n + f$, where f is the number of fixed points of the involution $^- : \Delta \to \Delta$. If $n + f \geq 2$ but not ($n = 0$ and $f = 2$), then the word problem for $\mathcal{M}(\Gamma, R)$ is* uNC^1*-hard under DLOGTIME-reductions. If $n + f < 2$ or ($n = 0$ and $f = 2$), then the word problem for $\mathcal{M}(\Gamma, R)$ is in* uTC^0*.*

Proof. If we do not remove the fixed points of the involution $^- : \Delta \to \Delta$, then the considerations from Section 3.2.3 imply that the word problem for $\mathcal{M}(\Gamma, R)$ is uTC^0-reducible to the word problem for $\mathcal{G} = F_n * \mathbb{Z}/2\mathbb{Z} * \cdots * \mathbb{Z}/2\mathbb{Z}$, where $*$ constructs the free product and we take f copies of $\mathbb{Z}/2\mathbb{Z}$ (each fixed point of the involution generates a copy of $\mathbb{Z}/2\mathbb{Z}$, and the remaining $2n$ many elements in Δ generate F_n). The case $n + f = 0$ is clear. If $n + f = 1$, then either $\mathcal{G} = \mathbb{Z}$ or $\mathcal{G} = \mathbb{Z}/2\mathbb{Z}$. For both groups the word problem is in uTC^0. If $n = 0$ and $f = 2$, then $\mathcal{G} = \mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$. Now $\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$ is a solvable group, see [171, Lem. 6.9]. Furthermore, if we choose two generators a and b of $\mathcal{G}$, where $a^2 = b^2 = 1$ in $\mathcal{G}$, then the number of elements of $\mathcal{G}$ definable by words over $\{a, b\}$ of length at most n grows only polynomially in n, i.e., $\mathcal{G}$ has a polynomial growth function. Now [171, Thm. 7.6] implies that the word problem for $\mathcal{G}$ is in uTC^0. Finally let $n + f \geq 2$ but not ($n = 0$ and $f = 2$). Then $\mathcal{G} = \mathcal{G}_1 * \mathcal{G}_2$, where either $\mathcal{G}_1 \not\cong \mathbb{Z}/2\mathbb{Z}$ or $\mathcal{G}_2 \not\cong \mathbb{Z}/2\mathbb{Z}$. Hence, $\mathcal{G}$ has F_2 as a subgroup, see e.g. the remark in [127, p 177]. Theorem 3.2.4 implies that the word problem for $\mathcal{G}$ and thus also the word problem for $\mathcal{M}(\Gamma, R)$ is uNC^1-hard under DLOGTIME-reductions. □

3.2.5 The general uniform case

The aim of this section is to prove that the uniform word problem for 2-homogeneous (not necessarily confluent) presentations is SL-complete. The main difficulty will be to prove membership in SL.

Throughout this section let (Γ, R) be an arbitrary 2-homogeneous presentation, which is not necessarily confluent. The following obvious fact will be used several times for the further consideration.

Lemma 3.2.16. *Let $a, b \in \Gamma$ such that $a \overset{*}{\leftrightarrow}_R b$ and define a homomorphism $h : \Gamma^* \to (\Gamma \setminus \{a\})^*$ by $h(a) = b$ and $h(c) = c$ for all $c \in \Gamma \setminus \{a\}$. Then for all $s, t \in \Gamma^*$ we have $s \overset{*}{\leftrightarrow}_R t$ if and only if $h(s) \overset{*}{\leftrightarrow}_{h(R)} h(t)$.*

A word $w = a_1 a_2 \cdots a_n \in \Gamma^*$, where $n \geq 1$ and $a_i \in \Gamma$ for $i \in \{1, \ldots, n\}$, is an *R-path from a_1 to a_n* if for all $i \in \{1, \ldots, n-1\}$, we have $(a_i a_{i+1}, \varepsilon) \in R$ or $(a_{i+1} a_i, \varepsilon) \in R$. Let $\triangleright$ and $\triangleleft$ be two new symbols. For an R-path $w = a_1 \cdots a_n$, the set $D_R(w) \subseteq \{\triangleright, \triangleleft\}^*$ contains all words of the form $d_1 \cdots d_{n-1}$ such that for all $i \in \{1, \ldots, n-1\}$ if $d_i = \triangleright$ (resp. $d_i = \triangleleft$), then $(a_i a_{i+1}, \varepsilon) \in R$ (resp. $(a_{i+1} a_i, \varepsilon) \in R$). Since R may contain two rules of the form (ab, ε) and (ba, ε), the set $D_R(w)$ may contain more than one word. We define a confluent and 2-homogeneous semi-Thue system over $\{\triangleright, \triangleleft\}$ by $\mathcal{Z} = \{(\triangleright\triangleright, \varepsilon), (\triangleleft\triangleleft, \varepsilon)\}$. Finally let $[\varepsilon]_{\mathcal{Z}} = \{s \in \{\triangleright, \triangleleft\}^* \mid s \overset{*}{\to}_{\mathcal{Z}} \varepsilon\}$. Note that every word in $[\varepsilon]_{\mathcal{Z}}$ has even length.

Example 3.2.17. *Let $\Gamma = \{a, b, c, d, e\}$ and $\mathrm{dom}(R) = \{ab, ba, bc, cd, ed\}$. Then $D_R(abcde) = \{d \triangleright \triangleright \triangleleft \mid d \in \{\triangleright, \triangleleft\}\}$.*

Lemma 3.2.18. *Let $a, b \in \Gamma$. Then $a \overset{*}{\leftrightarrow}_R b$ if and only if there exists an R-path w from a to b with $D_R(w) \cap [\varepsilon]_{\mathcal{Z}} \neq \emptyset$.*

Proof. First assume that $w = a_1 \cdots a_n$ is an R-path such that $a_1 = a$, $a_n = b$, and $d_1 \cdots d_{n-1} \in D_R(w) \cap [\varepsilon]_{\mathcal{Z}}$ (thus n is odd). The case $n = 1$ is clear. Thus, assume that $n \geq 3$, $s = d_1 \cdots d_{i-1} d_{i+2} \cdots d_{n-1} \in [\varepsilon]_{\mathcal{Z}}$, and $d_i = \triangleright = d_{i+1}$ (the case $d_i = \triangleleft = d_{i+1}$ is analogous). Hence, $(a_i a_{i+1}, \varepsilon), (a_{i+1} a_{i+2}, \varepsilon) \in R$ and $a_i \ {}_R\!\leftarrow a_i a_{i+1} a_{i+2} \to_R a_{i+2}$. Define the homomorphism h by $h(a_{i+2}) = a_i$ and $h(c) = c$ for all $c \in \Gamma \setminus \{a_{i+2}\}$. Then $w' = h(a_1) \cdots h(a_i) h(a_{i+3}) \cdots h(a_n)$ is an $h(R)$-path such that $s \in D_{h(R)}(w')$. Inductively we obtain $h(a) \overset{*}{\leftrightarrow}_{h(R)} h(b)$. Finally Lemma 3.2.16 implies $a \overset{*}{\leftrightarrow}_R b$.

Now assume that $a \overset{*}{\leftrightarrow}_R b$ and choose a derivation $a = u_1 \leftrightarrow_R u_2 \leftrightarrow_R \cdots u_{n-1} \leftrightarrow_R u_n = b$, where n is minimal. The case $a = b$ is clear, thus assume

that $a \neq b$ and hence $n \geq 3$. First we will apply the following transformation step to our chosen derivation: If the derivation contains a subderivation of the form $uv\ell_2w \;{}_R\!\leftarrow u\ell_1v\ell_2w \rightarrow_R u\ell_1vw$, where $(\ell_1, \varepsilon), (\ell_2, \varepsilon) \in R$, then we replace this subderivation by $uv\ell_2w \rightarrow_R uvw \;{}_R\!\leftarrow u\ell_1vw$. Similarly we proceed with subderivations of the from $u\ell_2vw \;{}_R\!\leftarrow u\ell_2v\ell_1w \rightarrow_R uv\ell_1w$. Since the iterated application of this transformation step is a terminating process, we finally obtain a derivation $\mathcal{D}$ from a to b, which does not allow further applications of the transformation described above. We proceed with the derivation $\mathcal{D}$. Note that $\mathcal{D}$ is also a derivation of minimal length from a to b. Since a and b belong to $\mathrm{IRR}(R)$, $\mathcal{D}$ must be of the form $a \overset{*}{\leftrightarrow}_R u \;{}_R\!\leftarrow v \rightarrow_R w \overset{*}{\leftrightarrow}_R b$ for some u, v, and w. The assumptions on $\mathcal{D}$ imply that there exist $s, t \in \Gamma^*$ and $(a_1a_2, \varepsilon), (a_2a_3, \varepsilon) \in R$ such that $u = sa_1t$, $v = sa_1a_2a_3t$, and $w = sa_3t$ (or $u = sa_3t$, $v = sa_1a_2a_3t$, and $w = sa_1t$, this case is analogous). Thus, $a_1 \overset{*}{\leftrightarrow}_R a_3$. Define the homomorphism h by $h(a_3) = a_1$ and $h(c) = c$ for all $c \in \Gamma \backslash \{a_3\}$. By applying h to our derivation $\mathcal{D}$ and removing the part $h(u) \;{}_{h(R)}\!\leftarrow h(v) \rightarrow_{h(R)} h(w) = h(u)$, we obtain $h(a) \overset{*}{\leftrightarrow}_{h(R)} h(b)$ by a derivation, which is shorter than $\mathcal{D}$. Inductively we can conclude that there exists an $h(R)$-path w' from $h(a)$ to $h(b)$ with $D_{h(R)}(w') \cap [\varepsilon]_{\mathcal{Z}} \neq \emptyset$. We inductively transform w' into an R-path w from a to b with $D_R(w) \cap [\varepsilon]_{\mathcal{Z}} \neq \emptyset$: We start with w'. Assume that we have already obtained a word $ucdv$ such that uc is an R-path, $(cd, \varepsilon) \in h(R)$ (the case $(dc, \varepsilon) \in h(R)$ is analogous), but $(cd, \varepsilon) \notin R$. Then there are three cases:

- $c = a_1$ and $(a_3d, \varepsilon) \in R$: we continue with the word $ua_1a_2a_3dv$.
- $d = a_1$ and $(ca_3, \varepsilon) \in R$: we continue with the word $uca_3a_2a_1v$.
- $c = a_1 = d$ and $(a_3a_3, \varepsilon) \in R$: we continue with $ua_1a_2a_3a_3a_2a_1v = uca_2a_3a_3a_2dv$.

This process results in an R-path w from a to b. Moreover, $D_R(w) \cap [\varepsilon]_{\mathcal{Z}} \neq \emptyset$, because the above transformation step replaces on the level of $\{\triangleright, \triangleleft\}$-words a single occurrence of $\triangleright$ in w' (resulting from $(cd, \varepsilon) \in h(R)$) by one of the sequences $\triangleright\triangleright\triangleright$, $\triangleright\triangleleft\triangleleft$, or $\triangleright\triangleright\triangleright\triangleleft\triangleleft$. □

Example 3.2.19. *In Example 3.2.17 we have* $\triangleleft\triangleright\triangleright\triangleleft \in D_R(abcde) \cap [\varepsilon]_{\mathcal{Z}}$. *Indeed, we have* $a \;{}_R\!\leftarrow eda \;{}_R\!\leftarrow ebcda \rightarrow_R eba \rightarrow_R e$.

Define the set $\mathcal{J}$ by $\mathcal{J} = ([\varepsilon]_{\mathcal{Z}} \backslash \{\varepsilon\}) \setminus ([\varepsilon]_{\mathcal{Z}} \backslash \{\varepsilon\})([\varepsilon]_{\mathcal{Z}} \backslash \{\varepsilon\})$, thus $\mathcal{J}$ is the minimal generating set for the submonoid $[\varepsilon]_{\mathcal{Z}}$ of $\{\triangleright, \triangleleft\}^*$. The following lemma follows immediately from the definition of $\mathcal{J}$.

Lemma 3.2.20. *We have* $\mathcal{J} \subseteq (\triangleright\{\triangleright, \triangleleft\}^*\triangleright) \cup (\triangleleft\{\triangleright, \triangleleft\}^*\triangleleft)$ *and* $[\varepsilon]_{\mathcal{Z}} = \mathcal{J}^*$.

Define a binary relation $\mathcal{T} \subseteq \Gamma \times \Gamma$ by $(a, b) \in \mathcal{T}$ if and only if there exists an R-path $w = a \cdots b$ from a to b with $|w|$ odd and furthermore there exist $c, d \in \Gamma$ such that either $(ac, \varepsilon), (db, \varepsilon) \in R$ or $(ca, \varepsilon), (bd, \varepsilon) \in R$. Note that $\mathcal{T}$ is symmetric.

Lemma 3.2.21. *Let* $a, b \in \Gamma$. *Then* $a \overset{*}{\leftrightarrow}_R b$ *if and only if* $(a, b) \in \overset{*}{\mathcal{T}}$.

Proof. For the if-direction it suffices to show that $a \overset{*}{\leftrightarrow}_R b$ if $(a, b) \in \mathcal{T}$. Thus, assume that there exists an R-path $w = a \cdots b$ from a to b with $|w|$ odd and furthermore there exist $c, d \in \Gamma$ such that $(ac, \varepsilon), (db, \varepsilon) \in R$ (the case that $(ca, \varepsilon), (bd, \varepsilon) \in R$ is analogous). Let $s \in D_R(w)$, thus $|s|$ is even. Since $(ac, \varepsilon), (db, \varepsilon) \in R$, also the word $w_i = (ac)^{|s|} w (db)^i$ is an R-path for every $i \geq 0$. We have $s_i = (\triangleright\triangleleft)^{|s|} s (\triangleleft\triangleright)^i \in D_R(w_i)$. Since $|s|$ is even and $|s| < |(\triangleright\triangleleft)^{|s|}|$, the unique normalform $\mathrm{NF}_{\mathcal{Z}}((\triangleright\triangleleft)^{|s|} s)$ of the prefix $(\triangleright\triangleleft)^{|s|} s \preceq s_i$ has the form $(\triangleright\triangleleft)^k$ for some $k \geq 0$. Thus, $s_k \in [\varepsilon]_{\mathcal{Z}}$ and $D_R(w_k) \cap [\varepsilon]_{\mathcal{Z}} \neq \emptyset$. By Lemma 3.2.18 we have $a \overset{*}{\leftrightarrow}_R b$.

Now let $a \overset{*}{\leftrightarrow}_R b$. By Lemma 3.2.18 there exists an R-path w from a to b and a word $s \in D_R(w) \cap [\varepsilon]_{\mathcal{Z}}$. Let $s = s_1 \cdots s_m$ where $m \geq 0$ and $s_i \in \mathcal{J}$. We can factorize w as $w = w_1 \cdots w_m$, where w_i is an R-path from a_i to a_{i+1} such that $s_i \in D_R(w_i)$ and $a_1 = a$, $a_{m+1} = b$. It suffices to show that $(a_i, a_{i+1}) \in \mathcal{T}$. Since $s_i \in \mathcal{J} \subseteq [\varepsilon]_{\mathcal{Z}}$, the length of s_i is even. Thus, $|w_i|$ is odd. Next $s_i \in (\triangleright\{\triangleright, \triangleleft\}^*\triangleright) \cup (\triangleleft\{\triangleright, \triangleleft\}^*\triangleleft)$ by Lemma 3.2.20. Let $s_i \in \triangleright\{\triangleright, \triangleleft\}^*\triangleright$, the other case is symmetric. Hence there exist rules $(a_i c, \varepsilon), (d a_{i+1}, \varepsilon) \in R$. Thus, indeed $(a_i, a_{i+1}) \in \mathcal{T}$. □

The preceding lemma is the key for proving that the uniform word problem for 2-homogeneous presentations is in SL. In general it is quite difficult to prove that a problem is contained in SL. A useful strategy, developed in [103] and applied in [197, 198] in a context similar to our situation, is based on a logical characterization of SL. In the following we consider finite structures of the form $\mathcal{A} = (\{0, \ldots, n-1\}, 0, \max, s, \mathcal{R})$, where $\max = n-1$, $s = \{(i, i+1) \mid 0 \leq i < n-1\}$, and $\mathcal{R}$ is an additional binary relations on $\{0, \ldots, n-1\}$. The logic FO+posSTC is the extension of first-order logic over the signature of $\mathcal{A}$ by the STC-operator, which has the following syntax: If $\phi(x, y)$ is a formula of FO+posSTC with two free variables, then $[\mathrm{STC}x, y\ \varphi(x, y)]$ is again a formula of FO+posSTC with two free variables, with the restriction that the STC-operator is not allowed to occur within the scope of a negation. The

semantics of STC is the following: Let $\varphi(x,y)$ be a formula of FO+posSTC with two free variables x and y, and let $\mathcal{A} = (\{0,\ldots,n-1\}, 0, \max, s, \mathcal{R})$ be a structure. Assume that $\varphi(x,y)$ describes the binary relation $\mathcal{S}$ over $\{0,\ldots,n-1\}$, i.e., $\mathcal{A} \models \varphi(i,j)$ if and only if $(i,j) \in \mathcal{S}$ for all $0 \leq i,j < n$. Then $\mathcal{A} \models [\mathrm{STC}x,y\ \varphi(x,y)](i,j)$ if and only if (i,j) belongs to the symmetric, transitive, and reflexive closure of $\mathcal{S}$, i.e., $(i,j) \in (\mathcal{S} \cup \mathcal{S}^{-1})^*$. In [103] it was shown that for every fixed sentence φ of FO+posSTC the following problem belongs to SL:

INPUT: A binary coded structure $\mathcal{A} = (\{0,\ldots,n-1\}, 0, \max, s, \mathcal{R})$
QUESTION: Does $\mathcal{A} \models \varphi$ hold?

Theorem 3.2.22. *The following problem is SL-complete:*
INPUT: A 2-homogeneous presentation (Γ, R) *and* $a, b \in \Gamma$.
QUESTION: Does $a \overset{*}{\leftrightarrow}_R b$ *hold?*

Proof. First we show containment in SL. Let (Γ, R) be a 2-homogeneous presentation and let $a, b \in \Gamma$. W.l.o.g. we may assume that $\Gamma = \{0,\ldots,n-1\}$ and $a = 0, b = n-1$. If the latter assumption does not hold, then it can be enforced by relabeling the alphabet symbols. This relabeling can be done in deterministic logspace and we can use the fact that $\mathrm{L}^{\mathrm{SL}} = \mathrm{SL}$. We identify the input (Γ, R, a, b) with the structure $\mathcal{A} = (\Gamma, 0, \max, s, \mathcal{R})$, where $\mathcal{R} = \{(c,d) \mid (cd,\varepsilon) \in R\}$. Now define formulas $\mathcal{S}(x,y)$ and $\mathcal{T}(x,y)$ as follows:

$$\mathcal{S}(x,y) \equiv \exists z\{(\mathcal{R}(x,z) \vee \mathcal{R}(z,x)) \wedge (\mathcal{R}(y,z) \vee \mathcal{R}(z,y))\}$$

$$\mathcal{T}(x,y) \equiv [\mathrm{STC}u,v\ \mathcal{S}(u,v)](x,y) \wedge \exists x', y' \left\{ \begin{array}{l} (\mathcal{R}(x,x') \wedge \mathcal{R}(y',y)) \vee \\ (\mathcal{R}(x',x) \wedge \mathcal{R}(y,y')) \end{array} \right\}$$

By Lemma 3.2.21, $a \overset{*}{\leftrightarrow}_R b$ if and only if $\mathcal{A} \models [\mathrm{STC}u,v\ \mathcal{T}(u,v)](0,\max)$. Thus, containment in SL follows from [103]. In order to show SL-hardness we use the SL-complete *undirected graph accessibility problem (UGAP)*, see also [119]:

INPUT: An undirected graph $G = (V,E)$ and two nodes $a, b \in V$.
QUESTION: Does there exist a path in G from a to b?

Let (G,a,b) be an instance of UGAP, where $G = (V,E)$ is an undirected graph and $a, b \in V$ are nodes. Of course we may assume that $V \cap E = \emptyset$. We define a 2-homogeneous presentation $(V \cup E, R)$, where

$$R = \{(ce,\varepsilon),\ (ec,\varepsilon) \mid c \in V, e \in E, \text{node } c \text{ is adjacent with edge } e\}.$$

We claim that there exists a path in G from a to b if and only if $a \overset{*}{\leftrightarrow}_R b$. First assume that there exists a path $a = a_1, a_2, \ldots, a_n = b$ with $(a_i, a_{i+1}) = e_i \in E$. The case $n = 1$ is clear. If $n > 1$, then by induction $a \overset{*}{\leftrightarrow}_R a_{n-1}$. Thus $a \overset{*}{\leftrightarrow}_R a_{n-1}\ {}_R\!\leftarrow a_{n-1}e_{n-1}a_n \rightarrow_R a_n = b$. Conversely assume that a and b belong to different connected components of G. Let V_a and E_a be the set of all nodes and edges, respectively, that belong to the connected component of a. Define a projection $\pi : V \cup E \to V_a \cup E_a$ by $\pi(x) = \varepsilon$ if $x \notin V_a \cup E_a$ and $\pi(x) = x$ if $x \in V_a \cup E_a$. If $a \overset{*}{\leftrightarrow}_R b$, then $a = \pi(a) \overset{*}{\leftrightarrow}_{\pi(R)} \pi(b) = \varepsilon$. But this is impossible, since $\pi(R)$ is 2-homogeneous and thus $u \overset{*}{\leftrightarrow}_{\pi(R)} \varepsilon$ implies that $|u|$ is even. □

In Theorem 3.2.22 we restrict the words that are checked for equality to single symbols. In order to deduce that the uniform word problem for 2-homogeneous presentations is in SL, we need only one more lemma:

Lemma 3.2.23. *Let (Γ, R) be a 2-homogeneous presentation, where w.l.o.g. $\Gamma = \{0, \ldots, n-1\}$. Let $h : \Gamma^* \to \Gamma^*$ be the homomorphism defined by $h(a) = \min\{b \in \Gamma \mid a \overset{*}{\leftrightarrow}_R b\}$. Then for all $u, v \in \Gamma^*$ we have $u \overset{*}{\leftrightarrow}_R v$ if and only if $h(u) \overset{*}{\leftrightarrow}_{h(R)} h(v)$. Furthermore the 2-homogeneous presentation $(h(\Gamma), h(R))$ is confluent.*

Proof. The first statement of the lemma follows from Lemma 3.2.16. For confluence note that all critical pairs of $h(R)$ are trivial: if

$$h(a)\ {}_{h(R)}\!\leftarrow h(a)h(b)h(c) \rightarrow_{h(R)} h(c),$$

then $a \overset{*}{\leftrightarrow}_R b$ and thus $h(a) = h(b)$. □

Theorem 3.2.24. *The uniform word problem for 2-homogeneous presentations is SL-complete.*

Proof. By Theorem 3.2.22 it remains to show containment in SL. W.l.o.g. we may assume that $\Gamma = \{0, \ldots, n-1\}$. Let h be the homomorphism from Lemma 3.2.23. We check whether $u \overset{*}{\leftrightarrow}_R v$ by essentially running the logspace algorithm for the uniform word problem for confluent and 2-homogeneous presentations from Section 3.2.3, but each time we read from the input-tape (the binary coding of) a symbol $a \in \Gamma$, we replace a by $h(a)$. Since $h(a) = \min\{b \in \Gamma \mid a \overset{*}{\leftrightarrow}_R b\}$, Theorem 3.2.22 implies that we can determine $h(a)$ by at most n queries to an SL-oracle. Since $\mathrm{L}^{\mathrm{SL}} = \mathrm{SL}$, the theorem follows. □

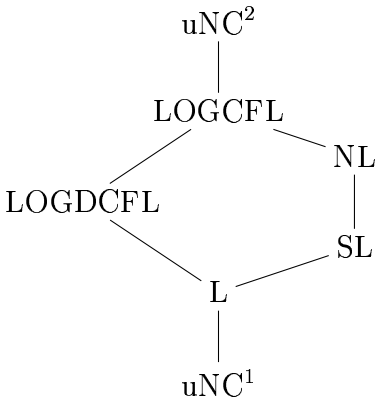


Figure 3.1: Complexity classes between uNC^1 and uNC^2

3.3 Length-reducing presentations

Book has shown in [23] that the word problem for every fixed finite, confluent, and length-reducing presentation can be decided in linear time. We will start this section with an investigation of the parallel complexity of this problem. For this we have to introduce two further complexity classes below uNC.

LOGCFL (resp. LOGDCFL) is the class of all problems that can be reduced in logarithmic space to a context-free language (resp. deterministic context-free language) [202]. An alternative characterization of these classes was given by Sudborough [202]: An AuxPDA is a pushdown automaton that has an auxiliary read-write tape. Let us denote with $\mathrm{AuxPDA}(s(n), t(n))$ the set of all languages that can be recognized with a nondeterministic AuxPDA with auxiliary space $O(s(n))$ in time $O(t(n))$. With $\mathrm{DAuxPDA}(s(n), t(n))$ we denote the corresponding deterministic class, see [49] for more details. Then L belongs to LOGCFL (resp. LOGDCFL) if and only if it belongs to $\mathrm{AuxPDA}(O(\log(n)), n^{O(1)})$ (resp. $\mathrm{DAuxPDA}(O(\log(n)), n^{O(1)})$) [202]. Further characterizations of these complexity classes can be found in [175, 213] (for LOGCFL) and [77] (for LOGDCFL).

It is known that LOGCFL is contained in uNC^2 [176], thus problems in LOGCFL admit efficient parallel algorithms. Figure 3.1 shows all known inclusions between the classes in the range from uNC^1 to uNC^2 that we have introduced so far.

A language $L \subseteq \Gamma^*$ is called *growing context-sensitive* if it can be gener-

ated by a context-sensitive grammar such that every production is strictly length-increasing. By [58] every fixed growing context-sensitive language is contained in LOGCFL.[4] Using this result, we can easily prove the following statement:

Theorem 3.3.1. *Let (Γ, R) be a fixed length-reducing and confluent presentation. Then the word problem for $\mathcal{M}(\Gamma, R)$ is in LOGCFL.*

Proof. Assume that (Γ, R) is length-reducing and confluent and let $u, v \in \Gamma^*$. Let $\overline{\Gamma} = \{\overline{a} \mid a \in \Gamma\}$ be a disjoint copy of Γ. For a word $s = a_1 \cdots a_n$, $a_i \in \Gamma$, let $\overline{s}^{\mathrm{rev}} = \overline{a}_n \cdots \overline{a}_1 \in \overline{\Gamma}^*$. Define the length-reducing semi-Thue system S by

$$S = R \cup \{\overline{s}^{\mathrm{rev}} \to \overline{t}^{\mathrm{rev}} \mid (s,t) \in R\} \cup \{a\overline{a} \to \varepsilon \mid a \in \Gamma\}.$$

Since R is confluent, we have $u \overset{*}{\leftrightarrow}_R v$ if and only if $u\overline{v}^{\mathrm{rev}} \overset{*}{\to}_S \varepsilon$. Moreover, the language $\{w \in (\Gamma \cup \overline{\Gamma})^* \mid w \overset{*}{\to}_S \varepsilon\}$ is easily seen to be growing context-sensitive. Since $u\overline{v}^{\mathrm{rev}}$ can be constructed in logarithmic space from u and v, the theorem follows from [58]. □

Another application of the inclusion of growing context-sensitive languages in LOGCFL concerns *hyperbolic groups*, see [52, 88]. Hyperbolic groups are usually defined via a certain hyperbolicity condition on the Cayley-graph. In [52] it was shown that a finitely generated group $\mathcal{G}$ is hyperbolic if and only if it has a finite presentation of the form (Γ, R), where R is length-reducing and confluent on the equivalence class of ε. The latter means that for all $s \in \Gamma^*$, $s \overset{*}{\leftrightarrow}_R \varepsilon$ if and only if $s \overset{*}{\to}_R \varepsilon$. Since R is length-reducing, the language $\{s \in \Gamma^* \mid s \overset{*}{\to}_R \varepsilon\}$ is growing context-sensitive. Hence we obtain the following result, which improves the uNC^2-upper bound from [38].

Theorem 3.3.2. *The word problem for every fixed hyperbolic group is in LOGCFL.*

It is not known whether the LOGCFL-upper bound in Theorem 3.3.1 is sharp in the sense that there exists a fixed length-reducing and confluent presentation with a LOGCFL-complete word problem. On the other hand, since there exits a fixed deterministic context-free language with a LOGDCFL-complete

[4] The result of [58] has been slightly improved in [35], where it was shown that every growing context-sensitive language is in AuxPDA$(O(\log(n)), n^{O(1)})$, where the read-only input tape of the AuxPDA is restricted to be one-way. In contrast to this, the uniform membership problem for growing context-sensitive grammars is NP-complete [33, 47, 115].

membership problem [202], Theorem 2.2 of [140] implies that there exists a fixed length-reducing and confluent presentation with a LOGDCFL-hard word problem. If we assume a further restriction on the class of presentations, we obtain a characterization of LOGDCFL. A presentation (Γ, R) is called *left-basic* if it satisfies the following two conditions:

- if $\ell \in \mathrm{dom}(R)$, $r \in \mathrm{ran}(R)$, and $r = u\ell v$, then $u = v = \varepsilon$.
- if $\ell \in \mathrm{dom}(R)$, $r \in \mathrm{ran}(R)$, $ur = \ell v$, and $|\ell| > |u|$, then $v = \varepsilon$.

The first condition means that a right-hand side does not strictly contain a left-hand side. The second condition means that the following kind of overlapping is not allowed:

u	$r \in \mathrm{ran}(R)$
$\ell \in \mathrm{dom}(R)$	$v \neq \varepsilon$

Note that every monadic presentation is left-basic.

Let us define the suffix-rewrite relation $\twoheadrightarrow_R$ by $s \twoheadrightarrow_R t$ if and only if $s = u\ell$ and $t = ur$ for some $u \in \Gamma^*$ and $(\ell, r) \in R$. The following simple fact is well-known.

Lemma 3.3.3. *If (Γ, R) is left-basic, then for every $u \in \mathrm{IRR}(R)$ and $a \in \Gamma$ we have $ua \xrightarrow{*}_R v$ if and only if $ua \overset{*}{\twoheadrightarrow}_R v$.*

Left-basic presentations have a close relationship to deterministic context-free languages, see [188] for an overview. The following theorem expresses the complexity theoretical side of this relationship.

Theorem 3.3.4. *The uniform word problem for terminating, confluent, and left-basic presentations is LOGDCFL-complete. Furthermore, there exists a fixed length-reducing, confluent, and left-basic presentation (Γ, R) such that the word problem for $\mathcal{M}(\Gamma, R)$ is LOGDCFL-complete.*

Proof. First we prove that the uniform word problem for the class of all terminating, confluent, and left-basic presentations is in LOGDCFL by presenting an algorithm that works on a deterministic AuxPDA in logarithmic space and polynomial time.

Thus, let us assume that the input consists of a triple $((\Gamma, R), s, t)$, where (Γ, R) is terminating, confluent, and left-basic, and $s, t \in \Gamma^*$. Let n be the

length of the binary coding of this input. Our algorithm will check whether $\mathrm{NF}_R(s) = \mathrm{NF}_R(t)$. For this, we will first show how to calculate $\mathrm{NF}_R(s)$ on a deterministic AuxPDA in logarithmic space and polynomial time. The basic idea of how to do this appeared many times in the literature, see e.g. [23]. The only slight complication in our situation is the desired uniformity in the presentation (Γ, R), for which we need the $O(\log(n))$ space bounded auxiliary store. The correctness of the following procedure follows from Lemma 3.3.3.

Our algorithm for computing $\mathrm{NF}_R(s)$ works in stages. At the beginning of a stage the pushdown contains a word from $\mathrm{IRR}(R)$ and the auxiliary store contains a pointer to a position i in the input word s. Note that a symbol $a \in \Gamma$ can be represented as a bit string of length $O(\log(n))$, thus the pushdown content is a sequence of blocks of length $O(\log(n))$, where every block represents a symbol from Γ. The stage begins by pushing the i-th symbol of s onto the pushdown (which is a bit string of length $O(\log(n))$) and incrementing the pointer to position $i+1$ in s. Now we have to check whether the pushdown content is of the form $\Gamma^* \,\mathrm{dom}(R)$. For this we have to scan every left-hand side of R using a second pointer to the input. Every $\ell \in \mathrm{dom}(R)$ is scanned in reverse order and thereby compared with the top of the push-down. During this phase, symbols are popped from the pushdown. If it turns out that the left-hand side that is currently scanned is not a suffix of the pushdown content, then these symbols must be "repopped". But this can be done, since the suffix of the pushdown content that was pushed so far is a suffix of the currently scanned left-hand side $\ell \in \mathrm{dom}(R)$, which is still available on the read-only input tape. If a left-hand side ℓ is found on top of the pushdown, then the corresponding right-hand side is pushed on the pushdown and we try to find again a left-hand side on top of the pushdown. If finally no left-hand side matches a suffix of the pushdown content, then we know that the pushdown content belongs to $\mathrm{IRR}(R)$ and we can proceed with the next stage. Finally, if the first pointer has reached the end of the input word s (or more precisely points to the first position following s), then the pushdown content is equal to $\mathrm{NF}_R(s)$. We will need also the following statement:

Claim: In the above procedure, after the i-th stage the pushdown has length $i \cdot (\alpha + 1)$, where $\alpha = \max\{|r| \mid r \in \mathrm{ran}(R)\}$. Moreover, every stage needs only polynomial time.

The first statement can be shown by induction on i. Since R is left-basic, it follows that if w is the pushdown content at the end of the $(i-1)$-th stage,

then the pushdown content at the end of the i-th stage either belongs to $w\Gamma$ or is of the form ur for some $r \in \mathrm{ran}(R)$ and some prefix u of w. Moreover, the i-th stage simulates at most $|w| \cdot |R|$ rewrite steps of R.

In order to check whether $s \overset{*}{\leftrightarrow}_R t$, we have to solve one more problem: If we would calculate $\mathrm{NF}_R(t)$ in the same way as above, then the pushdown would finally contain the word $\mathrm{NF}_R(s)\mathrm{NF}_R(t)$. But now there seems to be no way of checking, whether $\mathrm{NF}_R(s) = \mathrm{NF}_R(t)$. Thus, we have to apply another strategy. Note that for a fixed binary coded number $1 \leq i \leq |s|$, it is easy to modify our algorithm for calculating $\mathrm{NF}_R(s)$ such that some specified auxiliary storage cell S contains always the i-th symbol of the pushdown content (or some special symbol if the pushdown is shorter than i). For this we have to store the length of the pushdown, for which we need only space $O(\log(n))$. Moreover, also S only needs space $O(\log(n))$. Thus, at the end of our modified algorithm for computing $\mathrm{NF}_R(s)$, S contains the symbol $\mathrm{NF}_R(s)[i]$ (or some special symbol in case $|\mathrm{NF}_R(s)| < i$). Next, we flush the pushdown and repeat the same procedure with the other input word t and the same i, using another storage cell T. In this way we can check, whether $\mathrm{NF}_R(s)[i] = \mathrm{NF}_R(t)[i]$. Finally, we repeat this step for every $1 \leq i \leq \max\{|s| \cdot (\alpha + 1), |t| \cdot (\alpha + 1)\}$. The latter bound is the maximal pushdown-length that may occur, which follows from the above claim. Note that also i needs only space $O(\log(n))$. This concludes the description of our LOGDCFL-algorithm.

It remains to construct a fixed length-reducing, confluent, and left-basic presentation (Γ, R) such that the corresponding word problem is LOGDCFL-hard. In [202], Sudborough has shown that there exists a fixed deterministic context-free language $L \subseteq \Sigma^*$ with a LOGDCFL-complete membership problem. Let $\mathcal{A} = (Q, \Delta, \Sigma, \delta, q_0, \bot)$ be a deterministic pushdown automaton with $L = L(\mathcal{A})$, where Q is the set of states, $q_0 \in Q$ is the initial state, Δ is the pushdown alphabet, $\bot \in \Delta$ is the bottom symbol, and $\delta : \Delta \times Q \times \Sigma \to \Delta^* \times Q$ is the transition function. By [202, Lem. 7] we may assume that $\mathcal{A}$ makes no ε-moves and that $\mathcal{A}$ accepts L by empty store in state q_0. Let $m = \max\{|\gamma| \mid \delta(A, q, a) = (\gamma, p),\ q, p \in Q, A \in \Delta, a \in \Sigma\}$, thus m is the maximal length of a sequence that is pushed on the pushdown in one step. Let $\# \notin \Delta \cup Q \cup \Sigma$ be an additional symbol and let $\Gamma = \Delta \cup Q \cup \Sigma \cup \{\#\}$. Define the semi-Thue system R by

$$R = \{Aqa^m\# \to \gamma p \mid \delta(A, q, a) = (\gamma, p)\}.$$

It is easy to see that (Γ, R) is length-reducing, confluent, and left-basic.

Moreover, if $h : \Sigma^* \to (\Sigma \cup \{\#\})^*$ denotes the homomorphism defined by $h(a) = a^m\#$ (which can be computed in logarithmic space), then $w \in L$ if and only if $\bot q_0 h(w) \stackrel{*}{\to}_R q_0$ if and only if $\bot q_0 h(w) \stackrel{*}{\leftrightarrow}_R q_0$. □

In the rest of this section we will prove that the uniform word problem for length-reducing and confluent presentations is P-complete. Our lower bound proof as well as all other lower bound proofs in the rest of this chapter are based on generic reductions via Turing-machines. Thus, let us first fix our Turing-machine model. A *deterministic Turing-machine* is a tuple $\mathcal{T} = (Q, \Sigma, \delta, q_0, q_f)$, where Q is the finite set of states, $q_0 \in Q$ is the initial state, $q_f \in Q \setminus \{q_0\}$ is the unique final state, Σ is the finite tape alphabet containing the blank symbol $\Box$ ($\Sigma \cap Q = \emptyset$), and $\delta : (Q \setminus \{q_f\}) \times \Sigma \to Q \times \Sigma \times \{-1, +1\}$ is the transition function, where -1 ($+1$) means that the read-write head moves to the left (right). Note that $\mathcal{T}$ cannot perform any transition out of the final state q_f. We assume that $\mathcal{T}$ has a one-sided infinite tape, whose cells can be identified with the natural numbers $\mathbb{N} = \{1, 2, \ldots\}$. We also assume that in the first step $\mathcal{T}$ marks cell 1 so that $\mathcal{T}$ always knows when it reaches the left end of the input tape (then, $\mathcal{T}$ will move right in the next step). All these assumptions do not restrict the computational power of Turing-machines and will always be assumed in the following. An input for $\mathcal{T}$ is a word $w \in (\Sigma \setminus \{\Box\})^*$. A word of the form uqv, where $u, v \in \Sigma^*$ and $q \in Q$, encodes the configuration, where the machine is in state q, the read-write head is scanning cell $|u| + 1$, and cell i contains the symbol $(uv)[i]$ if $i \leq |uv|$, otherwise it contains the blank symbol $\Box$. We write $sqt \Rightarrow_{\mathcal{T}} upv$ if $\mathcal{T}$ can move in one step from the configuration sqt to the configuration upv, where $q, p \in Q$ and $st, uv \in \Sigma^*$. The language that is accepted by $\mathcal{T}$ is $L(\mathcal{T}) = \{w \in (\Sigma \setminus \{\Box\})^* \mid \exists u, v \in \Sigma^* : q_0 w \stackrel{+}{\Rightarrow}_{\mathcal{T}} uq_f v\}$. Note that $w \in L(\mathcal{T})$ if and only if $\mathcal{T}$ terminates on the input w.

Theorem 3.3.5. *Let Γ be a fixed alphabet with $|\Gamma| \geq 2$. Then the uniform word problem for the class of all confluent and length-reducing presentations of the form (Γ, R) is P-complete.*[5]

Proof. Membership in P is clear. For P-hardness, let us take a fixed deterministic Turing-machine $\mathcal{T} = (Q, \Sigma, \delta, q_0, q_f)$ with a P-complete acceptance problem (such a machine exists, take any machine that solves a P-complete

[5]This result already appeared in the PhD-theses of the author [123]. For completeness we will present a slightly simplified proof.

problem). Let $p(n)$ be a polynomial such that for every input w of length n, $\mathcal{T}$ finally reaches q_f (i.e., $w \in L(\mathcal{T})$) if and only if $\mathcal{T}$ reaches q_f after at most $p(n)$ many steps (and hence uses also space at most $p(n)$). For the following construction let us fix an input $w \in (\Sigma \backslash \{\Box\})^*$ for $\mathcal{T}$ and let $m = p(|w|)$. Let $\overline{\Sigma} = \{\overline{a} \mid a \in \Sigma\}$ be a disjoint copy of Σ and let $\# \notin Q \cup \Sigma \cup \overline{\Sigma}$ be an additional symbol. Define $\Gamma = Q \cup \Sigma \cup \overline{\Sigma} \cup \{\#\}$ and let R be the length-reducing semi-Thue system over Γ consisting of the following rules:

(1) $\#q^{3i}a \to \overline{b}\#p^{3(i-1)}$	if $\delta(q,a) = (p,b,+1)$, $1 \leq i \leq m+1$
(2) $\overline{c}\#q^{3i}a \to \#p^{3(i-1)}cb$	if $\delta(q,a) = (p,b,-1)$, $1 \leq i \leq m+1$, and $c \in \Sigma$
(3) $q_f x \to q_f$	for all $x \in \Gamma$
(4) $xq_f \to q_f$	for all $x \in \Gamma$

The rules in (1) and (2) simulate the machine $\mathcal{T}$. Here, the state q is represented by the word $\#q^{3i}$. The rules in (3) and (4) make q_f absorbing. Note that the state symbol is represented $3i$ times on the left–hand side and $3(i-1)$ times on the right–hand side. This makes R length–reducing. It is also easy to see that R can be computed from $|w|$ in logarithmic space, we only have to count up to m, for which we need space $\log(m) \in O(\log(|w|))$.

Claim: R is confluent. Furthermore $\mathcal{T}$ terminates on input w after $\leq m$ steps if and only if $\#q_0^{3(m+1)}w\Box^{m+1} \overset{*}{\leftrightarrow}_R q_f$.

For confluence note that only the rules in (3) and (4) generate nontrivial critical pairs. More precisely, in every nontrivial critical pair (s,t) of R, q_f occurs in both s and t. With the rules in (3) and (4), both s and t can be reduced to q_f. If $\mathcal{T}$ terminates on input w after $\leq m$ steps, then for some $j \geq 1$ and $u, v \in \Sigma^*$ it holds $\#q_0^{3(m+1)}w\Box^{m+1} \overset{*}{\to}_R \overline{u}\#q_f^{3j}v$. By applying the rules in (3) and (4), the word $\overline{u}\#q_f^{3j}v$ can be reduced to q_f. Now assume that $\mathcal{T}$ does not terminate on input w after $\leq m$ steps. By simulating $m+1$ steps of $\mathcal{T}$, we obtain $\#q_0^{3(m+1)}w\Box^{m+1} \overset{*}{\to}_R \overline{u}\#v \in \mathrm{IRR}(R)$ for some $u, v \in \Sigma^*$. But then $\#q_0^{3(m+1)}w\Box^{m+1} \overset{*}{\leftrightarrow}_R q_f$ cannot hold since also $q_f \in \mathrm{IRR}(R)$ and R is confluent. Thus, the claim is proved.

Now assume that $\Gamma = \{a_1, \ldots, a_k\}$ and consider the coding function defined by $h(a_i) = aba^{i+1}b^{k-i+2}$ for $1 \leq i \leq k$. The presentation $(\{a,b\}, h(R))$ is also length-reducing, confluent, and can be calculated from R (and hence from w) in logarithmic space. Moreover, $h(\#q_0^{3(m+1)}w\Box^{m+1}) \overset{*}{\leftrightarrow}_{h(R)} h(q_f)$ if and only if $\#q_0^{3(m+1)}w\Box^{m+1} \overset{*}{\leftrightarrow}_R q_f$ if and only if $w \in L(\mathcal{T})$. This proves the theorem. $\square$

3.4 Weight-reducing presentations

Weight-reducing presentations were investigated for instance in [61, 104, 148] and [33, 34] as a grammatical formalism. We will see that the word problem for a fixed weight-reducing and confluent presentation can be reduced to the word problem for a fixed length-reducing and confluent presentation. Thus, we obtain the following result:

Theorem 3.4.1. *Let (Γ, R) be a fixed weight-reducing and confluent presentation. Then the word problem for $\mathcal{M}(\Gamma, R)$ is in LOGCFL.*

Proof. Let (Γ, R) be a weight-reducing and confluent presentation. Let f be a weight-function such that $f(s) > f(t)$ for all $(s,t) \in R$. Let $\# \notin \Gamma$ and define the homomorphism $h : \Gamma^* \to (\Gamma \cup \{\#\})^*$ by $h(a) = \#^{f(a)-1}a$ for all $a \in \Gamma$ (note that $f(a) > 0$). Thus, $|h(a)| = f(a)$. Note that nontrivial overlappings between two words $h(a)$ and $h(b)$ are not possible. It follows that the presentation $(\Gamma \cup \{\#\}, h(R))$ is length-reducing and confluent. Moreover $u \overset{*}{\leftrightarrow}_R v$ if and only if $h(u) \overset{*}{\leftrightarrow}_{h(R)} h(v)$. Since $h(u)$ and $h(v)$ can be constructed in logarithmic space, the theorem follows from Theorem 3.3.1. □

Next we will consider the uniform word problem for weight-reducing and confluent presentations with a fixed alphabet. In order to get an upper bound for this problem we need the following lemma, which we state in a slightly more general form for later applications.

Lemma 3.4.2. *Let (Γ, R) be a finite presentation with $|\Gamma| = n$ and $\alpha = \max\{|s|_a \mid s \in \mathrm{dom}(R) \cup \mathrm{ran}(R), a \in \Gamma\}$. Let g be a weight-function with $g(s) \geq g(t)$ for all $(s,t) \in R$. Then there exists a weight-function f such that for all $(s,t) \in R$ and $a \in \Gamma$ we have:*

- *If $g(s) > g(t)$, then $f(s) > f(t)$.*
- *If $g(s) = g(t)$, then $f(s) = f(t)$.*
- *$f(a) \leq (n+1)(\alpha n)^n$.*

Proof. We use the following result about solutions of integer (in)equalities from [216]: Let A, B, C, D be $(m \times n)$-, $(m \times 1)$-, $(p \times n)$-, $(p \times 1)$-matrices, respectively, with integer entries. Let $r = \mathrm{rank}(A)$, $s = \mathrm{rank}\begin{pmatrix} A \\ C \end{pmatrix}$. Let M be an upper bound on the absolute values of all $(s-1) \times (s-1)$- and

$(s \times s)$-subdeterminants of the $(m+p) \times (n+1)$-matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$, which are formed with at least r rows from the matrix $(A \;\; B)$. Then the system $Ax = B \;\wedge\; Cx \geq D$ has an integer solution if and only if it has an integer solution x such that the absolute value of every entry of x is at most $(n+1)M$ [216].

Now let (Γ, R), n, α, and g be as specified in the lemma. Let $\Gamma = \{a_1, \ldots, a_n\}$ and $R = \{(s_i, t_i) \mid 1 \leq i \leq k\} \cup \{(u_i, v_i) \mid 1 \leq i \leq \ell\}$, where $g(s_i) = g(t_i)$ for $1 \leq i \leq k$ and $g(u_i) > g(v_i)$ for $1 \leq i \leq \ell$. Define the $(k \times n)$-matrix A by $A_{i,j} = |s_i|_{a_j} - |t_i|_{a_j}$ and define the $(\ell \times n)$-matrix C' by $C'_{i,j} = |u_i|_{a_j} - |v_i|_{a_j}$. Let $C = \begin{pmatrix} C' \\ \mathrm{Id}_n \end{pmatrix}$, where Id_n is the $(n \times n)$-identity matrix. Finally let $(j)^i$ be the i-dimensional column vector with all entries equal to j. Then the n-dimensional column vector $x = (x_j)_{1 \leq j \leq n}$ with $x_j = g(a_j)$ is a solution of the following system:

$$Ax = (0)^k \quad \wedge \quad Cx \geq (1)^{\ell+n} \tag{3.2}$$

Note that $r = \mathrm{rank}(A) \leq n$ and $s = \mathrm{rank} \begin{pmatrix} A \\ C \end{pmatrix} \leq n$. Moreover, all entries of the matrix $E = \begin{pmatrix} A & (0)^k \\ C & (1)^{\ell+n} \end{pmatrix}$ are bounded by α. Thus, the absolute value of every $(s-1) \times (s-1)$- or $(s \times s)$-subdeterminant of E is bounded by $s! \cdot \alpha^s \leq (\alpha n)^n$. By the result of [216], the system (3.2) has a solution $y = (y_j)_{1 \leq j \leq n}$ with $y_j \leq (n+1)(\alpha n)^n$ for all $1 \leq j \leq n$. If we define the weight-function f by $f(a_j) = y_j$, then f has all properties from the lemma. □

Theorem 3.4.3. *Let Γ be a fixed alphabet with $|\Gamma| \geq 2$. Then the uniform word problem for the class of all weight-reducing and confluent presentations of the form (Γ, R) is P-complete.*

Proof. Let (Γ, R) be a weight-reducing and confluent presentation with $|\Gamma| = n \geq 2$. Let $u, v \in \Gamma^*$. By Lemma 3.4.2 there exists a weight-function f such that $f(s) > f(t)$ for all $(s,t) \in R$ and $f(a) \leq (n+1)(\alpha n)^n$ for all $a \in \Gamma$. Thus, every derivation that starts from a word u has length at most $|u| \cdot (n+1) \cdot (\alpha n)^n$, which is polynomial in the input length $\|(\Gamma, R)\| + |uv|$ in case n is constant. Thus, $\mathrm{NF}_R(u)$ can be calculated in polynomial time and similarly for v. This proves the upper bound. P-hardness follows from Theorem 3.3.5. □

Finally for the class of all weight-reducing and confluent presentations the complexity of the uniform word problem increases to EXPTIME:

Theorem 3.4.4. *The uniform word problem for the class of all weight-reducing and confluent presentations is EXPTIME-complete.*

Proof. For the EXPTIME-upper bound we can use the arguments from the previous proof. Just note that this time the upper bound of $(n+1)(\alpha n)^n$ for the weight of a symbol is exponential in the length of the input. For the lower bound let $\mathcal{T} = (Q, \Sigma, \delta, q_0, q_f)$ be a deterministic Turing-machine such that for some polynomial p we have: If $w \in L(\mathcal{T})$, then $\mathcal{T}$, started on w, reaches the final state q_f after at most $2^{p(|w|)}$ many steps. Let $w \in (\Sigma \backslash \{\Box\})^*$ be an arbitrary input for $\mathcal{T}$. Let $m = p(|w|)$, and let $2^{(i)}$ be a new symbol and $\Sigma_i = \{a_i \mid a \in \Sigma\}$ be a disjoint copy of Σ for $0 \leq i \leq m$. Let R be the semi-Thue system over the alphabet

$$\Gamma = Q \cup \Sigma \cup \bigcup_{i=0}^{m} (\Sigma_i \cup \{2^{(i)}, A_i, B_i\}) \cup \{\#\}$$

that consists of the following rules:

(1)	$2^{(i)}ab \to a_m 2^{(i-1)} \cdots 2^{(1)} 2^{(0)} b$	for $0 \leq i \leq m$, $a \in \Sigma$, $b \in \Sigma$
(2)	$2^{(i)}a_k \to a_{k-1}2^{(i)}$	for $0 \leq i \leq m$, $1 \leq k \leq m$, $a \in \Sigma$
(3)	$\#a_k \to a\#$	for $0 \leq k \leq m$, $a \in \Sigma$
(4)	$\#x \to x$	for $x \in \Sigma \cup Q \backslash \{q_f\}$
(5)	$2^{(i)}q \to q$	for $0 \leq i \leq m$, $q \in Q \backslash \{q_f\}$
(6)	$2^{(i)}cqa \to cbp$	for $0 \leq i \leq m$, $c \in \Sigma$, $\delta(q,a) = (p,b,+1)$
(7)	$2^{(i)}cqa \to pcb$	for $0 \leq i \leq m$, $c \in \Sigma$, $\delta(q,a) = (p,b,-1)$
(8)	$A_i \to A_{i+1}A_{i+1}$	for $0 \leq i < m$
(9)	$B_i \to B_{i+1}B_{i+1}$	for $0 \leq i < m$
(10)	$A_m \to \#2^{(m)}$	
(11)	$B_m \to \Box$	
(12)	$xq_f \to q_f$	for $x \in \Gamma$
(13)	$q_f x \to q_f$	for $x \in \Gamma$

We claim that (Γ, R) is weight-reducing. For this we define the weight-

function f as follows:[6]

$$f(A_i) = 2 \cdot f(A_{i+1}) + 1 \text{ for } 0 \leq i < m \qquad f(A_m) = 2^m + 2$$
$$f(B_i) = 2 \cdot f(B_{i+1}) + 1 \text{ for } 0 \leq i < m \qquad f(B_m) = 2$$
$$f(x) = 1 \text{ for } x \in Q \cup \Sigma \cup \{\#\} \qquad f(2^{(i)}) = 2^i \text{ for } 0 \leq i \leq m$$
$$f(a_i) = 1 + \frac{i+1}{m+2} \text{ for } 0 \leq i \leq m,\ a \in \Sigma$$

Then, indeed $f(s) > f(t)$ for all $(s,t) \in R$. Confluence of (Γ, R) follows with the same argument as in the proof of Theorem 3.3.5. Finally we claim that $A_0 \Box q_0 w B_0 \overset{*}{\leftrightarrow}_R q_f$ if and only if $w \in L(\mathcal{T})$.

Before we prove this claim let us first explain the effect of the rules from R. For $0 \leq i \leq 2^m$ let $\beta(i) = 2^{(i_1)} \cdots 2^{(i_k)} \in \Gamma^*$, where $i_1 > \cdots > i_k$ and $i = 2^{i_1} + \cdots + 2^{i_k}$ (note that $\beta(0) = \varepsilon$). Let us call a word of the form $\#\beta(i) \in \Gamma^*$ a counter with value i. The effect of the rules in (1), (2), and (3) is to move counters to the right in words from Σ^*. If a whole counter moves one step to the right, its value is decreased by one. More generally, for all $u \in \Sigma^*$, $b \in \Sigma$, and all $|u| \leq i \leq 2^m$ we have $\#\beta(i)ub \overset{*}{\to}_R u\#\beta(i - |u|)b$. If a counter has reached the value 0, i.e., it only consists of the symbol $\#$, then the counter is deleted with a rule in (4). Also if a counter collides with a state symbol from Q at its right end, then the counter is deleted with the rules in (4) and (5). Note that such a collision may occur after an application of a rule in (7). The rules in (6) and (7) simulate the machine $\mathcal{T}$. In order to be weight-reducing, these rules consume the right-most symbol of the right-most counter. The rules in (8) and (10) produce 2^m many counters of the form $\#2^{(m)}$. Each of these counters can move at most 2^m cells to the right. But since $\mathcal{T}$ has to be simulated only for at most 2^m steps, the distance between the left end of the tape and the state symbol $q \in Q$ can be limited to 2^m as well. This implies that with each of the 2^m many counters that are produced from A_0, at least one step of $\mathcal{T}$ can be simulated. The rules in (9) and (11) produce 2^m many blank symbols, which is enough in order to simulate 2^m many steps of $\mathcal{T}$. Finally the rules in (12) and (13) make the final state q_f absorbing.

Now if $w \in L(\mathcal{T})$, then $A_0 \Box q_0 w B_0 \overset{*}{\to}_R (\#2^{(m)})^{2^m} \Box q_0 w \Box^{2^m} \overset{*}{\to}_R u q_f v \overset{*}{\to}_R q_f$ for some $u, v \in \Gamma^*$. On the other hand if $w \notin L(\mathcal{T})$, then $\mathcal{T}$ does not terminate on w. By simulating $\mathcal{T}$ until either all the 2^m many initial counters are completely consumed, or the state symbol moved to the

[6] Here we use rational weights, but of course they can be replaced by integer weights.

right end of the configuration word, we obtain an irreducible word $u \neq q_f$ with $A_0 \Box q_0 w B_0 \stackrel{*}{\rightarrow}_R (\#2^{(m)})^{2^m} \Box q_0 w \Box^{2^m} \stackrel{*}{\rightarrow}_R u \in \mathrm{IRR}(R)$. Since also $q_f \in \mathrm{IRR}(R)$ and R is confluent, $A_0 \Box q_0 w B_0 \stackrel{*}{\leftrightarrow}_R q_f$ cannot hold. □

Remark 3.4.5. *Since P is a proper subclass of EXPTIME, it follows from Theorem 3.4.3 and Theorem 3.4.4 that in general it is not possible to encode the alphabet of a weight-reducing and confluent presentation into a fixed alphabet with a polynomial blow-up such that the resulting presentation is still weight-reducing and confluent. For length-reducing presentations this is always possible, see the coding function from the proof of Theorem 3.3.5.*

Remark 3.4.6. *A problem that is closely related to the uniform word problem for weight-reducing and confluent presentations is the uniform membership problem for* quasi context-sensitive grammars. *A type-0 grammar $G = (N, T, S, \mathcal{P})$, where N is the set of nonterminals, T is the set of terminals, $S \in N$ is the start nonterminal, and $\mathcal{P} \subseteq (N \cup T)^* N (N \cup T)^* \times (N \cup T)^*$ is the finite set of productions, is called quasi context-sensitive, if there exists a weight-function $f : (N \cup T)^* \to \mathbb{N}$ with $f(u) \leq f(v)$ for all $(u, v) \in \mathcal{P}$, see [34]. In [34] it was shown that the uniform membership problem for quasi context-sensitive grammars is in EXPSPACE and NEXPTIME-hard. Using the technique from the proof of Theorem 3.4.4, we can show that this problem is in fact EXPSPACE-complete, see [124].*

3.5 Length-lexicographic presentations

In this section we consider length-lexicographic presentations, see for instance [106]. The complexity bounds that we will deduce in this section are the same that are known for preperfect presentations. A presentation (Γ, R) is *preperfect* if for all $s, t \in \Gamma^*$, $s \stackrel{*}{\leftrightarrow}_R t$ if and only if there exists $u \in \Gamma^*$ with $s \stackrel{*}{\mapsto}_R u$ and $t \stackrel{*}{\mapsto}_R u$, where the relation $\mapsto_R$ is defined by $v \mapsto_R w$ if $v \leftrightarrow_R w$ and $|v| \geq |w|$. Since every length-preserving presentation is preperfect and every linear bounded automaton can be easily simulated by a length-preserving semi-Thue system, there exists a fixed preperfect presentation with a PSPACE-complete word problem [25]. The following theorem may be seen as a stronger version of this well-known fact in the sense that a deterministic linear bounded automaton can be even simulated by a length-lexicographic, length-preserving, and confluent presentation. Recall that a

deterministic linear bounded automaton is a deterministic Turing-machine that operates in space n on an input of length n.

Theorem 3.5.1. *The uniform word problem for the class of all length-lexicographic and confluent presentations is PSPACE-complete. Moreover, there exists a fixed length-lexicographic and confluent presentation* (Γ, R) *with* $|\Gamma| = 2$ *such that the word problem for* $\mathcal{M}(\Gamma, R)$ *is PSPACE-complete.*

Proof. The first statement of the theorem is obvious. For the second statement let $\mathcal{T} = (Q, \Sigma, \delta, q_0, q_f)$ be a deterministic linear bounded automaton such that the question whether $w \in L(\mathcal{T})$ is PSPACE-complete. Such a linear bounded automaton exists, see e.g. [13]. Let $w \in (\Sigma \setminus \{\Box\})^*$ be an input for $\mathcal{T}$ with $|w| = n$. We may assume that $\mathcal{T}$ operates in phases, where a single phase consists of a sequence of $2 \cdot n$ transitions of the form $q_1 w_1 \overset{*}{\Rightarrow}_{\mathcal{T}} w_2 q_2 \overset{*}{\Rightarrow}_{\mathcal{T}} q_3 w_3$, where $w_1, w_2, w_3 \in \Sigma^*$ and $q_1, q_2, q_3 \in Q$. During the transition sequence $q_1 w_1 \overset{*}{\Rightarrow}_{\mathcal{T}} w_2 q_2$ only right-moves are made, whereas during the sequence $w_2 q_2 \overset{*}{\Rightarrow}_{\mathcal{T}} q_3 w_3$ only left-moves are made. A similar trick is used for instance also in [47]. W.l.o.g. $\mathcal{T}$ enters the final state q_f only when it reaches the left-most tape cell. Let $c > 0$ be a constant such that if $w \in L(\mathcal{T})$, then $\mathcal{T}$, started on w, reaches the final state q_f after at most $2^{c \cdot n}$ phases. As usual let $\overline{\Sigma}$ be a disjoint copy of Σ and similarly for $\overline{Q}$. Let $\Gamma = Q \cup \overline{Q} \cup \Sigma \cup \overline{\Sigma} \cup \{\triangleleft, 0, 1, \overline{1}\}$ and let R be the semi-Thue system over Γ that consists of the following rules:

$0\overline{q} \to \overline{q}\overline{1}$ for all $q \in Q$	$qa \to \overline{b}p$ if $\delta(q, a) = (p, b, +1)$
$1\overline{q} \to 0q$ for all $q \in Q$	$\overline{a}\,\overline{q} \to \overline{p}b$ if $\delta(q, a) = (p, b, -1)$
$q\overline{1} \to 1q$ for all $q \in Q$	$xq_f \to q_f$ for all $x \in \Gamma$
$q\triangleleft \to \overline{q}\triangleleft$ for all $q \in Q \setminus \{q_f\}$	$q_f x \to q_f$ for all $x \in \Gamma$

First we claim that (Γ, R) is length-lexicographic. For this choose a linear order $\succ$ on the alphabet Γ that satisfies $Q \succ 1 \succ 0 \succ \overline{\Sigma} \succ \overline{Q}$ (where, e.g., $Q \succ 1$ means that $q \succ 1$ for every $q \in Q$). Moreover, (Γ, R) is again confluent due to the absorbing symbol q_f. Finally we claim that $10^{c \cdot n} q_0 w\triangleleft \overset{*}{\leftrightarrow}_R q_f$ if and only if $w \in L(\mathcal{T})$. For $v = b_k \cdots b_0 \in \{0,1\}^*$ ($b_i \in \{0,1\}$) let $\mathrm{val}(v) = \sum_{i=0}^{k} b_i \cdot 2^i$. Note that for every $q \in Q$ and $s, t \in \{0,1\}^+$ with $s \neq 0^{|s|}$ it holds $s\overline{q} \overset{*}{\to}_R tq$ if and only if $|s| = |t|$ and $\mathrm{val}(t) = \mathrm{val}(s) - 1$. Now assume that $w \in L(\mathcal{T})$. Then $10^{c \cdot n} q_0 w\triangleleft \overset{*}{\to}_R v q_f u\triangleleft \overset{*}{\to}_R q_f$ for some $u \in \Sigma^{|w|}$ and $v \in \{0,1\}^*$. Now assume that $w \notin L(\mathcal{T})$. Then $\mathcal{T}$ does not terminate on w and we obtain $10^{c \cdot n} q_0 w\triangleleft \overset{*}{\to}_R 0^{c \cdot n + 1} \overline{q} u\triangleleft \overset{*}{\to}_R \overline{q}\overline{1}^{c \cdot n + 1} u\triangleleft \in \mathrm{IRR}(R)$, where

$u \in \Sigma^{|w|}$ and $q \in Q\backslash\{q_f\}$. Since also $q_f \in \mathrm{IRR}(R)$ and R is confluent, $10^{c \cdot n} q_0 w \triangleleft \overset{*}{\leftrightarrow}_R q_f$ cannot hold.

Finally, we have to encode the alphabet Γ into the alphabet $\{a, b\}$. For this let $\Gamma = \{a_1, \ldots, a_k\}$ and let $a_1 \succ a_2 \succ \cdots \succ a_k$ be the chosen linear order on Γ. Define a homomorphism $h : \Gamma^* \to \{a, b\}^*$ by $h(a_i) = ab^i ab^{2k+1-i}$ and let $a \succ b$. Then the presentation $(\{a, b\}, h(R))$ is also length-lexicographic and confluent, and for all $u, v \in \Gamma^*$, $u \overset{*}{\leftrightarrow}_R v$ if and only if $h(u) \overset{*}{\leftrightarrow}_{h(R)} h(v)$, see [27, p 60]. □

3.6 Weight-lexicographic presentations

The widest class of presentations that we study in this chapter are weight-lexicographic presentations. Let (Γ, R) be weight-lexicographic with $|\Gamma| = n$ and let $u \in \Gamma^*$. Thus, there exists a weight-function f with $f(s) \geq f(t)$ for all $(s, t) \in R$. By Lemma 3.4.2 we may assume that $f(a) \leq (n+1)(\alpha n)^n$ for all $a \in \Gamma$, where $\alpha = \max\{|u| \mid u \in \mathrm{dom}(R) \cup \mathrm{ran}(R)\}$. Thus, if $u \overset{*}{\to}_R v$, then $|v| \leq f(v) \leq f(u) \leq |u| \cdot (n+1)(\alpha n)^n$. Together with Theorem 3.5.1 it follows that the uniform word problem for weight-lexicographic and confluent presentations over a fixed alphabet of at least two symbols is PSPACE-complete. Moreover, there exists a fixed weight-lexicographic and confluent presentation with a PSPACE-complete word problem. For arbitrary weight-lexicographic and confluent presentations we have the following result.

Theorem 3.6.1. *The uniform word problem for the class of all weight-lexicographic and confluent presentations is EXPSPACE-complete.*

Proof. The EXPSPACE-upper bound can be shown by using the arguments from the preceding discussion. For the lower bound let $\mathcal{T} = (Q, \Sigma, \delta, q_0, q_f)$ be a deterministic Turing-machine, which uses for every input w at most space $2^{p(|w|)}$, where p is some polynomial. Similarly to the proof of Theorem 3.5.1 we may assume that $\mathcal{T}$ operates in phases. There exists a polynomial q such that if $w \in L(\mathcal{T})$, then $\mathcal{T}$, started on w, reaches q_f after at most $2^{2^{q(|w|)}}$ many phases. Let $w \in (\Sigma\backslash\{\Box\})^*$ be an arbitrary input for $\mathcal{T}$. Let $m = p(|w|)$, $n = q(|w|)$, and

$$\Gamma = Q \cup \overline{Q} \cup \Sigma \cup \overline{\Sigma} \cup \{\triangleleft, 0, 1, \overline{1}\} \cup \{A_i \mid 0 \leq i \leq n\} \cup \{B_i \mid 0 \leq i \leq m\}.$$

Let R be the semi-Thue system over Γ that consists of the following rules:

$0\overline{q} \to \overline{q}\overline{1}$	for all $q \in Q$
$1\overline{q} \to 0q$	for all $q \in Q$
$q\overline{1} \to 1q$	for all $q \in Q$
$qa \to \overline{b}p$	if $\delta(q,a) = (p,b,+1)$
$q\triangleleft \to \overline{q}\triangleleft$	for all $q \in Q \backslash \{q_f\}$
$\overline{a}\,\overline{q} \to \overline{p}b$	if $\delta(q,a) = (p,b,-1)$
$A_i \to A_{i+1}A_{i+1}$	for $0 \leq i < n$
$B_i \to B_{i+1}B_{i+1}$	for $0 \leq i < m$
$A_n \to 0$	
$B_m \to \Box$	
$xq_f \to q_f$	for all $x \in \Gamma$
$q_f x \to q_f$	for all $x \in \Gamma$

Note that the first six groups of rules are exactly the same groups that we used for the simulation of a linear bounded automaton in the proof of Theorem 3.5.1. In order to see that (Γ, R) is weight-lexicographic, we define the weight-function f by $f(x) = 1$ for all $x \in Q \cup \overline{Q} \cup \Sigma \cup \overline{\Sigma} \cup \{\triangleleft, 0, 1, \overline{1}, A_n, B_m\}$ and $f(A_i) = 2 \cdot f(A_{i+1})$, $f(B_j) = 2 \cdot f(B_{j+1})$ for $0 \leq i < n$, $0 \leq j < m$. Then the last two rules are weight-reducing and all other rules are weight-preserving. Now choose a linear order $\succ$ on Γ that satisfies $Q \succ 1 \succ 0 \succ \overline{\Sigma} \succ \overline{Q}$, $A_0 \succ A_1 \succ \cdots \succ A_n \succ 0$, and $B_0 \succ B_1 \succ \cdots \succ B_m \succ \Box$. Confluence of (Γ, R) follows with the usual arguments. Finally, $w \in L(\mathcal{T})$ if and only if $1A_0q_0wB_0\triangleleft \overset{*}{\leftrightarrow}_R q_f$. This can be shown by using the arguments from the proof of Theorem 3.5.1. Just note that this time from the prefix $1A_0$ we can generate the word 10^{2^n} which allows the simulation of 2^{2^n} many phases. Analogously to the proof of Theorem 3.4.4, the symbol B_0 generates enough blank symbols in order to satisfy the exponential space requirements of $\mathcal{T}$. $\Box$

The complexity results for word problems we have obtained in Section 3.3–3.6 are summarized in Table 3.1. The completeness results in the second row already hold for the alphabet $\{a, b\}$.

3.7 Confluence problems

The *confluence problem* for a class $\mathcal{P}$ of finite presentations is the following decision problem:

INPUT: A presentation $(\Gamma, R) \in \mathcal{P}$.

QUESTION: Is (Γ, R) confluent?

	length-red. & confluent	weight-red. & confluent	length-lex. & confluent	weight-lex. & confluent
word problem for fixed (Γ, R)	LOGCFL	LOGCFL	PSPACE, PSPACE-complete for some (Γ, R)	PSPACE, PSPACE-complete for some (Γ, R)
uniform word problem for fixed alph.	P-complete	P-complete	PSPACE-complete	PSPACE-complete
uniform word problem	P-complete	EXPTIME-complete	PSPACE-complete	EXPSPACE-complete

Table 3.1: Complexity results for word problems

The confluence problem for the class of all finite presentations is undecidable [13], whereas confluence is decidable for terminating presentations [154]. For length-reducing presentations the confluence problem is in P [26], the best known algorithm is the $O(\|(\Gamma, R)\|^3)$-algorithm from [105]. Furthermore in [122], the author has shown that the confluence problem for the class of all length-reducing presentations is P-complete. This was shown by using the following logarithmic space reduction from the uniform word problem for length-reducing and confluent presentations to the confluence problem for length-reducing presentations, see also [214]: Let (Γ, R) be length-reducing and confluent. Moreover, let A and B be two new symbols. Then for all $s, t \in \Gamma^*$ the length-reducing presentation

$$(\Gamma \cup \{A, B\},\ R \cup \{A^{|st|}B \to s,\ A^{|st|}B \to t\})$$

is confluent if and only if $s \overset{*}{\leftrightarrow}_R t$. Finally the alphabet $\Gamma \cup \{A, B\}$ can be reduced to the alphabet $\{a, b\}$ by using the coding function from the proof of Theorem 3.5.1. The same reduction also works for weight-reducing, length-lexicographic, and weight-lexicographic presentations, respectively. Thus, all lower bounds for uniform word problems from the preceding sections carry over to the corresponding confluence problems. Moreover, the same holds

	length-red.	weight-red.	length-lex.	weight-lex.
confluence problem for fixed alph.	P-complete	P-complete	PSPACE-complete	PSPACE-complete
confluence problem	P-complete	EXPTIME-complete	PSPACE-complete	EXPSPACE-complete

Table 3.2: Complexity results for confluence problems

for the given upper bounds as well: Our upper bound algorithms for uniform word problems are all based on the calculation of normal forms. But since every finite presentation has only polynomially many critical pairs, any upper bound for the calculation of normal forms also gives an upper bound for the confluence problem. The resulting complexity results are summarized in Table 3.2.

3.8 Automatic monoids

Automatic monoids were investigated for instance in [40, 97, 99, 157]. They generalize automatic groups, see [80]. Recall the notion of an α-automatic presentation from Section 2.5.

Let $\mathcal{M}$ be a finitely generated monoid and let Γ be a finite generating set. Recall the definition of the (left and right) Cayley-graph $\mathcal{C}_\beta(\mathcal{M}, \Gamma)$ ($\beta \in \{\ell, r\}$) from Section 2.6. We say that the triple (Γ, L, h) is an (α, β)-automatic presentation for $\mathcal{M}$ if the following holds:

- $L \subseteq \Gamma^*$ is a regular language,
- $h : L \to \mathcal{M}$ is surjective and maps every word $w \in L$ to the monoid element represented by w, and
- (Γ, L, h) is an α-automatic presentation of $\mathcal{C}_\beta(\mathcal{M}, \Gamma)$.

We say that $\mathcal{M}$ is (α, β)-*automatic* if there exists an (α, β)-automatic presentation for $\mathcal{M}$. Thus, we have four different basic notions of automaticity. Whereas for groups all these four variants are equivalent [97], one obtains 16 different notions of automaticity for monoids by combining the four basic variants of (α, β)-automaticity [97]. Here we will always work with the

strongest possible notion of automaticity: Our automatic monoids will be simultaneously (α,β)-automatic for all $\alpha,\beta \in \{\ell, r\}$.

Remark 3.8.1. *From the definition of an (α,β)-automatic monoid $\mathcal{M}$ it follows immediately that the structure $\mathcal{C}_\beta(\mathcal{M},\Gamma)$ is automatic. But the converse is false. The point is that for (α,β)-automaticity of $\mathcal{M}$, we require that the underlying set Γ of the α-automatic presentation (Γ, L, h) of $\mathcal{C}_\beta(\mathcal{M},\Gamma)$ has to be a generating set for $\mathcal{M}$ and that $h : L \to \mathcal{M}$ is a surjective restriction of the canonical homomorphism from Γ^* to $\mathcal{M}$. In fact, Hoffmann [97] has constructed a direct product $\mathcal{M} = \mathcal{M}_\ell \times \mathcal{M}_r$ such that $\mathcal{M}$ is not (α, r)-automatic for both $\alpha = \ell$ and $\alpha = r$. On the other hand, $\mathcal{M}_\alpha$ is (α, r)-automatic. Hence, there exists a finite generating set Γ_α of $\mathcal{M}_\alpha$ such that the Cayley-graph $\mathcal{C}(\mathcal{M}_\alpha, \Gamma_\alpha)$ is an automatic structure, which implies that also $\mathcal{C}(\mathcal{M}, \Gamma_\ell \cup \Gamma_r)$ is automatic.*

It is well-known that the word problem for an automatic group can be solved in quadratic time [80]. Moreover, the same algorithm also works for (α,β)-automatic monoids [40]. Here we will show that P is also a lower bound for the monoid case.

Theorem 3.8.2. *There exists a fixed length-lexicographic and confluent presentation (Γ, R) with the following properties, where $\mathcal{M} = \mathcal{M}(\Gamma, R)$ and $h : \mathrm{IRR}(R) \to \mathcal{M}$ is the bijection that maps a word $w \in \mathrm{IRR}(R)$ to the monoid element represented by w:*

- *$(\Gamma, \mathrm{IRR}(R), h)$ is an (α,β)-automatic presentation for $\mathcal{M}$ for all $\alpha,\beta \in \{\ell, r\}$, and*
- *the word problem for $\mathcal{M}$ is P-complete.*

Proof. Let $\mathcal{T} = (Q, \Sigma, \delta, q_0, q_f)$ be a fixed deterministic Turing-machine that accepts a P-complete language. Let $p(n)$ be a polynomial such that $\mathcal{T}$ terminates on an input $w \in L(\mathcal{T})$ after exactly $p(|w|)$ steps (this exact time bound can be easily enforced). Furthermore, we may assume that $\mathcal{T}$ is in the configuration $q_f\Box\cdots\Box$ when it terminates. Define $\Gamma = \Sigma \cup \overline{\Sigma} \cup Q \cup \{\$, \overline{\$}\}$ and let R be the following semi-Thue system over Γ:

$\overline{a}\,\overline{b}\,\overline{\$} \to \overline{a}\,\overline{\$}\,\overline{b}$	for $a, b \in \Sigma$
$\$\,a\,b \to a\,\$\,b$	for $a, b \in \Sigma$
$\$\,c\,q\,\overline{a}\,\overline{\$} \to c\,b\,p$	if $\delta(q,a) = (p, b, +1)$, $c \in \Sigma$
$\$\,c\,q\,\overline{a}\,\overline{\$} \to p\,\overline{c}\,\overline{b}$	if $\delta(q,a) = (p, b, -1)$, $c \in \Sigma$

It is easy to see that (Γ, R) is length-lexicographic and confluent. Next, let $w \in (\Sigma \setminus \{\Box\})^*$ be an arbitrary input for $\mathcal{T}$ and let $n = |w|, m = p(n)$. Then $w \in L(\mathcal{T})$ if and only if $\$^m \Box q_0 \overline{w} \overline{\Box}^m \overline{\$}^m \xrightarrow{*}_R \Box q_f \overline{\Box}^{m+n}$ if and only if $\$^m \Box q_0 \overline{w} \overline{\Box}^m \overline{\$}^m \overset{*}{\leftrightarrow}_R \Box q_f \overline{\Box}^{m+n}$. Thus, the word problem for $\mathcal{M}(\Gamma, R)$ is P-hard. It remains to show that for all $\alpha, \beta \in \{\ell, r\}$, $(\Gamma, \mathrm{IRR}(R), h)$ is an (α, β)-automatic presentation for $\mathcal{M}(\Gamma, R)$. Due to the symmetry present in R, it suffices to prove this only for $\beta = r$. Thus, we have to show that the relation

$$E_a = \{(u, v) \in \mathrm{IRR}(R) \times \mathrm{IRR}(R) \mid ua \xrightarrow{*}_R v\}$$

is α-automatic for all $a \in \Gamma$ and $\alpha \in \{\ell, r\}$. One should remark that all relations that appear in the following discussion have bounded length-difference. This allows to make use of Lemma 2.5.1.

Claim 1. Let $s \in \Sigma^*$ be a fixed word over the tape alphabet $\Sigma \subseteq \Gamma$. The following relation is α-automatic:

$$P_s = \{(u, v) \mid u, v \in (\Sigma \cup \{\$\})^* \cap \mathrm{IRR}(R),\ us \xrightarrow{*}_R v\}$$

Proof of Claim 1. Using induction on $|s|$ and the closure of α-automatic relations under relational composition, it suffices to consider the case $s = a \in \Sigma$. Let $a \in \Sigma$ and $u = u'\$a_n\$a_{n-1} \cdots \$a_2\$a_1$, where $u' \notin (\Sigma \cup \{\$\})^*\$\Sigma$. Then $ua \xrightarrow{*}_R u'a_n\$a_{n-1}\$ \cdots a_2\$a_1\$a \in \mathrm{IRR}(R)$. It follows easily that P_a is α-automatic. This proves Claim 1.

In the following let K be the regular set of all words from $\mathrm{IRR}(R)$ that do not end with a symbol from $\Sigma \cup \{\$\}$. By the previous claim and Lemma 2.5.1, the relation $\mathrm{Id}_K \cdot P_s = \{(tu, tv) \mid t \in K, (u, v) \in P_s\}$ is also α-automatic for every $s \in \Sigma^*$. Moreover, note that $t \in K$ and $u \in (\Sigma \cup \{\$\})^* \cap \mathrm{IRR}(R)$ implies that also $tu \in \mathrm{IRR}(R)$.

In order to prove that the relation E_a is α-automatic for every $a \in \Gamma$ and $\alpha \in \{\ell, r\}$, we can distinguish the following cases:

Case 1. $a \in \Sigma$

By the previous remark we have $E_a = \mathrm{Id}_K \cdot P_a$, which is α-automatic.

Case 2. $a \in \overline{\Sigma} \cup Q \cup \{\$\}$

By inspecting the rules of R, we see that $u \in \mathrm{IRR}(R)$ implies $ua \in \mathrm{IRR}(R)$. Thus, also E_a is α-automatic.

Case 3. $a = \overline{\$}$

Since $u \in \mathrm{IRR}(R) \setminus \Gamma^*\overline{\Sigma}$ implies $u\overline{\$} \in \mathrm{IRR}(R)$, we can write $E_{\overline{\$}} = E_1 \cup E_2 \cup E_3$, where

$$E_1 = \{(u, u\overline{\$}) \mid u, u\overline{\$} \in \mathrm{IRR}(R)\},$$
$$E_2 = \{(u, v) \in E_{\overline{\$}} \mid u\overline{\$} \in \mathrm{RED}(R),\ u \in \mathrm{IRR}(R) \cap \Gamma^*\overline{\Sigma} \setminus \Gamma^*\overline{\Sigma}\,\overline{\Sigma}\}, \text{ and}$$
$$E_3 = \{(u, v) \in E_{\overline{\$}} \mid u\overline{\$} \in \mathrm{RED}(R),\ u \in \mathrm{IRR}(R) \cap \Gamma^*\overline{\Sigma}\,\overline{\Sigma}\}.$$

Thus, it suffices to show that E_1, E_2, and E_3 are α-automatic. For E_1 this is clear. For E_2, let us consider a word $u \in \mathrm{IRR}(R) \cap \Gamma^*\overline{\Sigma} \setminus \Gamma^*\overline{\Sigma}\,\overline{\Sigma}$ such that $u\overline{\$} \in \mathrm{RED}(R)$. Then one of the following two cases holds:

Case 3.1. $u = u'\,\$\,c\,q\,\overline{a}$ such that $\$\,c\,q\,\overline{a}\,\overline{\$} \to p\,\overline{c}\,\overline{b}$ is a rule of R. Then $u\overline{\$} = u'\,\$\,c\,q\,\overline{a}\,\overline{\$} \to_R u'\,p\,\overline{c}\,\overline{b} \in \mathrm{IRR}(R)$.

Case 3.2. $u = u'\,\$\,c\,q\,\overline{a}$ such that $\$\,c\,q\,\overline{a}\,\overline{\$} \to c\,b\,p$ is a rule of R. Then $u\overline{\$} = u'\,\$\,c\,q\,\overline{a}\,\overline{\$} \to_R u'\,c\,b\,p$. Let us write $u' = u'_1u'_2$ such that $u'_1 \in K$ and $u'_2 \in (\{\$\} \cup \Sigma)^*$. Then $u'cbp = u'_1u'_2cbp \xrightarrow{*}_R u'_1v'_2p \in \mathrm{IRR}(R)$, where $u'_2cb \xrightarrow{*}_R v'_2 \in \mathrm{IRR}(R)$, i.e., $(u'_2, v'_2) \in P_{cb}$.

It follows that

$$E_2 = \{(u'\,\$\,c\,q\,\overline{a},\ u'\,p\,\overline{c}\,\overline{b}) \mid (\$\,c\,q\,\overline{a}\,\overline{\$},\ p\,\overline{c}\,\overline{b}) \in R,\ u'\,\$\,c\,q\,\overline{a} \in \mathrm{IRR}(R)\} \cup$$
$$\{(u'\$\,c\,q\,\overline{a},\ v'p) \mid (\$\,c\,q\,\overline{a}\,\overline{\$},\ c\,b\,p) \in R,\ (u', v') \in \mathrm{Id}_K \cdot P_{cb}\},$$

which is α-automatic.

It remains to show that E_3 is α-automatic. Consider a word $u \in \Gamma^*\overline{\Sigma}\,\overline{\Sigma}$. Write $u = u_1\overline{a}\,u_2$ such that $u_1 \notin \Gamma^*\overline{\Sigma}$ and $u_2 \in \overline{\Sigma}^+$. Then $u\overline{\$} = u_1\overline{a}\,u_2\overline{\$} \xrightarrow{*}_R u_1\overline{a}\,\overline{\$}\,u_2$. Let v such that $u_1\overline{a}\,\overline{\$} \xrightarrow{*}_R v$. Thus, $(u_1\overline{a}, v) \in E_1 \cup E_2$ and $u_1\overline{a}\,\overline{\$}\,u_2 \to_R v\,u_2 \in \mathrm{IRR}(R)$. It follows $E_3 = (E_1 \cup E_2) \cdot \mathrm{Id}_{\overline{\Sigma}^+}$. □

Corollary 3.8.3. *There exists a fixed monoid with a P-complete word problem, which is simultaneously (α, β)-automatic for all $\alpha, \beta \in \{\ell, r\}$.*

3.9 Open problems

Let us mention a few open problems concerning the complexity of word problems that might deserve further investigations.

- Is the word problem for the free group of rank 2 in uNC^1 or is it L-complete? The solution of this question has a strong impact on the results from Section 3.2. In [171], uniform circuits of polynomial size and depth $\log(n)\frac{\log\log(n)}{\log\log\log(n)}$ for the word problem for F_2 were constructed.

- Does there exist a length-reducing and confluent presentation with a LOGCFL-complete word problem?
- Is LOGCFL an optimal upper bound for the word problem of a hyperbolic group? Let us mention that the word problem of a hyperbolic group can be recognized in real time [98].
- What is the complexity of the word problem for monoids that can be presented by monadic and confluent (resp. erasing and confluent) presentations, both in the uniform and nonuniform setting? From Theorem 3.3.4 one obtains a LOGDCFL-upper bound. Equivalent questions were investigated in [15] in the context of McNaughton languages. In particular, [15, Thm. 50] implies that the uniform word problem for erasing and confluent presentations is L-hard. But the exact complexity of this problem remains open.
- Does there exist an automatic group with a P-complete word problem?

Another interesting open problem is the descriptive power of the presentations considered in Section 3.3–3.6. Let $\mathcal{C}_{\ell r}$ (resp. $\mathcal{C}_{wr}$, $\mathcal{C}_{\ell\ell}$, $\mathcal{C}_{w\ell}$) be the class of all monoids of the form $\mathcal{M}(\Gamma, R)$, where (Γ, R) is length-reducing (resp. weight-reducing, length-lexicographic, weight-lexicographic) and confluent. In [61] it was shown that the monoid $\mathcal{M}(a, b, c \mid ab = cc)$ is not contained in $\mathcal{C}_{\ell r}$, but it is clearly contained in $\mathcal{C}_{wr} \cap \mathcal{C}_{\ell\ell}$. Thus, $\mathcal{C}_{\ell r}$ is strictly contained in $\mathcal{C}_{wr}$ and $\mathcal{C}_{\ell\ell}$. Furthermore the trace monoid $\mathcal{M}(a, b \mid ab = ba)$ is contained in $\mathcal{C}_{\ell\ell} \backslash \mathcal{C}_{wr}$ [63, p 90]. Hence, if there exists a monoid in $\mathcal{C}_{wr} \backslash \mathcal{C}_{\ell\ell}$, then $\mathcal{C}_{wr}$ and $\mathcal{C}_{\ell\ell}$ are incomparable and both are proper subclasses of $\mathcal{C}_{w\ell}$. But we do not know whether this holds. A survey on the descriptive power of certain classes of presentations can be found in [129].

Another interesting class of presentations, for which (uniform) word problems were studied, is the class of commutative presentations (i.e., for all generators a and b, the pair (ab, ba) belongs to the rules), see for instance [42, 101, 102, 122, 137] for several decidability and complexity results. But there are still many interesting open problems, see for instance the remarks in [101, 102, 122].

Chapter 4

Logic over Cayley-graphs

4.1 Outline

In this chapter we study Cayley-graphs of monoids and groups (see Section 2.6) from the viewpoint of mathematical logic. A Cayley-graph can be seen as a relational structure, hence we can study its first-order and monadic second-order theory. Section 4.2 collects a few basic observations about theories of Cayley-graphs. In particular, we establish simple connections between the decidability of the word problem and the decidability of the first-order theory of the Cayley-graph. To obtain a deeper understanding, we have to consider more general classes of structures.

Section 4.3 is devoted to the study of graphs with decidable monadic second-order theories. Section 4.3.1 and 4.3.3 introduce necessary definitions concerning undirected and directed graphs. Our main tool for the investigation of graphs with decidable MSO theories are strong tree decompositions, which will be introduced in Section 4.3.2, moreover several combinatorial properties concerning these decompositions will be shown. Using these properties in combination with two results of Seese and Courcelle, respectively, we are able to prove the main result of Section 4.3 (Theorem 4.3.10): a connected graph of bounded degree, whose automorphism group has only finitely many orbits, has a decidable MSO theory if and only if it is context-free (i.e., is the transition graph of a pushdown automaton). One direction of this result was shown by Muller and Schupp [145] without the restriction on the automorphism group: every context-free graph has a decidable MSO theory. On the other hand, note that there exists a connected graph of bounded

degree with a decidable MSO theory, which is not context-free, see e.g. [145, p 72] for an example.

In [145], Muller and Schupp have also shown that the Cayley-graph of a finitely generated group $\mathcal{G}$ is a context-free graph if and only if $\mathcal{G}$ is a context-free group (see Section 2.4). Since Cayley-graphs of groups satisfy all the requirements of Theorem 4.3.10, we obtain also the converse implication, hence a group is context-free if and only if its Cayley-graph has a decidable MSO theory (Corollary 4.4.1 in Section 4.4.1). In order to obtain a similar result for first-order logic, we apply a technique developed by Ferrante and Rackhoff, which is based on Ehrenfeucht-Fraisse games. We introduce this method in a slight variant in Section 4.4.2. Based on this technique, we can show that a group has a decidable word problem if and only if the Cayley-graph of the group has a decidable first-order theory (Corollary 4.4.9 in Section 4.4.3).

The results mentioned in the previous paragraph do not carry over to monoids, see e.g. Proposition 4.2.3. Our main result about the monoid case states the preservation of the decidability of the first-order (resp. monadic second-order) theory under some well-known algebraic constructions. In order to obtain these results, we have to investigate a construction that works for arbitrary relational structures: In [219], Walukiewicz proved that the MSO theory of the tree-like unfolding (see Section 4.6.1) of a structure can be reduced to the MSO theory of the original structure, the original statement goes back to [186, 191, 201]. Using this deep result, we prove in Section 4.7 that the class of finitely generated monoids, whose Cayley-graphs have decidable MSO theories, is closed under finite free products (Theorem 4.7.5(2)). The same result also holds for first-order theories, in fact we will prove a more general preservation theorem in this case. For this, we have to generalize tree-like unfoldings. This leads us to the notion of a factorized unfolding: the tree-like unfolding of a structure $\mathcal{A}$ consists of the set of words over the set of elements of $\mathcal{A}$. This set of words is equipped with the natural tree structure. Hence the successors of any element of the tree can be identified with the elements of $\mathcal{A}$ and can therefore naturally be endowed with the structure of $\mathcal{A}$. Basically, a factorized unfolding is the quotient of this structure with respect to Mazurkiewicz's trace equivalence, in fact, it is a generalization of this quotient (see Section 4.6.2). In general, the monadic second-order theory of a factorized unfolding may be undecidable, even in case the underlying structure has a decidable monadic second-order theory. On the other hand, the first-order theory of a factorized unfolding can be reduced to the

first-order theory of the underlying structure (Theorem 4.6.6). Section 4.6.2 is devoted to the proof of this result. It, again, uses the technique of Ferrante and Rackoff [83] and a thorough analysis of factorized unfoldings using ideas from the theory of Mazurkiewicz traces. In particular, the first-order theory of a factorized unfolding is decidable in case the original structure has a decidable first-order theory. Based on this result, we will prove that the class of finitely generated monoids, whose Cayley-graphs have decidable first-order theories, is closed under finite graph products (Theorem 4.7.5(1)). The graph product is a well-known construction in mathematics, see e.g. [87, 94, 211].

Some of the results of this chapter can be also found in the extended abstract [116].

4.2 Basic results on Cayley-graphs

Similarly to the word problem, the chosen generating set has no influence on the decidability (or complexity) of the first-order (resp. monadic second-order) theory of the Cayley-graph.

Proposition 4.2.1. *Let* Γ *and* Σ *be finite generating sets of the same monoid* $\mathcal{M}$. *Then* $\mathrm{FOTh}(\mathcal{C}(\mathcal{M},\Gamma))$ *(resp.* $\mathrm{MSOTh}(\mathcal{C}(\mathcal{M},\Gamma))$*) is logspace-reducible to* $\mathrm{FOTh}(\mathcal{C}(\mathcal{M},\Sigma))$ *(resp.* $\mathrm{MSOTh}(\mathcal{C}(\mathcal{M},\Sigma))$*).*

Proof. The proposition follows from a simple first-order interpretation of $\mathcal{C}(\mathcal{M},\Gamma)$ in $\mathcal{C}(\mathcal{M},\Sigma)$. □

Whenever the specific generating set Γ will be not important, we will briefly write $\mathcal{C}(\mathcal{M})$ instead of $\mathcal{C}(\mathcal{M},\Gamma)$. The next proposition states a simple connection between the word problem and the first-order theory of the Cayley-graph.

Proposition 4.2.2. *Assume that* $\mathcal{M}$ *is a finitely generated monoid such that* $\exists\mathrm{FOTh}(\mathcal{C}(\mathcal{M}))$ *is decidable. Then the word problem for* $\mathcal{M}$ *is decidable.*

Proof. Choose a finite generating set Γ for $\mathcal{M}$. Two given words $u = a_0a_1\cdots a_{m-1}$ and $v = b_0b_1\cdots b_{n-1}$, where $a_i, b_j \in \Gamma$, represent different elements in $\mathcal{M}$ if and only if there exists $x \in \mathcal{M}$ such that the (unique) paths in $\mathcal{C}(\mathcal{M},\Gamma)$ starting in x and labeled by u and v, respectively, end in different nodes. This fact can be easily expressed by an existential sentence of

first-order logic:[1]

$$\exists x_0 \cdots \exists x_m \exists y_0 \cdots \exists y_n \left\{ \begin{array}{l} x_0 = y_0 \ \wedge \ x_m \neq y_n \ \wedge \\ \bigwedge\limits_{0 \leq i < m} x_i a_i = x_{i+1} \ \wedge \ \bigwedge\limits_{0 \leq i < n} y_i b_i = y_{i+1} \end{array} \right\}$$

□

On the other hand, the converse implication of Proposition 4.2.2 becomes false even for finitely presented monoids, whose word problem can be solved in linear time:

Proposition 4.2.3. *There exists a fixed finite, length-reducing, and confluent presentation* (Γ, R) *such that* $\exists\mathrm{FOTh}(\mathcal{C}(\mathcal{M}(\Gamma, R)))$ *is undecidable.*

Proof. By [150, Thm. 2.4] there exists a fixed finite, length-reducing, and confluent presentation (Γ, R) such that the common right-multiplier problem is undecidable for $\mathcal{M}(\Gamma, R)$,[2] which is the following problem:

INPUT: Words $u, v \in \Gamma^*$

QUESTION: Does there exist $x \in \mathcal{M}(\Gamma, R)$ with $xu = xv$ in $\mathcal{M}(\Gamma, R)$?

But this is an existential property of the Cayley-graph of $\mathcal{M}(\Gamma, R)$ that can be constructed effectively from u and v. This proves the proposition. □

4.3 Graphs

In order to obtain a characterization of the class of finitely generated groups $\mathcal{G}$ such that $\mathrm{MSOTh}(\mathcal{C}(\mathcal{G}))$ is decidable, it is necessary to study arbitrary graphs and their MSO theories. This is the aim of this section.

4.3.1 Undirected graphs

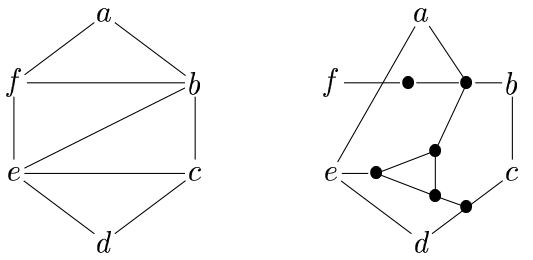
An *undirected graph* is a relational structure $G = (V, E)$, where V is called the set of nodes and $E \subseteq V \times V$ is a symmetric and irreflexive edge relation (thus, undirected graphs do not have self loops). All notions that were defined for arbitrary relational structures in Section 2.2 will be also used for undirected

[1] In the following, we will write $xa = y$ for a generator $a \in \Gamma$ in case there is an a-labeled edge from x to y in $\mathcal{C}(\mathcal{M}, \Gamma)$.

[2] In [150, Thm. 2.4] undecidability is stated for the common left-multiplier problem, but by reversing R, undecidability is also obtained for the common right-multiplier problem.

graphs. We will also use the notation $V(G) = V$ and $E(G) = E$. A *path* of length $n \geq 0$ in G between $u \in V$ and $v \in V$ is a sequence $[v_0, v_1, \ldots, v_n]$ of nodes such that $v_0 = u$, $v_n = v$, and $(v_i, v_{i+1}) \in E$ for all $0 \leq i < n$, it is a *closed path* if $u = v$, it is a *simple path* if $v_i \neq v_j$ for $i \neq j$. We write $d_G(u, v)$ for the distance between the nodes $u, v \in V$, i.e., $d_G(u, v)$ is the minimal length of a path between u and v. If such a path does not exist, we write $d_G(u, v) = \infty$. The *r-sphere, centered at* $v \in V$, is $S_G(r, v) = \{u \in V \mid d_G(v, u) \leq r\}$. For a k-tuple $\widetilde{v} = (v_1, \ldots, v_k) \in V^k$ we define $S_G(r, \widetilde{v}) = \bigcup_{i=1}^{k} S_G(r, v_i)$. The graph G is *connected* if $d_G(u, v) < \infty$ for all $u, v \in V$. The graph G is *acyclic* if G does not contain a closed path $[v_1, v_2, \ldots, v_n, v_1]$ such that $n \geq 3$ and $[v_1, v_2, \ldots, v_n]$ is simple. An acyclic graph is also called a *forest*, a connected forest is called a *tree*. Let $U \subseteq V$. The undirected graph $G{\restriction}U$ is called the *subgraph of G, induced by U*. The *diameter* $\mathrm{diam}_G(U)$ *of* U is the supremum in $\mathbb{N} \cup \{\infty\}$ of the set $\{d_G(u, v) \mid u, v \in U\}$. A *connected component* of G is an induced subgraph $G{\restriction}U$ such that $U = \{u \in V \mid d_G(v, u) < \infty\}$ for some node $v \in V$. The *degree* of a node $v \in V$ is the cardinality of the set $\{u \in V \mid (v, u) \in E\}$. The graph G is called *of bounded degree*, if there exists some $d \in \mathbb{N}$ such that each node $v \in V$ has degree at most d. In this case we also say that G is *of bounded degree* d.

Let $\pi = [v_1, v_2, \ldots, v_m, v_1]$ be a sequence of nodes $v_i \in V$ (which is not necessarily a path). With π we associate a closed convex polygon $\mathrm{Pol}(\pi)$ in the plane, whose boundary has m vertices $x_1, \ldots, x_m$, which are labeled in clockwise order with $v_1, \ldots, v_m$. For $M \geq 1$, an *M-triangulation* of π is a plane triangulation of $\mathrm{Pol}(\pi)$ with vertex set $\{x_1, x_2, \ldots, x_m\}$ and additional edges of the form (x_i, x_j) for $d_G(v_i, v_j) \leq M$, only. We say that G can be *M-triangulated*, if every *closed path* π of G can be M-triangulated [144, 220]. The example below shows a 3-triangulation of the sequence $[a, b, c, d, e, f]$ in the graph on the right. We have three additional edges in the triangulation, namely (b, e), (b, f), and (c, e).



A *tree decomposition* of $G = (V, E)$ is a pair (T, f), where T is a tree and

$f : V(T) \to 2^V \setminus \{\emptyset\}$ is a function such that the following holds:

- $\bigcup_{w \in V(T)} f(w) = V$,
- for every $(u, v) \in E$ there exists $w \in V(T)$ such that $u, v \in f(w)$, and
- if $w_1, w_3 \in V(T)$ and w_2 lies on the unique simple path from w_1 to w_3 in the tree T, then $f(w_1) \cap f(w_3) \subseteq f(w_2)$.

The supremum in $\mathbb{N} \cup \{\infty\}$ of the cardinalities $|f(w)|$, $w \in V(T)$, is called the *width* of the tree decomposition. We say that G has *tree-width* $\leq b$ if there exists a tree decomposition of width $\leq b$. Finally G has *finite tree-width* if it has tree-width $\leq b$ for some $b \in \mathbb{N}$. The notion of tree-width was introduced in [170], and plays a central role in Robertson's and Seymour's theory of graph minors, see e.g. [71] for an overview.

For the rest of Section 4.3, a strengthening of the notion of tree-width will be more important. The next section introduces this strengthening.

4.3.2 Strong tree decompositions

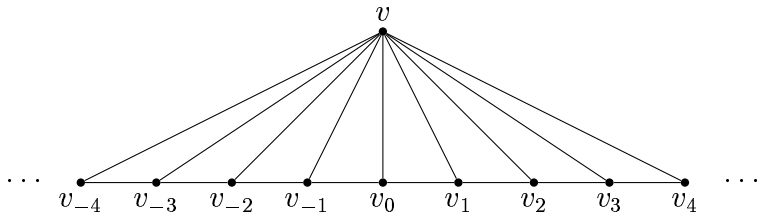
Let $G = (V, E)$ be an undirected graph and let P be a partition of V, i.e., $P \subseteq 2^V \setminus \{\emptyset\}$, $W_1 \cap W_2 = \emptyset$ for W_1 and W_2 distinct elements of P, and $\bigcup_{W \in P} W = V$. We define the *quotient graph* of G by P as the undirected graph

$$G/_P = (P,\ \{(W_1, W_2) \in P \times P \mid W_1 \neq W_2,\ (W_1 \times W_2) \cap E \neq \emptyset\}).$$

A *strong tree decomposition* of G is a partition P of V such that the quotient graph $G/_P$ is a forest. Note that if G is connected and P is a strong tree decomposition of G, then $G/_P$ must be also connected, i.e., it is a tree. The *width* of a strong tree decomposition P is defined as the supremum in $\mathbb{N} \cup \{\infty\}$ of the cardinalities $|W|$ for $W \in P$. We say that G has *strong tree-width* $\leq b$ if there exists a strong tree decomposition of width $\leq b$, G has *finite strong tree-width* if it has strong tree-width $\leq b$ for some $b \in \mathbb{N}$. The notion of strong tree-width is taken from [182].

Any strong tree decomposition $\{W_i \mid i \in J\}$ gives rise to a tree decomposition formed by the sets $W_i \cup W_j$ whenever $W_i \times W_j$ contains some edge of the graph. Thus, a graph of finite strong tree-width has finite tree-width as well. On the other hand, the converse implication is in general false:

Example 4.3.1. *Let G be the following graph of unbounded degree:*



A tree decomposition of G of width 3 is (T, f) with

$$T = (\mathbb{Z}, \{(n, n+1), (n+1, n) \mid n \in \mathbb{Z}\})$$

(this is a tree) and $f(n) = \{v, v_n, v_{n+1}\}$. On the other hand, for every partition P of the set of nodes of G such that every partition class of P is finite, it is easy to see that $G/_P$ has to contain a triangle. Note also that $\mathrm{Aut}(G)$ *has only two orbits on G.*

Our first result of this section, Theorem 4.3.4, states that at least for graphs of bounded degree, finite strong tree-width implies finite tree-width. The second and more important result is that under some conditions one can even find a strong tree decomposition of finite width such that all partition classes have a uniformly bounded diameter (Theorem 4.3.7).

In [206] it is shown that an arbitrary graph G has tree-width $\leq b$ if and only if every finite subgraph of G has tree-width $\leq b$. The corresponding statement for strong tree-width is also true, at least if we restrict to countable graphs:

Lemma 4.3.2. *Let G be countable. Then G has strong tree-width $\leq b$ if and only if every finite subgraph of G has strong tree-width $\leq b$.*

Proof. If G has strong tree-width $\leq b$, then it is easy to see that also every finite subgraph of G has strong tree-width $\leq b$. Now assume that every finite subgraph of G has strong tree-width $\leq b$. Since G is countable, we can choose an enumeration $v_0, v_1, \ldots$ of $V(G)$. Let $G_i = G{\restriction}\{v_0, \ldots, v_i\}$. We construct an infinite, finitely branching rooted tree T as follows: the nodes on the i-th level of T are precisely the strong tree decompositions of G_i of width $\leq b$. This set is finite and by assumption nonempty. The root of T is the trivial strong tree decomposition of the graph $(\{v_0\}, \emptyset)$. Now let P_i (resp. P_{i+1}) be a strong tree decomposition of G_i (resp. G_{i+1}) of width $\leq b$. We put an edge between P_i and P_{i+1} in the tree T, if P_i results from restricting P_{i+1} to the

nodes in $V_i = \{v_0, \ldots, v_i\}$, i.e., $P_i = \{W \cap V_i \mid W \in P_{i+1}, W \cap V_i \neq \emptyset\}$. Note that this defines indeed a tree T. Since T is infinite but finitely branching, König's Lemma implies that T has an infinite path $P_0, P_1, \ldots$, where P_i is a strong tree decomposition of G_i of width $\leq b$. By taking the limit along this sequence, we obtain a strong tree decomposition P of G of width $\leq b$. More precisely, let $W \in P$ if and only if there is $i \geq 0$ with $W \in P_i$ but there do not exist $j > i$ and $W' \in P_j$ with $W \subsetneq W'$. Then P is a partition of $V(G)$ and $G/_P$ is acyclic. □

To the knowledge of the author it is open whether Lemma 4.3.2 also holds for uncountable graphs. For the further discussion this problem has no relevance.

The following theorem was first stated in [22, Cor. 13]. It can be derived from a corresponding result for domino tree-width, which was independently shown in [72]. Later, a simplified proof was given in [21].

Theorem 4.3.3 (cf. [22]). *Let b and d be constants and let $\mathcal{S}$ be a set of finite graphs such that every $G \in \mathcal{S}$ is of bounded degree d and has tree-width $\leq b$. Then there exists a constant $c_{b,d}$ such that every $G \in \mathcal{S}$ has strong tree-width $\leq c_{b,d}$.*

In [21] it was shown that $c_{b,d} = (9b + 7)d(d + 1) - 2$ suffices in the previous theorem.

Theorem 4.3.4. *Let G be a graph of bounded degree. Then G has finite tree-width if and only if G has finite strong tree-width.*

Proof. If G has a strong tree decomposition P of width $\leq b$, then we can construct a tree decomposition of G of width $\leq 2b$ from all sets $W_1 \cup W_2$, where $W_1, W_2 \in P$ and $E(G) \cap (W_1 \times W_2) \neq \emptyset$.

Now assume that G has finite tree-width b and bounded degree d. First, we consider the case that G is connected. Since G is of bounded degree, G must be countable. Let $\mathcal{S}$ be the set of finite subgraphs of G. Then the tree-width of every graph in $\mathcal{S}$ is bounded by b [206]. Trivially, the degree of every graph in $\mathcal{S}$ is also bounded by d. From Theorem 4.3.3, we can infer that the strong tree-width of the graphs in $\mathcal{S}$ is bounded by some constant c. Hence, by Lemma 4.3.2, also G has strong tree-width $\leq c$. If G is not connected, then the above argument yields that all its connected components have strong tree-width $\leq c$. Hence c bounds the strong tree-width of G as well. □

For the further consideration let us fix a connected graph $G = (V, E)$. If V is partitioned into nonempty sets V_1 and V_2, then the set of edges

$$C = E \cap [(V_1 \times V_2) \cup (V_2 \times V_1)]$$

is a *cut* of G. If $|C| \leq 2k$, then C is called a *k-cut* of G (we choose $2k$ here, since for undirected graphs, edges always come in pairs). The sets V_1 and V_2 are called the *sides* of the cut C. If both $G{\restriction}V_1$ and $G{\restriction}V_2$ are connected subgraphs of G, then C is called a *tight cut*. The importance of tight cuts in our context comes from the following result of Dunwoody [74, Paragraph 2.5], later a simplified proof was given in [207, Prop. 4.1].

Lemma 4.3.5 (cf. [74, 207]). *Let G be a connected graph and let $k \in \mathbb{N}$. Then every edge of G is contained in only finitely many tight k-cuts of G.*

Let P be a strong tree decomposition of the graph G and let $e = (W_1, W_2) \in E(G/_P)$ be an edge of $G/_P$. Since $\{e\}$ is a cut of the tree $G/_P$, we can define a cut

$$\mathrm{cut}(e) = E \cap [(W_1 \times W_2) \cup (W_2 \times W_1)]$$

of $G/_P$. We say that P is *tight* if $\mathrm{cut}(e)$ is tight for all $e \in E(G/_P)$.

Lemma 4.3.6. *Let G be connected and of strong tree-width $\leq b$. Then there exists a tight strong tree decomposition of G of width $\leq b$.*

Proof. Let P be a strong tree decomposition of G of width $\leq b$. We first refine P maximally to a strong tree decomposition Q, where Q is *finer than* P – written $Q \preceq P$ – if for any $W \in Q$, there is $W' \in P$ with $W \subseteq W'$. Then, we will show that Q is tight.

So let $(P_\alpha)_{\alpha<\kappa}$ be some decreasing chain (with respect to $\preceq$) of strong tree decompositions P_α with $P_0 = P$, where κ is some ordinal. If we order the sets in $\bigcup_{\alpha<\kappa} P_\alpha$ under set inclusion, we obtain a disjoint union of finite trees (one for each $W \in P$). Then the set Q of all minimal elements in $\bigcup_{\alpha<\kappa} P_\alpha$ (with respect to $\subseteq$) is a partition of $V(G)$. Assume that there is a cycle in $G/_Q$, involving the nodes $U_1, \ldots, U_m \in Q$. Then there is some $\alpha < \kappa$ with $U_1, \ldots, U_m \in P_\alpha$, contradicting our assumption that P_α is a strong tree decomposition. Thus, Q is a strong tree decomposition of G. We have shown that any decreasing chain of strong tree decompositions is bounded from below. By Zorn's Lemma, this implies the existence of a minimal (with respect to $\preceq$) strong tree decomposition Q in $\{P' \mid P' \preceq P\}$.

Suppose Q is not tight, i.e., there is an edge $e \in E(G/_Q)$ such that $\mathrm{cut}(e)$ is not tight. Let $U \subseteq V(G)$ be one of the sides of $\mathrm{cut}(e)$ such that $G{\restriction}U$ is not connected and let $U_j \subseteq U$ $(j \in J)$ be the node sets of the connected components of $G{\restriction}U$. Let

$$P' = \{W \cap U_j \mid W \in Q,\ j \in J,\ W \cap U_j \neq \emptyset\} \cup \{W \in Q \mid W \subseteq V(G) \setminus U\}.$$

Then P' is a strong tree decomposition of G that is finer than Q, because if $e = (W_1, W_2)$, then at least the W_i with $W_i \subseteq U$ will be refined. We have obtained a contradiction. □

The next theorem is the main result of this section. Recall the notion of an orbit, which was defined in Section 2.2 for arbitrary relational structures.

Theorem 4.3.7. *Let G be a connected graph of bounded degree and of finite tree-width such that $\mathrm{Aut}(G)$ has only finitely many orbits on G. Then there exists a strong tree decomposition P of G of finite width and a constant c such that for all $W \in P$, $\mathrm{diam}_G(W) \leq c$.*

Proof. By Theorem 4.3.4, $G = (V, E)$ has strong tree-width $\leq b$ for some constant b. Thus, by Lemma 4.3.6 there exists a tight strong tree decomposition P of G of width $\leq b$. Hence, for all $e \in E(G/_P)$, $\mathrm{cut}(e)$ is a tight b^2-cut. In the following, for a cut $C \subseteq E$ let $V(C)$ denote the set of all $u \in V$ such that $(u, v) \in C$ for some $v \in V$. Let $\mathcal{O}_1, \ldots, \mathcal{O}_n$ be the orbits of $\mathrm{Aut}(G)$ on G. For every $1 \leq i \leq n$ choose a node $v_i \in \mathcal{O}_i$. Since every node v_i has only finitely many adjacent edges and, by Lemma 4.3.5, each of these edges is contained in only finitely many tight b^2-cuts, it follows that there are only finitely many tight b^2-cuts C with $V(C) \cap \{v_1, \ldots, v_n\} \neq \emptyset$. Let $\mathcal{C}$ be the set of all these cuts. Since $\mathcal{C}$ is finite and G is connected, we can define $d = \max\{\mathrm{diam}_G(V(C)) \mid C \in \mathcal{C}\} \in \mathbb{N}$.

Now, if $e \in E(G/_P)$, then $\mathrm{cut}(e)$ can be mapped via some $f \in \mathrm{Aut}(G)$ to some cut in $\mathcal{C}$ (take any node $v \in V(\mathrm{cut}(e))$, then v can be mapped via some $f \in \mathrm{Aut}(G)$ to some v_i, thus f maps $\mathrm{cut}(e)$ to some cut from $\mathcal{C}$). Thus, $\mathrm{diam}_G(V(\mathrm{cut}(e))) \leq d$.

Now let $W \in P$ and let $e_1, \ldots, e_m \in E(G/_P)$ be all those edges that are adjacent with W in $G/_P$. Let $V_i = V(\mathrm{cut}(e_i)) \cap W$. Thus, also $\mathrm{diam}_G(V_i) \leq d$ for all $1 \leq i \leq m$. Choose $u, v \in W$ with $u \neq v$. We will show that $d_G(u, v) \leq bd - 1$, which proves the theorem. Since G is connected, we can choose a simple path π in G between u and v of minimal length. Since $G/_P$

is a tree, we can split the path π into subpaths $\pi_1, \nu_1, \pi_2, \nu_2, \ldots, \pi_\ell, \nu_\ell, \pi_{\ell+1}$ ($\ell \geq 0$) such that

- π_1 starts in u, $\pi_{\ell+1}$ ends in v, and the final node of π_i (resp. ν_i) is the initial node of ν_i (resp. π_{i+1}),
- for all i, π_i is completely contained in W, and
- for all i, there exists j such that both the initial and final node of ν_i belong to V_j and are different, thus the length of ν_i is bounded by $d \geq 1$.

Since π is simple, the sum of the lengths of all paths π_i is at most $|W|-1-\ell \leq b-1-\ell$, and moreover $\ell \leq |W| \leq b$. It follows that the length of π is bounded by $b-1-\ell+\ell \cdot d = b-1+\ell(d-1) \leq b-1+b(d-1) = bd-1$. □

It is open whether every graph G that satisfies the conditions of Theorem 4.3.7 has a strong tree decomposition P of finite width such that moreover every partition class in P is connected.

4.3.3 Labeled directed graphs

Let Γ be some finite alphabet of labels. A Γ-*labeled directed graph* is a relational structure $G = (V, (E_a)_{a\in\Gamma})$, where V is the set nodes and $E_a \subseteq V \times V$ is the set of a-labeled directed edges. Note that self-loops are allowed in directed graphs. The Cayley-graph $\mathcal{C}(\mathcal{M}, \Gamma)$ of a monoid $\mathcal{M}$ with respect to the finite generating set Γ is for instance a Γ-labeled directed graph. Let us fix $G = (V, (E_a)_{a\in\Gamma})$ for the further discussion. We associate with G the unlabeled undirected graph

$$\mathrm{ud}(G) = (V, \bigcup_{a\in\Gamma}\{(u,v) \mid u \neq v,\ (u,v) \in E_a \text{ or } (v,u) \in E_a\}).$$

We say that G is connected (resp. of bounded degree) if $\mathrm{ud}(G)$ is connected (resp. of bounded degree). Note that for $U \subseteq V$, $G{\restriction}U$ and $G \setminus U$ are also Γ-labeled directed graphs. A *connected component* of G is a subgraph $G{\restriction}U$, where $\mathrm{ud}(G){\restriction}U$ is a connected component of $\mathrm{ud}(G)$. For a node $v \in V$ we call the structure (G, v) a *rooted graph*.

Assume now that G is connected and of bounded degree (and thus countable), and let $v_0 \in V$ be a distinguished node. Let $v \in V \setminus \{v_0\}$ and

$d = d_{\mathrm{ud}(G)}(v_0, v) - 1$. The unique connected component of $G \setminus S_{\mathrm{ud}(G)}(d, v_0)$ that contains the node $v \in V$ is denoted by $G(v)$. Furthermore, let

$$\Delta(v) = \{u \in V \mid u \text{ belongs to } G(v),\ d_{\mathrm{ud}(G)}(v_0, u) = d_{\mathrm{ud}(G)}(v_0, v)\}.$$

Two subgraphs $G(u)$ and $G(v)$ are *end-isomorphic* if there exists a (label-preserving graph-) isomorphism from $G(u)$ to $G(v)$, which bijectively maps $\Delta(u)$ to $\Delta(v)$. We say that the rooted graph (G, v_0) is *context-free* if there exist only finitely many $G(v)$ $(v \in V)$ that are pairwise not end-isomorphic. This notion was introduced in [145], where it was shown that if (G, v_0) is context-free, then (G, u) is context-free for every $u \in V$. Hence, in this case we can say that the graph G is context-free. By [145] the context-free graphs are exactly the transition graphs of pushdown automata. Moreover, by a reduction to Rabin's tree theorem [166], Muller and Schupp have shown that every context-free graph has a decidable MSO theory.

4.3.4 Monadic second-order logic over graphs

We begin this section with several known results related to MSO logic over graphs, see [55] for a more comprehensive exposition.

Let us fix a Γ-labeled directed graph $G = (V, (E_a)_{a \in \Gamma})$. Note that MSO logic as introduced in Section 2.2 only allows second-order quantifications over subsets of V. In order to allow also quantifications over sets of edges, we introduce, following [54], an extended representation of graphs. More precisely, we define the relational structure

$$G^{(e)} = (V \cup \bigcup_{a \in \Gamma} E_a,\ (\mathrm{inc}_a)_{a \in \Gamma}),$$

where $\mathrm{inc}_a = \{(e, u, v) \in E_a \times V \times V \mid e = (u, v) \in E_a\}$.[3] We have introduced this extended representation of graphs because of the following important result of Seese, see also [55, Thm. 5.8.10].

Theorem 4.3.8 (cf. [183]). *Let G be a Γ-labeled directed graph such that* $\mathrm{MSOTh}(G^{(e)})$ *is decidable. Then* $\mathrm{ud}(G)$ *has finite tree-width.*[4]

[3] Of course, we assume that $V \cap E_a = \emptyset$ for all $a \in \Gamma$.

[4] In [183] this theorem is stated only for unlabeled undirected graphs. But note that if $\mathrm{MSOTh}(G^{(e)})$ is decidable, then also $\mathrm{MSOTh}(\mathrm{ud}(G)^{(e)})$ is decidable.

This theorem holds even for classes of graphs. The converse of Seese's Theorem is not true: Using an undecidable subset of $\mathbb{N}$ it is easy to construct a tree with an undecidable first-order theory. On the other hand, Courcelle has shown that for every $b \in \mathbb{N}$ the class of *all* graphs of tree-width at most b has a decidable monadic second-order theory [53].

Note that if $\mathrm{MSOTh}(G^{(e)})$ is decidable, then also $\mathrm{MSOTh}(G)$ is decidable. For the reverse implication, restrictions on the graph G are necessary, e.g., for the complete graph on countably many nodes $G = K_{\aleph_0}$, $\mathrm{MSOTh}(G)$ is decidable but $\mathrm{MSOTh}(G^{(e)})$ is not. Courcelle has shown in [54] that for an undirected graph H of bounded degree, if $\mathrm{MSOTh}(H)$ is decidable, then also $\mathrm{MSOTh}(H^{(e)})$ is decidable.[5] Since the decidability of $\mathrm{MSOTh}(G)$ implies the decidability of $\mathrm{MSOTh}(\mathrm{ud}(G))$, Theorem 4.3.8 implies the following result:

Theorem 4.3.9 (cf. [54, 183]). *Let G be a Γ-labeled directed graph of bounded degree. If $\mathrm{MSOTh}(G)$ is decidable, then $\mathrm{ud}(G)$ has finite tree-width.*

Now we are ready to prove the main result of Section 4.3. It can be seen as a converse of Seese's Theorem for graphs with a high degree of symmetry.

Theorem 4.3.10. *Let G be a Γ-labeled connected graph of bounded degree such that $\mathrm{Aut}(G)$ has only finitely many orbits on G. Then the following properties are equivalent:*

(1) $\mathrm{MSOTh}(G)$ *is decidable.*

(2) $\mathrm{ud}(G)$ *has finite tree-width.*

(3) $\mathrm{ud}(G)$ *can be M-triangulated for some constant M.*

(4) G is context-free.

Proof. Since G is connected and of bounded degree, G must be countable. The implication (1) $\Rightarrow$ (2) is stated in Theorem 4.3.9, whereas the implication (4) $\Rightarrow$ (1) is shown in [145].

For (2) $\Rightarrow$ (3) assume that $\mathrm{ud}(G)$ has finite tree-width. Since any automorphism of G is also an automorphism of $\mathrm{ud}(G)$, the group $\mathrm{Aut}(\mathrm{ud}(G))$ has only finitely many orbits on $\mathrm{ud}(G)$. Hence, by Theorem 4.3.7, there exists a strong tree decomposition P of $\mathrm{ud}(G)$ of width $\leq b$ such that for all

[5] The results in [54] are stated for sets of finite graphs, but it is easy to see that the restriction to finite graphs is actually not crucial.

$W \in P$, $\mathrm{diam}_{\mathrm{ud}(G)}(W) \leq c$. Here b and c are fixed constants. Now consider a sequence $\pi = [v_0, v_1, \ldots, v_{m-1}, v_0]$ of nodes $v_i \in V(G)$. Let $W_i \in P$ be such that $v_i \in W_i$. Assume that for all $0 \leq i < m$, either $W_i = W_{i+1}$ or $(W_i, W_{i+1}) \in E(\mathrm{ud}(G)/_P)$ (here and in the following, all subscripts are interpreted modulo m). Note that this implies that $d_{\mathrm{ud}(G)}(v_i, v_{i+1}) \leq 2c+1$ for all $0 \leq i < m$ and that $T_\pi = (\mathrm{ud}(G)/_P)\restriction\{W_0, \ldots, W_{m-1}\}$ is a finite subtree of the tree $\mathrm{ud}(G)/_P$. By induction on m, we will construct a $(2c+1)$-triangulation of π, thus in particular every closed path of G can be $(2c+1)$-triangulated, which shows (3).

The following construction is quite similar to the proof of [19, Thm. 8]. The case that T_π consists of a single node is obvious, since this implies $d_{\mathrm{ud}(G)}(v_i, v_j) \leq c$ for all i, j. Thus, assume that T_π has at least two nodes. Let W be a leaf of T_π and let W' be the unique neighbor of W in T_π. Thus, there exists a subsequence $[v_i, v_{i+1}, \ldots, v_k]$ of π with $k - i \geq 2$, $v_i, v_k \in W'$, and $v_{i+1}, \ldots, v_{k-1} \in W$. Since $\mathrm{diam}_{\mathrm{ud}(G)}(W \cup W') \leq 2c+1$, we can find a $(2c+1)$-triangulation of the sequence $[v_i, v_{i+1}, \ldots, v_k, v_i]$. Moreover, by induction there exists a $(2c+1)$-triangulation of $[v_0, \ldots, v_i, v_k, \ldots, v_{m-1}, v_0]$. By gluing these two triangulations along the side $[v_i, v_k]$ (note that $d_{\mathrm{ud}(G)}(v_i, v_k) \leq 2c+1$), we obtain a $(2c+1)$-triangulation of π.

It remains to prove (3) $\Rightarrow$ (4). If $\mathrm{ud}(G)$ can be M-triangulated for some constant M, then by the argument given in the proof of Theorem 2.9 in [145] it follows that $\mathrm{diam}_{\mathrm{ud}(G)}(\Delta(v)) \leq 3 \cdot M$ for every $v \in V(G)$. Let $\mathcal{O}_1, \ldots, \mathcal{O}_n$ be the orbits of $\mathrm{Aut}(G)$ on G, and choose $v_i \in \mathcal{O}_i$ arbitrarily. Now if $v \in V(G)$ is arbitrary, then $\Delta(v)$ can be mapped injectively into some sphere $S_{\mathrm{ud}(G)}(3 \cdot M, v_i)$ for some $1 \leq i \leq n$. Hence, since every $G(v)$ is uniquely determined by $\Delta(v)$ and the set of those edges that connect nodes from $\Delta(v)$ with nodes from $G \setminus G(v)$, there exist only finitely many $G(v)$ that are pairwise not end-isomorphic.[6] □

Remark 4.3.11. *By [220, Remark 2], we can add the following two equivalent properties to the list of properties in Theorem 4.3.10, see [220] for the definition.*

- $\mathrm{ud}(G)$ *admits a uniformly spanning tree.*
- *All ends of* $\mathrm{ud}(G)$ *have finite diameter.*

[6] The latter argument appears in the proof of Theorem 2.9 in [145] for a vertex-transitive graph, i.e., a graph where $\mathrm{Aut}(G)$ has only one orbit on G.

Remark 4.3.12. *Let $G = (V, E)$ be a graph and let P be the partition of V given by the orbits of* $\mathrm{Aut}(G)$ *on G. In [189] it was shown that if G is context-free, then also $G/_P$ is context-free. Hence, a natural generalization of Theorem 4.3.10 would be the following statement: Let G be a connected graph of bounded degree such that the quotient graph $G/_P$ is context-free. Then G has a decidable MSO-theory if and only if G is context-free. But this is in fact false: Take $\mathbb{N}$ together with the successor relation and add to every number $m = \frac{1}{2}n(n+1)$ ($n \in \mathbb{N}$) a copy m' together with the edge (m, m'), whereas for every other number m we add two copies m' and m'' together with the edges (m, m') and (m, m''). The resulting graph is not context-free, but it has a decidable MSO theory [79] (see also [145]) and $G/_P$ is context-free.*

The next theorem generalizes Theorem 4.3.10 to graphs that are not necessarily connected and countable. For graphs G, G_1, G_2 and a cardinal α, αG denotes the graph that consists of α many disjoint copies of G, and $G_1 + G_2$ denotes the disjoint union of G_1 and G_2.

Theorem 4.3.13. *Let G be a graph of bounded degree but arbitrary cardinality such that* $\mathrm{Aut}(G)$ *has only finitely many orbits on G. Then* $\mathrm{MSOTh}(G)$ *is decidable if and only if there exist finitely many context-free graphs $G_1, \ldots, G_n$ and cardinals $\alpha_1, \ldots, \alpha_n$ such that $G = \alpha_1 G_1 + \cdots + \alpha_n G_n$.*

Proof. First assume that $G = \alpha_1 G_1 + \cdots + \alpha_n G_n$, where G_i is context-free. Thus, $\mathrm{MSOTh}(G_i)$ is decidable. With [191] (see also [89, 203]) we can deduce that also $\mathrm{MSOTh}(G)$ is decidable. Now assume that $\mathrm{MSOTh}(G)$ is decidable, and let G_i, $i \in J$, be the connected components of G. First, note that since every G_i is of bounded degree and connected, all G_i are countable. Furthermore, since $\mathrm{Aut}(G)$ has only finitely many orbits on G, there exist only finitely many pairwise nonisomorphic G_i. Thus, $G = \alpha_1 G_1 + \cdots + \alpha_n G_n$ for cardinals α_i. Moreover, also every $\mathrm{Aut}(G_i)$ has only finitely many orbits on G_i. Thus, in order to prove that every G_i is context-free, it suffices by Theorem 4.3.10 to show that $\mathrm{MSOTh}(G_i)$ is decidable for all $1 \leq i \leq n$.

The set of graphs $\{G_1, \ldots, G_n\}$ can be partitioned into classes $\mathcal{C}_1, \ldots, \mathcal{C}_m$ ($m \leq n$) such that $\mathrm{MSOTh}(G_i) = \mathrm{MSOTh}(G_j)$ if and only if $G_i, G_j \in \mathcal{C}_k$ for some k. Thus, for each class $\mathcal{C}_k$, we can select an MSO sentence ψ_k such that $G_i \models \psi_k$ if and only if $G_i \in \mathcal{C}_k$. Now we can reduce $\mathrm{MSOTh}(G_i)$ to $\mathrm{MSOTh}(G)$ as follows: Assume that $G_i \in \mathcal{C}_k$. Given an MSO sentence ϕ, we construct the following MSO sentence ϕ_k (recall that the existence of a finite

undirected path between two nodes can be expressed in MSO logic):

$$\phi_k \equiv \exists X \left\{ \begin{array}{l} \forall x, y \in X : d_{\mathrm{ud}(G)}(x,y) < \infty \wedge \\ \forall x \in X\ \forall y \notin X : d_{\mathrm{ud}(G)}(x,y) = \infty \wedge \\ \psi_k^X \ \wedge \ \phi^X \end{array} \right\}$$

Here, ψ_k^X denotes the formula that results from ψ_k by relativizing all quantifiers in ψ_k to the set of nodes X, and similarly for ϕ^X. Then $G_i \models \phi$ if and only if $G \models \phi_k$. □

4.4 Cayley-graphs of groups

In this section we will characterize those finitely generated groups such that the corresponding Cayley-graph has a decidable MSO (resp. first-order) theory.

4.4.1 Monadic second-order logic

Recall from Section 2.4 that a finitely generated group $\mathcal{G}$ is context-free, if the set of all words over the generators that represent the unit of $\mathcal{G}$ is a context-free language. In [145] it is shown that a finitely generated group is context-free if and only if its Cayley-graph (with respect to any generating set) is context-free. Together with Theorem 4.3.10 we can deduce the following result:

Corollary 4.4.1. *Let $\mathcal{G}$ be a finitely generated group. The following properties are equivalent:*

(1) $\mathrm{MSOTh}(\mathcal{C}(\mathcal{G}))$ *is decidable.*

(2) $\mathrm{ud}(\mathcal{C}(\mathcal{G}))$ *has finite tree-width.*

(3) $\mathcal{G}$ is context-free.

Remark 4.4.2. *If $\mathcal{G}$ is a finitely generated group such that the corresponding Cayley-graph has finite tree-width, then the corollary above implies that $\mathcal{G}$ is context-free. Hence, by [144], $\mathcal{G}$ is finitely presented. It seems to be hard to deduce this fact in a direct way.*

Further results on the geometric structure of context-free groups can be found in [57, 160, 187].

Remark 4.4.3. *A variant of the equivalence of (2) and (3) in Corollary 4.4.1 is stated in [19]: $\mathcal{G}$ is context-free if and only if there exists a (not necessarily strong) tree-decomposition (T, f) of $\mathcal{C}(\mathcal{G})$ of finite width such that for all $w \in V(T)$, the subgraph $\mathcal{C}(\mathcal{G}){\restriction}f(w)$ is connected (note that in contrast to our notation, in [19] such a tree decomposition is called strong).*

4.4.2 The method of Ferrante and Rackhoff

Before we continue with the investigation of first-order theories of Cayley-graphs of groups, we briefly interrupt with the discussion of a method of Ferrante and Rackhoff for proving upper bounds on the complexity of first-order theories.

Let $\mathcal{A} = (A, (R_i)_{i \in J})$ be a relational structure, where R_i has arity n_i. The *Gaifman-graph* $G_{\mathcal{A}}$ of the structure $\mathcal{A}$ is the following undirected graph:

$$G_{\mathcal{A}} = (A, \{(a, b) \in A \times A \mid \bigvee_{i \in J} \exists (c_1, \ldots, c_{n_i}) \in R_i \; \exists j, k : c_j = a \neq b = c_k\}).$$

We will mainly be interested in restrictions of the structure $\mathcal{A}$ to certain spheres in this graph. To ease notations, we will also write $S_{\mathcal{A}}(r, \widetilde{a})$ for $\mathcal{A}{\restriction}S_{G_{\mathcal{A}}}(r, \widetilde{a})$, i.e., $S_{\mathcal{A}}(r, \widetilde{a})$ is the substructure of $\mathcal{A}$ induced by the r-sphere around the tuple $\widetilde{a}$ in the Gaifman-graph of $\mathcal{A}$. Gaifman's theorem [86] in its strongest form expresses that any first-order formula is logically equivalent to a Boolean combination of "local sentences" (see [78] for a recent account of this result). For our use, the following weaker statement is sufficient, which is an immediate consequence of the main theorem in [86].

Theorem 4.4.4 (cf. [86]). *Let $\widetilde{a} = (a_1, a_2, \ldots, a_k)$ and $\widetilde{b} = (b_1, b_2, \ldots, b_k)$, where $a_i, b_i \in A$, such that*

$$(S_{\mathcal{A}}(7^n, \widetilde{a}), \widetilde{a}) \cong (S_{\mathcal{A}}(7^n, \widetilde{b}), \widetilde{b}).^7$$

Then, for any first-order formula $\varphi(x_1, x_2, \ldots, x_k)$ of quantifier-depth at most n, we have

$$\mathcal{A} \models \varphi(\widetilde{a}) \text{ if and only if } \mathcal{A} \models \varphi(\widetilde{b}).$$

[7] Thus, there exists a bijection $f : S_{\mathcal{A}}(7^n, \widetilde{a}) \to S_{\mathcal{A}}(7^n, \widetilde{b})$, which preserves all relations from $\mathcal{A}$ and such that $f(a_i) = b_i$ for $1 \leq i \leq k$.

A *norm function* on $\mathcal{A}$ is just a function $\lambda : A \to \mathbb{N}$. We write $\mathcal{A} \models \exists x \leq n : \varphi$ in order to express that there exists $a \in A$ such that $\lambda(a) \leq n$ and $\mathcal{A} \models \varphi(a)$, and similarly for $\forall x \leq n : \varphi$. Following Ferrante and Rackoff [83], we define H-bounded structures:

Definition 4.4.5. *Let λ be a norm function on $\mathcal{A}$. Let furthermore $H : \{(j,d) \in \mathbb{N} \times \mathbb{N} \mid j \leq d\} \to \mathbb{N}$ be a function such that the following holds: For any $j \leq d \in \mathbb{N}$, any $\widetilde{a} = (a_1, a_2, \ldots, a_{j-1}) \in A^{j-1}$ with $\lambda(a_i) \leq H(i,d)$, and any $a \in A$, there exists $a_j \in A$ with $\lambda(a_j) \leq H(j,d)$ and*

$$(S_{\mathcal{A}}(7^{d-j}, \widetilde{a}, a), \widetilde{a}, a) \cong (S_{\mathcal{A}}(7^{d-j}, \widetilde{a}, a_j), \widetilde{a}, a_j).$$

Then $\mathcal{A}$ (together with the norm function λ) is called H-bounded.

This is a slight variant of the definition in [83] that suits our needs much better than the original formulation. The following corollary to Theorem 4.4.4 was shown by Ferrante and Rackoff for their version of H-bounded structures.

Corollary 4.4.6 (cf. [83]). *Let $\mathcal{A}$ be a relational structure with norm λ and let $H : \{(j,d) \in \mathbb{N} \times \mathbb{N} \mid j \leq d\} \to \mathbb{N}$ be a function such that $\mathcal{A}$ is H-bounded. Then for any first-order formula $\varphi \equiv Q_1 x_1 \, Q_2 x_2 \cdots Q_d x_d : \psi$ where ψ is quantifier free and $Q_i \in \{\exists, \forall\}$, we have $\mathcal{A} \models \varphi$ if and only if*

$$\mathcal{A} \models Q_1 x_1 \leq H(1,d) \, Q_2 x_2 \leq H(2,d) \cdots Q_d x_d \leq H(d,d) : \psi.$$

Proof. For $j \leq d$, let ψ_j denote the formula $Q_j x_j \, Q_{j+1} x_{j+1} \cdots Q_d x_d : \psi$ and let φ_j stand for the sentence

$$Q_1 x_1 \leq H(1,d) \cdots Q_{j-1} x_{j-1} \leq H(j-1,d) \, \psi_j.$$

Thus, $\varphi_1 = \varphi$. We show that $\mathcal{A} \models \varphi_j$ if and only if $\mathcal{A} \models \varphi_{j+1}$, which then proves the corollary.

Let $\widetilde{a} = (a_1, \ldots, a_{j-1}) \in A^{j-1}$ with $\lambda(a_i) \leq H(i,d)$. First assume $Q_j = \exists$, i.e., $\psi_j \equiv \exists x_j : \psi_{j+1}$. If $\mathcal{A} \models \psi_j(\widetilde{a})$, then there is $a \in A$ with $\mathcal{A} \models \psi_{j+1}(\widetilde{a}, a)$. By our assumption on the norm function λ, we find $a_j \in A$ with $\lambda(a_j) \leq H(j,d)$ and

$$(S_{\mathcal{A}}(7^{d-j}, \widetilde{a}, a), \widetilde{a}, a) \cong (S_{\mathcal{A}}(7^{d-j}, \widetilde{a}, a_j), \widetilde{a}, a_j). \tag{4.1}$$

Since the quantifier depth of ψ_{j+1} is $d - j$, Theorem 4.4.4 implies $\mathcal{A} \models \psi_{j+1}(\widetilde{a}, a_j)$ and therefore $\mathcal{A} \models (\exists x_j \leq H(j,d) \, \psi_{j+1})(\widetilde{a})$. If, conversely, $\mathcal{A} \models (\exists x_j \leq H(j,d) \, \psi_{j+1})(\widetilde{a})$, we have trivially $\mathcal{A} \models \psi_j(\widetilde{a})$.

Assume now that $Q_j = \forall$, i.e., $\psi_j \equiv \forall x_j : \psi_{j+1}$. If $\mathcal{A} \models \psi_j(\widetilde{a})$, then of course also $\mathcal{A} \models (\forall x_j \leq H(j,d) : \psi_{j+1})(\widetilde{a})$. Now assume that $\mathcal{A} \models (\forall x_j \leq H(j,d) : \psi_{j+1})(\widetilde{a})$ and let $a \in A$ be arbitrary. We have to show that $\mathcal{A} \models \psi_{j+1}(\widetilde{a}, a)$. The case $\lambda(a) \leq H(j,d)$ is clear. Thus, assume that $\lambda(a) > H(j,d)$. Then there exists $a_j \in A$ with $\lambda(a_j) \leq H(j,d)$ and (4.1). Since $\lambda(a_j) \leq H(j,d)$, we have $\mathcal{A} \models \psi_{j+1}(\widetilde{a}, a_j)$. Finally, Theorem 4.4.4 implies $\mathcal{A} \models \psi_{j+1}(\widetilde{a}, a)$. □

4.4.3 First-order logic

Let us now consider first-order theories of Cayley-graphs of groups.

Theorem 4.4.7. *Let $\mathcal{G}$ be a finitely generated group such that the word problem of $\mathcal{G}$ belongs to* $\mathrm{ATIME}(a(n), t(n))$. *Then* $\mathrm{FOTh}(\mathcal{C}(\mathcal{G}))$ *belongs to* $\mathrm{ATIME}(n + a(2^{O(n)}), 2^{O(n)} + t(2^{O(n)}))$.

Proof. Choose a finite generating set Γ for $\mathcal{G}$. We want to apply Corollary 4.4.6, which requires to define the norm function λ and the bounding function H. For a group element $a \in \mathcal{G}$ let $\lambda(a) \in \mathbb{N}$ denote the smallest number n such that there exists a word $w \in \Gamma^*$ of length n, representing a. Thus, $\lambda(a)$ is the minimal length of a path from the identity 1 to a in the Cayley-graph $\mathcal{C} = \mathcal{C}(\mathcal{G})$. Next we define the function H by $H(j,d) = H(j-1,d) + 4 \cdot 7^{d-j}$ for $1 \leq j \leq d$ and set $H(0,d) = 0$. Thus, $H(j,d) \in 2^{O(d)}$.

Now let $j \leq d$ and $\widetilde{a} = (a_1, a_2, \ldots, a_{j-1}) \in \mathcal{G}^{j-1}$ with $\lambda(a_i) \leq H(i,d)$. Let furthermore $a \in \mathcal{G}$ with $\lambda(a) > H(j,d)$. The triangle inequality implies that the distance between a and every a_i in $\mathcal{C}$ is larger than $H(j,d) - H(i,d) \geq H(j,d) - H(j-1,d) = 4 \cdot 7^{d-j}$ for $i < j$. Hence, $S_{\mathcal{C}}(7^{d-j}, \widetilde{a}) \cap S_{\mathcal{C}}(7^{d-j}, a) = \emptyset$ and moreover there is no edge in the graph $\mathcal{C}$ between a node in $S_{\mathcal{C}}(7^{d-j}, \widetilde{a})$ and a node in $S_{\mathcal{C}}(7^{d-j}, a)$.

Now assume that $a_j \in \mathcal{G}$ is any group element with $\lambda(a_j) = H(j,d)$. Since $\mathrm{Aut}(\mathcal{C})$ has only one orbit on $\mathcal{C}$, we have $(S_{\mathcal{C}}(7^{d-j}, a), a) \cong (S_{\mathcal{C}}(7^{d-j}, a_j), a_j)$. Moreover, $\lambda(a_j) = H(j,d)$ implies that also $S_{\mathcal{C}}(7^{d-j}, \widetilde{a}) \cap S_{\mathcal{C}}(7^{d-j}, a_j) = \emptyset$, and that there are no edges between these two disjoint spheres. It follows that

$$(S_{\mathcal{C}}(7^{d-j}, \widetilde{a}, a), \widetilde{a}, a) \cong (S_{\mathcal{C}}(7^{d-j}, \widetilde{a}, a_j), \widetilde{a}, a_j).$$

Thus, indeed, the Cayley-graph $\mathcal{C}$ is H-bounded.

Let $\varphi \equiv Q_1 x_1 Q_2 x_2 \cdots Q_d x_d : \psi(x_1, \ldots, x_d)$ be a first-order sentence over the signature of $\mathcal{C}$ with d quantifiers $Q_i \in \{\exists, \forall\}$. Then, by Corollary 4.4.6, $\mathcal{C} \models \varphi$ if and only if

$$\mathcal{C} \models Q_1 x_1 \leq H(1,d)\, Q_2 x_2 \leq H(2,d) \cdots Q_d x_d \leq H(d,d) : \psi(x_1, \ldots, x_d).$$

Since $H(i,d) \in 2^{O(|\varphi|)}$, this implies the statement of the theorem: In order to verify the above statement, we guess (either existentially or universally) every $x_i \leq H(i,d)$. Every quantifier alternation leads to one additional alternation. After having guessed every x_i, all resulting identities in ψ have exponential length. These identities can be verified using the ATIME$(a(n), s(n))$-algorithm for the word problem, which leads to $a(2^{O(n)})$ many additional alternations. The time bound from the theorem follows analogously. □

Remark 4.4.8. *Recall that a problem is called* elementary decidable *if it can be solved in time* $O(2^{\cdot^{\cdot^{\cdot^{2^n}}}})$*, where the height of this tower of exponents is constant. By the previous theorem, if the word problem for* $\mathcal{G}$ *is elementary, then also* FOTh$(\mathcal{C}(\mathcal{G}))$ *is elementary decidable.*

Together with Proposition 4.2.2, Theorem 4.4.7 implies the following corollary.

Corollary 4.4.9. *Let* $\mathcal{G}$ *be a finitely generated group. Then the following properties are equivalent:*

(1) $\exists$FOTh$(\mathcal{C}(\mathcal{G}))$ *is decidable.*

(2) FOTh$(\mathcal{C}(\mathcal{G}))$ *is decidable.*

(3) The word problem of $\mathcal{G}$ *is decidable.*

4.5 Cayley-graphs of monoids

The results from the previous section do not carry over to monoids (see e.g. Proposition 4.2.3). In this section, we will prove several weaker results, that give at least an impression on the borderline between decidability and undecidability.

4.5.1 Monadic second-order logic

Note that Cayley-graphs of finitely generated groups are always of bounded degree. For Cayley-graphs of finitely generated monoids this is in general not true, there may be nodes of even infinite indegree, take for instance the monoid $\mathcal{M}(a, 0 \mid a0 = 0)$. On the other hand, these graphs are still deterministic: A Γ-labeled directed graph $G = (V, (E_a)_{a\in\Gamma})$ is called *deterministic* if for all $v \in V$ and all $a \in \Gamma$ there exists at most one $u \in V$ with $(v, u) \in E_a$.

Lemma 4.5.1. *Let G be a Γ-labeled deterministic graph. Then* $\mathrm{MSOTh}(G)$ *is decidable if and only if* $\mathrm{MSOTh}(G^{(e)})$ *is decidable.*

Proof. For the nontrivial direction note that a set $F \subseteq \bigcup_{a\in\Gamma} E_a$ of edges of $G = (V, (E_a)_{a\in\Gamma})$ can be represented by the tuple of node sets $(U_a)_{a\in\Gamma}$, where $U_a = \{v \in V \mid \exists u \in V : (v, u) \in E_a \cap F\}$. □

Thus, we obtain the following proposition, where the second statement follows together with Seese's Theorem 4.3.8.

Proposition 4.5.2. *Let $\mathcal{M}$ be a finitely generated monoid.*

- $\mathrm{MSOTh}(\mathcal{C}(\mathcal{M}))$ *is decidable if and only if* $\mathrm{MSOTh}(\mathcal{C}(\mathcal{M})^{(e)})$ *is decidable.*
- *If* $\mathrm{MSOTh}(\mathcal{C}(\mathcal{M}))$ *is decidable, then* $\mathrm{ud}(\mathcal{C}(\mathcal{M}))$ *has finite tree-width.*

The next result was shown in [110] for finite monadic presentations. Recall the definition of a left-basic presentation from Section 3.3. A finitely generated presentation (Γ, R) is called *regular* if R can be written as $R = \bigcup_{i=1}^{n} L_i \times R_i$ where both $L_i \subseteq \Gamma^*$ and $R_i \subseteq \Gamma^*$ are regular for $1 \leq i \leq n$.

Proposition 4.5.3. *Let (Γ, R) be a finitely generated presentation, which is terminating, confluent, left-basic, and regular, and let $\mathcal{M} = \mathcal{M}(\Gamma, R)$. Then* $\mathrm{MSOTh}(\mathcal{C}(\mathcal{M}))$ *is decidable.*

Proof. Let (Γ, R) be terminating, confluent, left-basic, and regular. A Γ-labeled directed graph $G = (V, (E_a)_{a\in\Gamma})$ is called *prefix-recognizable* [44], if is isomorphic to a graph of the form $(L, (F_a)_{a\in\Gamma})$, where L is a regular language over some finite alphabet Σ and every F_a is a finite union of relations of the form $\{(uv, uw) \mid u \in U, v \in V, w \in W\}$ where U, V, and W are regular languages over Σ.

It is easy to show that $\mathcal{C}(\mathcal{M}, \Gamma)$ is prefix-recognizable: Since R is terminating and confluent, $\mathcal{M}(\Gamma, R)$ is in one-to-one correspondence with $\mathrm{IRR}(R)$. Moreover, since R is confluent and left-basic, for all $u, v \in \mathrm{IRR}(R)$ and $a \in \Gamma$ we have $ua \overset{*}{\leftrightarrow}_R v$ if and only if $ua \overset{*}{\rightarrow}_R v$ if and only if $ua \overset{*}{\twoheadrightarrow}_R v$, where the last equivalence is stated in Lemma 3.3.3. With [44, Cor. 3.4] it follows that the graph $\mathcal{C}(\mathcal{M}, \Gamma) \cong (\mathrm{IRR}(R), (F_a)_{a\in\Gamma})$ with $F_a = \{(u, v) \in \mathrm{IRR}(R) \times \mathrm{IRR}(R) \mid ua \overset{*}{\twoheadrightarrow}_R v\}$ is prefix-recognizable. Hence, by [44], $\mathrm{MSOTh}(\mathcal{C}(\mathcal{M}))$ is decidable. □

4.5.2 First-order logic

We have already seen that the Cayley-graph of a finitely generated monoid may have an undecidable first-order theory, also in case the word problem is decidable (Proposition 4.2.3). On the decidability side, let us mention that Cayley-graphs of (α, r)-automatic monoids ($\alpha \in \{\ell, r\}$) have decidable first-order theories: These graphs are automatic and hence, by Theorem 2.5.3 they have a decidable first-order theory.

By [20], there exists an automatic structure with a nonelementary first-order theory. This complexity is already realized by Cayley-graphs of automatic monoids:

Theorem 4.5.4. *There exists a fixed length-lexicographic and confluent presentation (Γ, R) with the following properties, where $\mathcal{M} = \mathcal{M}(\Gamma, R)$ and $h : \mathrm{IRR}(R) \to \mathcal{M}$ is the bijection that maps a word $w \in \Gamma^*$ to the monoid element represented by w.*

- *$(\Gamma, \mathrm{IRR}(R), h)$ is an (α, β)-automatic presentation for $\mathcal{M}$ for all $\alpha, \beta \in \{\ell, r\}$.*
- *$\mathrm{FOTh}(\mathcal{C}(\mathcal{M}))$ is not elementary decidable.*

Proof. Let $\Gamma = \{a, b, \bar{a}, \bar{b}, \$_1, \$_2, \$_a\}$ and let the semi-Thue system R over Γ consist of the following rules, where $c \in \{a, b\}$:

$c\,\$_1 \to \$_1 c$	$c\,\$_2 \to \$_2 c$	$c\,\$_a \to \$_a c$
$\bar{c}\,\$_1 \to c$	$\bar{c}\,\$_2 \to \$_1 \bar{c}$	$\bar{a}\,\$_a \to a$
		$\bar{b}\,\$_a \to \$_a \bar{b}$

It is easy to see that R is length-lexicographic and confluent. Arguments similar to those from the proof of Theorem 3.8.2 show that $(\Gamma, \mathrm{IRR}(R), h)$

is an (α, β)-automatic presentation for $\mathcal{M}$. Thus, it remains to show that $\mathrm{FOTh}(\mathcal{C}(\mathcal{M}))$ is not elementary decidable. For this we reduce the first-order theory of finite words over $\{a, b\}$ to $\mathrm{FOTh}(\mathcal{C}(\mathcal{M}))$. The former theory is defined as follows: A word $w = a_1 a_2 \cdots a_n \in \{a, b\}^*$ of length n is identified with the relational structure $S_w = (\{1, \ldots, n\}, <, Q_a)$, where $<$ is the usual order on natural numbers and Q_a is the unary predicate such that $i \in Q_a$ if and only if $a_i = a$. Then the first-order theory of finite words (over $\{a, b\}$) consists of all first-order sentences ϕ over the signature $(<, Q_a)$ such that $S_w \models \phi$ for every word $w \in \{a, b\}^*$. It is known that the first-order theory of finite words is decidable but not elementary, see [85, 142, 169, 200]. A simplified proof can be found in [48, Example 8.1].

For our reduction first notice that $\mathrm{IRR}(R) = \{\$_1, \$_2, \$_a\}^* \{a, b, \bar{a}, \bar{b}\}^*$. Hence the latter set can be identified with the monoid $\mathcal{M}$. Now for $x \in \mathrm{IRR}(R)$ we have $x \in \{\$_1, \$_2, \$_a\}^* \{a, b\}^*$ if and only if $x\$_2\$_1 \neq x\$_1\$_1$ in $\mathcal{C}(\mathcal{M})$. This allows us to represent all words from $\{a, b\}^*$ in $\mathcal{C}(\mathcal{M})$. The fact that a word $w \in \{a, b\}^*$ is represented by infinitely many elements of $\mathcal{C}(\mathcal{M})$, namely by all elements from $\{\$_1, \$_2, \$_a\}^* w$ does not cause any problems, it is only important that every word $w \in \{a, b\}^*$ is represented at least once. In the sequel let us fix $x = vw$ with $v \in \{\$_1, \$_2, \$_a\}^*$ and $w \in \{a, b\}^*$. The set of all positions within the word w is in one-to-one correspondence with the set of all y such that $y\,\$_1 = x$ in $\mathcal{C}(\mathcal{M})$: the latter fact holds if and only if there exist $w_1, w_2 \in \{a, b\}^*$ and $c \in \{a, b\}$ such that $w = w_1 c w_2$ and $y = v w_1 \bar{c} w_2$. Thus, we can quantify over positions of the word w by quantifying in $\mathcal{C}(\mathcal{M})$ over all those y such that $y\,\$_1 = x$ in $\mathcal{C}(\mathcal{M})$. Assume that $y = v w_1 \bar{c} w_2$ and $w = w_1 c w_2$, i.e., y represents the position $|w_1| + 1$ of w. Then $c = a$ if and only if $y\,\$_a = x$ in $\mathcal{C}(\mathcal{M})$, thus we can express whether a position is labeled with the symbol a. It remains to express that a position is smaller than another one. Thus, assume that $y = v w_1 \bar{c} w_2$, $y' = v w_1' \bar{d} w_2'$, $w_1 c w_2 = w_1' d w_2' = w$, and $w_1 \neq w_1'$, i.e., the two positions represented by y and y' are different. Then $|w_1| < |w_1'|$ if and only if there exists a monoid element $z \in \mathcal{M}$ such that $z\,\$_1 = y$ and $z\,\$_2 = y'$ in $\mathcal{C}(\mathcal{M})$.

From the preceding discussion it follows that for every first-order sentence ψ over the signature $(<, Q_a)$ we can construct in polynomial time a first-order formula $\phi(x)$ over the signature of the Cayley-graph $\mathcal{C}(\mathcal{M})$ such that ψ belongs to the first-order theory of finite words if and only if $\mathcal{C}(\mathcal{M}) \models \forall x : \phi(x)$. This proves the theorem. □

Corollary 4.5.5. *There exists a fixed finitely generated monoid $\mathcal{M}$ such that:*

- $\mathcal{M}$ *is simultaneously* (α, β)*-automatic for all* $\alpha, \beta \in \{\ell, r\}$ *and*
- $\mathrm{FOTh}(\mathcal{C}(\mathcal{M}))$ *is not elementary decidable.*

Since the word problem for an automatic group can be solved in time $O(n^2)$, Corollary 4.4.7 implies that the nonelementary lower bound from the previous corollary cannot be realized by an automatic group. In fact, already for *right-cancellative automatic monoids*, we obtain elementary upper bounds:

Theorem 4.5.6. *Let* $\mathcal{M}$ *be a finitely generated right-cancellative* (α, r)*-automatic monoid (for* $\alpha = r$ *or* $\alpha = \ell$*). Then* $\mathrm{FOTh}(\mathcal{C}(\mathcal{M}))$ *can be decided in* $\mathrm{ATIME}(O(n), 2^{2^{2^{O(n)}}})$.

Note that every node in the Cayley-graph of a right-cancellative monoid has bounded indegree (if Γ is the set of generators, then the number of incoming edges is bounded by $|\Gamma|$). Thus, Cayley-graphs of finitely generated right-cancellative monoids have bounded degree. Therefore, Theorem 4.5.6 is an immediate corollary of the following much more general statement.

Theorem 4.5.7. *Let* $\mathcal{A}$ *be an automatic structure such that the Gaifman-graph* $G_{\mathcal{A}}$ *of the structure* $\mathcal{A}$ *has bounded degree. Then* $\mathrm{FOTh}(\mathcal{A})$ *can be decided in* $\mathrm{ATIME}(O(n), 2^{2^{2^{O(n)}}})$.

Proof. Let us fix an r-automatic presentation (Γ, L, h) for $\mathcal{A}$ and let the degree of $G_{\mathcal{A}}$ be bounded by d. By [109] we can assume that $h : L \to \mathcal{A}$ is injective and thus bijective. Let us define the norm $\lambda(a)$ for $a \in \mathcal{A}$ by $\lambda(a) = |h^{-1}(a)|$. By Corollary 4.4.6 it suffices to prove that $\mathcal{A}$ is H-bounded by some function H satisfying $H(j, d) \in 2^{2^{2^{O(d)}}}$ for all $j \leq d \in \mathbb{N}$ (the rest of the argumentation is analogous to the proof of Theorem 4.4.7). First we proof the following:

Claim 1. Let R be an n-ary relation of $\mathcal{A}$. Then the relation

$$R' = \{(u_1, \ldots, u_n) \in L^n \mid (h(u_1), \ldots, h(u_n)) \in R\}$$

has bounded length-difference (see Section 2.5).

Proof of Claim 1. Since (Γ, L, h) is an r-automatic presentation for $\mathcal{A}$, we know that the language

$$\{\nu_r(u_1, \ldots, u_n) \mid (u_1, \ldots, u_n) \in R'\}$$

can be recognized by a finite state automaton A_R, where ν_r is the coding function from Section 2.5. Let m be the number of states of A_R. Then R' must have length-difference bounded by $(n-1) \cdot m$. Because otherwise there would exist a tuple $(u_1, \ldots, u_n) \in R'$ such that (after reordering its components) $|u_i| \leq |u_{i+1}|$ for all $1 \leq i < n$ and $|u_{j+1}| - |u_j| > m$ for at least one j. Then a simple pumping argument shows that the automaton A_R accepts an infinite number of words of the form

$$\nu_r(u_1, \ldots, u_j, u'_{j+1}, \ldots, u'_n)$$

for $u'_{j+1}, \ldots, u'_n \in L$. Since the function h is a bijection between L and $\mathcal{A}$, it follows that the Gaifman-graph $G_\mathcal{A}$ has infinite degree, which is a contradiction.

An immediate consequence of Claim 1 is the next statement.

Claim 2. If there is an edge between a and b in the Gaifman-graph $G_\mathcal{A}$, then $|\lambda(a) - \lambda(b)| \leq \gamma$ for some constant $\gamma \geq 1$.

Claim 3. Let $r \in \mathbb{N}$ and $a \in \mathcal{A}$. Then there exists a finite automaton $A_{r,a}$ with $2^{2^{O(r)}}$ many states such that

$$L(A_{r,a}) = \{u \in L \mid (S_\mathcal{A}(r,a), a) \cong (S_\mathcal{A}(r, h(u)), h(u))\}.$$

Thus, the automaton $A_{r,a}$ accepts a word $u \in L$ if and only if the r-sphere around the element $h(u) \in \mathcal{A}$ represented by u is isomorphic to the r-sphere around a (with a mapped to $h(u)$). For the proof of this claim first notice that since $G_\mathcal{A}$ has bounded degree, $|S_\mathcal{A}(r,a)| \in 2^{O(r)}$. We will use this in order to describe the finite substructure $S_\mathcal{A}(r,a)$ by a formula of size $2^{O(r)}$ over the signature of $\mathcal{A}$:

First, for $0 \leq n \leq d$ (d is the maximal degree in the Gaifman-graph) let the formula $\delta_n(x)$ express that the degree of x in the Gaifman-graph $G_\mathcal{A}$ is exactly n. Thus, $\delta_n(x)$ is a fixed first-order formula over the signature of $\mathcal{A}$. Next take $m = |S_\mathcal{A}(r,a)|$ many variables $x_1, \ldots, x_m$, where x_i represents the element $a_i \in S_\mathcal{A}(r,a)$, ($a_i \neq a_j$ for $i \neq j$) and w.l.o.g. $a = a_1$. Then write down the conjunction of the following formulas, where R is an arbitrary relation of $\mathcal{A}$ and $0 \leq n \leq d$:

- $x_i \neq x_j$ for $i \neq j$,
- $R(x_{i_1}, \ldots, x_{i_n})$ if $(a_{i_1}, \ldots, a_{i_n}) \in R$,

- $\neg R(x_{i_1}, \ldots, x_{i_n})$ if $(a_{i_1}, \ldots, a_{i_n}) \notin R$, and
- $\delta_n(x_i)$ if the degree of a_i in $G_{\mathcal{A}}$ is precisely n.

Finally we quantify the variables $x_2, \ldots, x_m$ existentially. Let $\phi(x_1)$ be the resulting formula. We claim that $\mathcal{A} \models \phi(b)$ if and only if $(S_{\mathcal{A}}(r, a), a) \cong (S_{\mathcal{A}}(r, b), b)$. Only the use of the predicates $\delta_n(x_i)$ needs some explanation. If we would omit these predicates, then $\mathcal{A} \models \phi(b)$ would only express that $(S_{\mathcal{A}}(r, a), a)$ is isomorphic to some induced substructure of $(S_{\mathcal{A}}(r, b), b)$ (with a mapped to b). But by fixing the degree of every x_i we exclude the possibility that there exists $y \in S_{\mathcal{A}}(r, x_1)$ with $y \neq x_i$ for all $1 \leq i \leq m$.[8]

Now the automaton $A_{r,a}$ is obtained by translating the formula $\phi(x_1)$ into an automaton using the standard construction for automatic structures, see e.g. [109]: each of the predicates listed above can be translated into an automaton of fixed size. Since we have $2^{O(r)}$ such predicates, their conjunction can be described by a product automaton of size $2^{2^{O(r)}}$ working on $2^{O(r)}$ tracks (one for each variable x_i). Finally, the existential quantification over the variables $x_2, \ldots, x_m$ means that we have to project this automaton onto the track corresponding to the variable x_1. The resulting automaton is $A_{r,a}$, it still has $2^{2^{O(r)}}$ states and only one track. This proves Claim 3.

Using Claim 2 and 3 it is not difficult to prove that $\mathcal{A}$ is H-bounded for a suitable H. Assume that the size of $A_{r,a}$ is bounded by $2^{2^{cr}}$, where c is some fixed constant. Define the function H by $H(j, d) = H(j-1, d) + 2 \cdot \gamma \cdot 2^{2^{c \cdot 7^{d-j}}}$, where γ is the constant from Claim 2 and $H(0, d)$ is set to 0. Note that $H(d, d) \in 2^{2^{2^{O(d)}}}$. Now let $1 \leq j \leq d$ and $\widetilde{a} = (a_1, a_2, \ldots, a_{j-1}) \in \mathcal{A}^{j-1}$ with $\lambda(a_i) \leq H(i, d)$. Let furthermore $a \in \mathcal{A}$ with $\lambda(a) > H(j, d)$. Thus, $\lambda(a) - \lambda(a_i) > 2 \cdot \gamma \cdot 2^{2^{c \cdot 7^{d-j}}}$ for every $1 \leq i \leq j-1$, which by Claim 2 implies that the distance between a and every a_i in the Gaifman-graph is larger than $2 \cdot 2^{2^{c \cdot 7^{d-j}}}$. Thus, the spheres $S_{\mathcal{A}}(7^{d-j}, \widetilde{a})$ and $S_{\mathcal{A}}(7^{d-j}, a)$ are certainly disjoint and there is no edge between these two spheres.

Now consider the automaton $A_{7^{d-j},a}$ from Claim 3. It has at most $2^{2^{c \cdot 7^{d-j}}}$ states. Since $h(a)$ is accepted by $A_{7^{d-j},a}$, it accepts a word of length larger than $H(j, d) = H(j-1, d) + 2 \cdot \gamma \cdot 2^{2^{c \cdot 7^{d-j}}}$. Thus, a simple pumping argument

[8] The standard solution of this problem is to say that there does not exist $y \notin \{x_1, \ldots, x_m\}$ which is in $G_{\mathcal{A}}$ adjacent to some x_i with $d_{G_{\mathcal{A}}}(x_1, x_i) \leq r - 1$, see e.g. the proof of [204, Corollary 4.9]. But this would introduce a quantifier alternation that we want to avoid.

shows that $A_{7^{d-j},a}$ also accepts a word $w \in L$ with

$$H(j-1,d) + \gamma \cdot 2^{2^{c \cdot 7^{d-j}}} \leq |w| \leq H(j-1,d) + 2 \cdot \gamma \cdot 2^{2^{c \cdot 7^{d-j}}} = H(j,d)$$

(note that $\gamma \geq 1$). Let $a_j = h(w)$. Thus, $\lambda(a_j) \leq H(j,d)$. Moreover, since $\lambda(a_j) \geq H(j-1,d) + \gamma \cdot 2^{2^{c \cdot 7^{d-j}}}$, Claim 2 implies that the distance between a_j and a_i $(1 \leq i < j)$ in the Gaifman-graph is at least $2^{2^{c \cdot 7^{d-j}}}$. Thus, also the spheres $S_{\mathcal{A}}(7^{d-j}, \widetilde{a})$ and $S_{\mathcal{A}}(7^{d-j}, a_j)$ are disjoint and there is no edge between these two spheres. Finally, since by definition of the automaton $A_{7^{d-j},a}$ we have $(S_{\mathcal{A}}(7^{d-j}, a), a) \cong (S_{\mathcal{A}}(7^{d-j}, a_j), a_j)$, we obtain

$$(S_{\mathcal{A}}(7^{d-j}, \widetilde{a}, a), \widetilde{a}, a) \cong (S_{\mathcal{A}}(7^{d-j}, \widetilde{a}, a_j), \widetilde{a}, a_j).$$

Thus, $\mathcal{A}$ is H-bounded, which concludes the proof. □

Remark 4.5.8. *Another application of Theorem 4.5.7 concerns transition graphs of Turing-machines (such a graph has the set of all possible configurations as the set of nodes and an τ-labeled edge from configuration c_1 to configuration c_2, if the machine can move from c_1 to c_2 by applying transition τ). These graphs have bounded degree and are automatic. Thus, their first-order theories are elementary, i.e., "local properties" of Turing-machines are elementary decidable.*

Remark 4.5.9. *The proof of Theorem 4.5.7 shows also two other result. Assume that the premises of Theorem 4.5.7 are satisfied:*

(i) If moreover the Gaifman-graph $G_{\mathcal{A}}$ has polynomial growth, i.e., for every $a \in \mathcal{A}$, the size of the r-sphere $S_{\mathcal{A}}(r,a)$ is bounded by $r^{O(1)}$, then the size of the automaton $A_{r,a}$ from Claim 3 is bounded by $2^{(n^{O(1)})}$. It follows that $\mathrm{FOTh}(\mathcal{A})$ *can be decided in* $\mathrm{ATIME}(O(n), 2^{2^{(n^{O(1)})}})$.

(ii) $\exists\mathrm{FOTh}(\mathcal{A})$ *can be decided in SPACE$(2^{2^{O(n)}})$. To see this, take an existential sentence $\exists x_1 \cdots \exists x_n : \varphi(x_1, \ldots, x_n)$. Then every x_i can be restricted to elements of norm at most $2^{2^{2^{O(n)}}}$. Now, instead of guessing for every x_i a word $t_i \in \Gamma^*$ of length at most $2^{2^{2^{O(n)}}}$, we can first guess for every existentially quantified variable x_i the first symbol. Then we can partially evaluate the atomic predicates in ϕ by running the corresponding automata for one step. After this we forget the first symbols of the x_i and guess the second symbol for every variable x_i and so on. All we need to store beside the polynomially many states of the automata for the atomic predicates in ϕ are the lengths of the partial values for the variables x_i, which needs space $2^{2^{O(n)}}$.*

Remark 4.5.10. *It seems to be open, whether the upper bound in Theorem 4.5.7 is sharp. To the knowledge of the author, the worst case example with respect to complexity is the complete binary tree, i.e., the Cayley-graph of the free monoid $\{a, b\}^*$: it is automatic, of bounded degree, and its first-order theory is complete for* $\mathrm{ATIME}(O(n), 2^{O(n)})$*, see e.g. [48, Example 8.8].*

A precise characterization of the class of finitely generated monoids whose Cayley-graphs have decidable first-order theories (resp. MSO theories) of the kind we have obtained for the group case, remains open. In Section 4.7 we will prove that certain constructions for monoids preserve the decidability of the first-order theory (resp. MSO theory) of the Cayley-graph. Before doing this, we will first investigate a general unfolding operation that works for arbitrary relational structures.

4.6 Unfoldings

In this section we will introduce a construction on relational structures, called *factorized unfolding* (Definition 4.6.4). This construction results from the tree-like unfolding (Definition 4.6.1) by taking the quotient with respect to Mazurkiewicz's trace equivalence (see Section 2.7 for the relevant definitions concerning traces). Our main result about factorized unfoldings states that the first-order theory of a factorized unfolding can be reduced to the first-order theory of the initial structure. Later we will use factorized unfoldings for the study of Cayley-graphs of graph products.

4.6.1 Tree-like unfoldings

In [186] Semenov introduced the following construction, which he attributes to An. A. Muchnik and which generalizes a construction from [191, 201].

Definition 4.6.1. *Let $\mathcal{A} = (A, (R_i)_{1 \leq i \leq \kappa})$ be a relational structure with finitely many relations, where the relation R_i has arity n_i. On the set of finite words A^*, we define the following relations:*

$$\begin{aligned} \widehat{R_i} &= \{(ua_1, ua_2, \ldots, ua_{n_i}) \mid u \in A^*, (a_1, a_2, \ldots, a_{n_i}) \in R_i\} \\ \mathrm{suc} &= \{(u, ua) \mid u \in A^*, a \in A\} \\ \mathrm{cl} &= \{(ua, uaa) \mid u \in A^*, a \in A\} \end{aligned}$$

The relational structure $\widehat{\mathcal{A}} = (A^*, (\widehat{R_i})_{1 \leq i \leq \kappa}, \mathrm{suc}, \mathrm{cl})$ *is is called the* tree-like unfolding *of* $\mathcal{A}$.[9]

One can think of the structure $\widehat{\mathcal{A}}$ as a tree (A^*, suc) together with some additional relations. Any tuple of elements of A^* that appears in one of the additional relations is "local": the distance between any two entries in the tree (A^*, suc) is at most 2. The term tree-like unfolding comes from the fact that $\widehat{\mathcal{A}}$ is an extension of the tree (A^*, suc).

In [186], Semenov also sketched a proof of the following result, which he attributes again to An. A. Muchnik. A complete proof was given by Walukiewicz [219].

Theorem 4.6.2 (cf. [219]). $\mathrm{MSOTh}(\widehat{\mathcal{A}})$ *can be reduced to* $\mathrm{MSOTh}(\mathcal{A})$.

The relations of the tree-like unfolding are instances of a more general construction, which will be crucial for our notion of factorized unfoldings: Let φ be a first-order formula over the signature of $\mathcal{A}$ with $\sum_{i=1}^{n} k_i$ free variables, where $k_i \in \mathbb{N}$ ($k_i = 0$ is allowed). For words $u_i = a_{i,1}a_{i,2} \cdots a_{i,k_i}$ ($a_{i,j} \in A$) of length k_i ($1 \leq i \leq n$) we write $\mathcal{A} \models \varphi(u_1, u_2, \ldots, u_n)$ if

$$\mathcal{A} \models \varphi(a_{1,1}, \ldots, a_{1,k_1}, a_{2,1}, \ldots, a_{2,k_2}, \ldots, a_{n,1}, \ldots, a_{n,k_n}).$$

An n-ary relation R over A^* is *k-suffix definable* in $\mathcal{A}$ if there are $k_1, \ldots, k_n \leq k$ and a first-order formula φ over the signature of $\mathcal{A}$ with $\sum_{i=1}^{n} k_i$ free variables such that

$$R = \{(uu_1,\ uu_2,\ \ldots, uu_n) \mid u, u_i \in A^*, |u_i| = k_i, \mathcal{A} \models \varphi(u_1, u_2, \ldots, u_n)\}.$$

Note that a formula φ with m free variables can define many different k-suffix definable relations, namely one for each partition of the number m.

Obviously, all relations of $\widehat{\mathcal{A}}$ are 2-suffix definable in $\mathcal{A}$. On the other hand, there exist k-suffix definable relations such that adding them to $\widehat{\mathcal{A}}$ makes Theorem 4.6.2 fail: To see this, let

$$\mathrm{cp} = \{(ua, uba) \mid u \in A^*, a, b \in A\},$$

which is 2-suffix definable in $\mathcal{A}$.[10] Recall that $\preceq$ denotes the prefix order on A^*, thus it is the reflexive transitive closure of the relation suc from $\widehat{\mathcal{A}}$ and therefore MSO definable in $\widehat{\mathcal{A}}$.

[9] "cl" stands for "clone".

[10] "cp" stands for "copy".

Proposition 4.6.3. *Let* $S = \{(n, n+1) \mid i \in \mathbb{N}\}$ *be the successor relation on* $\mathbb{N}$. *Then* $\mathrm{MSOTh}(\mathbb{N}, S)$ *is decidable but* $\mathrm{FOTh}(\mathbb{N}^*, \widehat{S}, \preceq, \mathrm{cp})$ *is undecidable.*

Proof. The decidability of $\mathrm{MSOTh}(\mathbb{N}, S)$ was shown by Büchi [32]. For the undecidability of $\mathrm{FOTh}(\mathbb{N}^*, \widehat{S}, \preceq, \mathrm{cp})$ recall that it is undecidable whether a given two-counter machine (with zero-tests), started with empty counters, finally terminates. Thus, let us fix a two-counter machine $\mathcal{CM}$ with initial state q_0 and final state $q_f \neq q_0$. We will construct a first-order sentence $\phi_{\mathcal{CM}}$ such that $(\mathbb{N}^*, \widehat{S}, \preceq, \mathrm{cp}) \models \phi_{\mathcal{CM}}$ if and only if $\mathcal{CM}$, started in the configuration $(q_0, 0, 0)$ terminates. We can assume that the state space Q of $\mathcal{CM}$ is $\{1, \ldots, \lambda\}$. Then a computation of $\mathcal{CM}$ starting in $(q_0, 0, 0)$ can be encoded by a sequence of the form $q_0\, m_0\, n_0\, 0\, q_1\, m_1\, n_1\, 0 \cdots q_k\, m_k\, n_k \in \mathbb{N}^*$ such that $q_i \in Q$, $m_i, n_i \geq 1$, $m_0 = n_0 = 1$, and $(q_{i+1}, m_{i+1} - 1, n_{i+1} - 1)$ is a successor configuration of $(q_i, m_i - 1, n_i - 1)$ in $\mathcal{CM}$.

First, note that the relation suc from the tree-like unfolding is first-order definable in $(\mathbb{N}^*, \preceq)$. For a fixed $n \in \mathbb{N}$, it is also easy to write down a formula $\psi_n(x)$ such that $(\mathbb{N}^*, \widehat{S}, \preceq, \mathrm{cp}) \models \psi_n(w)$ if and only if $w = v\, n$ for some $v \in \mathbb{N}^*$: We start with $\psi_0(x) \equiv \neg\exists y : \widehat{S}(y, x) \wedge \exists z : \mathrm{suc}(z, x)$ and define inductively $\psi_n(x) \equiv \exists y : \psi_{n-1}(y) \wedge \widehat{S}(y, x)$. Next, using the prefix relation $\preceq$ and the formulas ψ_n for $0 \leq n \leq \lambda$, we can construct a first-order formula $\phi_0(x)$ such that $(\mathbb{N}^*, \widehat{S}, \preceq, \mathrm{cp}) \models \phi_0(w)$ if and only if $w \in \mathbb{N}^*$ has the form $q_0\, m_0\, n_0\, 0\, q_1\, m_1\, n_1\, 0 \cdots q_k\, m_k\, n_k$ with $q_i \in Q$, $m_i, n_i \geq 1$, $m_0 = n_0 = 1$, and $q_k = q_f$. Furthermore, we claim that there exists a formula $\phi_1(x)$ such that for every $w \in \mathbb{N}^*$ we have $(\mathbb{N}^*, \widehat{S}, \preceq, \mathrm{cp}) \models \phi_1(w)$ if and only if for every prefix $v\, p\, k\, \ell\, 0\, q\, m\, n \preceq w$ with $p, k, \ell, q, m, n \geq 1$, the configuration $(q, m-1, n-1)$ is a successor configuration of $(p, k-1, \ell-1)$, i.e., one of the finitely many transition rules of $\mathcal{CM}$ transforms the configuration $(p, k-1, \ell-1)$ into $(q, m-1, n-1)$. Let us consider one such transition rule, saying, e.g., that if $\mathcal{CM}$ is in state $q_1 \in Q$, then $\mathcal{CM}$ can move into state $q_2 \in Q$ and add 1 to the first counter. It suffices to construct a formula $\theta(x)$ such that for every word w of the form $v\, p\, m\, n\, 0\, q\, k\, \ell$ with $v \in \mathbb{N}^*$, we have $(\mathbb{N}^*, \widehat{S}, \preceq, \mathrm{cp}) \models \theta(w)$ if and only if $p = q_1$, $q = q_2$, $\ell = n$, and $k = m + 1$.

It is easy to express $p = q_1$ and $q = q_2$. Thus, it remains to express $k = m + 1$ (and $\ell = n$, which can be done analogously): This is the case if and only if there are $x_i, y_i \in \mathbb{N}^*$ for $0 \leq i \leq 4$ such that

$$\mathrm{suc}(x_4, w) \wedge y_4 = x_4 \wedge \bigwedge_{i=1}^{4} \Big(\mathrm{suc}(x_{i-1}, x_i) \wedge \mathrm{cp}(y_{i-1}, y_i) \wedge \widehat{S}(x_0, y_0)\Big).$$

Other transitions can be dealt with similarly. Hence, we have $(\mathbb{N}^*, \widehat{S}, \preceq, \mathrm{cp}) \models \exists x : \phi_0(x) \wedge \phi_1(x)$ if and only if $\mathcal{CM}$ reaches the final state q_f from the initial configuration $(q_0, 0, 0)$. This proves the theorem. □

Since $\preceq$ is MSO-definable in the presence of suc, the previous proposition implies that $\mathrm{MSOTh}(\mathbb{N}^*, \widehat{S}, \mathrm{suc}, \mathrm{cp})$ is undecidable. Thus, the presence of the relation cp makes Walukiewicz's result fail.

Recall that the underlying set of the tree-like unfolding of a structure $\mathcal{A}$ is the set of all finite words over the carrier set of $\mathcal{A}$. In factorized unfoldings that we introduce next, this underlying set consists of Mazurkiewicz traces.

4.6.2 Factorized unfoldings

Definition 4.6.4. *Let $\mathcal{A}$ be a relational structure with carrier set A. Let furthermore*

- *$I \subseteq A \times A$ be an independence relation that is first-order definable in $\mathcal{A}$,*
- *$\eta : \mathbb{M}(A, I) \to S$ be a monoid homomorphism into some finite monoid S such that $\eta^{-1}(q) \cap A$ is first-order definable in $\mathcal{A}$ for every $q \in S$, and*
- *R_i be a k_i-suffix definable relation in $\mathcal{A}$ for $1 \leq i \leq \kappa$.*

Then the structure $\mathcal{B} = (\mathbb{M}(A, I), (\eta^{-1}(q))_{q \in S}, (R_i/_I)_{1 \leq i \leq \kappa})$ is a factorized unfolding of $\mathcal{A}$*, it is also called the* factorized unfolding of $\mathcal{A}$ corresponding to I, η, and $(R_i)_{1 \leq i \leq \kappa}$.

Note that in contrast to the tree-like unfolding there are many different factorized unfoldings of $\mathcal{A}$.

The notion of a factorized unfolding is a proper generalization of the tree-like unfolding, also in case $I = \emptyset$ in Definition 4.6.4: By Proposition 4.6.3, the relation cp cannot be defined in the tree-like unfolding $\widehat{\mathcal{A}}$, but since it is 2-suffix definable it may be part of a factorized unfolding. On the other hand, for the relations $\eta^{-1}(q)$ in the above definition we have the following:

Lemma 4.6.5. *Let* $\mathrm{MSOTh}(\mathcal{A})$ *be decidable and $\eta : A^* \to S$ be a monoid morphism into a finite monoid S such that $\eta^{-1}(q) \cap A$ is MSO-definable in $\mathcal{A}$ for every $q \in S$. Then also* $\mathrm{MSOTh}(\widehat{\mathcal{A}}, (\eta^{-1}(q))_{q \in S})$ *is decidable.*

Proof. Since $P_q := \eta^{-1}(q) \cap A$ is MSO-definable in $\mathcal{A}$, the structure $\mathcal{B} = (\mathcal{A}, (P_q)_{q \in S})$ has a decidable MSO theory. Hence, by Theorem 4.6.2, also $\mathrm{MSOTh}(\widehat{\mathcal{B}})$ is decidable and it suffices to prove that $\eta^{-1}(q) \subseteq A^*$ is MSO-definable in $\widehat{B}$. This can be shown similar to Büchi's proof that recognizable sets are MSO-definable [32]: Note that given $x \in A^*$, we can express in MSO logic over $\widehat{\mathcal{B}}$ that $X \subseteq A^*$ is the set of all words on the unique path from ε to x in the tree (A^*, suc). Then $\eta(x) = q$ if and only if there is a partition $X = \bigcup_{s \in S} X_s$ such that $x \in X_q$, $\epsilon \in X_1$ (where 1 is the neutral element of S), and for all $(y, z) \in \mathrm{suc} \cap (X \times X)$: if $y \in X_{s_1}$, $z \in X_{s_2}$, and $s \in S$ such that $z \in \widehat{P_s}$, then $s_1 s = s_2$ in the monoid S. □

The following theorem is the main result of this section.

Theorem 4.6.6. *Let $\mathcal{A}$ be a relational structure and let $\mathcal{B}$ be the factorized unfolding of $\mathcal{A}$ corresponding to I, η, and $(R_i)_{1 \leq i \leq \kappa}$, where $\{I(a) \mid a \in A\} \subseteq 2^A$ is finite. Then any first-order sentence φ of quantifier alternation depth d over the signature of $\mathcal{B}$ can be transformed effectively into a sentence θ of quantifier alternation depth $d + O(1)$ and size $2^{2^{O(|\varphi|)}}$ over the signature of $\mathcal{A}$ such that $\mathcal{B} \models \varphi$ if and only if $\mathcal{A} \models \theta$.*

Remark 4.6.7. *The proof of Theorem 4.6.6 will show that not only the size of θ is bounded doubly exponential in the size of φ, but also the time needed to construct θ from φ is bounded doubly exponential in $|\varphi|$.*

Corollary 4.6.8. *Let $\mathcal{A}$ be a relational structure with a decidable first-order theory. Let $\mathcal{B}$ be the factorized unfolding of $\mathcal{A}$ corresponding to I, η, and $(R_i)_{1 \leq i \leq \kappa}$, where $\{I(a) \mid a \in A\} \subseteq 2^A$ is finite. Then $\mathrm{FOTh}(\mathcal{B})$ is decidable.*

Before we prove Theorem 4.6.6, let us first discuss several related results. The structure $(\mathbb{N}^*, \widehat{S}, \preceq, \mathrm{cp})$ from Proposition 4.6.3 has an undecidable first-order theory. Thus, allowing the relation $\preceq/_I$, which is the prefix order on traces, in factorized unfoldings would make Theorem 4.6.6 fail (already for $I = \emptyset$).

In Theorem 4.6.6, we also assume that there are only finitely many different sets $I(a)$. The reason is again that otherwise the result would fail: Let $V = \{(m, n) \in \mathbb{N}^2 \mid m \leq n\}$ and $E = \{(\ell, m, n) \in \mathbb{N}^3 \mid \ell, m \leq n\}$. On $A = V \cup E$, define the relation R by

$$R = \{((m, n),\ (\ell, m, n)) \mid \ell, m \leq n\} \cup \{((\ell, n),\ (\ell, m, n)) \mid \ell, m \leq n\}.$$

Thus, $\mathrm{dom}(R) = V$ and $\mathrm{ran}(R) = E$. Furthermore, let I be the set of pairs of distinct elements from $V \cup E$ that agree on their last component. Then there are infinitely many sets $I(a)$, but any of these sets is finite. We will consider the structure $\mathcal{A} = (A, R, I)$. The decidability of $\mathrm{FOTh}(\mathbb{N}, \leq)$ implies the decidability of $\mathrm{FOTh}(\mathcal{A})$.

Theorem 4.6.9. *Let* $\mathcal{B} = (\mathbb{M}(A, I), \mathrm{cl}/_I, \mathrm{suc}/_I, \widehat{R}/_I)$, *which is a factorized unfolding of* $\mathcal{A}$. *Then* $\mathrm{FOTh}(\mathcal{B})$ *is undecidable.*

Proof. We will reduce the first-order theory of all finite directed graphs (which is undecidable by [209]) to the first-order theory of $\mathcal{B}$.

The idea is to represent a finite graph by the finite (A, I)-clique $\max(s)$ of a trace $s \in \mathbb{M}(A, I)$, which by the definition of I is a subset of

$$V \cap (\mathbb{N} \times \{n\}) \ \cup \ E \cap (\mathbb{N}^2 \times \{n\})$$

for some $n \in \mathbb{N}$. Those elements from V (resp. E) in $\max(s)$ represent the nodes (resp. edges) of G. To represent the set $\max(s)$ in $\mathcal{B}$, we use the $\mathrm{cl}/_I$-relation. More precisely, for $s \in \mathbb{M}(A, I)$ let $G(s)$ denote the set of traces t such that $(s, t) \in \mathrm{cl}/_I$ in $\mathcal{B}$. In other words, $G(s)$ is the set of traces sa (with $a \in A$) such that $a \in \max(s)$. By f_s, we denote the bijection from $G(s)$ to $\max(s)$ given by $sa \mapsto a$ (this is well-defined, since $sa = sb$ implies $a = b$). Now let $t \in G(s)$. Then $f_s(t) \in V$ if and only if there is $u \in \mathbb{M}(A, I)$ with $(s, u) \in \mathrm{suc}/_I$ and $(t, u) \in \widehat{R}/_I$. Similarly, $f_s(t) \in E$ if and only if there is $u \in \mathbb{M}(A, I)$ with $(s, u) \in \mathrm{suc}$ and $(u, t) \in \widehat{R}/_I$. Now let $v, e \in G(s)$ with $f_s(v) \in V$ and $f_s(e) \in E$. Then $(f_s(v), f_s(e)) \in R$ if and only if $(v, e) \in \widehat{R}/_I$. Thus, we can write a formula $\mathrm{graph}(x)$ with one free variable x such that $\mathcal{B} \models \mathrm{graph}(s)$ if and only if $\max(s)$ is a directed graph, i.e., any element of $\max(s) \cap E$ is adjacent with two elements of $\max(s) \cap V$.

Now let φ be a sentence over the signature of directed graphs. Using the ideas explained above, we can construct a formula $\varphi'(x)$ with one free variable such that for any trace $s \in \mathbb{M}(A, I)$ satisfying $\mathrm{graph}(s)$, the graph of maximal elements of s satisfies φ if and only if $\mathcal{B} \models \varphi'(s)$. Thus, φ is true in all finite graphs (i.e., belongs to the theory of all finite graphs) if and only if $\mathcal{B} \models \forall x : \mathrm{graph}(x) \Rightarrow \varphi'(x)$. □

In order to prove the undecidability results in Theorem 4.6.3 and 4.6.9 we used infinite structures. Infinity is needed as the next theorem shows. Recall the definition of an automatic structure from Section 2.5.

Theorem 4.6.10. *Let $\mathcal{A}$ be a finite relational structure with universe A, and let $\mathcal{B} = (\mathbb{M}(A,I), (\eta^{-1}(q))_{q\in S}, (R_i/_I)_{1\le i\le\kappa})$ be any factorized unfolding of $\mathcal{A}$. Then the structure $(\mathcal{B}, \preceq/_I)$ is automatic and has therefore a decidable first-order theory.*

Proof. The free monoid $\mathcal{F}(A,I)^*$ generated by the set of (A,I)-cliques maps naturally onto $\mathbb{M}(A,I)$, let h denote the canonical homomorphism defined by $h(C) = [C]$. Let $\mathrm{FNF} \subseteq \mathcal{F}(A,I)^*$ denote the set of Foata normal forms. Then a word $C_1C_2\cdots C_n$ over $\mathcal{F}(A,I)$ belongs to FNF if and only if for every $1 \le i < n$ and every $a \in C_{i+1}$, there is $b \in C_i$ with $(a,b) \notin I$. Since A is finite, the set FNF is recognizable. Moreover, h maps FNF bijectively to $\mathbb{M}(A,I)$. We show that $(\mathcal{F}(A,I), \mathrm{FNF}, h)$ is an r-automatic presentation for $(\mathcal{B}, \preceq/_I)$.

Reading the Foata normal form of a trace t, a finite automaton with state space S can easily check whether $\eta(t) = q \in S$. The automaticity of the relations $R_i/_I$, $1 \le i \le \kappa$, can be reduced to the automaticity of the relations $\mathrm{suc}_a = \{(t, ta) \mid t \in \mathbb{M}(A,I)\}$, $a \in A$, as follows: Since A is finite and R_i is k-suffix definable for some k, we can write $R_i/_I$ as $R_i/_I = \bigcup_{(s_1,\ldots,s_n)\in F_i}\{(ts_1,\ldots,ts_n) \mid t \in \mathbb{M}(A,I)\}$, where $F_i \subseteq \mathbb{M}(A,I)^n$ is a *finite* relation. Now, for $s = [a_1a_2\cdots a_m]_I \in \mathbb{M}(A,I)$, $a_k \in A$, define

$$\mathrm{suc}_s(x_0, x_m) \;\equiv\; \exists x_1 \cdots \exists x_{m-1} \left\{ \bigwedge_{j=1}^{m} \mathrm{suc}_{a_j}(x_{j-1}, x_j) \right\}.$$

Then $(y_1,\ldots,y_n) \in R_i$ if and only if

$$\bigvee_{(s_1,\ldots,s_n)\in F_i} \exists x \left\{ \bigwedge_{1\le j\le n} \mathrm{suc}_{s_j}(x, y_j) \right\}.$$

Finally, we can use the closure of automatic relations under first-order definitions (Theorem 2.5.2).

Thus, it only remains to show that the relations suc_a, $a \in A$, and the prefix order $\preceq/_I$ are automatic: Let $v = B_1B_2\cdots B_m$ and $w = C_1C_2\cdots C_n$ be two Foata normal forms. Then for $a \in A$ we have $h(v)a = h(w)$ in $\mathbb{M}(A,I)$ if and only if $n \in \{m, m+1\}$ and there is j with $1 \le j \le n$ such that for all $1 \le i \le m$, we have

(1) if $i \ne j$, then $B_i = C_i$,

(2) $C_j = B_j \dot\cup \{a\}$ if $j \leq m$, and $C_n = \{a\}$ if $n = m+1 = j$, and

(3) if $i > j$, then $\{a\} \times B_i \subseteq I$.

Thus, reading v and w synchronously, a finite automaton can check whether $(h(v), h(w)) \in \mathrm{suc}_a$. Finally, we claim that the trace $h(v)$ is a prefix of $h(w)$ if and only if

(1) $m \leq n$,

(2) $B_i \subseteq C_i$ for $1 \leq i \leq m$, and

(3) for $i \leq j \leq m$, we have $(C_i \setminus B_i) \times B_j \subseteq I$.

If these three conditions hold, then in $\mathbb{M}(A, I)$ we have

$$\begin{aligned} h(w) &= [C_1]\,[C_2] \cdots [C_n] \\ &= [B_1]\,[C_1 \setminus B_1]\,[B_2]\,[C_2 \setminus B_2] \cdots [B_m]\,[C_m \setminus B_m][C_{m+1}] \cdots [C_n] \\ &= [B_1]\,[B_2] \cdots [B_m][C_1 \setminus B_1] \cdots [C_m \setminus B_m]\,[C_{m+1}] \cdots [C_n] \\ &= h(v)[C_1 \setminus B_1] \cdots [C_m \setminus B_m]\,[C_{m+1}] \cdots [C_n]. \end{aligned}$$

On the other hand, if $h(w) = [C_1]\,[C_2] \cdots [C_n] = [B_1]\,[B_2] \cdots [B_m]s = h(v)s$ for some trace s, then

$$\begin{aligned} C_1 = \min(h(w)) &= \min(h(v)) \cup \{a \in \min(s) \mid (a, h(v)) \in I\} \\ &= B_1 \cup \{a \in \min(s) \mid (a, h(v)) \in I\}. \end{aligned}$$

Since $(a, h(v)) \in I$ implies $a \notin B_1$, we have

$$C_1 \setminus B_1 = \{a \in \min(s) \mid (a, h(v)) \in I\}.$$

Thus, $(C_1 \setminus B_1) \times B_j \subseteq I$ for $j \geq 1$. Moreover $[C_2] \cdots [C_n] = [B_2] \cdots [B_m]s'$ for some s'. We can conclude inductively.

A finite automaton, reading v and w synchronously, can easily check (1) and (2) above. In order to check (3), it has to remember $\bigcup_{i \leq j} C_i \setminus B_i \subseteq A$ when j goes from 1 to m, which is possible since A is finite. Thus, the prefix order $\preceq /_I$ can be checked by a finite automaton. □

Remark 4.6.11. *We can also say something on the complexity of the decision procedure for the first-order theory of $(\mathcal{B}, \preceq /_I)$ in the previous theorem: Suppose that any two distinct letters from A are independent. Then*

one can reduce $\mathrm{FOTh}(\mathcal{B}, \preceq)$ *to Presburger's Arithmetic, which is decidable in* $\mathrm{ATIME}(O(n), 2^{2^{O(n)}})$ *[17] and hence elementary. On the other hand, the theory* $\mathrm{FOTh}(\{a,b\}^*, \mathrm{suc}_a, \mathrm{suc}_b, \preceq)$ *is not elementary decidable, see [48, Example 8.3].*

Note that Theorem 4.6.10 implies in particular that the prefix order on finite traces over a finite independence alphabet (A, I) has a decidable first-order theory.

Proof of Theorem 4.6.6

Let us fix the factorized unfolding $\mathcal{B} = (\mathbb{M}(A,I), (\eta^{-1}(q))_{q\in S}, (R_i/_I)_{1\leq i\leq \kappa})$ of $\mathcal{A}$. We will reduce the first-order theory of $\mathcal{B}$ to the first-order theory of $\mathcal{A}$. We first show how to reduce a "local" sentence over the signature of $\mathcal{B}$ to an equivalent sentence over the signature of $\mathcal{A}$. In the sequel, we will then show that every sentence over the signature of $\mathcal{B}$ can be reduced to such a local sentence. Assume that R_i is k_i-suffix definable in $\mathcal{A}$ and let $k = \max\{k_i \mid 1 \leq i \leq \kappa\}$. On $\mathbb{M}(A,I)$ we define a norm function $\lambda : \mathbb{M}(A,I) \to \mathbb{N}$ by $\lambda(t) = |t|$. According to Section 4.4.2, $\exists x \leq n : \phi(x)$ is an abbreviation for $\exists x : |x| \leq n \ \wedge \ \phi(x)$.

Proposition 4.6.12. *Let* $\psi(x_1, \ldots, x_d)$ *be a Boolean formula over the signature of* $\mathcal{B}$. *Let* $n_1, \ldots, n_d \in \mathbb{N}$, $n = \max\{n_1, \ldots, n_d\}$, *and* $Q_1, \ldots, Q_d \in \{\exists, \forall\}$. *Then we can effectively construct a sentence* θ *over the signature of* $\mathcal{A}$ *such that*

$$\mathcal{B} \models Q_1 x_1 \leq n_1 \, Q_2 x_2 \leq n_2 \cdots Q_d x_d \leq n_d : \psi(x_1, \ldots, x_d) \tag{4.2}$$

if and only if $\mathcal{A} \models \theta$. *Moreover,* θ *has quantifier alternation depth* $d + O(1)$ *and size bounded by* $n^d \cdot |\psi| \cdot 2^{O(n)}$.

Proof. We will encode a trace $x \in \mathbb{M}(A,I)$ with $|x| \leq n$ by a sequence $y_1 y_2 \cdots y_m$ of first-order variables $y_i \in A$ of length $m \leq n$, with the meaning that $x = [y_1 y_2 \cdots y_m]_I$. First, for every $m \leq n$, we have to construct a first-order formula in $2m$ free variables over the signature of $\mathcal{A}$, which expresses that $[y_1 y_2 \cdots y_m]_I = [z_1 z_2 \cdots z_m]_I$ in $\mathbb{M}(A,I)$. This can be done inductively as follows: If $m = 0$, then this formula is the truth value true. If $m > 0$, then $[y_1 y_2 \cdots y_m]_I = [z_1 z_2 \cdots z_m]_I$ if and only if

$$\bigvee_{i=1}^{m} \left(y_1 = z_i \ \wedge \ \bigwedge_{j=1}^{i-1} (z_i, z_j) \in I \ \wedge \ [y_2 \cdots y_m]_I = [z_1 \cdots z_{i-1} z_{i+1} \cdots z_m]_I \right)$$

(recall that by assumption the independence relation I can be defined by a fixed first-order formula over the signature of $\mathcal{A}$). The above recursive definition would lead to a formula of exponential size for $[y_1 y_2 \cdots y_m]_I = [z_1 z_2 \cdots z_m]_I$. Using a trick from Ferrante [82, Lem. 2] we can be a little bit more space economically: The above formula is equivalent to

$$\exists u_1 \cdots \exists u_{m-1} \left\{ \begin{array}{l} [y_2 \cdots y_m]_I = [u_1 \cdots u_{m-1}]_I \wedge \\ \displaystyle\bigvee_{i=1}^{m} \left(\begin{array}{l} y_1 = z_i \ \wedge \displaystyle\bigwedge_{j=1}^{i-1} (z_j = u_j \ \wedge \ (z_i, u_j) \in I) \wedge \\ \displaystyle\bigwedge_{j=i+1}^{m} z_j = u_{j-1} \end{array} \right) \end{array} \right\}.$$

Let s_m be the size of this formula with $2m$ free variables. Then s_m is bounded by $s_{m-1} + O(m^2)$. Thus, $s_n \in O(n^3)$. Moreover the quantifier alternation depth in the above formula is 0, since we only use existential quantifiers.

Now a bounded existential quantification $\exists x_i \leq n_i$ in (4.2) can be replaced by $\bigvee_{i=0}^{n_i} \exists y_1 \cdots \exists y_i$, where x is represented by the sequence $y_1 \cdots y_i$, and similarly for a universal quantifier. Since there are only d quantifiers in (4.2), these replacements increase the size of the formula at most by a factor n^d. Furthermore, the quantifier alternation depth is unchanged.

Next, consider an atomic formula $R/_I(x_1, \ldots, x_r)$ in ψ, where R is one of the k-suffix definable relations R_i $(1 \leq i \leq \kappa)$. Since R is k-suffix definable, we can assume that

$$R = \{(uu_1,\ uu_2,\ \ldots, uu_r) \mid u, u_i \in A^*, |u_i| = \ell_i, \mathcal{A} \models \phi(u_1, u_2, \ldots, u_r)\}$$

for some $\ell_i \leq k$, where ϕ is a fixed first-order formula over the signature of $\mathcal{A}$. Assume that the trace $x_i \in \mathbb{M}(A, I)$ is represented by the sequence $y_{i,1} \cdots y_{i,m_i}$ $(m_i \leq n)$. If for some $1 \leq i \leq r$, we have $m_i < \ell_i$, then we can replace $R/_I(x_1, \ldots, x_r)$ by the truth value false. The same can be done if $m_i - \ell_i \neq m_j - \ell_j$ for two different i, j. Thus, assume that $m_i - \ell_i = \ell \geq 0$ for all $1 \leq i \leq r$. Then we can replace $R/_I(x_1, \ldots, x_r)$ by the formula

$$\exists_{\substack{1 \leq i \leq r \\ 1 \leq j \leq \ell_i}} z_{i,j}\, \exists z_1 \cdots \exists z_\ell \left\{ \begin{array}{l} \displaystyle\bigwedge_{1 \leq i \leq r} [y_{i,1} \cdots y_{i,m_i}]_I = [z_1 \cdots z_\ell z_{i,1} \cdots z_{i,\ell_i}]_I \\ \wedge\ \phi(z_{1,1}, \ldots, z_{1,\ell_1}, \ldots, z_{r,1}, \ldots, z_{r,\ell_r}) \end{array} \right\},$$

which has fixed quantifier alternation depth and size bounded by $O(n^3)$. Similarly, an atomic formula of the form $x = y$ can be replaced by a formula of size $s_n \in O(n^3)$ and fixed quantifier alternation depth.

Finally, we want to express $\eta(x) = q$ for some $q \in S$. Assume that $x \in \mathbb{M}(A, I)$ is represented by the sequence $y_1 \cdots y_m$, where $m \leq n$. Then we can replace $\eta(x) = q$ by

$$\bigvee_{\substack{(q_1,\ldots,q_m)\in S^m \\ q_1\cdot q_2\cdots q_m = q}} \bigwedge_{1\leq i\leq m} \eta(y_i) = q_i.$$

Recall that $\eta(y_i) = q_i$ can be expressed by a fixed first-order formula over the signature of $\mathcal{A}$. Thus, the size of the above formula is bounded by $O(|S|^n)$ and its quantifier alternation depth is $O(1)$.

Altogether, any of the atomic subformulas in ψ gets replaced by a formula of quantifier alternation depth $O(1)$ and size bounded by $2^{O(n)}$. Hence, the size of the resulting sentence θ is bounded by $n^d \cdot |\psi| \cdot 2^{O(n)}$, and its quantifier alternation depth is bounded by $d + O(1)$. □

To reduce an arbitrary first-order sentence over the signature of $\mathcal{B}$ to a sentence of the form (4.2), we will use the technique developed by Ferrante and Rackoff, i.e., Corollary 4.4.6. In order to make it applicable, we next have to investigate the metric on $\mathbb{M}(A, I)$ that is induced by the structure $\mathcal{B}$.

Lemma 4.6.13. *Let $u = s[A_1]\cdots[A_m]y = tv \in \mathbb{M}(A, I)$ with $s[A_1]\cdots[A_m]$ in reversed Foata normal form and $|v| \leq m$. Then $t = s[A_1]\cdots[A_{m-|v|}]w$ for some $w \in \mathbb{M}(A, I)$.*

Proof. The lemma is shown by induction on $|v|$. The case $v = \varepsilon$ is trivial. Thus, let $v = av'$ for some $a \in A$, i.e., $u = s[A_1]\cdots[A_m]y = (ta)v'$. By induction we have $ta = s[A_1]\cdots[A_{m-|v|+1}]w'$ for some $w' \in \mathbb{M}(A, I)$. Then $a \in \max(s[A_1]\cdots[A_{m-|v|+1}]w')$. Since $s[A_1][A_2]\cdots[A_{m-|v|+1}]$ is in reversed Foata normal form, we obtain $a \in \max([A_{m-|v|+1}]w')$. Hence, there is a trace w satisfying $[A_{m-|v|+1}]w' = wa$, i.e., $t = s[A_1]\cdots[A_{m-|v|}]w$. □

Recall that for $r \in \mathbb{N}$ and $u \in \mathbb{M}(A, I)$ we denote by $S_{\mathcal{B}}(r, u)$ the substructure of $\mathcal{B}$ induced by the r-sphere around u in the Gaifman-graph $G_{\mathcal{B}}$. The distance function $d_{G_{\mathcal{B}}}$ of the Gaifman-graph $G_{\mathcal{B}}$ will be denoted by d in the following. Recall also that k was chosen such that every relation R_i is k-suffix definable.

Lemma 4.6.14. *Let $u = s[A_1]\cdots[A_{k\cdot r}]$ be in reversed Foata normal form. Then we have $S_{\mathcal{B}}(r, u) \subseteq s\,\mathbb{M}(A, I)$.*

Proof. Let us take $v \in \mathbb{M}(A,I)$ with $d(u,v) \leq r$. We have to show that $v \in s\,\mathbb{M}(A,I)$. By assumption there exists a path $u_0, u_1, \ldots u_m$ in the Gaifman-graph of $\mathcal{B}$ such that $u_0 = u$, $u_m = v$, and $m \leq r$. Inductively we will show that $u_i = s[A_1]\cdots[A_{kr-ki}]y_i$ for some y_i, thus $v = s[A_1]\cdots[A_{kr-km}]y_m \in s\,\mathbb{M}(A,I)$. The case $i = 0$ is clear. Now assume that $u_i = s[A_1]\cdots[A_{kr-ki}]y_i$ and $i < m$. Since (u_i, u_{i+1}) is an edge in the Gaifman-graph of $\mathcal{B}$ and all nonunary relations of $\mathcal{B}$ result from k-suffix definable relations, we have $u_i = s[A_1]\cdots[A_{kr-ki}]y_i = zw$ and $u_{i+1} = zw'$ for some $z, w, w' \in \mathbb{M}(A,I)$ with $|w| \leq k$. Lemma 4.6.13 implies that $u_{i+1} = s[A_1]\cdots[A_{kr-ki-k}]y'w'$ for some $y' \in \mathbb{M}(A,I)$. Thus, we can set $y_{i+1} = y'w'$. □

Thus, the r-sphere around $u = s[A_1][A_2]\cdots[A_{kr}]$ is contained in $s\,\mathbb{M}(A,I)$. The next lemma will be used to shorten s, i.e., to find v properly shorter than u that is the center of an isomorphic r-sphere. Let $s, t \in \mathbb{M}(A,I)$ be two traces. Since $\mathbb{M}(A,I)$ is cancellative, the mapping $f = f_{s,t} : s\,\mathbb{M}(A,I) \to t\,\mathbb{M}(A,I)$ defined by $f(su) = tu$ is a bijection. We will show that, under some assumptions on s and t, it is an isomorphism from $(S_{\mathcal{B}}(r,u), u)$ to $(S_{\mathcal{B}}(r,v), v)$.

Lemma 4.6.15. *Let $u = s[A_1]\cdots[A_{kr+k}]$ and $v = t[A_1]\cdots[A_{kr+k}]$ be in reversed Foata normal form and $\eta(s) = \eta(t)$. Then the mapping $f = f_{s,t}$ is an isomorphism from $(S_{\mathcal{B}}(r,u), u)$ to $(S_{\mathcal{B}}(r,v), v)$.*

Proof. Lemma 4.6.14 implies $S_{\mathcal{B}}(r,u) \subseteq s[A_1]\cdots[A_k]\mathbb{M}(A,I)$, thus f is defined on $S_{\mathcal{B}}(r,u)$. Since $\eta(f(x)) = \eta(x)$, f preserves all unary predicates $\eta^{-1}(q)$ for $q \in S$. Now assume that $(u_1, \ldots, u_n) \in R/_I$, where $u_i \in S_{\mathcal{B}}(r,u)$ and $R/_I$ is a relation of $\mathcal{B}$. Thus, R is k-suffix definable. Hence, there exist $y, w_1, \ldots, w_n \in \mathbb{M}(A,I)$ such that $u_i = yw_i$, $|w_i| \leq k$, and $(y'w_1, \ldots, y'w_n) \in R/_I$ for all $y' \in \mathbb{M}(A,I)$. Since $S_{\mathcal{B}}(r,u) \subseteq s[A_1]\cdots[A_k]\mathbb{M}(A,I)$, we have $yw_i = u_i = s[A_1]\cdots[A_k]v_i$ for some $v_i \in \mathbb{M}(A,I)$. Thus, Lemma 4.6.13 and $|w_i| \leq k$ implies $y = sy_i$ for some trace y_i. Since $\mathbb{M}(A,I)$ is cancellative, it follows $y_1 = \cdots = y_n =: z$. Thus, $f(u_i) = f(szw_i) = tzw_i$ and $(f(u_1), \ldots, f(u_n)) \in R/_I$. It follows that f maps $S_{\mathcal{B}}(r,u)$ injectively and structure preserving into $S_{\mathcal{B}}(r,v)$. Since we may exchange the roles of s and t, it follows that f maps $S_{\mathcal{B}}(r,u)$ bijectively to $S_{\mathcal{B}}(r,v)$. □

Now suppose that (A,I) has only finitely many neighborhoods, i.e., that the set $\{I(a) \mid a \in A\}$ is finite. Thus, also $\{D(a) \mid a \in A\}$ is finite, where $D = (A \times A) \setminus I$.

Lemma 4.6.16. *There exists a homomorphism $h : \mathbb{M}(A, I) \to Q$ into some finite monoid Q such that for all $s, t \in \mathbb{M}(A, I)$, we have:*

$$\text{if } h(s) = h(t) \text{ and } a \in \max(s), \text{ then } \exists b \in \max(t) : D(a) = D(b).$$

Proof. Let $\mathcal{D}$ be the powerset of $\{D(a) \mid a \in A\}$, thus $\mathcal{D}$ is finite. For $s \in \mathbb{M}(A, I)$, define $f(s) = \{D(a) \mid a \in \max(s)\} \in \mathcal{D}$. Let $s, t \in \mathbb{M}(A, I)$ such that $f(s) = f(t)$. We show that $f(sc) = f(tc)$ for all $c \in A$: Clearly, $c \in \max(sc) \cap \max(tc)$. Now let $a \in A \setminus \{c\}$. Then $a \in \max(sc)$ if and only if $(a, c) \in I$ and $a \in \max(s)$. Hence $f(sc) = \{D(c)\} \cup \{D(a) \mid D(a) \in f(s), c \notin D(a)\}$. Thus, indeed, $f(sc) = f(tc)$.

Now consider the image $f(\mathcal{D})$ of $\mathbb{M}(A, I)$ under f. Let Q be the transformation monoid of $f(\mathcal{D})$, i.e., $Q = (f(\mathcal{D})^{f(\mathcal{D})}, \circ)$ and define a mapping $h : A \to Q$ such that $h(a)(f(s)) = f(sa)$ which is well defined by the previous paragraph. Clearly $h(a) \circ h(b) = h(b) \circ h(a)$ for $(a, b) \in I$. Thus, we can extend h to a monoid homomorphism $h : \mathbb{M}(A, I) \to Q$ with $h(t)(f(s)) = f(st)$. Now suppose $h(s) = h(t)$. Then $a \in \max(s)$ implies $D(a) \in f(s) = h(s)(f(\varepsilon)) = f(t)$. Hence, there is $b \in \max(t)$ with $D(a) = D(b)$. □

Lemma 4.6.17. *Let h be the homomorphism from Lemma 4.6.16 and let $s, s', t, t' \in \mathbb{M}(A, I)$ with $h(s') = h(t')$, $s = s'[\max(s)]$, and $t = t'[\max(s)]$. Then $\max(t) = \max(s)$ and $\text{height}(t) = \text{height}(t') + 1$.*

Proof. Clearly, $\max(s) \subseteq \max(t)$. So let $a \in \max(t) \setminus \max(s)$. Since $t = t'[\max(s)]$, we get $a \in \max(t')$ and $(a, c) \in I$ for every $c \in \max(s)$. Since $h(s') = h(t')$, it follows $D(a) = D(b)$, i.e., $I(a) = I(b)$ for some $b \in \max(s')$. Thus, also $(b, c) \in I$ for every $c \in \max(s)$. But this implies $b \in \max(s)$, i.e., $(b, b) \in I$, a contradiction. Thus, indeed, $\max(t) = \max(s)$. This implies $\text{height}(t) = \text{height}(t'[\max(s)]) = \text{height}(t'[\max(t)]) = \text{height}(t') + 1$. □

By Lemma 4.6.16 we can find a homomorphism η' from $\mathbb{M}(A, I)$ into some finite monoid S' (namely $S \times Q$) such that the following implications hold:

- If $\eta'(s) = \eta'(t)$ and $a \in \max(s)$, then $\exists b \in \max(t) : D(a) = D(b)$.
- If $\eta'(s) = \eta'(t)$, then $\eta(s) = \eta(t)$.

Lemma 4.6.18. *Let $u \in \mathbb{M}(A, I)$ and $r, \ell \in \mathbb{N}$ such that $\ell \geq k(r + 1)$, $\text{height}(u) > k(r + 1) + |S'| + 1$, and $\text{height}(u) > \ell$. Then there exists $v \in \mathbb{M}(A, I)$ with*

$$\ell < \text{height}(v) \leq \ell + |S'| + 1 \quad \text{and} \quad (S_{\mathcal{B}}(r, u), u) \cong (S_{\mathcal{B}}(r, v), v).$$

Proof. Since $\text{height}(u) > k(r+1) + |S'| + 1$, there are (A,I)-cliques $A_i \subseteq A$ and $s \in \mathbb{M}(A,I)$ such that $u = s\,[A_1][A_2]\cdots[A_{k(r+1)}]$ is in reversed Foata normal form. Assume that $\text{height}(u) > \ell + |S'| + 1$ (otherwise we can set $v = u$). Thus, $\text{height}(s) > \ell - k(r+1) + |S'| + 1$. Let $s' \in \mathbb{M}(A,I)$ with $s = s'\,[\max(s)]$. Then $\text{height}(s') \geq \ell - k(r+1) + |S'| + 1$, and we can write $s' = s_1 s_2$ with $\text{height}(s_1) = \ell - k(r+1)$ and $\text{height}(s_2) > |S'|$. A simple pigeon hole argument shows that there exists $s_2' \in \mathbb{M}(A,I)$ such that $\eta'(s_2) = \eta'(s_2')$ and $\text{height}(s_2') \leq |S'|$. Define $t' = s_1 s_2'$ and $t = t'\,[\max(s)]$. Thus, $\eta'(s') = \eta'(t')$. Hence, by Lemma 4.6.17 we get $\text{height}(t) = \text{height}(t') + 1$ and $\max(t) = \max(s)$. This ensures in particular

$$\text{height}(t) = \text{height}(t') + 1 \leq \text{height}(s_1) + \text{height}(s_2') + 1 \leq \ell - k(r+1) + |S'| + 1$$

and

$$\text{height}(t) > \text{height}(t') \geq \text{height}(s_1) = \ell - k(r+1).$$

Now set $v = t[A_1][A_2]\cdots[A_{k(r+1)}]$. A repeated application of Lemma 4.6.17 implies $\text{height}(v) = \text{height}(t) + k(r+1)$ and therefore

$$\ell < \text{height}(v) \leq \ell + |S'| + 1.$$

Since $\eta'(s) = \eta'(t)$, we can apply Lemma 4.6.15, implying $(S_{\mathcal{B}}(r,u),u) \cong (S_{\mathcal{B}}(r,v),v)$, which finishes the proof. □

Proof of Theorem 4.6.6. We first show that the factorized unfolding $\mathcal{B}$ is H-bounded for a suitable function H. Recall that the norm of a trace t was defined as its length $|t|$.

Since $\{I(a) \mid a \in A\}$ is finite, there is $\alpha \in \mathbb{N}$ such that any (A,I)-clique contains at most α elements. We define $H(i,d)$ inductively: $H(1,d) = \alpha \cdot (k(7^{d-1}+1) + 2(|S'|+1))$ and $H(j,d) = \alpha \cdot (H(j-1,d) + 4 \cdot 7^{d-j} \cdot k + |S'| + 1)$ for $1 < j \leq d$. Then $H(j,d) \leq H(d,d)$ is bounded by $2^{O(d)}$ for $j \leq d$.

Let $d \in \mathbb{N}$ and $t \in \mathbb{M}(A,I)$ with $|t| > H(1,d)$. Let $\ell = k(7^{d-1}+1) + |S'| + 1 < H(1,d)/\alpha$ and $r = 7^{d-1}$. Then $\ell \geq k(r+1)$ and $\ell = k(r+1) + |S'| + 1 < H(1,d)/\alpha < |t|/\alpha \leq \text{height}(t)$. Hence, by Lemma 4.6.18, there is $t_1 \in \mathbb{M}(A,I)$ with $\text{height}(t_1) \leq \ell + |S'| + 1 \leq H(1,d)$ and $(S_{\mathcal{B}}(7^{d-1},t),t) \cong (S_{\mathcal{B}}(7^{d-1},t_1),t_1)$. This proves the base case for the H-boundedness of $\mathcal{B}$.

Next, let $1 < j \leq d \in \mathbb{N}$, $\widetilde{t} = (t_1, t_2, \ldots, t_{j-1}) \in \mathbb{M}(A,I)^{j-1}$ with $|t_i| \leq H(i,d)$ and $t \in \mathbb{M}(A,I)$ with $|t| > H(j,d)$. In order to apply Lemma 4.6.18, let $\ell = H(j-1,d) + 4 \cdot 7^{d-j} \cdot k$ and $r = 7^{d-j}$. Thus, $\ell \geq H(1,d) \geq k(7^{d-1}+1) \geq$

$k(r+1)$. Moreover, $|t| > H(j,d) > H(1,d) \geq \alpha \cdot (k(7^{d-j}+1)+|S'|+1)$. Hence, $\text{height}(t) \geq \frac{|t|}{\alpha} > k(r+1)+|S'|+1$. Furthermore, $|t| > H(j,d) > \alpha \cdot \ell$ implies $\text{height}(t) > \ell$. Thus, by Lemma 4.6.18, there exists $t_j \in \mathbb{M}(A,I)$ with

$$\ell < \text{height}(t_j) \leq \ell + |S'| + 1 \quad \text{and} \quad (S_{\mathcal{B}}(7^{d-j}, t), t) \cong (S_{\mathcal{B}}(7^{d-j}, t_j), t_j).$$

Thus, $\ell < |t_j| \leq \alpha \cdot (\ell + |S'| + 1)$. In the Gaifman-graph $G_{\mathcal{B}}$, the distance between t_i and t_j is at least $(|t_j| - |t_i|)/k$. Since $|t_i| \leq H(i,d) \leq H(j-1,d)$ for $1 \leq i < j$, we obtain $d(t_i, t_j) \geq (|t_j| - |t_i|)/k > (\ell - H(j-1,d))/k = 4 \cdot 7^{d-j}$. Hence the spheres $S_{\mathcal{B}}(7^{d-j}, \widetilde{t})$ and $S_{\mathcal{B}}(7^{d-j}, t_j)$ are disjoint and no edge in $G_{\mathcal{B}}$ connects elements from the former to elements from the latter. Clearly, the same holds for the spheres $S_{\mathcal{B}}(7^{d-j}, \widetilde{t})$ and $S_{\mathcal{B}}(7^{d-j}, t)$. Thus,

$$(S_{\mathcal{B}}(7^{d-j}, \widetilde{t}, t), \widetilde{t}, t) \cong (S_{\mathcal{B}}(7^{d-j}, \widetilde{t}, t_j), \widetilde{t}, t_j).$$

Thus, the factorized unfolding $\mathcal{B}$ is indeed H-bounded.

Now let $\varphi \equiv Q_1 x_1 Q_2 x_2 \cdots Q_d x_d : \psi(x_1, \ldots, x_d)$ be a first-order sentence over the signature of $\mathcal{B}$ with d quantifiers $Q_i \in \{\exists, \forall\}$. Then, by Corollary 4.4.6, $\mathcal{B} \models \varphi$ if and only if

$$\mathcal{B} \models Q_1 x_1 \leq H(1,d) Q_2 x_2 \leq H(2,d) \cdots Q_d x_d \leq H(d,d) : \psi(x_1, \ldots, x_d).$$

Since $H(i,d) \leq H(d,d) \in 2^{O(d)}$, Proposition 4.6.12 implies that this statement can be reduced to an equivalent statement on $\mathcal{A}$ of quantifier alternation depth $d + O(1)$ and size $2^{2^{O(|\varphi|)}}$. □

Remark 4.6.19. *Let $(R_i)_{i \in \mathbb{N}}$ be an enumeration of all relations that are k-suffix definable in $\mathcal{A}$ for some $k \in \mathbb{N}$. Note that there are only countably many such relations. Moreover, let $(L_i)_{i \in \mathbb{N}}$ be an enumeration of all subsets $\eta^{-1}(q) \subseteq \mathbb{M}(A,I)$ such that $\eta : \mathbb{M}(A,I) \to S$ is a homomorphism into a finite monoid S, $q \in S$, and $\eta^{-1}(p) \cap A$ is first-order definable in $\mathcal{A}$ for every $p \in S$. Again there are only countably many such subsets. Note that $L_i \in \text{REC}(\mathbb{M}(A,I))$ for every $i \in \mathbb{N}$. Then also the first-order theory of $\mathcal{B} = (\mathbb{M}(A,I), (L_i)_{i \in \mathbb{N}}, (R_i/_I)_{i \in \mathbb{N}})$ is decidable. The important point is that any first-order sentence over the signature of $\mathcal{B}$ can mention only finitely many relations $R_i/_I$ and L_i. Thus, it suffices to work in a suitable reduct of $\mathcal{B}$ with only finitely many relations, which can be handled by Theorem 4.6.6.*

4.7 Graph products

In this section we will introduce graph products of monoids. The graph product construction generalizes both the free product and the direct product. Graph products were introduced in [87]. Our main result will state that the decidability of the first-order theory (resp. MSO theory) of the Cayley-graph is preserved under graph products (resp. free products). Other closure results for graph products can be found for instance in [94, 121, 211, 212].

Let (Σ, I_Σ) be a finite independence alphabet, i.e., Σ is finite, and let $\mathcal{M}_\sigma = (M_\sigma, \circ_\sigma, 1_\sigma)$ be a finitely generated monoid for every $\sigma \in \Sigma$. As already mentioned in Section 2.3, the monoid $\mathcal{M}_\sigma$ can be presented by (A_σ, R_σ) where $A_\sigma = M_\sigma \setminus \{1_\sigma\}$ and

$$R_\sigma = \{(ab, c) \mid a, b, c \in A_\sigma, a \circ_\sigma b = c\} \cup \{(ab, \varepsilon) \mid a, b \in A_\sigma, a \circ_\sigma b = 1_\sigma\}.$$

Let $R = \bigcup_{\sigma\in\Sigma} R_\sigma$ and define an independence alphabet (A, I) by

$$A = \bigcup_{\sigma\in\Sigma} A_\sigma \quad \text{and} \quad I = \bigcup_{(\sigma,\tau)\in I_\Sigma} A_\sigma \times A_\tau,$$

where w.l.o.g. $A_\sigma \cap A_\tau = \emptyset$ for $\sigma \neq \tau$. The *graph product* $\mathbb{P}(\Sigma, I_\Sigma, (\mathcal{M}_\sigma)_{\sigma\in\Sigma})$ is defined by

$$\mathbb{P}(\Sigma, I_\Sigma, (\mathcal{M}_\sigma)_{\sigma\in\Sigma}) = \mathcal{M}(A, R \cup \{(ab, ba) \mid (a, b) \in I\}).$$

Special cases are the free product $*_{\sigma\in\Sigma}\mathcal{M}_\sigma$ (if $I_\Sigma = \emptyset$) and the *direct product* $\prod_{\sigma\in\Sigma} \mathcal{M}_\sigma$ (if $I_\Sigma = (\Sigma \times \Sigma) \setminus \{(\sigma, \sigma) \mid \sigma \in \Sigma\}$). Let us fix a graph product $\mathbb{P} = \mathbb{P}(\Sigma, I_\Sigma, (\mathcal{M}_\sigma)_{\sigma\in\Sigma})$ for the further discussion.

In the following we will consider R as a trace rewriting system over the trace monoid $\mathbb{M}(A, I)$, which is not finitely generated as soon as one of the monoids $\mathcal{M}_\sigma$ is infinite. From the definition of $\mathbb{P}$ we obtain $\mathbb{P} \cong \mathbb{M}(A, I)/\overset{*}{\leftrightarrow}_R$. The crucial fact for our further investigation is the following:

Lemma 4.7.1. *The trace rewriting system R over $\mathbb{M}(A, I)$ is confluent.*

Proof. Since R is terminating, it suffices to show that R is locally confluent. Thus, assume that $s \to_R s_1$ and $s \to_R s_2$. Hence, $s = t_i a_i b_i u_i$ and $s_i = t_i r_i u_i$ for $i = 1, 2$, where $(a_i b_i, r_i) \in R$. Thus, $r_i \in A \cup \{\varepsilon\}$. By applying Levi's Lemma 2.7.1 to the identity $t_1 a_1 b_1 u_1 = t_2 a_2 b_2 u_2$, we obtain the following diagram:

u_2	w_2	q_1	v_2
a_2b_2	p_2	t	q_2
t_2	v_1	p_1	w_1
	t_1	a_1b_1	u_1

Thus, $(w_1, w_2) \in I$. For the further arguments it is easy to see that we may assume $v_1 = v_2 = \varepsilon$. Assume that $a_i, b_i \in A_{\sigma_i}$. Let us first assume that $t \neq \varepsilon$. Thus, $\sigma_1 = \sigma_2 = \sigma$, $r_1, r_2 \in A_\sigma \cup \{\epsilon\}$, and $(c, w_i) \in I$ for all $c \in A_\sigma$ and $i = 1, 2$. Moreover, since $(p_1, p_2) \in I$ but both traces only contain symbols from A_σ, we have either $p_1 = \varepsilon$ or $p_2 = \varepsilon$ and similarly either $q_1 = \varepsilon$ or $q_2 = \varepsilon$. If $p_1 = p_2 = q_1 = q_2 = \varepsilon$ then $s_1 = s_2$. Otherwise, since a_1b_1 cannot be a proper factor of a_2b_2 and vice versa, we obtain up to symmetry the following diagram:

u_2	w_2	b_1	ε
a_2b_2	a_2	$b_2 = a_1$	ε
t_2	ε	ε	w_1
	t_1	a_1b_1	u_1

Thus, $s_1 = a_2w_2r_1w_1 = w_1a_2r_1w_2$ and $s_2 = w_1r_2w_2b_1 = w_1r_2b_1w_2$. Finally, by definition of the system R_σ it follows that a_2r_1 and r_2b_1 can be reduced to $a_2 \circ_\sigma b_2 \circ_\sigma b_1 = a_2 \circ_\sigma a_1 \circ_\sigma b_1$. This concludes the case $t \neq \varepsilon$.

Now assume that $t = \varepsilon$. Thus, we have the following diagram:

u_2	w_2	q_1	ε
a_2b_2	p_2	ε	q_2
t_2	ε	p_1	w_1
	t_1	a_1b_1	u_1

If also $p_1 = \varepsilon$, i.e.,

u_2	w_2	a_1b_1	ε
a_2b_2	p_2	ε	q_2
t_2	ε	ε	w_1
	t_1	a_1b_1	u_1

then $(w_1q_2, a_1) \in I$ implies $(w_1q_2, r_1) \in I$. We have to show that $s_1 = p_2w_2r_1w_1q_2$ and $s_2 = w_1r_2w_2a_1b_1$ can be reduced to the same trace. We have $s_2 \to_R w_1r_2w_2r_1$. Moreover with the independencies listed above, we obtain

$$s_1 = p_2w_2r_1w_1q_2 = p_2w_2w_1q_2r_1 = w_1p_2q_2w_2r_1 \to_R w_1r_2w_2r_1.$$

If one of the traces p_2, q_1, or q_2 is empty, then we can argue analogously. Thus, we may assume that p_1, p_2, q_1, and q_2 are nonempty. It follows $p_1 = a_1$, $q_1 = b_1$, $p_2 = a_2$, and $q_2 = b_2$. Then all traces from $\{w_1, w_2, a_1b_1, a_2b_2\}$ are pairwise independent, from which it follows again easily that s_1 and s_2 can be reduced to $w_1w_2r_1r_2$. □

Since R is also terminating, the previous lemma implies that $\mathbb{P}$ is in one-to-one correspondence with $\mathrm{IRR}(R) \subseteq \mathbb{M}(A, I)$, which is the set of all traces that do not contain a factor of the form ab with $a, b \in A_\sigma$ for some $\sigma \in \Sigma$.

For the further consideration, assume that $\mathcal{M}_\sigma$ is finitely generated by $\Gamma_\sigma \subseteq A_\sigma$. From the definition of $\mathbb{P}$ it is obvious that $\mathbb{P}$ is finitely generated by $\Gamma = \bigcup_{\sigma\in\Sigma} \Gamma_\sigma$, and we will choose this finite generating set for the further discussion.[11]

In the sequel, it will be useful to add the neutral element as a constant to the Cayley-graph. Thus, define the *rooted Cayley-graph* of a finitely generated monoid $\mathcal{M}$ as the rooted graph $(\mathcal{C}(\mathcal{M}), 1)$, where 1 is the neutral element of $\mathcal{M}$.

Our next goal is to define the rooted Cayley-graph $(\mathcal{C}(\mathbb{P}, \Gamma), 1)$ within the trace monoid $\mathbb{M}(A, I)$. For $a \in \Gamma$, let us define the edge-relation

$$F_a = \{(s, t) \in \mathrm{IRR}(R) \times \mathrm{IRR}(R) \mid sa \xrightarrow{*}_R t\}.$$

Since $\mathbb{P} \cong \mathbb{M}(A, I)/\overset{*}{\leftrightarrow}_R$ and R is confluent and terminating, we obtain the following lemma:

Lemma 4.7.2. $(\mathrm{IRR}(R), (F_a)_{a\in\Gamma}), \varepsilon)$ *is isomorphic to* $(\mathcal{C}(\mathbb{P}, \Gamma), 1)$.

Lemma 4.7.3. *For* $s, t \in \mathrm{IRR}(R) \subseteq \mathbb{M}(A, I)$ *and* $a \in \Gamma_\sigma \subseteq A_\sigma$ *we have* $(s, t) \in F_a$ *if and only if in* $\mathbb{M}(A, I)$*:*

- $t = sa$ *(and thus* $\max(s) \cap A_\sigma = \emptyset$*), or*
- $s = tb$ *for* $b \in A_\sigma$ *and* $b \circ_\sigma a = 1_\sigma$*, or*
- $s = ub$ *for* $u \in \mathrm{IRR}(R)$*,* $b \in A_\sigma$*,* $b \circ_\sigma a = c \neq 1_\sigma$*, and* $t = uc$.

[11] In fact, if the monoid $\mathcal{M}_\sigma$ is given by the presentation $(\Gamma_\sigma, S_\sigma)$, then, using Tietze transformations, it is easy to see that $\mathbb{P}$ can be presented by $(\bigcup_{\sigma\in\Sigma} \Gamma_\sigma, \bigcup_{\sigma\in\Sigma} S_\sigma \cup \{(ab, ba) \mid a \in \Gamma_\sigma, b \in \Gamma_\tau, (\sigma, \tau) \in I_\Sigma\})$.

Proof. If one of the three cases above holds, then it is easy to see that indeed $sa \xrightarrow{*}_R t$, i.e., $(s,t) \in F_a$. Now assume that $sa \xrightarrow{*}_R t \in \mathrm{IRR}(R)$. If $t \neq sa$, then $sa \to_R v \xrightarrow{*}_R t$ for some trace v. Thus, there exist $\tau \in \Sigma$, $(bc, r) \in R_\tau$, and $s_1, s_2 \in \mathbb{M}(A, I)$ such that $sa = s_1 bc s_2$. By applying Levi's lemma to this identity and using $s \in \mathrm{IRR}(R)$, we obtain the following diagram:

s_2	s_2	ε
bc	b	$a = c$
s_1	s_1	ε
	s	a

Thus, $\tau = \sigma$ and $(a, s_2) \in I$, which implies also $(b, s_2) \in I$. Thus, $s = ub$, for $u = s_1 s_2$. If $r = \varepsilon$, i.e., $b \circ_\sigma a = 1_\sigma$, then $v = u \in \mathrm{IRR}(R)$, thus $t = u$ and $s = tb$, i.e., the second case from the lemma holds. On the other hand, if $r = b \circ_\sigma a = c \neq 1_\sigma$, then $v = uc$, which again belongs to $\mathrm{IRR}(R)$ (otherwise, since $a, c \in A_\sigma$, also $s = ua \in \mathrm{RED}(R)$). Hence $t = v = uc$. □

We now define a structure

$$\mathcal{A} = (A, (A_\sigma)_{\sigma \in \Sigma}, (E_a)_{a \in \Gamma}, (a)_{a \in \Gamma}),$$

where for $a \in \Gamma_\sigma$, $\sigma \in \Sigma$, E_a consists of all pairs $(x, y) \in A_\sigma \times A_\sigma$ such that $x \circ_\sigma a = y$ in $\mathcal{M}_\sigma$. Thus, $\mathcal{A}$ is the disjoint union of the restricted Cayley-graphs $\mathcal{C}(\mathcal{M}_\sigma, \Gamma_\sigma) \backslash \{1_\sigma\}$, where moreover every generator $a \in \Gamma_\sigma$ is added as a constant, and every $A_\sigma \subseteq A$ is added as a unary predicate. We will apply Theorem 4.6.2 and 4.6.6 to the structure $\mathcal{A}$. For this, we now define a suitable factorized unfolding of $\mathcal{A}$. First, note that the independence relation $I \subseteq A \times A$ is first-order definable in $\mathcal{A}$ using the unary predicates A_σ and that $\{I(a) \mid a \in A\}$ is finite: If $a, b \in A_\sigma$, then $I(a) = I(b)$. In order to define a suitable homomorphism $\eta : \mathbb{M}(A, I) \to S$ into a finite monoid S, let us consider the finitely generated trace monoid $\mathbb{M}(\Sigma, I_\Sigma)$. The closure properties of recognizable trace languages (see Section 2.8) imply that

$$L = \bigcup_{\sigma \in \Sigma} \mathbb{M}(\Sigma, I_\Sigma) \sigma \sigma \mathbb{M}(\Sigma, I_\Sigma) \in \mathrm{REC}(\mathbb{M}(\Sigma, I_\Sigma)).$$

Hence, there exists a homomorphism $h : \mathbb{M}(\Sigma, I_\Sigma) \to S$ into a finite monoid S and a subset $F \subseteq S$ such that $L = h^{-1}(F)$. Now define $g : A \to \Sigma$ by $g(a) = \sigma$ if $a \in A_\sigma$. We can extend g to $g : \mathbb{M}(A, I) \to \mathbb{M}(\Sigma, I_\Sigma)$. Let $\eta = g \circ h$. Then $\eta^{-1}(F) = \mathrm{RED}(R)$ and $\eta^{-1}(S \setminus F) = \mathrm{IRR}(R)$. Note that

for every $q \in S$, the set $\eta^{-1}(q) \cap A_\sigma$ is either empty or A_σ. Thus, every set $\eta^{-1}(q) \cap A$ is first-order definable in $\mathcal{A}$.

From the previous discussion it follows that the structure

$$\mathcal{B} = (\mathbb{M}(A, I), (\eta^{-1}(q))_{q \in S}, \mathrm{suc}, (\widehat{A_\sigma}/_I)_{\sigma \in \Sigma}, (\widehat{E_a}/_I)_{a \in \Gamma}, (\widehat{a}/_I)_{a \in \Gamma})$$

(see Section 4.6.1 for the definition of the relation $\widehat{R}$) is a factorized unfolding of $\mathcal{A}$.[12] We next present a first-order interpretation of the rooted Cayley-graph $(\mathcal{C}(\mathbb{P}, \Gamma), 1)$ in $\mathcal{B}$.

Lemma 4.7.4. *$(\mathcal{C}(\mathbb{P}, \Gamma), 1)$ is first-order interpretable in $\mathcal{B}$.*

Proof. By Lemma 4.7.2 it suffices to show that $(\mathrm{IRR}(R), (F_a)_{a \in \Gamma}, \varepsilon)$ is first-order interpretable in $\mathcal{B}$. First, recall that $\mathrm{IRR}(R) = \eta^{-1}(S \setminus F)$. Moreover, ε is the only trace t such that there is no s with $(s, t) \in \mathrm{suc}$. Finally, by Lemma 4.7.3 we have $(s, t) \in F_a$ for $s, t \in \mathrm{IRR}(R)$ and $a \in \Gamma_\sigma$ if and only if in $\mathcal{B}$:

- $(s, t) \in \mathrm{suc}$, $s \notin \widehat{A_\sigma}/_I$, and $t \in \widehat{a}/_I$ (i.e., $t = sa$) or
- $(t, s) \in \mathrm{suc}$, $t \notin \widehat{A_\sigma}/_I$, $s \in \widehat{A_\sigma}/_I$, but there is no u with $(s, u) \in \widehat{E_a}/_I$ (note that if $s = vb$ with $b \in A_\sigma$ but there is no u with $(s, u) \in \widehat{E_a}/_I$, then $b \circ_\sigma a = 1_\sigma$), or
- $(s, t) \in \widehat{E_a}/_I$.

This proves the lemma. □

The following theorem is the main result of this section.

Theorem 4.7.5. *Let $\mathbb{P} = \mathbb{P}(\Sigma, I_\Sigma, (\mathcal{M}_\sigma)_{\sigma \in \Sigma})$, where $\mathcal{M}_\sigma$ is finitely generated.*

(1) If $\mathrm{FOTh}(\mathcal{C}(\mathcal{M}_\sigma), 1_\sigma)$ is decidable for all $\sigma \in \Sigma$, then also $\mathrm{FOTh}(\mathcal{C}(\mathbb{P}), 1)$ is decidable.

(2) If $I_\Sigma = \emptyset$ and $\mathrm{MSOTh}(\mathcal{C}(\mathcal{M}_\sigma), 1_\sigma)$ is decidable for all $\sigma \in \Sigma$, then also $\mathrm{MSOTh}(\mathcal{C}(\mathbb{P}), 1)$ is decidable.

[12] Here we identify the constant a with the unary relation $\{a\}$, thus $\widehat{a} = A^*a$.

Proof. Assume that $\mathcal{M}_\sigma$ is finitely generated by $\Gamma_\sigma \subseteq \mathcal{M}_\sigma \setminus \{1_\sigma\}$. Thus, $\mathbb{P}$ is finitely generated by $\Gamma = \bigcup_{\sigma\in\Sigma} \Gamma_\sigma$.

Let us first prove (1). If $\mathrm{FOTh}(\mathcal{C}(\mathcal{M}_\sigma, \Gamma_\sigma), 1_\sigma)$ is decidable, then, since every $a \in \Gamma_\sigma$ is first-order definable in the rooted Cayley-graph $(\mathcal{C}(\mathcal{M}_\sigma, \Gamma_\sigma), 1_\sigma)$, also the structure $(\mathcal{C}(\mathcal{M}_\sigma, \Gamma_\sigma)\setminus\{1_\sigma\}, (a)_{a\in\Gamma_\sigma})$ has a decidable first-order theory. By the Feferman-Vaught Theorem [81], the same holds for the disjoint union of these structures with the unary predicates $\mathcal{M}_\sigma \setminus \{1_\sigma\}$ added. But this is precisely the structure $\mathcal{A}$ from the previous discussion. We can therefore apply Corollary 4.6.8 and obtain that $\mathrm{FOTh}(\mathcal{B})$ is decidable. Since $(\mathcal{C}(\mathbb{P}, \Gamma), 1)$ is first-order interpretable in $\mathcal{B}$ (Lemma 4.7.4), it follows that the first-order theory of $(\mathcal{C}(\mathbb{P}), 1)$ is indeed decidable.

Now assume that $I_\Sigma = \emptyset$, i.e., $\mathbb{M}(A, I) = A^*$. The argumentation is similar to the first-order case: If $\mathrm{MSOTh}(\mathcal{C}(\mathcal{M}_\sigma, \Gamma_\sigma), 1_\sigma)$ is decidable for every $\sigma \in \Sigma$, then also $\mathrm{MSOTh}(\mathcal{A})$ is decidable, see e.g. [191]. Hence, by Lemma 4.6.5, also $\mathrm{MSOTh}(\widehat{\mathcal{A}}, (\eta^{-1}(q))_{q\in S})$ is decidable. But the structure $\mathcal{B}$ from the previous discussion (for $I_\Sigma = \emptyset$) is a reduct of this structure. Hence, $\mathrm{MSOTh}(\mathcal{B})$ is decidable, and the result follows again from Lemma 4.7.4. Note that the cl-predicate from $\widehat{A}$ is actually not needed here. □

Remark 4.7.6. *Concerning the complexity of* $\mathrm{FOTh}(\mathcal{C}(\mathbb{P}), 1)$*, note that Theorem 4.6.6 (more precisely Remark 4.6.7) allows us to reduce* $\mathrm{FOTh}(\mathcal{C}(\mathbb{P}), 1)$ *in doubly exponential time to* $\mathrm{FOTh}(\mathcal{A})$*. Recall that* $\mathcal{A}$ *is essentially the disjoint union of the Cayley-graphs of the monoids* $\mathcal{M}_\sigma$*. To the knowledge of the author, all known proofs for decomposition theorems (in the style of Feferman-Vaught's theorem) that allow to reduce the theory of a disjoint union (or direct product) to the theories of the factors, lead to a nonelementary blow-up in terms of complexity. Therefore, we are only able to give a nonelementary upper bound for* $\mathrm{FOTh}(\mathcal{C}(\mathbb{P}), 1)$ *even if all the theories* $\mathrm{FOTh}(\mathcal{C}(\mathcal{M}_\sigma), 1_\sigma)$ *can be decided in elementary time.*

For monadic second-order logic the situation is clear: Note that $\mathcal{C}(\mathbb{Z}/2\mathbb{Z})$ *is a graph with two nodes, thus its monadic second-order theory is in PSPACE (it is in fact PSPACE-complete). But in* $\mathcal{C}(\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/2\mathbb{Z})$ *we can define* $\mathbb{Z}$ *with the successor relation, which has a nonelementary MSO theory [142].*

An immediate corollary of Corollary 4.4.9 and Theorem 4.7.5(1) is the following result:

Corollary 4.7.7. *Let* $\mathcal{M}$ *be a graph product of automatic monoids and groups with decidable word problems. Then* $\mathrm{FOTh}(\mathcal{C}(\mathcal{M}))$ *is decidable.*

The author currently doesn't know any larger class of monoids with the latter property.

Theorem 4.7.5(1) does not generalize to MSO theories:

Proposition 4.7.8. *Let $\mathbb{P} = \mathbb{P}(\Sigma, I_\Sigma, (\mathcal{M}_\sigma)_{\sigma\in\Sigma})$, where $\mathcal{M}_\sigma$ is nontrivial and finitely generated by Γ_σ. If $\mathrm{MSOTh}(\mathcal{C}(\mathbb{P}), 1)$ is decidable, then:*

- *(Σ, I_Σ) does not contain an induced cycle of length 4 (also called C4),*
- *if $(\sigma, \tau) \in I_\Sigma$ and $\mathcal{M}_\sigma$ is infinite, then $\mathcal{M}_\tau$ is finite,*
- *if $(\sigma, \sigma_1), (\sigma, \sigma_2) \in I_\Sigma$, $\sigma_1 \neq \sigma_2$, and $\mathcal{M}_\sigma$ is infinite, then $(\sigma_1, \sigma_2) \in I_\Sigma$, and*
- *$\mathrm{MSOTh}(\mathcal{C}(\mathcal{M}_\sigma), 1_\sigma)$ is decidable for every $\sigma \in \Sigma$.*

Proof. If one of the first three conditions is not satisfied, then $\mathbb{P}$ contains a submonoid of the form $\mathcal{M}_1 \times \mathcal{M}_2$, where both $\mathcal{M}_1$ and $\mathcal{M}_2$ are infinite (note that we assume that every $\mathcal{M}_\sigma$ is nontrivial). Since $\mathcal{C}(\mathcal{M}_i)$ is infinite and has bounded outdegree, we find an infinite path $a_{i,1} \to a_{i,2} \to \cdots$ in $\mathcal{C}(\mathcal{M}_i)$. In $\mathcal{M}_1 \times \mathcal{M}_2 \subseteq \mathbb{P}$, these two paths generate an infinite grid. Hence the MSO theory of $\mathcal{C}(\mathbb{P})$ is undecidable, see Theorem 4.5.2.

Next we show that $\mathrm{MSOTh}(\mathcal{C}(\mathcal{M}_\sigma), 1_\sigma)$ is decidable for every $\sigma \in \Sigma$ in case $\mathrm{MSOTh}(\mathcal{C}(\mathbb{P}), 1)$ is decidable. Note that an MSO sentence φ holds in $(\mathcal{C}(\mathcal{M}_\sigma), 1_\sigma)$ if and only if its restriction to $\mathcal{M}_\sigma \subseteq \mathbb{P}$ holds in $(\mathcal{C}(\mathbb{P}), 1)$. It therefore suffices to show that $\mathcal{M}_\sigma$ is definable in the Cayley-graph of the graph product. But $x \in \mathbb{P}$ belongs to $\mathcal{M}_\sigma$, if and only if there exists a path from 1 to x in $\mathcal{C}(\mathbb{P})$ such that every edge along this path is labeled by a generator from Γ_σ. This property can be easily expressed in monadic second-order logic. □

4.8 Open problems

Various problems remain open for Cayley-graphs of finitely generated monoids. The most ambitious goal would be to obtain a complete (algebraic or combinatorial) characterization of those monoids such that the corresponding Cayley-graphs has a decidable first-order theory or MSO theory, respectively. But due to the missing symmetry in Cayley-graphs of monoids this problem might be too difficult. A promising class for further results might be cancellative monoids. Their Cayley-graphs have at least bounded degree. Is

there a cancellative monoid with a decidable word problem such that the corresponding Cayley-graph has an undecidable first-order theory? Is there a cancellative monoid such that its Cayley-graph has finite tree-width but an undecidable MSO theory?

We have shown that for Cayley-graphs of finitely generated groups the decidability of the full first-order theory is equivalent to the decidability of the existential first-order theory (Corollary 4.4.9). The corresponding problem for monoids is again open.

As already mentioned, in Theorem 4.7.5(1) it remains open whether the complexity of $\mathrm{FOTh}(\mathcal{C}(\mathbb{P}), 1)$ is bounded elementary in the complexity for the theories $\mathrm{FOTh}(\mathcal{C}(\mathcal{M}_\sigma), 1_\sigma)$, where the $\mathcal{M}_\sigma$ are the factors of the graph product. For MSO theories, the statements in Theorem 4.7.5(2) and Proposition 4.7.8 leave a gap. We conjecture that $\mathrm{MSOTh}(\mathcal{C}(\mathbb{P}), 1)$ is decidable if and only if the four conditions in Proposition 4.7.8 are satisfied. One might first try to prove this conjecture for graph products of finite monoids. In particular, if the independence relation (Σ, I_Σ) is a chain of four nodes (also called P4) and every node is labeled with a finite monoid, then it is not clear whether the corresponding graph product has a Cayley-graph with a decidable MSO theory.

Section 4.3.4 leaves also some open problems concerning general graphs. We have shown that a connected graph of bounded degree with only finitely many orbits has a decidable MSO theory if and only if it is context-free. Is there an analogous characterization for countable graphs of unbounded degree? Note that for graphs of unbounded degree, the decidability of $\mathrm{MSOTh}(G)$ does not necessarily imply the decidability of $\mathrm{MSOTh}(G^{(e)})$ – take for instance the complete graph on $\mathbb{N}$. One might conjecture that for a connected and countable graph G with only finitely many orbits, $\mathrm{MSOTh}(G)$ (resp. $\mathrm{MSOTh}(G^{(e)})$) is decidable if and only if G is prefix-recognizable [44] (resp. equational [53]) if and only if G has finite clique-width [56] (resp. finite tree-width).

Chapter 5

Word equations

5.1 Outline

In this chapter we will present several new results concerning theories of word equations over monoids. In Section 5.2 and 5.3 we will introduce the general framework of this chapter. Section 5.4 deals with existential theories. Our first main result (Theorem 5.4.3) states, roughly speaking, that, given a structure $\mathbb{A} = (A, \ldots)$ with a decidable existential theory together with an independence relation I, we can define on the trace monoid $\mathbb{M}(A, I)$ a suitable lifting of $\mathbb{A}$, whose existential theory is again decidable. Similarly to our main result on factorized unfoldings (Theorem 4.6.6) we have to require that the set $\{I(a) \mid a \in A\}$ is finite and that I is definable in $\mathbb{A}$ (in fact we will require a stronger condition). Based on this result, we will show in Section 5.4.3 that under some algebraic restriction on the factors of a graph product, the decidability of the existential theory of word equations is preserved under graph products (Theorem 5.4.10). This closure result remains also valid if we allow constraints for variables, which means that the value of a variable may be restricted to some specified set. More precisely, we will define an operation, which, starting from a class of constraints for each factor monoid of the graph product, constructs a class of constraints for the graph product. This construction is inspired by the notion of bipartite automata, which was introduced by Sakarovitch [177, 178] in order to study rational sets in free products. We will also present an upper bound for the space complexity of the existential theory of the graph product in terms of the space complexities for the existential theories of the factor monoids. This upper bound involves

an exponential blow-up. For a restricted class of graph products, we show in Section 5.4.4 that this exponential blow-up can be avoided. More precisely, we show that for a graph product of finite monoids, free monoids, and free groups, the existential theory of word equations can be decided in PSPACE (Corollary 5.4.16). This class of graph products strictly covers for instance trace monoids, free partially commutative groups (called semifree groups in [11, 12], right-angled Artin groups in [31], and graph groups in [73]), and plain groups [92]. If we allow constraints that are constructed from rational sets for the factor monoids using the general construction from Section 5.4.3, then we obtain PSPACE-completeness. Moreover, under certain restrictions on the graph underlying the graph product, PSPACE-completeness holds also in the case that (a suitable description) of the graph product is part of the input.

In Section 5.5 we will investigate positive theories of equations (a sentence is called positive if it is constructed from atomic formulas using only conjunctions, disjunctions, and quantifiers). We prove that the positive theory of word equations of a graph product of *groups* with recognizable constraints can be reduced to

- the positive theories with recognizable constraints of those factors of the graph product that are located in isolated nodes of the global dependence relation and
- the existential theories of those factors of the graph product that are located in nonisolated nodes of the global dependence relation.

As a corollary we obtain the decidability of the positive theory of a graph product of finite and free groups with recognizable constraints. This generalizes the well-known result of Makanin for free groups [132, 133]. The technical part relies on a generalization of the techniques introduced by Merzlyakov for free groups [141]. Our decision method leads only to a nonelementary algorithm for the positive theory, but additional restrictions on the graph underlying the graph product give us an elementary upper bound.

The results of this section are partially contained in [66].

5.2 Monoids with involution

A *monoid involution* on a monoid $\mathcal{M} = (M, \circ, 1)$ is an involution $\iota : M \to M$ such that $\iota(a \circ b) = \iota(b) \circ \iota(a)$ for all $a, b \in M$ and $\iota(1) = 1$. A typical example

of a monoid involution is taking the inverse in a group. A *partial monoid involution* on a monoid $\mathcal{M}$ is given by a submonoid $\mathcal{I}$ of $\mathcal{M}$ together with a monoid involution $\iota : \mathcal{I} \to \mathcal{I}$, the structure $(\mathcal{M}, \iota)$ is called a *monoid with partial involution*. If $(\mathcal{N}, \zeta)$ is another monoid with partial involution, then a homomorphism $f : (\mathcal{M}, \iota) \to (\mathcal{N}, \zeta)$ is a monoid homomorphism $f : \mathcal{M} \to \mathcal{N}$ such that furthermore $a \in \mathrm{dom}(\iota)$ implies $f(a) \in \mathrm{dom}(\zeta)$ and $\zeta(f(a)) = f(\iota(a))$. Let us discuss two important special cases of a monoid with partial involution:

- Let $\mathcal{U}$ be the *subgroup of all units* in $\mathcal{M}$, i.e., $a \in \mathcal{U}$ if there exists $b \in \mathcal{M}$ such that $a \circ b = b \circ a = 1$. On $\mathcal{U}$ we can define an involution $\iota : \mathcal{U} \to \mathcal{U}$ by setting $\iota(a) = a^{-1}$.
- Let (A, I) be an independence alphabet and let $B \subseteq A$ together with an involution $\iota : B \to B$. We say that $\iota : B \to B$ is *compatible* with I if $(a, b) \in I$ and $a, b \in B$ implies $(\iota(a), \iota(b)) \in I$. This allows to extend $\iota : B \to B$ to $\iota : \mathbb{M}(B, I) \to \mathbb{M}(B, I)$ by $\iota([a_1 \cdots a_n]_I) = [\iota(a_n) \cdots \iota(a_1)]_I$. The structure $(\mathbb{M}(A, I), \iota)$ is called a *trace monoid with partial involution*.

5.3 Theories of equations

Let $(\mathcal{M}, \iota)$ be a monoid with partial involution, where $\mathcal{M} = (M, \circ, 1)$ and $\iota : \mathcal{I} \to \mathcal{I}$ for a submonoid $\mathcal{I}$. We can view $(\mathcal{M}, \iota)$ as a relational structure by considering the multiplication $\circ$ as a ternary relation, ι as a binary relation and the constant 1 as a unary relation. Instead of $\iota(x, y)$ (resp. $\circ(x, y, z)$) we write $\iota(x) = y$ (resp. $x \circ y = z$ or briefly $xy = z$). Note that since ι may be only partially defined on M, $\iota(x) = y$ means in fact $x, y \in \mathcal{I} \wedge \iota(x) = y$.

We will also consider extensions $(\mathcal{M}, \iota, (R_i)_{i \in J})$ of the structure $(\mathcal{M}, \iota)$, where R_i is a relation of arbitrary arity over M. In case $\mathcal{C}$ is a class of subsets of M, we also write $(\mathcal{M}, \iota, \mathcal{C}, (R_i)_{i \in J})$ instead of $(\mathcal{M}, \iota, (L)_{L \in \mathcal{C}}, (R_i)_{i \in J})$ and call formulas of the form $x \in L$ for $L \in \mathcal{C}$ *constraints*. If we do not mention ι explicitly in the structure $(\mathcal{M}, \iota, (R_i)_{j \in J})$, then we assume that $\mathcal{I}$ is the group of units of $\mathcal{M}$ and ι is defined by taking inverses (in fact, in this case the predicate $\iota(x) = y$ is equivalent to $xy = yx = z \ \wedge \ z = 1$, hence it can be expressed by the other predicates).

In Section 5.5, we will also consider positive theories: The *positive theory* $\mathrm{posTh}(\mathcal{M}, \iota, (R_i)_{i \in J})$ is the set of all sentences in $\mathrm{FOTh}(\mathcal{M}, \iota, (R_i)_{i \in J})$

that do not use negations, i.e., that are built from atomic predicates using conjunctions, disjunctions, and existential and universal quantifications.

Remark 5.3.1. *Usually the first-order theory of a monoid (with partial involution) is defined by allowing arbitrary equations of the form $u = v$, where u and v are words over the variables, as atomic predicates. But this formulation is easily seen to be equivalent to our definition. Moreover, also constants from $\mathcal{M}$ are usually allowed in equations. We can deal with constants by including them as singleton subsets to the additional relations R_i.*

Note that if $\mathcal{M}$ is finitely generated by Γ, then constants from Γ suffice in order to define all monoid elements of $\mathcal{M}$. We call $\mathrm{FOTh}(\mathcal{M}, \iota, (a)_{a\in\Gamma})$ *the* first-order theory of $(\mathcal{M}, \iota)$ with constants. *On the other hand, the further investigations are not restricted to finitely generated monoids.*

A well-known example of a decidable theory of equations is Presburger's Arithmetic [164]. Translated into our framework, the results of [17] imply the following statement, where $\mathrm{RAT}(\mathbb{N})$ and $\mathrm{RAT}(\mathbb{Z})$ are the classes of *semi-linear sets* in $\mathbb{N}$ and $\mathbb{Z}$, respectively:

Proposition 5.3.2 (cf [17]). *For both $\mathcal{M} = \mathbb{N}$ and $\mathcal{M} = \mathbb{Z}$, the theory* $\mathrm{FOTh}(\mathcal{M}, \mathrm{RAT}(\mathcal{M}))$ *is complete for* $\mathrm{ATIME}(O(n), 2^{2^{O(n)}})$.

Remark 5.3.3. *It is known that* $\mathrm{FOTh}(\{a, b\}^*, a, b)$ *is undecidable [165], in fact already the $\forall\exists^3$-fragment of this theory is undecidable [76, 134]. Together with Presburger's result, it follows that the decidability of the full first-order theory of equations is not preserved under free products. For a restricted class of monoids, we will show such a closure result in Section 5.4.3 for the existential case, even for general graph products.*

The following result can be easily deduced from Proposition 5.3.2, basically because the free product $\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$ of two copies of $\mathbb{Z}/2\mathbb{Z}$ is isomorphic to the semi-direct product of $\mathbb{Z}$ by $\mathbb{Z}/2\mathbb{Z}$.

Corollary 5.3.4. *For $\mathcal{M} = \mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$, the theory* $\mathrm{FOTh}(\mathcal{M}, \mathrm{RAT}(\mathcal{M}))$ *is elementary decidable.*

Proof. Note that $\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/2\mathbb{Z} \cong \mathcal{M}(a, b \mid a^2 = b^2 = 1)$, thus every $s \in \mathcal{M}$ can be represented uniquely as $s = (ab)^i a^j$ where $i \in \mathbb{Z}$ and $j \in \{0, 1\}$ (note that $(ab)^{-1} = ba$ in $\mathcal{M}$). The subgroup K of $\mathcal{M}$ generated by ab is isomorphic to $\mathbb{Z}$. Furthermore let $Q \cong \mathbb{Z}/2\mathbb{Z}$ be the subgroup of $\mathcal{M}$ generated by a. It is

easy to see that $\mathcal{M}$ is the semidirect product of K by Q, thus $\mathcal{M} \simeq \mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$. An isomorphism $h : \mathcal{M} \to \mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$ can be defined by $h((ab)^i a^j) = (i, j)$, where $i \in \mathbb{Z}$ and $j \in \{0, 1\}$. In the following let $h(s) = (n_s, a_s)$. Thus, $st = u$ in $\mathcal{M}$ if and only if $n_u = n_s + (-1)^{a_s} n_t \wedge a_s + a_t \equiv a_u \mod 2$. Furthermore, it is easy to see that if $L \in \mathrm{RAT}(\mathcal{M})$, then $h(L) = L_0 \times \{0\} \cup L_1 \times \{1\}$, where $L_0, L_1 \subseteq \mathbb{Z}$ are semi-linear sets that can be constructed inductively from a rational expression for L, with at most an exponential size increase.

Now given a first-order sentence ϕ we take for every variable x in ϕ two new variables n_x and a_x. Every quantification $\exists x$ is replaced by $\exists n_x \in \mathbb{Z} \bigvee_{a_x \in \{0,1\}}$ (similarly for $\forall$-quantifications). An equation $x = 1$ is replaced by $n_x = 0 \wedge a_x = 0$. An equation $xy = z$ is replaced by $(n_z = n_x + (-1)^{a_x} n_y \wedge a_x + a_y \equiv a_z \mod 2)$. A constraint $x \in L$ with $L \in \mathrm{RAT}(\mathcal{M})$ is replaced by $(n_x \in L_0 \wedge a_x = 0) \vee (n_x \in L_1 \wedge a_x = 1)$ where $h(L) = L_0 \times \{0\} \cup L_1 \times \{1\}$. Finally by substituting for the variables a_x the values 0 and 1 we obtain a Presburger formula over $\mathbb{Z}$. Now the corollary follows from Proposition 5.3.2. □

5.4 Existential theories of graph products

Based on results from [69] for (finitely generated) trace monoids with partial involution (see Section 5.4.1), we will prove in Section 5.4.2 a general preservation theorem for existential theories. In Section 5.4.3 we will use this result in order to show that for a large class of monoids the decidability of the existential theory is preserved under graph products (see Section 4.7 for the definition of graph products). A better upper complexity bound for some special cases is presented in Section 5.4.4.

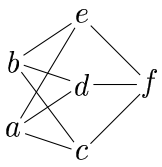
5.4.1 Trace monoids with partial involution

All our decidability results in this chapter are based on the main result from [69]. In order to state this result in its whole generality, we have to introduce the following graph theoretical concept: Let (A, I) be an independence alphabet. We define on A an equivalence relation $\sim_I$ by $a \sim_I b$ if and only if $I(a) = I(b)$. Note that $a \sim_I b$ implies $(a, b) \notin I$: if $I(a) = I(b)$ and $(a, b) \in I$, then also $(a, a) \in I$, which contradict the irreflexivity of I. An equivalence class B of $\sim_I$ is called a *thin clan* of (A, I), if $I(a) \neq \emptyset$ for some (and hence all) $a \in B$ [69]. The cardinality of the set of thin clans of (A, I) is denoted by

$c(A, I)$ – of course it may be infinite. The following facts are easy to verify:

- $c(A, I)$ is bounded by the cardinality of A.
- There exist at most one equivalence class of $\sim_I$, which is not a thin clan. It consists of all the isolated nodes of (A, I).
- The cardinality of a largest clique of (A, I) is at most $\max\{1, c(A, I)\}$.
- $c(A, I) \neq 1$, and $c(A, I) = 0$ if and only if $I = \emptyset$.

For the independence alphabet below, the equivalence classes of $\sim_I$ are $\{a, b\}$, $\{c, d, e\}$, and $\{f\}$, and they are all thin clans.



Now we can state the main result from [69].

Theorem 5.4.1. *For every $k \geq 0$, the following problem is in PSPACE:*

INPUT: A finite independence alphabet (A, I) with $c(A, I) \leq k$, an involution $\iota : B \to B$, $B \subseteq A$, which is compatible with I, and an existential sentence ϕ over the signature of $(\mathbb{M}(A, I), \iota, \mathrm{REC}(\mathbb{M}(A, I)))$ (with ι lifted to $\mathbb{M}(B, I)$).

QUESTION: Does $(\mathbb{M}(A, I), \iota, \mathrm{REC}(\mathbb{M}(A, I))) \models \phi$ hold?

If $c(A, I)$ is not bounded by a constant, then this problem is in EXPSPACE.

A few remarks should be made on Theorem 5.4.1.

- A recognizable set $L \in \mathrm{REC}(\mathbb{M}(A, I))$ has to be represented by a finite automaton for the regular language $\{u \in A^* \mid [u]_I \in L\}$. This is crucial. For instance, if recognizable trace languages are represented by loop-connected automata (see e.g. [147]), then already universality is EXPSPACE-complete for some fixed independence alphabet [147].
- Since every singleton subset belongs to $\mathrm{REC}(\mathbb{M}(A, I))$, constants are implicitly allowed in Theorem 5.4.1.

- In [69], Theorem 5.4.1 is only stated for a completely defined involution $\iota : A \to A$. But if the involution is only defined on $B \subsetneq A$, then we can introduce a new dummy symbol $\overline{a}$ for every $a \in A \backslash B$, extend the involution by $\iota(a) = \overline{a}$ and $\iota(\overline{a}) = a$, and restrict every variable to the original alphabet A, which is a recognizable constraint.

- The uniform EXPSPACE upper bound for the case that $c(A, I)$ is not bounded by a constant is not explicitly stated in the preliminary version [69], but it can be easily derived from the proof in [69].

Theorem 5.4.1 cannot be extended to the case of rational constraints: For $\mathcal{M} = \{a, b\}^* \times \{c, d\}^*$ it is undecidable whether for given $L_1, L_2 \in \mathrm{RAT}(\mathcal{M})$ it holds $L_1 \cap L_2 = \emptyset$, see [1]. A further investigation leads to the following characterization of Muscholl, see [146, Prop. 2.9.2 and 2.9.3].

Proposition 5.4.2. *Let $\mathcal{M} = \mathbb{M}(A, I)$ be a trace monoid with A finite. Then $\exists\mathrm{FOTh}(\mathcal{M}, \mathrm{RAT}(\mathcal{M}))$ is decidable if and only if $\mathcal{M}$ is a free product of free commutative monoids, i.e., $\mathcal{M} = *_{i=1}^{n} \mathbb{N}^{k_i}$ for $n, k_1, \ldots, k_n \in \mathbb{N}$.*

5.4.2 A general preservation theorem

The aim of this section is to prove a general preservation theorem for existential theories. We will apply this result in the next section to existential theories of graph products.

For the further discussion let us fix a set A together with a partial involution ι on A and a countably infinite subset $\mathcal{C} \subseteq 2^A$. Let $\mathbb{A} = (A, \iota, (L)_{L \in \mathcal{C}})$. Moreover, we have given an independence relation $I \subseteq A \times A$ and additional predicates R_j $(1 \leq j \leq m)$ of arbitrary arity on A such that:

(1) ι is compatible with I,

(2) the set $\{I(a) \mid a \in A\}$ is finite,

(3) $\mathrm{dom}(\iota) \subseteq A$ as well as every equivalence class of $\sim_I$ belong to $\mathcal{C}$, and

(4) $\exists\mathrm{FOTh}(\mathbb{A}, (R_j)_{1 \leq j \leq m})$ is decidable.

Due to (1), we can lift ι to a partial involution on $\mathbb{M}(A, I)$. Moreover, (2) and (3) imply that I is definable by a Boolean formula over $(A, (L)_{L \in \mathcal{C}})$, because I is a finite union of Cartesian products of equivalence classes of $\sim_I$.

From the unary predicates in $\mathcal{C}$ we construct a set $\mathcal{L}(\mathcal{C}, I) \subseteq 2^{\mathbb{M}(A,I)}$ as follows: A $\mathcal{C}$-automaton $\mathcal{A}$ is a finite automaton in the usual sense, except that every edge of $\mathcal{A}$ is labeled with some language $L \in \mathcal{C}$. The language $L(\mathcal{A}) \subseteq A^*$ is defined in the obvious way: $a_1 a_2 \cdots a_n \in L(\mathcal{A})$ ($a_i \in A$) if and only if there exists a path $q_0 \xrightarrow{L_1} q_1 \xrightarrow{L_2} q_2 \cdots \xrightarrow{L_{n-1}} q_{n-1} \xrightarrow{L_n} q_n$ in $\mathcal{A}$ such that q_0 is the initial state of $\mathcal{A}$, q_n is a final state of $\mathcal{A}$, and $a_i \in L_i$ for $1 \leq i \leq n$. We say that $\mathcal{A}$ is I-closed if $[u]_I = [v]_I$ and $u \in L(\mathcal{A})$ implies $v \in L(\mathcal{A})$. In the following, we will identify $L(\mathcal{A})$ with the set of traces $\{[u]_I \mid u \in L(\mathcal{A})\}$. Then $L \subseteq \mathbb{M}(A, I)$ belongs to $\mathcal{L}(\mathcal{C}, I)$ if there exists an I-closed $\mathcal{C}$-automaton $\mathcal{A}$ with $L(\mathcal{A}) = L$. In the following, we will briefly write $\mathcal{L}(\mathcal{C})$ instead of $\mathcal{L}(\mathcal{C}, I)$. For effectiveness statements, it is necessary that languages in $\mathcal{C}$ have some finite representation. Then, also languages from $\mathcal{L}(\mathcal{C})$ have a canonical finite representation.

Since $A \subseteq \mathbb{M}(A, I)$, we can view every relation R_j also as a relation over the trace monoid $\mathbb{M}(A, I)$. This is done in the following theorem, which is the main result of this section:

Theorem 5.4.3. *Let $\mathbb{A}$, I, and $(R_j)_{1 \leq j \leq m}$ be as above. Then*

$$\exists\mathrm{FOTh}(\mathbb{M}(A, I), \iota, \mathcal{L}(\mathcal{C}), (R_j)_{1 \leq j \leq m}) \tag{5.1}$$

is decidable. Moreover, if $\exists\mathrm{FOTh}(\mathbb{A}, (R_j)_{j \in J})$ is decidable in NSPACE$(s(n))$, then (5.1) can be decided in NSPACE$(2^{O(n)} + s(n^{O(1)}))$.

Reducing the number of generators

The main difficulty in the proof of Theorem 5.4.3 is to reduce the infinite set of generators of $\mathbb{M}(A, I)$ to a finite set of generators B. In the sequel, we will restrict to some reduct $(A, \iota, (L)_{L \in \mathcal{D}})$, where $\mathcal{D} \subseteq \mathcal{C}$ is finite and contains $\mathrm{dom}(\iota)$ as well as every equivalence class of $\sim_I$. We denote this reduct by $\mathbb{A}$ as well. For the following consideration it is useful to fix some enumeration $L_0, \ldots, L_k$ of $\mathcal{D}$, where $\mathrm{dom}(\iota) = L_0$ and $L_1, \ldots, L_\ell$ is an enumeration of the equivalence classes of $\sim_I$. Thus, $\{L_1, \ldots, L_\ell\}$ is a partition of A. Moreover there exists a fixed independence relation I' on $\{1, \ldots, \ell\}$ such that $I = \bigcup_{(i,j) \in I'} L_i \times L_j$.

Given another structure $\mathbb{B} = (B, \zeta, (K_i)_{0 \leq i \leq k})$ (with ζ an involution on B, $K_i \subseteq B$, and $K_0 = \mathrm{dom}(\zeta)$), a mapping $f : A \to B$ is a *strong homomorphism* from $\mathbb{A}$ to $\mathbb{B}$ if for all $a \in A$:

- $a \in L_i$ if and only if $f(a) \in K_i$ for all $0 \leq i \leq k$ and
- $f(\iota(a)) = \zeta(f(a))$ if $a \in \mathrm{dom}(\iota)$.

Lemma 5.4.4. *We can effectively construct a* finite *structure*

$$\mathbb{B} = (B, \zeta, (K_i)_{0 \leq i \leq k})$$

(with ζ a partial involution on B, $K_i \subseteq B$, and $\mathrm{dom}(\zeta) = K_0$) such that $|B| \leq 2^{k+1}(2^{k+1}+2)$ and there exist strong homomorphisms $f : \mathbb{A} \to \mathbb{B}$ and $g : \mathbb{B} \to \mathbb{A}$ with f surjective.[1]

Proof. First we will define B and $f : A \to B$ such that every L_i is a finite union of preimages $f^{-1}(c)$ ($c \in B$), i.e., f saturates every L_i. Moreover,

(†) $f(a) = f(a')$ and $a, a' \in \mathrm{dom}(\iota)$ will imply $f(\iota(a)) = f(\iota(a'))$, and

(‡) $f(a) = f(\iota(a))$ will imply $a' = \iota(a')$ for some a' with $f(a) = f(a')$.

Figure 5.1 visualizes the construction. For simplicity we assume that $k = 2$. The sets L_1 and L_2 are represented by the left half and lower half, respectively, of the whole square, which represents A. The right half (resp. upper half) represents $A \setminus L_1$ (resp. $A \setminus L_2$), the big inner circle represents $\mathrm{dom}(\iota) = L_0$, and the thin lines represent the partial involution ι on A. The 22 regions that are bounded by thick lines represent the preimages $f^{-1}(b)$ ($b \in B$) and hence the elements of B. Of course, the sets $f^{-1}(b)$ will be infinite in general.

Let $[k] = \{0, \ldots, k\}$. We realize B as a subset of $2^{[k]} \cup (2^{[k]} \times 2^{[k]}) \cup (2^{[k]} \times \{0,1\})$.[2] For a subset $\alpha \subseteq [k]$ define $L^\alpha = \bigcap_{i \in \alpha} L_i \cap \bigcap_{i \notin \alpha} A \setminus L_i$. If $\alpha \subseteq [k]$ is such that $0 \notin \alpha$ (i.e., $L^\alpha \cap \mathrm{dom}(\iota) = \emptyset$) and $L^\alpha \neq \emptyset$, then we put α into B and define the function f on L^α by $f(L^\alpha) = \alpha$. Note that we can check effectively whether $L^\alpha \neq \emptyset$, we just have to decide whether $\mathbb{A} \models \exists x : x \in L^\alpha$. For instance the four outer regions in Figure 5.1 would be represented by $\{1,2\}$, $\{1\}$, $\{2\}$, and $\emptyset$. If $0 \in \alpha$, i.e., $L^\alpha \subseteq \mathrm{dom}(\iota)$, then L^α has to be split into possibly several preimages of f. To represent them in B, take a second subset $\beta \subseteq [k]$ with $0 \in \beta$. In case $\alpha \neq \beta$ we check whether $L^\alpha \cap \iota(L^\beta) \neq \emptyset$, i.e., $\mathbb{A} \models \exists x \in L^\alpha \; \exists y \in L^\beta : x = \iota(y)$. If this is true,

[1] Effectiveness in this context means that given a finite set $\mathcal{D} \subseteq \mathcal{C}$, we can construct the finite structure $\mathbb{B}$ effectively.

[2] The specific representation of B is not really important, we only need some finite representation.

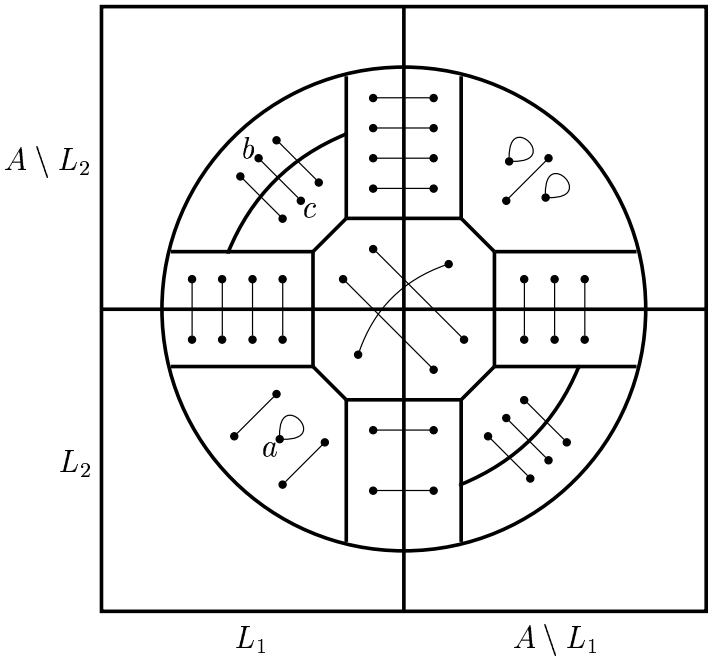


Figure 5.1: The construction from the proof of Lemma 5.4.4

then we put (α, β) and (β, α) into B and define $f(L^\alpha \cap \iota(L^\beta)) = (\alpha, \beta)$ and $f(L^\beta \cap \iota(L^\alpha)) = (\beta, \alpha)$. Now assume that $\alpha = \beta$. We proceed with testing whether $\mathbb{A} \models \exists x \in L^\alpha : x = \iota(x)$. If this holds, then we put (α, α) into B and define $f(L^\alpha \cap \iota(L^\alpha)) = (\alpha, \alpha)$. For instance, the region containing a in Figure 5.1 is represented by $(\{0,1,2\}, \{0,1,2\})$. On the other hand, if $\mathbb{A} \models \neg\exists x \in L^\alpha : x = \iota(x)$, then we check whether $L^\alpha \cap \iota(L^\alpha) \neq \emptyset$, i.e., $\mathbb{A} \models \exists x, y \in L^\alpha : \iota(x) = y$. If this holds, then due to (‡) the set $L^\alpha \cap \iota(L^\alpha)$ has to be split into precisely two preimages C_0 and C_1 of f, where $\iota(a) \in C_i$ for all $a \in C_{1-i}$. These two classes can be represented by the pairs $(\alpha, 0)$ and $(\alpha, 1)$, which we put into B. We set $f(C_i) = (\alpha, i)$. For instance the two regions containing b and $c = \iota(b)$ in Figure 5.1 are represented by $(\{0,1\}, 0)$ and $(\{0,1\}, 1)$ (it does not matter which of the two possible assignments is chosen). This completes the construction of the alphabet B as well as the definition of the surjection f. The size bound $|B| \leq 2^{k+1}(2^{k+1} + 2)$ follows immediately from the construction.

We define the involution ζ on B as follows: If $\alpha, \beta \in 2^{[k]}$ are such that $(\alpha, \beta) \in B$, then we define $\zeta(\alpha, \beta) = (\beta, \alpha)$. If $\alpha \in 2^{[k]}$ is such that $(\alpha, 0), (\alpha, 1) \in B$, then $\zeta(\alpha, i) = (\alpha, 1 - i)$ for $i \in \{0, 1\}$. We define the set

$K_i \subseteq B$ by

$$\begin{aligned} K_i =& \{\alpha \in B \mid \alpha \in 2^{[k]}, i \in \alpha\} \cup \{(\alpha, \beta) \in B \mid \alpha, \beta \in 2^{[k]}, i \in \alpha\} \cup \\ & \{(\alpha, j) \in B \mid \alpha \in 2^{[k]}, j \in \{0,1\}, i \in \alpha\} \end{aligned}$$

This finishes the construction of $\mathbb{B}$. Clearly $K_i = f(L_i)$, $B \setminus K_i = f(A \setminus L_i)$, and $\zeta(f(a)) = f(\iota(a))$, i.e., $f : \mathbb{A} \to \mathbb{B}$ is a strong homomorphism.

We have defined $f : A \to B$ such that if $\zeta(b) = b$, then there exists $a \in f^{-1}(b)$ with $\iota(a) = a$ (see (‡)). This allows to select $g(b) \in f^{-1}(b)$ for every $b \in B$ such that $\iota(g(b)) = g(\zeta(b))$. Moreover, since $g(b) \in f^{-1}(b)$, we have $b \in K_i$ if and only if $g(b) \in L_i$. Thus, $g : \mathbb{B} \to \mathbb{A}$ is a strong homomorphism as well. □

Note that since the strong homomorphism f is surjective in the previous lemma and $\{L_1, \ldots, L_\ell\}$ is a partition of A, also $\{K_1, \ldots, K_\ell\}$ is a partition of B.

Now assume that we have given a third structure $\mathbb{C} = (C, \xi, (\Lambda_i)_{0 \leq i \leq k})$, where C is finite, ξ is a partial involution on C, $\Lambda_i \subseteq C$ for $0 \leq i \leq k$, $\mathrm{dom}(\xi) = \Lambda_0$, and $\{\Lambda_1, \ldots, \Lambda_\ell\}$ is a partition of C (with $\Lambda_i = \emptyset$ allowed). In the sequel, an *embedding of* $\mathbb{C}$ *in* $\mathbb{A}$ is an injective strong homomorphism $h : \mathbb{C} \to \mathbb{A}$.

For the independence relation I let $f(I) = \{(f(a), f(a')) \mid (a, a') \in I\}$ in the following lemma.

Lemma 5.4.5. *Given* $\mathbb{C}$ *as above, we can effectively construct a* finite *structure* $\mathbb{B} = (B, \zeta, (K_i)_{0 \leq i \leq k})$ *(with* ζ *a partial involution on* B*,* $K_i \subseteq B$*, and* $\mathrm{dom}(\zeta) = K_0$*) together with an independence relation* $J \subseteq B \times B$ *such that:*

- $C \subseteq B$*,*
- $|B| \leq 2^{k+1}(2^{k+1} + 2) + |C|$*,*
- ζ *is compatible with* J*, and*
- *for every embedding* $h : \mathbb{C} \to \mathbb{A}$ *there exist strong homomorphisms* $f : \mathbb{A} \to \mathbb{B}$ *and* $g : \mathbb{B} \to \mathbb{A}$ *such that* $f(I) \subseteq J$*,* $g(J) \subseteq I$*, and* $f(h(c)) = c$*,* $g(c) = h(c)$ *for all* $c \in C$*.*

Proof. By Lemma 5.4.4 we can effectively construct a finite structure $\mathbb{B}' = (B', \zeta', (K'_i)_{0 \leq i \leq k})$ such that $\mathrm{dom}(\zeta') = K'_0$, $|B'| \leq 2^{k+1}(2^{k+1} + 2)$, and there

exist strong homomorphisms $f' : \mathbb{A} \to \mathbb{B}'$ and $g' : \mathbb{B}' \to \mathbb{A}$ with f' surjective. Note that $\{K'_1, \ldots, K'_\ell\}$ must be a partition of B'. Now we define the structure $\mathbb{B} = (B, \zeta, (K_i)_{0 \le i \le k})$ by $B = B' \dot\cup C$, $\zeta = \zeta' \dot\cup \xi$, and $K_i = K'_i \dot\cup \Lambda_i$ for $0 \le i \le k$. The given size bound for $|B|$ in the lemma follows from $|B'| \le 2^{k+1}(2^{k+1}+2)$. Since $\{K_1, \ldots, K_\ell\}$ is a partition of B, we can define the independence relation J on B by $J = \bigcup_{(i,j)\in I'} K_i \times K_j$.

Given an embedding $h : \mathbb{C} \to \mathbb{A}$, we define $f : A \to B$ by $f(h(c)) = c$ for $c \in C$ (since h is injective, this is well-defined) and $f(a) = f'(a)$ for $a \in A \setminus h(C)$. We define $g : B \to A$ by $g(b) = g'(b)$ for $b \in B'$ and $g(c) = h(c)$ for $c \in C$. Since $h : \mathbb{C} \to \mathbb{A}$ and $f' : \mathbb{A} \to \mathbb{B}'$ are strong homomorphisms, the following properties are easy to verify for all $a \in A$ and $b \in B = B' \cup C$:

- $a \in L_i$ if and only if $f(a) \in K_i$ and $b \in K_i$ if and only if $g(b) \in L_i$.
- $f(\iota(a)) = \zeta(f(a))$ and $g(\zeta(b)) = \iota(g(b))$ (for the first identity note that $a \in h(C)$ implies $\iota(a) \in h(C)$).

Thus, $f : \mathbb{A} \to \mathbb{B}$ and $g : \mathbb{B} \to \mathbb{A}$ are strong homomorphisms with $f(h(c)) = c$ and $g(c) = h(c)$ for all $c \in C$. Moreover, since $I = \bigcup_{(i,j)\in I'} L_i \times L_j$ and $J = \bigcup_{(i,j)\in I'} K_i \times K_j$, the first point above implies that $(a, a') \in I$ if and only if $(f(a), f(a')) \in J$ and $(b, b') \in J$ if and only if $(g(b), g(b')) \in I$. In particular, $f(I) \subseteq J$ and $g(J) \subseteq I$.

In order to see that ζ is compatible with J assume that $(a, b) \in J$ and $a, b \in \mathrm{dom}(\zeta)$. Then $(g(a), g(b)) \in I$ and $g(a), g(b) \in \mathrm{dom}(\iota)$. Since ι is compatible with I, we obtain $(\iota(g(a)), \iota(g(b))) = (g(\zeta(a)), g(\zeta(b))) \in I$. Hence, $(\zeta(a), \zeta(b)) \in J$. □

Proof of Theorem 5.4.3

For the proof of Theorem 5.4.3 let us take a Boolean formula θ over the signature of $(\mathbb{M}(A, I), \iota, \mathcal{L}(\mathcal{C}), (R_j)_{1 \le j \le m})$. We have to decide whether θ is satisfiable in the structure, $(\mathbb{M}(A, I), \iota, \mathcal{L}(\mathcal{C}), (R_j)_{1 \le j \le m})$. For this, we will present a nondeterministic algorithm that constructs a finitely generated trace monoid with partial involution $(\mathbb{M}(B, J), \zeta)$ and a Boolean formula ϕ' over the signature of $(\mathbb{M}(B, J), \zeta, \mathrm{REC}(\mathbb{M}(B, J)))$ such that θ is satisfiable in $(\mathbb{M}(A, I), \iota, \mathcal{L}(\mathcal{C}), (R_j)_{1 \le j \le m})$ if and only if for at least one outcome of our nondeterministic algorithm, ϕ' is satisfiable in $(\mathbb{M}(B, J), \zeta, \mathrm{REC}(\mathbb{M}(B, J)))$. This allows to apply Theorem 5.4.1.

Assume that every $\mathcal{C}$-automaton in θ only uses sets among the finite set $\mathcal{D} \subseteq \mathcal{C}$. Assume that also $\mathrm{dom}(\iota)$ as well as every $\sim_I$-equivalence class belongs to $\mathcal{D}$. Let $\mathcal{D} = \{L_0, \ldots, L_k\}$, where $L_0 = \mathrm{dom}(\iota)$ and $L_1, \ldots, L_\ell$ is an enumeration of the $\sim_I$-equivalence classes of (A, I). Note that $k \in O(|\theta|)$.

First we may push negations to the level of atomic subformulas in θ. Moreover, disjunctions may be eliminated by nondeterministically guessing one of the two corresponding disjuncts. Thus, we may assume that θ is a conjunction of atomic predicates and negated atomic predicates. We replace every negated equation $xy \neq z$ by $xy = z' \wedge z \neq z'$, where z' is a new variable. Similarly an equation $\iota(x) \neq y$ is replaced by $\iota(x) = z \wedge z \neq y$. Thus, we may assume that all negated predicates in θ are of the form $x \neq y$, $x \notin L$, and $\neg R_j(x_1, \ldots, x_n)$ for variables x, y, x_i and $L \in \mathcal{L}(\mathcal{D})$.

We can write θ as a conjunction $\phi \wedge \psi$, where ψ contains all predicates of the form $(\neg)R_j(x_1, \ldots, x_n)$. Let $x \neq y$ be a negated equation in ϕ, where x and y are variables. Since $x \neq y$ is interpreted in the trace monoid $\mathbb{M}(A, I)$, we can replace $x \neq y$ by either

$$x = zau \ \wedge \ y = zbv \ \wedge \ a, b \in L \ \wedge \ a \neq b \quad \text{or}$$

$$x = zu \ \wedge \ y = zv \ \wedge \ u \in L\,\mathbb{M}(A, I) \ \wedge \ v \notin L\,\mathbb{M}(A, I),$$

where $L \in \mathcal{D}$ is an equivalence class of $\sim_I$ that is guessed nondeterministically. In the first case, we add $a, b \in L \ \wedge \ a \neq b$ to the "$\mathbb{A}$-local" part ψ. In the second case, we have to construct an I-closed $\mathcal{D}$-automaton for $L\,\mathbb{M}(A, I)$, which is easy, since all $\sim_I$-equivalence classes belong to $\mathcal{D}$. Thus, in the sequel we may assume that ϕ does not contain negated equations.

So far, we have obtained a conjunction $\phi \ \wedge \ \psi$, where ϕ is interpreted in $(\mathbb{M}(A, I), \iota, \mathcal{L}(\mathcal{D}))$ and ψ is interpreted in the base structure $(\mathbb{A}, (R_j)_{1 \leq j \leq m})$. The formula ϕ does not contain negated equations. Let Ξ be the set of all variables that occur in $\phi \ \wedge \ \psi$, and let $\Omega \subseteq \Xi$ contain all variables that occur in the $\mathbb{A}$-local part ψ. Thus, all variables from Ω are implicitly restricted to $A \subseteq \mathbb{M}(A, I)$. Note that variables from Ω may of course also occur in ϕ. In case ϕ contains a constraint $x \in L$ with $L \in \mathcal{L}(\mathcal{D})$ and $x \in \Omega$, then we can guess $L' \in \mathcal{D}$ with $L \cap L' \neq \emptyset$ and replace $x \in L$ by the constraint $x \in L'$, which will be shifted to ψ. Hence, we may assume that for every constraint $x \in L$ that occurs in ϕ, we have $x \in \Xi \setminus \Omega$.

Next, for every variable $x \in \Omega$ we guess whether $x \in L_0 = \mathrm{dom}(\iota)$ or $x \notin \mathrm{dom}(\iota)$ holds and add the corresponding (negated) constraint to ψ. In case $x \in \mathrm{dom}(\iota)$ was guessed, we add a new variable $\overline{x}$ to Ω and add the

equation $\iota(x) = \overline{x}$ to ψ. Next, we guess for all different variables $x, y \in \Omega$ (here Ω refers to the new set of variables including the added copys $\overline{x}$), whether $x = y$ or $x \neq y$. In case $x = y$ is guessed, we can eliminate for instance y. Thus, we may assume that for all different variables $x, y \in \Omega$ the negated equation $x \neq y$ belongs to ψ. Finally, for every set L_i with $1 \leq i \leq k$ and every $x \in \Omega$ we guess whether $x \in L$ or $x \notin L$ holds and add the corresponding constraint to ψ. We denote the resulting formula by ψ as well.

Most of the guessed formulas ψ will be not satisfiable in $(\mathbb{A}, (R_j)_{1 \leq j \leq m})$ (e.g., if $L_i \cap L_j = \emptyset$ and the constraints $x \in L_i$ and $x \in L_j$ were guessed). But since $\exists\mathrm{FOTh}(\mathbb{A}, (R_j)_{1 \leq j \leq m})$ is decidable, we can effectively check whether the guessed formula ψ is satisfiable. If it is not satisfiable, then we reject on the corresponding computation path. Let us fix a specific guess, which results in a satisfiable formula ψ, for the further consideration.

Now we define a finite structure $\mathbb{C} = (\widetilde{\Omega}, \xi, (\Lambda_i)_{0 \leq i \leq k})$ as follows: Let $\widetilde{\Omega} = \{\widetilde{x} \mid x \in \Omega\}$ be a disjoint copy of the set of variables Ω. For $0 \leq i \leq k$ let Λ_i be the set of all $\widetilde{x} \in \widetilde{\Omega}$ such that $x \in L_i$ belongs to ψ. Finally, we define the partial involution ξ on $\widetilde{\Omega}$ as follows: The domain of ξ is Λ_0 and $\xi(\widetilde{x}) = \widetilde{y}$ in case $\iota(x) = y$ or $\iota(y) = x$ belongs to the conjunction ψ. Since ψ is satisfiable and $\{L_1 \ldots, L_\ell\}$ is a partition of A, it follows that $\{\Lambda_1 \ldots, \Lambda_\ell\}$ is a partition of $\widetilde{\Omega}$ (with $\Lambda_i = \emptyset$ allowed). Thus, $\mathbb{C}$ satisfies all the requirements from Lemma 5.4.5 and we can apply Lemma 5.4.5 to the structures $\mathbb{A}$ and $\mathbb{C}$. Hence, from $\mathbb{C}$ we can effectively determine a finite structure $\mathbb{B} = (B, \zeta, (K_i)_{0 \leq i \leq k})$ together with an independence relation $J \subseteq B \times B$ such that $\widetilde{\Omega} \subseteq B$, ζ is compatible with J, and for every embedding $h : \mathbb{C} \to \mathbb{A}$ there exist strong homomorphisms $f : \mathbb{A} \to \mathbb{B}$ and $g : \mathbb{B} \to \mathbb{A}$ with $f(I) \subseteq J$, $g(J) \subseteq I$, and $f(h(\widetilde{x})) = \widetilde{x}$, $g(\widetilde{x}) = h(\widetilde{x})$ for every $x \in \Omega$. We also obtain a size bound of $|\widetilde{\Omega}| + 2^{O(k)} \subseteq 2^{O(|\theta|)}$ for $|B|$. We denote the lifting of ζ to $\mathbb{M}(B, J)$ by ζ as well.

Recall that we have to check whether there exist assignments $\kappa : \Omega \to A$ and $\lambda : \Xi \setminus \Omega \to \mathbb{M}(A, I)$ such that κ satisfies ψ in $(\mathbb{A}, (R_j)_{1 \leq j \leq m})$ and $\kappa \cup \lambda$ satisfies ϕ in $(\mathbb{M}(A, I), \iota, \mathcal{L}(\mathcal{D}))$. We have already verified that the conjunction ψ is satisfiable in $(\mathbb{A}, (R_j)_{1 \leq j \leq m})$. For the following consideration let us fix an arbitrary assignment $\kappa : \Omega \to A$ that satisfies ψ in $(\mathbb{A}, (R_j)_{1 \leq j \leq m})$.[3] Then κ defines an embedding $h : \mathbb{C} \to \mathbb{A}$ by $h(\widetilde{x}) = \kappa(x)$ for $x \in \Omega$. Therefore there exist strong homomorphisms $f : \mathbb{A} \to \mathbb{B}$ and $g : \mathbb{B} \to \mathbb{A}$ with

[3] We do not have to determine this assignment explicitly, only its existence is important.

$f(\kappa(x)) = \widetilde{x}$ and $g(\widetilde{x}) = \kappa(x)$ for every $x \in \Omega$ and all the other properties from Lemma 5.4.5. Since f and g preserve the involution on A and B, respectively, and $f(I) \subseteq J$, $J \subseteq g(I)$, we obtain the following homomorphisms between trace monoids with partial involution:

- $f : (\mathbb{M}(A, I), \iota) \to (\mathbb{M}(B, J), \zeta)$
- $g : (\mathbb{M}(B, J), \zeta) \to (\mathbb{M}(A, I), \iota)$.

Given a $\mathcal{D}$-automaton $\mathcal{A}$, we define a new automaton $\mathcal{A}'$ by replacing every edge $p \xrightarrow{L_i} q$ in $\mathcal{A}$ by $p \xrightarrow{K_i} q$ (and changing nothing else). Recall that $K_i \subseteq B$. Since $\mathcal{A}$ is I-closed, $\mathcal{A}'$ is easily seen to be J-closed. Moreover, since B is finite, $L(\mathcal{A}') \subseteq \mathbb{M}(B, J)$ is a recognizable trace language. Recall that for every $0 \leq i \leq k$, we have $a \in L_i$ if and only if $f(a) \in K_i$ and $b \in K_i$ if and only if $g(b) \in L_i$. Thus, the following statement is obvious:

Lemma 5.4.6. *Let $t \in \mathbb{M}(A, I)$ and $u \in \mathbb{M}(B, J)$:*

- *$t \in L(\mathcal{A})$ if and only if $f(t) \in L(\mathcal{A}')$.*
- *$u \in L(\mathcal{A}')$ if and only if $g(u) \in L(\mathcal{A})$.*

Next, we transform the conjunction ϕ into a conjunction ϕ', which will be interpreted over $(\mathbb{M}(B, J), \zeta, \mathrm{REC}(\mathbb{M}(B, J)))$, by replacing in ϕ every occurrence of a variable $x \in \Omega$ by the constant $\widetilde{x} \in \widetilde{\Omega} \subseteq B$. Thus, ϕ' contains constants from $\widetilde{\Omega}$ and variables from $\Xi \backslash \Omega$, which range over the trace monoid $\mathbb{M}(B, J)$. Moreover, every constraint $x \in L(\mathcal{A})$ (resp. $x \notin L(\mathcal{A})$) in ϕ is replaced by $x \in L(\mathcal{A}')$ (resp. $x \notin L(\mathcal{A}')$) (note that $x \in \Xi \setminus \Omega$). Thus, all constraint languages in ϕ' are recognizable trace languages.

Lemma 5.4.7. *The following two statements are equivalent:*

(1) There exists an assignment $\lambda : \Xi \backslash \Omega \to \mathbb{M}(A, I)$ such that $\kappa \cup \lambda$ satisfies ϕ in $(\mathbb{M}(A, I), \iota, \mathcal{L}(\mathcal{D}))$.

(2) There exists an assignment $\lambda' : \Xi \backslash \Omega \to \mathbb{M}(B, J)$ that satisfies ϕ' in $(\mathbb{M}(B, J), \zeta, \mathrm{REC}(\mathbb{M}(B, J)))$.

Proof. First, assume that (1) holds. We claim that (2) holds with $\lambda' = f \circ \lambda$. Consider a constraint $x \in L(\mathcal{A}')$ (resp. $x \notin L(\mathcal{A}')$) of ϕ'. Then $x \in \Xi \setminus \Omega$ and $x \in L(\mathcal{A})$ (resp. $x \notin L(\mathcal{A})$) is a constraint of ϕ. Thus, $\lambda(x) = (\kappa \cup$

$\lambda)(x) \in L(\mathcal{A})$ (resp. $\lambda(x) \notin L(\mathcal{A})$), which implies $\lambda'(x) = f(\lambda(x)) \in L(\mathcal{A}')$ (resp. $\lambda'(x) \notin L(\mathcal{A}')$) by Lemma 5.4.6. Now let $u' = v'$ be an equation of ϕ', which results from the equation $u = v$ of ϕ. The only difference between $u = v$ and $u' = v'$ is that every occurrence of every variable $x \in \Omega$ in $u = v$ is replaced by the constant $\widetilde{x}$ in $u' = v'$. The assignment $\kappa \cup \lambda$ is a solution of $u = v$ in $(\mathbb{M}(A, I), \iota)$. Since f is a homomorphism between trace monoids with partial involution, $f \circ (\kappa \cup \lambda) = f \circ \kappa \cup f \circ \lambda = f \circ \kappa \cup \lambda'$ is a solution of $u = v$ in $(\mathbb{M}(B, J), \zeta)$. Since $f(\kappa(x)) = \widetilde{x}$ for every $x \in \Omega$, the mapping λ' is a solution of $u' = v'$ in $(\mathbb{M}(B, J), \zeta)$.

Now assume that (2) holds. We claim that (1) holds with $\lambda = g \circ \lambda'$. Let $x \in L(\mathcal{A})$ (resp. $x \notin L(\mathcal{A})$) be a constraint of ϕ. Then $x \in \Xi \setminus \Omega$ and $x \in L(\mathcal{A}')$ (resp. $x \notin L(\mathcal{A}')$) is a constraint of ϕ', hence $\lambda'(x) \in L(\mathcal{A}')$ (resp. $\lambda'(x) \notin L(\mathcal{A}')$). Lemma 5.4.6 implies that $\lambda(x) = g(\lambda'(x)) \in L(\mathcal{A})$ (resp. $\lambda(x) \notin L(\mathcal{A})$). Now consider an equation $u = v$ of ϕ and let $u' = v'$ be the corresponding equation of ϕ'. Thus, λ' is a solution of $u' = v'$ in $(\mathbb{M}(B, J), \zeta)$. Let the function π map every variable $x \in \Omega$ to the constant $\widetilde{x} \in B$. By construction of $u' = v'$, $\lambda' \cup \pi$ is a solution of $u = v$ in $(\mathbb{M}(B, J), \zeta)$. Since g is a homomorphism between trace monoids with partial involution and $g(\pi(x)) = \kappa(x)$ for every $x \in \Omega$, the mapping $g \circ (\lambda' \cup \pi) = \lambda \cup \kappa$ is a solution of $u = v$ in $(\mathbb{M}(A, I), \iota)$. □

For the previous lemma it is crucial that the conjunction ϕ does not contain negated equations, because the homomorphisms f and g are not injective in general, and therefore do not preserve inequalities.

Since Lemma 5.4.7 holds for every $\kappa : \Omega \to A$ that satisfies ψ in the structure $(\mathbb{A}, (R_j)_{1 \leq j \leq m})$, and we already know that such an assignment exists, it only remains to check whether $(\mathbb{M}(B, J), \zeta, \mathrm{REC}(\mathbb{M}(B, J))) \models \phi'$. By Theorem 5.4.1 this can be done effectively. This finishes the proof of Theorem 5.4.3.

For the upper complexity bound in Theorem 5.4.3 one has to notice the following two points:

- The size of the new alphabet B is bounded by $2^{O(|\theta|)}$ and size of the formula ϕ' is bounded by $|\theta|^{O(1)}$, where θ is the initial formula. Moreover, $c(B, J) = c(A, I)$, where the latter is a fixed finite constant. This allows to apply the complexity statement from Theorem 5.4.1 in order to check in $\mathrm{NSPACE}(2^{O(|\theta|)})$ whether ϕ' is satisfiable in $(\mathbb{M}(B, J), \zeta, \mathrm{REC}(\mathbb{M}(B, J)))$.

- During the construction of B and ϕ', we had to check the validity of existential formulas of size $|\theta|^{O(1)}$ in the structure $(\mathbb{A}, (R_j)_{1\leq j\leq m})$, which can be done in NSPACE$(s(|\theta|^{O(1)}))$.

5.4.3 Closure under graph products

In this section we will apply Theorem 5.4.3 in order to show that under some restrictions, the decidability of the existential theory is preserved by graph products. Concerning graph products we will use the notation from Section 4.7 in the following.

We fix a graph product $\mathbb{P} = \mathbb{P}(\Sigma, I_\Sigma, (\mathcal{M}_\sigma)_{\sigma\in\Sigma})$ for the further discussion, where $\mathcal{M}_\sigma = (M_\sigma, \circ_\sigma, 1_\sigma)$. Let $A_\sigma = M_\sigma \setminus \{1_\sigma\}$ and define

$$A = \bigcup_{\sigma\in\Sigma} A_\sigma \quad \text{and} \quad I = \bigcup_{(\sigma,\tau)\in I_\Sigma} A_\sigma \times A_\tau,$$

where w.l.o.g. $A_\sigma \cap A_\tau$ for $\sigma \neq \tau$. In Section 4.7 we have defined the trace rewriting system

$$R = \bigcup_{\sigma\in\Sigma} \{ab \to c \mid a, b, c \in A_\sigma, a \circ_\sigma b = c\} \cup \{ab \to \varepsilon \mid a, b \in A_\sigma, a \circ_\sigma b = 1_\sigma\}$$

over $\mathbb{M}(A, I)$. We have seen that R is confluent (Lemma 4.7.1) and that $\mathbb{P}$ is in one-to-one correspondence with $\mathrm{IRR}(R) \subseteq \mathbb{M}(A, I)$. Let $U_\sigma \subseteq \mathcal{M}_\sigma$ be the group of units of $\mathcal{M}_\sigma$ and let $\mathbb{U}$ be the set of units of $\mathbb{P}$. It is easy to see that $\mathbb{U} = \mathbb{P}(\Sigma, I_\Sigma, (U_\sigma)_{\sigma\in\Sigma})$. On $V_\sigma = U_\sigma \setminus \{1_\sigma\} \subseteq A_\sigma$ we define an involution $\iota_\sigma : V_\sigma \to V_\sigma$ by taking the inverse in the group U_σ. Let $V = \bigcup_{\sigma\in\Sigma} V_\sigma$ and $\iota = \bigcup_{\sigma\in\Sigma} \iota_\sigma$. Then the involution $\iota : V \to V$ is compatible with I, hence it can be extended to a partial monoid involution $\iota : \mathbb{M}(V, I) \to \mathbb{M}(V, I)$. Moreover, the set of traces $\mathbb{M}(V, I) \cap \mathrm{IRR}(R)$ is in one-to-one correspondence with the group of units $\mathbb{U}$.

Constraints

Our announced closure result will also include constraints. In this paragraph we present a general construction that defines a class of constraints in the graph product $\mathbb{P}$, starting from a constraint class for every factor monoid $\mathcal{M}_\sigma$. The following construction is inspired by [177, 178].

For every $\sigma \in \Sigma$ let $\mathcal{C}_\sigma \subseteq 2^{M_\sigma}$ be a class of languages and let $\mathcal{D}_\sigma = \{L \setminus \{1_\sigma\} \mid L \in \mathcal{C}_\sigma\}$. It is not required that $\mathcal{D}_\sigma \subseteq \mathcal{C}_\sigma$. Let $\mathcal{C} = \bigcup_{\sigma\in\Sigma} \mathcal{C}_\sigma$ and

$\mathcal{D} = \bigcup_{\sigma\in\Sigma} \mathcal{D}_\sigma \subseteq 2^A$. Recall the definition of the class $\mathcal{L}(\mathcal{D}, I) = \mathcal{L}(\mathcal{D}) \subseteq 2^{\mathbb{M}(A,I)}$ from Section 5.4.2. We define the class $\mathcal{IL}(\mathcal{C}, I, R) \subseteq 2^{\mathbb{M}(A,I)}$ by

$$\mathcal{IL}(\mathcal{C}, I, R) = \{L \cap \mathrm{IRR}(R) \mid L \in \mathcal{L}(\mathcal{D}, I)\}.$$

In the following, we will briefly write $\mathcal{IL}(\mathcal{C})$ for $\mathcal{IL}(\mathcal{C}, I, R)$. Using the one-to-one correspondence between $\mathbb{P}$ and $\mathrm{IRR}(R)$, we may view $L \cap \mathrm{IRR}(R)$ also as a subset of $\mathbb{P}$, hence $\mathcal{IL}(\mathcal{C}) \subseteq 2^{\mathbb{P}}$.

The main result

Throughout this section we will assume that the following two requirements hold:

Assumption 5.4.8. *If $a, b \in M_\sigma$ satisfy $a \circ_\sigma b = 1_\sigma$, then $a, b \in U_\sigma$.*

For example groups, free monoids, and finite monoids satisfy all this requirement,[4] whereas $\mathcal{M}(a, b \mid ab = \varepsilon)$ does not.

Assumption 5.4.9. *$\exists\mathrm{FOTh}(\mathcal{M}_\sigma, \mathcal{C}_\sigma)$ is decidable and $U_\sigma \in \mathcal{C}_\sigma$.*

The following theorem is the main result of this section.

Theorem 5.4.10. *Let (Σ, I_Σ) be a finite independence alphabet. Let $\mathcal{M}_\sigma$ be a monoid and $\mathcal{C}_\sigma \subseteq 2^{\mathcal{M}_\sigma}$ be a class of languages such that Assumption 5.4.8 and Assumption 5.4.9 hold. Then, for $\mathcal{C} = \bigcup_{\sigma\in\Sigma} \mathcal{C}_\sigma$,*

$$\exists\mathrm{FOTh}(\mathbb{P}(\Sigma, I_\Sigma, (\mathcal{M}_\sigma)_{\sigma\in\Sigma}), \mathcal{IL}(\mathcal{C})) \tag{5.2}$$

is also decidable. Moreover, if each of the theories $\exists\mathrm{FOTh}(\mathcal{M}_\sigma, \mathcal{C}_\sigma)$ is in NSPACE$(s(n))$, then (5.2) can be decided in NSPACE$(2^{O(n)} + s(n^{O(1)}))$.

Before we go into the details of the proof of Theorem 5.4.10 let us first present an application. The existential theory of a finite monoid is decidable for trivial reasons. By Makanin's result, the existential theory with constants of a free monoid is also decidable. Finally, by [184], also the existential theory with constants of a torsion-free hyperbolic group is decidable.[5] Note

[4] For a finite monoid note that $a \circ b = 1$ implies that the mapping $x \mapsto b \circ x$ is injective, hence it is surjective. Thus, there exists c with $b \circ c = 1$.

[5] Rips and Sela have shown in [168] that it is decidable whether a word equation is solvable over a torsion-free hyperbolic group. In [184], Sela extended the approach of [168] such that also negated equations can be handled.

that every free group is torsion-free hyperbolic. Since finite monoids, free monoids, and groups in general all satisfy Assumption 5.4.8, we obtain the following corollary:

Corollary 5.4.11. *Let $\mathbb{P}$ be a graph product of finite monoids, free monoids, and torsion-free hyperbolic groups, and let Γ be a finite generating set for $\mathbb{P}$. Then $\exists\mathrm{FOTh}(\mathbb{P}, (a)_{a\in\Gamma})$ is decidable.*

We begin the proof of Theorem 5.4.10 with a few simple observations.

Assumption 5.4.8 implies that if $a \circ_\sigma b = 1_\sigma$ for $a, b \in A_\sigma$ in the monoid $\mathcal{M}_\sigma$, then $a, b \in V_\sigma$ and $\iota_\sigma(a) = a^{-1} = b$, thus

$$R = \bigcup_{\sigma\in\Sigma} \{ab \to c \mid a, b, c \in A_\sigma, a \circ_\sigma b = c\} \cup \{a(a^{-1}) \to \varepsilon \mid a \in V_\sigma\}.$$

We may assume that $M_\sigma \in \mathcal{C}_\sigma$, i.e., $A_\sigma \in \mathcal{D}$, for every $\sigma \in \Sigma$ without violating Assumption 5.4.9.[6] Hence, since every equivalence class of $\sim_I$ is a union of some of the A_σ, we may assume that these classes belong to $\mathcal{D}$ as well. Finally, since $V_\sigma \in \mathcal{D}$ and $V = \bigcup_{\sigma\in\Sigma} V_\sigma$, also $V \in \mathcal{D}$ may be assumed.

Note that $A_\sigma \in \mathcal{D}$ also implies that $\mathrm{IRR}(R) \in \mathcal{L}(\mathcal{D})$: The closure properties of recognizable trace languages (Section2.8) imply that the set $K = \bigcup_{\sigma\in\Sigma} \mathbb{M}(\Sigma, I_\Sigma)\sigma\sigma\mathbb{M}(\Sigma, I_\Sigma)$ belongs to $\mathrm{REC}(\mathbb{M}(\Sigma, I_\Sigma))$. Thus, $L = \Sigma^* \setminus \{u \in \Sigma^* \mid [u]_{I_\Sigma} \in K\} \subseteq \Sigma^*$ is a regular language. In order to define a $\mathcal{D}$-automaton for $\mathrm{IRR}(R)$, we just have to replace in a finite automaton for L every label σ by A_σ.

It follows that every constraint $x \in \mathcal{IL}(\mathcal{C})$ can be written as $x \in L_1 \wedge x \in L_2$ with $L_1, L_2 \in \mathcal{L}(\mathcal{D})$.

Isolating the structure of the $\mathcal{M}_\sigma$

In this paragraph we finish the proof of Theorem 5.4.10. Assume that for every $\sigma \in \Sigma$ the theory $\exists\mathrm{FOTh}(\mathcal{M}_\sigma, \mathcal{C}_\sigma)$ is decidable in $\mathrm{NSPACE}(s(n))$. Then the same holds for the theory $\exists\mathrm{FOTh}(A_\sigma, \circ_\sigma, \iota_\sigma, (L)_{L\in\mathcal{D}_\sigma})$, where $\circ_\sigma$ is considered as a tenary relation that is restricted to A_σ. Let $\mathbb{A} = (A, \iota, (L)_{L\in\mathcal{D}})$. We obtain that also $\exists\mathrm{FOTh}(\mathbb{A}, (\circ_\sigma)_{\sigma\in\Sigma})$ is decidable in $\mathrm{NSPACE}(s(n))$. Now we apply Theorem 5.4.3 to the structure $\mathbb{A}$ together with the independence

[6] Note that a constraint of the form $x \in U_\sigma$ could be eliminated by $\exists y : x \circ_\sigma y = 1_\sigma$, but this is not possible for constraints $x \notin U_\sigma$, since we would introduce a universal quantifier in this way. Therefore we assume explicitly that $U_\sigma \in \mathcal{C}_\sigma$.

relation I and the additional relations $\circ_\sigma$. Clearly, $(\mathbb{A}, (\circ_\sigma)_{\sigma\in\Sigma})$ and I satisfy the requirements from Section 5.4.2. It follows that

$$\exists\mathrm{FOTh}(\mathbb{M}(A,I), \iota, \mathcal{L}(\mathcal{D}), (\circ_\sigma)_{\sigma\in\Sigma})$$

is decidable in $\mathrm{NSPACE}(2^{O(n)} + s(n^{O(1)}))$.

Let θ be a Boolean formula with atomic predicates of the form $xy = z$ and $x \in L$, where $L \in \mathcal{IL}(\mathcal{C})$.[7] We have to check, whether there exists an assignment for the variables in θ to elements in $\mathbb{P}$ that satisfies θ.

The rest of the section shows that θ can be transformed in polynomial time into an equivalent existential statement over $(\mathbb{M}(A,I), \iota, \mathcal{L}(\mathcal{D}), (\circ_\sigma)_{\sigma\in\Sigma})$. Thus, in some sense we isolate the structure of the factor monoids $\mathcal{M}_\sigma$ into the "$\mathcal{M}_\sigma$-local" $\circ_\sigma$-predicates.

First, we may push negations to the level of atomic subformulas in θ. We replace every negated equation $xy \neq z$ by $xy = z' \wedge z \neq z'$, where z' is a new variable. Thus, we may assume that all negated predicates in θ are of the form $x \neq y$ and $x \notin L$ for variables x and y.

Recall that $\mathbb{P} \cong \mathbb{M}(A,I)/\overset{*}{\leftrightarrow}_R$ and that R is confluent and terminating. Hence, if $\omega(s)$ denotes the unique trace from $\mathrm{IRR}(R)$ that represents $s \in \mathbb{P}$, then for all $s, t, u \in \mathbb{P}$ and $L \in \mathcal{IL}(\mathcal{C})$, we have:

- $s = t$ if and only if $\omega(s) = \omega(t)$,
- $st = u$ in $\mathbb{P}$ if and only if $\omega(s)\omega(t) \overset{*}{\to}_R \omega(u)$, and
- $s \in L$ if and only if $\omega(s) \in L$.

For the last point note that $\mathcal{IL}(\mathcal{C})$, viewed as a subset of $2^{\mathbb{M}(A,I)}$, is contained in $2^{\mathrm{IRR}(R)}$.

Hence, if we add for every variable x in θ the constraint $x \in \mathrm{IRR}(R)$ (recall that $\mathrm{IRR}(R) \in \mathcal{L}(\mathcal{D})$) and replace every equation $xy = z$ in θ by the rewriting constraint $xy \overset{*}{\to}_R z$, then we obtain a formula, which is satisfiable in the trace monoid $\mathbb{M}(A,I)$ if and only if the original formula θ is satisfiable in $\mathbb{P}$. Using the following lemma, we can replace the rewriting constraints $xy \overset{*}{\to}_R z$ by ordinary equations plus $\circ_\sigma$-predicates.

Lemma 5.4.12. *There exists a fixed positive Boolean formula*

$$\psi(x, y, z, x_1, \ldots, x_m)$$

over the signature of $(\mathbb{M}(A,I), \iota, (\circ_\sigma)_{\sigma\in\Sigma})$ *such that*

[7] Atomic predicates of the form $x = 1$ are not necessary since $\{1\} \in \mathcal{IL}(\mathcal{C})$.

- *the size of* ψ *is bounded by* $|\Sigma|^{O(c(\Sigma,I_\Sigma))}$,[8] *and*
- *for all* $x, y, z \in \mathrm{IRR}(R)$ *we have* $xy \xrightarrow{*}_R z$ *in* $\mathbb{M}(A,I)$ *if and only if* $(\mathbb{M}(A,I), \iota, (\circ_\sigma)_{\sigma\in\Sigma}) \models \exists x_1 \cdots \exists x_m : \psi(x, y, z, x_1, \ldots, x_m)$.

Proof. Recall that $\mathcal{F}(A,I)$ is the set of all independence cliques in (A,I). For the further reasoning it is important to note that $a, b \in A_\sigma$ and $(a,c) \in I$ implies $(b,c) \in I$.

First we show that for all $x, y, z \in \mathrm{IRR}(R)$, $xy \xrightarrow{*}_R z$ in $\mathbb{M}(A,I)$ if and only if there exist $p, r, s, t, u \in \mathrm{IRR}(R)$ and $C_1, C_2 \in \mathcal{F}(A,I)$ such that in $(\mathbb{M}(A,I), \iota)$

$$[C_1][C_2] \xrightarrow{*}_R u, \quad \iota(p) = r, \quad x = s\,[C_1]\,p, \quad y = r\,[C_2]\,t, \quad z = s\,u\,t. \tag{5.3}$$

If (5.3) holds, then $xy \xrightarrow{*}_R z$ follows immediately. Now assume that $xy \xrightarrow{*}_R z$. We can choose $p \in \mathbb{M}(A,I)$ of maximal length such that $x = x'p$ and $y = \iota(p)y'$. Let $C_1 = \max(x') \in \mathcal{F}(A,I)$, $C_2 = \min(y') \in \mathcal{F}(A,I)$, and $[C_1][C_2] \xrightarrow{*}_R u \in \mathrm{IRR}(R)$. Hence, there are s and t with $x = s[C_1]p$, $y = \iota(p)[C_2]t$, and $xy \xrightarrow{*}_R sut \xrightarrow{*}_R z$. Note that $p, s, t, u, [C_1], [C_2] \in \mathrm{IRR}(R)$. Due to the choice of p, only rules of the form $(ab, c) \in R$, where $a \in C_1$, $b \in C_2$, and $a, b, c \in A_\sigma$ for some $\sigma \in \Sigma$, can be applied to the trace $[C_1][C_2]$. Thus, $\mathrm{alph}(u) = C_1 \cup C_2$, and if $(a,u) \in I$ for $a \in A$, then also $(a, C_1) \in I$. We claim that $sut \in \mathrm{IRR}(R)$, which implies $z = sut$ and hence (5.3). Assume that there exist $ab \in \mathrm{dom}(R)$ and traces q_1, q_2 such that $sut = q_1abq_2$. By Levi's Lemma 2.7.1 we obtain up to symmetry one of the following two diagrams (recall that $s, u, t \in \mathrm{IRR}(R)$):

q_2	s_2	u_2	t_2
ab	a	ε	b
q_1	s_1	u_1	t_1
	s	u	t

q_2	s_2	u_2	t_2
ab	a	b	ε
q_1	s_1	u_1	t_1
	s	u	t

Assume that $a, b \in A_\sigma$ ($\sigma \in \Sigma$). Let us first consider the left diagram. Since $(a, u_1) \in I$, $(b, u_2) \in I$, and $u = u_1u_2$, we obtain $(a,u) \in I$ and thus $(a, C_1) \in I$. Furthermore, from the diagram we obtain also $(b, s_2) \in I$. Thus, $(a, s_2) \in I$, which implies $a \in \max(s)$. Together with $(a, C_1) \in I$ it follows that $a \in \max(s[C_1]) = C_1$, which contradicts $(a, C_1) \in I$.

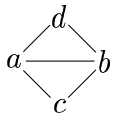
[8] This is a fixed constant in our situation. But latter we will apply an analogous lemma in a situation, where (Σ, I_Σ) belongs to the input.

Now let us consider the right diagram. Again we have $a \in \max(s)$. Furthermore, since $b \in \min(u) \cap A_\sigma$, there are two possibilities: either there exists $a' \in C_1 \cap A_\sigma$ or $b \in C_2$ and $(b, C_1) \in I$. If $a' \in C_1 \cap A_\sigma$, then $s[C_1]$ would contain the factor $aa' \in \mathrm{dom}(R_\sigma)$, which contradicts $x = s[C_1]p \in \mathrm{IRR}(R)$. If $b \in C_2$ and $(b, C_1) \in I$, then also $(a, C_1) \in I$, which implies $a \in \max(s[C_1]) = C_1$, again a contradiction.

Thus, $xy \xrightarrow{*}_R z$ is equivalent to (5.3). It remains to replace the additional rewriting constraints of the form $[C_1][C_2] \xrightarrow{*}_R u$, where $C_1, C_2 \in \mathcal{F}(A, I)$, by equations of the form $x' \circ_\sigma y' = z'$. Since $C_i \in \mathcal{F}(A, I)$ we can write down a disjunction over all independence cliques C_1' and C_2' in (Σ, I_Σ), with the meaning that $C_i' = \{\sigma \in \Sigma \mid C_i \cap A_\sigma \neq \emptyset\}$, and replace C_i in (5.3) by $x_{i,1}x_{i,2}\cdots x_{i,n_i}$, where $n_i = |C_i'| \leq |\Sigma|$ and $x_{i,j}$ is a new variable. Moreover, we add the constraints $x_{i,j} \in A_{\sigma(i,j)}$, where $C_i' = \{\sigma(i,j) \mid 1 \leq j \leq n_i\}$. Since there are at most $|\Sigma|^{c(\Sigma,I_\Sigma)+1}$ many cliques in (Σ, I_Σ), this results in a disjunction of $|\Sigma|^{O(c(\Sigma,I_\Sigma))}$ many conjunctions of size $O(|\Sigma|)$. Finally the rewriting constraint $x_{1,1}\cdots x_{1,n_1}x_{2,1}\cdots x_{2,n_2} \xrightarrow{*}_R u$ is equivalent to a conjunction of at most $|\Sigma|$ many equations of the form $x' \circ_\sigma y' = z'$ with $x', y', z' \in A_\sigma$ and a single equation over $\mathbb{M}(A, I)$. □

Let us illustrate the last step in the previous proof with an example:

Example 5.4.13. *Assume that* $\Sigma = \{a, b, c, d\}$ *and the independence relation* I_Σ *looks as follows:*



Then the rewriting constraint

$$x_a x_b x_c \, x_a' x_b' x_d \xrightarrow{*}_R u,$$

where $x_a, x_a' \in A_a$, $x_b, x_b' \in A_b$, $x_c \in A_c$, *and* $x_d \in A_d$, *is equivalent to*

$$x_a \circ_a x_a' = y_a \;\wedge\; x_b \circ_b x_b' = y_b \;\wedge\; u = y_a y_b x_c x_d$$

with the additional constraints $y_a \in A_a$ *and* $y_b \in A_b$. *Here, the equation* $u = y_a y_b x_c x_d$ *is interpreted in the trace monoid* $\mathbb{M}(A, I)$.

By applying Lemma 5.4.12 to every rewriting constraint $xy \xrightarrow{*}_R z$, we obtain an equivalent formula over $(\mathbb{M}(A, I), \iota, (\circ_\sigma)_{\sigma\in\Sigma}, \mathcal{L}(\mathcal{D}))$. Since (Σ, I_Σ) is assumed to be fixed, the size of the resulting conjunction increased only by a constant factor. This concludes the proof of Theorem 5.4.10.

5.4.4 Graph products of finite monoids, free monoids, and free groups

In this section we briefly consider a restricted class of graph products, for which we derive better complexity bounds. More precisely, we consider graph products of the form $\mathbb{P} = \mathbb{P}(\Sigma, I_\Sigma, (\mathcal{M}_\sigma)_{\sigma\in\Sigma})$, where for every $\sigma \in \Sigma$, the monoid $\mathcal{M}_\sigma$ is either a finite monoid, or a free monoid, or a free group. In fact, it is enough (and convenient) to assume that every $\mathcal{M}_\sigma$ is either isomorphic to $\mathbb{N}$ or to $\mathbb{Z}$, or $\mathcal{M}_\sigma$ is finite. If all $\mathcal{M}_\sigma$ are equal to $\mathbb{N}$, then we obtain *trace monoids*. If all $\mathcal{M}_\sigma$ are equal to $\mathbb{Z}$, then we obtain *free partially commutative groups*, which are also known as *semifree groups* [11, 12] or *graph groups* [73]. Free groups and free commutative groups arise as the extreme cases. If $I_\Sigma = \emptyset$ and all $\mathcal{M}_\sigma$ are finite or free groups, then we obtain *plain groups* in the sense of Haring-Smith [92].

Let us fix a graph product $\mathbb{P} = \mathbb{P}(\Sigma, I_\Sigma, (\mathcal{M}_\sigma)_{\sigma\in\Sigma})$, where for all $\sigma \in \Sigma$, the monoid $\mathcal{M}_\sigma$ is either finite, or is isomorphic to $\mathbb{N}$ or $\mathbb{Z}$. We select a natural set of generators Γ_σ for $\mathcal{M}_\sigma$ as follows. If $\mathcal{M}_\sigma$ is finite, let us define $\Gamma_\sigma = \mathcal{M}_\sigma \backslash \{1_\sigma\}$. If $\mathcal{M}_\sigma$ is isomorphic to $\mathbb{N}$, let $\Gamma_\sigma = \{a_\sigma\}$ for a generator a_σ. Finally if $\mathcal{M}_\sigma$ is isomorphic to $\mathbb{Z}$, then $\Gamma_\sigma = \{a_\sigma, a_\sigma^{-1}\}$, where a_σ generates $\mathcal{M}_\sigma$ as a group. Let $\Gamma = \bigcup_{\sigma\in\Sigma} \Gamma_\sigma$ and $I = \bigcup_{(\sigma,\tau)\in I_\Sigma} \Gamma_\sigma \times \Gamma_\tau$. Since every Γ_σ is closed under taking the inverse in the group of units of $\mathcal{M}_\sigma$, we have a partial involution ι on Γ that can be lifted to a partial involution on the finitely generated trace monoid $\mathbb{M} = \mathbb{M}(\Gamma, I)$. This trace monoid will replace the monoid $\mathbb{M}(A, I)$ from the previous section. Define the trace rewriting system R on $\mathbb{M}$ by

$$R = \bigcup_{\sigma\in\Sigma} \{ab \to c \mid \mathcal{M}_\sigma \text{ is finite, } a, b, c \in \Gamma_\sigma, \text{ and } a \circ_\sigma b = c\} \cup \bigcup_{\sigma\in\Sigma} \{a(a^{-1}) \to \varepsilon \mid a \in \Gamma_\sigma, a \text{ invertible in } \mathcal{M}_\sigma\}.$$

Thus, R is a finite trace rewriting system over the finitely generated trace monoid $\mathbb{M}$ and $\mathbb{P} \cong \mathbb{M}/\overset{*}{\leftrightarrow}_R$. Completely analogous to Lemma 4.7.1 we can show that this new system R is again confluent.

Let $\omega : \mathbb{P} \to \mathrm{IRR}(R)$ be the function that maps $s \in \mathbb{P}$ to the unique irreducible trace from $\mathbb{M}$, representing s. Following [66], we choose for $\mathbb{P}$ the constraint class $\mathrm{NRAT}(\mathbb{P}) = \{L \subseteq \mathbb{P} \mid \omega(L) \in \mathrm{REC}(\mathbb{M})\} \subseteq 2^{\mathbb{P}}$.[9] In [66] it

[9] "NRAT" stands for "normalized rational".

was shown that NRAT($\mathbb{P}$) is an effective Boolean algebra. We represent a set $L \in \mathrm{NRAT}(\mathbb{P})$ by a finite automaton for $\{u \in \Gamma^* \mid [u]_I \in \omega^{-1}(L)\}$.

It is not hard to see that the class NRAT($\mathbb{P}$) can be also obtained by applying the $\mathcal{IL}$-operation from the previous section to the classes $\mathrm{RAT}(\mathcal{M}_\sigma)$ for $\sigma \in \Sigma$. Hence, Theorem 5.4.10 implies that $\exists\mathrm{FOTh}(\mathbb{P}, \mathrm{NRAT}(\mathbb{P}))$ is decidable in exponential space. In this section we will improve the complexity to PSPACE. In order to obtain existing results for free monoids as special cases, we will also put a description of the graph product $\mathbb{P}$ into the input. This description contains the adjacency matrix of (Σ, I_Σ), and for each node σ either the multiplication table of $\mathcal{M}_\sigma$ if $\mathcal{M}_\sigma$ is finite or a bit indicating whether $\mathcal{M}_\sigma \cong \mathbb{N}$ or $\mathcal{M}_\sigma \cong \mathbb{Z}$.

We proceed analogously to the proof of Theorem 5.4.10: We may assume that we have given a positive Boolean combination θ of predicates of the form $xy = z$, $x \neq y$, $x \in L$, and $x \notin L$, where x, y, and z are variables and L belongs to NRAT($\mathbb{P}$). In order to change from $\mathbb{P}$ to the finitely generated trace monoid $\mathbb{M}$, we add for every variable x the constraint $x \notin \mathrm{RED}(R)$[10] and replace every equation $xy = z$ by $xy \xrightarrow{*}_R z$. These rewriting constraints can be replaced by ordinary equations using the following lemma, which can be shown analogously to Lemma 5.4.12.

Lemma 5.4.14. *There exists a positive Boolean formula*

$$\psi(x, y, z, x_1, \ldots, x_m)$$

over the signature of $(\mathbb{M}, \iota)$ *such that*

- *the size of* ψ *is bounded by* $|\Gamma|^{O(c(\Gamma,I))}$ *and*
- *for all* $x, y, z \in \mathrm{IRR}(R)$ *we have* $xy \xrightarrow{*}_R z$ *in* $\mathbb{M}$ *if and only if* $(\mathbb{M}, \iota) \models \exists x_1 \cdots \exists x_m : \psi(x, y, z, x_1, \ldots, x_m)$.

Note that in contrast to Lemma 5.4.12 we can avoid the $\circ_\sigma$-predicates in the present situation: We would only obtain "local equations" of the form $x \circ_\sigma y = z$ for the case that $\mathcal{M}_\sigma$ is a finite monoid, which allows to substitute all possible values from $\mathcal{M}_\sigma$ for x, y, and z.

Since $c(\Gamma, I) = c(\Sigma, I_\Sigma) \leq k$, the above transformations can be done in time polynomial for every fixed k. Thus, we have shown the following result:

[10] Of course this constraint is equivalent to $x \in \mathrm{IRR}(R)$, but we prefer the negated constraint $x \notin \mathrm{RED}(R)$, since an automaton for $\{u \in \Gamma^* \mid [u]_I \in \mathrm{RED}(R)\}$ can be easily constructed in polynomial time: $\{u \in \Gamma^* \mid [u]_I \in \mathrm{RED}(R)\} = \bigcup_{ab \in \mathrm{dom}(R)} \Gamma^* a I(a)^* b \Gamma^*$.

Theorem 5.4.15. *For every $k \geq 0$ there is a deterministic polynomial time algorithm such that:*

- *The input consists of a graph product $\mathbb{P} = \mathbb{P}(\Sigma, I_\Sigma, (\mathcal{M}_\sigma)_{\sigma\in\Sigma})$ (coded as described above) with $\mathcal{M}_\sigma$ finite or isomorphic to $\mathbb{N}$ or $\mathbb{Z}$ and $c(\Sigma, I_\Sigma) \leq k$, and an existential sentence θ over the signature of $(\mathbb{P}, \mathrm{NRAT}(\mathbb{P}))$.*
- *On a given input $(\Sigma, I_\Sigma, (\mathcal{M}_\sigma)_{\sigma\in\Sigma}, \theta)$, the algorithm produces an existential sentence ϕ over the signature of $(\mathbb{M}, \iota, \mathrm{REC}(\mathbb{M}))$ such that*

$$(\mathbb{P}, \mathrm{NRAT}(\mathbb{P})) \models \theta \text{ if and only if } (\mathbb{M}, \iota, \mathrm{REC}(\mathbb{M})) \models \phi.$$

Corollary 5.4.16. *The following problem is PSPACE-complete for every fixed $k \geq 0$:*
INPUT: As in Theorem 5.4.15.
QUESTION: Does θ belong to $\exists\mathrm{FOTh}(\mathbb{P}, \mathrm{NRAT}(\mathbb{P}))$?
If $c(\Sigma, I_\Sigma)$ is not bounded by a constant, then this problem is in EXPSPACE.

Proof. The PSPACE (resp. EXPSPACE) upper bound follows from Theorem 5.4.1 and Theorem 5.4.15 with $c(\Sigma, I_\Sigma) = c(\Gamma, I)$. PSPACE-hardness follows from the PSPACE-hardness of $\exists\mathrm{FOTh}(\{a, b\}^*, \mathrm{RAT}(\{a, b\}^*))$, see [114, Lem. 3.2.3] and [162, Thm. 1]. □

Remark 5.4.17. *Corollary 5.4.16 encompasses corresponding statements from [65, 67, 68, 69, 91, 131, 132, 162].*

5.5 Positive theories of graph products

In this section we consider positive theories of graph products. Assume that $\mathbb{P} = \mathbb{P}(\Sigma, I_\Sigma, (\mathcal{G}_\sigma)_{\sigma\in\Sigma})$ is a graph product such that every $\mathcal{G}_\sigma$ is a *finitely generated group*. Let Γ_σ be a finite generating set for $\mathcal{G}_\sigma$. Then $\mathbb{P}$ is generated by $\Gamma = \bigcup_{\sigma\in\Sigma} \Gamma_\sigma$. Let $D_\Sigma = (\Sigma \times \Sigma) \setminus I_\Sigma$ the dependence relation corresponding to I_Σ. A node $\sigma \in \Sigma$ is called an *isolated node of the dependence alphabet* (Σ, D_Σ) if $D_\Sigma(\sigma) = \{\sigma\}$. Throughout this section, we make the following two assumptions:

Assumption 5.5.1. *For every isolated node σ of (Σ, D_Σ), the positive theory $\mathrm{posTh}(\mathcal{G}_\sigma, (a)_{a\in\Gamma_\sigma}, \mathrm{REC}(\mathcal{G}_\sigma))$ is decidable.*

Assumption 5.5.2. *For every nonisolated node σ of (Σ, D_Σ), the existential theory* $\exists\mathrm{FOTh}(\mathcal{G}_\sigma, (a)_{a\in\Gamma_\sigma}, \mathrm{REC}(\mathcal{G}_\sigma))$ *is decidable.*

Since we restrict to finitely generated groups, we obtain finite representations for recognizable constraints. More precisely, since $\mathbb{P}$ is a group, it follows that $L \in \mathrm{REC}(\mathbb{P})$ if and only if there exists a surjective group homomorphism $\rho : \mathbb{P} \to S$ onto a finite group S such that $L = \rho^{-1}(\rho(L))$. Thus, L can be represented by the finite group S, the homomorphism ρ and $F \subseteq S$ with $L = \rho^{-1}(F)$. To represent ρ, it suffices to specify its value $\rho(a)$ for every generator $a \in \Gamma$.

The aim of this section is to proof of the following result:

Theorem 5.5.3. *Let $\mathbb{P} = \mathbb{P}(\Sigma, I_\Sigma, (\mathcal{G}_\sigma)_{\sigma\in\Sigma})$ by a graph product such that Assumption 5.5.1 and Assumption 5.5.2 hold. Then* $\mathrm{posTh}(\mathbb{P}, (a)_{a\in\Gamma}, \mathrm{REC}(\mathbb{P}))$ *is decidable.*

Since the theory of a finite group is of course decidable, and the same holds for the theory of $\mathbb{Z}$ with rational constraints (Proposition 5.3.2), we obtain the following corollary:

Corollary 5.5.4. *Let $\mathbb{P}$ be a graph product of finite groups and free groups. Then* $\mathrm{posTh}(\mathbb{P}, (a)_{a\in\Gamma}, \mathrm{REC}(\mathbb{P}))$ *is decidable.*

Remark 5.5.5. *Note that Corollary 5.5.4 cannot be extended by allowing monoids for the factors of the graph product. Already the positive $\forall\exists^3$-theory of the free monoid $\{a, b\}^*$ is undecidable [76, 134]. Similarly, Corollary 5.5.4 cannot be extended by replacing* $\mathrm{REC}(\mathbb{P})$ *by* $\mathrm{RAT}(\mathbb{P})$*, since the latter class contains a free monoid $\{a, b\}^*$ in case $\mathbb{P} = F_2$ is the free group of rank 2.*

The proof of Theorem 5.5.3 consists of the following steps:

- In a first step, we will reduce $\mathrm{posTh}(\mathbb{P}, (a)_{a\in\Gamma}, \mathrm{REC}(\mathbb{P}))$ to the positive theories $\mathrm{posTh}(\mathbb{P}_i, (a)_{a\in\Gamma_i}, \mathrm{REC}(\mathbb{P}_i))$, $1 \le i \le n$, where the $\mathbb{P}_i$ result from the connected components of the dependence alphabet (Σ, D_Σ). Thus, $\mathbb{P} = \prod_{i=1}^n \mathbb{P}_i$. After this step, we may assume that (Σ, D_Σ) is connected and (by Assumption 5.5.1) contains at least two nodes.

- Next, we will reduce $\mathrm{posTh}(\mathbb{P}, (a)_{a\in\Gamma}, \mathrm{REC}(\mathbb{P}))$ (where the underlying dependence alphabet (Σ, D_Σ) is connected and contains at least two nodes) to $\exists\mathrm{FOTh}(\mathbb{P} * F, (a)_{a\in\Gamma\cup K}, \mathrm{REC}(\mathbb{P} * F) \cup \mathcal{C})$. Here $F = F(K)$ is the free group generated by the finite set K, and $\mathcal{C}$ contains all

subgroups of $\mathbb{P} * F$ of the form $\mathbb{P} * F(K')$ for $K' \subseteq K$. This second step in inspired by techniques of Makanin and Merzlyakov [133, 141] developed for free groups. The proof of the main technical lemma is shifted into Section 5.5.3.

- The last step consists of an application of Theorem 5.4.10. In order to apply this theorem to $\exists\mathrm{FOTh}(\mathbb{P} * F, (a)_{a\in\Gamma\cup K}, \mathrm{REC}(\mathbb{P} * F) \cup \mathcal{C})$, we have to "decompose" the constraints (Lemma 5.5.8).

5.5.1 Simplifying the graph product $\mathbb{P}$

In a first step we may assume that no finite group $\mathcal{G}_\sigma$, $\sigma \in \Sigma$, is a direct product of two finite nontrivial groups, since otherwise we could replace σ by two independent nodes. In particular, if $\mathcal{G}_\sigma$ is not $\mathbb{Z}/2\mathbb{Z}$, then there must exist $a \in \mathcal{G}_\sigma$ such that $a^2 \neq 1_\sigma$, i.e., $a \neq a^{-1}$ in $\mathcal{G}_\sigma$. Next, assume that the dependence alphabet (Σ, D_Σ) consists of two nonempty disjoint components (Σ_1, D_1) and (Σ_2, D_2), which define graph products $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively. Then $\mathbb{P} = \mathbb{P}_1 \times \mathbb{P}_2$. Furthermore by Mezei's Theorem, see e.g. [18], every $L \in \mathrm{REC}(\mathbb{P})$ is effectively a finite union of sets of the form $L_1 \times L_2$ with $L_i \in \mathrm{REC}(\mathbb{P}_i)$. Since the corresponding statement for singleton sets (i.e., constants from Γ) holds as well, we may apply the following Proposition 5.5.6, which is a decomposition lemma in the style of the Feferman Vaught Theorem [81].

Proposition 5.5.6. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be monoids with classes $\mathcal{C}_1 \subseteq 2^{\mathcal{M}_1}$ and $\mathcal{C}_2 \subseteq 2^{\mathcal{M}_2}$. Let $\mathcal{C}$ be a class of subsets of $\mathcal{M}_1 \times \mathcal{M}_2$ such that each $L \in \mathcal{C}$ is effectively a finite union of sets of the form $L_1 \times L_2$ with $L_1 \in \mathcal{C}_1$ and $L_2 \in \mathcal{C}_2$. If both* (pos)FOTh$(\mathcal{M}_1, \mathcal{C}_1)$ *and* (pos)FOTh$(\mathcal{M}_2, \mathcal{C}_2)$ *are decidable, then* (pos)FOTh$(\mathcal{M}_1 \times \mathcal{M}_2, \mathcal{C})$ *is decidable, too.*

Proof. Let $\phi(x_1, \ldots, x_n)$ be a first-order formula with free variables $x_1, \ldots, x_n$ and atomic subformulas of the form $x = 1$, $xy = z$, and $x \in L$ with x, y, and z variables and $L \in \mathcal{C}$. For every variable x that appears in ϕ let $x^{(1)}$ and $x^{(2)}$ be two new variables. Then we replace each quantification $\exists x$ (resp. $\forall x$) in ϕ by $\exists x^{(1)}\exists x^{(2)}$ (resp. $\forall x^{(1)}\forall x^{(2)}$). Furthermore every equation $xy = z$ is replaced by the conjunction $x^{(1)}y^{(1)} = z^{(1)} \wedge x^{(2)}y^{(2)} = z^{(2)}$, and similarly for $x = 1$. Finally, given a constraint $x \in L$ in ϕ, where $L = \bigcup_{i=1}^n L_{i,1} \times L_{i,2}$ with $L_{i,1} \in \mathcal{C}_1$ and $L_{i,2} \in \mathcal{C}_2$, we replace this constraint by $\bigvee_{i=1}^n (x^{(1)} \in L_{i,1} \wedge x^{(2)} \in L_{i,2})$. Let us call the resulting formula $\phi'(x_1^{(1)}, x_1^{(2)}, \ldots, x_n^{(1)}, x_n^{(2)})$.

If all quantified variables of the form $x^{(i)}$ only range over $\mathcal{M}_i$ in ϕ', then for all $c_i = (a_i, b_i) \in \mathcal{M}_1 \times \mathcal{M}_2$ $(1 \leq i \leq n)$ we have $\phi(c_1, \ldots, c_n)$ if and only if $\phi'(a_1, b_1, \ldots, a_n, b_n)$. We claim that ϕ' is logically equivalent to a formula of the form $\bigvee_{j=1}^{m}(\theta_j^{(1)} \wedge \theta_j^{(2)})$, where for $i \in \{1, 2\}$ the formula $\theta_j^{(i)}$ only contains variables of the form $x^{(i)}$. Note that this proves the proposition. The claim above can be shown by induction on the quantifier-depth of ϕ. The case that ϕ is quantifier-free is clear. Assume that $\phi \equiv \exists x : \chi$. Hence, ϕ' is of the form $\exists x^{(1)} \exists x^{(2)}\ \chi'$. By induction we can assume that χ' is logically equivalent to a formula $\bigvee_{j=1}^{m}(\theta_j^{(1)} \wedge \theta_j^{(2)})$, where for $i \in \{1, 2\}$ the formula $\theta_j^{(i)}$ only contains variables of the form $y^{(i)}$. Thus, $\exists x^{(1)} \exists x^{(2)}\ \chi'$ is equivalent to $\bigvee_{j=1}^{m}(\exists x^{(1)}\ \theta_j^{(1)} \ \wedge\ \exists x^{(2)}\ \theta_j^{(2)})$. In the case of a universal quantification we can conclude similarly, but we first have to transform the formula $\bigvee_{j=1}^{m} \theta_j^{(1)} \wedge \theta_j^{(2)}$ into a formula of the form $\bigwedge_{j=1}^{k} \psi_j^{(1)} \vee \psi_j^{(2)}$, where $\psi_j^{(i)}$ only contains variables of the form $y^{(i)}$. This is of course possible but may cause an exponential size increase. Finally note that the construction above does not introduce negations and thus can be also used for positive formulas. □

Note that the construction from the previous proof may lead to a nonelementary blow-up with respect to formula size, since each quantifier alternation leads to an exponential size increase. This will be the main bottle neck in our proof of Theorem 5.5.3.

By Proposition 5.5.6 and Assumption 5.5.1 we may assume that the graph (Σ, D_Σ) is connected and contains at least two nodes. By Corollary 5.3.4 we can also exclude the case that Σ contains exactly two D_Σ-adjacent nodes, which are both labeled by $\mathbb{Z}/2\mathbb{Z}$. Thus, we may assume that either the graph (Σ, D_Σ) contains a path consisting of three different nodes or one of the groups $\mathcal{G}_\sigma$ has a generator $g \in \mathcal{G}_\sigma$ with $g^{-1} \neq g \neq 1_\sigma$. Hence, there exist three different generators $a \in \mathcal{G}_{\sigma(a)} \setminus \{1_{\sigma(a)}\}$, $b \in \mathcal{G}_{\sigma(b)} \setminus \{1_{\sigma(b)}\}$, and $c \in \mathcal{G}_{\sigma(c)} \setminus \{1_{\sigma(c)}\}$ such that

- $\sigma(a) \neq \sigma(b)$ and $(\sigma(a), \sigma(b)) \in D_\Sigma$,
- $\sigma(b) \neq \sigma(c)$ and $(\sigma(b), \sigma(c)) \in D_\Sigma$, and finally
- either $\sigma(a) \neq \sigma(c)$ or $a \neq a^{-1} = c$ in $\mathcal{G}_{\sigma(a)}$.

Thus, the dependency between a, b, and c being used is

$$a \text{ —} b \text{ —} c.$$

In Section 5.5.3, a, b, and c will always refer to these three elements.

5.5.2 Reducing to the existential theory

Our strategy for reducing the positive theory of $\mathbb{P}$ to an existential theory is based on [133, 141], but the presence of partial commutation and recognizable constraints makes the construction more involved: Given a positive sentence θ, which is interpreted over $\mathbb{P}$, we construct an *existential sentence* θ', which is interpreted over a free product $\mathbb{P} * F$ of $\mathbb{P}$ and a free group F, such that θ is true in $\mathbb{P}$ if and only if θ' is true in $\mathbb{P} * F$. Roughly speaking, θ' results from θ by replacing the universally quantified variables by the generators of the free group F.

Assume that we have given a positive Boolean combination ϕ of equations (with constants) and recognizable constraints $x_i \in L_i$ $(1 \leq i \leq n)$, where the latter are represented via surjective homomorphisms $\rho_i : \mathbb{P} \to S_i$ such that $L_i = \rho_i^{-1}(\rho_i(L_i))$. Let $S = \prod_{i=1}^n S_i$ and define $\rho(x) = (\rho_1(x), \ldots, \rho_n(x))$ for $x \in \mathbb{P}$. Now we can replace every constraint $x_i \in L_i$ by constraints of the form $\rho(x_i) = q$ for $q \in S$. Note that the number of these constraints is bounded exponentially in the size of the description of ϕ. Thus, we may assume that all recognizable constraints in our initial positive formula are given in the form $\rho(x) = q$ for $q \in S$ and a fixed surjective homomorphism $\rho : \mathbb{P} \to S$ onto a finite group S.

Let K be a finite set of new constants, $K \cap \Gamma = \emptyset$. Recall that $F(K)$ is the free group generated by K. For the free product $\mathbb{P} * F(K)$ we will write $\mathbb{P}[K]$ in the following. Instead of $\mathbb{P}[\{k_1, \ldots, k_n\}]$, we write $\mathbb{P}[k_1, \ldots, k_n]$. Similarly, instead of $\mathbb{P}[K \cup \{k\}]$ we write $\mathbb{P}[K, k]$. In the sequel we also have to deal with formulas, where the constraints are given by different extensions of our basic homomorphism $\rho : \mathbb{P} \to S$ to $\mathbb{P}[K]$. For this we introduce the following notation: Let $\mathcal{G}$ be an arbitrary group, and let $\varrho : \mathcal{G} \to S$ be a group homomorphism onto some finite group S. Let $K = \{k_1, \ldots, k_n\}$ and $q_1, \ldots, q_n \in S$. Then $\varrho_{q_1,\ldots,q_n}^{k_1,\ldots,k_n} : \mathcal{G}[K] \to S$ denotes the unique extension of ϱ, defined by $\varrho_{q_1,\ldots,q_n}^{k_1,\ldots,k_n}(k_i) = q_i$. Similarly, if ϕ is some Boolean combination of equations and constraints of the form $\varrho(x) = q$, then $\phi_{q_1,\ldots,q_n}^{k_1,\ldots,k_n}$ denotes the formula that results from ϕ by replacing every constraint $\varrho(x) = q$ by $\varrho_{q_1,\ldots,q_n}^{k_1,\ldots,k_n}(x) = q$. Let us now fix a formula[11]

$$\theta(\widetilde{z}) \;\equiv\; \forall x_1 \exists y_1 \cdots \forall x_n \exists y_n \, \phi(x_1, \ldots, y_n, y_1, \ldots, y_n, \widetilde{z}),$$

[11] In the following symbols with a tilde like $\widetilde{x}$ will denote sequences of arbitrary length over some set that will be always clear form the context. If say $\widetilde{a} = (a_1, \ldots, a_m)$, then $\widetilde{a} \in A$ means $a_1 \in A, \ldots, a_m \in A$.

with ϕ a positive Boolean formula over the signature of $(\mathbb{P}, (a)_{a\in\Gamma}, \mathrm{REC}(\mathbb{P}))$ such that all recognizable constraints are given in the form $\rho(x) = q \in S$ for our fixed homomorphism $\rho : \mathbb{P} \to S$. Choose for every universally quantified variable x_i in θ a new constant k_i and let $K = \{k_1, \ldots, k_n\}$. The following theorem yields the reduction from the positive to the existential theory.

Theorem 5.5.7. *Let $\theta(\widetilde{z}) \equiv \forall x_1 \exists y_1 \cdots \forall x_n \exists y_n\, \phi(x_1, \ldots, x_n, y_1, \ldots, y_n, \widetilde{z})$ be as above. For all $\widetilde{u} \in \mathbb{P}$ we have $\theta(\widetilde{u})$ in $\mathbb{P}$ if and only if*

$$\bigwedge_{q_1\in S} \exists y_1 \cdots \bigwedge_{q_n\in S} \exists y_n \left\{ \begin{array}{l} \bigwedge\limits_{1\le i\le n} y_i \in \mathbb{P}[k_1, \ldots, k_i] \wedge \\ \phi^{k_1,\ldots,k_n}_{q_1,\ldots,q_n}(k_1, \ldots, k_n, y_1, \ldots, y_n, \widetilde{u}) \end{array} \right\} \text{ in } \mathbb{P}[K]. \quad (5.4)$$

Proof. We prove the theorem by induction on n. The case $n = 0$ is clear. If $n > 0$, then inductively we can assume that for all $s_1, t_1, \widetilde{u} \in \mathbb{P}$ we have

$$\forall x_2 \exists y_2 \cdots \forall x_n \exists y_n\, \phi(s_1, x_2, \ldots, x_n, t_1, y_2, \ldots, y_n, \widetilde{u}) \quad \text{in } \mathbb{P}$$

if and only if

$$\bigwedge_{q_2\in S} \exists y_2 \cdots \bigwedge_{q_n\in S} \exists y_n \left\{ \begin{array}{l} \bigwedge\limits_{2\le i\le n} y_i \in \mathbb{P}[k_2, \ldots, k_i] \wedge \\ \phi^{k_2,\ldots,k_n}_{q_2,\ldots,q_n}(s_1, k_2, \ldots, k_n, t_1, y_2, \ldots, y_n, \widetilde{u}) \end{array} \right\}$$

is true in $\mathbb{P}[k_2, \ldots, k_n]$. Thus, for all $\widetilde{u} \in \mathbb{P}$ we have

$$\forall x_1 \exists y_1 \cdots \forall x_n \exists y_n\, \phi(x_1, \ldots, x_n, y_1, \ldots, y_n, \widetilde{u}) \quad \text{in } \mathbb{P}$$

if and only if

$$\forall x_1 \in \mathbb{P}\ \exists y_1 \bigwedge_{q_2\in S} \exists y_2 \cdots \bigwedge_{q_n\in S} \exists y_n \left\{ \begin{array}{l} \bigwedge\limits_{1\le i\le n} y_i \in \mathbb{P}[k_2, \ldots, k_i] \wedge \\ \phi^{k_2,\ldots,k_n}_{q_2,\ldots,q_n}(x_1, k_2, \ldots, k_n, y_1, \ldots, y_n, \widetilde{u}) \end{array} \right\}$$

is true in $\mathbb{P}[k_2, \ldots, k_n]$. Note that if we transform this formula into prenex normal form, in the resulting existential formula the constraints are given by different extensions of the basic homomorphism ρ.

Since ρ is surjective, we can replace the universal quantifier $\forall x_1 \in \mathbb{P}$ by $\bigwedge_{q_1\in S} \forall x \in \rho^{-1}(q_1)$. Hence, by the following Lemmas 5.5.10 and 5.5.11 the above formula is true in $\mathbb{P}[k_2, \ldots, k_n]$ if and only if

$$\bigwedge_{q_1\in S} \exists y_1 \bigwedge_{q_2\in S} \exists y_2 \cdots \bigwedge_{q_n\in S} \exists y_n \left\{ \begin{array}{l} \bigwedge\limits_{1\le i\le n} y_i \in \mathbb{P}[k_1, k_2, \ldots, k_i] \wedge \\ \phi^{k_1,k_2,\ldots,k_n}_{q_1,q_2,\ldots,q_n}(k_1, k_2, \ldots, k_n, y_1, \ldots, y_n, \widetilde{u}) \end{array} \right\}$$

is true in $\mathbb{P}[k_1, k_2, \ldots, k_n] = \mathbb{P}[K]$. Whereas Lemma 5.5.10 is only valid for positive sentences, but has a quite simple proof, Lemma 5.5.11 holds for arbitrary formulas, but its proof is quite involved. □

To complete the proof of Theorem 5.5.3, we apply Theorem 5.4.10 to the group $\mathbb{P}[K]$, which is a graph product as well: Add every $k \in K$ as an isolated node to the independence alphabet (Σ, I_Σ) and label it with $F(k) \cong \mathbb{Z}$. For every $\sigma \in \Sigma$ let $\mathcal{C}_\sigma = \mathrm{REC}(\mathcal{G}_\sigma) \cup \{\{a\} \mid a \in \Gamma_\sigma\}$ and for every $k \in K$ let $\mathcal{C}_k = \mathrm{RAT}(F(k))$, which contains $\mathrm{REC}(F(k))$ and every singleton subset. Let $\mathcal{C} = \bigcup_{\tau \in \Sigma \cup K} \mathcal{C}_\tau$. By Assumption 5.5.2 (note that (Σ, D_Σ) does not contain isolated nodes by the simplifications from the previous section), $\exists\mathrm{FOTh}(\mathcal{G}_\sigma, \mathcal{C}_\sigma)$ is decidable for every $\sigma \in \Sigma$. By Proposition 5.3.2, for every $k \in K$, $\exists\mathrm{FOTh}(F(k), \mathcal{C}_k)$ is decidable as well. Thus, in order to apply Theorem 5.4.10, it suffices to show that all constraint sets and constants (viewed as singleton sets) in (5.4) belong to $\mathcal{IL}(\mathcal{C})$. For the constants this is clear – they all belong to $\Gamma \cup K$. Also $\mathbb{P}[k_1, \ldots, k_i] \in \mathcal{IL}(\mathcal{C})$ is easy to see. Thus, the following lemma finishes the proof of Theorem 5.5.3.

Lemma 5.5.8. *Let $\mathbb{P} = \mathbb{P}(\Sigma, I_\Sigma, (\mathcal{M}_\sigma)_{\sigma \in \Sigma})$ be an arbitrary graph product of monoids $\mathcal{M}_\sigma$. Then $\mathrm{REC}(\mathbb{P}) \subseteq \mathcal{IL}(\mathcal{C})$ for $\mathcal{C} = \bigcup_{\sigma \in \Sigma} \mathrm{REC}(\mathcal{M}_\sigma)$.*

Proof. Let $\mathcal{D} = \bigcup_{\sigma \in \Sigma} \{L \setminus \{1_\sigma\} \mid L \in \mathrm{REC}(\mathcal{M}_\sigma)\}$. Assume that $L \in \mathrm{REC}(\mathbb{P})$ and let $\varrho : \mathbb{P} \to S$ be a surjective homomorphism onto the finite monoid S such that $L = \varrho^{-1}(F)$ for $F \subseteq S$. We define A_σ, A, I, and R in the same way as at the beginning of Section 5.4.3. Note that if we restrict ϱ to $\mathcal{M}_\sigma \subseteq \mathbb{P}$, we obtain a homomorphism from $\mathcal{M}_\sigma$ to S. Thus,

$$\Delta_\sigma = \{A_\sigma \cap \varrho^{-1}(q) \mid q \in S\} \setminus \{\emptyset\} \subseteq \mathcal{D}.$$

Let $\Delta = \bigcup_{\sigma \in \Sigma} \Delta_\sigma$ and $I_\Delta = \bigcup_{(\sigma,\tau) \in I_\Sigma} \Delta_\sigma \times \Delta_\tau$. Hence, (Δ, I_Δ) is a finite independence alphabet. Define the homomorphism $\pi : \mathbb{M}(\Delta, I_\Delta) \to S$ by $\pi(A_\sigma \cap \varrho^{-1}(q)) = q$. Since $(B, C) \in I_\Delta$ implies $\pi(B)\pi(C) = \pi(C)\pi(B)$ in S, this defines indeed a homomorphism. Thus, $\pi^{-1}(F) \in \mathrm{REC}(\mathbb{M}(\Delta, I_\Delta))$. Let $\mathcal{A}$ be a finite state automaton that accepts $\{w \in \Delta^* \mid [w]_{I_\Delta} \in \pi^{-1}(F)\}$. Since every edge of $\mathcal{A}$ is labeled with a set from $\Delta \subseteq \mathcal{D}$, we can interpret $\mathcal{A}$ also as a $\mathcal{D}$-automaton, which is moreover I-closed. It is easy to see that $L(\mathcal{A})$, viewed as a subset of $\mathbb{M}(A, I)$, equals $h^{-1}(\varrho^{-1}(F))$, where $h : \mathbb{M}(A, I) \to \mathbb{P}$ is the canonical homomorphism that maps a trace $t \in \mathbb{M}(A, I)$ to the element of $\mathbb{P}$ represented by t.

Now consider $t \in \mathbb{M}(A, I)$. Then $h(t) = h(\mathrm{NF}_R(t))$. Thus, $t \in L(\mathcal{A}) = h^{-1}(\varrho^{-1}(F))$ if and only if $\mathrm{NF}_R(t) \in L(\mathcal{A})$. It follows that $L(\mathcal{A}) \cap \mathrm{IRR}(R) \subseteq \mathbb{M}(A, I)$ is precisely the set of irreducible traces that represents our initial language $L \subseteq \mathbb{P}$. Thus, $L \in \mathcal{IL}(\mathcal{C})$. □

Remark 5.5.9. *Concerning the complexity, it can be shown that our proof of Theorem 5.5.3 leads to a nonelementary blow-up due to the construction in our proof of Proposition 5.5.6. On the other hand, if we restrict to connected graphs (Σ, D_Σ), then Proposition 5.5.6 becomes superfluous. Due to Corollary 5.3.4 and the complexity statement in Theorem 5.4.10, we obtain an elementary reduction from the positive theory to the theories in Assumption 5.5.1 and 5.5.2.*

For the further consideration let us fix a set of constants K and a further constant $k \notin K$. Moreover, let $K_i \subseteq K$ for $1 \leq i \leq m$. Fix also $q \in S$ and a sequence $\widetilde{u} = (u_1, \ldots, u_N)$ of elements $u_i \in \mathbb{P}$.

Lemma 5.5.10. *Assume that $\phi(x, y_1, \ldots, y_m, \widetilde{z})$ is a positive Boolean formula with constraints of the form $\varrho(y) = p$ for (possibly different) extensions $\varrho : \mathbb{P}[K] \to S$ of $\rho : \mathbb{P} \to S$ and $p \in S$. If*

$$\exists y_1 \cdots \exists y_m \left\{ \begin{array}{l} \bigwedge\limits_{1 \leq i \leq m} y_i \in \mathbb{P}[K_i, k] \wedge \\ \phi_q^k(k, y_1, \ldots, y_m, \widetilde{u}) \end{array} \right\} \text{ in } \mathbb{P}[K, k],$$

then

$$\forall x \in \mathbb{P} \cap \rho^{-1}(q)\ \exists y_1 \cdots \exists y_m \left\{ \begin{array}{l} \bigwedge\limits_{1 \leq i \leq m} y_i \in \mathbb{P}[K_i] \wedge \\ \phi(x, y_1, \ldots, y_m, \widetilde{u}) \end{array} \right\} \text{ in } \mathbb{P}[K].$$

Proof. Assume that

$$\exists y_1 \cdots \exists y_m \left\{ \begin{array}{l} \bigwedge\limits_{1 \leq i \leq m} y_i \in \mathbb{P}[K_i, k] \wedge \\ \phi_q^k(k, y_1, \ldots, y_m, \widetilde{u}) \end{array} \right\} \text{ in } \mathbb{P}[K, k].$$

Then there are $t_i \in \mathbb{P}[K_i, k]$, $1 \leq i \leq m$, such that $\phi_q^k(k, t_1, \ldots, t_m, \widetilde{u})$ is true in $\mathbb{P}[K, k]$. Now choose an arbitrary $s \in \mathbb{P} \cap \rho^{-1}(q)$ and define a homomorphism $h : \mathbb{P}[K, k] \to \mathbb{P}[K]$ by $h(k) = s$ and $h(x) = x$ for $x \in \mathbb{P}[K]$. Since

$\rho(s) = q$, $h(\widetilde{u}) = \widetilde{u} \in \mathbb{P}$, $h(s) = s \in \mathbb{P}$, and ϕ_q^k is positive, the statement $\phi(s, h(t_1), \ldots, h(t_m), \widetilde{u})$ is true in $\mathbb{P}[K]$. Thus, indeed

$$\forall x \in \mathbb{P} \cap \rho^{-1}(q)\ \exists y_1 \cdots \exists y_m \left\{ \begin{array}{l} \bigwedge\limits_{1 \le i \le m} y_i \in \mathbb{P}[K_i] \wedge \\ \phi(x, y_1, \ldots, y_m, \widetilde{u}) \end{array} \right\} \text{ in } \mathbb{P}[K].$$

□

Note that the assertion of Lemma 5.5.10 does not hold in general, if ϕ involves negations. For example $\forall x : x \neq 1$ is false, but $k \neq 1$ is true. On the other hand, the converse implication of Lemma 5.5.10 is true for arbitrary formulas:

Lemma 5.5.11. *Assume that $\phi(x, y_1, \ldots, y_m, \widetilde{z})$ is a not necessarily positive Boolean formula with constraints of the form $\varrho(y) = p$ for (possibly different) extensions $\varrho : \mathbb{P}[K] \to S$ of $\rho : \mathbb{P} \to S$ and $p \in S$. If*

$$\forall x \in \mathbb{P} \cap \rho^{-1}(q)\ \exists y_1 \cdots \exists y_m \left\{ \begin{array}{l} \bigwedge\limits_{1 \le i \le m} y_i \in \mathbb{P}[K_i] \wedge \\ \phi(x, y_1, \ldots, y_m, \widetilde{u}) \end{array} \right\} \text{ in } \mathbb{P}[K],$$

then

$$\exists y_1 \cdots \exists y_m \left\{ \begin{array}{l} \bigwedge\limits_{1 \le i \le m} y_i \in \mathbb{P}[K_i, k] \wedge \\ \phi_q^k(k, y_1, \ldots, y_m, \widetilde{u}) \end{array} \right\} \text{ in } \mathbb{P}[K, k].$$

The statement of Lemma 5.5.11 will be shown by a reduction to the underlying trace monoid with involution. For this, we need one more lemma. First, we have to introduce a few notations.

In Lemma 5.5.11, the groups $\mathbb{P}$, $\mathbb{P}[K_i]$, $\mathbb{P}[K_i, k]$, $\mathbb{P}[K]$, and $\mathbb{P}[K, k]$ appear $(1 \le i \le m)$. Similarly to Section 5.4.3 we define $A_\sigma = \mathcal{G}_\sigma \setminus \{1_\sigma\}$ for $\sigma \in \Sigma$, $A = \bigcup_{\sigma \in \Sigma} A_\sigma$, and $I = \bigcup_{(\sigma,\tau) \in I_\Sigma} A_\sigma \times A_\tau$. Let $\mathbb{M} = \mathbb{M}(A, I)$. Since every $\mathcal{G}_\sigma$ is a group, we can define a total involution ι on A by taking the inverse in each group $\mathcal{G}_\sigma$, which can be lifted to a total involution on $\mathbb{M}$. Next, take for each constant $\kappa \in K \cup \{k\}$ a new copy $\overline{\kappa}$. Let $\overline{K} = \{\overline{\kappa} \mid \kappa \in K\}$ and similarly for $\overline{K_i}$. We extend the involution ι on A to $A \cup K \cup \overline{K} \cup \{k, \overline{k}\}$ by setting $\iota(\kappa) = \overline{\kappa}$ and $\iota(\overline{\kappa}) = \kappa$ for $\kappa \in K \cup \{k\}$. Then ι can be also lifted to the free product $\mathbb{M} * (K \cup \overline{K} \cup \{k, \overline{k}\})^*$, which will be the largest trace monoid in our further investigation. We will use the following abbreviations in the sequel: $\mathbb{M}[K_i] = \mathbb{M} * (K_i \cup \overline{K_i})^*$, $\mathbb{M}[K_i, k] = \mathbb{M} * (K_i \cup \overline{K_i} \cup \{k, \overline{k}\})^*$,

$\mathbb{M}[K] = \mathbb{M} * (K \cup \overline{K})^*$, and $\mathbb{M}[K,k] = \mathbb{M} * (K \cup \overline{K} \cup \{k, \overline{k}\})^*$. Finally let R be the trace rewriting system on $\mathbb{M}[K,k]$ defined by

$$R = \bigcup_{\sigma \in \Sigma} \{ab \to c \mid a,b,c \in A_\sigma,\ a \circ_\sigma b = c\} \cup \{a(a^{-1}) \to \varepsilon \mid a \in A_\sigma\} \cup \bigcup_{\kappa \in K \cup \{k\}} \{\kappa\overline{\kappa} \to \varepsilon, \overline{\kappa}\kappa \to \varepsilon\}$$

Then R is confluent and $\mathbb{M}[K,k]/\overset{*}{\leftrightarrow}_R \cong \mathbb{P}[K,k]$. Similarly, if we restrict R to traces from $\mathbb{M}[K]$ (resp. $\mathbb{M}$), then $\mathbb{M}[K]/\overset{*}{\leftrightarrow}_R \cong \mathbb{P}[K]$ (resp. $\mathbb{M}/\overset{*}{\leftrightarrow}_R \cong \mathbb{P}$).

Let $\widetilde{w} = (w_1, \ldots, w_N)$, where $w_i \in \mathbb{M} \cap \mathrm{IRR}(R)$ is the unique irreducible trace representing the fixed group element $u_i \in \mathbb{P}$. In the following, we identify a homomorphism $\varrho : \mathbb{P}[K] \to S$ with $h \circ \varrho : \mathbb{M}[K] \to S$, where $h : \mathbb{M}[K] \to \mathbb{P}[K]$ is the canonical homomorphism that maps a trace t to the group element represented by t. Moreover, for $\varrho : \mathbb{M}[K] \to S$, we denote with $\varrho_q^k : \mathbb{M}[K,k] \to S$ the extension of ϱ, defined by $\varrho_q^k(k) = q$ and $\varrho_q^k(\overline{k}) = q^{-1}$.

Similarly to Section 5.4.3, in the following lemma $\circ_\sigma$ denotes the tenary relation $\{(a,b,c) \mid a,b,c \in A_\sigma, a \circ_\sigma b = c\} \subseteq \mathbb{M}^3$ for $\sigma \in \Sigma$.

Lemma 5.5.12. *Let $\chi(x, y_1, \ldots, y_m, \widetilde{z})$ be a not necessarily positive Boolean formula over the signature of $(\mathbb{M}[K], \iota, (a)_{a \in \Gamma \cup K}, \mathrm{REC}(\mathbb{M}[K]), (\circ_\sigma)_{\sigma \in \Sigma})$ such that all recognizable constraints in χ have the form $\varrho(y) = p$ for (possibly different) extensions $\varrho : \mathbb{M}[K] \to S$ of $\rho : \mathbb{M} \to S$ and $p \in S$. If*

$$\forall x \in \mathrm{IRR}(R) \cap \rho^{-1}(q)\ \exists y_1 \cdots \exists y_m \left\{ \begin{array}{l} \bigwedge_{1 \le i \le m} y_i \in \mathbb{M}[K_i] \cap \mathrm{IRR}(R) \\ \wedge\ \chi(x, y_1, \ldots, y_m, \widetilde{w}) \end{array} \right\} \text{ in } \mathbb{M}[K],$$

then there are $s_1, s_2 \in \mathbb{M} \cap \mathrm{IRR}(R)$ with $\rho(s_1)q\rho(s_2) = q$ in S and

$$\exists y_1 \cdots \exists y_m \left\{ \begin{array}{l} \bigwedge_{1 \le i \le m} y_i \in \mathbb{M}[K_i, k] \cap \mathrm{IRR}(R) \\ \wedge\ \chi_q^k(s_1 k s_2, y_1, \ldots, y_m, \widetilde{w}) \end{array} \right\} \text{ in } \mathbb{M}[K,k].$$

The proof of Lemma 5.5.12 is the main technical difficulty and shifted to the next section. Using Lemma 5.5.12, we can finish the proof of Lemma 5.5.11: Assume that

$$\forall x \in \mathbb{P} \cap \rho^{-1}(q)\ \exists y_1 \cdots \exists y_m \left\{ \begin{array}{l} \bigwedge_{1 \le i \le m} y_i \in \mathbb{P}[K_i] \wedge \\ \phi(x, y_1, \ldots, y_m, \widetilde{u}) \end{array} \right\} \text{ in } \mathbb{P}[K].$$

By restricting every variable in ϕ to $\mathbb{M}[K] \cap \mathrm{IRR}(R)$ and replacing every equation $xy = z$ by $xy \xrightarrow{*}_R z$, we obtain a true statement over $\mathbb{M}[K]$. Similarly to Lemma 5.4.12, we can replace every rewriting constraint $xy \xrightarrow{*}_R z$ by a formula $\psi(x, y, z, \ldots)$ over the signature of $(\mathbb{M}[K], \iota, (\circ_\sigma)_{\sigma\in\Sigma})$. This transformation introduces only new existentially quantified variables ($\widetilde{y}$ below). We obtain a formula χ over the signature of $(\mathbb{M}[K], \iota, \mathrm{REC}(\mathbb{M}[K]), (\circ_\sigma)_{\sigma\in\Sigma})$ such that

$$\forall x \in \mathrm{IRR}(R) \cap \rho^{-1}(q)\ \exists y_1 \cdots \exists y_m\ \exists \widetilde{y} \left\{ \begin{array}{l} \bigwedge_{1\le i\le m} y_i \in \mathbb{M}[K_i] \cap \mathrm{IRR}(R) \\ \wedge\ \widetilde{y} \in \mathbb{M}[K] \cap \mathrm{IRR}(R) \\ \wedge\ \chi(x, y_1, \ldots, y_m, \widetilde{y}, \widetilde{w}) \end{array} \right\}$$

is true in $\mathbb{M}[K]$. Thus, by Lemma 5.5.12 there exist $s_1, s_2 \in \mathbb{M} \cap \mathrm{IRR}(R)$ such that $\rho(s_1)q\rho(s_2) = q$ in S and

$$\exists y_1 \cdots \exists y_m \exists\, \widetilde{y} \left\{ \begin{array}{l} \bigwedge_{1\le i\le m} y_i \in \mathbb{M}[K_i, k] \cap \mathrm{IRR}(R) \\ \wedge\ \widetilde{y} \in \mathbb{M}[K, k] \cap \mathrm{IRR}(R) \\ \wedge\ \chi^k_q(s_1 k s_2, y_1, \ldots, y_m, \widetilde{y}, \widetilde{w}) \end{array} \right\} \text{ in } \mathbb{M}[K, k].$$

By doing the above transformation from $\mathbb{P}[K]$ to $\mathbb{M}[K]$ backwards, it follows that

$$\exists y_1 \cdots \exists y_m \left\{ \bigwedge_{1\le i\le m} y_i \in \mathbb{P}[K_i, k]\ \wedge\ \phi^k_q(s_1 k s_2, y_1, \ldots, y_m, \widetilde{u}) \right\} \text{ in } \mathbb{P}[K, k],$$

where $s_i \in \mathbb{M} \cap \mathrm{IRR}(R)$ is identified with the group element it represents. Let us define a group homomorphism $h : \mathbb{P}[K, k] \to \mathbb{P}[K, k]$ by $h(k) = s_1^{-1} k s_2^{-1}$ and $h(x) = x$ for $x \in \mathbb{P}[K]$. First, note that h is injective (the homomorphism defined by $g(k) = s_1 k s_2$ defines an inverse). Thus, the truth value of all (negated) equations is preserved by h. Moreover, $\rho(s_1)q\rho(s_2) = q$ and hence, $\varrho^k_q(s_1^{-1} k s_2^{-1}) = \rho(s_1)^{-1} q \rho(s_2)^{-1} = q = \varrho^k_q(k)$ for every extension ϱ of ρ. Thus, all recognizable constraints are also preserved by h. Finally, $h(s_1 k s_2) = s_1 s_1^{-1} k s_2^{-1} s_2 = k$ in $\mathbb{P}[K, k]$. Hence, applying h to the above statement yields

$$\exists y_1 \cdots \exists y_m \left\{ \bigwedge_{1\le i\le m} y_i \in \mathbb{P}[K_i, k]\ \wedge\ \phi^k_q(k, y_1, \ldots, y_m, \widetilde{u}) \right\} \text{ in } \mathbb{P}[K, k].$$

5.5.3 Proof of Lemma 5.5.12

Recall that $\mathbb{M} \subseteq \mathbb{M}[K_i] \subseteq \mathbb{M}[K] \subseteq \mathbb{M}[K,k]$, $q \in S$, and $\widetilde{w} = (w_1, \ldots, w_N)$ with $w_i \in \mathbb{M} \cap \mathrm{IRR}(R)$ are already fixed. Let $D = (A \cup K \cup \overline{K} \cup \{k, \overline{k}\})^2 \setminus I$ be the dependence relation corresponding to $\mathbb{M}[K,k]$. The involution ι is totally defined on $\mathbb{M}[K,k]$. In the following we will write $\overline{t}$ instead of $\iota(t)$. Recall that (Σ, D_Σ) is assumed to be connected with $|\Sigma| > 1$. Let $\chi(x, y_1, \ldots, y_m, \widetilde{z})$ be an arbitrary Boolean formula with atomic predicates of the form $xy = z$, $x = \overline{y}$, $x = t$, $x \circ_\sigma y = z$, and $\varrho(x) = p$, where x, y, and z are variables, $t \in \mathbb{M}[K]$ is a constant (w.l.o.g. $|t| \leq 1$), $p \in S$, and $\varrho : \mathbb{M}[K] \to S$ is some extension of our basic homomorphism $\rho : \mathbb{M} \to S$. Since ρ was derived from a corresponding group homomorphism on $\mathbb{P}$, $s \overset{*}{\leftrightarrow}_R t$ implies $\rho(s) = \rho(t)$ for $s, t \in \mathbb{M}$. Let $W = \{w_1, \overline{w}_1, \ldots, w_N, \overline{w}_N\}$ and let d be the number of equations of the form $xy = z$ that occur in χ. Choose λ such that $|S|$ divides $\lambda - 1$ and $\lambda \geq 2d + 1$.

We start with the definition of some specific traces. A *chain* is a trace $t = a_1 a_2 \cdots a_\kappa$ such that $a_i \in A_{\sigma_i}$ $(1 \leq i \leq \kappa)$ and $[\sigma_1, \sigma_2, \ldots, \sigma_\kappa]$ is a path in the dependence graph (Σ, D_Σ). Thus, $t \in \mathbb{M} \cap \mathrm{IRR}(R)$ and $(a_i, a_{i+1}) \in D$ for $1 \leq i \leq \kappa - 1$.

Recall that we have fixed symbols $a, b, c \in A$ at the end of Section 5.5.1 such that $(a,b), (b,c) \in D$ and either a, b, and c belong to pairwise different A_σ or $a \neq \overline{a} = c$. It is possible that $(a,c) \in I$.

Lemma 5.5.13. *There exists a trace $\ell \in \mathbb{M} \cap \mathrm{IRR}(R)$ such that $\rho(\ell) = h$, $(\ell, t) \in D$ for every $t \neq \varepsilon$, and $\min(\ell) = \max(\ell) = \{a\}$.*

Proof. First, for every $x \in A$ we construct a trace $t(x) \in \mathbb{M} \cap \mathrm{IRR}(R)$ such that $\min(t(x)) = \{x\}$, $\max(t(x)) = \{\overline{x}\}$, and $\rho(t(x)) = 1$. If a, b, and c belong to pairwise different alphabets A_σ, then we set $t(x) = x\, s\, (b\, a)^{|S|} (c\, b)^{|S|}\, \overline{s}\, \overline{x}$, where $x\, s\, b$ is a chain from x to b, which exists since (Σ, D_Σ) is connected. On the other hand, if $a \neq \overline{a} = c$, then choose a chain $x\, s\, a$ and define $s(x) = x\, s\, (a\, b\, a)^{|S|}\, \overline{s}\, \overline{x}$. Since $a \neq \overline{a}$, we have $a^2 \to_R a'$ for some $a' \in A$. Then in $\mathbb{P}$ the element $s(x)$ equals $x\, s\, a\, (b\, a')^{|S|-1}\, b\, a\, \overline{s}\, \overline{x} \in \mathrm{IRR}(R)$ and we can choose the latter trace for $t(x)$.

Now we construct ℓ as follows:

- Select a trace $s = b_1 b_2 \cdots b_\kappa \in \mathrm{IRR}(R)$, $b_i \in A$, with $\rho(b_1 \cdots b_\kappa) = q$. Recall that ρ was assumed to be surjective, hence s exists.

- Let $u_1, \ldots, u_{\kappa+1} \in \mathbb{M} \cap \mathrm{IRR}(R)$ be chains, visiting every subset A_σ ($\sigma \in \Sigma$), such that the trace $au_1b_1u_2b_2 \cdots u_\kappa b_\kappa u_{\kappa+1}a$ is also a chain. These u_i exist, since (Σ, D_Σ) is connected.

- If $u_i = c_1 \cdots c_{\kappa_i}$ with $c_j \in A$, then define $v_i = t(c_1) \cdots t(c_{\kappa_i})$ (thus $\rho(v_i) = 1$).

- Finally, let $\ell = t(a)v_1b_1v_2b_2 \cdots v_\kappa b_\kappa v_{\kappa+1}t(\bar{a})$.

The construction implies that ℓ has indeed the desired properties. □

For the rest of the section let $\ell \in \mathbb{M}$ be some trace satisfying the properties from the previous lemma.

A *trace system of degree* n is a tuple $\mathcal{R} = (r_0, \ldots, r_\lambda)$ of $\lambda + 1$ traces $r_i = t_i(ba)^{|S|}(cb)^{|S|}$ with $t_i \in \{(ba)^{|S|}, (bc)^{|S|}\}^*$, $|t_i| = 2n|S|$ for some n large enough. The value of n will be made more precise later. Note that the traces r_i are irreducible and almost chains (only the single factor ac leads to commutation) with $\rho(r_i) = 1$. There are $2^{n(\lambda+1)}$ trace systems of degree n. We append the suffix $(ba)^{|S|}(cb)^{|S|}$ to every t_i in order to assure that every r_i starts and ends with b.

An *overlapping of two traces* $u, v \in \mathbb{M}$ is a trace s with $u = ts$ and $v = st'$ for some $t, t' \in \mathbb{M}$. The trace system $\mathcal{R} = (r_0, \ldots, r_\lambda)$ has *no long overlapping*, if

- the traces $r_0, \ldots, r_\lambda$ are pairwise different (then also $r_0, \overline{r}_0, \ldots, r_\lambda, \overline{r}_\lambda$ are pairwise different), and

- for all $0 \leq i, j \leq \lambda$, $u \in \{r_i, \overline{r}_i\}$, and $v \in \{r_j, \overline{r}_j\}$ we have: if s is an overlapping of u and v with $|s| \geq \frac{|r_i|-|\ell|}{2} = (n+2)|S| - \frac{|\ell|}{2}$, then $s = u = v$.

Note that this implies in particular that if $r_i\ell r_{i+1} = urv$ with $r \in \{r_j, \overline{r}_j\}$, then either $u = \varepsilon$ and $r_i = r$ or $v = \varepsilon$ and $r_j = r$, i.e., r cannot be properly contained in $r_i\ell r_{i+1}$.

The following lemma can be derived by standard techniques that random strings are incompressible, the formal proof is therefore omitted. The idea is that if $\mathcal{R}$ has a long overlapping, then, in case n is large enough, the description of $\mathcal{R}$ can be compressed to less than $n(\lambda + 1)$ bits. But this cannot happen for all systems $\mathcal{R}$.

Lemma 5.5.14. *There exists n_0 (depending only on λ and $|S|$) such that for all $n \geq n_0$ there exists a trace system of degree n without long overlapping.*

Remark 5.5.15. *Later, we will use $\mathcal{R}$ to construct a trace s, which can be replaced by the trace $s_1 k s_2$ in Lemma 5.5.12. An explicit construction of s without using the notion of random strings is given in [66].*

Let us fix a trace system $\mathcal{R} = (r_0, \ldots, r_\lambda)$ of degree n without long overlapping, where $2|r_i| + |\ell| = 4(n+2)|S| + |\ell| > |w|$ for all $w \in W$. For every $1 \leq i \leq \lambda$ define the length-reducing trace rewriting system

$$T_i = \{r_{i-1}\,\ell\,r_i \to r_{i-1}\,k\,r_i,\ \overline{r_i}\,\overline{\ell}\,\overline{r_{i-1}} \to \overline{r_i}\,\overline{k}\,\overline{r_{i-1}}\}.$$

We consider T_i as a trace rewriting system over our largest trace monoid $\mathbb{M}[K, k]$. Note that $W \cup A \subseteq \mathrm{IRR}(T_i)$ for all $w \in W$ by the choice of n, and that $s \to_{T_i} t$ implies also $\overline{s} \to_{T_i} \overline{t}$.

Lemma 5.5.16. *Every trace rewriting system T_i is confluent.*

Proof. Since T_i is terminating, we have to verify that T_i is locally confluent. Assume that $t \;{}_{T_i}\!\leftarrow s \to_{T_i} u$, where t (resp. u) results from s by an application of the rule $r_{i-1}\ell r_i \to r_{i-1} k r_i$, all other cases can be dealt analogously. Thus, there exist traces $t_1, t_2, u_1, u_2 \in \mathbb{M}[K, k]$ such that

$$s = t_1 r_{i-1}\,\ell\,r_i t_2 = u_1 r_{i-1}\,\ell\,r_i u_2 \text{ and } t = t_1 r_{i-1}\,k\,r_i t_2,\ u = u_1 r_{i-1}\,k\,r_i u_2.$$

Now we apply Levi's Lemma 2.7.1 to the identity $t_1 r_{i-1}\,\ell\,r_i t_2 = u_1 r_{i-1}\,\ell\,r_i u_2$. Recall that every r_j starts and ends with b. Hence, nonempty prefixes (resp. suffixes) of r_{i-1} (resp. r_i) are dependent. Moreover, ℓ is dependent from every nonempty trace. Thus, we obtain up to symmetry one of the following two diagrams:

u_2	ε	s_2	t_2
$r_{i-1}\,\ell\,r_i$	ε	$r_{i-1}\,\ell\,r_i$	ε
u_1	t_1	s_1	ε
	t_1	$r_{i-1}\,\ell\,r_i$	t_2

u_2	ε	ε	u_2
$r_{i-1}\,\ell\,r_i$	ε	s	s_2
u_1	t_1	s_1	v
	t_1	$r_{i-1}\,\ell\,r_i$	t_2

In the first case, $s_1 = \varepsilon = s_2$ and thus $t = u$. In the second case, we may assume that $s_1 \neq \varepsilon \neq s_2$, since otherwise we obtain a special case of the first diagram. Furthermore, if $s = \varepsilon$, then

$$t \to_{T_i} t_1\,r_{i-1}\,k\,r_i\,v\,r_{i-1}\,k\,r_i\,u_2 \;{}_{T_i}\!\leftarrow u.$$

Thus, assume that also $s \neq \varepsilon$. Since $r_{i-1}\,\ell\,r_i = s_1\,s = s\,s_2$ with $s_1 \neq \varepsilon \neq s_2$, and $\mathcal{R}$ has no long overlapping, there exist traces r and r' such that $s_1 = r_{i-1}\,\ell\,r$, $s_2 = r'\,\ell\,r_i$, $r_i = r\,s$, $r_{i-1} = s\,r'$. Since $(v,s) \in I$, we obtain

$$\begin{aligned} t = t_1 r_{i-1}\,k\,r_i t_2 &= t_1\,r_{i-1}\,k\,r\,s\,v\,r'\,\ell\,r_i\,u_2 \\ &= t_1\,r_{i-1}\,k\,r\,v\,s\,r'\,\ell\,r_i\,u_2 \\ &\to_{T_i} t_1\,r_{i-1}\,k\,r\,v\,s\,r'\,k\,r_i\,u_2 \\ &= t_1\,r_{i-1}\,k\,r\,s\,v\,r'\,k\,r_i\,u_2 \\ &\,{}_{T_i}\!\leftarrow t_1\,r_{i-1}\,\ell\,r\,s\,v\,r'\,k\,r_i\,u_2 \\ &= t_1\,r_{i-1}\,\ell\,r\,v\,s\,r'\,k\,r_i\,u_2 = u_1 r_{i-1}\,k\,r_i u_2 = u. \end{aligned}$$

Thus, T_i is confluent. □

The previous lemma implies that for every $1 \leq i \leq \lambda$, every trace $s \in \mathbb{M}[K,k]$ has a unique normal form $\mathrm{NF}_{T_i}(s) \in \mathrm{IRR}(T_i)$. In the following, we briefly write $\mathrm{NF}_i(s)$ for $\mathrm{NF}_{T_i}(s)$. The following lemma is easy to verify. For the last point note that $\rho(\ell) = q = \rho_q^k(k)$.

Lemma 5.5.17. *For every $1 \leq i \leq \lambda$ and $s \in \mathbb{M}[K]$ we have:*

- $\mathrm{NF}_i(s) = s$ *if* $|s| \leq 1$ *or* $s \in W$, *in particular, if* $a \circ_\sigma b = c$ *for* $a,b,c \in A_\sigma$ *($\sigma \in \Sigma$), then also* $\mathrm{NF}_i(a) \circ_\sigma \mathrm{NF}_i(b) = \mathrm{NF}_i(c)$,
- $\overline{\mathrm{NF}_i(s)} = \mathrm{NF}_i(\overline{s})$, *and*
- $\varrho(s) = \varrho_q^k(\mathrm{NF}_i(s))$ *for every extension* $\varrho : \mathbb{M}[K] \to S$ *of* $\rho : \mathbb{M} \to S$.

Thus, every normal form mapping NF_i preserves constants, the involution $^-$, and recognizable constraints. On the other hand, concatenation in $\mathbb{M}[K]$ is in general not preserved, but the following statement will suffice:

Lemma 5.5.18. *Let $u,v \in \mathbb{M}[K]$. There are at most two $i \in \{1,\ldots,\lambda\}$ such that $\mathrm{NF}_i(u)\mathrm{NF}_i(v) \neq \mathrm{NF}_i(uv)$.*

Proof. Assume that $1 \leq i \leq \lambda$ is such that $\mathrm{NF}_i(u)\mathrm{NF}_i(v) \in \mathrm{RED}(T_i)$. We only consider the case that $\mathrm{NF}_i(u)\mathrm{NF}_i(v) = s r_i \ell r_{i+1} t$ for some $s,t \in \mathbb{M}[K]$. Due to the dependencies between nonempty suffixes and prefixes of r_i, ℓ, and r_{i+1}, we obtain one of the following three diagrams (where $r'_j \neq \varepsilon \neq r''_j$ for $j \in \{i-1,i\}$):

$\mathrm{NF}_i(v)$	s_2	r''_{i-1}	ℓ	r_i	t
$\mathrm{NF}_i(u)$	s_1	r'_{i-1}	ε	ε	ε
	s	r_{i-1}	ℓ	r_i	t

$\mathrm{NF}_i(v)$	ε	ε	ε	r''_i	t_2
$\mathrm{NF}_i(u)$	s	r_{i-1}	ℓ	r'_i	t_1
	s	r_{i-1}	ℓ	r_i	t

$\mathrm{NF}_i(v)$	s_2	ε	ℓ_2	r_i	t_2
$\mathrm{NF}_i(u)$	s_1	r_{i-1}	ℓ_1	ε	t_1
	s	r_{i-1}	ℓ	r_i	t

Since every r_j starts and ends with b, it follows that $(s_2, b) \in I$ (resp. $(t_1, b) \in I$) in the first and third (resp. second and third) diagram. Let π denote the homomorphism that projects onto the subalphabet $\{a, \overline{a}, b, \overline{b}, c, \overline{c}, k, \overline{k}\}$. Thus, $\pi(s_2) = \pi(t_1) = \varepsilon$. It follows that one of the following three cases holds, where $x, y \in \{a, \overline{a}, b, \overline{b}, c, \overline{c}, k, \overline{k}\}^*$ and $\ell' = \pi(\ell)$:

- $\pi(\mathrm{NF}_i(u)) = xr$ and $\pi(\mathrm{NF}_i(v)) = r'\ell' r_i y$ where $r_{i-1} = rr'$
- $\pi(\mathrm{NF}_i(u)) = xr_{i-1}\ell' r$ and $\pi(\mathrm{NF}_i(v)) = r'y$, where $r_i = rr'$
- $\pi(\mathrm{NF}_i(u)) = xr_{i-1}\ell'_1$ and $\pi(\mathrm{NF}_i(v)) = \ell'_2 r_i y$, where $\ell' = \ell'_1\ell'_2$

But then there are also $x', y' \in \{a, \overline{a}, b, \overline{b}, c, \overline{c}\}^*$ with

- $\pi(u) = x'r$ and $\pi(v) = r'\ell' r_i y'$ where $r_{i-1} = rr'$ or
- $\pi(u) = x'r_{i-1}\ell' r$ and $\pi(v) = r'y'$, where $r_i = rr'$ or
- $\pi(u) = x'r_{i-1}\ell'_1$ and $\pi(v) = \ell'_2 r_i y'$, where $\ell' = \ell'_1\ell'_2$

The traces x' and y' result from x and y, respectively, by replacing every occurrence of k and $\overline{k}$, respectively, by ℓ' and $\overline{\ell'}$, respectively. Thus $\pi(u) = x'z_1$, $\pi(v) = z_2 y'$, and $z_1 z_2 = r_{i-1}\ell' r_i$. Now assume that this holds for three different i_1, i_2, and i_3. Then it is easy to see that two of the three traces $r_{j-1}\ell' r_j$ ($j \in \{i_1, i_2, i_3\}$) have a "long" overlapping, contradicting the fact that $\mathcal{R}$ has no long overlapping. □

Since moreover $\mathrm{NF}_i(u) = \mathrm{NF}_i(v)$ implies $u = v$ for all $u, v \in \mathbb{M}[K]$, we obtain the following lemma – recall that $\lambda \geq 2d + 1$, where d is the number of equations in the formula χ.

Lemma 5.5.19. *Let $x_j, y_j, z_j \in \mathbb{M}[K]$ for $1 \leq j \leq d$. Then there exists $1 \leq i \leq \lambda$ such that for all $1 \leq j \leq d$, we have $x_j y_j = z_j$ if and only if $\mathrm{NF}_i(x_j)\mathrm{NF}_i(y_j) = \mathrm{NF}_i(z_j)$.*

Now we are able to prove Lemma 5.5.12: Assume that

$$\forall x \in \mathrm{IRR}(R) \cap \rho^{-1}(q)\ \exists y_1 \cdots \exists y_m \left\{ \begin{array}{l} \bigwedge\limits_{1 \leq i \leq m} y_i \in \mathbb{M}[K_i] \cap \mathrm{IRR}(R) \\ \wedge\ \chi(x, y_1, \ldots, y_m, \widetilde{w}) \end{array} \right\} \text{ in } \mathbb{M}[K].$$

Let $s = r_0\, \ell\, r_1\, \ell \cdots r_{\lambda-1}\, \ell\, r_\lambda \in \mathbb{M} \cap \mathrm{IRR}(R)$. Since $\rho(r_i) = 1$ and λ was chosen such that $|S|$ is a divisor of $\lambda - 1$, we have $\rho(s) = \rho(\ell^\lambda) = q^\lambda = q$. Thus, there exist traces $t_i \in \mathbb{M}[K_i] \cap \mathrm{IRR}(R)$, $1 \leq i \leq m$, with $\chi(s, t_1, \ldots, t_m, \widetilde{w})$ in $\mathbb{M}[K]$. By Lemma 5.5.17 and Lemma 5.5.19 there exists $1 \leq j \leq \lambda$ such that $\chi_q^k(\mathrm{NF}_j(s), \mathrm{NF}_j(t_1), \ldots, \mathrm{NF}_j(t_m), \widetilde{w})$ in $\mathbb{M}[K, k]$. Note that we can write $\mathrm{NF}_j(s) = s_1\, k\, s_2$ for $s_1, s_2 \in \mathbb{M} \cap \mathrm{IRR}(R)$ such that $\rho(s_1)q\rho(s_2) = \rho(s) = q$. Thus,

$$\exists y_1 \cdots \exists y_m \left\{ \begin{array}{l} \bigwedge\limits_{1 \leq i \leq m} y_i \in \mathbb{M}[K_i, k] \cap \mathrm{IRR}(R) \\ \wedge\ \chi_q^k(s_1 k s_2, y_1, \ldots, y_m, \widetilde{w}) \end{array} \right\} \text{ in } \mathbb{M}[K, k].$$

5.6 Open problems

Concerning existential theories, the following problems might deserve further investigations:

- Is the additional exponential summand in Theorem 5.4.10 unavoidable?
- Is Assumption 5.4.8 necessary in Theorem 5.4.10?
- Is the existential theory of a context-free (i.e., virtually free) group decidable? Is at least the solvability of a single equation decidable for a context-free group? Note that every context-free group is hyperbolic, but the fact that the existential theory of a torsion-free hyperbolic group is decidable [184] does not help with respect to this problem: By a result of Stallings [196], a torsion-free context-free group is already a free group.

- Is the existential theory of an automatic group undecidable? At least for asynchronous automatic groups (see [80]) this is the case, in fact already conjugacy is in general undecidable for asynchronous automatic groups [14]. Further results concerning undecidable existential theories for groups and monoids can be found in [167, 174].

For positive theories it remains open whether an elementary reduction is possible in Theorem 5.5.3 in case (Σ, D_Σ) is not connected. One might also investigate, whether the positive theory of a torsion-free hyperbolic group is decidable. Further results on positive theories can be found in [173, 210].

Finally, one may hope to get decidability results for full first-order theories of restricted graph products, like for instance graph groups. One approach might be to generalize the techniques developed by Kharlampovich and Myasnikov for solving Tarski's problem on free groups to broader classes. We should mention here that elementary decision procedures cannot be expected for full first-order theories. By a result of Semenov [185], the full first-order theory of a free group of rank 2 is nonelementary.

Chapter 6

Conclusion

In this work we considered three problem domains for infinite monoids:

- the word problem
- first-order and monadic second-order logic over Cayley-graphs
- word equations and theories of word equations

For each of these three topics we proved new complexity and decidability results – but many exciting open problems and further research directions remain (see also Section 3.9, 4.8, and 5.6):

Concerning the complexity of word problems, our results from Chapter 3 mainly apply to monoids. This is not really surprising: It is in general much harder to encode computation steps in a faithful way in groups – a phenomenon, which is also documented by the complicated proof of the Novikov-Boone-Theorem on the undecidability of the word problem for groups. In comparison to this, the undecidability proof for the word problem for monoids is straight-forward. On the complexity theoretical side, this means that it is in general quite hard to prove lower bounds for word problems on groups. In fact, some of the results that were shown for monoids in Chapter 3 are unlikely (under reasonable assumptions from complexity theory) to hold for groups. For instance, a group that can be presented by a weight-reducing and confluent presentation is already context-free [62, 128], hence its word problem can be decided in logarithmic space [171, Thm. 6.8]. Thus, LOGDCFL-hardness, which was shown for the corresponding monoid case (Theorem 3.3.4), is unlikely, because it would imply that deterministic

context-free languages can be recognized in logarithmic space. On the other hand, some of the results from Chapter 3 might be transfered to groups as well, for instance the P-completeness of the word problem for automatic monoids (Theorem 3.8.2) may also hold for automatic groups.

We studied theories of Cayley-graphs by investigating more general classes of structures, in particular infinite graphs. During the last years, infinite graphs and their logical properties received a lot of attention in the verification community, see [30, 36, 205] for an overview. Infinite graphs are the canonical model for systems with an infinite number of system states, like unbounded stacks, unbounded communication channels, or systems defined by unrestricted parameters: Nodes represent system states and edges represent transitions between them. Our results from Section 4.3.4 on graphs with decidable MSO theories can be seen as a contribution to this research direction. Moreover, due to their generality, Cayley-graphs of infinite monoids may be in particular useful for the specification of infinite state systems: It is well-known that the theory of finite automata can be based on finite monoids – monoid elements are states and transitions between them result from right-multiplication. If we replace finite monoids by infinite monoids in this approach, we obtain basically Cayley-graphs of infinite monoids, which can therefore be viewed as an infinite version of finite automata. In this context, closure results in the spirit of Theorem 4.7.5 are in particular useful, because they allow a modular specification and verification of system properties.

Word equations are a very general concept, and recently found applications in several other fields. For instance, in [179] it was shown with the help of word equations that the problem of recognizing string graphs is NP-complete. This problem was open for more than 20 years and has applications in graph drawing for instance. By extending the fundamental decidability results of Makanin [131, 132, 133] and at the same time improving the complexity of the corresponding algorithms, the range of further applications will be certainly extended. Concerning complexity, the major open problem currently, is whether the solvability of word equations in a free monoid is NP-complete. NP-hardness of this problem follows easily from the NP-hardness of integer programming, thus it remains to prove membership in NP.

The investigation of word equations may also benefit from a better understanding of Cayley-graphs. In Section 5.6 we asked, whether solvability of equations over context-free and automatic groups, respectively, is decidable. In order to answer these questions, it might be helpful to study the Cayley-

graphs of the corresponding groups in more detail – an approach that was applied successfully in [168] for instance.

Index

Bibliography

[1] I. J. Aalbersberg and H. J. Hoogeboom. Characterizations of the decidability of some problems for regular trace languages. *Mathematical Systems Theory*, 22:1–19, 1989.

[2] S. I. Adjan. *Defining relations and algorithmic problems for groups and semigroups*, volume 85 of *Proceedings of the Steklov Institute of Mathematics*. American Mathematical Society, 1967.

[3] C. Alvarez and R. Greenlaw. A compendium of problems complete for symmetric logarithmic space. *Electronic Colloquium on Computational Complexity*, Report No. TR96-039, 1996.

[4] A. V. Anisimov. Group languages. *Kibernetika*, 4:18–24, 1971. In Russian; English translation in *Cybernetics 4*, 594–601, 1973.

[5] R. Armoni, A. Ta-Shma, A. Widgerson, and S. Zhou. An $O(\log(n)^{4/3})$ space algorithm for (s,t) connectivity in undirected graphs. *Journal of the Association for Computing Machinery*, 47(2):294–311, 2000.

[6] J. Avenhaus and K. Madlener. Subrekursive Komplexität bei Gruppen: I. Gruppen mit vorgeschriebener Komplexität. *Acta Informatica*, 9(2):87–104, 1978. In German.

[7] F. Baader and W. Snyder. Unification theory. In J. Robinson and A. Voronkov, editors, *Handbook of Automated Reasoning*, volume I, pages 447–533. Elsevier Science Publishers, 2001.

[8] L. Babai. Automorphism groups, isomorphism, reconstruction. In R. L. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics*, volume II, chapter 27, pages 1447–1540. Elsevier Science Publishers, 1995.

[9] D. A. M. Barrington and J. Corbet. On the relative complexity of some languages in NC^1. *Information Processing Letters*, 32:251–256, 1989.

[10] D. A. M. Barrington, N. Immerman, and H. Straubing. On uniformity within NC^1. *Journal of Computer and System Sciences*, 41:274–306, 1990.

[11] A. Baudisch. Kommutationsgleichungen in semifreien Gruppen. *Acta Mathematica Academiae Scientiarum Hungaricae*, 29:235–249, 1977. In German.

[12] A. Baudisch. Subgroups of semifree groups. *Acta Mathematica Academiae Scientiarum Hungaricae*, 38:19–28, 1981.

[13] G. Bauer and F. Otto. Finite complete rewriting systems and the complexity of the word problem. *Acta Informatica*, 21:521–540, 1984.

[14] G. Baumslag, S. Gersten, M. Shapiro, and H. Short. Automatic groups and amalgams. *Journal of Pure Applied Algebra*, 76(3):229–316, 1991.

[15] M. Beaudry, M. Holzer, G. Niemann, and F. Otto. McNaughton families of languages. *Theoretical Computer Science*, 290(3):1581–1628, 2003.

[16] B. Benninghofen, S. Kemmerich, and M. M. Richter. *Systems of Reductions*. Number 277 in Lecture Notes in Computer Science. Springer, 1987.

[17] L. Berman. The complexity of logical theories. *Theoretical Computer Science*, 11:71–77, 1980.

[18] J. Berstel. *Transductions and context–free languages*. Teubner Studienbücher, Stuttgart, 1979.

[19] A. Blumensath. Prefix-recognizable graphs and monadic second-order logic. Technical Report 2001-06, RWTH Aachen, Department of Computer Science, 2001.

[20] A. Blumensath and E. Grädel. Automatic structures. In *Proceedings of the 15th Annual IEEE Symposium on Logic in Computer Science (LICS'2000)*, pages 51–62. IEEE Computer Society Press, 2000.

[21] H. L. Bodlaender. A note on domino treewidth. *Discrete Mathematics & Theoretical Computer Science*, 3(4):141–150, 1999.

[22] H. L. Bodlaender and J. Engelfriet. Domino treewidth. *Journal of Algorithms*, 24(1):94–123, 1997.

[23] R. V. Book. Confluent and other types of Thue systems. *Journal of the Association for Computing Machinery*, 29(1):171–182, 1982.

[24] R. V. Book. Homogeneous Thue systems and the Church–Rosser property. *Discrete Mathematics*, 48:137–145, 1984.

[25] R. V. Book, M. Jantzen, B. Monien, C. P. Ó'Dúnlaing, and C. Wrathall. On the complexity of word problems in certain Thue systems. In J. Gruska and M. Chytil, editors, *Proceedings of the 10th International Symposium on Mathematical Foundations of Computer Science (MFCS'81), Štrbské Pleso (Czechoslovakia)*, number 118 in Lecture Notes in Computer Science, pages 216–223. Springer, 1981.

[26] R. V. Book and C. P. Ó'Dúnlaing. Testing for the Church–Rosser property (note). *Theoretical Computer Science*, 16:223–229, 1981.

[27] R. V. Book and F. Otto. *String–Rewriting Systems*. Springer, 1993.

[28] W. W. Boone. The word problem. *Annals of Mathematics (2)*, 70:207–265, 1959.

[29] W. W. Boone, F. B. Cannonito, and R. C. Lyndon. *Word Problems*. North-Holland, 1973.

[30] A. Bouajjani. Languages, rewriting systems, and verification of infinite-state systems. In *Proceedings of the 28th International Colloquium on Automata, Languages and Programming (ICALP 01), Crete (Greece)*, number 2076 in Lecture Notes in Computer Science, pages 24–39. Springer, 2001.

[31] N. Brady and J. Meier. Connectivity at infinity for right angled Artin groups. *Transactions of the American Mathematical Society*, 353:117–132, 2001.

[32] J. Büchi. On a decision method in restricted second order arithmetics. In E. Nagel et al., editors, *Proceedings of the International Congress on Logic, Methodology and Philosophy of Science*, pages 1–11. Stanford University Press, Stanford, 1960.

[33] G. Buntrock and K. Loryś. On growing context-sensitive languages. In W. Kuich, editor, *Proceedings of the 19th International Colloquium on Automata, Languages and Programming (ICALP 92), Vienna (Austria)*, number 623 in Lecture Notes in Computer Science, pages 77–88. Springer, 1992.

[34] G. Buntrock and K. Loryś. The variable membership problem: Succinctness versus complexity. In P. Enjalbert, E. W. Mayr, and K. W. Wagner, editors, *Proceedings of the 11th Annual Symposium on Theoretical Aspects of Computer Science (STACS 94), Caen (France)*, number 775 in Lecture Notes in Computer Science, pages 595–606. Springer, 1994.

[35] G. Buntrock and F. Otto. Growing context-sensitive languages and Church-Rosser languages. *Information and Computation*, 141:1–36, 1998.

[36] O. Burkart, D. Caucal, F. Moller, and B. Steffen. Verification on infinite structures. In J. A. Bergstra, A. Ponse, and S. A. Smolka, editors, *Handbook of process algebra*, pages 545–623. Elsevier, 2001.

[37] S. R. Buss. The Boolean formula value problem is in ALOGTIME. In *Proceedings of the 19th Annual Symposium on Theory of Computing (STOC 87)*, pages 123–131. ACM Press, 1987.

[38] J.-y. Cai. Parallel computation over hyperbolic groups. In *Proceedings of the 24th Annual Symposium on Theory of Computing (STOC 92)*, pages 106–115. ACM Press, 1992.

[39] H. Calbrix and T. Knapik. A string-rewriting characterization of Muller and Schupp's context-free graphs. In V. Arvind and R. Ramanujam, editors, *Proceedings of the 18th International Conference on Foundations of Software Technology and Theoretical Computer Science*, number 1530 in Lecture Notes in Computer Science, pages 331–342. Springer, 1999.

[40] C. M. Campbell, E. F. Robertson, N. Ruškuc, and R. M. Thomas. Automatic semigroups. *Theoretical Computer Science*, 250(1-2):365–391, 2001.

[41] F. Cannonito. Hierarchies of computable groups and the word problem. *Journal of Symbolic Logic*, 31:376–392, 1966.

[42] E. W. Cardoza. Computational complexity of the word problem for commutative semigroups. Technical Report MAC Technical Memorandum 67, MIT, 1975.

[43] P. Cartier and D. Foata. *Problèmes combinatoires de commutation et réarrangements.* Number 85 in Lecture Notes in Mathematics. Springer, 1969.

[44] D. Caucal. On infinite transition graphs having a decidable monadic theory. In F. M. auf der Heide and B. Monien, editors, *Proceedings of the 23nd International Colloquium on Automata, Languages and Programming (ICALP'96), Paderborn (Germany)*, number 1099 in Lecture Notes in Computer Science, pages 194–205. Springer, 1996.

[45] A. Cayley. On the theory of groups. *Proceedings of the London Mathematical Society (1)*, 9:126–133, 1878.

[46] A. K. Chandra, D. C. Kozen, and L. J. Stockmeyer. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133, 1981.

[47] S. Cho and D. T. Huynh. The complexity of membership for deterministic growing context-sensitive grammars. *International Journal of Computer Mathematics*, 37:185–188, 1990.

[48] K. J. Compton and C. W. Henson. A uniform method for proving lower bounds on the computational complexity of logical theories. *Annals of Pure and Applied Logic*, 48:1–79, 1990.

[49] S. A. Cook. Characterizations of pushdown machines in terms of time-bounded computers. *Journal of the Association for Computing Machinery*, 18(1):4–18, 1971.

[50] S. A. Cook. Deterministic CFL's are accepted simultaneously in polynomial time and log squared space. In *Proceedings of the 11th Annual*

Symposium on Theory of Computing (STOC 79), pages 338–345. ACM Press, 1979.

[51] S. A. Cook. A taxonomy of problems with fast parallel algorithms. *Information and Control*, 64:2–22, 1985.

[52] M. Coornaert, T. Delzant, and A. Papadopoulos. *Géométrie et théorie des groupes*. Number 1441 in Lecture Notes in Mathematics. Springer, 1990.

[53] B. Courcelle. The monadic second-order logic of graphs, II: Infinite graphs of bounded width. *Mathematical Systems Theory*, 21:187–221, 1989.

[54] B. Courcelle. The monadic second-order logic of graphs VI: On several representations of graphs by relational structures. *Discrete Applied Mathematics*, 54:117–149, 1994.

[55] B. Courcelle. The expression of graph properties and graph transformations in monadic second-order logic. In G. Rozenberg, editor, *Handbook of graph grammars and computing by graph transformation, Volume 1 Foundations*, pages 313–400. World Scientific, 1997.

[56] B. Courcelle and S. Olari. Upper bounds on the clique-width of graphs. *Discrete Applied Mathematics*, 101:77–114, 2000.

[57] G. D'Agostino. Cayley graphs of virtually free groups. *International Journal of Algebra and Computation*, 3(2):189–199, 1993.

[58] E. Dahlhaus and M. K. Warmuth. Membership for growing context-sensitive grammars is polynomial. *Journal of Computer and System Sciences*, 33:456–472, 1986.

[59] M. Dehn. Über die Toplogie des dreidimensionalen Raumes. *Mathematische Annalen*, 69:137–168, 1910. In German.

[60] W. Dicks and M. J. Dunwoody. *Groups Acting on Graphs*. Cambridge University Press, 1989.

[61] V. Diekert. Some properties of weight-reducing presentations. Report TUM-I8710, Technical University Munich, 1987.

[62] V. Diekert. Some remarks on presentations by finite Church-Rosser Thue systems. In F.-J. Brandenburg, G. Vidal-Naquet, and M. Wirsing, editors, *Proceedings of the 17th Annual Symposium on Theoretical Aspects of Computer Science (STACS 87), Passau (Germany)*, number 247 in Lecture Notes in Computer Science, pages 272–285. Springer, 1987.

[63] V. Diekert. *Combinatorics on Traces*. Number 454 in Lecture Notes in Computer Science. Springer, 1990.

[64] V. Diekert. Makanin's algorithm. In *Algebraic Combinatorics on Words*, pages 342–390. Cambridge University Press, 2001.

[65] V. Diekert, C. Gutiérrez, and C. Hagenah. The existential theory of equations with rational constraints in free groups is PSPACE-complete. In A. Ferreira and H. Reichel, editors, *Proceedings of the 18th Annual Symposium on Theoretical Aspects of Computer Science (STACS 2001), Dresden (Germany)*, number 2010 in Lecture Notes in Computer Science, pages 170–182. Springer, 2001.

[66] V. Diekert and M. Lohrey. Existential and positive theories of equations in graph products. In H. Alt and A. Ferreira, editors, *Proceedings of the 19th Annual Symposium on Theoretical Aspects of Computer Science (STACS 2002), Juan les Pins (France)*, number 2285 in Lecture Notes in Computer Science, pages 501–512. Springer, 2002.

[67] V. Diekert and M. Lohrey. A note on the existential theory of equations in plain groups. *International Journal of Algebra and Computation*, 12(1 & 2):1–7, 2002.

[68] V. Diekert, Y. Matiyasevich, and A. Muscholl. Solving word equations modulo partial commutations. *Theoretical Computer Science*, 224(1–2):215–235, 1999.

[69] V. Diekert and A. Muscholl. Solvability of equations in free partially commutative groups is decidable. In F. Orejas, P. G. Spirakis, and J. van Leeuwen, editors, *Proceedings of the 28th International Colloquium on Automata, Languages and Programming (ICALP 2001), Crete (Greece)*, number 2076 in Lecture Notes in Computer Science, pages 543–554. Springer, 2001.

[70] V. Diekert and G. Rozenberg, editors. *The Book of Traces.* World Scientific, 1995.

[71] R. Diestel. *Graph Theory, Second Edition.* Springer, 2000.

[72] G. Ding and B. Oporowski. Some results on tree decomposition of graphs. *Journal of Graph Theory*, 20:481–499, 1995.

[73] C. Droms. Graph groups, coherence and three-manifolds. *Journal of Algebra*, 106(2):484–489, 1985.

[74] M. J. Dunwoody. Cutting up graphs. *Combinatorica*, 2(1):15–23, 1981.

[75] M. J. Dunwoody. The accessibility of finitely presented groups. *Inventiones Mathematicae*, 81:449–457, 1985.

[76] V. G. Durnev. Undecidability of the positive $\forall\exists^3$-theory of a free semigroup. *Sibirsky Matematicheskie Jurnal*, 36(5):1067–1080, 1995. English translation.

[77] P. W. Dymond and W. L. Ruzzo. Parallel RAMs with owned global memory and deterministic context-free language recognition. *Journal of the Association for Computing Machinery*, 47:16–45, 2000.

[78] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory.* Springer, 1991.

[79] C. C. Elgot and M. O. Rabin. Decidability and undecidability of extensions of second (first) order theory of (generalized) successor. *Journal of Symbolic Logic*, 31(2):169–181, 1966.

[80] 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, Boston, 1992.

[81] S. Feferman and R. L. Vaught. The first order properties of products of algebraic systems. *Fundamenta Mathematicae*, 47:57–103, 1959.

[82] J. Ferrante. *Some upper and lower bounds on decision procedures in logic.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1974.

[83] J. Ferrante and C. Rackoff. *The Computational Complexity of Logical Theories*. Number 718 in Lecture Notes in Mathematics. Springer, 1979.

[84] C. Frougny and J. Sakarovitch. Synchronized rational relations of finite and infinite words. *Theoretical Computer Science*, 108(1):45–82, 1993.

[85] M. Fürer. *Nicht-elementare untere Schranken in der Automatentheorie*. PhD thesis, ETH Zürich, 1978. In German.

[86] H. Gaifman. On local and nonlocal properties. In J. Stern, editor, *Logic Colloquium '81*, pages 105–135. North Holland, 1982.

[87] E. R. Green. *Graph Products of Groups*. PhD thesis, The University of Leeds, 1990.

[88] M. Gromov. Hyperbolic groups. In S. M. Gersten, editor, *Essays in Group Theory*, number 8 in MSRI Publ., pages 75–263. Springer, 1987.

[89] Y. Gurevich. Monadic second-order theories. In J. Barwise and S. Feferman, editors, *Model-Theoretic Logics*, pages 479–506. Springer, 1985.

[90] C. Gutiérrez. Satisfiability of word equations with constants is in exponential space. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS 98)*, pages 112–119. IEEE Computer Society Press, 1998.

[91] C. Gutiérrez. Satisfiability of equations in free groups is in PSPACE. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing (STOC'2000)*, pages 21–27. ACM Press, 2000.

[92] R. H. Haring-Smith. Groups and simple languages. *Transactions of the American Mathematical Society*, 279:337–356, 1983.

[93] T. Herbst and R. M. Thomas. Group presentations, formal languages and characterizations of one-counter groups. *Theoretical Computer Science*, 112:187–213, 1983.

[94] S. Hermiller and J. Meier. Algorithms and geometry for graph products of groups. *Journal of Algebra*, 171:230–257, 1995.

[95] W. Hesse. Division is in uniform TC^0. In F. Orejas, P. G. Spirakis, and J. van Leeuwen, editors, *Proceedings of the 28th International Colloquium on Automata, Languages and Programming (ICALP 2001), Crete (Greece)*, number 2076 in Lecture Notes in Computer Science, pages 104–114. Springer, 2001.

[96] W. Hodges. *Model Theory*. Cambridge University Press, 1993.

[97] M. Hoffmann. *Automatic semigroups*. PhD thesis, University of Leicester, Department of Mathematics and Computer Science, 2000.

[98] D. Holt. Word-hyperbolic groups have real-time word problem. *International Journal of Algebra and Computation*, 10:221–227, 2000.

[99] J. F. P. Hudson. Regular rewrite systems and automatic structures. In *Semigroups, Automata and Languages*, pages 145–152. World Scientific, 1998.

[100] G. Huet and D. Lankford. On the uniform halting problem for term rewriting systems. Report Lab. Report No. 283, INRIA, Le Chesnay, France, 1978.

[101] D. T. Huynh. Complexity of the word problem for commutative semigroups of fixed dimension. *Acta Informatica*, 22:421–432, 1985.

[102] D. T. Huynh. The complexity of the membership problem for two subclasses of polynomial ideals. *SIAM Journal on Computing*, 15(2):581–594, 1986.

[103] N. Immerman. Languages that capture complexity classes. *SIAM Journal on Computing*, 16(4):760–778, 1987.

[104] M. Jantzen. Confluent string rewriting. In *EATCS Monographs on theoretical computer science*, volume 14. Springer, 1988.

[105] D. Kapur, M. S. Krishnamoorthy, R. McNaughton, and P. Narendran. An $O(|T|^3)$ algorithm for testing the Church-Rosser property of Thue systems. *Theoretical Computer Science*, 35(1):109–114, 1985.

[106] D. Kapur and P. Narendran. The Knuth-Bendix completion procedure and Thue systems. *SIAM Journal on Computing*, 14(3):1052–1072, 1985.

[107] R. M. Keller. Parallel program schemata and maximal parallelism I. Fundamental results. *Journal of the Association for Computing Machinery*, 20(3):514–537, 1973.

[108] O. Kharlampovich and A. Myasnikov. Tarski's problem about the elementary theory of free groups has a positive solution. *Electronic Research Announcements of the American Mathematical Society*, 4:101–108, 1998.

[109] B. Khoussainov and A. Nerode. Automatic presentations of structures. In *LCC: International Workshop on Logic and Computational Complexity*, number 960 in Lecture Notes in Computer Science, pages 367–392, 1994.

[110] T. Knapik and H. Calbrix. The graphs of finite monadic semi-Thue systems have a decidable monadic second-order theory. In C. S. Calude and M. J. Dinneen, editors, *Combinatorics, Computation and Logic'99*, pages 273–285. Springer, 1999.

[111] T. Knapik and H. Calbrix. Thue specifications and their monadic second-order properties. *Fundamenta Informaticae*, 39:305–325, 1999.

[112] D. E. Knuth and P. B. Bendix. Simple Word Problems in Universal Algebras. In J. Leech, editor, *Computational Problems in Abstract Algebras*, pages 263–297. Pergamon Press, Elmsford N.Y., 1967.

[113] A. Kościelski and L. Pacholski. Makanin's algorithm is not primitive recursive. *Theoretical Computer Science*, 191(1-2):145–156, 1998.

[114] D. Kozen. Lower bounds for natural proof systems. In *Proceedings of the 18th Annual Symposium on Foundations of Computer Science (FOCS 77)*, pages 254–266. IEEE Computer Society Press, 1977.

[115] K. Krithivasan and P. Narendran. On the membership problem for some grammars. Technical Report CS-TR-1787, University of Maryland, 1987.

[116] D. Kuske and M. Lohrey. Decidable theories of Cayley-graphs. In H. Alt and M. Habib, editors, *Proceedings of the20th Annual Symposium on Theoretical Aspects of Computer Science (STACS 2003),*

Berlin (Germany), number 2607 in Lecture Notes in Computer Science, pages 463–474. Springer, 2003.

[117] P. Le Chenadec. *Canonical Forms in Finitely Presented Algebras.* Pitman-Wiley, 1986.

[118] A. A. Letičevskiĭ and L. B. Smikun. On a class of groups with solvable problem of automata equivalence. *Dokl. Akad. Nauk SSSR*, 227:36–38, 1976. In Russian; English translation in *Sov. Math., Dokl. 17*, 341-344, 1976.

[119] H. R. Lewis and C. H. Papadimitriou. Symmetric space-bounded computation. *Theoretical Computer Science*, 19(2):161–187, 1982.

[120] R. J. Lipton and Y. Zalcstein. Word problems solvable in logspace. *Journal of the Association for Computing Machinery*, 24(3):522–526, 1977.

[121] J. Loeffler, J. Meier, and J. Worthington. Graph products and Cannon pairs. *International Journal of Algebra and Computation*, 12(6):747–754, 2002.

[122] M. Lohrey. Complexity results for confluence problems. In M. Kutylowski, L. Pacholski, and T. Wierzbicki, editors, *Proceedings of the 24th International Symposium on Mathematical Foundations of Computer Science (MFCS'99), Szklarska Poreba (Poland)*, number 1672 in Lecture Notes in Computer Science, pages 114–124. Springer, 1999.

[123] M. Lohrey. *Das Konfluenzproblem für Spurersetzungssysteme.* PhD thesis, Universität Stuttgart, 1999. In German.

[124] M. Lohrey. Word problems and confluence problems for restricted semi-Thue systems. In L. Bachmair, editor, *Proceedings of the 11th International Conference on Rewrite Techniques and Applications (RTA 2000), Norwich (UK)*, number 1833 in Lecture Notes in Computer Science, pages 172–186. Springer, 2000.

[125] M. Lohrey. Word problems for 2-homogeneous monoids and symmetric logspace. In J. Sgall, A. Pultr, and P. Kolman, editors, *Proceedings of the 26th International Symposium on Mathematical Foundations of Computer Science (MFCS 2001), Marianske Lazne (Czech Republic)*,

number 2136 in Lecture Notes in Computer Science, pages 500–511. Springer, 2001.

[126] R. C. Lyndon. On Dehn's algorithm. *Mathematical Annales*, 166:208–228, 1966.

[127] R. C. Lyndon and P. E. Schupp. *Combinatorial Group Theory.* Springer, 1977.

[128] K. Madlener and F. Otto. Groups presented by certain classes of finite length-reducing string-rewriting systems. In P. Lescanne, editor, *Proceedings of the 2nd International Conference on Rewrite Techniques and Applications (RTA 87), Bordeaux (France)*, number 256 in Lecture Notes in Computer Science, pages 133–144. Springer, 1987.

[129] K. Madlener and F. Otto. About the descriptive power of certain classes of finite string-rewriting systems. *Theoretical Computer Science*, 67(2-3):143–172, 1989.

[130] W. Magnus, A. Karrass, and D. Solitar. *Combinatorial Group Theory.* Wiley, 1966.

[131] G. S. Makanin. The problem of solvability of equations in a free semigroup. *Math. Sbornik*, 103:147–236, 1977. In Russian; English translation in *Math. USSR Sbornik 32, 1977.*

[132] G. S. Makanin. Equations in a free group. *Izv. Akad. Nauk SSR*, Ser. Math. 46:1199–1273, 1983. In Russian; English translation in *Math. USSR Izvestija 21, 1983.*

[133] G. S. Makanin. Decidability of the universal and positive theories of a free group. *Izv. Akad. Nauk SSSR*, Ser. Mat. 48:735–749, 1984. In Russian; English translation in *Math. USSR Izvestija, 25*, 75–88, 1985.

[134] S. S. Marchenkov. Unsolvability of the positive $\forall\exists$-theory of a free semi-group. *Sibirsky Matematicheskie Jurnal*, 23(1):196–198, 1982.

[135] A. Markov. On the impossibility of certain algorithms in the theory of associative systems. *Doklady Akademii Nauk SSSR*, 55, 58:587–590, 353–356, 1947.

[136] Y. V. Matiyasevich. *Hilbert's Tenth Problem*. MIT Press, Cambridge, Massachusetts, 1993.

[137] E. W. Mayr and A. R. Meyer. The complexity of the word problems for commutative semigroups and polynomial ideals. *Advances in Mathematics*, 46:305–329, 1982.

[138] A. Mazurkiewicz. Concurrent program schemes and their interpretations. DAIMI Rep. PB 78, Aarhus University, Aarhus, 1977.

[139] J. D. McKnight. Kleene quotient theorems. *Pacific Journal of Mathematics*, 14:1343–1352, 1964.

[140] R. McNaughton, P. Narendran, and F. Otto. Church-Rosser Thue systems and formal languages. *Journal of the Association for Computing Machinery*, 35(2):324–344, 1988.

[141] Y. I. Merzlyakov. Positive formulas on free groups. *Algebra i Logika Sem.*, 5(4):25–42, 1966. In Russian.

[142] A. R. Meyer. Weak monadic second order theory of one successor is not elementary recursive. In *Proceedings of the Logic Colloquium (Boston 1972–73)*, number 453 in Lecture Notes in Mathematics, pages 132–154. Springer, 1975.

[143] C. F. Miller III. Decision problems for groups – survey and reflections. In G. Baumslag and C. F. Miller III, editors, *Algorithms and classification in combinatorial group theory*, pages 1–59. Springer, 1992.

[144] D. E. Muller and P. E. Schupp. Groups, the theory of ends, and context-free languages. *Journal of Computer and System Sciences*, 26:295–310, 1983.

[145] D. E. Muller and P. E. Schupp. The theory of ends, pushdown automata, and second-order logic. *Theoretical Computer Science*, 37(1):51–75, 1985.

[146] A. Muscholl. *Decision and Complexity Issues on Concurrent Systems*. Habilitation thesis, Universität Stuttgart, 1999.

[147] A. Muscholl and D. Peled. Message sequence graphs and decision problems on Mazurkiewicz traces. In M. Kutylowski, L. Pacholski, and T. Wierzbicki, editors, *Proceedings of the 24th International Symposium on Mathematical Foundations of Computer Science (MFCS'99), Szklarska Poreba (Poland)*, number 1672 in Lecture Notes in Computer Science, pages 81–91. Springer, 1999.

[148] P. Narendran and F. Otto. Elements of finite order for finite weight-reducing and confluent Thue systems. *Acta Informatica*, 25:573–591, 1988.

[149] P. Narendran and F. Otto. Preperfectness is undecidable for Thue systems containing only length-reducing rules and a single commutation rule. *Information Processing Letters*, 29:125–130, 1988.

[150] P. Narendran and F. Otto. Some results on equational unification. In M. E. Stickel, editor, *Proceedings of the 10th International Conference on Automated Deduction (CADE 90), Kaiserslautern (Germany)*, number 449 in Lecture Notes in Computer Science, pages 276–291. Springer, 1990.

[151] M. H. A. Newman. On theories with a combinatorial definition of "equivalence". *Annals of Mathematics*, 43:223–243, 1943.

[152] J. Nielsen. Die Isomorphien der allgemeinen unendlichen Gruppe mit zwei Erzeugenden. *Mathematische Annalen*, 78:385–397, 1918. In German.

[153] N. Nisan and A. Ta-Shma. Symmetric logspace is closed under complement. *Chicago Journal of Theoretical Computer Science*, 1995.

[154] M. Nivat and M. Benois. Congruences parfaites et quasi–parfaites. *Seminaire Dubreil*, 25(7–01–09), 1971–1972.

[155] P. S. Novikov. On the algorithmic unsolvability of the word problem in group theory. *American Mathematical Society, Translations, II. Series*, 9:1–122, 1958.

[156] E. Ochmański. Regular behaviour of concurrent systems. *Bulletin of the European Association for Theoretical Computer Science (EATCS)*, 27:56–67, 1985.

[157] F. Otto, A. Sattler-Klein, and K. Madlener. Automatic monoids versus monoids with finite convergent presentations. In T. Nipkow, editor, *Proceedings of the 9th International Conference on Rewriting Techniques and Applications (RTA-98), Tsukuba (Japan)*, number 1379 in Lecture Notes in Computer Science, pages 32–46. Springer, 1998.

[158] F. Otto and L. Zhang. Decision problems for finite special string-rewriting systems that are confluent on some congruence class. *Acta Informatica*, 28:477–510, 1991.

[159] C. H. Papadimitriou. *Computational Complexity*. Addison Wesley, 1994.

[160] L. Pélecq. Automorphism groups of context-free graphs. *Theoretical Computer Science*, 165(2):275–293, 1996.

[161] N. Pippenger. On simultaneous resource bounds. In *Proceedings of the 20th Annual Symposium on Foundations of Computer Science (FOCS 79)*, pages 307–311. IEEE Computer Society Press, 1979.

[162] W. Plandowski. Satisfiability of word equations with constants is in PSPACE. In *Proceedings of the 40th Annual Symposium on Foundations of Computer Science (FOCS 99)*, pages 495–500. IEEE Computer Society Press, 1999.

[163] E. Post. Recursive unsolvability of a problem of Thue. *Journal of Symbolic Logic*, 12(1):1–11, 1947.

[164] M. Presburger. Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt. In *Comptes Rendus du Premier Congrès des Mathématicienes des Pays Slaves*, pages 92–101, Warsaw, 1929.

[165] W. V. O. Quine. Concatenation as a basis for arithmetic. *Journal of Symbol Logic*, 11(4):105–114, 1946.

[166] M. O. Rabin. Decidability of second-order theories and automata on infinite trees. *Transactions of the American Mathematical Society*, 141:1–35, 1969.

[167] N. N. Repin. Some simply presented groups for which an algorithm recognizing solvability of equations is impossible. *Voprosy Kibernetiki (Moskva)*, 134:167–175, 1988. In Russian.

[168] E. Rips and Z. Sela. Canonical representatives and equations in hyperbolic groups. *Inventiones Mathematicae*, 120:489–512, 1995.

[169] E. L. Robertson. Structure of complexity in the weak monadic second-order theories of the natural numbers. In *Proceedings of the 6th Annual Symposium on Theory of Computing (STOC 74)*, pages 161–171. ACM Press, 1974.

[170] N. Robertson and P. D. Seymour. Graph minors II. Algorithmic aspects of tree-width. *Journal of Algorithms*, 78:309–322, 1986.

[171] D. Robinson. *Parallel Algorithms for Group Word Problems.* PhD thesis, University of California, San Diego, 1993.

[172] J. J. Rotman. *An Introduction to the Theory of Groups (fourth edition).* Springer, 1995.

[173] B. V. Rozenblat. Positive theories of free inverse semigroups. *Siberian Mathematical Journal*, 20:910–918, 1980. English translation.

[174] B. V. Rozenblat. Diophantine theories of free inverse semigroups. *Siberian Mathematical Journal*, 26:860–865, 1985. English translation.

[175] W. L. Ruzzo. Tree–size bounded alternation. *Journal of Computer and System Sciences*, 21:218–235, 1980.

[176] W. L. Ruzzo. On uniform circuit complexity. *Journal of Computer and System Sciences*, 22:365–383, 1981.

[177] J. Sakarovitch. On regular trace languages. *Theoretical Computer Science*, 52:59–75, 1987.

[178] J. Sakarovitch. The "last" decision problem for rational trace languages. In I. Simon, editor, *Proceedings of the 1st Latin American Symposium on Theoretical Informatics (LATIN'92)*, number 583 in Lecture Notes in Computer Science, pages 460–473. Springer, 1992.

[179] M. Schaefer, E. Sedgwick, and D. Stefankovic. Recognizing string graphs in NP. In *Proceedings of the 34th Annual Symposium on Theory of Computing (STOC 2002)*, pages 1–6. ACM Press, 2002.

[180] K. U. Schulz. Makanin's algorithm for word equations — Two improvements and a generalization. In K. U. Schulz, editor, *Word Equations and Related Topics*, number 572 in Lecture Notes in Computer Science, pages 85–150. Springer, 1991.

[181] P. E. Schupp. Groups and graphs: Groups acting on trees, ends, and cancellation diagrams. *Mathematical Intelligencer*, 1:205–222, 1979.

[182] D. Seese. Tree-partite graphs and the complexity of algorithms. In L. Budach, editor, *Proceedings of Fundamentals of Computation Theory (FCT'85), Cottbus (GDR)*, number 199 in Lecture Notes in Computer Science, pages 412–421, 1985.

[183] D. Seese. The structure of models of decidable monadic theories of graphs. *Annals of Pure and Applied Logic*, 53:169–195, 1991.

[184] Z. Sela. Diophantine geometry over groups VIII: The elementary theory of a hyperbolic group. Available via `http://www.ma.huji.ac.il/zlil/`, 2002.

[185] A. L. Semenov. An interpretation of free algebras in free groups. *Soviet Math. Dokl.*, 21:952–955, 1980.

[186] A. L. Semenov. Decidability of monadic theories. In M. Chytil and V. Koubek, editors, *Proceedings of the 11th International Symposium of Mathematical Foundations of Computer Science (MFCS'84), Praha (Czechoslovakia)*, number 176 in Lecture Notes in Computer Science, pages 162–175. Springer, 1984.

[187] G. Sénizergues. An effective version of Stalling's theorem in the case of context-free groups. In A. Lingas, R. G. Karlsson, and S. Carlsson, editors, *Proceedings of the 20th International Colloquium on Automata, Languages and Programming (ICALP 93), Lund (Sweden)*, number 700 in Lecture Notes in Computer Science, pages 478–495. Springer, 1993.

[188] G. Sénizergues. Formal languages and word-rewriting. In H. Comon and J.-P. Jouannaud, editors, *Term Rewriting, French Spring School*

of Theoretical Computer Science, Font Romeux (France), number 909 in Lecture Notes in Computer Science, pages 75–94. Springer, 1993.

[189] G. Sénizergues. Semi-groups acting on context-free graphs. In F. M. auf der Heide and B. Monien, editors, *Proceedings of the 23th International Colloquium on Automata, Languages and Programming (ICALP'96), Paderborn (Germany)*, number 1099 in Lecture Notes in Computer Science, pages 206–218. Springer, 1996.

[190] J.-P. Serre. *Trees*. Springer, 1980.

[191] S. Shelah. The monadic theory of order. *Annals of Mathematics, II. Series*, 102:379–419, 1975.

[192] P. V. Silva and B. Steinberg. A geometric characterization of automatic monoids. Technical Report CMUP 2000-03, University of Porto, 2001.

[193] P. V. Silva and B. Steinberg. Extensions and submonoids of automatic monoids. *Theoretical Computer Science*, 289:727–754, 2002.

[194] H.-U. Simon. Word problems for groups and contextfree recognition. In *Proceedings of Fundamentals of Computation Theory (FCT'79), Berlin/Wendisch-Rietz (GDR)*, pages 417–422. Akademie-Verlag, 1979.

[195] L. B. Smikun. Über den Zusammenhang kontextfreier Gruppen und Gruppen mit lösbarem Problem der Äquivalenz von Automaten. *Kibernetika*, 5:33–37, 1976. In Russian.

[196] J. R. Stallings. *Group Theory and Three-Dimensional Manifolds*. Number 4 in Yale Mathematical Monographs. Yale University Press, 1971.

[197] I. A. Stewart. Complete problems for symmetric logspace involving free groups. *Information Processing Letters*, 40:263–267, 1991.

[198] I. A. Stewart. Refining known results on the generalized word problem for free groups. *International Journal of Algebra and Computation*, 2:221–236, 1992.

[199] I. A. Stewart and R. M. Thomas. Formal languages and the word problem for groups. In *Groups St Andrews*, number 261 in London Mathematical Society Lecture Notes Series, pages 689–700. Cambridge University Press, 1997.

[200] L. Stockmeyer. *The complexity of decision problems in automata and logic*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1974.

[201] J. Stupp. The lattice-model is recursive in the original model. The Hebrew University, Jerusalem, 1975.

[202] I. H. Sudborough. On the tape complexity of deterministic context–free languages. *Journal of the Association for Computing Machinery*, 25(3):405–414, 1978.

[203] W. Thomas. Ehrenfeucht games, the composition method, and the monadic theory of ordinal words. In J. Mycielski, G. Rozenberg, and A. Salomaa, editors, *Structures in Logic and Computer Science, A Selection of Essays in Honor of A. Ehrenfeucht*, number 1261 in Lecture Notes in Computer Science, pages 118–143. Springer, 1997.

[204] W. Thomas. Languages, automata, and logic. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages, volume III*, pages 389–455. Springer, 1997.

[205] W. Thomas. A short introduction to infinite automata. In W. Kuich, G. Rozenberg, and A. Salomaa, editors, *Proceedings of the 5th International Conference on Developments in Language Theory (DLT 2001), Vienna (Austria)*, number 2295 in Lecture Notes in Computer Science, pages 130–144. Springer, 2001.

[206] C. Thomassen. Configurations in graphs of large minimum degree, connectivity, or chromatic number. In *Proceedings of the 3rd International Conference on Combinatorial Mathematics*, number 555 in Annals of the New York Academy of Sciences, pages 402–412, 1989.

[207] C. Thomassen and W. Woess. Vertex-transitive graphs and accessibility. *Journal of Combinatorial Theory, Series B*, 58:248–268, 1993.

[208] A. Thue. Probleme über die Veränderungen von Zeichenreihen nach gegebenen Regeln. *Skr. Vid. Kristiania, I Math. Natuv. Klasse*, No. 10, 34 S., 1914. In German.

[209] B. A. Trakhtenbrot. Impossibility of an algorithm for the decision problem in finite classes. *American Mathematical Society, Translations, II. Series*, 23:1–5, 1950.

[210] Y. M. Vazhenin and B. V. Rozenblat. Decidability of the positive theory of a free countably generated semigroup. *Mathematics of USSR Sbornik.*, 44(1):109–116, 1983. English translation.

[211] A. Veloso da Costa. Graph products of monoids. *Semigroup Forum*, 63(2):247–277, 2001.

[212] A. Veloso da Costa. On graph products of automatic monoids. *R.A.I.R.O. — Informatique Théorique et Applications*, 35(5):403–417, 2001.

[213] H. Venkateswaran. Properties that characterize LOGCFL. *Journal of Computer and System Sciences*, 43:380–404, 1991.

[214] R. M. Verma, M. Rusinowitch, and D. Lugiez. Algorithms and reductions for rewriting problems. *Fundamenta Informaticae*, 46(3):257–276, 2001.

[215] H. Vollmer. *Introduction to Circuit Complexity.* Springer, 1999.

[216] J. von Zur Gathen and M. Sieveking. A bound on solutions of linear integer equalities and inequalities. *Proceedings of the American Mathematical Society*, 72(1):155–158, 1978.

[217] S. Waack. The parallel complexity of some constructions in combinatorial group theory. *Journal of Information Processing and Cybernetics EIK*, 26:265–281, 1990.

[218] S. Waack. On the parallel complexity of linear groups. *R.A.I.R.O. — Informatique Théorique et Applications*, 25(4):265–281, 1991.

[219] I. Walukiewicz. Monadic second-order logic on tree-like structures. *Theoretical Computer Science*, 275(1–2):311–346, 2002.

[220] W. Woess. Graphs and groups with tree-like properties. *Journal of Combinatorial Theory, Series B*, 47:361–371, 1989.